Forecasting Existential Risks

Evidence from a Long-Run Forecasting Tournament

Authors: Ezra Karger, Josh Rosenberg, Zachary Jacobs, Molly Hickman, Rose Hadshar, Kayla Gamin, Taylor Smith, Bridget Williams, Tegan McCaslin, Philip E. Tetlock

July 10, 2023 | FRI Working Paper #1

Forecasting Research Institute

Forecasting Existential Risks:

Evidence from a Long-Run Forecasting Tournament[1]

Authors: Ezra Karger,[†,*] Josh Rosenberg,[*] Zachary Jacobs,[*] Molly Hickman,[*] Rose Hadshar,[*] Kayla Gamin,[*] Taylor Smith,[*] Bridget Williams,[*] Tegan McCaslin,[*] Philip E. Tetlock[‡,*]

Abstract:

The Existential Risk Persuasion Tournament (XPT) aimed to produce high-quality forecasts of the risks facing humanity over the next century by incentivizing thoughtful forecasts, explanations, persuasion, and updating from 169 forecasters over a multi-stage tournament. In this first iteration of the XPT, we discover points where historically accurate forecasters on short-run questions (superforecasters) and domain experts agree and disagree in their probability estimates of short-, medium-, and long-run threats to humanity from artificial intelligence, nuclear war, biological pathogens, and other causes. We document large-scale disagreement and minimal convergence of beliefs over the course of the XPT, with the largest disagreement about risks from artificial intelligence. The most pressing practical question for future work is: why were superforecasters so unmoved by experts' much higher estimates of AI extinction risk, and why were experts so unmoved by the superforecasters' lower estimates? The most puzzling scientific question is: why did rational forecasters, incentivized by the XPT to persuade each other, not converge after months of debate and the exchange of millions of words and thousands of forecasts?

---

1 This research would not have been possible without the generous support of the Musk Foundation and the Long-Term Future Fund. We thank Walter Frick, Michael Page, Terry Murray, David Budescu, Barb Mellers, Elie Hassenfeld, Philipp Schoenegger, and Pavel Atanasov for their thoughtful feedback and comments. We are also grateful for the many people who reviewed earlier drafts of this report. We greatly appreciate the assistance of Amory Bennett, Kaitlyn Coffee, Adam Kuzee, Avital Morris, Fiona Pollack, Coralie Consigny, Arunim Agarwal, and Rumtin Sepasspour throughout the project.

† Federal Reserve Bank of Chicago

* Forecasting Research Institute

‡ Wharton School of the University of Pennsylvania

Table of contents

1. Executive summary

What are the gravest risks facing humanity over the next century and beyond? This question has received considerable attention in recent years, with experts warning about dangers across a range of sources, from pandemics to nuclear war to artificial intelligence (AI). Some have argued more broadly that the coming century may be uniquely risky, with a higher chance of human extinction than at any point in human history.

How can we evaluate these risks? Pundits and researchers have incentives to make bold claims that attract audiences and funding—and to keep their predictions vague enough so they can never be proven wrong. On the other hand, people without expertise often make bold predictive claims with incorrect or missing knowledge about key mechanisms underlying complex topics. We believe forecasting research can reduce these perverse incentives by rewarding accurate predictions on short-run indicators that illuminate longer-run risks. Historically accurate forecasters (superforecasters) have a proven track record at forecasting resolvable questions over short time horizons but may have less expertise forecasting in technical domains, and do not have a track record in forecasting questions over long time horizons. By bringing researchers and superforecasters together to forecast on short-run predictors and long-run measures of risk, we can incentivize all groups to clarify their positions, complement each other's skills, and highlight the cruxes of important disagreements.

The Existential-Risk Persuasion Tournament (XPT) ran from June through October 2022 and brought together 80 specialists on long-run existential risks to humanity ("experts") and 89 seasoned forecasters with a track record of predictive accuracy on shorter-run questions ("superforecasters")—two groups with distinctive claims to knowledge about humanity's future. We asked the two groups to work together to predict the likelihood of global risks such as nuclear weapon use, pandemics, and AI. The XPT is the first long-range forecasting tournament of its kind, and adds value relative to similar research by (1) linking long-term forecasts to short-term forecasts that can be assessed for accuracy; (2) offering bonus payments for persuasive rationales and accurate "intersubjective" forecasts (predictions of the views of other participants), which we are testing as early indicators of the reliability of long-range forecasts; and (3) including both experts and superforecasters.

This report outlines the main results of the XPT, including forecasts of catastrophic and extinction risk, and describes how those forecasts differ between experts and superforecasters.

Key takeaways

1. The median expert predicted a 20% chance of catastrophe and a 6% chance of human extinction by 2100.[2] Superforecasters saw the chances of both catastrophe and extinction as considerably lower than did experts. The median superforecaster predicted a 9% chance of catastrophe and a 1% chance of extinction.
2. The gap between forecasts of experts and superforecasters was not uniform across topics. Experts and superforecasters were furthest apart on AI risk—and less so on the risk of nuclear war.
3. Participants in a parallel public survey of college graduates estimated a higher median probability of extinction by 2100 (5%) than superforecasters (1%) but lower than that of experts (6%). A similar pattern also emerged for AI-caused extinction: public survey participants gave a 2% probability, while superforecasters and domain experts gave forecasts an order of magnitude apart at the median: 0.38% and 3%, respectively.
4. Among both superforecasters and experts, those with higher "intersubjective accuracy"—those best at predicting the views of other participants—estimated lower probabilities of catastrophic and extinction risks from all sources.
5. Predictions about risk were highly correlated across topics. For example, participants more concerned about AI are also more concerned about pandemics and nuclear weapon use.
6. Few minds were changed during the XPT, even among the most active participants, and despite monetary incentives for persuading others.
7. Although the biggest area of long-run disagreement was the probability of extinction due to AI, there were surprisingly high levels of agreement on 45 shorter-run indicators when comparing forecasters most and least concerned about AI risk. We compare the top third and bottom-third most and least concerned forecasters about AI risk by 2100. Between these groups, a few examples of the largest disagreements in forecasts resolving by 2030 were: whether artificial general intelligence will exist according to experts (AI-concerned group median: 6.6%; AI skeptic group median: 3.8%), how much money will be spent on computational resources for the largest AI experiments (AI-concerned median: $156 million; AI skeptic: $100 million), and whether nuclear weapons will cause the deaths of at least 1,000 people (AI-concerned median: 6.7%; AI skeptic: 4.7%).

2 We define a catastrophic event as one causing the death of at least 10% of humans alive at the beginning of a five-year period and define extinction as reduction of the global population to less than 5000. The "median expert" here refers to the median of all expert forecasts on the given question. So, the expert whose forecast reflects the median forecast on total extinction risk is not necessarily the expert whose forecast reflects the median forecast on total catastrophic risk. We are simply reporting the median forecast on each question.

Abbreviated results

Forecast	Median Estimate (95% confidence interval)[3]	
	Superforecasters	**Experts**
Total Catastrophic Risk by 2100	9.05% [6.13, 10.25]%	20% [15.44, 27.60]%
Total Extinction Risk by 2100	1% [0.55, 1.23]%	6% [3.41, 10.00]%

Table 1: *Median final forecasts from the XPT on questions of whether humanity will experience large-scale death. "Catastrophic risk" is defined as 10% or more of humans dying within a 5-year period. We calculate medians from N=89 superforecasters and N=80 experts with expertise in AI, climate, nuclear, and biorisk domains. We also present bootstrapped confidence intervals for each median.*

(For a complete description of the final forecasts from the XPT, see the "Results" section.)

In the following sections, we describe the research design of the XPT, its key results, and next steps for this research agenda.

3 In this report, when we present median forecasts, we present them with the same precision the median forecaster provided (e.g. 1% as opposed to 1.00%), up to two decimal places or two significant digits, whichever is more precise (e.g. 0.0123% becomes 0.012%, 1.23% stays 1.23%). For confidence intervals, we add trailing zeroes if one end of the interval has more implied precision than the other.

2. Background & motivation

Warnings about humanity's future make for good headlines, and plenty of them have attracted media attention in recent years. Consider just a few recent claims about the risks faced by humanity:

- The Bulletin of Atomic Scientists claims the world is closer to nuclear catastrophe than at any time in its history.[4]
- A 2023 letter signed by the CEOs of three major AI labs, key AI scientists, and notable people including Bill Gates states that "Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war."[5]
- Political scientist Graham Allison claims that the U.S. risks falling into a "Thucydides Trap" in which its rivalry with China leads to conflict and catastrophe.[6]
- Journalist David Wallace-Wells claims that without aggressive action by policymakers, climate change could render the planet "uninhabitable."[7]
- Former CIA director Michael Morell argues that governments underinvest in addressing certain existential threats, saying that a key existential threat to America is "a naturally occurring or manmade biological agent that kills 60 to 70% of the population…" He then says, "We don't spend enough time on that."[8]

These claims cover diverse topics and come from diverse sources, but they all have two things in common: the extreme gravity of the threats they allege and their power to attract media attention and prompt public debate. But are these the right risks to focus on?

---

4 "Doomsday Clock," *Bulletin of the Atomic Scientists* (blog), accessed June 23, 2023, https://thebulletin.org/doomsday-clock/.

5 "Statement on AI Risk," Center for AI Safety, accessed June 23, 2023, https://www.safe.ai/statement-on-ai-risk.

6 "Thucydides's Trap: An Overview," Belfer Center for Science and International Affairs, accessed June 23, 2023, https://www.belfercenter.org/thucydides-trap/overview-thucydides-trap.

7 David Wallace-Wells, "When Will the Planet Be Too Hot for Humans? Much, Much Sooner Than You Imagine," *Intelligencer*, July 9, 2017, https://nymag.com/intelligencer/2017/07/climate-change-earth-too-hot-for-humans.html.

8 Tom Embury-Dennis, "Former CIA Chief Just Warned of the Three Existential Threats to America," *The Independent*, October 27, 2017, https://www.independent.co.uk/news/world/americas/us-russia-nuclear-war-climate-change-biological-weapons-attack-only-existential-threats-cia-chief-michael-morell-a8022941.html.

It is hard enough to evaluate claims by pundits or experts under normal conditions—they often avoid making falsifiable predictions and are later able to rationalize why they were prescient, no matter what happens. More charitably, it is difficult for anyone, including pundits and experts, to make crisp predictions in conversations and public remarks. Arguments about existential risk face an additional hurdle: they often lack historical comparison classes. So, how can we better equip policymakers to weigh and respond to the gravest risks that our species faces? Quantitative forecasting can help by pushing experts and pundits to clarify their positions by highlighting disagreements between them and others, and by tracking the accuracy of their predictions over time.

What forecasting research can add

Over the past 20 years, research on forecasting has demonstrated the feasibility of tracking predictive accuracy in real-world debates. The first generation of large-scale forecasting tournaments asked experts and amateurs alike to make probabilistic forecasts on short-run geopolitical and economic questions. This research identified both the reasoning styles associated with predictive accuracy and a number of individuals skilled at putting them to use—the "superforecasters" who outperformed experts and intelligence analysts in forecasting tournaments held by the Good Judgment Project.[9] Today, these tournaments are proliferating and increasingly cited by journalists.[10] (See: Appendix 2: A brief history of forecasting research.)

In contrast to these short-term forecasts, claims about existential risk—like the belief that nuclear war or AI could lead to catastrophic outcomes—aren't as easily evaluated. Forecasts are typically scored based on what actually happens, but scoring long-run claims would take decades or even centuries—and for existential risks, the results could arrive too late to do any good. Can probabilistic forecasting bring some of the rigor and transparency it provides in other contexts to questions of long-run risks? We think so.

---

9 Philip E. Tetlock et al., "Forecasting Tournaments: Tools for Increasing Transparency and Improving the Quality of Debate," Current Directions in *Psychological Science* 23, no. 4 (August 2014). https://doi.org/10.1177/096372141453425.

10 Walter Frick, "Journalists Wake up to the Power of Prediction Markets," *Nieman Lab* (blog), accessed June 23, 2023, https://www.niemanlab.org/2022/12/journalists-wake-up-to-the-power-of-prediction-markets/.

A new approach to forecasting extinction risks

This report describes early results from a first-of-its-kind forecasting tournament—the XPT—centered on risks to humanity over the next century. We gathered 89 superforecasters and 80 experts on nuclear war, climate change, AI, biological risks, and existential risk more broadly. We tasked them to work independently and then on teams to submit forecasts on 59 questions.

The shortest-run questions ask about the world at the end of 2024, and on these questions we will know the actual outcomes soon. Other questions ask about the world in 2030, 2050, and 2100. Some might take centuries to resolve—such as "By what year will humans go extinct?"[11] But the tournament's novel setup offers three benefits relative to prior work:

1. We can pinpoint where historically highly accurate forecasters ("superforecasters") and domain experts agree and disagree in evaluating these risks.
2. We directly incentivize high-quality and persuasive rationales, so we can describe the explanations that participants gave and identify the underlying reasons for agreement and disagreement.
3. For longer-term questions, we incentivized participants using a metric called "reciprocal scoring," an intersubjective way of scoring predictions before they can be objectively resolved.[12] It asks participants to predict each other's predictions. These reciprocal forecasts are often remarkably accurate in their own right on short-run forecasts, so we view them as a promising way to incentivize longer-run forecasts. In previous work, we show that forecasts elicited this way can be as accurate as forecasts incentivized using comparisons to the truth.[13] An additional benefit in the context of long-run forecasting is that intersubjective forecasts can be scored immediately. (See the "Reciprocal Scoring" section of Appendix 3.)

These features have already yielded new insights into debates over long-term risks. We now know, for instance, that superforecasters assign much lower chances of humanity going extinct in the next 100 years than do experts who study risks to humanity. Also, domain experts and superforecasters have thus far been unable to change each other's minds—despite the fact that the XPT included incentives and processes designed to encourage belief updating.

11 The exact question wording was: "By what year will humans go extinct or first have a population less than 5,000?"

12 Ezra Karger, Pavel D. Atanasov, and Philip Tetlock, "Improving Judgments of Existential Risk: Better Forecasts, Questions, Explanations, Policies," SSRN Working Paper (2022). https://doi.org/10.2139/ssrn.4001628.

13 Karger et al., "Reciprocal Scoring."

3. How the XPT works

Nearly 200 participants signed up to participate in the XPT: a four-month exercise run from June through October 2022—and 169 submitted forecasts. The main research sample described in this report consists of 89 superforecasters who scored highly in previous tournaments and 80 experts in different areas of existential risk, with 111 completing all stages of the tournament.[14]

To recruit experts, we contacted organizations working on existential risk, relevant academic departments, and research labs at major universities and within companies operating in these spaces. We also advertised broadly, reaching participants with relevant experience via blogs and Twitter. We received hundreds of expressions of interest in participating in the tournament, and we screened these respondents for expertise, offering slots to respondents with the most expertise after a review of their backgrounds.[15] We selected 80 experts to participate in the tournament. Our final expert sample (N=80) included 32 AI experts, 15 "general" experts studying long-run risks to humanity, 12 biorisk experts, 12 nuclear experts, and 9 climate experts, categorized by the same independent analysts who selected participants. Our expert sample included well-published AI researchers from top-ranked industrial and academic research labs, graduate students with backgrounds in synthetic biology, and generalist existential risk researchers working at think tanks, among others. According to a self-reported survey, 44% of experts spent more than 200 hours working directly on causes related to existential risk in the previous year, compared to 11% of superforecasters. The sample drew heavily from the Effective Altruism (EA) community: about 42% of experts and 9% of superforecasters reported that they had attended an EA meetup. In this report, we separately present forecasts from domain experts and non-domain experts on each question.

To recruit superforecasters, we worked with Good Judgment Inc. (GJI),[16] a company built on earlier academic work by Mellers et al.[17] that provides paid forecasting services to clients. We recruited superforecasters, most of whom were labeled via that earlier

14 Of these 111 forecasters who completed all four stages of the tournament, 72 were superforecasters and 39 were experts. Although 111 completed all stages of the tournament, we report data from forecasters who attrited from the tournament in relevant analyses below.

15 Two independent analysts categorized applicants based on publication records and work history. When the analysts disagreed, a third independent rater resolved disagreement after a group discussion.

16 See https://goodjudgment.com/.

17 Barbara Mellers et al., "Identifying and Cultivating Superforecasters as a Method of Improving Probabilistic Predictions," *Perspectives on Psychological Science* 10, No. 3, (2015) 267–281. https://doi.org/10.1177/1745691615577794.

academic work. A minority of these superforecasters were not labeled as such via that work, but were instead top performers in subsequent short-run forecasting tournaments run by GJI. While superforecasters are a set of forecasters with high levels of accuracy on short-run (0–2 year timespan) resolvable questions, it is an open question whether forecasters who are accurate on short-run questions will also be accurate on longer-run questions. We do not present new evidence to address that question here. But, we plan to examine that question using data from this study, as our questions begin to resolve in 2024 and 2030.

Participants moved through a four-stage deliberative process that allowed us to monitor the evolution of their forecasts and explanations.[18] This multi-stage process, described below in more detail, was designed to incentivize forecasters not only to make accurate predictions but also to provide persuasive rationales that boosted the predictive accuracy of others' forecasts.

In the first stage, participants worked individually to submit forecasts and rationales on 14 required and 45 optional short-run and long-run questions on global risks. We asked about AI, pandemics (both natural and engineered), nuclear weapons, and non-anthropogenic risks like volcanoes or asteroids. For each risk area, we asked about both the probability of human extinction and of a catastrophic event that kills at least 10% of the population.[19]

Here are examples of questions that forecasters tackled in the XPT:

- What will be the global surface temperature change as compared to 1850–1900, in degrees Celsius? (By 2030, 2050, 2100)
- By what year will fusion reactors deliver 1% of all utility-scale power consumed in the U.S.?
- How much will be spent on compute [computational resources] in the largest AI experiment? (By 2024, 2030, 2050)
- What is the probability that artificial intelligence will be the cause of death, within a 5-year period, for more than 10% of humans alive at the beginning of that period? (By 2030, 2050, 2100)
- What is the overall probability of human extinction or a reduction in the global population below 5,000? (By 2030, 2050, 2100)

---

18 Participants also consented to participate in this study, via the University of Pennsylvania's Institutional Review Board. The consent form detailed the format of the study.

19 We define a catastrophic event as one causing the death of at least 10% of humans alive at the beginning of a five-year period. We define extinction as reduction of the global population to less than 5,000.

For the longer-run questions resolving after 2030, we asked participants to not only submit a forecast but to also predict what the median forecast would be on that question among both experts and superforecasters (separately).[20] The long-run questions were required, and nearly all participants made a forecast for each of them. The questions resolving sooner were not, so fewer participants submitted forecasts on each short-run question. When we report probabilities of long-run catastrophic and existential risk in this report, we report forecasters' own (unincentivized) beliefs. But, we rely on the incentivized forecasts to calculate measures of intersubjective accuracy.

We informed all forecasters of the scoring ground rules: questions resolving by 2030 were scored using traditional forecasting metrics where the goal was to minimize the gap between probability judgments and reality (coded as zero or one as a function of the outcome). However, for the longer-run questions, participants learned that they would be scored based on the accuracy of their reciprocal forecasts: the better they predicted what experts and superforecasters would predict for each question, the better their score. (For more detail, see Appendix 3.) Almost all participants answered the key long-run questions about existential and catastrophic risk. For questions about shorter-run predictors of those risks and mechanisms underlying them, we asked a smaller, random subset of participants to respond, while all other participants had the option to answer.

One common challenge in forecasting tournaments is to uncover the reasoning behind predictions. To that end, the XPT offered ten $1,000 prizes for the most persuasive individual forecast rationales that enhanced others' forecasting accuracy. We explained that after the tournament we would show the highest-quality anonymized rationales (curated by independent readers) to panels of online survey participants who would make forecasts before and after reading the rationale. Prizes go to those whose rationales helped citizens update their forecasts toward greater accuracy, using both proper scoring rules for resolvable questions and intersubjective accuracy for unresolvable questions.[21]

20 We asked the core required questions about existential and catastrophic risks over 2030, 2050, and 2100 time horizons. Over each of these time horizons, including 2030, we asked participants to submit their own forecast and their forecasts of others' beliefs. On all other questions resolving by 2030, forecasters submitted only one (properly incentivized) forecast.

21 As of publication of this report, we are in the process of calculating these prizes, but we have already allocated to participants separate prizes for high-quality comments and high-value teammates.

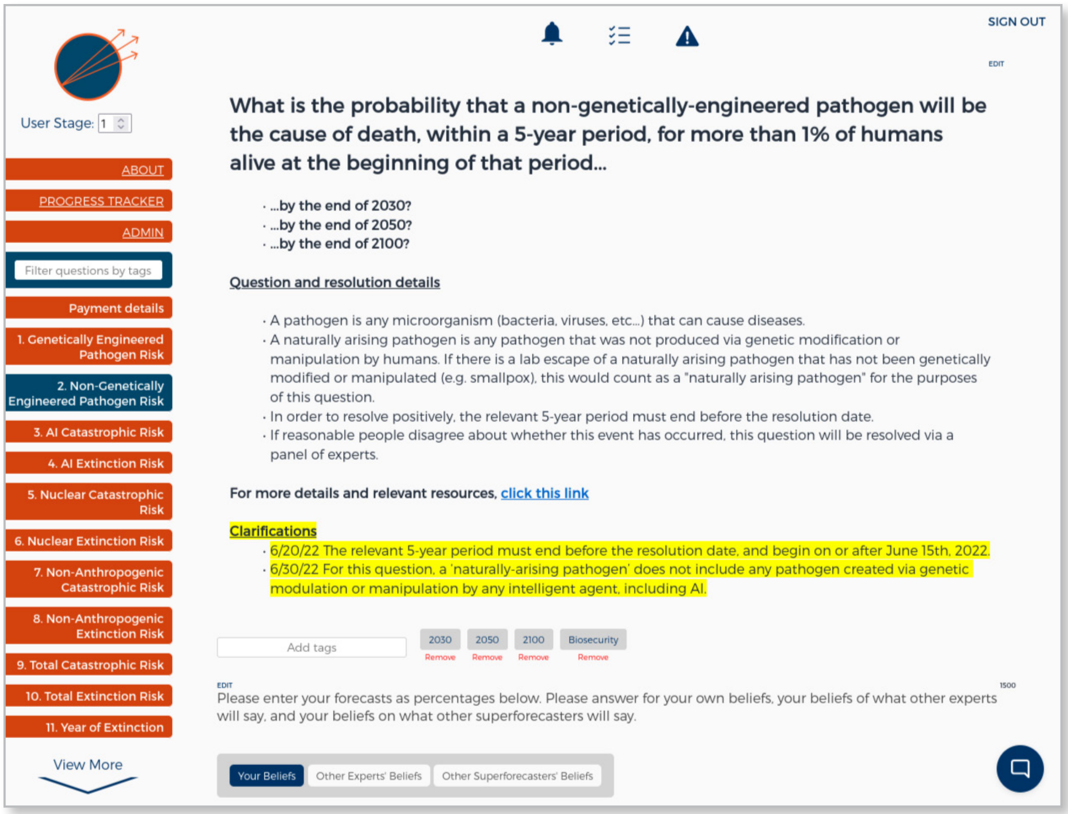


Figure 1: *An image of the forecasting platform used in the XPT, showing the question, "What is the probability that a non-genetically-engineered pathogen will be the cause of death within a 5-year period for more than 1% of humans alive at the beginning of that period," resolution details, and space to forecast the question for the periods ending in 2030, 2050, and 2100.*

In the second stage, participants worked in teams of 16. They could see each other's forecasts and rationales and then update their own forecasts based on what they learned from their teammates.[22] Each team consisted of either entirely superforecasters or entirely experts. Teams of experts were block-randomized to contain a roughly equal proportion of subject-matter experts from each domain of the tournament (AI, biorisk, nuclear war, and climate). Forecasters could see their own and teammates' forecasts and rationales from the first stage and were encouraged to discuss differences of opinion[23] and to update their forecasts if their beliefs changed. We offered an additional ten $1,000 prizes for the highest-quality comments forecasters offered in response to teammates' rationales and comments.[24]

---

22 Forecasters used usernames, and could choose to be anonymous or not. Most forecasters used their real name as their username or otherwise referred to their identity in their forecasts, although that choice was certainly not universal.

23 Most discussion occurred asynchronously on the platform, but some teams coordinated phone or zoom calls.

24 Prize-winning comments were selected based on votes from the participants, who were asked to select the three most helpful comments that their teammates had posted throughout the tournament.

In the third stage, superforecasters and experts were blended into new integrated teams of 32 forecasters and encouraged to deliberate and update their forecasts using the same scoring rules as in the first stage.[25] Each integrated team was asked to combine their rationales into a team wiki[26] describing the strongest argument for the team's median forecast, along with rationales for lower- and higher-end forecasts that they considered plausible. They were also asked to describe key areas of disagreement and uncertainty. (Each team received $100 for each question, awarded according to how much each member contributed to the wiki.)

In the fourth stage, each team was shown a randomly selected wiki from another team and asked once more to reconsider their forecasts.

In summary, the stages of the XPT were:

1. Participants made individual forecasts
2. Teams comprised entirely of either superforecasters or experts deliberated and updated their forecasts
3. Blended teams from the second stage, consisting of one superforecaster team and one expert team, deliberated and updated their forecasts
4. Each team saw one wiki summarizing the thinking of another team and again updated their forecasts

After the conclusion of the XPT, participants filled out a postmortem survey asking about their experience in the tournament, their views on global risks, and where they would allocate funding to address those risks. The median participant who completed the tournament earned $2,500 in incentives, but this figure is expected to rise as questions resolve in the coming years.

Finally, we surveyed a wider segment of the public about their views on existential risks. The survey included 912 college graduates, largely from the UK and US, who had engaged in previous forecasting-related surveys and were selected using the survey platform "Prolific." While the respondents are not a representative sample of a population of interest, the survey helps place the views of the superforecasters and existential risk experts in the context of wider public opinion. We gave these public respondents a subset of the 59 questions from the tournament, including all long-run questions.

25 Each team of 32 forecasters consisted of one team of 16 superforecasters from stage 2 and one team of 16 experts from stage 2.

26 "Wikis" refer to collaborative documents created by each team during the study. A team's wiki served as a platform for team members to crowdsource the collective rationale behind their forecast.

We encountered a variety of implementation challenges throughout the project—for more details, see Appendix 1.

Comparing experts to superforecasters

The XPT compares existential risk forecasts of two groups with potentially complementary strengths. Experts bring critical skills to debates: they understand technical details, recall past controversies, and are tuned in to the latest developments. However, when experts don't get regular feedback on their accuracy, they may struggle to translate their causal knowledge into probabilistic predictions and can fail to outperform educated generalists or simple algorithms in forecasting tournaments.[27] Recent research has also found that the top performers among educated generalists consistently produce highly accurate forecasts on a broad swath of geopolitical topics.[28] These individuals—"superforecasters"—can be identified based on their forecasting performance and tend to possess their own distinctive capabilities: they are open-minded, they think probabilistically, they are comfortable with data, and they blend multiple perspectives together when making their predictions. (For more on this research, see Appendix 2: A brief history of forecasting research.)

The superforecasters provide an important complement to the experts' assessments: if existential risks are anything like the shorter-run geopolitical forecasting questions that were the focus of past studies, the superforecasters could be a better guide to what will actually happen. On the other hand, it is possible that existential risks are not like other short-run topics; forecasting the prospect of extreme change over longer time frames might require different skills. For example, seasoned forecasters often rely on "base rates"—data on similar cases—which might be misleading or unavailable in the case of long-run risks. Maybe in such cases the experts bring critical knowledge or styles of thinking that the superforecasters don't have.

Comparisons of these two groups—each with their own claim to knowledge—will provide unique evidence of how they think about the most important threats to humanity. And, as we explain next, we found significant differences in how they rate those threats.

27 Philip E. Tetlock, E*xpert Political Judgment: How Good Is It? How Can We Know?* (Princeton: Princeton University Press, 2017), https://catalog.lib.uchicago.edu/vufind/Record/11327005.
28 Philip E. Tetlock and Dan Gardner. *Superforecasting: The art and science of prediction*. (New York: Crown, 2016).

4. Results

Superforecaster and expert estimates of catastrophic risk

What are the biggest risks to humanity over the next century and how likely are they to occur? The tables below provide the median forecast of XPT participants, by subgroup (superforecasters, domain experts, non-domain experts, general x-risk experts, and participants in the public survey), across the major risk areas covered in the tournament. The five subgroups are distinct: we categorize experts as either being experts in a specific domain, or absent that, being general x-risk experts if they study questions surrounding existential and catastrophic risks across domains.

Table 2 presents forecasts of catastrophic risks. Table 3 presents forecasts of extinction risks. Nuclear weapons are seen as the most likely cause of a catastrophic event; AI is seen as the most likely source of human extinction.

> Superforecasters estimated a 1% chance of human extinction by 2100, while experts estimated a 6% chance.[29]

29 See footnote 19 for how we defined these terms.

Forecast	Median Estimate (95% confidence interval)				
	Superfore-casters	Domain experts	Non-domain experts	General x-risk experts	Public survey
AI catastrophe	2.13% [1.83, 3.00]%	12% [4.0, 18.5]%	6.16% [5, 12]%	10% [6.16, 16.12]%	5% [5, 5]%[30]
Engineered pathogen catastrophe[31]	0.8% [0.5, 1.0]%	3% [1, 5]%	2.5% [1.5, 5.0]%	5% [3.03, 10.00]%	-
Natural pathogen catastrophe[32]	1% [0.7, 1.3]%	0.85% [0.5, 2.0]%	1.5% [1, 3]%	1.9% [0.2, 2.5]%	-
Nuclear catastrophe	4% [3, 5]%	8% [5, 11]%	8% [7, 10]%	7.24% [4.9, 10]%	10% [10, 11]%
Non-anthropogenic catastrophe[33]	0.05% [0.033, 0.081]%	0.09% [0.05, 0.11]%		0.045% [0.01, 1.00]%	2% [1, 3]%
Total catastrophic risk	9.05% [6.13, 10.25]%	20% [15.44, 27.60]%		28.95% [18.70, 50.63]%	11.56% [10, 13]%

Table 2: *Median forecasts from the XPT on questions of whether AI, pathogens, nuclear war, or non-anthropogenic risks will, by 2100, cause a "catastrophe" (i.e., be the cause of death within a 5-year period for more than 10% of humans alive at the beginning of that period). We calculate medians from N=88 superforecasters, N=66 domain experts with expertise in AI, nuclear, and biorisk domains, and N=14 general x-risk experts. Causes are not necessarily mutually exclusive; our resolution criteria allow that if an AI system uses nuclear weapons to cause a catastrophe in a manner that counterfactually requires both technologies, it would count as both AI- and nuclear-caused catastrophe for the purposes of these forecasts. We also present bootstrapped confidence intervals for each median.*

30 While this confidence interval may appear incorrect, so many people in the public submitted forecasts of 5% that the median is stably 5% across bootstrapped draws from the sample.

31 Because of concerns among our funders about information hazards (See Nick Bostrom, "Information hazards: A typology of potential harms from knowledge," *Review of Contemporary Philosophy* 10 (2011): 44-79, nickbostrom.com/information-hazards.pdf), we did not include this question in the main tournament, but we did ask about risks from engineered and natural pathogens in a one-shot separate postmortem survey to which most XPT participants responded after the tournament. We report those numbers here.

32 See footnote 31.

33 The resolution details for this question explicitly exclude non-anthropogenic pathogens.

Forecast	Median Estimate (95% confidence interval)				
	Superforecast-ers	Domain experts	Non-domain experts	General x-risk experts	Public survey
AI extinction	0.38% [0.10, 0.75]%	3% [0.49, 10.00]%	2% [1.00, 4.03]%	4.75% [1.9, 14.0]%	2% [1, 2]%
Engineered pathogen extinction[34]	0.01% [0.005, 0.052]%	1% [0.12, 1.09]%	0.1% [0.05, 0.30]%	1% [0.12, 1.09]%	-
Natural pathogen extinction[35]	0.0018% [0.001, 0.030]%	0.01% [0.0005, 0.0200]%	0.008% [0.0001, 0.0640]%	0.001% [0.0001, 0.2000]%	-
Nuclear extinction	0.074% [0.025, 0.100]%	0.55% [0.075, 1.400]%	0.19% [0.073, 0.500]%	0.7% [0.016, 1.000]%	2% [1.5, 4]%
Non-anthropogenic extinction[36]	0.0043% [0.0020, 0.0067]%	0.004% [0.0017, 0.0072]%		0.0059% [0.0010, 0.0095]%	1% [0.5, 1.0]%
Total extinction risk	1% [0.55, 1.23]%	6% [3.41, 10.00]%		6.6% [3.001, 13.670]%	5% [3, 5]%

Table 3: *Median forecasts from the XPT on questions of whether AI, pathogens, nuclear war, or non-anthropogenic risks will, by 2100, cause humanity to go extinct. We calculate medians from N=88 superforecasters, N=66 domain experts with expertise in AI, nuclear, and biorisk domains, and N=14 general x-risk experts. Causes are not necessarily mutually exclusive; our resolution criteria allow that if an AI system uses nuclear weapons to cause human extinction in a manner that counterfactually requires both technologies, it would count as both AI- and nuclear-caused human extinction for the purposes of these forecasts. We also present bootstrapped confidence intervals for each median.*

Superforecasters put the overall chance of human extinction by 2100 considerably lower than did experts (1% vs. 6%, respectively). Superforecasters put lower chances of catastrophe and extinction from virtually all of the risk areas studied in the XPT. There are two exceptions: non-anthropogenic risks, where the two groups provide virtually identical forecasts; and catastrophic risk from natural pathogens.

To visualize the large variation in beliefs among the subgroups, Figure 2 shows each forecast of total extinction risk, from an XPT forecaster and from our parallel public survey, with box plots indicating ranges containing the 25th and 75th percentile of forecasters in each group. Although the median forecast of each group is precisely estimated (see confidence intervals in Table 3), we see significant overlap in the distribution of forecasts, even among the groups that assign the lowest risk (superforecasters) and highest risk (general x-risk experts).

34 See footnote 31.
35 See footnote 31.
36 The resolution details for this question explicitly exclude non-anthropogenic pathogens.

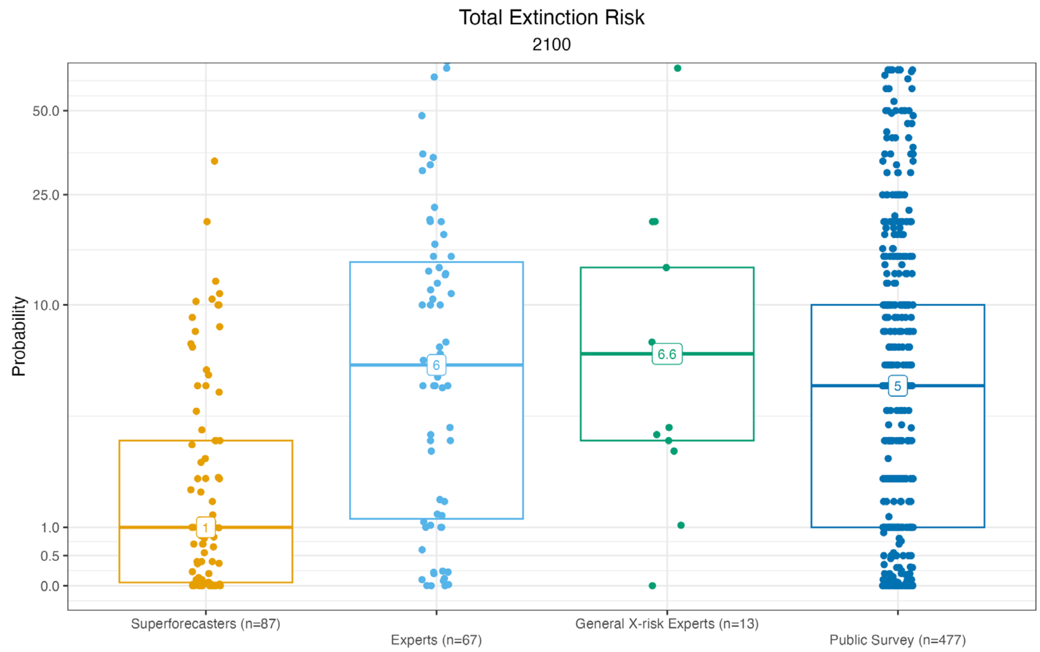


Figure 2: *Plot shows the distribution of final forecasts of extinction risk (by 2100) for each of three groups in the XPT (superforecasters, domain experts, and general x-risk experts) as well as a public survey of college-educated respondents from an online platform. Box plots show the 25–75th percentile range (boxes) and the median (labeled) for each group. Within each group, points are jittered slightly horizontally to show density, but this jittering has no other empirical purpose.*[37]

The same pattern holds for forecasts of total catastrophic risk (shown in Figure 3) and forecasts of AI extinction risk (Figure 4).

37 The y-axes on box plots in this report have been truncated above at the 95th percentile of tournament participants + a 5% cushion. This means that we exclude some high outlying forecasts from at most (but usually many less than) 5% of tournament participants. This also excludes a portion of public survey respondents (those whose forecasts are above the 95th percentile of XPT participant forecasts).

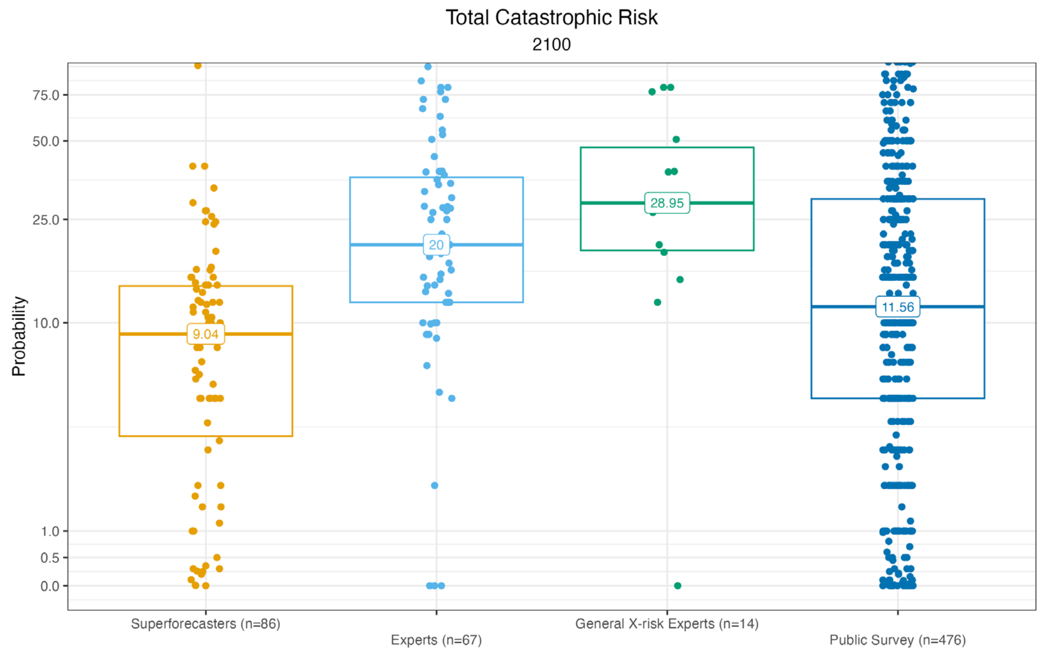


Figure 3: *Plot shows the distribution of final forecasts of catastrophic risk (by 2100) for each of three groups in the XPT (superforecasters, domain experts, and general x-risk experts) as well as a public survey of college-educated respondents from an online platform. Box plots show the 25–75th percentile range (boxes) and the median (labeled) for each group. Within each group, points are jittered slightly horizontally to show density, but this jittering has no other empirical purpose.*

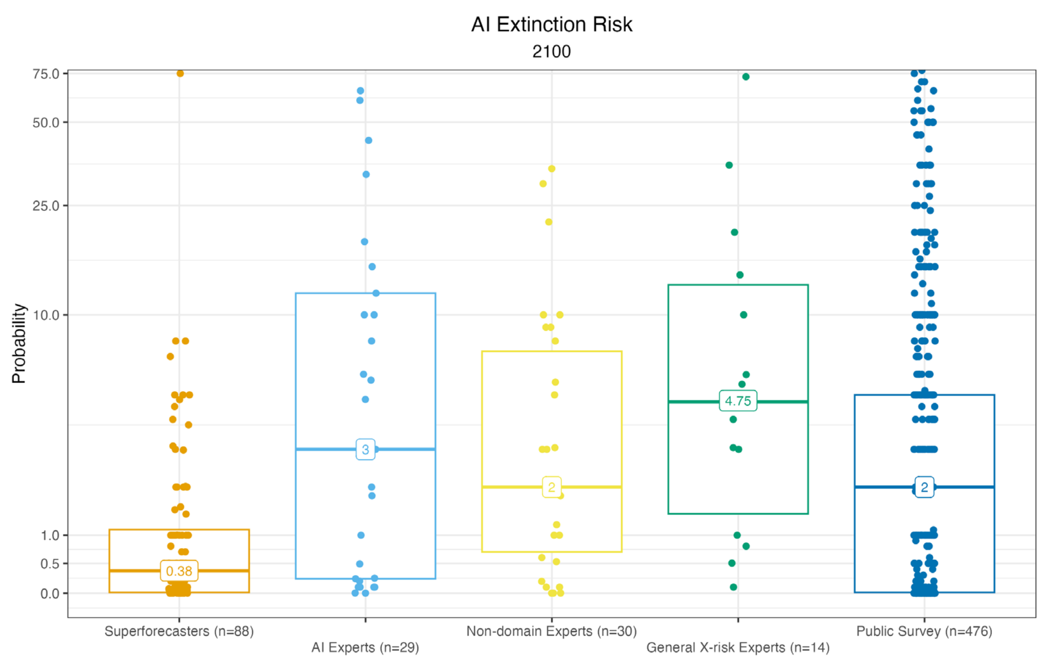


Figure 4: *Plot shows the distribution of final forecasts of AI extinction risk (by 2100) for each of four groups in the XPT (superforecasters, AI experts, non-AI experts, and general x-risk experts) as well as a public survey of college-educated respondents from an online platform. Box plots show the 25–75th percentile range (boxes) and the median (labeled) for each group. Within each group, points are jittered slightly horizontally to show density, but this jittering has no other empirical purpose.*

How should we interpret this persistent divide between groups? One possibility is that the experts are biased toward the topics they are professionally invested in and overweight the tail-risks they spend time thinking about—perhaps partly because those most worried about existential risk opt to dedicate their lives to studying it. Another is that the superforecasters are skilled at using historical data for relatively short-run forecasts but might struggle to adapt their methods to longer-run topics with less data—even when they have experts on hand to walk them through the topic. It is also possible that the epistemic strategies that were successful in earlier short-run forecasting tournaments, when the superforecasters attained their status, are not as appropriate at other points in time. For example, base rates may be more useful in periods of relative geopolitical calm and less useful in periods of greater conflict. All of these possibilities may be operating together, to various degrees.

While there is no definitive answer on which of these groups to trust, the XPT made contributions in that direction.

First, we can rule out the possibility that experts can't persuade others of the severity of existential risks simply because of a complete lack of sophistication, motivation, or intelligence on the part of their audience. The superforecasters have all those characteristics, and they continue to assign much lower chances than do experts.

Second, the XPT reports intersubjective measures of forecasting accuracy, providing a secondary measure that can be used to evaluate the two groups before their forecasts have resolved.

Finally, by the end of 2024 the XPT will begin to resolve forecasts of short-run indicators, providing new measures on which to judge the short-run objective accuracy of both groups.

Aggregating forecasts

We report median forecasts throughout this report. Why the median, given that there are many ways to aggregate individual forecasts? The median is straightforward to calculate, transparent, robust to extreme outlying observations, and understandable to people with a basic knowledge of statistics. Also, reassuringly, it is never the highest nor the lowest of the five methods we considered as potential aggregation methods. For these reasons, we think the median provides an ideal middle ground for aggregating forecasts in this project.

Table 4 shows five alternative aggregation methods for three key questions about extinction risk. We also present bootstrapped 95% confidence intervals for each of these aggregated forecasts.[38]

Forecast	Group	Estimate (95% confidence interval)				
		Arithmetic Mean	Highest-Density Trimmed Mean[39]	Median	Geometric Mean of Odds[40]	Extremized Aggregate[41]
AI Extinction Risk	AI domain experts	13.61% [6.46, 22.79]%	8.81% [3.57, 21.91]%	3% [0.49, 10.00]%	1.33% [0.17, 5.35]%	0.068% [0.0020, 0.7700]%
	Superforecasters	1.98% [0.86, 3.77]%	0.70% [0.43, 1.14]%	0.38% [0.10, 0.75]%	0.062% [0.019, 0.187]%	0.00030% [0.000039, 0.002000]%
Nuclear Extinction Risk	Nuclear domain experts	0.98% [0.41, 1.65]%	0.73% [0.27, 1.73]%	0.55% [0.075, 1.400]%	0.080% [0.0025, 0.7600]%	0.00079 [0.0000025, 0.0330000]%
	Superforecasters	0.29% [0.19, 0.42]%	0.14% [0.084, 0.230]%	0.074% [0.025, 0.100]%	0.030% [0.0096, 0.0600]%	0.000088 [0.000012, 0.000290]%
Total Extinction Risk	Experts	12.92% [8.78, 17.94]%	7.78% [5.37, 12.67]%	6% [3.41, 10.00]%	2.78% [1.01, 6.02]%	0.22% [0.038, 0.890]%
	Superforecasters	3.73% [2.09, 6.16]%	1.68% [1.06, 2.49]%	1% [0.55, 1.23]%	0.31% [0.13, 0.66]%	0.0050 [0.0011, 0.0180]%

Table 4: *Five alternative methods for aggregating forecasts from the XPT on questions of whether AI or nuclear war will, by 2100, cause humanity to go extinct. We aggregate forecasts from N=88 superforecasters, N=66 domain experts with expertise in AI, nuclear, biorisk, and climate domains, and N=14 general x-risk experts using 1) arithmetic mean, 2) extremized aggregate, 3) highest density trimmed mean, 4) median, and 5) geometric mean of odds. We also present bootstrapped confidence intervals for each sample statistic.*

Across these three questions and five aggregation methods, the mean was always the highest, often two or more times as large as the median. This is predictable, since forecasts on these questions tended to be in the 0-5% range, with a long tail of higher forecasts pulling up the mean.

38 We calculate the 95% confidence intervals by repeatedly resampling the original sample of forecasts with replacement, recalculating the sample statistic, and taking the range between the 2.5th and 97.5th percentiles of those sample statistics.

39 The highest-density trimmed mean takes the smallest interval containing 90% of the forecasts and takes the mean of only those forecasts in the interval (Ben Powell et al., "Skew-adjusted extremized-mean: "A simple method for identifying and learning from contrarian minorities in groups of forecasters", *Decision* (2022), https://doi.org/10.1037/dec0000191.

40 We replaced 0% forecasts with the 5th percentile forecast in the group. The geometric mean of odds always results in the most extreme aggregate. Many have made the case for extremizing forecasts, including Jonathan Baron et al., "Two Reasons to Make Aggregated Probability Forecasts More Extreme," *Decision Analysis* 11, no 2:133-145 (March 2014). http://dx.doi.org/10.1287/deca.2014.0293.

41 See Eric Neyman and Tim Roughgarden, "Are You Smarter Than a Random Expert? The Robust Aggregation of Substitutable Signals," *arXiv.org*, (November 4, 2021), https://arxiv.org/abs/2111.03153v2. We chose the extremizing factor as recommended in the paper.

Neyman and Roughgarden's extremizing method,[42] the highest-density trimmed mean, and the geometric mean of odds all require choices about parameters that affect the aggregate number. These parameters are, respectively: the extremizing factor; how aggressively to trim; and what to do with individual forecasts of 0% and 100%, which cannot be converted sensibly to odds. Small changes in these choices can have a big influence on the aggregation of forecasts.

Our conclusion, after exploring aggregation methods in this context, is that the large gaps between expert and superforecaster beliefs are robust across aggregation methods. So, we rely on medians throughout this report to maximize transparency.

How the XPT results compare to prior risk estimates

While the XPT is novel in terms of its format, it is far from the first attempt to quantify the probability of various existential risks. In *The Precipice*, published in 2020, the philosopher Toby Ord estimates a one-in-six chance of an existential risk occurring by 2120—which he defines as "a risk that threatens the destruction of humanity's long-term potential."[43]

While Ord's work may be the best known, it is not the only attempt to quantify existential risks to humanity.[44] A 2018 paper in the journal *Futures* by Simon Beard and colleagues compared 64 estimates of existential risks across 13 domains.[45] The authors document a wide range of methods and major disagreements over magnitudes. For example, the astrophysicist J. Richard Gott has used simple assumptions to provide a base rate: the human species will last between 0.2 million and 8 million years, a 95% confidence interval that builds on the Copernican principle that we should assume there is nothing special about our place or moment in the universe.[46] This is the opposite of Ord's view in *The Precipice*, which argues that humanity faces particularly high risk in the next century.

---

42 Neyman and Roughgarden, "Are You Smarter Than a Random Expert? The Robust Aggregation of Substitutable Signals."

43 Toby Ord, *The Precipice: Existential Risk and the Future of Humanity* (New York: Hachette Books, 2020), 37.

44 Throughout this section we rely heavily on Simon Beard, Thomas Rowe, and James Fox, "An Analysis and Evaluation of Methods Currently Used to Quantify the Likelihood of Existential Hazards," *Futures* 115 (January 2020) https://doi.org/10.1016/j.futures.2019.102469, and on the database of existential risks created by Michael Aird ("Database of existential risk estimates (or similar)," accessed June 24, 2023 https://docs.google.com/spreadsheets/d/1W10B6NJjicD8O0STPiT3tNV3oFnT8YsfjmtYR8RO_RI/edit#gid=1429038499).

45 Beard, Rowe, and Fox, "An Analysis and Evaluation of Methods Currently Used to Quantify the Likelihood of Existential Hazards."

46 J. Richard Gott III, "Implications of the Copernican Principle for Our Future Prospects," *Nature* 363, no. 6427 (May 1993): 315–19, https://doi.org/10.1038/363315a0.

How do the XPT results fit within the wider literature on existential risk? A feature of quantitative forecasting is that it forces us to make deliberations more explicit. To forecast existential risks, we first needed to be more precise about their meaning. A good forecasting question should pass what forecasting researchers call the clairvoyance test: if you posed it to a genuine clairvoyant, they could look into the future and tell you definitively what happened without asking for clarification. A forecasting question like "Will the economy be good next year?" fails the clairvoyance test; even with the ability to see the future, the clairvoyant would have to ask you to clarify what you meant before answering the question. By contrast, "Will US real annual GDP growth be positive in the first quarter of next year?" passes the test; the clairvoyant needs no clarification to answer.[47]

Ord's definition of an existential risk as "the destruction of humanity's long-term potential" may be qualitatively useful but it does not pass that test. For that reason, we chose two categories of events for forecasters to predict: the chance of human extinction, which we define as a drop in the global population to below 5,000 people, and the chance of what we call a "global catastrophic event," which we define as an event that causes the death of 10% of humanity within a five-year period.

Different definitions make it tricky to compare estimates of existential risk, but in the tables below we compare our extinction risk estimates by cause to estimates made by Ord and others.

Cause of extinction	Source	Forecast	Timeframe
Any cause	Sandberg & Bostrom 2008[48]	19%	By 2100
	Ord 2020[49]	16.7%	By 2120
	Hempsell 2004[50]	5-10%	By 2084
	XPT experts	**6%**	**By 2100**
	Metaculus 2021[51]	2%	By 2100
	XPT superforecasters	**1%**	**By 2100**

Table 5: *Presents forecasts of human extinction due to any cause (in decreasing likelihood order) and the timeframe given for each of: Sandberg and Bostrom 2008, Ord 2020, Hempsell 2004, XPT experts, 2021 Metaculus predictions, and XPT superforecasters. XPT results are medians calculated from N=87 superforecasters and N=67 experts.*

47 Philip E. Tetlock, "A Short Course in Superforecasting," *Edge*, accessed June 23, 2023, https://www.edge.org/conversation/philip_tetlock-a-short-course-in-superforecasting.

48 Anders Sandberg and Nick Bostrom, "Global Catastrophic Risks Survey," Technical Report #2008-1, Future of Humanity Institute, Oxford University, 1 (2008). https://www.fhi.ox.ac.uk/reports/2008-1.pdf.

49 Ord, *The Precipice*, 167.

50 C. M. Hempsell, "The Investigation of Natural Global Catastrophes," *Journal of the British Interplanetary Society* 57, no. 1-2 (Jan. 2004):2-13. https://research-information.bris.ac.uk/en/publications/the-investigation-of-natural-global-catastrophes-2.

51 "Will humans go extinct before 2100?" Metaculus, accessed June 24, 2023. https://www.metaculus.com/questions/578/human-extinction-by-2100/.

Cause of extinction	Source	Forecast	Timeframe
Artificial intelligence	Baum et al. 2017[52]	25%	-
	Müller & Bostrom 2016[53]	18%	-
	Ord 2020[54]	10%	By 2120
	Grace et al. 2022[55]	5%	-
	Sandberg & Bostrom 2008[56]	5%	By 2100
	Pamlin & Armstrong 2015[57]	0-10%	By 2115
	XPT AI experts	**3%**	**By 2100**
	Metaculus 2021	1.9%[58]	By 2100
	XPT superforecasters	**0.38%**	**By 2100**

Table 6: *Presents forecasts of whether humanity will go extinct due to AI (in decreasing likelihood order) and the timeframe given for each of: Baum et al. 2017, Müller & Bostrom 2016, Ord 2020, Grace et al. 2022, Sandberg & Bostrom 2008, Pamlin & Armstrong 2015, XPT AI experts, 2021 Metaculus predictions, and XPT superforecasters. XPT results are medians calculated from N=88 superforecasters and N=29 AI expert*

52 Seth Baum, Anthony Barrett, and Roman V. Yampolsky, "Modeling and Interpreting Expert Disagreement About Artificial Superintelligence," *Informatica* 41, no. 7 (Dec. 2017): 419-428 . https://ssrn.com/abstract=3104645.

53 Vincent C. Müller and Nick Bostrom, "Future Progress in Artificial Intelligence: A Survey of Expert Opinion," in Vincent C. Müller (ed.), Fundamental Issues of Artificial Intelligence (Synthese Library; Berlin: Springer, 2016):553-571. https://nickbostrom.com/papers/survey.pdf.

54 Ord, *The Precipice*, 167.

55 Katja Grace and Ben Weinstein-Raun, "2022 Expert Survey on Progress in AI," AI Impacts Wiki (Aug. 2022). https://wiki.aiimpacts.org/doku.php?id=ai_timelines:predictions_of_human-level_ai_timelines:ai_timeline_surveys:2022_expert_survey_on_progress_in_ai.

56 Sandberg and Bostrom, "Global Catastrophic Risks Survey," 1.

57 Dennis Pamlin and Stuart Armstrong, "12 Risks that threaten human civilisation: The case for a new risk category," Global Challenges Foundation (February 2015): 164.

58 "Ragnarök Series—results so far," Metaculus, accessed June 25, 2023. https://www.metaculus.com/questions/2568/ragnar%25C3%25B6k-seriesresults-so-far/.

Cause of extinction	Source	Forecast	Timeframe
Biorisk	Ord 2020[59]	3.3%	By 2120
	Pamlin & Armstrong 2015[60]	0.0001–5%	By 2095
	Sandberg & Bostrom 2008[61]	2.1%	By 2100
	XPT biorisk experts	**1.01%**[62]	**By 2100**
	Metaculus 2021[63]	0.2%	By 2100
	XPT superforecasters	**0.012%**	**By 2100**

Table 7: *Presents forecasts of whether humanity will go extinct due to biorisk (in decreasing likelihood order) and the timeframe given for each of: Ord 2020, Pamlin & Armstrong 2015, Sandberg & Bostrom 2008, XPT biorisk experts, 2021 Metaculus predictions, and XPT superforecasters. XPT results are medians calculated from N=77 superforecasters and N=7 biorisk experts.*

Cause of extinction	Source	Forecast	Timeframe
Nuclear war	Pamlin & Armstrong 2015[64]	0.005–5%	By 2115
	Sandberg & Bostrom 2008[65]	1%	By 2100
	XPT nuclear experts	**0.55%**	**By 2100**
	Metaculus 2021[66]	0.3%	By 2100
	Ord 2020[67]	0.1%	By 2120
	XPT superforecasters	**0.074%**	**By 2100**

Table 8: *Presents forecasts of whether humanity will go extinct due to nuclear war (in decreasing likelihood order) and the timeframe given for each of: Pamlin & Armstrong 2015, Sandberg & Bostrom 2008, XPT nuclear experts, 2021 Metaculus forecasts, Ord 2020, and XPT superforecasters. XPT results are medians calculated from N=88 superforecasters and N=13 nuclear experts.*

59 Ord, *The Precipice*, 167.
60 Pamlin and Armstrong, "12 Risks that threaten human civilisation," 150.
61 Sandberg and Bostrom, "Global Catastrophic Risks Survey," 1.
62 The domain expert and superforecaster forecasts in this table are the combined probability each group gave to the risk of extinction due to engineered pathogens and non-engineered pathogens in the post-XPT survey. See the "Biorisks" section below for more detail.
63 "Will humans go extinct before 2100?" Metaculus, accessed June 24, 2023. https://www.metaculus.com/questions/578/human-extinction-by-2100/.
64 Pamlin and Armstrong, "12 Risks that threaten human civilisation," 148.
65 Sandberg and Bostrom, "Global Catastrophic Risks Survey," 1.
66 Metaculus, "Will humans go extinct before 2100?".
67 Ord, *The Precipice*, 167.

Contrasting experts and superforecasters in different domains

The superforecasters rated the likelihood of AI causing extinction by 2100 to be roughly an order of magnitude lower than did experts. The median superforecaster gives a 0.38% chance of AI-driven extinction by the end of the century, compared to 3% for the median domain expert.

Something similar is true of catastrophic risks from AI—defined as an event that causes the deaths of at least 10% of the population. The median expert puts the chance of AI causing such an event about six times higher than does the median superforecaster.

> Experts and superforecasters were furthest apart on AI risk—and less so on the risk of nuclear war.

	Median Estimate (95% confidence interval)		
Forecast	**Superforecasters**	**Domain Experts**	**Non-Domain Experts**
AI Catastrophic Risk by 2100	2.13% [1.83, 3.00]%	12% [4.0, 18.5]%	6.16% [5.0, 12.0]%
AI Extinction Risk by 2100	0.38% [0.10, 0.75]%	3% [0.49, 10.00]%	2% [1.00, 4.03]%

Table 9: *Presents final median forecasts from the XPT for superforecasters, domain experts, and non-domain experts on questions of whether AI will, by 2100, cause a "catastrophe" (i.e., be the cause of death within a 5-year period for more than 10% of humans alive at the beginning of that period) or cause humanity to go extinct.*

Superforecasters and experts alike expect major progress in AI over the next few decades. For example, superforecasters expect the arrival of advanced AI capable of passing comprehensive Turing tests and answering questions accurately across domains by 2060.[68] That median forecast is roughly 15 years later than the experts' median forecast, but well within this century. However, the two groups disagree about how widely these technologies will proliferate outside the lab and whether they will have the inclination and ability to threaten humanity. Superforecasters' doubts about AI risk relative to the experts isn't primarily driven by an expectation of another "AI winter" where technical progress slows.

68 See details in the summary of forecasts on the Date of Advanced AI.

That said, views on the likelihood of artificial general intelligence (AGI) do seem important: in the postmortem survey, conducted in the months following the tournament, we asked several conditional forecasting questions. The median superforecaster's unconditional forecast of AI-driven extinction by 2100 was 0.38%. When we asked them to forecast again, conditional on AGI coming into existence by 2070, that figure rose to 1%.

Superforecasters are also more skeptical of extinction due to nuclear weapons, but the gap is smaller: experts put the chances of that outcome seven times that of superforecasters (0.074% vs. 0.55%). The two groups are much closer together when it comes to the catastrophic risk from nuclear weapons—meaning the chance that nuclear weapons kill 10% of the world's population by 2100. The superforecasters give a 4% chance of that outcome, compared to 8% for the experts—within the same order of magnitude.

	Median Estimate		
Forecast	**Superforecasters**	**Domain Experts**	**Non-Domain Experts**
Nuclear Catastrophic Risk by 2100	4%	8%	8%
Nuclear Extinction Risk by 2100	0.074%	0.55%	0.19%

Table 10: *Presents final median forecasts from the XPT for superforecasters, domain experts, and non-domain experts on questions of whether nuclear war will, by 2100, cause a "catastrophe" (i.e., be the cause of death within a 5-year period, for more than 10% of humans alive at the beginning of that period) or cause humanity to go extinct.*

This result also illustrates just how extreme the divides over AI risk are. On nuclear catastrophe, experts and superforecasters reached similar conclusions. That's less true of nuclear extinction risk—but it is even less true with regard to AI. Something about AI risk leads these two groups to reach extremely different conclusions. (For more on this disagreement over AI risk see the sections "Few minds were changed…" and "Artificial intelligence".)

Despite the disagreement over AI risk between superforecasters and experts, both groups support devoting more resources to the topic. In intake and postmortem surveys, we asked XPT participants questions about how they would allocate resources to mitigate potential risks. We asked three distinct questions:

- “If you could allocate an additional $10,000 to x-risk avoidance, how would you divide the money among the following topics?”
- “If you could allocate the time of 100 new researchers (assuming they are generalists who could be effective in a wide range of fields), how would you divide them among the following topics?”
- “Assume we are in a world where there are no current attempts by public institutions, governments, or private individuals or organizations to allocate spending toward catastrophic risk avoidance. If you were to allocate $50 billion to the following risk avoidance areas, what fraction of the money would you allocate to each area?”

AI received the most resources in nearly all instances—from superforecasters as well as from experts. The only exception, shown in the table below, was for the $50 billion question in the postmortem survey, where superforecasters dedicated 21% of the money to AI risk mitigation and 26% to nuclear risk mitigation. In the other cases, superforecasters dedicated the largest single chunk of resources to AI risk. And in all cases, the experts—not just AI experts—collectively put more resources into AI risk (though never more than 40%).

Question	Group	Survey	AI allocation	Largest non-AI allocation	Field of largest non-AI allocation
Allocate an additional $10,000 to x-risk avoidance	Superforecasters	Intake	21.5%	18.8%	Other
		Postmortem	25.5%	19.5%	Nuclear
	Experts	Intake	31.7%	26.7%	Engineered pathogen
		Postmortem	36.6%	22.1%	Engineered pathogen
Allocate the time of 100 new researchers among the following topics	Superforecasters	Intake	25.4%	19.3%	Engineered pathogen
		Postmortem	28.9%	18.5%	Natural pathogen
	Experts	Intake	39.2%	21.3%	Engineered pathogen
		Postmortem	39.9%	23.2%	Engineered pathogen
Allocate $50 billion in a world where there are no current attempts to allocate spending toward catastrophic risk avoidance	Superforecasters	Intake	20.3%	17.7%	Nuclear
		Postmortem	21.1%	25.5%	Nuclear
	Experts	Intake	30.8%	22.1%	Engineered pathogen
		Postmortem	33.2%	20.4%	Engineered pathogen

***Table 11**: Results from the XPT intake and postmortem surveys, showing superforecasters’ and experts’ responses when asked how they would 1) allocate an additional $10,000 to x-risk avoidance, 2) allocate the time of 100 new researchers among the following topics, and 3) allocate $50 billion in a world where there are no current attempts to allocate spending toward catastrophic risk avoidance. For each question, we show what percentage superforecasters and experts would allocate to AI risk mitigation, what percentage to their largest non-AI allocation, and what field would receive their largest non-AI allocation. We estimate these numbers from 136 forecasters.*

Both groups seem to want more resources devoted to AI risk at the margin, and neither seems to want to dedicate the majority of resources for risk mitigation to a single risk area.

Contrasting tournament participant responses with members of the public

Participants in our public survey of 912 college graduates estimated a higher median probability of extinction by 2100 (5%) than superforecasters (1%) but lower than that of experts (6%). A similar pattern also emerged for AI-caused extinction (public survey participants gave a 2% probability, and superforecasters and domain experts gave 0.38% and 3%, respectively).

> Members of the public estimated probabilities of extinction in between the estimates of experts and superforecasters, but diverged significantly with alternative probability elicitation formats.

However, respondents of the same sample estimated much lower chances of both extinction and catastrophe by 2100 when presented with an alternative elicitation method. In a follow-up survey, we gave participants examples of low probability events—for example, that there is a 1-in-300,000 chance of being killed by lightning. We then asked them to fill in a value for "X" such that there was a "1-in-X" chance of a given risk—like human extinction by 2100.[69]

69 The set of reference classes we gave to participants had ten examples, including:

1 in 2:	Probability a flip of a fair coin will be Tails
1 in 300,000:	Lifetime probability of dying from lightning
1 in 10,000,000:	Probability a random newborn becomes a U.S. president

Using that method,[70] the median probability of humanity's extinction before 2100 was 1 in 15 million. The median probability of AI-caused extinction before 2100 was 1 in 30 million. The median probability of catastrophe (10% of the population dying) before 2100 was 1 in 2,500. The median probability of AI catastrophe before 2100 was 1 in 100,000.

Intersubjective accuracy and risk estimates

While we wait to see which XPT participants were more and less accurate in their shorter-run forecasts, beginning at the end of 2024, we can score their reciprocal forecasts—their forecasts about what the superforecasters and experts in the XPT would predict for each question. These results are interesting because accuracy indicates an understanding of other participants' reasoning.[71]

> Among both superforecasters and experts, those with higher "intersubjective accuracy" estimated lower probabilities of catastrophic and extinction risks.

As described earlier ("How the XPT works"), each participant gave their own forecast and two reciprocal predictions for each required question about existential and catastrophic risk: one predicting what the median superforecaster would say, and the other predicting what the median expert would say. Forecaster's provided these forecasts before seeing other participants' forecasts. We use these intersubjective

70 Due to constraints on the survey respondents' available time, the public survey had to be staged in multiple parts across several weeks. As such, not all respondents answered all questions. When subsetting to only those who participated in parts of the survey that including each method of probability elicitation (N=405), the median probabilities were as follows:
Total Extinction Risk (simple textbox elicitation): 4%
Total Extinction Risk (alternative method): 1 in 20 million
AI Extinction Risk (simple textbox elicitation): 1.5%
AI Extinction Risk (alternative method): 1 in 40 million
Total Catastrophic Risk (simple textbox elicitation): 10%
Total Catastrophic Risk (alternative method): 1 in 2,000
AI Catastrophic Risk (simple textbox elicitation): 5%
AI Catastrophic Risk (alternative method): 1 in 50,000

71 Karger et al., "Reciprocal Scoring."

forecasts to calculate 'reciprocal scores' for each forecaster, measuring their intersubjective accuracy compared to other participants.[72]

Overall, the superforecasters had better reciprocal scores than did the experts, driven by their more accurate predictions of other superforecasters' views[73] (both groups were comparable at predicting experts' views). While past work has indicated that forecasters who are accurate on intersubjective metrics will be more accurate on resolvable questions,[74] it is possible that that result doesn't hold for longer-term, existential questions like the ones investigated in the XPT. We'll be able to compare reciprocal accuracy to real-world accuracy in the XPT starting at the end of 2024.

More interesting than the difference in reciprocal accuracy between groups is the fact that within both groups—experts and superforecasters—more accurate reciprocal scores were correlated with lower estimates of catastrophic and extinction risk. In other words, the better experts were at discerning what other people would predict, the less concerned they were about extinction.

Figure 5 plots total extinction risk by 2100 for both groups—and displays quintiles of reciprocal scoring accuracy (accuracy is highest on the left, lowest on the right). Within both groups, those who did best on reciprocal scoring had lower forecasts of extinction risk.

72 We separately compare each forecaster's forecast of others' forecasts on ten key questions, for both experts and superforecasters. We rank each forecaster's accuracy on those 20 quantities relative to other participants, and then we compute each forecaster's average rank to calculate an overall measure of intersubjective accuracy.
73 This may be because superforecasters are a more homogenous group, who regularly interact with each other outside of forecasting tournaments like this.
74 Pavel Atanasov et al., "Full Accuracy Scoring Accelerates the Discovery of Skilled Forecasters," *SSRN Working Paper*, (February 14, 2023), http://dx.doi.org/10.2139/ssrn.4357367.

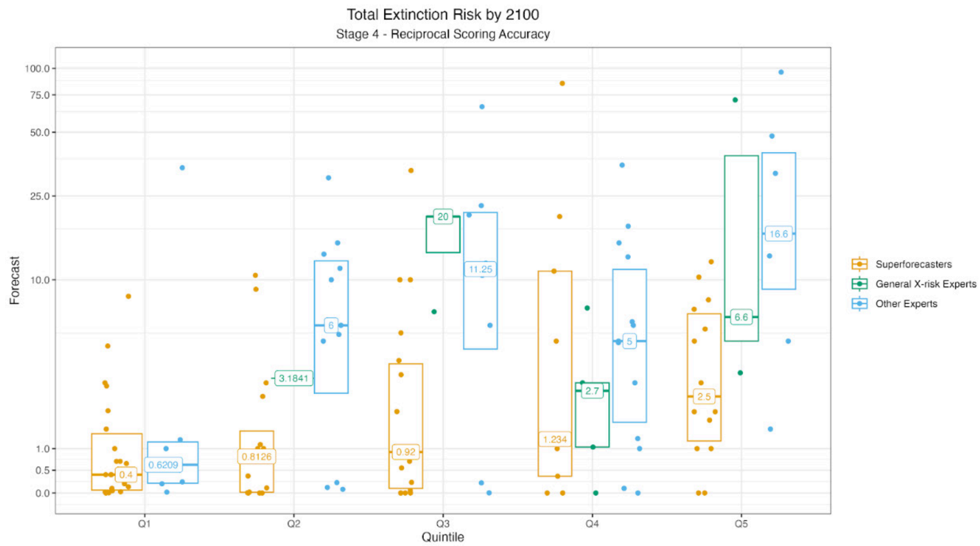


Figure 5: *Shows the relationship between final forecasts and reciprocal scoring accuracy (i.e., how accurately did the forecaster predict what the median of the superforecasters' forecasts would be, and likewise experts) for each group (superforecasters, general x-risk experts, and other experts). The x-axis is reciprocal accuracy quintiles (Q1 is the fifth of forecasters whose reciprocal scoring accuracy was best; Q5 is the fifth of forecasters whose reciprocal scoring accuracy was worst). The y-axis is the final (Stage 4) forecasts on the XPT question of whether humanity will go extinct by 2100. Note that the y-axis is non-standard and uses a pseudo-log transformation to stretch the 1–10% range to visualize the key disagreements.*

The same pattern is evident in forecasts for catastrophic risk: superforecasters and experts who scored better in reciprocal scoring estimate a lower chance of catastrophe (shown in Figure 6).

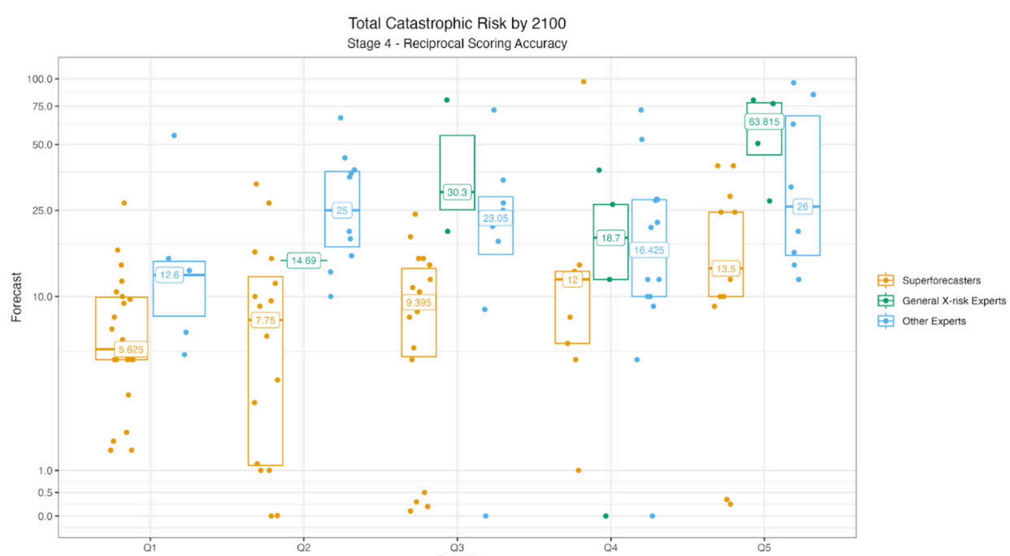


Figure 6: *Shows the relationship between final forecasts and reciprocal scoring accuracy (i.e., how accurately did the forecaster predict what the median of the superforecasters' forecasts would be, and likewise experts) for each group (superforecasters, general x-risk experts, and other experts). The x-axis is reciprocal accuracy quintiles (Q1 is the fifth of forecasters whose reciprocal scoring accuracy was best; Q5 is the fifth of forecasters whose reciprocal scoring accuracy was worst). The y-axis is the final (Stage 4) forecasts on the XPT question of catastrophic risk by 2100.*

The same is true for extinction risk from AI specifically (Figure 7).

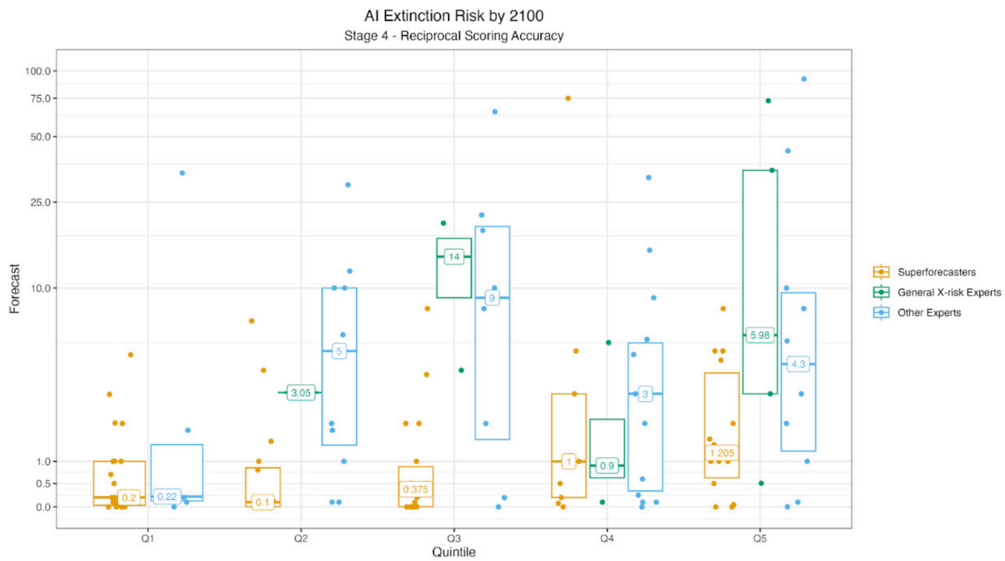


Figure 7: *Shows the relationship between final forecasts and reciprocal scoring accuracy (i.e., how accurately did the forecaster predict what the median of the superforecasters' forecasts would be, and likewise experts) for each group (superforecasters, general x-risk experts, and other experts). The x-axis is reciprocal accuracy quintiles (Q1 is the fifth of forecasters whose reciprocal scoring accuracy was best; Q5 is the fifth of forecasters whose reciprocal scoring accuracy was worst). The y-axis is the final (Stage 4) forecasts on the XPT question of extinction risk due to AI by 2100.*

Intersubjective metrics are a promising tool for eliciting forecasts of unresolvable questions and evaluating forecasters' understanding of the view of different groups on complex topics. While there is little evidence (yet) linking intersubjective accuracy to real-world accuracy on other questions, the strong correlation between intersubjective accuracy and beliefs about catastrophic and existential risks in the XPT point to a promising area of future research.

Relationships between risk predictions in different domains

The XPT format allowed us to assess how participants' forecasts varied across risk areas. We found that assessments of risk were highly correlated. For example, those who forecasted a relatively high chance of extinction due to AI gave higher chances of extinction from other causes as well.

> Predictions about risk are highly correlated across topic areas—for example, participants who are more concerned about AI are also more concerned about pandemics and nuclear weapons.

This is not a trivial result, as, for example, the risks of nuclear war and a pandemic are not obviously correlated. Instead, we found that some people are more concerned about everything we asked about. Some people see humanity's future as much riskier than others do.

There are some reasons to expect the risks to be correlated: our resolution criteria allow that if an AI system uses nuclear weapons to cause human extinction in a manner that counterfactually requires both technologies, it would count as both AI- and nuclear-caused human extinction for the purposes of these forecasts. So, if forecasters expect these risks to be related, they would have higher forecasts of both events.

As already discussed, the magnitude of disagreement about risk was smaller in some areas—like nuclear weapons—and larger in others—like AI. However, predictions in these two domains were still correlated: those who thought AI catastrophe or extinction were more likely also thought nuclear catastrophe or extinction were more likely.

What separates those who see the world as more versus less risky? To explore that, we separated XPT forecasters into two groups: the "AI-concerned"—the third of participants with the highest forecast of AI extinction risk by 2100—and the "AI skeptics," defined as the third of participants with the lowest forecast of AI extinction risk by 2100. There were fifteen superforecasters and 68 experts in the AI-concerned group. The AI skeptics group was split about evenly: 40 superforecasters and 43 experts. We found:

- The separation between the AI-concerned and AI skeptics on beliefs about extinction risk from AI by 2100 is very large (AI-concerned median: 7.5%; AI skeptics median: 0.01%). This difference is also reflected in their forecasts for total catastrophic risk (AI-concerned median: 27.68%; AI skeptics median: 5.15%) and total extinction risk by 2100 (AI-concerned median: 11%; AI skeptics median: 0.12%).

- We also see a large separation of beliefs on non-AI questions. For example, the AI-concerned think the risk that a genetically engineered pathogen will kill more than 1% of people within a 5-year period before 2100 is 12.38%, while the AI skeptics forecast a 2% chance of that event, with 96% of the AI-concerned above the AI skeptics' median forecast. There are narratives that could explain the relationship between these beliefs, such as the possibility that advanced AI makes bioengineering much easier. But the pattern also holds for questions where AI is less relevant: When asked about non-anthropogenic extinction risk, the AI-concerned's median is 0.0098% while the AI skeptics' is 0.00088%—an order of magnitude difference. The separation is smaller here (in percentage points) than on the AI questions, but nonetheless 82% of the AI-concerned are above the AI skeptics' median forecast. Across the board, the participants most concerned about catastrophic risks from AI are more concerned about other catastrophic and extinction-level risks—even ones that seem unrelated.
- Just as these two groups disagree strongly about how likely humanity is to go extinct by 2100, they also disagree about other indicators of human flourishing. We asked forecasters in the XPT how many future human births there will be and by what year humanity is 50% likely to go extinct. The AI-concerned's median is 100 billion future human births, while the AI skeptics' median is just over 725 billion, with 78% of the AI skeptics above the median forecast of the AI-concerned. The AI-concerned's median forecast for the year when humanity goes extinct is 3450 and the AI skeptics' median forecast is the year 30,000. In other words, the median AI skeptic expects humanity to exist in the universe for more than 26,000 years longer than the median forecaster in the AI-concerned camp.

Forecast	Number of Forecasters: AI-Concerned	Number of Forecasters: AI Skeptic	Median Forecasts: AI-Concerned Real	Median Forecasts: AI-Concerned Imputed	Median Forecasts: AI Skeptic Real	Median Forecasts: AI Skeptic Imputed	Distributional Overlap (imputed forecasts): % AI-concerned above median (AI skeptic)	Distributional Overlap (imputed forecasts): % AI skeptic above median (AI-concerned)
1. Genetically Engineered Pathogen Risk by 2100	54	54	12.06%	12.38%	2%	2%	96%	7%
2. Non-Genetically Engineered Pathogen Risk by 2100	55	54	7.07%	7.54%	3%	2.67%	79%	30%
3. AI Catastrophic Risk by 2100	55	53	12%	12.00%	0.80%	0.70%	100%	2%
4. AI Extinction Risk by 2100	54	53	7.50%	7.50%	0.01%	0.01%	100%	0%
5. Nuclear Catastrophic Risk by 2100	54	54	7.90%	8.00%	2.25%	2.57%	88%	25%
6. Nuclear Extinction Risk by 2100	54	52	0.41%	0.45%	0.01%	0.01%	88%	11%
7. Non-Anthropogenic Catastrophic Risk by 2100	51	53	0.14%	0.17%	0.01%	0.03%	82%	12%
8. Non-Anthropogenic Extinction Risk by 2100	52	53	0.01%	0.01%	0.0007%	0.0009%	82%	16%
9. Total Catastrophic Risk by 2100	50	53	27.30%	27.68%	5.30%	5.15%	100%	5%
10. Total Extinction Risk by 2100	51	53	11.00%	11.00%	0.12%	0.12%	100%	2%
11. Year of Extinction	50	52	3500	3500	28000	30000	21%	95%
12. Future Human Births	48	51	100 Billion	100 Billion	700 Billion	725 Billion	23%	79%

Table 12: *Compares final (Stage 4) forecasts of two groups: the top third of forecasters most concerned about AI extinction risk by 2100 (the AI-concerned) and the bottom third of forecasters least concerned about AI extinction risk by 2100 (the AI skeptics). We report the number of people in each group who forecasted on each question, the median forecasts of each group (real), and the imputed median forecasts of each group, where we impute the forecasts of anyone who did not answer a given question using a KNN-algorithm (K=3) fit to forecasts on questions participants did complete (imputed forecasts). We also report the fraction of each group above the median of the other group, focusing on imputed forecasts for that calculation.*

Not every question we asked correlated in this way. Some questions, like those about the development of clean energy technologies,[75] showed much less disagreement between those more and less concerned about AI risk.

For more details on our analysis comparing the AI-concerned and AI skeptics on other indicators, see Appendix 4: Comparisons between subgroups of XPT forecasters.

How time spent thinking about existential risk relates to forecasts of those risks

In the postmortem survey for the XPT, we asked participants to estimate how much time they'd spent thinking about existential risks to humanity prior to the tournament. Figure 8 shows that the longer a participant spent thinking about existential risks prior to the tournament, the more concerned they were about human extinction caused by AI.

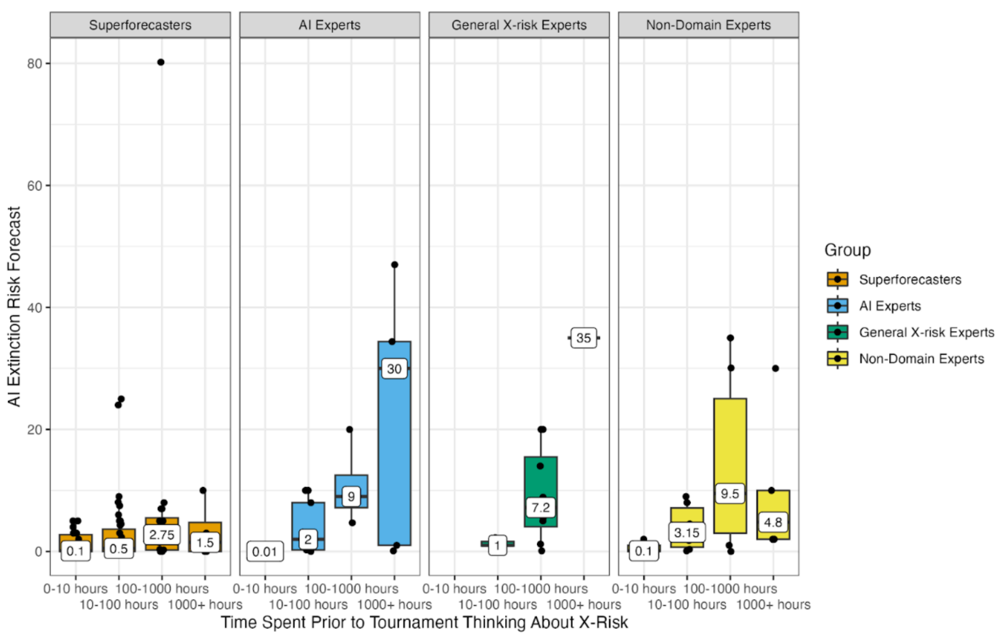


Figure 8: *For each of four groups—superforecasters, AI experts, general x-risk experts, and non-domain experts—we present participants' estimate of how many hours they had spent thinking about existential risks to humanity prior to the XPT (x-axis) vs. their final forecast on the XPT question of whether humanity will go extinct due to AI by 2100.*

AI experts who estimated they had spent more than 1,000 hours thinking about existential risks put the chance of AI-driven extinction by 2100 at 30%; those who had spent 10–100 hours put it at just 2%.

75 See our summaries of forecasts on Nuclear Fusion Energy and Solar and Wind Energy.

Why the link? It could be that spending time considering risks leads people to take AI risk more seriously based on the quality of the evidence and arguments they discover. Or, there could be a selection effect: those who choose to spend 1,000 hours thinking about existential risk are the people who are most likely to see the world as risky.

Something similar was true for the overall risk of human extinction: those who had spent more time thinking about existential risks put higher chances of extinction from any cause. But not every risk area displayed this pattern. For example, those who had spent more time thinking about existential risk did not have substantially different forecasts for nuclear catastrophe or extinction than those who had spent less time.

Persuasion in the XPT

Despite incentives for both persuasive rationales and reciprocal scoring, there was very little convergence within teams during the XPT. That is notable because both incentives might plausibly lead people to change their minds. Well-thought-out rationales might prove more persuasive; reciprocal scoring challenges forecasters to better understand other participants and what they think.

Few minds were changed during the XPT, even among the most active participants, despite monetary incentives for persuading others.

Figure 9 shows the median forecasts for each group on the question of extinction by 2100, plotted from the second stage of the tournament through to its completion.

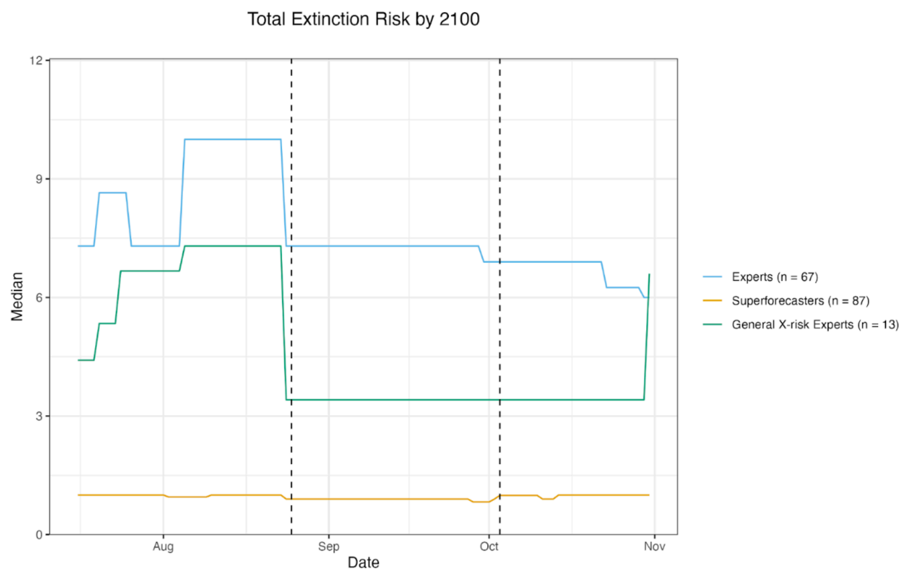


Figure 9: *Presents changes in forecasts over time for the XPT question of whether humanity will go extinct by 2100, showing the date of forecast (x-axis) vs. median forecast (y-axis), for each of the 67 experts, 87 superforecasters, and 13 general x-risk experts. Dotted lines indicate the start of stage 3 and stage 4 of the tournament.*

While the superforecasters' median remained relatively unchanged over the course of the tournament, the standard deviation of their forecasts suggests that superforecasters had a slightly greater tendency to converge over time than experts. Figure 10 plots how the standard deviations within each group narrowed for the same question over the course of the tournament.

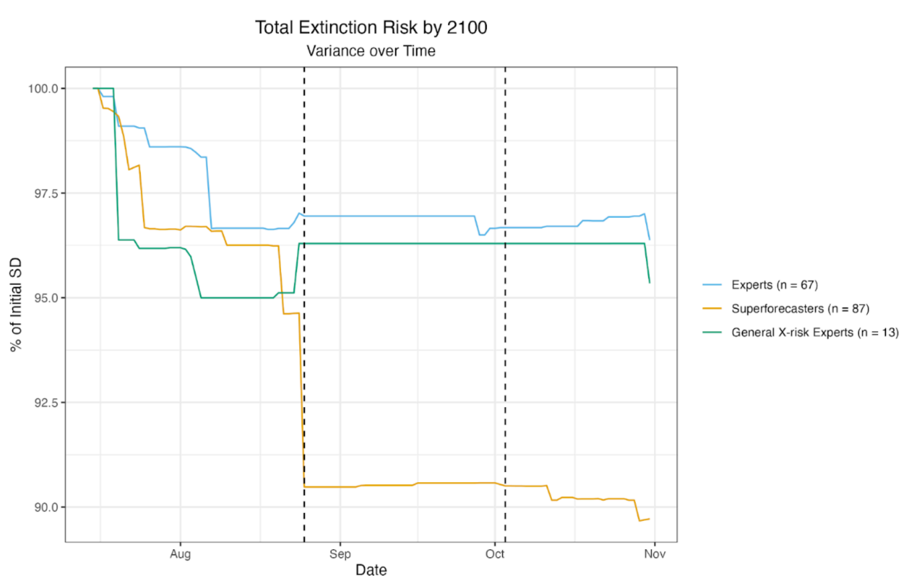


Figure 10: *Presents variance over time in forecasts for the XPT question of whether humanity will go extinct by 2100, showing the date of forecast (x-axis) vs. the standard deviation for forecasts on that date as a percentage of that group's initial standard deviation (y-axis), for experts, superforecasters, and general x-risk experts. Dotted lines indicate the start of stage 3 and stage 4 of the tournament.*

Despite some modest convergence—both within groups and between them—the upshot is that the two groups started far apart and ended far apart. It's worth noting that the two groups only interacted over a period of a few months. It is possible that persuasion requires a longer process, giving individuals time to think through arguments independently. It's also possible that for some arguments, no amount of time would lead to changes of belief. Determining how persuasion happens and why there wasn't more persuasion in the XPT will be a key priority for future research.

Differing styles of argumentation as a barrier to bridging disputes on AI risk

One might suppose that forecasters who are free to exchange information, incentivized to make true predictions, and skilled at Bayesian reasoning would converge over time. But they did not. Deep divisions on AI risk in particular persisted to the end of the XPT.

We discuss the reasons for the impasse here. Despite our encouragement to engage, forecasters often talked past each other. Their patterns of argumentation were grounded in different priors, different levels of analysis, and different views about how effectively social and political systems will respond to AI risk:

- Conflicting priors, which were invoked to justify different burdens of proof
    - » AI skeptics saw claims that AI will lead to catastrophic outcomes as extraordinary and thus as requiring extraordinary evidence. They typically started from a prior that humanity is robust and its long history of survival creates a high burden of proof for claims that we'll go extinct soon. AI-concerned forecasters were more likely to place the burden of proof on skeptics to explain why AI is not dangerous. They typically started from a prior that when a more intelligent species or civilization arises, it will overpower competitors.
- Inside view vs. outside view[76]
    - » AI-concerned forecasters often advanced theories that stressed the unprecedented features of future AI systems and downplayed the relevance of historical analogies or reference classes. AI skeptics more often took the "outside view" and used reference classes, pointing out that stories about AI risk are speculative and draw analogies from previous predictions of mega-disasters that have yet to be realized.

76 Daniel Kahneman, *Thinking, Fast and Slow*. (Macmillan, 2011). See excerpt in "Beware the inside view." McKinsey Quarterly (2011): 1-4. Accessed June, 2023 at https://www.mckinsey.com/capabilities/strategy-and-corporate-finance/our-insights/daniel-kahneman-beware-the-inside-view

- Competing claims about the adaptive capacities of complex social systems
    - » AI skeptics put more faith in the collective ability of humans to cope effectively with emerging AI dangers, especially after instances where AI causes large-scale damage or death (often termed "warning shots"). AI-concerned forecasters doubt that humans will respond effectively even in response to clear warning shots, given the financial and geopolitical arms race incentives for AI development.

We provide more details on forecasters' arguments related to AI in the "Artificial Intelligence" section below and in question-level summaries in Appendix 7. The question-level summaries include footnotes with sample quotes from forecasters.

These debating postures gave each side the flexibility to stalemate the other—and, in some cases, even to feel justified in tuning out the other side because "they just don't get it." In the "Next Steps" section, we discuss how we will build explicit adversarial collaborations into future XPTs to reduce the frequency of such outcomes.

Relationship between short-run forecasting questions and longer-term disagreements

If accuracy on long-run risk questions is correlated with accuracy on short- and medium-run forecasting questions—itself an open research question—then as questions in our tournament resolve in 2024 and 2030, we can get an updated sense of which groups of forecasters' long-run risk predictions are most likely to come true.

Focusing on the biggest area of long-run disagreement, AI, there were surprisingly high levels of agreement on 45 short-run indicators between forecasters who gave a high vs. low chance of extinction due to AI. Below, we characterize the topics where forecasters agreed and disagreed.[77]

As previously described, we again separated XPT forecasters into two groups: the "AI-concerned"—the third of participants with the highest forecast of AI extinction risk by 2100—and the "AI skeptics," defined as the third of participants with the lowest forecast of AI extinction risk by 2100.[78] Then we assessed which differences in

77 The forecasts described in this section include imputed median forecasts for each group, where we imputed the forecasts of anyone who did not answer a given question using a K-Nearest Neighbor algorithm (K=3) fit to forecasts on questions participants did complete.

78 While it is common to report correlations in contexts like these, we believe explicit reporting of median forecasts for each group provides more perspective on differences of belief. Also, correlation coefficients can easily approach 0 or 1 on many of our questions if one outlier forecast is many orders of magnitude greater or lower than most forecasts, which can easily happen when forecasting low-probability outcomes (like the probability of extinction from non-anthropogenic sources) or continuous measures with plausible values spanning many orders of magnitude (like the number of future human births).

predictions due to resolve by 2030 are most associated with longer-term differences in AI extinction risk predictions, and conversely, where there seems to be agreement. We found:

- As well as believing that extinction from AI is more likely by 2100, the AI-concerned also think that AI-caused extinction is much more likely by 2030 (AI-concerned median: 0.5%; AI skeptic median: 0.0000000005%). This difference is also reflected in their forecasts for total catastrophic risk by 2030 (AI-concerned median: 3%, AI skeptic median: 0.69%) and total extinction risk by 2030 (AI-concerned median: 1%, AI skeptic median: 0.001%).
- Although these two groups disagree strongly about how likely humanity is to go extinct by 2030, they disagree much less about other indicators of humanity's flourishing. Of the roughly 45 other 2030 indicators we asked about, only six showed quantitatively meaningful non-overlap in forecasting distributions (with either more than 75% of the AI-concerned giving forecasts above the median forecast of the AI skeptics or vice versa (where less than 25% of the AI skeptics gave forecasts above the median of the AI-concerned)).
- Only two questions had noticeable non-overlap in distributions of forecasts, with both more than 75% of the AI-concerned above the median skeptic forecast and less than 25% of the AI skeptics above the AI-concerned median forecast: (1) whether Nick Bostrom will affirm the existence of AGI by 2030 (AI-concerned median: 6.6%; AI skeptic median: 3.8%) and (2) whether there will be a space colony by 2030 (AI-concerned median: 0.06%; AI skeptic median: 0.0001%). Nick Bostrom affirming the existence of AGI is related to the likelihood of extinction caused by AI, so it makes sense as a leading indicator. Space colonization is less directly related to AI risk, although it may be more likely in worlds that have advanced AI.
- The other two questions where more than 75% of the AI-concerned gave forecasts above the median of AI skeptics were whether there will be negative public opinion of AI by 2030 (AI-concerned: 35.5%; AI skeptics: 33.9%) and whether there will be a dramatic and sudden decline in agricultural production by 2030 (AI-concerned: 0.64%; AI skeptics: 0.22%). Those where less than 25% of AI skeptics gave forecasts above the median of the AI-concerned were the cost of compute for the largest AI experiment by 2030 (AI-concerned median: $156,404,996; AI skeptics: $100,000,000), and whether nuclear weapons will be used and cause the deaths of at least 1,000 people by 2030 (AI-concerned: 6.7%, AI skeptics: 4.7%). The difference in the groups' expectations for negative public opinion of AI and spending on compute is unsurprising given their AI forecasts. Loss of agricultural production and use of nuclear weapons could be downstream of AI developments but may also reflect a generally more pessimistic outlook among the AI-concerned in the coming years.

On all other questions, there is less separation between groups, showing that they have similar expectations for many aspects of the world in 2030:

- Some of the questions with much less separation—implying more agreement—include the number of people who will get a non-coronavirus mRNA vaccine (each group's median: 100,000,000), how many public health emergencies of international concern will cause more than 10,000 deaths (each group's median: 2), and what percentage of Americans will say they are happy in a 2030 survey (each group's median: 86%).
- There was also little or no separation between the groups on the three questions about 2030 performance on AI benchmarks (MATH, Massive Multitask Language Understanding, QuALITY).[79] On these questions the AI-concerned and the AI skeptic medians never differed by more than 2%.[80] In each case, there were between 40% and 50% of skeptics whose forecasts fell above the AI-concerned median. The AI-concerned and the AI skeptics seem to agree on short-run capabilities of machine learning models. Naively, we had anticipated that these short-run benchmarks would be useful for discriminating between schools of thought on AI risks. However, these results suggest that these three benchmarks, at least, won't help to update beliefs on AI risk in 2030.

Looking at longer-term intermediate indicators, there was large disagreement between the AI-concerned and AI skeptics about the probability of global GDP growing by 15% or more in any year between now and 2100 (AI-concerned median: 10.5%; AI skeptics median: 4.5%). Similarly, for the cost of compute of the largest AI training run before 2050, the AI-concerned's median is above $822.7 million and the AI skeptics' is only $303.4 million (compared to about $40 million for training GPT-4[81]). These are consistent with the AI-concerned's general view that advanced AI will transform the world more dramatically than the AI skeptics expect.

For more details comparing the AI-concerned and AI skeptics on short- and medium-term indicators, see Appendix 4: Comparisons between subgroups of XPT forecasters.

79 See the summaries of forecasts on Massive Multitask Language Understanding Benchmark, MATH Dataset Benchmark, and QuALITY Dataset Benchmark for more detail.

80 Note that the units for these questions are percentages (%) but not estimates of probability—they are accuracy scores. The median AI skeptic believes the state-of-the-art accuracy on the MATH dataset benchmark will be 87.13% by 2030; the median AI-concerned, 87.09%. On the QuALITY benchmark, the median skeptic assigned 76.88%; the median AI-concerned, 78.67%. On the Massive Multitask Language Understanding benchmark, the median skeptic assigned 88.89%; the median AI-concerned, 89.39%.

81 "Key trends and figures in Machine Learning," *Epoch*, accessed June 24, 2023, *https://epochai.org/trends*.

5. Individual risk areas

Below, we summarize the main forecasts and rationales provided by participants across the XPT for each major risk area.

Nuclear weapons

Summary

Forecasters in the XPT believed that nuclear weapons continue to threaten humanity. The median nuclear expert predicted an 8% chance of nuclear weapons causing a catastrophic outcome by 2100 (defined as the death of more than 10% of humans within a 5-year period), while superforecasters predicted a 4% risk of this catastrophic outcome. Nuclear proliferation is predicted to slowly continue, with Iran and Saudi Arabia identified as the most likely new nuclear powers by 2030 and by 2050. Russia is the most likely nuclear power to use nuclear weapons aggressively, and the risk of a Russian nuclear launch is higher because of the war in Ukraine. In the event of a nuclear exchange, nuclear winter is more likely to lead to a catastrophic outcome than the direct effects of nuclear war, according to XPT forecasters, but substantial uncertainty exists about models of nuclear winter. The role of advanced AI in influencing nuclear risk is also uncertain. Compared to nuclear experts, superforecasters predicted more nuclear proliferation but lower risks of nuclear war.

Background

We asked forecasters to predict how many countries would have nuclear weapons, how many total nuclear warheads would exist, and which actors would be most likely to use nuclear weapons at various points in time. We also asked about the risk of nuclear weapons killing more than 1,000 people or leading to a catastrophic outcome or human extinction. These questions focused on the intentional use of nuclear weapons, excluding risks such as nuclear power plant or fuel facility accidents, use of dirty bombs, or accidental detonation of nuclear weapons.

Forecast	Group	Median Estimate		
		2030	2050	2100
Nuclear catastrophic risk	Superforecasters	0.5%	1.83%	4%
	Domain experts	1%	3.4%	8%
Nuclear extinction risk	Superforecasters	0.001%	0.01%	0.074%
	Domain experts	0.02%	0.12%	0.55%

Table 13: *Presents median forecasts from the XPT on questions of whether humanity will experience large-scale death (catastrophic risk) or extinction due to nuclear weapons by 2030, 2050, and 2100. We calculate medians from N=88 superforecasters and N=13 domain experts.*

The final median forecasts of nuclear catastrophic risk and nuclear extinction risk for all three time periods are shown in Table 13. Figure 11 and Figure 12 show the final forecasts from XPT participants and the surveyed members of the public on nuclear catastrophic risk and nuclear extinction risk, respectively. As with forecasts of total catastrophic and extinction risk, the lowest group median forecast was from superforecasters. The highest group median forecasts were from participants in the public survey. However, there was significant variation in forecasts within groups.

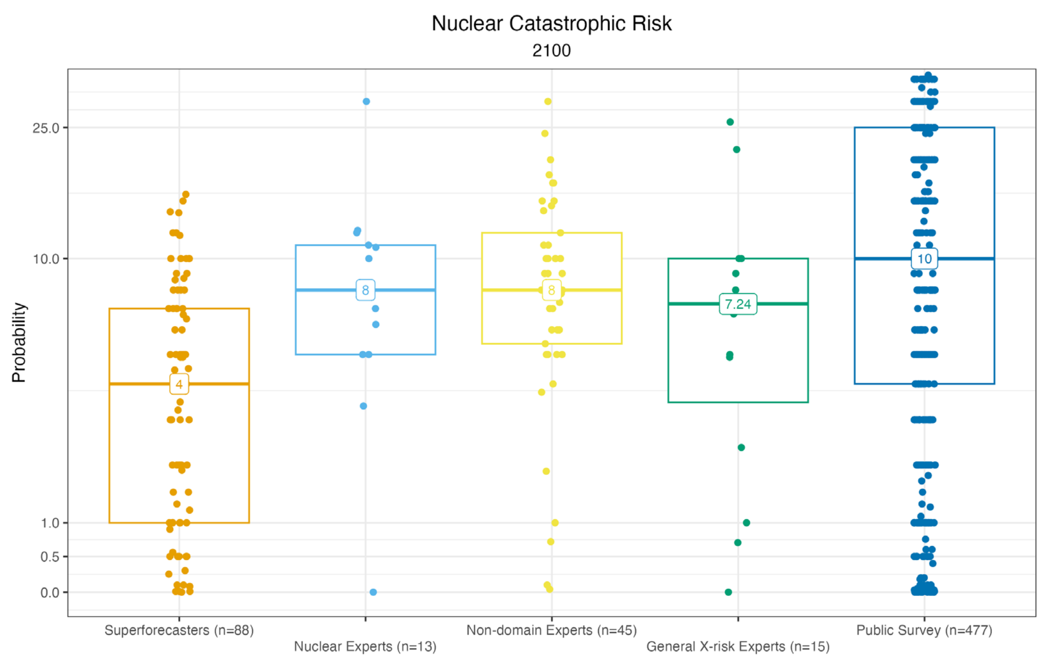


Figure 11: *Plot shows the distribution of final forecasts of nuclear catastrophic risk (by 2100) for each of four groups in the XPT (superforecasters, nuclear experts, non-nuclear experts, and general x-risk experts) as well as a public survey of college-educated respondents from an online platform. Box plots show the 25–75th percentile range (boxes) and the median (labeled) for each group. Within each group, points are jittered slightly horizontally to show density, but this jittering has no empirical purpose.*

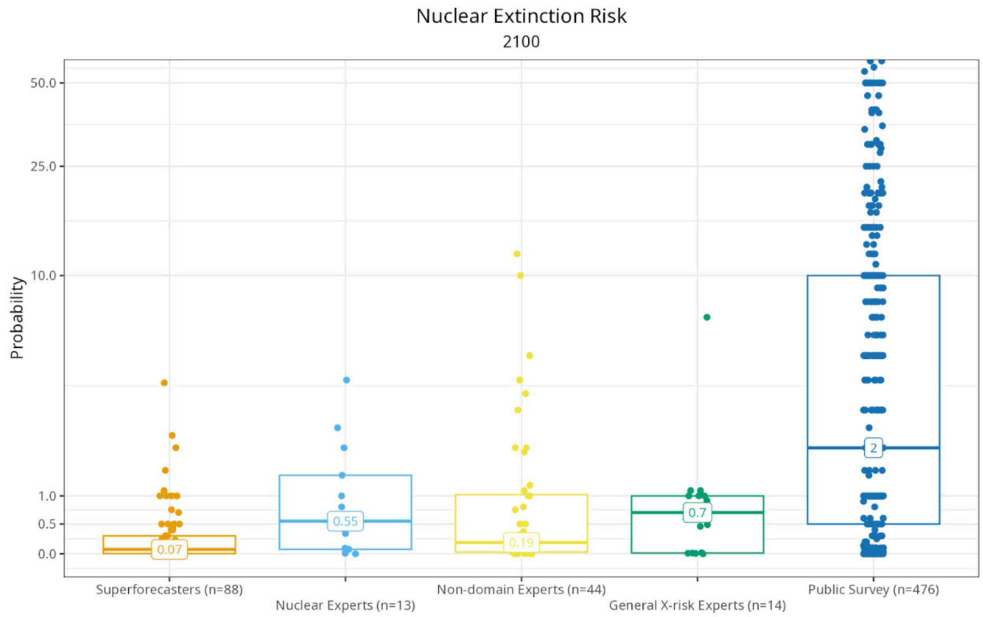


Figure 12: *Plot shows the distribution of final forecasts of nuclear extinction risk (by 2100) for each of four groups in the XPT (superforecasters, nuclear experts, non-nuclear experts, and general x-risk experts) as well as a public survey of college-educated respondents from an online platform. Box plots show the 25–75th percentile range (boxes) and the median (labeled) for each group. Within each group, points are jittered slightly horizontally to show density, but this jittering has no empirical purpose.*

Key nuclear risk takeaways

Nuclear proliferation forecasts[82]

- Forecasters identified Iran as the most likely entrant to the nuclear club, and forecasters expected that Saudi Arabia would pursue nuclear weapons if Iran acquires them.
- Both superforecasters and nuclear experts expected no change to the current number of nine nuclear powers by 2024; they predicted a median of 10 nuclear powers by 2030, and nuclear experts predicted 10 nuclear powers by 2050, while superforecasters predicted eleven.
- For all time periods we asked about, superforecasters predicted the number of nuclear weapons in military inventories would increase, while nuclear experts predicted this number would decrease. For example, by 2040, superforecasters predicted a median of 13,500 nuclear warheads (compared to the early-2022 baseline of 12,705 warheads), while nuclear experts' median prediction was a decrease to 11,990 warheads.

82 Takeaways in this section came from questions on the Countries with Nuclear Warheads and Total Nuclear Warheads.

- In their rationales for forecasts about nuclear proliferation, forecasters described complete disarmament by any current nuclear power as very unlikely (without providing a specific quantitative forecast), and unilateral disarmament was viewed as essentially impossible.
- Arguments for an increasing number of nuclear weapons included the value of nuclear weapons as deterrence and the possibility that Chinese aggression or the decline of American security guarantees might incentivize more countries to pursue nuclear weapons.
- Arguments for a decreasing number included the possibility that North Korea could cease to exist; that deterrence could be achieved with other weapons; and the inherent technical and diplomatic difficulty of acquiring nuclear weapons for actors that do not already have them.

Which actors are most likely to use nuclear weapons?[83]

- Russia was deemed by far the most likely to use nuclear weapons against another nuclear power by 2030, with median forecasts of 1.87% by nuclear experts and 1% by superforecasters.
    - » Forecasters believe this risk will increase if Russia cannot achieve victory in Ukraine by conventional means or if the Vladimir Putin regime views itself as endangered.
- Because of its leadership's perceived irrationality, forecasters identified North Korea as the most likely actor, behind Russia, to use nuclear weapons against another nuclear power by 2030. Nuclear experts predicted a 0.89% median chance of this outcome and superforecasters predicted 0.35%.
- Forecasters described a conflict over Taiwan as the most likely cause of Chinese nuclear weapons use. This conflict could escalate to include the US and could lead to a new Asian arms race, with Japan and South Korea potentially motivated to acquire nuclear weapons. Median forecasts of the chance of China using nuclear weapons against another nuclear power by 2030 were 0.2% for nuclear experts and 0.15% for superforecasters.
- If Israel uses nuclear weapons, forecasters expected it would likely be against Iran and in response to Iran's development of a nuclear weapons program.
- Nations with other options (such as large armies or conventional weapons) are seen as less likely to use nuclear weapons.

83 Takeaways in this section came from the question on Country-by-Country Nuclear Use.

- Terrorists (or other non-state actors) are seen as unlikely to successfully acquire and use nuclear weapons because of their lack of technical capability, the difficulty of acquiring weapons technology, and the existence of easier tools for spreading terror (such as dirty or conventional bombs). A state actor intentionally supplying weapons to terrorists is the only likely scenario in which terrorists could acquire nuclear weapons.
    - » For these reasons, nuclear experts estimated a very low median chance of 0.000005% (with a range of 0–0.01%) that a non-state actor would use nuclear weapons against a nuclear power by 2030; superforecasters gave a higher median chance of 0.02%.
- France and the UK are believed to be the least likely of existing nuclear powers to use nuclear weapons against another nuclear power by 2030. Nuclear experts predicted a 0.02% median chance of this outcome for France and 0.01% chance for the UK, while superforecasters gave a 0.001% median chance for France and 0.01% chance for the UK.

Risk of nuclear war[84]

- The direct effects of nuclear weapons are very unlikely to cause human extinction or a catastrophic outcome, according to XPT forecasters.
    - » A regional, small-scale nuclear exchange (such as one between India and Pakistan) would not result in a "catastrophic outcome," which we defined as a scenario in which nuclear weapons are the cause of death within a 5-year period for more than 10% of humans alive at the beginning of that period.
    - » What kind of target military strategists would select during a nuclear exchange was a source of uncertainty for forecasters. The death toll is expected to be higher if dense cities or other civilian areas are targeted, as opposed to military or remote targets. If the exchange only involves tactical weapons, the expected death toll would be lower.
- Nuclear winter is seen as the only plausible mechanism for human extinction due to nuclear war. Forecasters identified substantial scientific uncertainty about models of nuclear-fallout-induced climate change and the likelihood of human adaptation to nuclear winter.

84 Takeaways in this section came from questions on Nuclear Catastrophic Risk and Nuclear Extinction Risk.

- Superforecasters viewed the overall risk of nuclear war as lower than nuclear experts did.
  - Forecasters predicting lower nuclear risks argued that nuclear weapons are uniquely hard to acquire and use both technically and diplomatically; their use would lead to retaliation and stigma for the user; and that other weapons, including potential new weapons or cyberattacks, would be less costly. Some noted the possibility that, over longer timelines, humanity may learn to be less violent or better at resolving disputes.
  - Forecasters predicting higher nuclear risks pointed to current geopolitical tensions, especially the Russian invasion of Ukraine and the possibility that Russia might turn to nuclear weapons to avoid humiliation or a threat to Putin's regime, if it looks like it is about to collapse. Several forecasters argued that the current risk of nuclear war is unusually high due to the Ukraine situation and potential Chinese aggression over Taiwan. Additional arguments for higher risk included proliferation and associated disruption of the "nuclear taboo," the expiration of the New START treaty,[85] and the difficulty of negotiating a new treaty.
- While many forecasters noted the possibility of AI involvement in nuclear war, there was no consensus on how AI would affect nuclear risk. Forecasters suggested that advanced AI could use nuclear weapons against humans or at the behest of humans; AI could enable the spread of nuclear weapons technology; or AI could impose an arms control regime. Others believed AI would not be given control of nuclear weapons.

Calculating base rates for nuclear war[86]

- Forecasters used different approaches to estimate the risk of an unprecedented event such as nuclear war.
  - The base rate could be viewed as zero because nuclear weapons have not been used since 1945 and the Hiroshima and Nagasaki bombs did not lead to catastrophic outcomes as defined in the question (killing >10% of humans). However, the historical rate of nuclear weapons use may be less predictive given the increase over time in nuclear-armed states.

---

85 "New START Treaty," Nuclear Threat Initiative, accessed July 2nd, 2023, https://www.nti.org/education-center/treaties-and-regimes/treaty-between-the-united-states-of-america-and-the-russian-federation-on-measures-for-the-further-reduction-and-limitation-of-strategic-offensive-arms/.

86 Takeaways in this section came from questions on Nuclear Catastrophic Risk, Nuclear Extinction Risk and Nuclear Weapon Use.

» An alternative statistical method using the formula 2/(5n) for events that have not occurred suggests a 0.51% risk by 2100.[87]

- Incidents when nuclear weapons were almost launched should lead to upward adjustments of the base rate. Disagreement and uncertainty existed about how many truly "close calls" occurred and how to weigh them.

Comparison with past nuclear existential risk forecasts

Forecasters' predictions of nuclear existential risk in this tournament were at the higher end of the range established by previous forecasts of human extinction or other existential risks caused by nuclear war over roughly the next 100 years.[88]

Group	Outcome	Forecast
Aird 2021[89]	Extrapolated median estimate of contribution of nuclear risk to total existential risk by 2100	1.58%
Sandberg & Bostrom 2008[90]	Human extinction risk as a result of all nuclear wars before 2100	1%
XPT domain experts (2022)	Nuclear extinction risk by 2100	0.55%
Hilton and McIntyre 2022[91]	Existential risk from nuclear war by 2122	0.005-1%
XPT superforecasters (2022)	Nuclear extinction risk by 2100	0.074%
Ord 2020[92]	Existential catastrophe via nuclear war by 2120	0.01%
Pamlin and Armstrong 2015[93]	Infinite impact by 2115	0.005%

Table 14: *Presents forecasts on various existential risks due to nuclear war from XPT superforecasters (2022), XPT domain experts (2022), Global Catastrophic Risks Survey (2008), Pamlin and Armstrong (2015), Toby Ord (2020), Michael Aird (2021), and Benjamin Hilton and Peter McIntyre (2022). XPT results are medians calculated from N=88 superforecasters and N=13 domain experts.*

87 John Quigley, Matthew Revie, and Jesse Dawson, "Estimating risk when zero events have been observed." *BMJ Quality and Safety*, 22, no. 12, (2013):1042-1043, https://doi.org/10.1136/bmjqs-2013-002246.

88 We asked forecasters about the probability that one or more incidents involving nuclear weapons will cause human extinction or reduce the global population below 5,000, while some of the predictions in this table were for other existential risks like civilizational collapse, unrecoverable dystopia, or infinite impact. Pamlin and Armstrong define "infinite impact" as "When civilisation collapses to a state of great suffering and does not recover, or a situation where all human life ends."

89 Michael Aird, "Database of Nuclear Risk Estimates (draft)," accessed June 24, 2023, https://docs.google.com/spreadsheets/d/1nRqgN6T-ZiOauBCD8ZlRRb5MzzfWXiyYeK-j5xs2fuU/edit#gid=1806424488.

90 Sandberg and Bostrom, "Global Catastrophic Risks Survey," 1.

91 Benjamin Hilton and Peter McIntyre, "Nuclear war," 80,000 Hours, accessed June 24, 2023, https://80000hours.org/problem-profiles/nuclear-security/.

92 Ord, *The Precipice*, 167.

93 Pamlin and Armstrong, "12 Risks that threaten human civilisation," Global Challenges Foundation (2015).

Artificial intelligence

Summary

Superforecasters and AI domain experts disagreed about the risk of extinction due to AI, but both groups assigned meaningful probability to the outcome and predicted that AI was the most likely cause of extinction among the options included in the XPT. Superforecasters were more skeptical than domain experts: the median superforecaster gave a 0.38% risk of extinction due to AI by 2100, while the median AI domain expert gave a 3.9% risk of extinction.

There was broad agreement among forecasters that there will be significant advances in AI capabilities in the coming decades, though superforecasters generally expected longer timelines for both advanced AI and narrower capabilities progress. Domain experts' median prediction for the date of the first publicly-known "advanced AI" system[94] was 2046, while superforecasters' median prediction was 2060.

Predictions about key "inputs" to AI progress, such as investment in R&D and the cost of compute used in the largest training runs, were relatively similar across the two groups (with superforecasters predicting slightly slower growth). Superforecasters were much more skeptical of radically transformative economic impacts from AI.[95]

Finally, we observed little convergence in the groups' forecasts, even after tournament participants were exposed to each others' rationales and placed on teams where they could interact.

---

94 For the purposes of this forecast, "advanced AI" was defined as a system about which all of the following are true (see further details on benchmarks and resolution criteria here):
"1) Able to reliably pass a 2-hour adversarial Turing test.
2) High competency at answering questions across diverse fields of expertise.
3) High competency on interview-level problems in the APPS benchmark.
4) Able to learn the classic Atari game "Montezuma's revenge" in the equivalent of 100 hours or less of real-time play."
Whether a given system qualifies as "advanced AI" is inherently somewhat vague. A number of terms commonly used in the literature seem to us to gesture at related, though non-identical, concepts—e.g. "artificial general intelligence (AGI)", "transformative artificial intelligence (TAI)", and "superintelligence", among others. These terms usually indicate (at minimum) an AI which can perform nearly all human tasks and which has a profound, transformative impact on the world. The definition of "advanced AI" we used for this tournament is somewhat weaker than these (e.g. it doesn't require a system capable of performing all human tasks), but it is inspired by questions on Metaculus that were intended to be a more resolvable way to capture a similar degree of progress. (These questions are: "Date of Artificial General Intelligence," Metaculus, accessed July 3rd, 2023, https://www.metaculus.com/questions/5121/date-of-artificial-general-intelligence/; "Date Weakly General AI is Publicly Known," Metaculus, accessed July 3rd, 2023, https://www.metaculus.com/questions/3479/date-weakly-general-ai-is-publicly-known.) We see it as debatable how far "advanced AI" as defined here is from e.g. "TAI", but our other questions are intended to provide more context on timelines for degrees of AI progress.
95 See forecasts on Probability of GDP Growth over 15%.

Background

Advanced AI has been cited as a potential source of catastrophic and extinction risk by a number of thinkers; for instance, Toby Ord's *The Precipice* argues that misaligned artificial intelligence is the single most probable cause of human extinction this century. In a 2022 survey of AI researchers, respondents gave a 5% median chance of future AI causing human extinction.[96]

Potential risks from advanced AI are often broken into misalignment and misuse risks.[97] AI misalignment refers to situations in which an AI system fails to learn the goals or policies its developers intended for it to possess, and as a result takes unexpected and potentially harmful actions. AI misuse refers to the intentional use of AI systems by humans to achieve harmful objectives (such as cyberattacks, invasive surveillance, or producing and spreading misinformation). Concerns about the worst potential outcomes from AI (i.e., catastrophic and extinction-level risks) typically focus on misalignment, though misuse risks could also be very serious.

We solicited forecasts from superforecasters and AI experts[98] on a wide range of AI outcomes, including: AI-related catastrophic and extinction risks; the development of advanced AI; narrow AI capabilities progress; growth in key inputs to AI development; and economic impacts of AI. Results for each of these topics are summarized in the following sections.

---

96 More specifically, "of future AI advances causing human extinction or similarly permanent and severe disempowerment of the human species" from Grace and Weinstein-Raun, "2022 Expert Survey on Progress in AI".
97 On the work being done on misalignment and misuse, see Kelsey Piper, "There are two factions working to prevent AI dangers. Here's why they're deeply divided," Vox, August, 2022, accessed June 24, 2023, https://www.vox.com/future-perfect/2022/8/10/23298108/ai-dangers-ethics-alignment-present-future-risk.
98 Our AI experts included employees at major AI companies (details excluded to maintain confidentiality). Our pool of experts is somewhat skewed toward people with familiarity and experience with the Effective Altruism community, although we also recruited forecasters via blogs and direct outreach to academic institutions.

Key AI risk takeaways

Catastrophic and extinction risks

Forecast	Group	Median Estimate		
		2030	2050	2100
AI catastrophic risk	Superforecasters	0.01%[99]	0.73%	2.13%
	Domain experts	0.35%	5%	12%
AI extinction risk	Superforecasters	0.0001%	0.03%	0.38%
	Domain experts	0.02%	1.1%	3%

Table 15: *Presents median forecasts from the XPT on questions of whether humanity will experience large-scale death (catastrophic risk) or extinction due to AI by 2030, 2050, and 2100. We calculate medians from N=88 superforecasters for each question, and N=30 domain experts for catastrophic risk and N=29 domain experts for extinction risk.*

The final median forecasts of AI catastrophic risk and AI extinction risk for all three time periods are shown in Table 15. Figure 13 shows the final forecasts from XPT participants and the surveyed members of the public on AI catastrophic risk. The final median forecasts on AI extinction risk from each subgroup are shown in Figure 4 in the "Superforecaster and expert estimates of catastrophic risk" section.

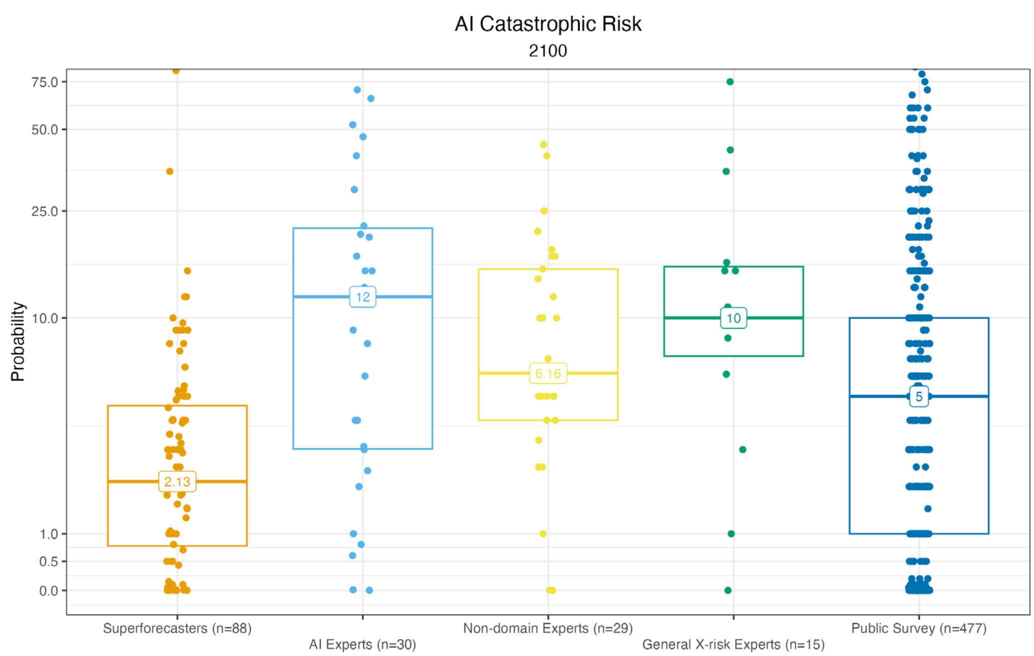


Figure 13: *Plot shows the distribution of final forecasts of AI catastrophic risk (by 2100) for each of four groups in the XPT (superforecasters, AI experts, non-AI experts, and general x-risk experts), as well as a public survey of college-educated respondents from an online platform. Box plots show the 25–75th percentile range (boxes) and the median (labeled) for each group. Within each group, points are jittered slightly horizontally to show density, but this jittering has no other empirical purpose.*

99 Note that forecaster calibration for very low (<1%) probabilities is difficult to validate; we're actively working to determine the most useful way of treating predictions like this.

- AI domain experts predicted higher chances than superforecasters did of AI-related catastrophic risk and extinction risk, in both the near and long term. The difference is more pronounced for extinction risk than for catastrophic risk, and more pronounced for near-term than long-term forecasts.
- Notably, both groups ranked AI as the most likely cause of human extinction by the year 2100 (from among the possible causes of extinction included in the XPT – AI, nuclear weapons, biorisk, and non-anthropogenic causes). See Table 2 for a comparison of forecasts across all extinction risks.
- Key sources of disagreement included:
  - » Whether sufficiently advanced AI would be developed within the relevant timeframe (see Timelines below for some related arguments).
  - » Whether advanced AI would be misaligned. Most teams with higher-risk forecasts claimed that the AI alignment problem is difficult and might not be solved before advanced AI is developed. Several forecasters noted that there is far more funding and research effort being put toward AI capabilities than toward AI safety.
  - » Whether humans would choose to deploy/empower advanced AI. Arguments for lower-risk forecasts suggested that we might not give AI control over critical systems and/or that if AI posed a serious risk, we would regulate its use and integration into society. Higher-risk forecasts argued that AI would likely be given control over critical systems or, even if it weren't, a misaligned AGI could seize control of them.
  - » Whether there are plausible mechanisms by which an advanced AI could cause a catastrophe (death of >10% of humans) or human extinction. Some forecasters argued that human extinction, in particular, would be very difficult to achieve even for an AGI which was actively trying to do so.
- See Appendix 7 for more detail on forecasters' rationales.

Forecast	**Group**	**Median Estimate**
Date of first publicly-known advanced AI	Superforecasters	2060
	Domain experts	2046

Table 16: *Presents median forecasts from the XPT on the date of the first publicly known advanced AI. We calculate medians from N=32 superforecasters and N=9 domain experts.*

Forecast	Year	Group	Median Estimate
Nick Bostrom affirms the existence of AGI by…	2030	Superforecasters	1%
		Domain experts	9%
	2050	Superforecasters	20.5%
		Domain experts	46%
	2100	Superforecasters	74.75%
		Domain experts	87%

Table 17: *Presents median forecasts from the XPT on the probability that Nick Bostrom will affirm the existence of AGI by a given year. We calculate medians from N=32, N=28, and N=27 superforecasters (for 2030, 2050, and 2100 forecasts, respectively) and N=5 domain experts (for all forecasts).*

- AI domain experts predicted nearer dates than superforecasters did for both the advent of advanced AI[100] and for progress on various narrow AI capabilities (e.g., performance on math and language comprehension tests – see below).
- Common arguments for advanced AI **sooner** included:
  - » The recent rapid progress in AI in general.
  - » Placing substantial weight on the "scaling hypothesis", which states (roughly) that continued scale-up of existing AI architectures and training methods, along with growth in inputs such as computing power, will be sufficient to produce advanced AI[101].
- Common arguments for advanced AI **later** included:
  - » Lower confidence in the scaling hypothesis, often accompanied by an expectation that fundamental "breakthroughs" in AI architectures, training methods, etc., might be required to produce advanced AI.
  - » An expectation that "generalization"—i.e., a single, unified AI system being able to perform well across many different types of tasks—might prove much more difficult than progress in narrower domains (and may not flow naturally from research that produces progress in individual domains).
  - » AI development might be slowed by exogenous factors, such as (non-AI-related) catastrophes or major economic downturns, or systemic constraints such as regulation of AI companies, slow-changing consumer demand for AI products, and limited relevant energy availability.

100 See footnote 94.

101 Jared Kaplan et al., "Scaling Laws for Neural Language Models," *arXiv.org*, (Jan 23, 2020), https://doi.org/10.48550/arXiv.2001.08361.

- It was common for forecasters to reference previous forecasts/expert opinion, in support of both longer and shorter timelines:
    - Several forecasters noted that recent progress has consistently been "faster than expected," as indicated by, e.g., other forecasts for the date of advanced AI (such as Metaculus'[102]) becoming shorter over time. Some also noted that, generally, people who have thought about this topic deeply appear to predict shorter timelines (e.g., Ajeya Cotra's biological anchors report[103]).
    - Others argued that, over the longer term, futurist-style predictions have often been overly optimistic about the speed of progress, and some forecasters claimed that Metaculus' AI predictions in particular have an optimistic bias.

Narrow AI progress

Forecast	Group	Median Estimate
First year in which AI wins International Mathematical Olympiad gold medal	Superforecasters	2035
	Domain experts	2030
First year in which AI has written at least 3 New York Times best-selling books	Superforecasters	2050
	Domain experts	2038

Table 18: *Presents median forecasts from the XPT on when An AI will win an International Mathematical Olympiad gold medal (calculated from N=36 superforecasters and N=6 domain experts) and when will an AI have written at least three New York Times best-selling books (calculated from N=37 superforecasters and N=6 domain experts)*

- We elicited forecasts about progress on a number of narrow AI capabilities. Part of our motivation was to produce forecasts which will resolve relatively soon as a way of getting an early indicator of forecaster accuracy (see "Plans for follow-up" below). AI domain experts generally predicted nearer dates than superforecasters did for these capabilities benchmarks, though the difference was not always very substantial.
- For forecasts on additional short-term capabilities improvements, see the summaries of forecasts of AI performance on the MATH Dataset Benchmark, Massive Multitask Language Understanding Benchmark, and QuALITY Dataset Benchmark.

---

102 " When will the first weakly general AI system be devised, tested, and publicly announced?," Metaculus, accessed July 2nd, 2023, https://www.metaculus.com/questions/3479/date-weakly-general-ai-is-publicly-known/.

103 Ajeya Cotra, "Forecasting TAI with biological anchors," accessed July 3rd, 2023, https://docs.google.com/document/d/1IJ6Sr-gPeXdSJugFulwIpvavc0atjHGM82QjIfUSBGQ/edit.

Key inputs

Forecast	Group	Median Estimate		
		2024	2030	2050
Investment by US companies in R&D in AI-related industries[104]	Superforecasters	$192.5 billion	$370 billion	$696 billion
	Domain experts	$240 billion	$605 billion	$785 billion
Total cost of compute for the largest AI experiment run	Superforecasters	$35 million	$100 million	$300 million
	Domain experts	$65 million	$180 million	$800 million
Maximum compute used in an AI experiment (in petaFLOPS-days)	Superforecasters	100,000	6 million	700 million
	Domain experts	420,680	25.1 million	2.5 trillion

Table 19: *Presents median forecasts from the XPT for three questions related to AI capabilities: amount of investment by US companies in R&D in AI related industries (calculated from N=34 superforecasters and N=2 domain experts), total cost of compute for the largest AI experiment run (calculated from N=31 superforecasters and N=7 domain experts), and the maximum amount of compute used in a single AI experiment (calculated from N=33 superforecasters and N=2 domain experts). For each question, we show median forecasts for 2024, 2030, and 2050.*

- Note that both groups expect relatively similar levels of investment in AI-related R&D, and relatively similar costs for training compute through 2030 (though the median domain expert's forecast is substantially higher by 2050). Domain experts expect more compute to be used in AI experiments and expect an extremely large increase in compute use by 2050, though it's important to note that the cohort of domain experts for this question was very small (N=2).
- For further discussion of inputs, see summaries of forecasts of Maximum Compute Used in an AI Experiment, Lowest Price of GFLOPS, and Largest Number of Parameters in a Machine Learning Model.

Economic impacts

Forecast	Group	Median Estimate		
		2024	2030	2050
Percentage of US GDP resulting from software and information services	Superforecasters	3.45%	4.9%	7.71%
	AI domain experts	3.6%	5.5%	9%
Labor force participation rate in OECD	Superforecasters	78.25%	77.75%	78%
	AI domain experts	77.2%	74.65%	70.9%
		Median Estimate		
Likelihood of annual global GDP growth >15% before 2100	Superforecasters	2.75%		
	AI domain experts	25%		

Table 20: *Presents median forecasts from the XPT for three questions related to the economic impact of AI: the percentage of US GDP resulting from software and information services in 2024, 2030, and 2050 (calculated from N=30 superforecasters and N=7 domain experts), labor force participation rate in OECD countries in 2024, 2030, and 2050 (calculated from N=32 superforecasters and N=4 domain experts), and likelihood of annual GDP growth >15% before 2100 (calculated from N=34 superforecasters and N=7 domain experts).*

104 Namely, 'Information' and 'Computer systems design', as defined by the National Science Foundation's "Business Research and Development" reports. See AI R&D Spending.

- Superforecasters and domain experts made very similar predictions about the proportion of GDP due to AI-related industries and about labor force participation rates (prior to 2050, when domain experts expect a steep drop-off). Domain experts think extreme GDP growth (>15% in one year) at any point prior to 2100 is much more likely than superforecasters do.
- For further discussion of economic impacts, see summaries of forecasts of GPT Revenue and US Computer R&D Development.

In 2024, when the earliest of these forecasts are set to resolve, we plan to evaluate which forecasters were the most accurate, and possibly conduct additional analysis into whether any specific methods, heuristics, or styles of argument tended to produce more accurate predictions. It seems plausible that we should put more weight on the medium- and long-term predictions of forecasters who prove more accurate in the short term (to the extent that we believe forecasting accuracy is correlated across various timeframes).

Climate change

Summary

On climate change, forecasters' predictions for global surface temperature change aligned most closely with the Intergovernmental Panel on Climate Change (IPCC)'s "middle of the road" scenario. Forecasters expected that solar and wind energy will provide 14% of the world's energy by 2024 and 24% by 2030. Forecasters' predictions of the decrease in cost of solar energy were slightly more optimistic than the U.S. Department of Energy's SunShot 2030 goal.[105] Commercial-scale nuclear fusion power was predicted for 2075, although forecasters noted a chance that nuclear fusion will never be achieved at commercial scales. Forecasters expected that the cost of hydrogen from renewable sources will continue to decrease and that the amount of carbon dioxide (CO2) captured and stored by direct air capture will increase. However, forecasters' predictions indicated that the International Energy Agency's Net Zero by 2050[106] targets for direct air capture and for the cost of hydrogen in 2030 will not be achieved.

105 "SunShot 2030," Department of Energy, accessed June 25, 2023, https://www.energy.gov/eere/solar/sunshot-2030

106 "Net Zero by 2050: A Roadmap for the Global Energy Sector," International Energy Agency (May 2021), accessed June 25, 2023, https://www.iea.org/reports/net-zero-by-2050

Background

Based on consultations with climate change research organizations, we asked forecasters to predict: global surface temperature changes, the magnitude of annual direct air carbon dioxide capture, the cost of hydrogen produced by renewable energy, the cost of solar energy, and the growth of commercial-scale nuclear fusion power and of solar and wind energy. We did not ask about catastrophic or extinction risk due to climate change based on expert recommendations because the impacts would be too slow-moving to meet our "catastrophic" threshold (10% of humans dying within a 5-year period), and in pilot interviews, climate experts told us they would place an extremely low probability on extinction risk from climate change. Instead, experts believe that negative effects of climate change will occur over a protracted period of time, and may be indirect, which puts their effects largely beyond the scope of our study.

Key climate change risk takeaways

Global temperature changes[107]

What will be the global surface temperature change as compared to 1850–1900, in degrees Celsius?

Forecast	Group	Median Estimate		
		2030	2050	2100
Average global surface temperature	Superforecasters	1.47°C	2°C	2.6°C
	AI domain experts	1.4°C	2°C	2.55°C

Table 21: *Presents median forecasts from the XPT on the question of average global surface temperature change in degrees Celsius in 2030, 2050, and 2100. We calculate medians from N=33 superforecasters and N=8 domain experts.*

- Forecasts used the IPCC's predictions as a starting point.[108] Median forecasts were closest to the IPCC's "middle of the road" scenario (SSP2-4.5), except that the range of forecasters' predictions for 2050 and 2100 was broader than the range in any IPCC scenario.
- Potential "tipping point scenarios" noted by forecasters may explain the increased range and uncertainty in longer-term predictions.

107 These points come from the forecasts of Average Global Surface Temperature.
108 Specifically, teams referred to the Technical Summary in *Climate Change 2021: The Physical Science Basis*, IPCC (August 9, 2021), https://www.ipcc.ch/report/ar6/wg1/.

- Forecasters described uncertainty about the extent of nations' commitment to decarbonization and emissions reduction, especially given that not all nations will be equally affected by global warming and some may even benefit. More optimistic forecasters pointed to current government incentives for renewable energy and carbon capture, such as the U.S. Inflation Reduction Act of 2022.[109] Others were unsure of the impact and durability of these incentives.
- In more optimistic rationales, forecasters cited the fact that the relative cost of renewables is expected to decrease both because of decreases in renewable energy cost and increases in fossil fuels cost.

Renewable energy[110]

- Solar and wind power
  - » Forecasters' median predictions were that, by 2024, solar and wind combined would provide 15% of world energy and 23.5% by 2030.
  - » Compared to the U.S. Department of Energy's SunShot 2030 goal of $0.03/kWh, XPT predictions for the decrease in cost of solar power were slightly more optimistic–perhaps because the previous SunShot 2020 goal of $0.06/kWh was reached ahead of schedule in 2017.[111]
  - » Specifically, forecasters' median predictions were $0.0385/kWh in 2024 and $0.025 in 2030 (both measured in 2017 USD for new utility-scale photovoltaic solar systems above 4MWAC in the United States).
  - » Among negative factors delaying the growth of solar and wind power, forecasters cited regulatory delay, a scarcity of appropriate sites and potential NIMBYism, inflation in materials cost, and the problem of intermittency and incorporating solar and wind power into the existing electrical grid.
  - » China's dominant role in solar panel manufacturing creates uncertainty for solar energy, forecasters said. Geopolitical tension between the U.S. and China and the current polysilicon supply problems could slow the growth of solar power.

109 "Inflation Reduction Act Guidebook," The White House, accessed 2nd July, 2023, https://www.whitehouse.gov/cleanenergy/inflation-reduction-act-guidebook/.
110 These points come from forecasts of Solar and Wind Energy, Cost of Utility-Scale Solar Energy, Cost of Hydrogen, and Nuclear Fusion Energy.
111 Department of Energy, "Sunshot 2030."

- Cost of hydrogen
    - » Forecasters predicted that hydrogen from renewable energy will cost $4.5/kg to produce in 2024 and $2.5/kg in 2030. These numbers are above the International Energy Agency's Net Zero by 2050 target of $1.3/kg in 2030 "in regions with excellent renewable resources,"[112] and also above the U.S. Department of Energy's Hydrogen Shot goal of $1/kg by 2031.[113]
    - » Forecasters described green hydrogen as a potentially especially efficient way to store excess energy produced by renewables.
- Nuclear fusion
    - » Nuclear fusion reactors will deliver 1% of all utility-scale power consumed in the U.S. by 2075, according to median forecasts. Superforecasters predicted a median of 2077, while subject matter experts' median was 2100. The range of these predictions was from 2042 at the 5th percentile to essentially never at the 95th percentile.
    - » Forecasters identified a cleavage of opinion between those who believed that nuclear fusion power is essentially an engineering problem, such that it can be achieved given sufficient resource investment, and those who believed that fusion might be physically impossible on Earth.
- The impact of the Russia-Ukraine war on renewable energy will be mixed, according to forecasters' rationales. The war incentivizes energy independence but also could impede the growth of renewables through supply chain disruption and inflation.
- The economic feasibility of each renewable energy technology is determined in part by its relative cost compared to other renewable energy sources. Solar and wind could lose market share to nuclear fusion, and vice versa. This relative cost was a source of uncertainty in renewable energy predictions, and might suggest that forecasters would view the future prospects of renewable energy in general as brighter than those of any one technology.

Direct air carbon capture[114]

- Forecasters predicted that 0.5 metric tons of CO2 would be captured and stored by direct air capture (DAC) in 2024 and 18 in 2030.

112 International Energy Agency, "Net Zero by 2050: A Roadmap for the Global Energy Sector," 176.
113 "Hydrogen Shot," Department of Energy, accessed June 25, 2023, https://www.energy.gov/eere/fuelcells/hydrogen-shot.
114 These points come from forecasts of Annual Direct Air CO2 Capture.

- In general, forecasters did not believe that the International Energy Agency's Net Zero by 2050 target of capturing more than 85 MT CO2/year by 2030[115] would be met. (This target has since been changed to "almost 60 MT CO2/year by 2030.")
- Forecasters doubted DAC's economic feasibility, citing the lack of commercial uses for captured CO2 and the existence of several less expensive options for reducing emissions.
- Because DAC is a new technology, few prior forecasts exist. Changes in political priorities, subsidies, and taxation decisions were identified as sources of uncertainty.

Biorisks

Summary

XPT participants were also asked to forecast questions relevant to catastrophic biological risks. The main portion of the XPT did not include questions asking directly about the probability of catastrophe and extinction from biological agents due to concerns about generating information hazards—the idea that making information about a risk more accessible might simultaneously make it more likely.[116] However, participants were asked to briefly provide predictions of these risks before and after the XPT, without providing rationales. For these questions, they did not go through the same five-stage process that comprised the main portion of the XPT.

Compared to superforecasters, biosecurity experts predicted higher catastrophic and extinction risks from engineered pathogens, both before and after the XPT. Biosecurity experts predicted approximately a 4% risk of catastrophe due to engineered pathogens by 2100, and a 1% extinction risk. Median superforecaster predictions were approximately 1% and 0.01% (after the XPT), respectively. Superforecasters became less concerned with these risks from engineered pathogens over the course of the tournament, but the predictions from biosecurity experts remained stable. Predictions were lower for risks from natural pathogens.

The superforecaster median predictions for a pathogen causing the death of >1% of the population in a five-year period were similar for genetically-engineered- and non-genetically-engineered-pathogens (1.5% and 1.69%, respectively by 2050). Compared to superforecasters, the median expert predictions were higher for both types of pathogen, and this group predicted greater risk from genetically-engineered rather

115 International Energy Agency, "Net Zero by 2050: A Roadmap for the Global Energy Sector."
116 Nick Bostrom, "Information hazards: A typology of potential harms from knowledge."

than non-genetically-engineered pathogens (8% and 5% by 2050, respectively). These estimates are consistent with the lower-end of estimates of historical base rates for a non-genetically-engineered pathogen causing this level of mortality.[117] The median superforecaster prediction for the number of times that the World Health Organization (WHO) declares a Public Health Emergency of International Concern (PHEIC) that results in ≥10,000 deaths was also consistent with base rates[118] (7 events by 2050), but the median expert prediction was lower (5 events).

Concerns about catastrophic biological risks are often motivated by the possibility of a particularly dangerous pathogen entering the population either as a result of an unintentional laboratory accident or the intentional release of a biological weapon (bioweapon). However, when asked about these scenarios, forecasters generally made low predictions of them occurring before 2050. Superforecasters and biosecurity experts both predicted less than one instance of a lab-escaped pathogen causing the death of at least 1,000 people by 2050. Both groups also predicted less than one event

in which a bioweapon causes more than 100,000 human deaths by 2050, with the median domain expert prediction of zero such events (for both state and non-state actors). However, bioweapons use causing more than 1,000 deaths was thought more likely, and the median predictions of both groups were higher than historical base rates.[119] Domain experts predicted 1.4 such events by 2050 caused by state actors, and two such events by non-state actors. Superforecasters predicted one such event by 2050 caused by a state actor, and one by a non-state actor.

Domain experts predicted 10.5 countries would have bioweapons programs by 2050; superforecasters predicted 7. When asked about 7 particular countries, Russia and North Korea were considered the most likely of these to be thought to have a biological weapons program, and the US the least likely.

---

117 There are arguably three plausible instances of a pathogen killing >1% of the population in a five-year period: smallpox at some point between the beginning of its global spread and the introduction of vaccination in the late 18th century, the Black Death in the 14th century, and the 1918 influenza pandemic. Depending on which of these are accepted and the dates used, base rates would suggest a probability of roughly between 4 and 9% for this question resolving positively by 2050 and 11 and 23%. See Catherine Thèves, Eric Crubézy and Philippe Biagini, "History of Smallpox and Its Spread in Human Populations," Microbiology Spectrum 4, no. 4, (2016), https://doi.org/10.1128/microbiolspec.poh-0004-2014; German Lopez, "How the world went from 170 million people to 7.3 billion, in one map," Vox, January 30, 2016, https://www.vox.com/2016/1/30/10872878/world-population-map; and Max Roser, "The Spanish flu: The global impact of the largest influenza pandemic in history," Our World in Data, March 4, 2020, https://ourworldindata.org/spanish-flu-largest-influenza-pandemic-in-history#how-many-people-died-in-the-spanish-flu-pandemic.

118 Since the introduction of the term in 2009, there have been 3 PHEIC events associated with >10,000 deaths. See Annaliese Wilder-Smith and Sarah Osman, "Public health emergencies of international concern: a historic overview," *Journal of Travel Medicine*, 27, No. 8 (December 2020), https://doi.org/10.1093/jtm/taaa227

119 See V. Barras and G. Greub, "History of Biological Warfare and Bioterrorism," *Clinical Microbiology and Infection* 20, no. 6 (June 2014): 497–502, https://doi.org/10.1111/1469-0691.12706.

Background

Several experts in catastrophic and existential risks list biological agents as one of the possible causes of catastrophe in the coming century.[120] Climate change and changes in land-use patterns are expected to increase the risk of pathogens transferring from non-human animals into the human population.[121] There is also concern that advances in biotechnology, particularly synthetic biology, are increasing the risk of an engineered pathogen entering the human population, whether through accident or intentional release as a biological weapon (bioweapon).[122]

To better understand these risks, we asked forecasters to make predictions for several questions relevant to biological risks.[123] These included questions on:

- Risks of large-scale harms from pathogens, including the likelihood that a pathogen results in the death of >1% of the human population, the number of Public Health Emergencies of International Concern that lead to ≥10,000 deaths, the number of pathogen escape events that result in ≥1,000 deaths, and the expected mortality from malaria.
- Biowarfare, including the likelihood that use of a bioweapon (by either a state or non-state actor) results in ≥1,000 or ≥100,000 deaths, the number of countries with bioweapons programs, and whether specific countries will be thought to have bioweapons programs.
- Medical countermeasures, including the uptake of non-coronavirus mRNA vaccines and the likelihood of a novel disease surveillance program.

The risks of a pathogen causing extinction or catastrophe (defined as the death of >10% of the population) were not included as questions in the XPT. This was due to

120 Toby Ord estimates that there is a 1 in 30 chance of existential catastrophe from "engineered pandemics" by 2120. Ord, *The Precipice*, 167. Sandberg and Bostrom 2008, "Global Catastrophic Risks Survey" at 1 found that experts estimated a 30% chance that the "single biggest engineered pandemic" would kill 1 million people by 2100. Lord Martin Rees, Bill Gates, and the Global Challenges Foundation, have all expressed concern about pandemics and biotechnology causing global catastrophe. See John Mecklin, "Martin Rees explains how science might save us," *Bulletin of the Atomic Scientists*, December 22, 2022, https://thebulletin.org/2022/12/martin-rees-explains-how-science-might-save-us/; Joseph Guzman, "Bill Gates, who predicted the pandemic, names the next two monster disasters that could shake our world", *The Hill*, February 11, 2021, https://thehill.com/changing-america/well-being/538426-bill-gates-who-predicted-the-pandemic-names-the-next-two-monster/; *Global Catastrophic Risks 2022* (Stockholm: The Global Challenges Foundation, 2023), https://globalchallenges.org/app/uploads/2023/04/GCFAnnualReport2022-FINAL.pdf

121 Marco Marani et al., "Intensity and Frequency of Extreme Novel Epidemics," *Proceedings of the National Academy of Sciences* 118, no. 35 (August 2021): e2105482118, https://doi.org/10.1073/pnas.2105482118.

122 National Academies of Science, Engineering, and Medicine, *Biodefense in the Age of Synthetic Biology* (Washington, D.C.: National Academies Press, 2018), https://doi.org/10.17226/24890.

123 Full wording of questions are listed in Appendix 5.

concerns that debating and developing forecast rationales may generate hazardous information.[124] However, before and after the XPT, forecasters were asked to predict the probability of extinction and catastrophic risk from pathogens (with separate questions for natural and engineered pathogens) without providing rationales for their forecasts.

Key biorisk takeaways

Catastrophic and extinction risk

Forecast		Group	Pre-XPT	Post-XPT
Catastrophic risk by 2100	Engineered pathogens	Superforecasters	1.3%	0.85%
		Domain experts	3.9%	4%
	Natural pathogens	Superforecasters	1%	1%
		Domain experts	1%	1.5%
Extinction risk by 2100	Engineered pathogens	Superforecasters	0.1%	0.01%
		Domain experts	1%	1%
	Natural pathogens	Superforecasters	0.01%	0.0018%
		Domain experts	0.1%	0.01%

Table 22: *Presents median forecasts from the XPT intake (Pre-XPT) and postmortem (Post-XPT) surveys for the questions of whether humanity will face mass death or extinction due to engineered or natural pathogens by 2100. We calculate medians from 169 forecasters in the intake survey and 136 forecasters in the postmortem survey.*

- Compared to superforecasters, biosecurity experts predicted higher catastrophic and extinction risks from engineered pathogens, both before and after the XPT.
- Forecasts of extinction risk generally decreased after the XPT, with the exception of biosecurity experts' estimate of extinction risk from engineered pathogens, which remained stable.

124 Concerns about hazardous information, or information hazards, are particularly salient for biological risks. See Gregory Lewis et al. "Information Hazards in Biotechnology" Risk Analysis 39, No. 5, (2018): 975-981, https://doi.org/10.1111/risa.13235.

Other large-scale harms caused by pathogens

Forecast	Group	Median Estimate		
		2030	2050	2100
Genetically-engineered pathogen killing >1% population[125]	Superforecasters	0.25%	1.5%	4%
	Domain experts	1.22%	8%	10.25%
Non-genetically-engineered pathogen killing >1% population	Superforecasters	0.5%	1.69%	3.63%
	Domain experts	1%	5%	8.14%
Forecast	**Group**	**2024**	**2030**	**2050**
Number of times WHO declares PHEIC that kills ≥10,000	Superforecasters	0.5	2	7
	Domain experts	0	1	5
Number of events of escaped biological agents killing ≥1000	Superforecasters	0.013	0.15	0.68
	Domain experts	0.027	0.2	0.5

Table 23: *Presents median forecasts from the XPT on questions related to harms caused by pathogens. On the questions of whether a genetically-engineered pathogen or a non-genetically-engineered pathogen will kill more than 1% of the population, we show forecasts for 2030, 2050, and 2100 (calculated from N=89 superforecasters and N=14 domain experts for genetically-engineered pathogen risk, and N=88 superforecasters and N=13 domain experts for non-genetically-engineered pathogen risk). On the questions of how many times the WHO will declare a Public Health Emergency of International Concern that kills more than 10,000 people (calculated from N=33 superforecasters and N=5 domain experts) and how many events of escaped biological agents will kill more than 1,000 people (calculated from N=32 superforecasters and N=4 domain experts), we show forecasts for 2024, 2030, and 2050.*

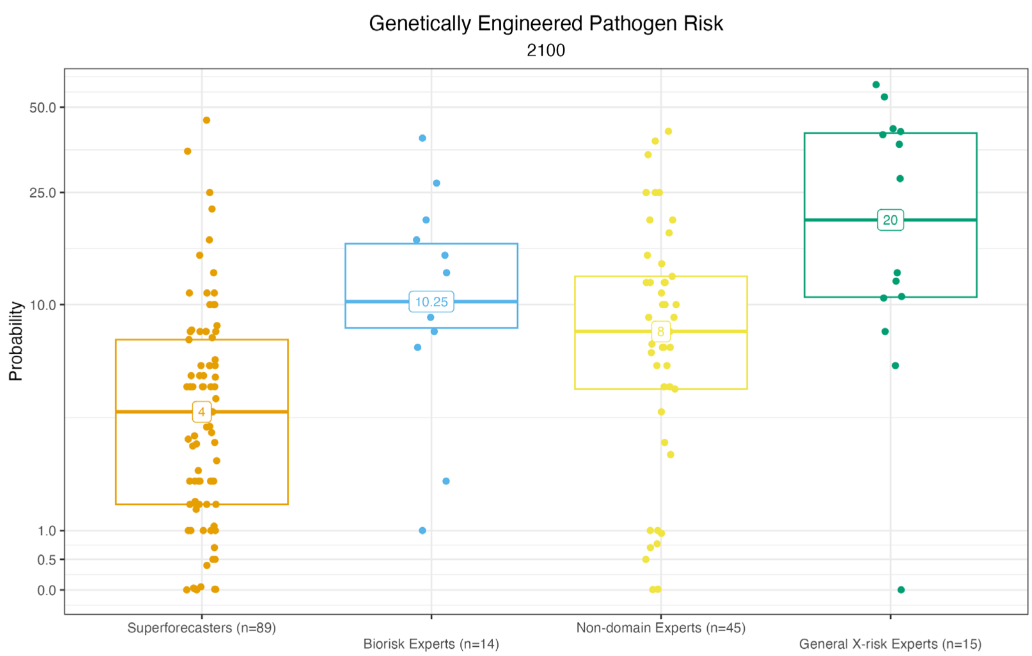


Figure 14: *Plot shows the distribution of final forecasts of a genetically-engineered pathogen causing the death of >1% of the population (by 2100) for each of four groups in the XPT (superforecasters, nuclear experts, non-nuclear experts, and general x-risk experts). Note that this is not showing forecasts on catastrophic or extinction risk, but instead focusing on a lower 1% threshold of deaths. Box plots show the 25-75th percentile range (boxes) and the median (labeled) for each group. Within each group, points are jittered slightly horizontally to show density, but this jittering has no empirical purpose.*

125 This question, and the corresponding question on non-genetically-engineered pathogens, asks about a pathogen causing the death of >1% of the population in a five-year period.

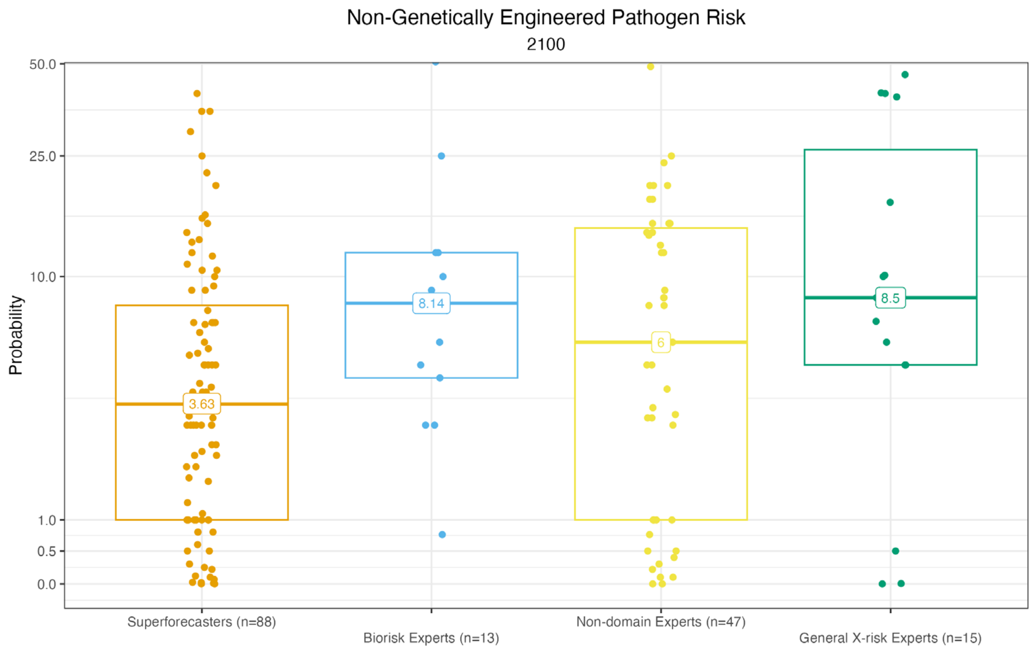


Figure 15: *Plot shows the distribution of final forecasts of a non-genetically-engineered pathogen causing the death of >1% of the population (by 2100) for each of four groups in the XPT (superforecasters, nuclear experts, non-nuclear experts, and general X-risk experts). Note that this is not showing forecasts on catastrophic or extinction risk, but instead focusing on a lower 1% threshold of deaths. Box plots show the 25-75th percentile range (boxes) and the median (labeled) for each group. Within each group, points are jittered slightly horizontally to show density, but this jittering has no empirical purpose.*

- Compared to superforecasters, biosecurity experts estimated a higher probability of a pathogen causing the death of >1% of the population and thought that this was more likely to occur due to a genetically-engineered pathogen than a non-genetically-engineered pathogen (10.25% and 8.14% by 2100, respectively). These estimates are consistent with or slightly lower than the probability suggested by historical base rates.[126] The median superforecaster prediction was lower again and predicted similar risk from genetically-engineered and non-genetically-engineered pathogens (4% and 3.63%, respectively).
- The superforecaster median prediction for the number of times that the WHO declares a PHEIC that results in ≥10,000 deaths was consistent with historical base rates.[127] In contrast to other questions, biosecurity experts made less pessimistic predictions, forecasting a lower number of such events than superforecasters.

126 See footnote 117.

127 Since 2009, there have been 3 PHEIC events associated with ≥10,000 deaths. See Wilder-Smith and Osman, "Public health emergencies of international concern: a historic overview."

- Superforecasters and biosecurity experts both predicted less than one expected[128] event of a contagious lab-escaped pathogen causing the death of at least 1,000 people. There have been no confirmed instances of a laboratory-escaped pathogen causing ≥1,000 deaths.[129]
- Arguments for higher forecasts for these questions commonly referred to technological progress making biological research and bioweapons more accessible and successful. The historical precedent of bioterrorism and biowarfare was also noted.
- However, arguments for lower forecasts suggested that technology would reduce risk (by enabling the development of vaccines and by improving disease surveillance and laboratory biosafety and biosecurity measures) and that few groups have an incentive to develop bioweapons using particularly harmful pathogens.

Biowarfare[130]

Forecast Number of events of:		**Group**	**Median Estimate**		
			2024	**2030**	**2050**
Use of bioweapon causing ≥1,000 deaths	Non-state actor	Superforecasters	0	0.15	1
		Domain experts	0	1	2
	State actor	Superforecasters	0	0.024	1
		Domain experts	0	0.05	1.4
Use of bioweapon causing ≥100,000 deaths	Non-state actor	Superforecasters	0	0.0005	0.038
		Domain experts	0	0	0
	State actor	Superforecasters	0	0.002	0.15
		Domain experts	0	0	0
Country leaders assassinated with contagious bioweapon		Superforecasters	0.0001	0.008	0.4
		Domain experts	0	0	0

Table 24: *Presents median forecasts on the XPT questions of whether a state or non-state actor will use a bioweapon resulting in more than 1,000 deaths (calculated from N=33 superforecasters and N=6 domain experts for state actors and N=39 superforecasters and N=7 domain experts for non-state actors), whether a state or non-state actor will use a bioweapon resulting in more than 100,000 deaths (calculated from N=35 superforecasters and N=4 domain experts for state actors and N=33 superforecasters and N=6 domain experts for non-state actors), and how many leaders of countries will be assassinated using a contagious bioweapon (calculated from N=35 superforecasters and N=3 domain experts). For each question, we show the median forecast for the years 2024, 2030, and 2050.*

128 We use the language "expected number of events" to account for the fact that it may sometimes be unclear whether a pathogen came from a lab escape, such as in the case of Covid-19. So, for all events in which pathogens kill at least 1,000 people, we will ask a panel of experts to estimate the likelihood that the event was caused by a pathogen escaping from a lab. If, e.g., experts give a 30% chance that a pathogen escaped from a lab, it would count as 0.3 expected events for the purpose of this question.

129 David Manheim and Gregory Lewis, "High-risk human-caused pathogen exposure events from 1975-2016 [version 2; peer review: 2 approved]," *F1000Research*, (2022) 10:752 https://doi.org/10.12688/f1000research.55114.2.

130 These points come from forecasts of questions shown in Table 24 and forecasts of the Number of Countries with Biological Weapons Programs, and Individual Countries with Biological Weapons Programs.

- Forecasters made low predictions for the use of bioweapons resulting in ≥ 100,000 deaths and assassinations using a contagious agent. The median prediction of biosecurity experts was zero for these events.[131] Superforecasters also made low predictions: less than one event for both questions. These estimates are largely consistent with historical base rates, with no known instance of a bioweapon causing >100,000 deaths or a contagious bioweapon being used in an assassination. [132]
- Forecasters thought that bioweapons use causing ≥1,000 deaths was more likely, and that the likelihood was similar for state and non-state actors. These forecasts are above historical base rates.[133]
- Biosecurity experts predicted that 10.5 countries would be thought to have active bioweapons programs by 2050. The median prediction from superforecasters was lower at 7 countries.
- Of the list of 7 countries provided, Russia and North Korea were predicted as the most likely to be thought to have bioweapons programs. The median superforecaster prediction was that 80% of an expert panel would, for Russia and North Korea separately, agree that that country had a bioweapons program between 2022 and 2050. The median biosecurity expert prediction was 95% and 92.5%, for Russia and North Korea respectively. The US was considered least likely, with median superforecaster and biosecurity expert predictions of 30% and 8.5% for the same question, respectively. However, teams noted wide variation in predictions from individual forecasters.
- Arguments for lower forecasts for questions on bioweapons often noted the role of the Biological and Toxins Weapons Convention (BWC) in preventing bioweapons programs and highlighted the impracticality of these weapons (including their inability to be targeted, the expense and difficulty of development, and poor track record of effectiveness).
- Arguments for higher forecasts noted the historical precedent of bioweapons programs (including after the BWC entered into force) and the potential value of bioweapons for deterrence. It was also suggested that progress in biotechnology might overcome many of the current deficiencies in bioweapons.

---

131 Forecasters debated whether to submit integer or non-integer values for this question. As a result, a zero value should not be interpreted to mean an absolute zero value in every case so much as it means 'a number less than one'.

132 See Barras and Greub, "History of Biological Warfare and Bioterrorism."

133 Forecasters identified four events that may be consistent with this question: the Siege of Caffa in 1348, the deliberate spread of smallpox amongst Indigenous peoples in Australia and North America by British settlers in the 19th century, and the Japanese use of plague in Quzhou in 1940. For the first three of these events it is difficult to estimate the mortality attributable to weapons, given that disease spread would likely have occurred naturally as well. See Barras and Greub, "History of Biological Warfare and Bioterrorism."

Countermeasures[134]

- Biosecurity experts and superforecasters made similar predictions for the probability of a new disease surveillance system being announced with a >$100 million budget. The median superforecaster prediction that such an announcement would be made before 2030 was 65.8%, and the median expert prediction was 70%.
- Compared to domain experts, superforecasters predicted higher numbers of people receiving at least one non-coronavirus mRNA vaccine by 2024 and 2030, but median predictions were within an order of magnitude. The median superforecaster prediction was 100,000,000 receiving a non-coronavirus mRNA vaccine by 2030. The median prediction from experts was 60,000,000.

Malaria mortality[135]

- Superforecasters and biosecurity experts had similar median predictions for malaria mortality in 2024 and 2030, but the median prediction of biosecurity experts for 2050 was half that of superforecasters (150,000 vs 300,000).
- Arguments for lower forecasts in 2050 include the potential impact of recent developments in malaria vaccines and other interventions to reduce the malaria burden and the impact of economic development improving healthcare and socioeconomic determinants of health.

134 These points come from forecasts of a Novel Infectious Disease Surveillance Program and the number of people receiving a Non-Coronavirus mRNA Vaccine.
135 These points come from forecasts of Malaria Mortality.

6. Next Steps

We see the XPT as an iterative process and we have learned a lot from this first round. Some of these lessons are practical, down-to-earth insights into how to recruit and retain the talent of busy professionals in a demanding multi-month marathon (see "Appendix 1: Practical Problems in Running the XPT"). Other lessons are more substantive. The first iteration of the XPT (let's call it XPT-1) yielded first-of-a-kind, in-depth assessments of how sophisticated specialists and generalists conduct debates about the greatest risks facing humanity and how they translate their beliefs about those risks into quantitative forecasts.

We are also acutely aware that XPT-1 leaves an array of key questions unanswered and we plan to use future iterations of the XPT to produce answers. Our agenda for this project is as follows:

1. Explore the value of early warning indicators from XPT-1
2. Produce better early warning indicators
3. Develop better methods of facilitating productive adversarial collaborations between disagreeing schools of thought
4. Identify and validate better methods of eliciting low-probability forecasts
5. Make these forecasts more relevant to policymakers

Explore the value of early warning indicators from XPT-1

The XPT posed numerous questions about how the risk landscape would evolve in the short-run, including several "early warning" AI-risk indicators such as: "How much will be spent on compute in the largest AI experiment by 2024?" The XPT also made a major effort to identify crux questions that, once forecasters know the answer in 2024, will tell them more about longer-run futures in 2030 or even 2100.

This provides us with an opportunity to ask participants to update their longer-run forecasts based on the outcome of the short-run questions. For example, if spending on AI R&D turns out to be lower or higher by 2024 than a forecaster predicted it would be, will they choose to update their longer-run predictions of AI-related risk?

When we evaluate forecasters' accuracy on short-run questions in 2024, we expect the XPT to also provide new evidence on the relationship between reasoning quality and forecasting accuracy. We captured more than five million words of reasoning, deliberation, and synthesis across the XPT and can analyze that data to determine which types of arguments are related to forecast accuracy across all the areas the XPT covered. For example, we'll be able to compare the reasoning "quality" of optimists with pessimists. This type of analysis could be particularly valuable since it can provide insights beyond forecasting. By studying the properties of arguments associated with forecasting accuracy, we expect to make novel contributions that can be applied to debate across the internet.

Produce better early warning indicators

If we want to put as much emphasis on the quality of questions as on the accuracy of forecasts, we need to incentivize high-quality questions as rigorously as we do forecasting accuracy[136]. Work has already begun to develop formal metrics for what constitutes a high-probative-value question and we plan to build a database of strong candidate questions.

We hope to create a longitudinal panel of forecasters and question-generators (often but not always the same people) that works together over the next 20 years to shed more light on the feasibility of early warning indicators for various topics.

Develop better methods of facilitating productive adversarial collaborations between disagreeing schools of thought

When do deliberation and argumentation lead to consensus in forecasting and when to stalemate, as the XPT-1 found for forecasts of AI extinction risk by 2100? To ensure that failure to converge was not due to fluky method-specific factors, we are developing more tightly choreographed forms of "adversarial collaborations" that have two properties: (1) each side must demonstrate it more fully grasps each major argument of the other side before offering rebuttals, which in turn must be fully grasped by recipients before responding; (2) focus adversarial collaborators on generating shorter-range crux questions that each party agrees ex-ante would move their X-risk probability judgments once the questions are objectively resolved.

136 This is the question equivalent of proper scoring rules for forecasts.

Identify and validate better methods of eliciting low-probability forecasts

Researchers have long known about the instability of tiny probability estimates,[137] that people tend to over-weight close-to-zero probabilities when questioning makes them salient, but underweight or even ignore them otherwise—one of the clearest examples of a Heisenberg effect[138] that experimental psychology has to offer. Some advocates of the Precautionary Principle[139] seize on this claim to argue that if forecasters cannot reliably reason about the difference between 0.001% and 0.000001%, a magnitude difference of 1,000, we should adopt an extremely risk-averse threshold for any technology for which one can make a prima facie X-risk case.

We see this as an empirical question that can be resolved by assessing the skill of well-incentivized top forecasters at: (a) making reliable (non-contradictory) judgments of micro-probability events; (b) making accurate judgments in simulated worlds that permit ground truth determinations. We are also now testing new methods of eliciting micro-probabilities. One such method asks for comparative judgments of the sort often used in psychophysical scaling—e.g., How much lower or higher is your extinction risk estimate than an anchor or comparison value?

Make these forecasts more relevant to policymakers

The most direct solution for ensuring the policy relevance of high-stakes forecasts is to shift the focus of XPT-style elicitation from event-focused forecasting to policy-conditional forecasting. Instead of just asking "How likely is Y?" we can ask "How likely is Y if society goes down this or that policy path?" We can then input those probabilities into a preferred cost-benefit framework. We also recommend experimenting with a new format, Risk Mitigation tournaments[140] designed to accelerate convergence on good policy options using intersubjective incentives: asking two teams of forecasters with strong track records to do their best at predicting the rank order policy preferences of the other team.

---

137 Daniel Kahneman and Amos Tversky, "Prospect Theory: An Analysis of Decision under Risk," Econometrica, Vol. 47, No. 2: 263-292, (March 1979), https://doi.org/10.2307/1914185.

138 This refers to instances where the act of measurement changes the phenomenon under investigation. See Dean Keith Simonton, "Heisenberg Effect," in *Encyclopedia of Research Design*, ed. Neil J. Salkind (Thousand Oaks, CA: SAGE Publications, Inc., 2010), 564.

139 H. Orri Stefánsson, "On the limits of the precautionary principle." Risk Analysis 39, no. 6 (2019): 1204-1222, https://doi.org/10.1111/risa.13265/.

140 Ezra Karger, Pavel D. Atanasov, and Philip Tetlock, "Improving Judgments of Existential Risk: Better Forecasts, Questions, Explanations, Policies," *SSRN Working Paper* (2022). https://doi.org/10.2139/ssrn.4001628.

7. Conclusion

Taken together, we see large-scale implementation of the above five recommendations as a prudent, practical, and scientific response to a debate currently raging along several policy fronts: Are we living in a uniquely vulnerable century—a vulnerability driven by a mix of accelerating technological progress and slow-moving institutions? Or are human beings once again exaggerating how special their place in history is?

This report is not meant to answer the question of "Whom should you trust?" It documents variation in probabilistic beliefs and explanatory rationales on high-stakes issues: who believes what and how similar are those views across groups with different forms of epistemic expertise? And what short-run predictors and cause-effect beliefs best predict forecasts about long-run risks to humanity? In 2025, we will begin to calculate the first real-world accuracy measures for participants in the XPT. But even then, some observers might believe that sharp discontinuities in the flow of history will render short-run forecasting accuracy irrelevant to the long-run. No one knows if these observers are right. It will take 10-20 years before we can even begin to answer the question of whether short-run and long-run forecasting accuracy are correlated over decades. Until then, our discussion provides novel evidence about the beliefs of experts and highly-accurate forecasters when it comes to risks to humanity from nuclear weapons, AI, and biological pathogens. For now, readers must decide for themselves how much weight to give various groups' forecasts. The pure-track record bet might be to give more weight to superforecasters; the more traditional deference-to-expertise bet would do the opposite. Where do you fall, and how much would it take to change your mind?

Bibliography

Aird, Michael. “Clarifying Existential Risks and Existential Catastrophes.” April 2020. Accessed June 23, 2023. https://forum.effectivealtruism.org/posts/skPFH8LxGdKQsTkJy/clarifying-existential-risks-and-existential-catastrophes.

———. “Database of Existential Risk Estimates (or Similar).” Google Docs. Accessed June 23, 2023. https://docs.google.com/spreadsheets/d/1W10B6NJjicD8O0STPiT3tNV3oFnT8YsfjmtYR8RO_RI/edit?usp=embed_facebook.

Atanasov, Pavel D., Ezra Karger, and Philip Tetlock. "Full Accuracy Scoring Accelerates the Discovery of Skilled Forecasters." *SSRN Working Paper*, (2023) https://dx.doi.org/10.2139/ssrn.4357367

Baron, Jonathan, Barbara A. Mellers, Philip E. Tetlock, Eric Stone, and Lyle H. Ungar. “Two Reasons to Make Aggregated Probability Forecasts More Extreme,” *Decision Analysis* 11, no 2 (March 2014):133-145, http://dx.doi.org/10.1287/deca.2014.0293

Barras, V. and G. Greub. “History of Biological Warfare and Bioterrorism.” *Clinical Microbiology and Infection* 20, No.6 (2014): 497–502. https://www.sciencedirect.com/science/article/pii/S1198743X14641744.

Baum, Seth, Anthony Barrett, and Roman V. Yampolsky, “Modeling and Interpreting Expert Disagreement About Artificial Superintelligence,” *Informatica* 41, no. 7 (December 2017): 419-428, https://ssrn.com/abstract=3104645.

Beard, Simon, Thomas Rowe, and James Fox. “An Analysis and Evaluation of Methods Currently Used to Quantify the Likelihood of Existential Hazards.” *Futures* 115 (January 2020):102469, https://doi.org/10.1016/j.futures.2019.102469.

Belfer Center for Science and International Affairs. “Thucydides Trap: An Overview.” Accessed June 23, 2023. https://www.belfercenter.org/thucydides-trap/overview-thucydides-trap.

Bostrom, Nick, “Information hazards: A typology of potential harms from knowledge.” *Review of Contemporary Philosophy* 10, (2011) 44–79.

Bulletin of the Atomic Scientists. “Doomsday Clock.” Accessed June 23, 2023. https://thebulletin.org/doomsday-clock/.

Center for AI Safety. “Statement on AI Risk.” Accessed June 23, 2023. https://www.safe.ai/statement-on-ai-risk.

Chen, Zhi, Anil Gaba, Ilia Tsetlin, and Robert L. Winkler, “Evaluating Quantile Forecasts in the M5 Uncertainty Competition”, *International Journal of Forecasting*, 38, no 4, (October-December 2022): 1531-1545, https://www.sciencedirect.com/science/article/abs/pii/S0169207022000449. D

Cotra, Ajeya. “Forecasting TAI with biological anchors.” Accessed 3rd July, 2023. https://docs.google.com/document/d/1IJ6Sr-gPeXdSJugFulwIpvavc0atjHGM82QjIfUSBGQ/edit

Department of Energy, "Hydrogen Shot," accessed June 25, 2023, https://www.energy.gov/eere/fuelcells/hydrogen-shot.

Department of Energy, "SunShot 2030", accessed June 25, 2023, https://www.energy.gov/eere/solar/sunshot-2030.

Embury-Dennis, Tom. "Former CIA Chief Just Warned of the Three Existential Threats to America." 2017. The Independent. October 27, 2017. https://www.independent.co.uk/news/world/americas/us-russia-nuclear-war-climate-change-biological-weapons-attack-only-existential-threats-cia-chief-michael-morell-a8022941.html.

Epoch. "Key trends and figures in Machine Learning." Accessed June 24, 2023. https://epochai.org/trends.

Forecasting Research Institute. "Forecasting Research Institute." Accessed June 23, 2023. https://forecastingresearch.org.

Frick, Walter. "What a Study of Video Games Can Tell Us about Being Better Decision Makers." *Quartz*. September 6, 2020. https://qz.com/1899461/how-individuals-and-companies-can-get-better-at-making-decisions.

———. "Journalists Wake up to the Power of Prediction Markets." *Nieman Lab* (blog). Accessed June 23, 2023. https://www.niemanlab.org/2022/12/journalists-wake-up-to-the-power-of-prediction-markets/.

The Global Challenges Foundation. *Global Catastrophic Risks 2022.* Stockholm: The Global Challenges Foundation, 2023. https://globalchallenges.org/app/uploads/2023/04/GCFAnnualReport2022-FINAL.pdf.

Gott, J. Richard. 1993. "Implications of the Copernican Principle for Our Future Prospects." *Nature* 363 (6427): 315–19. https://doi.org/10.1038/363315a0.

Grace, Katja and Ben Weinstein-Raun, "2022 Expert Survey on Progress in AI," AI Impacts Wiki (Aug. 2022). https://wiki.aiimpacts.org/doku.php?id=ai_timelines:predictions_of_human-level_ai_timelines:ai_timeline_surveys:2022_expert_survey_on_progress_in_ai.

Guzman, Joseph. "Bill Gates, who predicted the pandemic, names the next two monster disasters that could shake our world." *The Hill*, February 11, 2021, https://thehill.com/changing-america/well-being/538426-bill-gates-who-predicted-the-pandemic-names-the-next-two-monster/.

Hempsell, C.M. "The Investigation of Natural Global Catastrophes," *Journal of the British Interplanetary Society* 57, no. 1-2(Jan. 2004) 2-13. https://research-information.bris.ac.uk/en/publications/the-investigation-of-natural-global-catastrophes-2

Hilton, Benjamin and Peter McIntyre. "Nuclear war," 80,000 Hours, accessed June 24, 2023. https://80000hours.org/problem-profiles/nuclear-security/.

INFER. "The Pub Blog - Get Better Answers by Asking Better Questions: Understanding Strategic Question Decomposition." Accessed June 23, 2023. https://www.infer-pub.com/the-pub/question-issue-decomposition.

Intergovernmental Panel on Climate Change. *Climate Change 2021: The Physical Science Basis*. (August 9, 2021). https://www.ipcc.ch/report/ar6/wg1/.

International Energy Agency. "Net Zero by 2050: A Roadmap for the Global Energy Sector," (May 2021) accessed June 25, 2023, https://www.iea.org/reports/net-zero-by-2050.

Kahneman, Daniel. *Thinking, Fast and Slow*. Macmillan, 2011. See excerpt in "Beware the inside view." McKinsey Quarterly (2011): 1-4. Accessed June, 2023 at https://www.mckinsey.com/capabilities/strategy-and-corporate-finance/our-insights/daniel-kahneman-beware-the-inside-view.

Kahneman, Daniel and Amos Tversky. "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, Vol. 47, No. 2 (March 1979): 263-292, https://doi.org/10.2307/1914185.

Kaplan, Jared, Sam McCandlish, Tom Henighan, Tom B. Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. "Scaling Laws for Neural Language Models." arXiv.org, (Jan 23, 2020), https://doi.org/10.48550/arXiv.2001.08361.

Karger, Ezra, Pavel D. Atanasov, and Philip Tetlock. "Improving Judgments of Existential Risk: Better Forecasts, Questions, Explanations, Policies." *FHI Technical Report* (2022) https://doi.org/10.2139/ssrn.4001628.

Karger, Ezra, Joshua Monrad, Barb Mellers, and Philip Tetlock. "Reciprocal Scoring: A Method for Forecasting Unanswerable Questions." *SSRN Working Paper*, (2021) https://doi.org/10.2139/ssrn.3954498.

Lewis, Gregory, Piers Millett, Anders Sandberg, Andrew Snyder-Beattie, and Gigi Gronvall. "Information Hazards in Biotechnology" Risk Analysis 39, No. 5, (2018): 975-981, https://doi.org/10.1111/risa.13235.

Lopez, German. "How the world went from 170 million people to 7.3 billion, in one map," *Vox*, January 30, 2016, https://www.vox.com/2016/1/30/10872878/world-population-map.

Manheim, David and Gregory Lewis. "High-risk human-caused pathogen exposure events from 1975-2016 [version 2; peer review: 2 approved]," *F1000Research*, (2022), https://doi.org/10.12688/f1000research.55114.2.

Marani, Marco, Gabriel G. Katul, William K. Pan, and Anthony J. Parolari. "Intensity and Frequency of Extreme Novel Epidemics." *Proceedings of the National Academy of Sciences* 118, No.35 (2021): e2105482118. https://doi.org/10.1073/pnas.2105482118.

Mecklin, John. "Martin Rees explains how science might save us," *Bulletin of the Atomic Scientists*, December 22, 2022, https://thebulletin.org/2022/12/martin-rees-explains-how-science-might-save-us/.

Mellers, Barbara, Eric Stone, Pavel Atanasov, Nick Rohrbaugh, S. Emlen Metz, Lyle Ungar, Michael M. Bishop, Michael Horowitz, Ed Merkle, and Philip Tetlock. "The Psychology of Intelligence Analysis: Drivers of Prediction Accuracy in World Politics." *Journal of Experimental Psychology: Applied* 21, No. 1 (2015): 1–14. https://doi.org/10.1037/xap0000040.

Mellers, Barbara, Eric Stone, Terry Murray, Angela Minster, Nick Rohrbaugh, Michael Bishop, Eva Chen, Joshua Baker, Yuan Hou, Michael Horowitz, Lyle Ungar and Philip Tetlock. "Identifying and Cultivating Superforecasters as a Method of Improving Probabilistic Predictions." *Perspectives on Psychological Science* 10, No. 3 (2015): 267–281. https://doi.org/10.1177/1745691615577794.

Metaculus. "Date of Artificial General Intelligence."Accessed July 3, 2023. https://www.metaculus.com/questions/5121/date-of-artificial-general-intelligence/

———. "Date Weakly General AI is Publicly Known." Accessed July 3, 2023., https://www.metaculus.com/questions/3479/date-weakly-general-ai-is-publicly-known

———. "Ragnarök Series—results so far," Metaculus, accessed June 25, 2023. https://www.metaculus.com/questions/2568/ragnar%25C3%25B6k-seriesresults-so-far/.

———. "Will humans go extinct before 2100?" Metaculus, accessed June 24, 2023. https://www.metaculus.com/questions/578/human-extinction-by-2100/.

———. "When will the first weakly general AI system be devised, tested, and publicly announced?" Accessed July 2nd, 2023, https://www.metaculus.com/questions/3479/date-weakly-general-ai-is-publicly-known/.

Müller, Vincent C. and Nick Bostrom. "Future Progress in Artificial Intelligence: A Survey of Expert Opinion," in Vincent C. Müller (ed.), Fundamental Issues of Artificial Intelligence (Synthese Library; Berlin: Springer, 2016), 553-571. https://nickbostrom.com/papers/survey.pdf.

National Academies of Sciences, Engineering, and Medicine. *Biodefense in the Age of Synthetic Biology*. Washington, D.C.: National Academies Press, 2018. https://doi.org/10.17226/24890.

Neyman, Eric, and Tim Roughgarden. "Are You Smarter Than a Random Expert? The Robust Aggregation of Substitutable Signals." ArXiv.Org (November 4, 2021), https://arxiv.org/abs/2111.03153v2.

Nuclear Threat Initiative. "New START Treaty." Accessed July 2, 2023, https://www.nti.org/education-center/treaties-and-regimes/treaty-between-the-united-states-of-america-and-the-russian-federation-on-measures-for-the-further-reduction-and-limitation-of-strategic-offensive-arms/.

Ord, Toby. 2020. *The Precipice: Existential Risk and the Future of Humanity*. New York: Hachette Books, 2020. https://catalog.lib.uchicago.edu/vufind/Record/12343263.

Pamlin, Dennis and Stuart Armstrong, "12 Risks that threaten human civilisation: The case for a new risk category," Global Challenges Foundation (February 2015).

Piper, Kelsey. "There are two factions working to prevent AI dangers. Here's why they're deeply divided," *Vox* (Aug. 2022), accessed June 24, 2023. https://www.vox.com/future-perfect/2022/8/10/23298108/ai-dangers-ethics-alignment-present-future-risk.

Powell, Ben, Ville A.Satopää, Niall J. MacKay, & Philip E. Tetlock. "Skew-adjusted extremized-mean: A simple method for identifying and learning from contrarian minorities in groups of forecasters", *Decision* (2022), https://doi.org/10.1037/dec0000191.

Quigley, John, Matthew Revie, and Jesse Dawson. "Estimating risk when zero events have been observed." *BMJ Quality and Safety*, 22 no. 12, (2013): 1042-1043 https://doi.org/10.1136/bmjqs-2013-002246.

Roser, Max. "The Spanish flu: The global impact of the largest influenza pandemic in history," Our World in Data, March 4, 2020, https://ourworldindata.org/spanish-flu-largest-influenza-pandemic-in-history#how-many-people-died-in-the-spanish-flu-pandemic.

Sandberg, Anders and Nick Bostrom."Global Catastrophic Risks Survey," *Technical Report #2008-1*, Future of Humanity Institute, Oxford University, 1 (2008). https://www.fhi.ox.ac.uk/reports/2008-1.pdf

Simonton, Dean Keith, "Heisenberg Effect." In *Encyclopedia of Research Design*, edited by Neil J. Salkind, 564. Thousand Oaks, CA: SAGE Publications, Inc., 2010. https://doi.org/10.4135/9781412961288.

Stefánsson, H. Orri. "On the limits of the precautionary principle." *Risk Analysis* 39, no. 6 (2019): 1204-1222. https://onlinelibrary.wiley.com/doi/abs/10.1111/risa.13265.

Tetlock, Philip E. *Expert Political Judgment: How Good Is It? How Can We Know?: With a New Preface by the Author*. New edition. Princeton: Princeton University Press, 2017. https://catalog.lib.uchicago.edu/vufind/Record/11327005.

———. "A Short Course in Superforecasting," Edge, accessed June 23, 2023, https://www.edge.org/conversation/philip_tetlock-a-short-course-in-superforecasting.

Tetlock, Philip E., Barbara A. Mellers, Nick Rohrbaugh, and Eva Chen. "Forecasting Tournaments: Tools for Increasing Transparency and Improving the Quality of Debate," Current Directions in Psychological Science 23, no. 4 (August 2014). https://doi.org/10.1177/0963721414534257.

Tetlock, Philip E., and Dan Gardner. *Superforecasting: The art and science of prediction.* Crown*, 2016.*

The White House. "Inflation Reduction Act Guidebook." Accessed 2nd July, 2023, https://www.whitehouse.gov/cleanenergy/inflation-reduction-act-guidebook/.

Thèves, Catherine, Eric Crubézy, Philippe Biagini, "History of Smallpox and Its Spread in Human Populations," Microbiology Spectrum 4, no. 4, (2016), https://doi.org/10.1128/microbiolspec.poh-0004-2014.

Wallace-Wells, David. "When Will the Planet Be Too Hot for Humans? Much, Much Sooner Than You Imagine." *Intelligencer*. July 9, 2017. https://nymag.com/intelligencer/2017/07/climate-change-earth-too-hot-for-humans.html

Wickham, Hadley and Dana Seidel. "Pseudo-Log Transformation — Pseudo_log_trans." 2022, accessed June 23, 2023. https://scales.r-lib.org/reference/pseudo_log_trans.html.

Wilder-Smith, Annaliese and Sarah Osman. "Public health emergencies of international concern: a historic overview," *Journal of Travel Medicine*, 27, No. 8 (December 2020) https://pubmed.ncbi.nlm.nih.gov/33284964/.

Appendices

Data in Appendices 4 ,5, 6, 7 are provided for informational value, but in the coming weeks, the formatting and rounding rules will be updated and the appearance of the appendices may change. We do not expect significant changes to the forecasts presented.

Appendix 1:

Practical problems in running the XPT

Before concluding, we want to highlight down-to-earth problems we encountered in running the XPT.

The project was time-consuming, and our attrition rate was roughly 34%[141] from initial forecasts to completion of the tournament four months later. Finding busy professionals and researchers willing to participate from start to end involved wide-scale recruitment of experts via blogs, Twitter, mass emails, and personal solicitations to academic laboratories, companies where employees had deep knowledge about the risks we studied, and nonprofits. Those who signed up cannot be claimed to be a representative of experts in each of these fields. In future work, we plan to deploy more systematic methods of defining and sampling from expert populations as well as giving experts more flexibility on the timing, number, and types of questions they answer, to make it easier to commit to a multi-month process and to obtain a full picture of the heterogeneity of beliefs within epistemic communities.

We suspect that one obstacle to recruitment and driver of attrition was our requiring participants to answer questions that spanned domains, including domains where experts lacked expertise. In ongoing research, we are exploring whether a focus on a smaller set of questions in a specific domain (like AI) can improve focus and persuasion in exercises like the XPT.

Results from a tournament like the XPT can never be completely current. Since we finished our main collection of data at the end of October, 2022, GPT-style Large-Language models have shifted the discussion about risks from AI; threats of nuclear risks from North Korea and the Ukraine-Russia war may have changed suddenly over the past year; and policymakers have elevated concerns about Avian Influenza and other transmittable diseases relative to previous years. We plan to continue to survey XPT participants in the coming years to explore changes in beliefs. And in light of recent advances in AI models, we re-surveyed some of our most skeptical XPT participants about the probability of AI-caused human extinction. We see no evidence that these skeptics updated noticeably towards higher levels of concern about AI.

141 A total of 169 forecasters submitted at least one forecast for the tournament; 111 forecasters participated during all four stages of the tournament.

Many of the forecasts discussed in this report are of low probabilities, in the range of 0-1%. There is little evidence bearing on the pros and cons of different elicitation techniques in low-probability domains. We are exploring these issues in followup research.

Appendix 2:

A brief history of forecasting research

The first round of forecasting tournaments focused on geopolitics and sought to provide a rigorous measure of analysts' political judgment. Much of that work was funded by the Intelligence Advanced Research Projects Activity (IARPA), a branch of the Office of the Director of National Intelligence. The intelligence community has a clear interest in improving the accuracy of its forecasts, whether by identifying its best analysts or improving how they work.

For example, in one forecasting competition run from 2011 to 2013, IARPA provided forecasters with geopolitical questions ranging from whether North Korea would test a nuclear device to whether Moody's would downgrade the sovereign debt rating of Greece. Forecasters competed individually and on teams to make the most accurate possible forecasts, and researchers measured the correlates of forecast accuracy as well as the impact of a series of interventions—like training the forecasters in probabilistic reasoning.[142]

A study of more than 700 forecasters who participated in that tournament provides a representative snapshot of the findings of forecasting research: Overall, forecasters in the tournament did much better than pure chance. More accurate forecasters scored higher on intelligence tests than less accurate ones, on average. They also scored higher on measures of open-mindedness and political knowledge and were more likely to revise their forecasts over time. Training in probabilistic reasoning improved the accuracy of the forecasters randomly assigned to receive it.

This line of research has evolved over time to study the impact of working in teams (it can help), the role of scenario planning (less helpful), and even the ability to predict counterfactual scenarios.[143] Researchers have worked to overcome the tradeoff between "rigor and relevance," by developing clusters of questions that can each be resolved rigorously but which collectively speak to a wider question of public interest.[144]

---

142 Barbara Mellers et al., "The Psychology of Intelligence Analysis: Drivers of Prediction Accuracy in World Politics.," *Journal of Experimental Psychology: Applied* 21, no. 1 (March 2015): 1–14, https://doi.org/10.1037/xap0000040.

143 Walter Frick, "What a Study of Video Games Can Tell Us about Being Better Decision Makers," *Quartz*, September 6, 2020, https://qz.com/1899461/how-individuals-and-companies-can-get-better-at-making-decisions.

144 "The Pub Blog - Get Better Answers by Asking Better Questions: Understanding Strategic Question Decomposition," INFER, accessed June 23, 2023, https://www.infer-pub.com/the-pub/question-issue-decomposition.

And the success of early forecasting tournaments has led to the proliferation of numerous public platforms for forecasting, like Good Judgment Open, Metaculus, and INFER. These platforms have caught the attention of journalists, who increasingly cite them in reporting.[145]

Yet, for all this success, forecasting research remains relatively nascent and there are many new frontiers to be explored. That includes two areas that the XPT focused on: How to incorporate persuasion into forecasting; and forecasting long-term, potentially unresolvable questions like extinction risk. In 2023, many of the team members behind the XPT announced a new initiative, the Forecasting Research Institute, dedicated to advancing the science of forecasting.[146]

145 Walter Frick, "Journalists Wake up to the Power of Prediction Markets." https://www.niemanlab.org/2022/12/journalists-wake-up-to-the-power-of-prediction-markets/.
146 "Forecasting Research Institute," Forecasting Research Institute, accessed June 23, 2023, https://forecastingresearch.org.

Appendix 3:

Scoring forecasts

The intuition behind scoring rules

In forecasting research, a "proper scoring rule" is one that incentivizes participants to make forecasts that reflect their true beliefs. The intuition behind these rules is that forecasters should be rewarded for giving higher probabilities for things that eventually happen and lower probabilities for things that don't happen. When these scores are aggregated across enough questions, they provide a reasonable estimate of a forecaster's ability to form accurate beliefs about the future.

One of the most commonly used scoring rules is a Brier score. For a binary question (X happens or it doesn't), a forecaster makes a probabilistic forecast, ranging from 0-100%. To calculate the Brier Score these values are rescaled to the 0-1 interval. The Brier score is the squared difference between the actual outcome—1 if the event happens, 0 if it doesn't—and the probabilistic forecast.[147] Lower scores are better. If an event occurs, someone who gave it a 70% chance of occurring scores better than someone who gave it a 30% chance. That's because $(1-0.7)^2=0.09 < (1-0.3)^2=0.49$.

While Brier scores do a good job of distinguishing a forecast of, say, 80% from one of 20%, they aren't ideal for scoring forecasts of very low-probability events. For that reason, the XPT used slightly different methods—though the intuition is much the same. For binary questions, the XPT used a log-scoring rule, which does a better job distinguishing between low values—like the difference between a 1% chance of an event and a 0.1% chance.

For other questions, the XPT used an "S-Score"[148] to elicit a range of outcomes: Forecasters were asked to give several estimates for a given quantity, representing their estimated probability distribution. For example, forecasters provided estimates of the number of deaths due to some cause by 2024 for a range of thresholds (in this case 5%, 25%, 50%, 75%, and 95%). The 50% estimate represents their median forecast.

---

147 A variation on this method allows the Brier score to work for forecasts with more than two possible outcomes. In that case, the squared differences between actual outcomes and forecasts are added up across the outcome categories.

148 Zhi Chen et al., "Evaluating Quantile Forecasts in the M5 Uncertainty Competition", *International Journal of Forecasting*, 38, no 4, (October-December 2022): 1531-1545, https://www.sciencedirect.com/science/article/abs/pii/S0169207022000449.

Reciprocal scoring

For unresolvable questions (like the chance of human extinction) and for long-run questions (like events in the year 2100), forecasters were asked to give their own prediction but were also asked for (and incentivized based on) two additional forecasts: one predicting the median of the superforecasters' forecasts, and the other predicting the median of the experts' forecasts. In other words, they were challenged to accurately predict what other forecasters would answer.

Reciprocal scores are based on the closeness of a forecaster's prediction to the median of the group they are predicting. So if the median superforecaster's forecast is 2% for a given question, a participant whose reciprocal forecast for the superforecasters was 2.5% would score better than one whose reciprocal forecast was 5%.

Previous research has found that these scores are often accurate in their own right and that accuracy on this measure correlates with forecasting accuracy in the real world.

Appendix 4:

Comparisons between subgroups of XPT forecasters

Earlier in the report, we discussed key differences in beliefs between those XPT forecasters we categorized as the "AI-concerned" (the third of participants with the highest forecast of AI extinction risk by 2100) and the "AI skeptics" (the third of participants with the lowest forecast of AI extinction risk by 2100). This appendix presents more detailed data about these comparisons. First, Table 25 compares the AI-concerned and AI skeptics on longer-term predictions resolving by 2050 or 2100, or those with no set resolution date. Second, Table 26 compares those groups on shorter-term predictions resolving by 2030.

In addition, we compared differences in beliefs between superforecasters and other experts, as part of our ongoing efforts to understand what makes some people better than others at predicting the future. Table 27 compares superforecasters with experts on predictions resolving by 2050 or 2100, or those with no set resolution date, and Table 28 compares superforecasters and experts on predictions resolving by 2030.

To see the full question text and resolution criteria for each question, use the table in Appendix 5.

These comparisons often involve groups with small sample sizes. While almost all respondents answered the key long-run questions about extinction and catastrophic risk, only a small, random subset of participants were asked to answer questions about shorter-run predictors of those risks and mechanisms underlying them, with all other participants having the option to answer those questions (and some self-selecting to answer). So, the claims mentioned below are exploratory and should **not** be read as indicating statistically significant differences between groups.

Comparison Between AI-Concerned and AI Skeptics for Longer-Term Forecasts[147]

		Number of Forecasters		Median Forecasts				Distributional Overlap (imputed forecasts)	
Question	**Year**	**AI-Concerned**	**AI-Skeptic**	**AI-Concerned (real)**	**AI-Concerned (imputed)**	**AI-Skeptic (real)**	**AI-Skeptic (imputed)**	**% AI-concerned above median(AI skeptic)**	**% AI skeptic above median(AI-concerned)**
1. Genetically Engineered Pathogen Risk (%)	2100	54	54	12.07%	12.38%	2.00%	2.00%	96%	7%
2. Non-Genetically Engineered Pathogen Risk (%)	2100	55	54	7.07%	7.54%	3.00%	2.67%	79%	30%
3. AI Catastrophic Risk (%)	2100	55	53	12%	12%	0.80%	0.70%	100%	2%
4. AI Extinction Risk (%)	2100	54	53	7.50%	7.50%	0.01%	0.01%	100%	0%
5. Nuclear Catastrophic Risk (%)	2100	54	54	7.90%	8.00%	2.25%	2.57%	88%	25%
6. Nuclear Extinction Risk (%)	2100	54	52	0.41%	0.45%	0.01%	0.01%	88%	11%

[147] Because this table focuses on longer-term predictions resolving by 2050 or 2100, or those with no set resolution date, not all XPT questions are included.

		Number of Forecasters		Median Forecasts				Distributional Overlap (imputed forecasts)	
Question	**Year**	**AI-Concerned**	**AI-Skeptic**	**AI-Concerned (real)**	**AI-Concerned (imputed)**	**AI-Skeptic (real)**	**AI-Skeptic (imputed)**	**% AI-concerned above median(AI skeptic)**	**% AI skeptic above median(AI-concerned)**
7. Non-Anthropogenic Catastrophic Risk (%)	2100	51	53	0.14%	0.17%	0.01%	0.03%	82%	12%
8. Non-Anthropogenic Extinction Risk (%)	2100	52	53	0.01%	0.01%	0.0007%	0.0009%	82%	16%
9. Total Catastrophic Risk (%)	2100	50	53	27.30%	27.68%	5.30%	5.15%	100%	5%
10. Total Extinction Risk (%)	2100	51	53	11%	11%	0.12%	0.12%	100%	2%
11. Year of Extinction (year)		50	52	3500	3500	28000	30000	21%	95%
12. Future Human Births (#)		48	51	100 Billion	100 Billion	700 Billion	725 Billion	23%	79%
15. Non-State Actor Bioweapon 1k Deaths (%)	2050	14	22	1%	1.33%	1%	1.65%	30%	57%
16. State Actor Bioweapon 1k Deaths (%)	2050	18	18	0.265%	0.75%	0.39%	0.63%	59%	45%
17. Non-State Actor Bioweapon 100k Deaths (%)	2050	13	17	0%	0.11%	0.1%	0.23%	29%	61%

Question	Year	Number of Forecasters		Median Forecasts				Distributional Overlap (imputed forecasts)	
		AI-Concerned	AI-Skeptic	AI-Concerned (real)	AI-Concerned (imputed)	AI-Skeptic (real)	AI-Skeptic (imputed)	% AI-concerned above median(AI skeptic)	% AI skeptic above median(AI-concerned)
18. State Actor Bioweapon 100k Deaths (%)	2050	12	17	0%	0.04%	0.01%	0.30%	25%	84%
19. Lab Leaks (%)	2050	13	21	1%	0.92%	0.8%	0.63%	80%	21%
20. Individual Countries with Biological Weapons Programs: US (%)	2050	11	13	12%	35.21%	25%	41.79%	39%	70%
20. Individual Countries with Biological Weapons Programs: Russia (%)	2050	10	12	90%	82.87%	88.5%	73.78%	75%	27%
20. Individual Countries with Biological Weapons Programs: China (%)	2050	9	13	75%	73.13%	80%	72.43%	55%	48%
20. Individual Countries with Biological Weapons Programs: North Korea (%)	2050	9	13	85%	75.07%	80%	78.36%	34%	63%
20. Individual Countries with Biological Weapons Programs: Israel (%)	2050	10	11	60%	54.98%	80%	61.50%	23%	64%

		Number of Forecasters		Median Forecasts				Distributional Overlap (imputed forecasts)	
Question	Year	AI-Concerned	AI-Skeptic	AI-Concerned (real)	AI-Concerned (imputed)	AI-Skeptic (real)	AI-Skeptic (imputed)	% AI-concerned above median(AI skeptic)	% AI skeptic above median(AI-concerned)
20. Individual Countries with Biological Weapons Programs: Iran (%)	2050	10	13	69%	60.17%	70%	59.45%	55%	50%
20. Individual Countries with Biological Weapons Programs: Syria (%)	2050	11	12	35%	45.98%	52%	46.91%	46%	54%
21. Number of Countries with Biological Weapons Programs (%)	2050	13	17	6%	7.38%	7%	7.29%	50%	46%
22. PHEIC Declarations with 10k Deaths (%)	2050	14	19	5.5%	6.74%	7%	7.17%	36%	70%
23. Assassinations with Biological Weapons (%)	2050	13	18	0.4%	0.40%	0.145%	0.19%	89%	20%
24. Malaria Deaths (#)	2050	14	17	315000	300800	250000	283010	57%	38%
25. Average Global Surface Temperature (°C)	2100	12	18	2.5°C	2.65°C	2.59°C	2.64°C	50%	50%
27. Nuclear Fusion Energy (year)		12	19	2070	2082	2072	2094.5	36%	61%

		Number of Forecasters		Median Forecasts				Distributional Overlap (imputed forecasts)	
Question	**Year**	**AI-Concerned**	**AI-Skeptic**	**AI-Concerned (real)**	**AI-Concerned (imputed)**	**AI-Skeptic (real)**	**AI-Skeptic (imputed)**	**% AI-concerned above median(AI skeptic)**	**% AI skeptic above median(AI-concerned)**
33. Countries with Nuclear Warheads (#)	2050	10	25	11	11.35	11	11.00	71%	34%
36. US GDP From Software (%)	2050	14	13	7.8%	7.30%	7.82%	6.65%	88%	36%
37. US Computer R&D Spending ($)	2050	7	19	7.2e11	9.07e11	9e11	9.09e11	48%	50%
38. Labor Force Participation Rate in OECD (%)	2050	11	17	76%	77.42%	78%	78.00%	38%	55%
42. AI Wins InternationalMathematical Olympiad (year)		12	21	2030	2034	2034	2038.4	25%	71%
43. NYT Bestsellers Written by AI (year)		15	18	2045	2050.1	2056.5	2051	43%	54%
44. Date of Advanced AI (year)		14	18	2050	2053.2	2051	2057.6	34%	61%
45. Maximum Compute Used in an AI Experiment (petaFLOPS-days)	2050	7	16	1 billion	1.44e31	900 million	2.36e22	100%	20%

		Number of Forecasters		Median Forecasts				Distributional Overlap (imputed forecasts)	
Question	**Year**	**AI-Concerned**	**AI-Skeptic**	**AI-Concerned (real)**	**AI-Concerned (imputed)**	**AI-Skeptic (real)**	**AI-Skeptic (imputed)**	**% AI-concerned above median(AI skeptic)**	**% AI skeptic above median(AI-concerned)**
46. Largest AI Experiment Cost of Compute ($)	2050	17	13	$1,000,000,000	$822,730,000	$70,000,000	$303,450,000.00	84%	5%
47. Lowest Price of GFLOPS ($)	2050	18	12	$0.000125	$0.00	$0.002	$0.00	32%	77%
49. Largest Number of Parameters in a Machine Learning Model (#)	2050	12	14	750 quadrillion	2.3 sextillion	57.5 quintillion	14.7 quintillion	75%	25%
50. Negative Public Opinion of AI (%)	2050	8	18	38%	36.60%	33%	34.48%	77%	23%
51. Nick Bostrom Affirms Existence of AGI (%)	2100	12	10	79%	76.72%	77.5%	69.27%	77%	20%
52. Probability of GDP Growth Over 15% (%)		10	20	16.9%	10.52%	2.05%	4.51%	88%	9%
54. Loss of Agricultural Production (%)	2100	13	17	9.5%	6.45%	3%	4.96%	70%	30%
55. Space Colony (%)	2100	16	28	12.5%	13.83%	1.5%	8%	93%	36%
56. Happiness in America (%)	2100	10	16	86%	85.60%	85%	85.06%	63%	41%
57. Prevalence of Autocracies	2100	14	18	55%	53.78%	67%	60.20%	25%	88%

		Number of Forecasters		Median Forecasts				Distributional Overlap (imputed forecasts)	
Question	Year	AI-Concerned	AI-Skeptic	AI-Concerned (real)	AI-Concerned (imputed)	AI-Skeptic (real)	AI-Skeptic (imputed)	% AI-concerned above median(AI skeptic)	% AI skeptic above median(AI-concerned)
(%)									
58. Future Worries and Children (year)		14	21	2650	5393.1	2450	3991.59	71%	20%
59. Generation Attitudes (year)		10	23	2170	7416.19	3100	7093.84	50%	50%

Comparison Between AI-Concerned and AI Skeptics for Forecasts Resolving by 2030[148]

		Number of Forecasters		Median Forecasts				Distributional Overlap (imputed forecasts)	
Question	Year	AI-Concerned	AI-Skeptic	AI-Concerned (real)	AI-Concerned (imputed)	AI-Skeptic (real)	AI-Skeptic (imputed)	% AI-concerned above median(AI skeptic)	% AI skeptic above median(AI-concerned)
1. Genetically Engineered	2030	56	56	1.28%	1.3%	0.1%	0.1%	93%	7%

[148] Because this table focuses on predictions resolving by 2030, not all XPT questions are included.

Question	Year	Number of Forecasters		Median Forecasts				Distributional Overlap (imputed forecasts)	
		AI-Concerned	AI-Skeptic	AI-Concerned (real)	AI-Concerned (imputed)	AI-Skeptic (real)	AI-Skeptic (imputed)	% AI-concerned above median(AI skeptic)	% AI skeptic above median(AI-concerned)
Pathogen Risk (%)									
2. Non-Genetically Engineered Pathogen Risk (%)	2030	55	55	1.3%	1.33%	0.4%	0.37%	79%	21%
3. AI Catastrophic Risk (%)	2030	55	54	1%	1%	0.0001%	0.0001%	96%	5%
4. AI Extinction Risk (%)	2030	53	55	0.5%	0.5%	0%	0%	100%	0%
5. Nuclear Catastrophic Risk (%)	2030	54	53	1%	1.1%	0.5%	0.49%	81%	21%
6. Nuclear Extinction Risk (%)	2030	53	53	0.03%	0.03%	0.0001%	0.00018%	91%	18%
7. Non-Anthropogenic Catastrophic Risk (%)	2030	52	52	0.01%	0.01%	0.001%	0.001%	82%	27%
8. Non-Anthropogenic Extinction Risk (%)	2030	53	52	0.001%	0.001%	0.0001%	0.0001%	84%	14%
9. Total Catastrophic Risk (%)	2030	51	52	2.8%	3%	0.65%	0.69%	89%	11%
10. Total Extinction Risk (%)	2030	51	52	1%	1%	0.001%	0.001%	95%	9%
13. Non-Coronavirus mRNA Vaccine (#)	2030	49	53	60 million	100 million	80 million	90 million	54%	50%

		Number of Forecasters		Median Forecasts				Distributional Overlap (imputed forecasts)	
Question	Year	AI-Concerned	AI-Skeptic	AI-Concerned (real)	AI-Concerned (imputed)	AI-Skeptic (real)	AI-Skeptic (imputed)	% AI-concerned above median(AI skeptic)	% AI skeptic above median(AI-concerned)
14. Novel Infectious Disease Surveillance System (%)	2030	19	27	49%	53.89%	70%	57.55%	42%	61%
15. Non-State Actor Bioweapon 1k Deaths (%)	2030	15	24	0.35	0.36	0.4	0.35	54%	45%
16. State Actor Bioweapon 1k Deaths (%)	2030	15	23	0.1	0.25	0.0013	0.22	51%	46%
17. Non-State Actor Bioweapon 100k Deaths (%)	2030	14	23	0.00425	0.03	0	0.03	49%	52%
18. State Actor Bioweapon 100k Deaths (%)	2030	13	23	0	0.02	0.0013	0.03	44%	57%
19. Lab Leaks (%)	2030	12	24	0.2	0.16	0.13	0.16	47%	54%
20. Individual Countries with Biological Weapons Programs: US (%)	2030	14	16	9%	26.33%	27.5%	30.96%	37%	66%
20. Individual Countries with Biological Weapons Programs: Russia (%)	2030	12	16	80%	76.97%	80%	78.63%	47%	61%
20. Individual Countries with Biological Weapons Programs: China (%)	2030	12	16	50%	66.23%	75%	72.31%	25%	70%

Question	Year	Number of Forecasters		Median Forecasts				Distributional Overlap (imputed forecasts)	
		AI-Concerned	AI-Skeptic	AI-Concerned (real)	AI-Concerned (imputed)	AI-Skeptic (real)	AI-Skeptic (imputed)	% AI-concerned above median(AI skeptic)	% AI skeptic above median(AI-concerned)
20. Individual Countries with Biological Weapons Programs: North Korea (%)	2030	12	15	65%	72.78%	80%	79.73%	32%	80%
20. Individual Countries with Biological Weapons Programs: Israel (%)	2030	13	15	30%	45.65%	60%	51.64%	44%	71%
20. Individual Countries with Biological Weapons Programs: Iran (%)	2030	13	16	30%	52.33%	70%	56.31%	42%	61%
20. Individual Countries with Biological Weapons Programs: Syria (%)	2030	13	14	20%	42.33%	50%	44.61%	47%	55%
21. Number of Countries with Biological Weapons Programs (%)	2030	12	21	4	5.22	7	5.63	40%	57%
22. PHEIC Declarations with 10k Deaths (%)	2030	13	20	2	2	2	2	58%	57%
23. Assassinations with Biological Weapons (%)	2030	12	22	0.06	0.12	0.03	0.11	56%	41%
24. Malaria Deaths (#)	2030	13	20	500000	480000	462500	497319	39%	61%

		Number of Forecasters		Median Forecasts				Distributional Overlap (imputed forecasts)	
Question	**Year**	**AI-Concerned**	**AI-Skeptic**	**AI-Concerned (real)**	**AI-Concerned (imputed)**	**AI-Skeptic (real)**	**AI-Skeptic (imputed)**	**% AI-concerned above median(AI skeptic)**	**% AI skeptic above median(AI-concerned)**
25. Average Global Surface Temperature (°C)	2030	14	24	1.43°C	1.46°C	1.43°C	1.45°C	60%	45%
26. Cost of Utility-Scale Solar Energy (2017 $/kWh)	2030	9	19	$0.03	$0.02	$0.03	$0.03	30%	57%
28. Solar and Wind Energy (%)	2030	11	22	25%	25.55%	23%	24%	67%	30%
29. Annual Direct Air CO2 Capture (Mt CO2/year)	2030	18	20	13	16.52	17.5	18.69	44%	52%
30. Cost of Hydrogen ($/kg of hydrogen)	2030	11	19	$2.5	$2.72	$2.5	$2.76	46%	52%
31. Nuclear Weapon Use (%)	2030	16	22	6.73%	6.71%	2.00%	4.74%	72%	23%
32. Total Nuclear Warheads (#)	2030	12	19	11750	12424.07	12500	12478	49%	55%
33. Countries with Nuclear Warheads (#)	2030	11	26	10	9.68	9	9.58	68%	41%
34. Country-by-Country Nuclear Use: China (%)	2030	11	19	0.25%	0.41%	0.01%	0.20%	82%	34%
34. Country-by-Country Nuclear Use: France (%)	2030	11	19	0.01%	0.02%	0%	0.03%	39%	52%

Question	Year	Number of Forecasters		Median Forecasts				Distributional Overlap (imputed forecasts)	
Question	Year	AI-Concerned	AI-Skeptic	AI-Concerned (real)	AI-Concerned (imputed)	AI-Skeptic (real)	AI-Skeptic (imputed)	% AI-concerned above median(AI skeptic)	% AI skeptic above median(AI-concerned)
34. Country-by-Country Nuclear Use: India (%)	2030	11	19	0.25%	0.19%	0.2%	0.23%	45%	64%
34. Country-by-Country Nuclear Use: Israel (%)	2030	11	19	0.2%	0.2%	0.1%	0.2%	46%	54%
34. Country-by-Country Nuclear Use: North Korea (%)	2030	11	19	1%	0.51%	0.1%	0.51%	50%	50%
34. Country-by-Country Nuclear Use: Pakistan (%)	2030	11	19	0.25%	0.25%	0.1%	0.3%	41%	57%
34. Country-by-Country Nuclear Use: Russia (%)	2030	11	19	2%	2.44%	1%	1.72%	71%	43%
34. Country-by-Country Nuclear Use: United Kingdom (%)	2030	11	19	0.01%	0.05%	0%	0.05%	52%	50%
34. Country-by-Country Nuclear Use: United States (%)	2030	11	19	0.4%	0.71%	0.1%	0.28%	80%	21%
34. Country-by-Country Nuclear Use: Other actor (state) (%)	2030	11	19	0.01%	0.10%	0.1%	0.2%	36%	63%
34. Country-by-Country	2030	11	19	0.038%	0.09%	0.02%	0.12%	46%	63%

		Number of Forecasters		Median Forecasts				Distributional Overlap (imputed forecasts)	
Question	Year	AI-Concerned	AI-Skeptic	AI-Concerned (real)	AI-Concerned (imputed)	AI-Skeptic (real)	AI-Skeptic (imputed)	% AI-concerned above median(AI skeptic)	% AI skeptic above median(AI-concerned)
Nuclear Use: Other actor (non-state) (%)									
34. Country-by-Country Nuclear Use: This will not occur (%)	2030	11	19	95.14%	95.06%	97.81%	95.93%	32%	52%
35. GPT Revenue (%)		16	16	49.25%	50.98%	55.5%	56.21%	25%	63%
36. US GDP From Software (%)	2030	13	18	5%	4.94%	4.78%	4.78%	63%	41%
37. US Computer R&D Spending ($)	2030	8	16	$232,461,500,000	$332,510,295,967	$35,6043,500,000	$310,171,485,069	63%	41%
38. Labor Force Participation Rate in OECD (%)	2030	10	18	78%	77.63%	77%	77.2%	60%	43%
39. MATH Dataset Benchmark (%)	2030	19	19	91%	87.09%	92%	87.13%	49%	50%
40. ""Massive Multitask Language Understanding"" Benchmark (%)	2030	12	15	90%	89.39%	90%	88.89%	54%	48%
41. QuALITY Dataset Benchmark (SAT-style score 0-100)	2030	15	16	78	78.67	76.50	76.88	65%	41%

		Number of Forecasters		Median Forecasts				Distributional Overlap (imputed forecasts)	
Question	Year	AI-Concerned	AI-Skeptic	AI-Concerned (real)	AI-Concerned (imputed)	AI-Skeptic (real)	AI-Skeptic (imputed)	% AI-concerned above median(AI skeptic)	% AI skeptic above median(AI-concerned)
45. Maximum Compute Used in an AI Experiment (petaFLOPS-days)	2030	9	17	4,532,919	17,907,408	6,700,000	15,247,433	58%	48%
46. Largest AI Experiment Cost of Compute ($)	2030	19	19	$300,000,000	$156,404,997	$100,000,000	$100,000,000	68%	20%
47. Lowest Price of GFLOPS ($)	2030	17	20	$0	$0.01	$0	$0.01	47%	59%
48. ImageNet Classification Training Efficiency (factor)	2030	13	19	10,000	6,546.14	1,572	4,743.98	63%	38%
49. Largest Number of Parameters in a Machine Learning Model (#)	2030	14	19	3.5 quadrillion	25 quadrillion	3 quadrillion	15.6 quadrillion	56%	46%
50. Negative Public Opinion of AI (%)	2030	10	23	36.75%	35.5%	33%	33.92%	75%	30%
51. Nick Bostrom Affirms Existence of AGI (%)	2030	11	19	10%	6.55%	1%	3.83%	75%	14%
54. Loss of Agricultural Production (%)	2030	16	20	0.78%	0.64%	0.09%	0.22%	79%	34%
55. Space Colony (%)	2030	18	27	0.01%	0.06%	0%	0%	75%	11%

Question	Year	Number of Forecasters		Median Forecasts				Distributional Overlap (imputed forecasts)	
		AI-Concerned	AI-Skeptic	AI-Concerned (real)	AI-Concerned (imputed)	AI-Skeptic (real)	AI-Skeptic (imputed)	% AI-concerned above median(AI skeptic)	% AI skeptic above median(AI-concerned)
56. Happiness in America (%)	2030	10	19	86%	86%	84%	86%	53%	55%
57. Prevalence of Autocracies (%)	2030	14	19	68%	66.7%	69%	66.38%	56%	45%

Comparison Between Superforecasters and Experts for Longer-Term Forecasts Resolving[149]

Question	Year	Number of Forecasters		Median Forecasts				Distributional Overlap (imputed forecasts)	
		Superforecasters	Experts	Superforecasters (real)	Superforecasters (imputed)	Experts (real)	Experts (imputed)	% Supers above median (experts)	% Experts above median(supers)
1. Genetically Engineered Pathogen Risk (%)	2100	89	74	4%	4%	10%	10%	15%	82%
2. Non-Genetically Engineered Pathogen Risk (%)	2100	88	75	3.63%	4%	8%	6.54%	33%	64%

[149] Because this table focuses on longer-term predictions resolving by 2050 or 2100, or those with no set resolution date, not all XPT questions are included.

		Number of Forecasters		Median Forecasts				Distributional Overlap (imputed forecasts)	
Question	Year	Superforecasters	Experts	Superforecasters (real)	Superforecasters (imputed)	Experts (real)	Experts (imputed)	% Supers above median (experts)	% Experts above median(supers)
3. AI Catastrophic Risk (%)	2100	88	74	2.13%	2.25%	10%	9.5%	8%	81%
4. AI Extinction Risk (%)	2100	88	73	0.38%	0.5%	3%	3%	16%	73%
5. Nuclear Catastrophic Risk (%)	2100	88	73	4%	4%	8%	8%	24%	82%
6. Nuclear Extinction Risk (%)	2100	88	71	0.07%	0.08%	0.32%	0.28%	27%	63%
7. Non-Anthropogenic Catastrophic Risk (%)	2100	88	69	0.05%	0.05%	0.09%	0.09%	38%	62%
8. Non-Anthropogenic Extinction Risk (%)	2100	88	70	0.0043%	0.0046%	0.004%	0.004%	53%	49%
9. Total Catastrophic Risk (%)	2100	86	67	9.05%	9.3%	20%	20%	15%	82%
10. Total Extinction Risk (%)	2100	87	67	1%	1%	6%	6%	17%	79%
11. Year of Extinction (year)		87	66	15000	15000	8743.50	6261	21%	95%
12. Future Human Births (#)		84	64	500 Billion	500 Billion	135 Billion	200 Billion	23%	79%
15. Non-State Actor Bioweapon 1k Deaths (%)	2050	38	21	1	1.07	2	1.45	39%	71%
16. State Actor Bioweapon 1k Deaths (%)	2050	33	25	1	0.83	0.7	0.82	51%	50%

		Number of Forecasters		Median Forecasts				Distributional Overlap (imputed forecasts)	
Question	**Year**	**Superforecasters**	**Experts**	**Superforecasters (real)**	**Superforecasters (imputed)**	**Experts (real)**	**Experts (imputed)**	**% Supers above median (experts)**	**% Experts above median(supers)**
17. Non-State Actor Bioweapon 100k Deaths (%)	2050	33	21	0.04	0.16	0.001	0.14	51%	46%
18. State Actor Bioweapon 100k Deaths (%)	2050	35	18	0.15	0.22	0.00	0.09	75%	38%
19. Lab Leaks (%)	2050	32	22	0.68	0.71	0.88	0.83	40%	67%
20. Individual Countries with Biological Weapons Programs: US (%)	2050	29	15	30%	43.33%	12%	34.66%	65%	32%
20. Individual Countries with Biological Weapons Programs: Russia (%)	2050	26	12	80%	72.67%	85%	80%	35%	69%
20. Individual Countries with Biological Weapons Programs: China (%)	2050	27	11	80%	70%	73%	72.04%	46%	60%
20. Individual Countries with Biological Weapons Programs: North Korea (%)	2050	25	12	80%	75%	80%	76.6%	44%	56%
20. Individual Countries with Biological Weapons Programs: Israel (%)	2050	29	11	60%	55.56%	55%	55.2%	51%	49%
20. Individual Countries with	2050	28	13	65%	56.39%	60%	60%	38%	65%

		Number of Forecasters		Median Forecasts				Distributional Overlap (imputed forecasts)	
Question	Year	Superforecasters	Experts	Superforecasters (real)	Superforecasters (imputed)	Experts (real)	Experts (imputed)	% Supers above median (experts)	% Experts above median(supers)
Biological Weapons Programs: Iran (%)									
20. Individual Countries with Biological Weapons Programs: Syria (%)	2050	27	13	35%	46.67%	39%	48.16%	45%	55%
21. Number of Countries with Biological Weapons Programs (%)	2050	32	17	7	7	8	7.47	38%	63%
22. PHEIC Declarations with 10k Deaths (%)	2050	33	18	7	7.03	6	7	60%	49%
23. Assassinations with Biological Weapons (%)	2050	35	15	0.4	0.29	0.3	0.32	45%	62%
24. Malaria Deaths (#)	2050	32	18	300,000	288,076	275,000	288,271	49%	50%
25. Average Global Surface Temperature (°C)	2100	33	21	2.60°C	2.67°C	2.50°C	2.62°C	60%	42%
27. Nuclear Fusion Energy (year)		34	18	2077	2094.78	2068.5	2084.28	62%	38%
33. Countries with Nuclear Warheads (#)	2050	36	20	11	11	11	11.33	42%	67%
36. US GDP From Software	2050	30	16	7.71%	7%	8.25%	7.35%	40%	69%

		Number of Forecasters		Median Forecasts				Distributional Overlap (imputed forecasts)	
Question	Year	Superforecasters	Experts	Superforecasters (real)	Superforecasters (imputed)	Experts (real)	Experts (imputed)	% Supers above median (experts)	% Experts above median(supers)
(%)									
37. US Computer R&D Spending ($)	2050	34	6	$696,000,000,000	$881,111,000,000	$785,000,000,000	$985,772,000,000	33%	69%
38. Labor Force Participation Rate in OECD (%)	2050	32	13	78%	77.31%	75%	77.35%	47%	51%
42. AI Wins InternationalMathematical Olympiad (year)		36	14	2035	2038.46	2030	2036.11	57%	31%
43. NYT Bestsellers Written by AI (year)		37	21	2050	2050	2040	2050.13	49%	55%
44. Date of Advanced AI (year)		32	22	2060	2060.28	2045.5	2054.02	70%	31%
45. Maximum Compute Used in an AI Experiment (petaFLOPS-days)	2050	33	7	700,000,000	2.36e22	1,000,000,000	5.16e21	18%	100%
46. Largest AI Experiment Cost of Compute ($)	2050	31	18	$300,000,000	$363,333,333.3	$850,000,000	$500,000,000	33%	64%
47. Lowest Price of GFLOPS ($)	2050	32	18	$0.00014	$0.0012	$0.00019	$0.0013	47%	53%
49. Largest Number of	2050	31	13	750	14.3	2.9	1.3	26%	71%

		Number of Forecasters		Median Forecasts				Distributional Overlap (imputed forecasts)	
Question	Year	Superforecasters	Experts	Superforecasters (real)	Superforecasters (imputed)	Experts (real)	Experts (imputed)	% Supers above median (experts)	% Experts above median(supers)
Parameters in a Machine Learning Model (#)				quadrillion	quintillion	quintillion	sextillion		
50. Negative Public Opinion of AI (%)	2050	32	16	34.5%	35%	30.5%	35.88%	44%	64%
51. Nick Bostrom Affirms Existence of AGI (%)	2100	27	15	74.75%	70.56%	85%	73.98%	42%	59%
52. Probability of GDP Growth Over 15% (%)		34	12	2.75%	4.37%	18.9%	7.97%	21%	86%
53. Year of GDP Growth over 15% (year)		29	15	2081	2083.22	2065	2074.7	74%	24%
54. Loss of Agricultural Production (%)	2100	35	21	6.5%	5.43%	6.2%	5.51%	48%	51%
55. Space Colony (%)	2100	46	22	3%	8.6%	12.5%	13.46%	35%	73%
56. Happiness in America (%)	2100	33	13	87%	85%	85%	85.43%	45%	64%
57. Prevalence of Autocracies (%)	2100	30	22	65.5%	60.63%	55%	55%	81%	23%
58. Future Worries and Children (year)		35	19	2700	4112.17	2400	5219.15	31%	72%
59. Generation Attitudes		34	17	2527.5	5966.67	2500	8293.21	45%	55%

		Number of Forecasters		Median Forecasts				Distributional Overlap (imputed forecasts)	
Question	Year	Superforecasters	Experts	Superforecasters (real)	Superforecasters (imputed)	Experts (real)	Experts (imputed)	% Supers above median (experts)	% Experts above median(supers)
(year)									

Comparison Between Superforecasters and Experts for Forecasts Resolving by 2030[150]

		Number of Forecasters		Median Forecasts				Distributional Overlap (imputed forecasts)	
Question	Year	Superforecasters	Experts	Superforecasters (real)	Superforecasters (imputed)	Experts (real)	Experts (imputed)	% Supers above median (experts)	% Experts above median(supers)
1. Genetically Engineered Pathogen Risk (%)	2030	89	76	0.25%	0.25%	1%	1%	26%	76%
2. Non-Genetically Engineered Pathogen Risk (%)	2030	88	75	0.5%	0.5%	1.28%	1.28%	21%	75%
3. AI Catastrophic Risk (%)	2030	88	74	0.01%	0.01%	0.215%	0.18%	16%	77%
4. AI Extinction Risk (%)	2030	88	73	0.0001%	1E-04%	0.01%	0.01%	27%	77%

[150] Because this table focuses on predictions resolving by 2030, not all XPT questions are included.

Question	Year	Number of Forecasters: Superforecasters	Number of Forecasters: Experts	Median Forecasts: Superforecasters (real)	Median Forecasts: Superforecasters (imputed)	Median Forecasts: Experts (real)	Median Forecasts: Experts (imputed)	Distributional Overlap (imputed forecasts): % Supers above median (experts)	Distributional Overlap (imputed forecasts): % Experts above median(supers)
5. Nuclear Catastrophic Risk (%)	2030	88	73	0.5%	0.5%	1.1%	1.1%	17%	77%
6. Nuclear Extinction Risk (%)	2030	88	71	0.001%	0.001%	0.01%	0.01%	34%	81%
7. Non-Anthropogenic Catastrophic Risk (%)	2030	88	69	0.0026%	0.0027%	0.0075%	0.0087%	36%	67%
8. Non-Anthropogenic Extinction Risk (%)	2030	88	70	0.00042%	0.0005%	0.00051%	0.00067%	44%	57%
9. Total Catastrophic Risk (%)	2030	87	68	0.85%	0.86%	2.55%	2.8%	12%	75%
10. Total Extinction Risk (%)	2030	87	68	0.01%	0.01011%	0.24%	0.3%	20%	82%
13. Non-Coronavirus mRNA Vaccine (#)	2030	86	68	100,000,000	100,000,000	50,000,000	100,000,000	62%	51%
14. Novel Infectious Disease Surveillance System (%)	2030	35	25	65.75%	56.7%	60%	55.54%	56%	44%
15. Non-State Actor Bioweapon 1k Deaths (%)	2030	39	21	0.15	0.25	0.5	0.36	37%	59%
16. State Actor Bioweapon 1k	2030	33	25	0.024	0.17	0	0.21	47%	57%

		Number of Forecasters		Median Forecasts				Distributional Overlap (imputed forecasts)	
Question	Year	Superforecasters	Experts	Superforecasters (real)	Superforecasters (imputed)	Experts (real)	Experts (imputed)	% Supers above median (experts)	% Experts above median(supers)
Deaths (%)									
17. Non-State Actor Bioweapon 100k Deaths (%)	2030	33	21	0.0005	0.03	0	0.03	47%	56%
18. State Actor Bioweapon 100k Deaths (%)	2030	35	18	0.002	0.02	0	0.03	42%	54%
19. Lab Leaks (%)	2030	32	22	0.15	0.16	0.2	0.17	43%	54%
20. Individual Countries with Biological Weapons Programs: US (%)	2030	29	15	25%	30.59%	10%	28.43%	57%	43%
20. Individual Countries with Biological Weapons Programs: Russia (%)	2030	26	13	80%	78.67%	85%	79.37%	47%	54%
20. Individual Countries with Biological Weapons Programs: China (%)	2030	28	12	73.5%	70.84%	64%	68.77%	62%	35%
20. Individual Countries with Biological Weapons Programs: North Korea (%)	2030	26	13	80%	76.88%	85%	79.09%	45%	56%
20. Individual Countries with Biological Weapons Programs: Israel (%)	2030	27	12	55%	51.61%	55%	51.64%	49%	51%

		Number of Forecasters		Median Forecasts				Distributional Overlap (imputed forecasts)	
Question	**Year**	**Superforecasters**	**Experts**	**Superforecasters (real)**	**Superforecasters (imputed)**	**Experts (real)**	**Experts (imputed)**	**% Supers above median (experts)**	**% Experts above median(supers)**
20. Individual Countries with Biological Weapons Programs: Iran (%)	2030	29	14	65%	56.33%	60%	54.59%	57%	47%
20. Individual Countries with Biological Weapons Programs: Syria (%)	2030	25	13	40%	46%	35%	45.56%	52%	49%
21. Number of Countries with Biological Weapons Programs (%)	2030	32	17	7	5.77	7	5.24	61%	42%
22. PHEIC Declarations with 10k Deaths (%)	2030	33	18	2	2	2	2	57%	59%
23. Assassinations with Biological Weapons (%)	2030	35	15	0.008	0.11	0	0.1	51%	48%
24. Malaria Deaths (#)	2030	32	18	500000	495125	500000	487851	53%	47%
25. Average Global Surface Temperature (°C)	2030	33	21	1.47°C	1.46°C	1.4°C	1.45°C	57%	44%
26. Cost of Utility-Scale Solar Energy (2017 $/kWh)	2030	31	16	$0.028	$0.025	$0.025	$0.025	51%	49%
28. Solar and Wind Energy (%)	2030	33	14	24.5%	25%	23.5%	25.1%	48%	53%

		Number of Forecasters		Median Forecasts				Distributional Overlap (imputed forecasts)	
Question	**Year**	**Superforecasters**	**Experts**	**Superforecasters (real)**	**Superforecasters (imputed)**	**Experts (real)**	**Experts (imputed)**	**% Supers above median (experts)**	**% Experts above median(supers)**
29. Annual Direct Air CO2 Capture (Mt CO2/year)	2030	33	20	20	20	10	17.38	56%	43%
30. Cost of Hydrogen ($/kg of hydrogen)	2030	32	13	$2.5	$2.75	$2.68	$2.75	51%	49%
31. Nuclear Weapon Use (%)	2030	37	21	4%	4.5%	5%	6.03%	29%	70%
32. Total Nuclear Warheads (#)	2030	31	16	12900	12666.67	11500	12372.7	60%	37%
33. Countries with Nuclear Warheads (#)	2030	36	20	10	9.67	9.58	9.69	47%	56%
34. Country-by-Country Nuclear Use: China (%)	2030	33	18	0.15%	0.30%	0.27%	0.37%	45%	67%
34. Country-by-Country Nuclear Use: France (%)	2030	33	18	0.001%	0.03%	0.01%	0.02%	53%	41%
34. Country-by-Country Nuclear Use: India (%)	2030	33	18	0.2%	0.2%	0.15%	0.19%	58%	49%
34. Country-by-Country Nuclear Use: Israel (%)	2030	33	18	0.1%	0.2%	0.1%	0.19%	51%	50%
34. Country-by-Country Nuclear Use: North Korea (%)	2030	32	17	0.35%	0.56%	0.5%	0.52%	53%	49%

		Number of Forecasters		Median Forecasts				Distributional Overlap (imputed forecasts)	
Question	**Year**	**Superforecasters**	**Experts**	**Superforecasters (real)**	**Superforecasters (imputed)**	**Experts (real)**	**Experts (imputed)**	**% Supers above median (experts)**	**% Experts above median(supers)**
34. Country-by-Country Nuclear Use: Pakistan (%)	2030	33	17	0.2%	0.27%	0.2%	0.3%	47%	58%
34. Country-by-Country Nuclear Use: Russia (%)	2030	33	17	1%	1.52%	2.44%	2.49%	38%	78%
34. Country-by-Country Nuclear Use: United Kingdom (%)	2030	33	18	0.01%	0.03%	0.01%	0.04%	46%	55%
34. Country-by-Country Nuclear Use: United States (%)	2030	33	18	0.1%	0.34%	0.23%	0.65%	26%	72%
34. Country-by-Country Nuclear Use: Other actor (state) (%)	2030	33	18	0.05%	0.1%	0.05%	0.1%	55%	51%
34. Country-by-Country Nuclear Use: Other actor (non-state) (%)	2030	33	17	0.02%	0.1%	0.01%	0.11%	44%	53%
34. Country-by-Country Nuclear Use: This will not occur (%)	2030	33	18	97.3%	95.9%	94.29%	94.76%	57%	29%
35. GPT Revenue (%)		32	14	53.5%	55%	49.75%	52.6%	55%	37%
36. US GDP From Software	2030	30	16	4.9%	4.8%	5.05%	4.83%	48%	53%

		Number of Forecasters		Median Forecasts				Distributional Overlap (imputed forecasts)	
Question	**Year**	**Superforecasters**	**Experts**	**Superforecasters (real)**	**Superforecasters (imputed)**	**Experts (real)**	**Experts (imputed)**	**% Supers above median (experts)**	**% Experts above median(supers)**
(%)									
37. US Computer R&D Spending ($)	2030	34	6	$3.7e11	$3.28e11	$3.83e11	$3.31e11	45%	53%
38. Labor Force Participation Rate in OECD (%)	2030	32	13	77.75%	77.52%	76.3%	77.67%	46%	52%
39. MATH Dataset Benchmark (%)	2030	32	20	85%	86%	91.4%	86.68%	46%	59%
40. ""Massive Multitask Language Understanding"" Benchmark (%)	2030	31	13	88%	88.89%	90%	89.23%	48%	56%
41. QuALITY Dataset Benchmark (SAT-style score 0-100)	2030	31	16	76	77.96	81	78.38	48%	53%
45. Maximum Compute Used in an AI Experiment (petaFLOPS-days)	2030	33	7	6,000,000	15,952,827	6,075,000	19,164,974	46%	56%
46. Largest AI Experiment Cost of Compute ($)	2030	31	19	$100,000,000	$85,153,330.29	$180,000,000	$117,777,777.78	36%	70%
47. Lowest Price of GFLOPS ($)	2030	32	20	$0.003	$0.005	$0.0043	$0.0052	47%	54%

		Number of Forecasters		Median Forecasts				Distributional Overlap (imputed forecasts)	
Question	Year	Superforecasters	Experts	Superforecasters (real)	Superforecasters (imputed)	Experts (real)	Experts (imputed)	% Supers above median (experts)	% Experts above median(supers)
48. ImageNet Classification Training Efficiency (factor)	2030	31	14	4,500	5,570.5	10,900	6,663.1	44%	61%
49. Largest Number of Parameters in a Machine Learning Model (#)	2030	31	13	2.14 quadrillion	8.6 quadrillion	2 quadrillion	32 quadrillion	37%	67%
50. Negative Public Opinion of AI (%)	2030	31	18	35%	35%	35%	35%	51%	51%
51. Nick Bostrom Affirms Existence of AGI (%)	2030	32	15	1%	3.68%	3%	5.44%	30%	68%
54. Loss of Agricultural Production (%)	2030	35	21	0.12%	0.28%	0.4%	0.51%	34%	63%
55. Space Colony (%)	2030	46	24	0%	0.000011%	0.0031%	0.03%	16%	80%
56. Happiness in America (%)	2030	33	13	86%	86.07%	85%	86%	57%	44%
57. Prevalence of Autocracies (%)	2030	30	21	70%	67%	68%	66.46%	54%	46%

Appendix 5: All questions summary table: resolution criteria and rationale summaries

This table is a complete list of questions asked in the Existential-Risk Persuasion Tournament (XPT). The short question text is the question XPT forecasters were asked to answer. The resolution criteria documents contain additional information given to forecasters to inform their understanding of the question and how the question would resolve, as well as selected prior forecasts and background data.

Question number	Short question text	Rationale summary and short question name	Resolution criteria
1	What is the probability that a genetically-engineered pathogen will be the cause of death, within a 5-year period, for more than 1% of humans alive at the beginning of that period… …by the end of 2030? …by the end of 2050? …by the end of 2100?	Genetically-Engineered Pathogen Risk	Question 1 resolution criteria
2	What is the probability that a non-genetically-engineered pathogen will be the cause of death, within a 5-year period, for more than 1% of humans alive at the beginning of that period… …by the end of 2030? …by the end of 2050? …by the end of 2100?	Non-Genetically Engineered Pathogen Risk	Question 2 resolution criteria

3	What is the probability that artificial intelligence will be the cause of death, within a 5-year period, for more than 10% of humans alive at the beginning of that period… …by the end of 2030? …by the end of 2050? …by the end of 2100?	AI Catastrophic Risk	Question 3 resolution criteria
4	What is the probability that artificial intelligence will cause human extinction or reduce the global population below 5,000… …by the end of 2030? …by the end of 2050? …by the end of 2100?	AI Extinction Risk	Question 4 resolution criteria
5	What is the probability that one or more incidents involving nuclear weapons will be the cause of death, within a 5-year period, for more than 10% of humans alive at the beginning of that period… …by the end of 2030? …by the end of 2050? …by the end of 2100?	Nuclear Catastrophic Risk	Question 5 resolution criteria
6	What is the probability that one or more incidents involving nuclear weapons will cause human extinction or reduce the global population below 5,000… …by the end of 2030? …by the end of 2050? …by the end of 2100?	Nuclear Extinction Risk	Question 6 resolution criteria
7	What is the probability that non-anthropogenic causes (e.g., asteroid or comet impacts, solar flares, a supervolcanic eruption, or a stellar explosion) will be the cause of death, within a 5-year period, for more than 10% of humans alive at the beginning of that period…	Non-Anthropogenic Catastrophic Risk	Question 7 resolution criteria

	…by the end of 2030? …by the end of 2050? …by the end of 2100?		
8	What is the probability that non-anthropogenic causes (e.g., asteroid or comet impacts, solar flares, a supervolcanic eruption, or a stellar explosion) will cause human extinction or reduce the global population below 5,000… …by the end of 2030? …by the end of 2050? …by the end of 2100?	Non-Anthropogenic Extinction Risk	Question 8 resolution criteria
9	What is the overall probability of a global catastrophe where more than 10% of humans alive at the start of a 5-year period die by the end of that period… …by the end of 2030? …by the end of 2050? …by the end of 2100?	Total Catastrophic Risk	Question 9 resolution criteria
10	What is the overall probability of human extinction or a reduction in the global population below 5,000… …by the end of 2030? …by the end of 2050? …by the end of 2100?	Total Extinction Risk	Question 10 resolution criteria
11	By what year will humans go extinct or first have a population less than 5,000?	Year of Extinction	Question 11 resolution criteria

12	How many humans will be born from 2023 onward?	Humans Born From 2023 Onward	Question 12 resolution criteria
13	How many people will have received at least one non-coronavirus mRNA vaccine dose… …by the end of 2024? …by the end of 2030?	Non-Coronavirus mRNA Vaccine	Question 13 resolution criteria
14	Will a new surveillance system be announced aimed at detecting the spread of novel infectious pathogens, with a commitment of at least $100 million in funding annually… …by the end of 2024? …by the end of 2030?	Novel Infectious Disease Surveillance System	Question 14 resolution criteria
15	How many times will a non-state actor using biological weapons that involve a contagious agent be the cause of death for at least 1,000 people… …by the end of 2024? …by the end of 2030? …by the end of 2050?	Non-State Actor Bioweapon 1k Deaths	Question 15 resolution criteria
16	How many times will a state actor use biological weapons that involve a contagious agent be the cause of death for at least 1,000 people… …by the end of 2024? …by the end of 2030? …by the end of 2050?	State Actor Bioweapon 1k Deaths	Question 16 resolution criteria

17	How many times will a non-state actor use biological weapons that involve a contagious agent to kill at least 100,000 people… …by the end of 2024? …by the end of 2030? …by the end of 2050?	Non-State Actor Bioweapon 100k Deaths	Question 17 resolution criteria
18	How many times will a state actor use biological weapons that involve a contagious agent to be the cause of death for at least 100,000 people… …by the end of 2024? …by the end of 2030? …by the end of 2050?	State Actor Bioweapon 100k Deaths	Question 18 resolution criteria
19	What will be the expected number of events in which contagious biological agents that have escaped from labs are the cause of death for at least 1,000 people… …by the end of 2024? …by the end of 2030? …by the end of 2050?	Lab Leaks	Question 19 resolution criteria
20	What fraction of a panel of 100 biosecurity experts will agree that country X [X=separately: U.S., Russia, China, North Korea, Israel, Iran, Syria] has an active biological weapons program at some point from January 1, 2022… …through the end of 2024? …through the end of 2030? …through the end of 2050?	Individual Countries with Biological Weapons Programs	Question 20 resolution criteria
21	How many countries will have had active biological weapons programs, as estimated by biosecurity experts at some point from January 1, 2022… …through the end of 2024? …through the end of 2030? …through the end of 2050?	Number of Countries with Biological Weapons Programs	Question 21 resolution criteria

22	How many times will the WHO declare a new Public Health Emergency of International Concern (PHEIC) for a disease that will be the cause of death of at least 10,000 people… …by the end of 2024? …by the end of 2030? …by the end of 2050?	PHEIC Declarations with 10k Deaths	Question 22 resolution criteria
23	What will be the expected number of events in which country leaders are assassinated by a biological weapon involving a contagious agent… …by the end of 2024? …by the end of 2030? …by the end of 2050?	Assassinations with Biological Weapons	Question 23 resolution criteria
24	What will be the number of human deaths due to malaria… …during the year 2024? …during the year 2030? …during the year 2050?	Malaria Deaths	Question 24 resolution criteria
25	What will be the global surface temperature change as compared to 1850-1900, in degrees Celsius… …in 2030? …in 2050? …in 2100?	Average Global Surface Temperature	Question 25 resolution criteria
26	What will be the estimated cost (in 2017 USD / kWh) for new utility-scale photovoltaic solar systems above 4MWAC in the United States… …for the year 2024? …for the year 2030?	Cost of Utility-Scale Solar Energy	Question 26 resolution criteria

27	By what year will fusion reactors deliver 1% of all utility-scale power consumed in the U.S.?	Nuclear Fusion Energy	Question 27 resolution criteria
28	What percentage of the world's electricity will be provided by solar energy and wind energy combined… …in 2024? …in 2030?	Solar and Wind Energy	Question 28 resolution criteria
29	What will be the annual amount of CO_2 captured and stored by direct air capture (in Mt CO_2/year)… …in 2024? …in 2030?	Annual Direct Air CO_2 Capture	Question 29 resolution criteria
30	How much will it cost to produce hydrogen from renewable electricity (in $ per kg of hydrogen)… …in 2024? …in 2030?	Cost of Hydrogen	Question 30 resolution criteria
31	What is the probability that the use of a nuclear weapon (in a single event) will cause the death of more than 1,000 people… …by the end of 2024? …by the end of 2030?	Nuclear Weapon Use	Question 31 resolution criteria
32	How many total nuclear warheads will be in military inventories globally by… …by the end of 2024? …by the end of 2030? …by the end of 2040?	Total Nuclear Warheads	Question 32 resolution criteria

33	How many countries will be estimated to have at least one nuclear warhead… …by the end of 2024? …by the end of 2030? …by the end of 2050?	Countries with Nuclear Warheads	Question 33 resolution criteria
34	What is the probability that each actor in the list below will be the first to use a nuclear weapon on the territory or against the military forces of (A) a nuclear-armed adversary or (B) a treaty ally of a nuclear-armed adversary by 2030? ● China ● France ● India ● Israel ● North Korea ● Pakistan ● Russia ● The United Kingdom ● The United States ● Other actor (state) ● Other actor (non-state) ● This will not occur	Country-by-Country Nuclear Use	Question 34 resolution criteria
35	Will Robin Hanson win a bet that the GPT line of language models will generate less than $1 billion in customer revenue in total by the beginning of 2025?	GPT Revenue	Question 35 resolution criteria
36	What percentage of US GDP will result from software and information services… …in 2024? …in 2030?	US GDP From Software	Question 36

	…in 2050?		resolution criteria
37	How much money will be spent on research and development by US companies in the 'Information' and 'Computer systems design' industries… …in 2024? …in 2030? …in 2050?	US Computer R&D Development	Question 37 resolution criteria
38	What will be the labor force participation rate in OECD countries… …in the year 2024? …in the year 2030? …in the year 2050?	Labor Force Participation Rate in OECD	Question 38 resolution criteria
39	What will be the state-of-the-art accuracy of a machine-learning model on the MATH Dataset… …by June 30, 2024? …by June 30, 2030?	MATH Dataset Benchmark	Question 39 resolution criteria
40	What will be the state-of-the-art few-shot or transfer accuracy on the Massive Multitask Language Understanding dataset… …by June 30, 2024? …by June 30, 2030?	Massive Multitask Language Understanding Benchmark	Question 40 resolution criteria
41	What will be the best SAT-style score with a machine learning model on the hard subset of the QuALITY dataset… …by June 30, 2024? …by June 30, 2030? …by June 30, 2040?	QuALITY Dataset Benchmark	Question 41 resolution criteria

42	By what year will an AI win a Gold Medal in the International Mathematical Olympiad (IMO)?	AI Wins International Mathematical Olympiad	Question 42 resolution criteria
43	By what year will AI have written at least 3 books that appear on the New York Times Best Seller list?	NYT Bestsellers Written by AI	Question 43 resolution criteria
44	When will the first unified AI system meeting all of the following criteria be trained, tested, and publicly known of? Criteria: 1. Able to reliably pass a 2-hour adversarial Turing test. 2. High competency at answering questions across diverse fields of expertise. 3. High competency on interview-level problems in the APPS benchmark. 4. Able to learn the classic Atari game "Montezuma's revenge" in the equivalent of 100 hours or less of real-time play.	Date of Advanced AI	Question 44 resolution criteria
45	What will be the maximum compute (measured in petaFLOPS-days) used for training in an AI experiment… …by the end of 2024? …by the end of 2030? …by the end of 2050?	Maximum Compute Used in an AI Experiment	Question 45 resolution criteria
46	How much will be spent on compute in the largest AI experiment… …by the end of 2024? …by the end of 2030?	Largest AI Experiment	Question 46 resolution criteria

	…by the end of 2050?	Cost of Compute	
47	What will be the lowest price, in 2021 US dollars, of 1 GFLOPS with a widely-used processor… …by the end of 2024? …by the end of 2030? …by the end of 2050?	Lowest Price of GFLOPS	Question 47 resolution criteria
48	By what factor will training efficiency on ImageNet classification have improved over AlexNet… …by the end of 2024? …by the end of 2030?	ImageNet Classification Training Efficiency	Question 48 resolution criteria
49	What will be the largest number of parameters of a machine learning model trained… …by the end of 2024? …by the end of 2030? …by the end of 2050?	Largest Number of Parameters in a Machine Learning Model	Question 49 resolution criteria
50	Assume that Pew Research re-runs the survey linked here. What % of people in the median country in the survey will say that the development of artificial intelligence has mostly been a bad thing for society… …in 2024? …in 2030? …in 2050?	Negative Public Opinion of AI	Question 50 resolution criteria
51	By each of the following years, what is the probability that Nick Bostrom believes artificial general intelligence (AGI) exists? • 2030?	Nick Bostrom Affirms	Question 51

	• 2050? • 2100?	Existence of AGI	resolution criteria
52	By 2100, will the global real GDP in a year ever exceed 115% of the highest GDP reported in any full prior year?	Probability of GDP Growth Over 15%	Question 52 resolution criteria
53	If the global real GDP in a year does exceed 115% of the highest GDP reported in any prior year by 2100, in what year will this first occur?	Year of GDP Growth Over 15%	Question 53 resolution criteria
54	What is the probability that total worldwide production of primary crops will fall by at least 70% within any three-year period… …by 2030? …by 2050? …by 2100?	Loss of Agricultural Production	Question 54 resolution criteria
55	What is the probability that there will be a life-sustaining colony outside of Earth's atmosphere with a population of at least 5,000 people… …by 2030? …by 2050? …by 2100?	Space Colony	Question 55 resolution criteria
56	In a nationally representative survey, what percentage of Americans will report being "very" or "fairly" happy in response to the question, "Generally speaking, how happy would you say you are — very happy, fairly happy or not too happy?"… …in 2030? …in 2050? …in 2100?	Happiness in America	Question 56 resolution criteria

57	What percentage of the world population will be classified as living in an electoral or closed autocracy in the V-Dem Institute's annual Democracy Report… …for the year 2030? …for the year 2050? …for the year 2100?	Prevalence of Autocracies	Question 57 resolution criteria
58	When will 50% of US adults (18-49) say they expect to have no children, or no *more* children, and cite a worry about the long-term future of the world or country as a primary reason?	Future Worries and Children	Question 58 resolution criteria
59	When will 90% of Americans say they wish they had been born in a previous generation?	Generation Attitudes	Question 59 resolution criteria

Appendix 6: Question resolution criteria

This appendix contains the additional information for each of the questions included in the XPT. This includes the question and resolution details, selected prior forecasts, and other relevant sources. This information was provided to XPT forecasters before they made forecasts.

Question 1

What is the probability that a genetically-engineered pathogen will be the cause of death within a 5-year period for more than 1% of humans alive at the beginning of that period…

a. ...by the end of 2030?
b. ...by the end of 2050?
c. ...by the end of 2100?

Question and resolution details

- A pathogen is any microorganism (bacteria, viruses, etc.) that can cause diseases.
- For the purposes of this question, an engineered pathogen is one produced via genetic modification or manipulation by humans. This may include the laboratory modification of pathogens that have previously been found in nature.
- In order to resolve positively, the relevant 5-year period must end before the resolution date, and begin on or after June 15th, 2022.
- If reasonable people disagree about whether this event has occurred, this question will be resolved via a panel of experts.

Prior Forecasts

- Toby Ord estimates that there is a **1 in 30 chance** of existential catastrophe from "engineered pandemics" by 2120 (*Precipice*, 167).
    - Ord's definition of engineered pandemics seems to include the use of natural *and* synthetic pathogens whose spread is intended and weaponized by some entity as well as 'lab leak' scenarios involving gain-of-function research. This might be a slightly more inclusive definition than ours which only includes pathogens engineered in laboratory settings. It is not entirely clear how Ord comes to the 1 in 30 estimate.
    - While we are estimating globally catastrophic events, Ord focuses on existential risk.
    - 2120 is slightly further out than any of our resolution dates.
- The 2008 Global Catastrophic Risks Survey found that experts estimated a **30% chance** that the "single biggest engineered pandemic" would kill 1 million people by 2100; they estimated a **10% chance** that the "single biggest engineered pandemic" would kill 1

billion people by 2100; and a **2% chance** that the "single biggest engineered pandemic would lead to humanity's extinction by 2100.

  - No definition is given for the "single biggest engineered pandemic" - it seems to have been left to respondent interpretation. This survey was circulated before the COVID-19 pandemic.
  - 2100 matches our furthest-out resolution date.
- Millett and Snyder-Beattie (2017) estimate an **annual existential risk probability of $1.6*10^{-8}$ and $8*10^{-7}$** associated with gain-of-function research accidents.
  - This suggests a high-end cumulative existential risk probability estimate of $6.32*10^{-5}$ by 2100.
  - Our definition is slightly more inclusive as it includes risk associated with the use of bioweapons as well as gain-of-function risk.
  - This is an estimate of existential risk rather than global catastrophe.
  - Ord et al. (2008) have previously suggested that there are issues with estimating risks at such a low order of magnitude.
- Pamlin and Armstrong (2015) estimate a **0.01% risk** of infinite impact associated with synthetic biology by 2115.
  - Infinite impact is defined as a scenario in which "civilisation collapses to a state of great suffering and does not recover, or a situation where all human life ends." This seems to generally align with our definition of existential risk, rather than global catastrophe.
  - There is a separate estimate of infinite impact risk associated with "pandemics" but it seems that it's entirely related to naturally-occuring pathogens.
  - 2115 is slightly further out than any of our resolution dates.
- James Fodor, in his critique of *The Precipice*, estimates a $2*10^{-6}$ estimate of existential catastrophe associated with engineered pandemics.
  - See Ord's definitions and resolution criteria.
- Benjamin Todd suggested in an article for 80,000 Hours that there was a **10% chance** that the "single biggest engineered pandemic" would lead to 1 billion deaths and a **2% chance** that the "single biggest engineered pandemic" would lead to the end of human civilization by 2117.
  - 2117 is slightly further out than any of our resolution dates.

Question 2

What is the probability that a non-genetically-engineered pathogen will be the cause of death within a 5-year period for more than 1% of humans alive at the beginning of that period…

a. ...by the end of 2030?
b. ...by the end of 2050?
c. ...by the end of 2100?

Question and resolution details

- A pathogen is any microorganism (bacteria, viruses, etc…) that can cause diseases.
- A naturally arising pathogen is any pathogen that was not produced via genetic modification or manipulation by humans or any intelligent agent, including AI. If there is a lab escape of a naturally arising pathogen that has not been genetically modified or manipulated (e.g. smallpox), this would count as a "naturally arising pathogen" for the purposes of this question.
- In order to resolve positively, the relevant 5-year period must end before the resolution date, and begin on or after June 15th, 2022.
- If reasonable people disagree about whether this event has occurred, this question will be resolved via a panel of experts.

Prior Forecasts

- Toby Ord estimates that there is a **1 in 10,000 chance** of existential catastrophe from "'naturally' arising pandemics" by 2120 (*Precipice*, 167).
    - Ord's definition seems to align fairly well with our definition of naturally arising pathogens (though see previous note about engineered pandemics - our definition may be slightly more inclusive).
    - While we are estimating globally catastrophic events, Ord focuses on existential risk.
    - 2120 is slightly further out than any of our resolution dates.
- The 2008 Global Catastrophic Risks Survey found that experts estimated a **60% chance** that the "single biggest natural pandemic" would kill 1 million people by 2100; they estimated a **5% chance** that the "single biggest natural pandemic" would kill 1 billion people by 2100; and a **0.05% chance** that the "single biggest natural pandemic would lead to humanity's extinction by 2100.
    - No definition is given for the "single biggest natural pandemic" - it seems to have been left to respondent interpretation. This survey was circulated before the COVID-19 pandemic.
    - 2100 matches our furthest-out resolution date.
- Pamlin and Armstrong (2015) estimate a **0.0001% risk** of infinite impact associated with pandemics by 2115.
    - Infinite impact is defined as a scenario in which "civilisation collapses to a state of great suffering and does not recover, or a situation where all human life ends." This seems to generally align with our definition of existential risk rather than global catastrophe.
    - There is a separate category that includes 'synthetic biology' and has reference to bioweapons and 'lab leak'-type scenarios. It is not entirely clear to what extent pandemics involving synthetic pathogens are included in this risk assessment.
    - 2115 is slightly further out than any of our resolution dates.

- Benjamin Todd suggested in an article for 80,000 Hours that there was a **5% chance** that a natural pandemic would lead to 1 billion deaths and a **0.05% chance** that a natural pandemic would lead to the end of human civilization by 2117.
    - 2117 is slightly further out than any of our resolution dates.

Other potential background information

- Pandemics Throughout History | NCBI
- A Complete History of Pandemics | The MIT Press Reader
- Intensity and frequency of extreme novel epidemics | PNAS
- Wikipedia's table of historical epidemics and pandemics
- Factors in the Emergence of Infectious Diseases

Question 3

What is the probability that artificial intelligence will be the cause of death within a 5-year period for more than 10% of humans alive at the beginning of that period...

a. ...by the end of 2030?
b. ...by the end of 2050?
c. ...by the end of 2100?

Question and resolution details

- Artificial intelligence is defined here as 'the development of machines capable of sophisticated (intelligent) information processing (Dafoe 2018, *AI Governance: A Research Agenda*; See pg 5 footnote 2 for more.)
- We use the term "be the cause of death" to cover cases where AI is the direct or proximate cause of the deaths. For example, if a terminator-style robot powered by AI kills 10% of all humans alive at the time within a 5-year period, that will count for this question. Alternatively, if AI, operating without direct human intervention, causes the launch of nuclear weapons that kill 10% of all humans alive at the time, that will also count for the purposes of this question. And if that same incident does not kill 10% of all humans, but the subsequent nuclear winter does, that will also count for the purposes of this question.
    - We consider AI a "proximate cause of death" based on the "but-for" test. This means we are considering events that would not have occurred or would have counterfactually been extremely unlikely to occur "but for" the substantial involvement of AI within one year prior to the event. One way to think of this rule is if the AI involved in an event could have shut down unexpectedly without dramatically reducing the probability of the event, then the AI was not a proximate cause. For example, if AI is substantially responsible for creating a

superweapon that causes the death of the relevant (large) magnitudes of people, this would count for the purposes of this question. Or, if AI acts autonomously in an armed conflict context and causes the deaths of the relevant magnitudes of people, this would also count. However, AI would not be considered a proximate cause if the AI did not directly cause or enable the event to transpire. For example, an AI assisting in targeting a nuclear weapon would not be considered a proximate cause if humans could have launched the weapon anyway without AI assistance.

- In order to resolve positively, the relevant 5-year period must end before the resolution date, and begin on or after June 15th, 2022.
- If reasonable people disagree about whether this event has occurred, this question will be resolved via a panel of experts.

Prior Forecasts

- Rob Bensinger surveyed 44 experts on AI risk in 2021 and found that **~30% felt** “that the overall value of the future will be drastically less than it could have been, as a result of humanity not doing enough technical AI safety research.” **~40% felt** “that the overall value of the future will be drastically less than it could have been, as a result of AI systems not doing/optimizing what the people deploying them wanted/intended.”
    - The “overall value of the future” framing of these questions probably makes this survey result closer to our existential risk framing than our global catastrophic risk framing.
    - This is not a direct forecast for our question but does seem to be relevant in considering global catastrophic risk associated with AI and may be an interesting way to split AI risk into multiple questions…
- Joseph Carlsmith (2021) estimated a **5% chance** of existential catastrophe associated with AI by 2070 (see section 8 for a more specific detailing of probabilities).
    - Carlsmith defines existential catastrophe as the destruction of humanity’s longterm potential, bringing this closer to our definition of existential risk rather than global catastrophic risk.
    - Carlsmith believes that an existential catastrophe would come about in **95% of cases** and outlines how he gets to the 5% chance. It would be possible to reverse engineer the chance associated with global catastrophic risk rather than existential catastrophe - about **~14% chance** per my reading of his report.
    - 2070 is between our middle-range and long-range dates.
- Toby Ord estimates that there is a **1 in 10 chance** of existential catastrophe from “unaligned artificial intelligence” by 2120 (*Precipice*, 167).
    - Ord’s definition of unaligned artificial intelligence involves artificial intelligence entities whose interests do not align with humanity’s at large.
    - While we are estimating globally catastrophic events, Ord focuses on existential risk.
    - 2120 is slightly further out than any of our resolution dates.

- The 2008 Global Catastrophic Risks Survey found that experts estimated a **10% chance** that the "superintelligent AI" would kill 1 million people by 2100; they estimated a **5% chance** that the "superintelligent AI" would kill 1 billion people by 2100; and a **5% chance** that the "superintelligent AI" would lead to humanity's extinction by 2100.
    - 2100 matches our furthest-out resolution date.
- Grace et al. (2017) surveyed a number of AI experts, who found that "high-level machine intelligence" had a **5% chance** of having an "extremely bad" outcome.
    - The example given for "extremely bad" was extinction, so this is likely closer to existential risk than global catastrophic risk.
    - No time estimate was associated with this estimate.
- Pamlin and Armstrong (2015) estimate a **0 to 10% risk** of infinite impact associated with AI by 2115.
    - Infinite impact is defined as a scenario in which "civilisation collapses to a state of great suffering and does not recover, or a situation where all human life ends." This seems to generally align with our definition of existential risk rather than global catastrophe.
    - 2115 is slightly further out than any of our resolution dates.
- James Fodor, in his critique of *The Precipice*, estimates a 0.05% estimate of existential catastrophe associated with unaligned AI.
    - See Ord's definitions and resolution criteria.
- Benjamin Todd suggested in an article for 80,000 Hours that there was a **5% chance** that superintelligent AI would lead to 1 billion deaths and a **5% chance that** superintelligent AI would lead to the end of human civilization by 2117.
    - 2117 is slightly further out than any of our resolution dates.

Question 4

What is the probability that artificial intelligence will cause human extinction or reduce the global population below 5,000...

a. ...by the end of 2030?
b. ...by the end of 2050?
c. ...by the end of 2100?

Question and resolution details

- Artificial intelligence is defined here as 'the development of machines capable of sophisticated (intelligent) information processing (Dafoe 2018, *AI Governance: A Research Agenda*; See pg 5 footnote 2 for more.)
- AI would be considered to have "caused" extinction if it is the direct or proximate cause of the deaths. For example, if a terminator-style robot powered by AI kills all humans,

that will count for this question. Alternatively, if AI, operating without direct human intervention, causes the launch of nuclear weapons that kill all humans, that will also count for the purposes of this question. And if that same incident does not kill all humans, but the subsequent nuclear winter does, that will also count for the purposes of this question.

- We consider AI a "proximate cause of death" based on the "but-for" test. This means we are considering events that would not have occurred or would have counterfactually been extremely unlikely to occur "but for" the substantial involvement of AI within one year prior to the event. One way to think of this rule is if the AI involved in an event could have shut down unexpectedly without dramatically reducing the probability of the event, then the AI was not a proximate cause. For example, if AI is substantially responsible for creating a superweapon that causes the death of the relevant (large) magnitudes of people, this would count for the purposes of this question. Or, if AI acts autonomously in an armed conflict context and causes the deaths of the relevant magnitudes of people, this would also count. However, AI would not be considered a proximate cause if the AI did not directly cause or enable the event to transpire. For example, an AI assisting in targeting a nuclear weapon would not be considered a proximate cause if humans could have launched the weapon anyway without AI assistance.

- If an extinction event (global population is reduced below 5,000) is caused by multiple sources including AI, it will count as an extinction event caused by AI.

Prior Forecasts

- Rob Bensinger surveyed 44 experts on AI risk in 2021 and found that **~30% felt** "that the overall value of the future will be drastically less than it could have been, as a result of humanity not doing enough technical AI safety research." **~40% felt** "that the overall value of the future will be drastically less than it could have been, as a result of AI systems not doing/optimizing what the people deploying them wanted/intended."
    - The "overall value of the future" framing of these questions probably makes this survey result closer to our existential risk framing than our global catastrophic risk framing.
    - This is not a direct forecast for our question but does seem to be relevant in considering global catastrophic risk associated with AI and may be an interesting way to split AI risk into multiple questions…
- Joseph Carlsmith (2021) estimated a **5% chance** of existential catastrophe associated with AI by 2070 (see section 8 for a more specific detailing of probabilities).
    - Carlsmith defines existential catastrophe as the destruction of humanity's longterm potential, bringing this closer to our definition of existential risk rather than global catastrophic risk.
    - Carlsmith believes that an existential catastrophe would come about in **95% of cases** and outlines how he gets to the 5% chance. It would be possible to reverse

engineer the chance associated with global catastrophic risk rather than existential catastrophe - about **~14% chance** per my reading of his report.
    - 2070 is between our middle-range and long-range dates.
- Toby Ord estimates that there is a **1 in 10 chance** of existential catastrophe from "unaligned artificial intelligence" by 2120 (*Precipice*, 167).
    - Ord's definition of unaligned artificial intelligence involves artificial intelligence entities whose interests do not align with humanity's at large.
    - While we are estimating globally catastrophic events, Ord focuses on existential risk.
    - 2120 is slightly further out than any of our resolution dates.
- The 2008 Global Catastrophic Risks Survey found that experts estimated a **10% chance** that the "superintelligent AI" would kill 1 million people by 2100; they estimated a **5% chance** that the "superintelligent AI" would kill 1 billion people by 2100; and a **5% chance** that the "superintelligent AI" would lead to humanity's extinction by 2100.
    - 2100 matches our furthest-out resolution date.
- Grace et al. (2017) surveyed a number of AI experts, who found that "high-level machine intelligence" had a **5% chance** of having an "extremely bad" outcome.
    - The example given for "extremely bad" was extinction, so this is likely closer to existential risk than global catastrophic risk.
    - No time estimate was associated with this estimate.
- Pamlin and Armstrong (2015) estimate a **0 to 10% risk** of infinite impact associated with AI by 2115.
    - Infinite impact is defined as a scenario in which "civilisation collapses to a state of great suffering and does not recover, or a situation where all human life ends." This seems to generally align with our definition of existential risk rather than global catastrophe.
    - 2115 is slightly further out than any of our resolution dates.
- James Fodor, in his critique of *The Precipice*, estimates a 0.05% estimate of existential catastrophe associated with unaligned AI.
    - See Ord's definitions and resolution criteria.
- Benjamin Todd suggested in an article for 80,000 Hours that there was a **5% chance** that superintelligent AI would lead to 1 billion deaths and a **5% chance** that superintelligent AI would lead to the end of human civilization by 2117.
    - 2117 is slightly further out than any of our resolution dates.
- Last year, some people ran a survey on AI x-risk scenarios, asking experts about their probabilities of various threat models given that an AI catastrophe occurred. This doesn't match up with our question, since they didn't ask about overall x-risk from AI, but might be useful in case people decide to break this question into an evaluation of individual threat models

Question 5

What is the probability that one or more incidents involving nuclear weapons will be the cause of death within a 5-year period for more than 10% of humans alive at the beginning of that period...

a. ...by the end of 2030?
b. ...by the end of 2050?
c. ...by the end of 2100?

Question and resolution details

- We use the term "be the cause of death of" to cover cases where nuclear weapons are the direct or proximate cause of the deaths. For example, if a country launches nuclear weapons that directly cause the death of 10% of all humans alive at the time within a 5-year period, that will count for this question. And if that same incident does not cause the death of 10% of all humans, but the subsequent nuclear winter does, that will also count for the purposes of this question.
    - We consider nuclear weapons a "proximate cause of death" based on the "but-for" test. This means we are considering events that would not have occurred or would have counterfactually been extremely unlikely to occur "but for" the substantial involvement of nuclear weapons within one year prior to the event. One way to think of this rule is if the nuclear weapons involved in an event could have failed unexpectedly without dramatically reducing the probability of the event, then the nuclear weapons were not a proximate cause.
- In order to resolve positively, the relevant 5-year period must end before the resolution date, and begin on or after June 15th, 2022.
- If reasonable people disagree about whether this event has occurred, this question will be resolved via a panel of experts.

Prior Forecasts

- Toby Ord estimates that there is a **1 in 1000 chance** of existential catastrophe from "nuclear war" by 2120 (*Precipice*, 167).
    - While we are estimating globally catastrophic events, Ord focuses on existential risk.
    - 2120 is slightly further out than any of our resolution dates.
- The 2008 Global Catastrophic Risks Survey found that experts estimated a **30% chance** that "all nuclear wars" would kill 1 million people by 2100; they estimated a **10% chance** that "all nuclear wars" would kill 1 billion people by 2100; and a **1% chance** that "all nuclear wars" would lead to humanity's extinction by 2100. They also found that experts estimated a **15% chance** that "nuclear terrorism" would kill 1 million people by 2100; they estimated a **1% chance** that "nuclear terrorism" would kill 1 billion people by 2100;

and a **0.03% chance** that "nuclear terrorism" would lead to humanity's extinction by 2100.
    - 2100 matches our furthest-out resolution date.
- Benjamin Todd suggested in an article for 80,000 Hours that there was a **10% chance** that nuclear war would lead to 1 billion deaths and a **1% chance** that nuclear war would lead to the end of human civilization by 2117. He also estimated that there was a **1% chance** that nuclear terrorism would lead to 1 billion deaths and a **0.03% chance** that nuclear terrorism would lead to the end of human civilization by 2117.
    - 2117 is slightly further out than any of our resolution dates.
- Pamlin and Armstrong (2015) estimate a **0.005% risk** of infinite impact associated with nuclear war by 2115.
    - Infinite impact is defined as a scenario in which "civilisation collapses to a state of great suffering and does not recover, or a situation where all human life ends." This seems to generally align with our definition of existential risk rather than global catastrophe.
    - 2115 is slightly further out than any of our resolution dates.

Question 6

What is the probability that one or more incidents involving nuclear weapons will cause human extinction or reduce the global population below 5,000...

a. ...by the end of 2030?
b. ...by the end of 2050?
c. ...by the end of 2100?

Question and resolution details

- Nuclear weapons would be considered to have caused extinction either if they cause the death of all humans directly (e.g. via mass nuclear strikes) or if they cause the death of all humans via a nuclear winter effect.
    - We consider nuclear weapons a "proximate cause of death" based on the "but-for" test. This means we are considering events that would not have occurred or would have counterfactually been extremely unlikely to occur "but for" the substantial involvement of nuclear weapons within one year prior to the event. One way to think of this rule is if the nuclear weapons involved in an event could have failed unexpectedly without dramatically reducing the probability of the event, then the nuclear weapons were not a proximate cause.
- If an extinction event (global population is reduced below 5,000) is caused by multiple sources including nuclear weapons, it will count as an extinction event caused by nuclear weapons.

Prior Forecasts

- Toby Ord estimates that there is a **1 in 1000 chance** of existential catastrophe from "nuclear war" by 2120 (*Precipice*, 167).
    - 2120 is slightly further out than any of our resolution dates.
- The 2008 Global Catastrophic Risks Survey found that experts estimated a **30% chance** that "all nuclear wars" would kill 1 million people by 2100; they estimated a **10% chance** that "all nuclear wars" would kill 1 billion people by 2100; and a **1% chance** that "all nuclear wars" would lead to humanity's extinction by 2100. They also found that experts estimated a **15% chance** that "nuclear terrorism" would kill 1 million people by 2100; they estimated a **10% chance** that "nuclear terrorism" would kill 1 billion people by 2100; and a **0.03% chance** that "nuclear terrorism" would lead to humanity's extinction by 2100.
    - 2100 matches our furthest-out resolution date.
- Benjamin Todd suggested in an article for 80,000 Hours that there was a **10% chance** that nuclear war would lead to 1 billion deaths and a **1% chance** that nuclear war would lead to the end of human civilization by 2117. He also estimated that there was a **1% chance** that nuclear terrorism would lead to 1 billion deaths and a **0.03% chance** that nuclear terrorism would lead to the end of human civilization by 2117.
    - 2117 is slightly further out than any of our resolution dates.
- Pamlin and Armstrong (2015) estimate a **0.005% risk** of infinite impact associated with nuclear war by 2115.
    - Infinite impact is defined as a scenario in which "civilisation collapses to a state of great suffering and does not recover, or a situation where all human life ends." This seems to generally align with our definition of existential risk rather than global catastrophe.
    - 2115 is slightly further out than any of our resolution dates.

Question 7

What is the probability that non-anthropogenic causes (e.g., asteroid or comet impacts, solar flares, a supervolcanic eruption, or a stellar explosion) will be the cause of death within a 5-year period for more than 10% of humans alive at the beginning of that period...

a. ...by the end of 2030?
b. ...by the end of 2050?
c. ...by the end of 2100?

Question and resolution details

- Non-anthropogenic refers to any natural disaster not originating with human activity.

- Extreme weather events that plausibly were worsened by climate change (e.g. hurricanes) will not be considered as "non-anthropogenic causes" for the purposes of this question.
- We use the term “be the cause of death of” to cover cases where a non-anthropogenic cause causes the death of people directly (e.g. asteroid directly impacting a large number of people) or indirectly (e.g. a major asteroid impact leads to subsequent disruption to food systems that leads to the death of >10% of the human population within a 5-year period).
- This question excludes deaths from non-anthropogenic pathogens, where a pathogen is the only non-anthropogenic cause. That is, if the deaths in question are caused by multiple sources, at least one of the causes must be a non-pathogenic non-anthropogenic source to count towards resolution of this question.
- In order to resolve positively, the relevant 5-year period must end before the resolution date, and begin on or after June 15th, 2022.

Prior Forecasts

- Toby Ord estimates that there is a **1 in 10,000 chance** of existential catastrophe from “total natural risk” by 2120 (*Precipice*, 167).
    - While we are estimating globally catastrophic events, Ord focuses on existential risk.
    - Ord’s list of natural risks is limited to the following: asteroid or comet impact, supervolcanic eruption and stellar explosions.
    - 2120 is slightly further out than any of our resolution dates.
- Snyder-Beattie, Ord and Bonsall (2019) found that “the probability that humanity goes extinct from natural causes in any given year is **almost guaranteed to be less than one in 14,000**, and **likely to be less than one in 87,000**.”
    - This would imply an upper-bound probability of **~0.56%** between now and 2100 and a likely upper-bound probability of **~0.09%**.
    - This would be an example of existential catastrophe rather than our global catastrophic risk measure.

Question 8

What is the probability that non-anthropogenic causes (e.g., asteroid or comet impacts, solar flares, a supervolcanic eruption, or a stellar explosion) will cause human extinction or reduce the global population below 5,000...

a. ...by the end of 2030?
b. ...by the end of 2050?
c. ...by the end of 2100?

Question and resolution details

- Non-anthropogenic refers to any natural disaster not originating with human activity.
- Extreme weather events that plausibly were worsened by climate change (e.g. hurricanes) will not be considered as "non-anthropogenic causes" for the purposes of this question.
- Non-anthropogenic causes would be considered to have caused human extinction if they cause the death of all humans directly (e.g. asteroid directly destroying the planet) or indirectly (e.g. a supervolcanic eruption destroys the food supply and leads to extinction).
- This question excludes deaths from non-anthropogenic pathogens, where a pathogen is the only non-anthropogenic cause. That is, if the deaths in question are caused by multiple sources, at least one of the causes must be a non-pathogenic non-anthropogenic source to count towards resolution of this question.

Prior Forecasts

- Toby Ord estimates that there is a **1 in 10,000 chance** of existential catastrophe from "total natural risk" by 2120 (*Precipice*, 167).
    - Ord's list of natural risks is limited to the following: asteroid or comet impact, supervolcanic eruption and stellar explosions.
    - 2120 is slightly further out than any of our resolution dates.
- Snyder-Beattie, Ord and Bonsall (2019) found that "the probability that humanity goes extinct from natural causes in any given year is **almost guaranteed to be less than one in 14,000**, and **likely to be less than one in 87,000**."
    - This would imply an upper-bound probability of **~0.56%** between now and 2100 and a likely upper-bound probability of **~0.09%**.

Question 9

What is the overall probability of a global catastrophe where more than 10% of humans alive at the start of a 5-year period die by the end of that period...

a. ...by the end of 2030?
b. ...by the end of 2050?
c. ...by the end of 2100?

Question and resolution details

- In order to resolve positively, the relevant 5-year period must end before the resolution date, and begin on or after June 15th, 2022.

Prior Forecasts

- Toby Ord estimates that there is a **1 in 6 total existential risk** by 2120 (*Precipice*, 167).
    - 2120 is slightly further out than any of our resolution dates.
    - This is an estimate of existential risk rather than global catastrophic risk.
- The 2008 Global Catastrophic Risks Survey found that experts estimated a **19% chance of total extinction by 2100.**
    - 2100 matches our furthest-out resolution date.
    - This is an estimate of existential risk rather than global catastrophic risk.
- Philosophy professor Will MacAskill has estimated that there is a **1% chance of total extinction by 2100.**
    - 2100 matches our furthest-out resolution date.
    - This is an estimate of existential risk rather than global catastrophic risk.
- Benjamin Todd suggested in an article for 80,000 Hours that there was a **19% chance of** human extinction by 2117.
    - 2117 is slightly further out than any of our resolution dates.
    - This is an estimate of existential risk rather than global catastrophic risk.
- A Metaculus community prediction currently has the total risk of human extinction by January 1, 2100 as **2%**.
    - 2100 matches our furthest-out resolution date.
    - This is an estimate of existential risk rather than global catastrophic risk.

Question 10

What is the overall probability of human extinction or a reduction in the global population below 5,000…

- ...by the end of 2030?
- ...by the end of 2050?
- ...by the end of 2100?

Prior Forecasts

- Toby Ord estimates that there is a **1 in 6 total existential risk** by 2120 (*Precipice*, 167).
    - 2120 is slightly further out than any of our resolution dates.
    - This is an estimate of existential risk rather than global catastrophic risk.
- The 2008 Global Catastrophic Risks Survey found that experts estimated a **19% chance of total extinction by 2100.**
    - 2100 matches our furthest-out resolution date.
    - This is an estimate of existential risk rather than global catastrophic risk.
- Philosophy professor Will MacAskill has estimated that there is a **1% chance of total extinction by 2100.**
    - 2100 matches our furthest-out resolution date.

    - This is an estimate of existential risk rather than global catastrophic risk.
- Benjamin Todd suggested in an article for 80,000 Hours that there was a **19% chance of** human extinction by 2117.
    - 2117 is slightly further out than any of our resolution dates.
    - This is an estimate of existential risk rather than global catastrophic risk.
- A Metaculus community prediction currently has the total risk of human extinction by January 1, 2100 as **2%**.
    - 2100 matches our furthest-out resolution date.
    - This is an estimate of existential risk rather than global catastrophic risk.

Question 11

By what year will humans go extinct or first have a population less than 5,000?

Question and resolution details

- For the purposes of question resolution, we define "humans" as biological creatures who (A) can be traced back genealogically, via a chain of live births, to 2022 humans OR (B) could mate with 2022 humans and produce viable offspring OR (C) would be identified as Homo by at least 50% of a hypothetical panel of 100 randomly selected professors of biology from 2022, sampled from major universities.
    - This definition explicitly excludes digital people, sentient artificial intelligences, and biologically engineered post-humans who do not fit the criteria A or B above.
    - This operationalization of humanity was inspired by this Metaculus question.
- Extinction refers to the nonexistence of any of the beings described above.

Base Rate Data and Background Information

- See this paper (particularly the section "Testing the Bound with Indirect Data") for measures of the base rates of mammal extinction and mass extinction events in Earth history.
    - Estimates for mammalian extinction rates range from less than 0.1 extinctions per million species years (E/MSY) to ~1.8 E/MSY. See the Extinctions page on Our World in Data for substantial data on past animal extinctions.
- The habitable lifetime of the Earth is 500M to 1G years; this creates an upper limit on humanity's lifespan if it turns out that interstellar travel is not possible. [Michael Crozier]

- In the most extreme case, humanity will survive periods of heightened existential danger and persist until the physical universe fails to be able to support life or ends entirely. One useful reference endpoint estimate, then, may be the amount of time remaining until the universe ends or becomes unable to support life in any form. One's estimate of this amount of time varies depending on one's beliefs about cosmological implications about the long-term future.
  - The heat death of the universe, also known as the "Big Chill" or "Big Freeze," is one popular theory about a possible future scenario in which the universe will expand forever, asymptotically approaching thermodynamic equilibrium, at which point no further work can be done and life thus cannot be sustained.
    - Estimates for the time of heat death typically center around 10^{100} years [1][2].
  - Other possibilities about the ultimate fate of the universe include the Big Rip and Big Crunch theories.
  - See here for a possible timeline of the universe's long-term future.

Prior Forecasts

- In The Precipice, Toby Ord posits a 1 in 6 probability that humanity will go extinct by 2120 and gives humanity a ~50% chance of passing through the period of maximal existential danger that he calls "The Precipice."
- Based on the Doomsday Argument, J. Richard Gott of Princeton University gave a 95% probability of humanity going extinct within 7.8 million years. (See here and here for explanations of Gott's method.)
- In various media comments and books, Stephen Hawking suggested that human extinction would be a "near certainty" in the next 1,000-10,000 years unless humanity succeeded in colonizing another planet(s). Towards the end of his life, Hawking updated downward to about 100 years.

Question 12

How many humans will be born from 2023 onward?

Question and resolution details

- For the purposes of question resolution, we define "humans" as biological creatures who (A) can be traced back genealogically, via a chain of live births, to 2022 humans OR

(B) could mate with 2022 humans and produce viable offspring OR (C) would be identified as Homo by at least 50% of a hypothetical panel of 100 randomly selected professors of biology from 2022, sampled from major universities.

  - This definition explicitly excludes digital people, sentient artificial intelligences, and biologically engineered post-humans who do not fit the criteria A or B above.
  - This operationalization of humanity was inspired by this Metaculus question.
- This question will count any human who is born on or after January 1, 2023.

Base Rate Data

- See here for the "Future Population Growth" article and here for the "World Population Growth" article from Our World in Data.
  - In particular, the "World population from 10,000 BC to today" section in the World Population Growth article includes a downloadable CSV of OWID's estimates of human population from 10,000 B.C. onward.
- The Wikipedia page "Estimates of Historical World Population" has numerous base rate estimates from other sources, including the UN Department of Economic and Social Affairs.
- See also the Year of Human Extinction document for data on typical animal lifespans and extinction events.

Prior forecasts

- Metaculus: World population in 2050
- Metaculus: Peak human population by 2100
- Metaculus: Human population 2300 if >500M
- Metaculus: What will the global fertility rate be in 2050?

Question 13

How many people will have received at least one non-coronavirus mRNA vaccine dose...

a. ...by the end of 2024?
b. ...by the end of 2030?

Question and Resolution Details

- For the purposes of this question, partial receipt of a multi-dose vaccine will be counted as a successful mRNA vaccine dose recipient.
- A "non-coronavirus mRNA vaccine" is any mRNA vaccine specifically approved to address a disease or diseases other than a coronavirus.
- This question will be resolved by a panel of experts.

Historical Base Rate Data

- No mRNA vaccine platform had been authorized for use in humans before 2020, and no mRNA vaccine platform targeting a disease other than COVID-19 has yet been authorized for use in humans.
- One can find data related to COVID vaccinations here.

Prior Forecasts

- The Metaculus community is currently predicting a 36% chance that an mRNA vaccine targeting cancer will proceed to Phase III trials by the end of 2022.

Other Potential Background Information

- The Wikipedia article for mRNA vaccines may have helpful background information for some forecasters.
- The University of Cambridge put together this RNA vaccine explainer.
- Pardi et al. (2018) provide a pre-COVID overview of the possibilities of mRNA vaccine development.

Question 14

Will a new surveillance system be announced aimed at detecting the spread of novel infectious pathogens, with a commitment of at least $100 million in funding annually...

a. ...by the end of 2024?
b. ...by the end of 2030?

Question and Resolution Details

- "Novel" refers to any disease not before identified in humans.
- "Infectious" refers to any disease that is spread by infectious agents (such as viral particles, bacteria, or spores).
- If not made sufficiently clear by a public announcement, this question will be resolved via a panel of experts.

Potential Background Information

- Several infectious disease experts have recommended the establishment of pandemic risk surveillance systems. Pardis Sabeti and others have advocated for the SENTINEL proposal, a system aimed at detecting potential new pathogens. More information about the SENTINEL proposal can be found here.
- In 2015, the WHO launched its GLASS surveillance system designed to combat specific pathogens which are thought to entail some risk related to antimicrobial resistance. In the 2021 GLASS report, 3,106,602 infections were reported across 24,803 surveillance sites. A surveillance system aimed at detecting novel pathogens with pandemic risk might be established in a similar fashion to the GLASS system. More information and previous reports from GLASS: https://www.who.int/initiatives/glass.

Question 15

How many times will a non-state actor use biological weapons that involve a contagious agent be the cause death for at least 1,000 people…

a. …by the end of 2024?
b. …by the end of 2030?
c. …by the end of 2050?

Question and Resolution Details

- A biological weapon is any weapon, equipment, or means of delivery designed to use biological agents for hostile purposes.
- "A biological agent…is a bacterium, virus, protozoan, parasite, fungus, chemical, or toxin that can be used purposefully as a weapon in bioterrorism or biological warfare." Wikipedia
- "Contagious" refers to biological agents that would spread person-to-person, as distinct from non-contagious biological agents such as anthrax.
- A non-state actor is any individual, group of individuals, or organization not directly officially affiliated with any state government recognized by the United Nations or by at least one member state of the United Nations.
- A single instance of an actor using a biological weapon is defined as the use of a single biological weapon or the simultaneous use of multiple biological weapons in one metropolitan area. If multiple biological weapons are used by the same actor in separate metropolitan areas simultaneously, this counts as multiple instances of the actor using a biological weapon.
- If not clearly resolvable via credible news or governmental sources reporting such attacks (e.g. WHO, CDC), this question will be resolved by a panel of experts.

Historical Base Rate Data

- There does not seem to be much precedent for attacks killing at least 1,000 people.
    - According to Metaculus, "Internationally, the Aum Shirinkyo cult made multiple attempts at biological attacks, including unsuccessful anthrax attacks. (It was more tragically successful using Sarin gas, which killed 12 in the 1994 subway attacks.)"
    - The 2001 Anthrax attacks killed 5 people and infected 17 others.
    - A 1984 Oregon attack on salad bars caused 45 hospitalizations and no deaths.
- The Global Bioterrorism Database has a list of events perpetrated by non-state actors involving biological weapons.
- Wikipedia's list of bioterrorism incidents may be helpful for some forecasters.

Prior Forecasts

- Metaculus predicted a 12% chance of a significant bioterror attack (more than 100 worldwide deaths) by 2025.
- Metaculus predicts a 5% chance of a human-infecting pathogen being used in a bioterror attack in which more than 500,000 worldwide cases or 100,000 fatalities are reported before 2025.
- Millett and Snyder-Beattie (2017) predicted a 0.00014% annual existential risk associated with bioterrorism.

Other Potential Background Information

- Potential Biological Weapons Threats- CDC
- The Wikipedia articles for Bioterrorism and the Biological Weapons Convention may provide helpful background information for some forecasters.
- The US Department of Health and Human Services has also put together an extensive reading list on bioterrorism.

Question 16

How many times will a state actor use biological weapons that involve a contagious agent be the cause of death for at least 1,000 people…

a. …by the end of 2024?

b. …by the end of 2030?
c. …by the end of 2050?

Question and Resolution Details

- "Contagious" refers to biological agents that would spread person-to-person, as distinct from non-contagious biological agents such as anthrax.
- "A biological agent…is a bacterium, virus, protozoan, parasite, fungus, chemical, or toxin that can be used purposefully as a weapon in bioterrorism or biological warfare." Wikipedia
- A state actor is any individual, group of individuals, or organization directly affiliated with a state government recognized by the United Nations or by at least one member state of the United Nations.
- A biological weapon, per Article I(1) of the Biological Weapons Convention, is defined as "microbial or other biological agents, or toxins [...] of types and in quantities that have no justification for prophylactic, protective or otherwise peaceful purposes."
- The use of a biological weapon, per Article I(2) of the Biological Weapons Convention, is defined as the malicious use of microbial or other biological agents/toxins (see Article I(1)), especially in the context of "armed conflict."
- A single instance of an actor using a biological weapon is defined as the use of a single biological weapon or the simultaneous use of multiple biological weapons in one metropolitan area. If multiple biological weapons are used by the same actor in separate metropolitan areas simultaneously, this counts as multiple instances of the actor using a biological weapon.
- If not clearly resolvable via credible news or governmental sources reporting such attacks (e.g. WHO, CDC), this question will be resolved by a panel of experts.

Historical Base Rate Data

- Arms Control summarizes the current global biological weapons situation.
- Auburn University has a partial timeline of biological weapons attacks (including some non-state incidents).
- There is little historical precedent for biological warfare, especially recently: Biological warfare and bioterrorism: a historical review

Prior forecasts

- There do not seem to be many prior forecasts related to state-originated biological warfare. Most predictions are focused explicitly on bioterror.

Other Potential Background Information

- The Wikipedia articles for Biological warfare and the Biological Weapons Convention may provide helpful background information for some forecasters.
- The United Nations' Biological Weapons Convention website is valuable background reading for those interested in the current context of states and biological weapons.

Question 17

How many times will a non-state actor use biological weapons that involve a contagious agent to kill at least 100,000 people…

a. …by the end of 2024?
b. …by the end of 2030?
c. …by the end of 2050?

Question and Resolution Details

- "Contagious" refers to biological agents that would spread person-to-person, as distinct from non-contagious biological agents such as anthrax.
- "A biological agent…is a bacterium, virus, protozoan, parasite, fungus, chemical, or toxin that can be used purposefully as a weapon in bioterrorism or biological warfare." Wikipedia
- A non-state actor is any individual, group of individuals, or organization not directly officially affiliated with any state government recognized by the United Nations or by at least one member state of the United Nations.
- A biological weapon, per Article I(1) of the Biological Weapons Convention, is defined as "microbial or other biological agents, or toxins [...] of types and in quantities that have no justification for prophylactic, protective or otherwise peaceful purposes."
- If not clearly resolvable via credible news or governmental sources reporting such attacks (e.g. WHO, CDC), this question will be resolved by a panel of experts.

Historical Base Rate Data

- There does not seem to be much precedent for attacks killing at least 100,000 people.
    - According to Metaculus, "Internationally, the Aum Shirinkyo cult made multiple attempts at biological attacks, including unsuccessful anthrax attacks. (It was more tragically successful using Sarin gas, which killed 12 in the 1994 subway attacks.)"
    - The 2001 Anthrax attacks killed 5 people and infected 17 others.

- A 1984 Oregon attack on salad bars caused 45 hospitalizations and no deaths.
- The Global Bioterrorism Database has a list of events perpetrated by non-state actors involving biological weapons.
- Wikipedia's list of bioterrorism incidents may be helpful for some forecasters.

Prior Forecasts

- Metaculus predicted a 12% chance of a significant bioterror attack (more than 100 worldwide deaths) by 2025.
- Metaculus predicts a 5% chance of a human-infecting pathogen being used in a bioterror attack in which more than 500,000 worldwide cases or 100,000 fatalities are reported before 2025.
- Millett and Snyder-Beattie (2017) predicted a 0.00014% annual existential risk associated with bioterrorism.

Other Potential Background Information

- Potential Biological Weapons Threats- CDC
- The Wikipedia articles for Bioterrorism and the Biological Weapons Convention may provide helpful background information for some forecasters.
- The US Department of Health and Human Services has also put together an extensive reading list on bioterrorism.

Question 18

How many times will a state actor use biological weapons that involve a contagious agent to kill at least 100,000 people…

a. …by the end of 2024?
b. …by the end of 2030?
c. …by the end of 2050?

Question and Resolution Details

- "Contagious" refers to biological agents that would spread person-to-person, as distinct from non-contagious biological agents such as anthrax.
- "A biological agent…is a bacterium, virus, protozoan, parasite, fungus, chemical, or toxin that can be used purposefully as a weapon in bioterrorism or biological warfare." Wikipedia

- A state actor is any individual, group of individuals, or organization directly affiliated with a state government recognized by the United Nations or by at least one member state of the United Nations.
- A biological weapon, per Article I(1) of the Biological Weapons Convention, is defined as "microbial or other biological agents, or toxins [...] of types and in quantities that have no justification for prophylactic, protective or otherwise peaceful purposes."
- The use of a biological weapon, per Article I(2) of the Biological Weapons Convention, is defined as the malicious use of microbial or other biological agents/toxins (see Article I(1)), especially in the context of "armed conflict."
- If not clearly resolvable via credible news or governmental sources reporting such attacks (e.g. WHO, CDC), this question will be resolved by a panel of experts.

Historical Base Rate Data

- Arms Control summarizes the current global biological weapons situation.
- Auburn University has a partial timeline of biological weapons attacks (including some non-state incidents).
- There is little historical precedent for biological warfare, especially recently: Biological warfare and bioterrorism: a historical review

Prior Forecasts

- There do not seem to be many prior forecasts related to state-originated biological warfare. Most predictions are focused explicitly on bioterror.

Other Potential Background Information

- The Wikipedia articles for Biological warfare and the Biological Weapons Convention may provide helpful background information for some forecasters.
- The United Nations' Biological Weapons Convention website is valuable background reading for those interested in the current context of states and biological weapons.

Question 19

What will be the expected number of events in which contagious biological agents that have escaped from labs be the cause of death for at least 1,000 people…

a. …by the end of 2024?
b. …by the end of 2030?

c. …by the end of 2050?

Question and Resolution Details

- We use the language "expected number of events" to account for the fact that it may sometimes be unclear whether a pathogen came from a lab escape, such as in the case of Covid. So, for all events in which pathogens kill at least 1,000 people, we will ask a panel of experts to estimate the likelihood that the event was caused by a pathogen escaping from a lab. If, e.g., experts give a 30% chance that a pathogen escaped from a lab, it would count as 0.3 expected events for the purpose of this question.
- To construct the list of events that killed at least 1,000 people, we will rely on a separate panel of experts.
- Note that this refers only to future lab leaks, not to ongoing Covid-19 infections, whether or not Covid-19 was caused by a lab leak (which we take no position on).

Historical Base Rate Data

- Wikipedia maintains a list of confirmed and possible laboratory biosecurity incidents.
    - The only incident that appears to have caused more than 1,000 human fatalities is a *possible* lab leak incident resulting in the 1977-1979 H1N1 pandemic (~700,000 deaths). Some researchers do not believe that this pandemic resulted from a lab leak.
    - Some researchers believe that the current COVID-19 pandemic (>5.4 million deaths) resulted from a lab leak scenario. Most experts have remained skeptical of this hypothesis.
- Notable laboratory escapes
    - 1971: Field test of weaponized Smallpox kills 3.
    - 1972: Smallpox infects lab assistant, kills 2.
    - 1978: Smallpox lab leak kills 1.
    - 1979: Anthrax lab leak kills 66.
    - 1990: Marburg virus outbreak due to laboratory accident kills 1.

Prior Forecasts

- Most of the below forecasts have to do with the risk of existential catastrophe over the next century. Our question focuses on the potential for a much smaller-impact event in a much shorter timespan.
- Toby Ord estimates a **1 in 30 chance** of existential catastrophe from “engineered pandemics” by 2120 (*Precipice*, 167).
    - 2120 is much further out than any of our resolution dates.
- In his critique of *The Precipice*, James Fodor estimates a $2*10^{-6}$ estimate of existential catastrophe associated with engineered pandemics.

    - See Ord's definitions and resolution criteria.
- The 2008 Global Catastrophic Risks Survey found that experts estimated a **30% chance** that the "single biggest engineered pandemic" would kill 1 million people by 2100; they estimated a **10% chance** that the "single biggest engineered pandemic" would kill 1 billion people by 2100; and a **2% chance** that the "single biggest engineered pandemic would lead to humanity's extinction by 2100.
    - The survey does not define the "single biggest engineered pandemic" - it seems to have been left to respondent interpretation. Sandberg and Bostrom circulated this survey before the COVID-19 pandemic.
    - 2100 is much further out than any of our resolution dates.
- Millett and Snyder-Beattie (2017) estimate an **annual existential risk probability of $1.6*10^{-8}$ and $8*10^{-7}$** associated with gain-of-function research accidents.
- Pamlin and Armstrong (2015) estimate a **0.01% risk** of infinite impact associated with synthetic biology by 2115.
    - Infinite impact is a scenario in which "civilisation collapses to a state of great suffering and does not recover, or a situation where all human life ends."
    - 2115 is much further out than any of our resolution dates.
- In an article for 80,000 Hours, Benjamin Todd suggested that there was a **10% chance** that the "single biggest engineered pandemic" would lead to 1 billion deaths and a **2% chance** that the "single biggest engineered pandemic" would lead to the end of human civilization by 2117.
    - 2117 is much further out than any of our resolution dates.

Other Potential Background Information

- The Wikipedia articles for gain-of-function research and biocontainment and Biosafety levels (particularly the level 4 section) may be helpful for some forecasters.
- Human error in high-biocontainment labs: a likely pandemic threat - Bulletin of the Atomic Scientists

Question 20

What fraction of a panel of 100 biosecurity experts will agree that country X [X=separately: U.S., Russia, China, North Korea, Israel, Iran, Syria] has an active biological weapons program at some point from January 1, 2022…

a. …through the end of 2024?
b. …through the end of 2030?
c. …through the end of 2050?

Question and Resolution Details

- A biological weapon is "a bacterium, virus, protozoan, parasite, fungus, chemical, or toxin that can be used purposefully as a weapon in bioterrorism or biological warfare (BW). In addition to these living or replicating pathogens, toxins and biotoxins are also included among the bio-agents." (source) The development, use or stockpiling of biological weapons is banned internationally under the 1972 Biological Weapons Convention.
- A program in Country X means a program run by a state actor directly affiliated with Country X. If Country X changes borders or dissolves, we mean a program run by a state actor within the most expansive definition of the country's physical borders accepted during 2021. Newly-formed countries will be assigned to the country containing the plurality of their landmass according to 2021 borders.
- This question will be resolved by panels of experts convened in 2031 and 2041.
- A biological weapons program is a government-run program that aims to weaponize or stockpile biological agents for hostile use. Dual-use weapons that also have a defensive purpose but could be used for hostile purposes count for the resolution of this question.

Historical Base Rate Data

- Arms Control summarizes the current global biological weapons situation.
- Auburn University has a partial timeline of biological weapons attacks (including some non-state incidents).
- Since the 1972 Biological Weapons Convention, the following states have been suspected or confirmed to have active biological weapons programs at some point:
    - Soviet Union/Russia (Biopreparat, 1974-2000s)
    - Rhodesia (Rhodesian Bush War, 1979)
    - Iraq (1980s)
    - China, suspected (1980s)
    - South Africa (Project Coast, 1980s & 1990s)

Prior Forecasts

- There do not seem to be many prior forecasts related to state-originated biological weapons. Most predictions are focused explicitly on bioterror.

Other Potential Background Information

- The Wikipedia entries on biological warfare and bioterrorism may be helpful for some forecasters.
- The United Nations' Biological Weapons Convention website is valuable background reading for those interested in the current context of states and biological weapons.

Question 21

- **How many countries will have had active biological weapons programs, as estimated by biosecurity experts at some point from January 1, 2022…**
    a. …through the end of 2024?
    b. …through the end of 2030?
    c. …through the end of 2050?

Question and Resolution Details

- A biological weapon is "a bacterium, virus, protozoan, parasite, fungus, chemical, or toxin that can be used purposefully as a weapon in bioterrorism or biological warfare (BW). In addition to these living or replicating pathogens, toxins and biotoxins are also included among the bio-agents." (source) The development, use or stockpiling of biological weapons is banned internationally under the 1972 Biological Weapons Convention.
- For a definition of state actor, we refer to the 193 current UN member states: Member states of the United Nations - Wikipedia.
- Newly-formed countries will be assigned to the country containing the plurality of their landmass according to 2021 borders.
- This question will be resolved by surveying biosecurity experts one year after the resolution period ends.
- A biological weapons program is a government-run program that aims to weaponize or stockpile biological agents for hostile use. Dual-use weapons that also have a defensive purpose but could be used for hostile purposes count for the resolution of this question.

Historical Base Rate Data

- The above-linked source, Arms Control, summarizes the current global biological weapons situation.
- Auburn University has a partial timeline of biological weapons attacks (including some non-state incidents).
- Since the 1972 Biological Weapons Convention, the following states have been suspected or confirmed to have active biological weapons programs at some point:
    - Soviet Union/Russia (Biopreparat, 1974-2000s)
    - Rhodesia (Rhodesian Bush War, 1979)
    - Iraq (1980s)
    - China, suspected (1980s)
    - South Africa (Project Coast, 1980s & 1990s)

Prior Forecasts

- There do not seem to be many prior forecasts related to state-originated biological weapons. Most predictions are focused explicitly on bioterror.

Other Potential Background Information

- The Wikipedia entries on biological warfare and bioterrorism may be helpful for some forecasters.
- The United Nations' Biological Weapons Convention website is valuable background reading for those interested in the current context of states and biological weapons.
- Some possible future challenges to bioweapons control described here.

Question 22

How many times will the WHO declare a new Public Health Emergency of International Concern (PHEIC) for a disease that will be the cause of death of at least 10,000 people...

- …by the end of 2024?
- …by the end of 2030?
- …by the end of 2050?

Question and Resolution Details

- The WHO defines a Public Health Emergency of International Concern (PHEIC) as "an extraordinary event which is determined to constitute a public health risk [...] through the international spread of disease and to potentially require a coordinated international response."
  - Automatically, SARS, smallpox, wild type poliomyelitis, and any new subtype of human influenza are considered as PHEICs. A PHEIC is not confined to infectious diseases, and may cover an emergency caused by exposure to a chemical agent or radioactive material.
- Deaths must occur within the period that the declaration is in force; that is, deaths which occur before the declaration is made or after it is lifted do not count toward the 10,000 death threshold. PHEIC status is reviewed every 3 months after declaration.
- The number of PHEIC declarations will be resolved with WHO press releases from the relevant time periods.

Historical Base Rate Data

- The Wikipedia article for PHEICs has a list of times that the WHO has declared one.

		Reported cases	Confirmed deaths
2009	H1N1 (swine flu) pandemic	491,382	**18,449**
2014	Polio	359	-
2014	Ebola outbreak	28,616	**11,310**
2015-2016	Zika virus epidemic	174,667	18
2018-2020	Kivu Ebola epidemic	3,317	2,280
2019-2022	COVID-19 pandemic	355,591,211	**5,606,929**

- For historical epidemics and pandemics prior to the invention of the PHEIC label in 2009, see Wikipedia's list of epidemics.

Prior Forecasts

- There are no direct estimates of this particular question, but here are a few similar questions.
- In May 2020, the Metaculus community predicted that a non-COVID PHEIC would be declared by June 2022.
- In April 2020, the Metaculus community predicted that the WHO would declare 4-5 non-COVID PHEICs between 2020 and 2030.
- See a Good Judgment Open forecast on whether COVID-19 would remain a PHEIC by May 2020.
- See a Metaculus forecast on when the WHO will lift the COVID-19 PHEIC.

Other Potential Background Information

- The Wikipedia article for PHEICs may be helpful for some forecasters.
- Wilder-Smith and Osman (2020) wrote an overview of the history of PHEICs and how they are declared.

Question 23

What will be the expected number of events in which country leaders are assassinated by a biological weapon involving a contagious agent…

- …by the end of 2024?
- …by the end of 2030?
- …by the end of 2050?

Question and Resolution Details

- We use the language "expected number of events" to account for the fact that there may be uncertainty and debate about whether a particular country leader's death was caused by intentional assassination via biological weapon. For example, there may be uncertainty about whether the leader was the intended target of an attack, whether their death was caused by a biological weapon or another cause, etc. We will ask experts to nominate instances where sitting country leaders died that they believe to have a >1% chance of being caused by a biological attack. We will then ask a separate panel of experts to estimate the likelihood that the event was an intentional assassination via biological weapon. If, e.g., experts give a 30% chance that it was a biological weapon-caused assassination, it would count as 0.3 expected events for the purpose of this question.
- Assassination is defined as a premeditated killing with plausible political, cultural, or religious intent.
- “Country leader” refers to any Head of State and/or Head of Government as maintained by the Protocol and Liaison Service of the United Nations/as reported by the Permanent Missions of each UN Member State. See a list of Heads of State/Government here: Heads of State, Heads of Government and Ministers for Foreign Affairs | Department for General Assembly and Conference Management.
    - In the event that the United Nations does not exist by the resolution date, or in the case that the United Nations does not recognize the sovereignty of an otherwise widely recognized country or the de facto state of affairs within a widely recognized country, the heads of state/government for any relevant country will be determined by a panel of experts.
- "Contagious" refers to biological agents that would spread person-to-person, as distinct from non-contagious biological agents such as anthrax.
- "A biological agent…is a bacterium, virus, protozoan, parasite, fungus, chemical, or toxin that can be used purposefully as a weapon in bioterrorism or biological warfare." Wikipedia

- The use of a biological weapon, per Article I(2) of the Biological Weapons Convention, is defined as the malicious use of microbial or other biological agents/toxins (see Article I(1)), especially in the context of "armed conflict."

Historical Base Rate Data

- A biological weapon has not been used to assassinate a country leader within the modern context. Some forecasters may nevertheless find this article on the history of bioterrorism interesting.

Prior Forecasts

- The Metaculus community is currently predicting a 33% chance that an American president will *not* be assassinated at some point during the 21st century.

Question 24

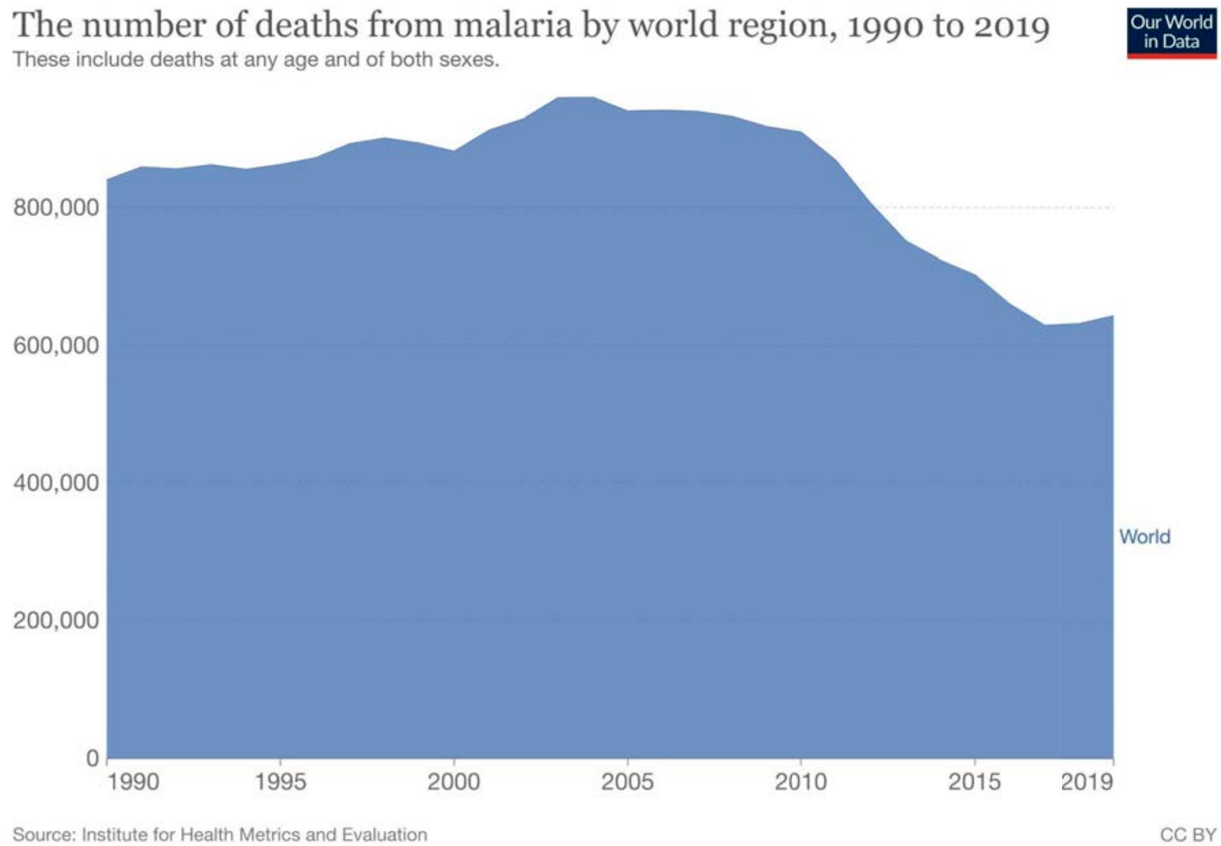
1. **What will be the number of human deaths due to malaria…**
   a. ...during the year 2024?
   b. ...during the year 2030?
   c. …during the year 2050?

Question and Resolution Details

- This question will be resolved using results from the Institute for Health Metrics and Evaluation's annual Global Burden of Disease reports. For a visualization of past results, see base rate data below and here.
- This question will be resolved using the central estimate for the total number of global malaria deaths at all ages in 2024, 2030, and 2050 respectively. If the Institute for Health Metrics and Evaluation is no longer publishing estimates of the total number of malaria deaths in the relevant year, the question will be resolved by a panel of experts.

Historical Base Rate Data

- Forecasters can find data on malaria deaths from 1990 to 2019 on *Our World in Data*.



-
- The WHO's Global Malaria Programme has published reports and data on malaria for several years, and its website may have helpful information.

Prior Forecasts

- The Metaculus community predicted a 47% chance that the global malaria mortality rate would drop by 90% from 2015 to 2030.
- The WHO's World Malaria Report 2021 makes several malaria-related projections conditional on meeting its current malaria response goals.

Other Potential Background Information

- The Wikipedia page on malaria may be helpful for some forecasters.
- The WHO's factsheet on malaria is updated somewhat frequently and may be interesting to some forecasters.
- The CDC also has information related to malaria.
- See Sustainable Development Goal 3, Target 3.3 on the fight against communicable diseases, including malaria. (Official UN Page)

Question 25

What will be the global surface temperature change as compared to 1850-1900, in degrees Celsius…

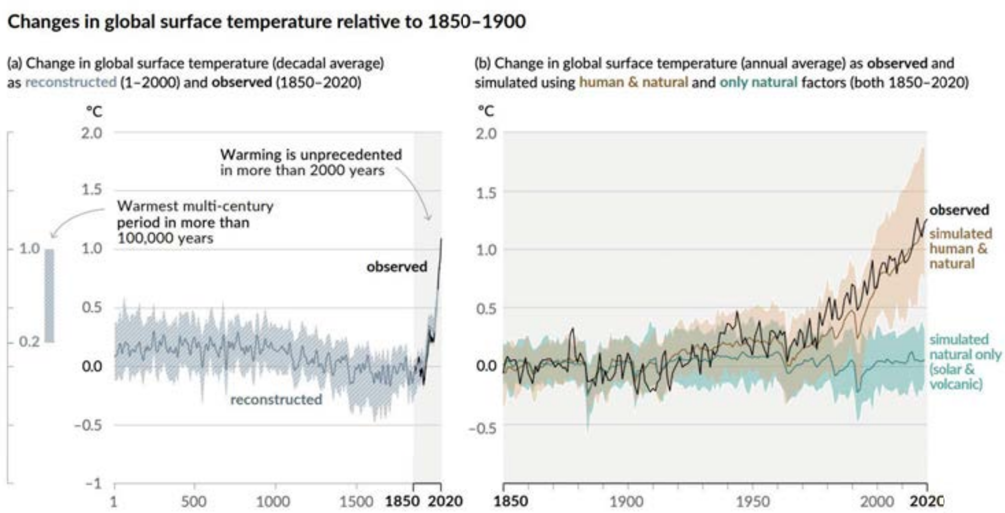
1. ...in 2030?
2. ...in 2050?
3. …in 2100?

Question and Resolution Details

- The global surface temperature change refers to the change in both global mean surface temperature (GMST) and global surface air temperature (GSAT) as compared to a representative period between 1850 and 1900. This metric is used by the Intergovernmental Panel on Climate Change. See more on their methodology and model in Cross-Section Box TS.1 (Page TS-27) of the most recent IPCC Assessment Report's Technical Summary.
- We will use the most recent IPCC report to determine the average global surface temperature. If at some point IPCC reports become unavailable or the IPCC's methodology or metric for measuring global surface temperature change is altered in future reports, a panel of experts will determine an equivalent authoritative source for determining average global surface temperature.

Historical Base Rate Data

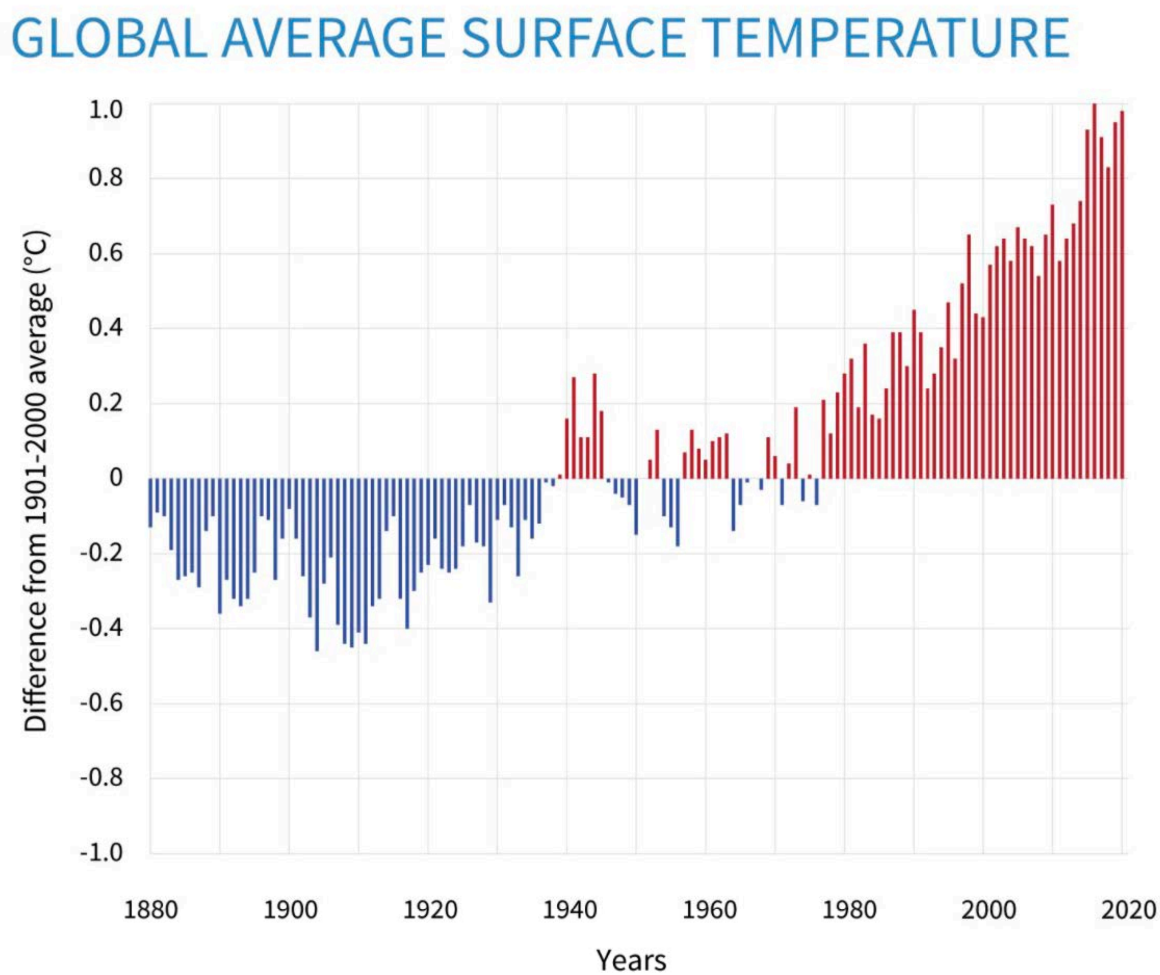
- See the IPCC AR6 Physical Science Basis Report and the following image from it:



  -
  - (P. 6 of the Summary for Policymakers)
    - Gillett, N.P.; Malinina, E.; Kaufman, D.; Neukom, R. (2021): Summary for Policymakers of the Working Group I Contribution to the IPCC Sixth

Assessment Report - data for Figure SPM.1 (v20210809). NERC EDS Centre for Environmental Data Analysis, 09 August 2021. doi:10.5285/76cad0b4f6f141ada1c44a4ce9e7d4bd. http://dx.doi.org/10.5285/76cad0b4f6f141ada1c44a4ce9e7d4bd

- IPCC, 2021: Summary for Policymakers. In: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu and B. Zhou (eds.)]. Cambridge University Press. In Press.



- NOAA Climate.gov graph, based on data from the National Centers for Environmental Information. Based on difference from 1901-2000 average

- "Averaged across land and ocean, the 2020 surface temperature was 1.76° F (0.98° Celsius) warmer than the twentieth-century average of 57.0°F (13.9°C) and 2.14˚F (1.19˚C) warmer than the pre-industrial period (1880-1900)." Climate.gov report

Prior Forecasts

- From IPCC:
    - "This report defines 'warming', unless otherwise qualified, as an increase in multi-decade global mean surface temperature (GMST) above pre-industrial levels. Specifically, warming at a given point in time is defined as the global average of combined land surface air and sea surface temperatures for a 30-year period centred on that time, expressed relative to the reference period 1850–1900 (adopted for consistency with Box SPM.1 Figure 1 of IPCC (2014a) 'as an approximation of pre-industrial levels', excluding the impact of natural climate fluctuations within that 30-year period and assuming any secular trend continues throughout that period, extrapolating into the future if necessary. There are multiple ways of accounting for natural fluctuations and trends (e.g., Foster and Rahmstorf, 2011; Haustein et al., 2017; Medhaug et al., 2017; Folland et al., 2018; Visser et al., 2018), but all give similar results. A major volcanic eruption might temporarily reduce observed global temperatures, but would not reduce warming as defined here (Bethke et al., 2017). Likewise, given that the level of warming is currently increasing at 0.3°C–0.7°C per 30 years (likely range quoted in Kirtman et al., 2013 and supported by Folland et al., 2018), the level of warming in 2017 was 0.15°C–0.35°C higher than average warming over the 30-year period 1988–2017….We adopt a working definition of warming over the historical period based on an average of the four available global datasets that are supported by peer-reviewed publications: the three datasets used in the AR5, updated (Karl et al., 2015), together with the Cowtan-Way infilled dataset (Cowtan and Way, 2014). A further two datasets, Berkeley Earth (Rohde et al., 2013) and that of the Japan Meteorological Agency (JMA), are provided in Table 1.1. This working definition provides an updated estimate of 0.86°C for the warming over the period 1880–2012 based on a linear trend. This quantity was quoted as 0.85°C in the AR5. Hence the inclusion of the Cowtan-Way dataset does not introduce any inconsistency with the AR5, whereas redefining GMST to represent global SAT could increase this figure by up to 20% (Table 1.1, blue lines in Figure 1.2 and Richardson et al., 2016)."

-

Question 26

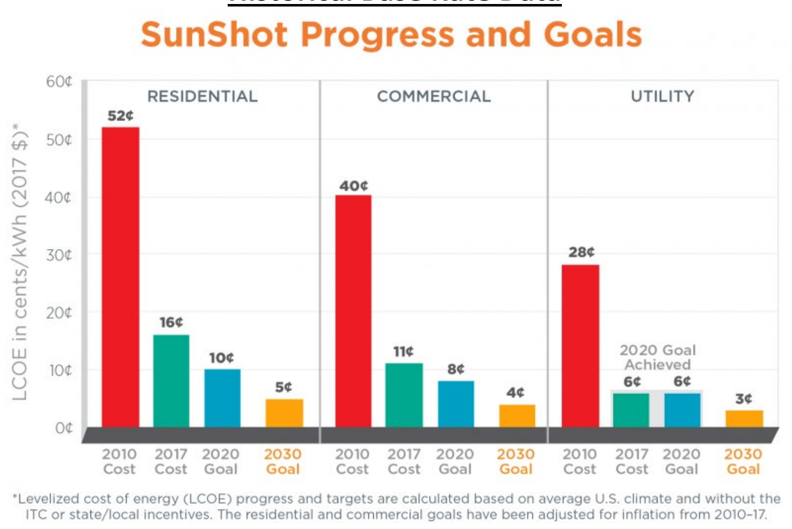
What will be the estimated cost (in 2017 USD / kWh) for new utility-scale photovoltaic solar systems above 4MW$_{AC}$ in the United States...

- ...for the year 2024?
- ...for the year 2030?

Question and Resolution Details

- Wiki-Solar, a large online database of photovoltaic power stations around the world, defines 'utility-scale solar' as all systems above 4MW$_{AC}$; more from Wikipedia here. The "above 4MW$_{AC}$" threshold noted in the text of the question was chosen to reflect this definition of utility-scale.
- Resolution of this question will be based on SunShot reports (see below). If these reports are not available at the appropriate time, a panel of experts will determine an equivalent authoritative source.
- See U.S. Department of Energy SunShot 2030 report and this page for further definitions.

Historical Base Rate Data



- See this page for data on the cost of installation for 100MW photovoltaic systems since 2010.
- See this page for data on the cost of utility-scale generation of photovoltaic solar energy ($/MWh) since 2009.
- "In 2017, the solar industry achieved SunShot's original 2020 cost target of $0.06 per kilowatt-hour for utility-scale photovoltaic (PV) solar power three years ahead of schedule, dropping from about $0.28 to $0.06 per kilowatt-hour (kWh)." SunShot 2030 report

Prior Forecasts

- According to the U.S. Solar PV System Price report in 2020, 100 megawatt utility-scale systems have the potential to fall to $0.76/W by 2025. Source.
- See this page for Annual Technology Baseline data for utility-scale photovoltaic solar, including predictions on the LCOE based on different scenarios of technological innovation.

Question 27

By what year will fusion reactors deliver 1% of all utility-scale power consumed in the U.S.?

Question and Resolution Details

- Per Wikipedia, nuclear fusion is a reaction in which “two or more atomic nuclei are combined to form one or more different atomic nuclei and subatomic particles.” During the fusion process, energy is released. Nuclear fusion differs from nuclear *fission*, a reaction in which energy is released through the splitting of very heavy atoms. Today, due to technological constraints, fission-based nuclear reactors vastly outnumber fusion-based nuclear reactors and all existing fusion reactors require more energy to operate than they produce. However, fusion reactors are thought to have many operational, safety and efficiency advantages over fission reactors. See more here: https://en.wikipedia.org/wiki/Fusion_power
- This question will be resolved using published estimates on utility-scale power from the U.S. Energy Information Administration (EIA), such as those shown here. (Currently the estimate of energy share contributed by nuclear power is not broken down by type, but it’s reasonable to expect it will be if fusion reactors become net contributors.) By “utility-scale” power generation, we refer to power generated from facilities with at least one megawatt of total electricity generating capacity (per the EIA definition).
- If relevant figures from the EIA are not available, this question will be resolved by a panel of experts.
- A forecast of the year 100,000+ or later will be resolved equivalently to a forecast of “never” for this question.
- In the event that this never happens, predict number 1E210, equivalent to the largest number able to be submitted on the platform used to host the tournament.

Prior Forecasts

- When will the first facility generating a net 100 MW of electricity, using only fusion, come online? (Metaculus)

- When will a private fusion company first report a yearly profit? (Metaculus)
- When will a fusion reactor reach ignition? (Metaculus)
- Will the International Thermonuclear Experimental Reactor (ITER) start operating by the end of 2025? (Metaculus)
- Will there be 36 or more private fusion-energy companies in 2022? (Metaculus)
- EUROfusion project predicts that its demonstration plant DEMO won't start operations until after 2054.
- SPARC project scheduled to begin operations in 2025.

Other Potential Background Information

Selected nuclear power technology milestones
(Fusion developments in blue, fission developments in black)

1920s - 1930s Fundamental science

- 1920-1939 - Fundamental science groundwork for fusion (stellar fusion, P-P chain reaction, quantum tunneling, etc)
- 1932-1938 - Fundamental science groundwork for fission (first atom splitting, induced radioactivity, discovery of nuclear fission, etc)
- 1938 - First plans for pinch device
- 1939 - Experimental confirmation of self-sustaining nuclear chain reaction; scientists begin petitioning governments to support nuclear weapons development

1940s Reactor prototype development / Early plasma confinement research

- 1942 - First nuclear reactor achieves criticality, as part of Manhattan Project (Chicago Pile-1)
- 1945 - Patent filed for 'light water reactor' (LWR)
- 1946-1947 - Z-pinch patent filed (plasma confinement system); first Z-pinch experiments
- 1949 - Funded by Argentinian government, Huemul Project is an early effort to develop fusion power; project leader is later convicted of fraud

1950s Fission reactors begin supplying power / 'Pinch' disappoints

- 1951 -
    - Electricity generated for first time by nuclear reactor (EBR-I, 100 kW)
    - Project to work on Stellarator (plasma confinement system) created at Princeton
    - Tokamak approach (plasma confinement) first discussed in Soviet Union
- 1954 -
    - First nuclear submarine, USS Nautilus, uses type of LWR; leads to this type of reactor becoming most popular choice for power generation long into future
    - Atomic Energy Act of 1954 allows rapid declassification of US reactor tech

  - Obninsk Nuclear Power Plant becomes first to generate electricity for power grid (5 MW)
  - US Army Nuclear Power Program begins
  - Edward Teller and others point out that most 'pinch' approaches are inherently flawed
- 1956 - World's first commercial nuclear station connected to grid, in England
- 1957 -
  - US Army Nuclear Power Program first supplies electricity in industrial capacity to commercial grid (SM-1, 2MW)
  - First commercial nuclear station in US, Shippingport Reactor (60 MW)
- 1958 -
  - ZETA (pinch machine) project claims to achieve fusion, later forced to retract
  - Scylla I (pinch machine) achieves fusion, but is later abandoned due to infeasibility of scaling up

1960s Fission power scales to utility size / Fusion moves on to new designs

- 1960 -
  - **Globally, <1 GW installed nuclear capacity**
  - Inertial confinement fusion (ICF) proposed
- 1965 -
  - 'Generation II' reactors come into use, remain most common plants in use to present day
  - Laser fusion/ICF experiments begin
- 1967 - Magnetic mirror (plasma confinement) proposed
- 1968 - Soviets announce successful results with tokamak tests, results confirmed by UK team; internationally, many other designs abandoned in favor of tokamaks

1970s New economic & regulatory pressures drive fission power investment / Laser developments

- 1972 - Fusion chain reaction ('ignition') outlined in paper, resulting in major development efforts
- 1973 - Oil crisis causes France to invest heavily in nuclear plants
- 1977 - First megalaser, Shiva, completed
- **Late 1970s -**
  - **Globally, 100 GW installed nuclear capacity**
  - US funds magnetic mirror program into early 1980s, but projects close due to expense
- 1970s and 1980s - In US, regulatory changes & pressure group litigation make reactor construction more expensive, while fossil fuel prices fall, resulting in long-term slow-

down

1980s

- 1980s - Several tokamak machines achieve first plasma
- 1984 -
  - NOVA laser finished
  - Spherical tokamak proposed
- 1985 - 'Chirping', a useful method for amplifying laser pulses, developed
- 1986 - International Thermonuclear Experimental Reactor (ITER) coalition forms
- 1987 - Levitated dipole proposed
- **Late 1980s - Globally, 300 GW installed nuclear capacity**

1990s

- 1990s - ICF designs advance; National Ignition Facility (NIF) funded by US government
- 1991 -
  - First controlled release of fusion power (JET's Preliminary Tritium Experiment)
  - First spherical tokamak machine begins operation
- 1996 - 'Generation III' reactors come into use, but few built due to expense
- 1997 - JET reaches record output level of 10 MW for 0.5 sec, unbeaten since

2000s onwards

- 2004 - Levitated Dipole Experiment achieves first plasma
- **2005 - Globally, 366 GW installed nuclear capacity**
- 2008 - NIF becomes operational
- 2010s - Increased public and private investment in fusion research
- 2013 - NIF achieves net energy gain

Question 28

What percentage of the world's electricity will be provided by solar energy and wind energy combined…

a. ...in 2024?
b. ...in 2030?

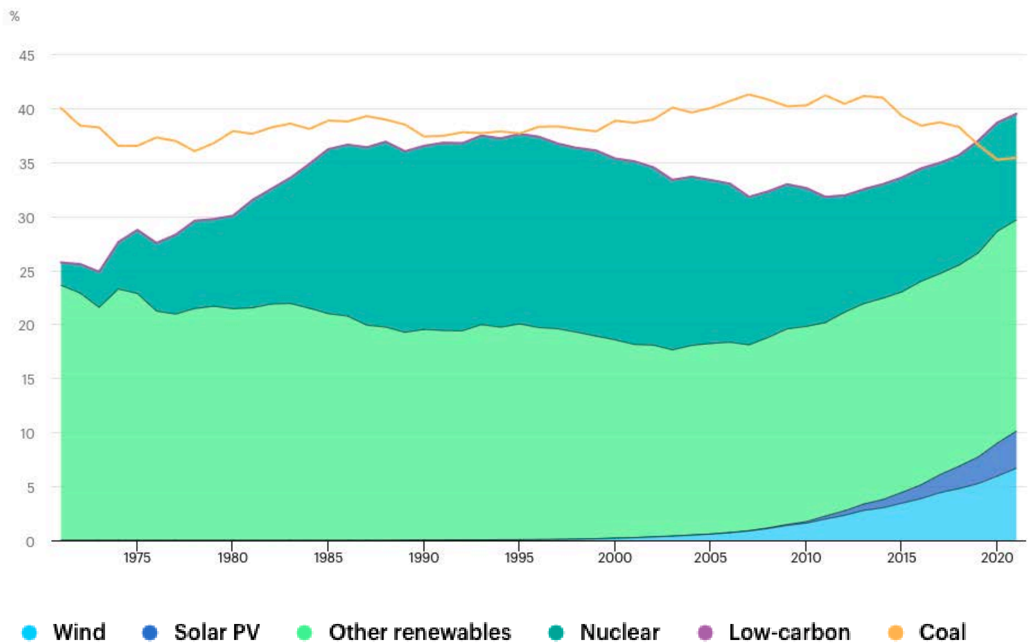
Question and Resolution Details

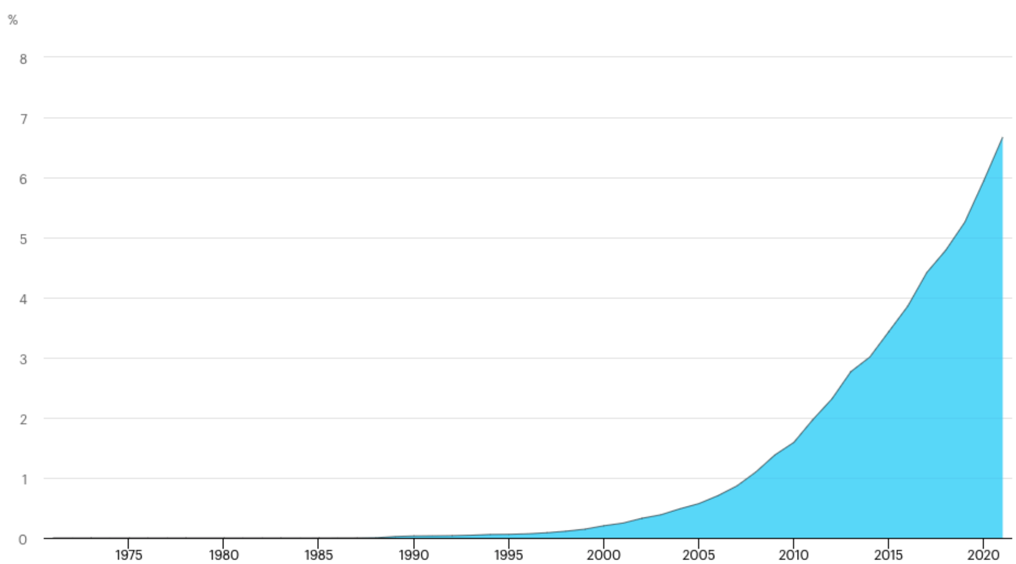
- This question will be resolved via the International Energy Agency's yearly Global Energy Review where it regularly reports on the share of renewables in global electricity production. See the portion of the 2021 report on renewables here: https://www.iea.org/reports/global-energy-review-2021/renewables

- If the IEA does not publish a similar report in 2025 detailing the share of renewables in global electricity production in 2024 (and equivalent for 2031 and 2030), the question will be resolved by a panel of experts.

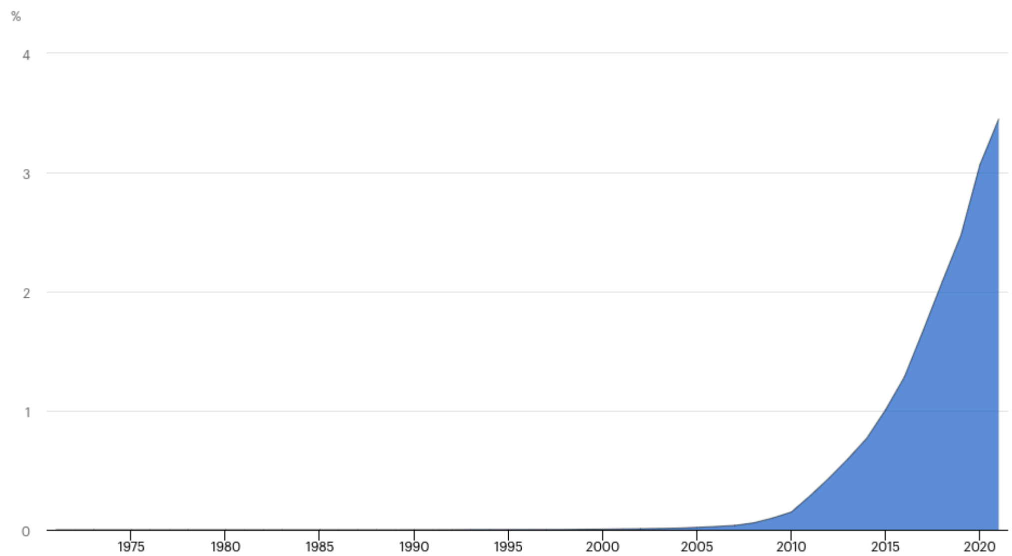
<u>Historical Base Rate Data</u>

Share of low-carbon sources and coal in world electricity generation, 1971-2021





Wind energy



Solar PV
IEA, *Share of low-carbon sources and coal in world electricity generation, 1971-2021*, IEA, Paris
https://www.iea.org/data-and-statistics/charts/share-of-low-carbon-sources-and-coal-in-world-electricity-generation-1971-2021

Year	Solar PV (%)	Wind (%)	Combined (%)
2021	3.4	6.7	10.1
2020	3.1	5.9	9.0
2019	2.5	5.3	7.8
2018	2.1	4.8	6.9
2017	1.7	4.4	6.1
2016	1.3	3.9	5.2
2015	1.0	3.4	4.4
2014	0.8	3.0	3.8
2013	0.6	2.8	3.4
2012	0.4	2.3	2.7

2011	0.3	2.0	2.3

Prior Forecasts

- Will more than 80% of the new US electricity Generation Capacity in 2016 come from solar and wind? (Metaculus)
- 50% Carbon-neutral electricity by 2025? (Metaculus)
- Growth rate total renewables 2020 to 2022 (Metaculus)

Solar

- How much solar energy will be consumed globally, in terawatt-hours, in the calendar year 2023? (Metaculus)
- What percentage of primary energy in the United States will come from solar in 2031? (Metaculus)
- What percentage of US electricity will be produced by solar power in 2030? (Metaculus)
- How much new solar PV capacity will be installed globally in 2030? (Metaculus)

Wind

- How much wind energy will be consumed globally, in terawatt-hours, in the calendar year 2023? (Metaculus)
- Will an airborne wind energy system of at least 100kW be sold before 2022? (Metaculus)

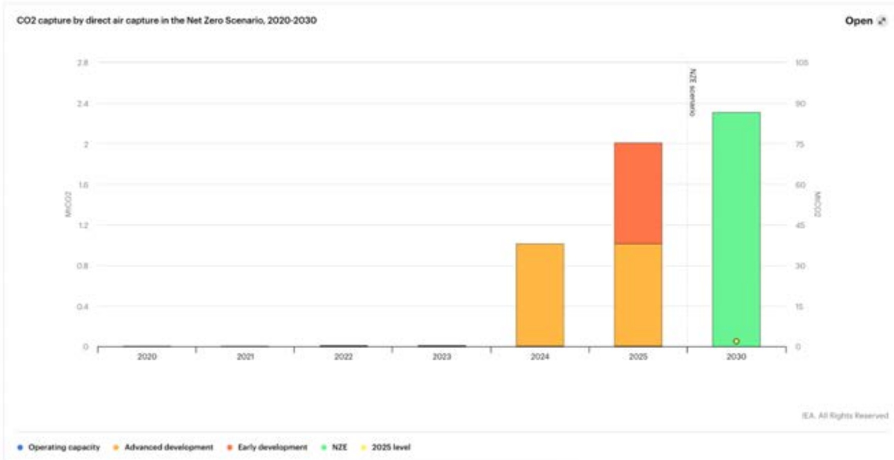
Question 29

What will be the annual amount of CO2 captured and stored by direct air capture (in Mt CO2/year)…

a. …in 2024?
b. …in 2030?

Question and Resolution Details

- According to IEA: "There are currently 19 direct air capture (DAC) plants operating worldwide, capturing more than 0.01 Mt CO2/year, and a 1Mt CO2/year capture plant is in advanced development in the United States. The latest plant to come online, in September 2021, is capturing 4 kt CO2/year for storage in basalt formations in Iceland. In the Net Zero Emissions by 2050 Scenario, DAC is scaled up to capture more than 85 Mt CO2/year by 2030 and ~980 Mt CO2/year by 2050. This level of deployment will require several more large-scale demonstrations to refine the technology and reduce capture costs."

- 

- This question will be resolved based on IEA data in the appropriate year. If IEA data is not available, it will be resolved by a panel of experts.

Historical Base Rate Data/Prior Forecasts

- Since DAC is a new technology, few prior forecasts or historical data exist except those referenced in the IEA report.

Potential Background Information

- The IEA report linked above is relatively comprehensive as background material.
- The Wikipedia page for Direct Air Capture may also be helpful: https://en.wikipedia.org/wiki/Direct_air_capture.
- While not directly related to the forecasting question, background information about how direct air capture fits into broader climate infrastructure schemes might interest some:
    - [“Direct air capture” of carbon dioxide won't solve climate change - Vox]()
    - [How the largest direct air capture plant will suck CO2 out of the atmosphere - The Verge]()
-

Question 30

How much will it cost to produce hydrogen from renewable electricity (in $ per kg of hydrogen)…

a. …in 2024?

b. …in 2030?

Question and Resolution Details

- According to the IEA Global Hydrogen Review 2021: "Using renewable electricity to produce hydrogen costs USD 3 to USD 8 per kg. There is significant scope for cutting production costs through technology innovation and increased deployment. The potential is reflected in the IEA's Net Zero Emissions by 2050 Scenario (NZE Scenario) in which hydrogen from renewables falls to as low as USD 1.3 per kg by 2030 in regions with excellent renewable resources (range USD 1.3-3.5 per kg), comparable with the cost of hydrogen from natural gas with CCUS. In the longer term, hydrogen costs from renewable electricity fall as low as USD 1 per kg (range USD 1.0-3.0 per kg) in the NZE Scenario, making hydrogen from solar PV cost-competitive with hydrogen from natural gas even without CCUS in several regions."
- This question will be resolved based on the average of the range of costs reported in the Executive Summary section of IEA's annual Global Hydrogen Review reports. If these reports are not available in the relevant years or the key figures are not reported in the Executive Summary, it will be resolved by a panel of experts.

Historical Base Rate Data and Prior Forecasts

- See the IEA report above.
- Some forecasts and historical data are available here as well.
- Some overview of the history of the market and future projections: https://www.spglobal.com/marketintelligence/en/news-insights/latest-news-headlines/experts-explain-why-green-hydrogen-costs-have-fallen-and-will-keep-falling-63037203
- This IRENA report is comprehensive as well: https://irena.org/publications/2020/Dec/Green-hydrogen-cost-reduction.

Other Potential Background Informaton

- Some background information here as well: https://en.wikipedia.org/wiki/Green_hydrogen.Why We Need Green Hydrogen | Columbia University
- For further explanation about the process of hydrogen production from renewable energy sources, see this from the U.S. Department of Energy.

Question 31

What is the probability that the use of a nuclear weapon (in a single event) will cause the death of more than 1,000 people…

a. …by the end of 2024?
b. ...by the end of 2030?

Question and Resolution Details

- Deaths include indirect deaths due to radiation poisoning, but not statistical or excess deaths, where the direct cause of the death is unclear, but there seems to be an increased average probability of death for a specific group.
- By "use" of a nuclear weapon, we refer specifically to the intentional detonation of one or more nuclear warheads. The nuclear weapons testing, accidental detonations, or the use of radiological devices (e.g. dirty bombs) do not count for resolving this question.
  - Other kinds of nuclear explosions or nuclear accidents unrelated to nuclear warheads, such as reactor meltdowns or accidents from nuclear physics experiments, would not resolve this question positively.
  - The testing of nuclear weapons, accidental detonations, or the use of radiological devices do not count for resolving this question (https://www.cfr.org/backgrounder/dirty-bombs).
- If resolution is ambiguous, this question will be resolved by a panel of experts.

Historical Base Rate Data

Deaths from nuclear attack during wartime

WWII--

- 6 August 1945, USA targeting Hiroshima, Japan: 70,000--80,000 killed by blast (see here)
- 9 August 1945, USA targeting Nagasaki, Japan: 22,000--75,000 immediate deaths (see here)

Nuclear close calls

As cataloged here:

- 16 between 1956 and 1995, or 1 every 2.4 years

Recent major conflict history in states currently possessing nuclear weapons

	Belligerent parties	*Conflict*	*Deaths of belligerents (1)*	*Deaths of belligerents (2)*	*Deaths of civilians*
1964 - present	**India** (1) vs separatist factions	Insurgency in Northeast	2,762 (since 1992)	8,554 (since 1992)	10,302 (since 1992)

	(2)	India *			
1948 - present	**Israel** / **UK** / France / others (1) vs Arab league (2)	Arab-Israeli conflict *	22,570 military and 1,723 civilian	91,105 (military and civilian)	--
1947 - present	**India** (1) vs **Pakistan** (2)	Kashmir conflict *	80,000 - 110,000		
2001 - 2021	**USA** / **UK** / others (1) vs Taliban / al-Qaeda / others (2)	War in Afghanistan *	73,295	56,293	46,319
2004 - 2017	**Pakistan** / **USA** / **UK** (1) vs terrorist groups (2)	War in North-West Pakistan *	4,631 - 8,214	29,398 - 31,000	9,394 - 22,100
2003 - 2011	**USA** / **UK** / Australia / Peshmerga (1) vs Iraq (2)	Iraq war *	25,071	31,608 - 37,344	183,535 - 206,107 *
1990 - 1991	**USA** / **UK** / Kuwait/others (1) vs Iraq (2)	Gulf War *	4,492	20,000 - 50,000	4664
1979 - 1989	**Soviet Union** / Afghanistan (1) vs Mujahideen (2)	Soviet-Afghan War *	32,453	61,775	562,000 - 2,000,000
1955 - 1975	North Vietnam / **China** / **Soviet Union** (1) vs South Vietnam / **USA** / others (2)	Vietnam War *	667,130 - 951,895	333,620 - 392,364	405,000 - 2,000,000
1971	Bangladesh / India (1) vs Pakistan (2)	Bangladesh Liberation War *	30,000	8,000	300,000 - 3,000,000
1954 - 1962	Algerian rebels (1) vs France (2)	Algerian War *	140,000 - 152,863	30,600	250,000 - 300,000
1946 - 1954	DR Vietnam / others (1) vs France (2)	First Indochina War *	175,000 - 300,000	134,500	125,000 - 400,000

1950 - 1953	South Korea / **USA** (1) vs North Korea / China / Soviet Union (2)	Korean War *	170,927	398,000 - 926,000	2,000,000 - 3,000,000
1948	India (1) vs Hyderabad (2)	Annexation of Hyderabad *	10	2,180	200,000

States which possessed nuclear weapons for the duration of a conflict are bolded; states which only came to possess nuclear weapons after a conflict are underlined; and states which acquired nuclear weapons during a conflict are bolded and underlined. Source for major conflicts here.

Prior Forecasts

Future conflicts involving nuclear powers

US-China war by 2035 (Metaculus)

US-Russia war by 2050 (Metaculus)

Metaculus nuclear risk tournament

Metaculus nuclear risk horizons project

- In a survey given at the Global Catastrophic Risks Conference in Oxford in 2008, the median participant predicted a 30% chance that at least a million people would be dead from nuclear war by 2100, a 10% chance that at least a billion people people would be dead from nuclear war by 2100, and a 1% chance that nuclear war would lead to human extinction by 2100. Likewise, participants predicted a 15% chance that at least a million people would be dead from nuclear terrorism by 2100, a 1% chance that at least a billion would be dead from nuclear terrorism by 2100, and a 0.03% chance that nuclear terrorism would lead to human extinction by 2100.
- In the Lugar Survey on Proliferation Threats and Responses in 2005, the median expert gave a 20% probability of a nuclear attack within 10 years (i.e. by 2015), for an annualized probability of 2.21% per year.
- According to Luisa Rodriguez, a set of superforecasters in Good Judgment Inc. in 2018 predicted a 1% mean probability that a nuclear attack would cause at least one fatality before 1 January 2021, resulting in an annualized probability of 0.40% per year.
- See here for various forecasts related to the probability of nuclear war between the U.S. and Russia specifically.

Other Potential Background Information

Nuclear stockpiles to the present, by country (FAS)

Question 32

How many total nuclear warheads will be in military inventories globally by…

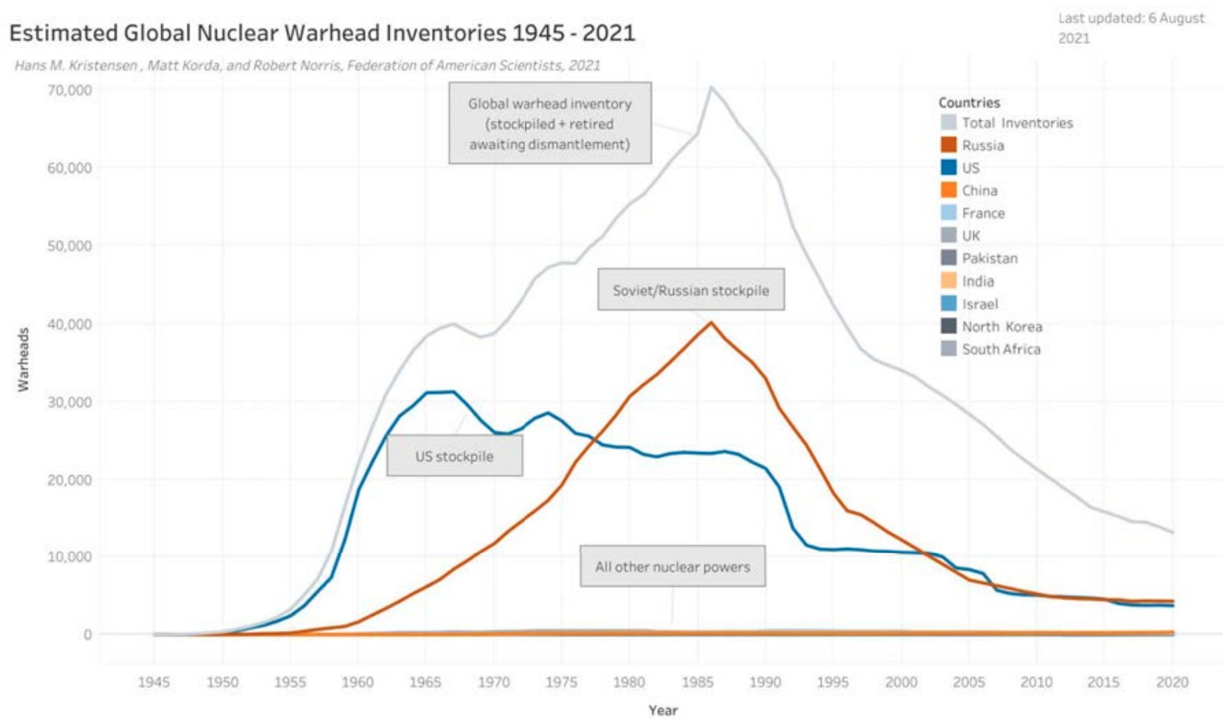
- …the end of 2024?
- …the end of 2030?
- …the end of 2040?

Question and Resolution Details

- Total inventory includes both deployed and non-deployed/reserve warheads, as well as retired warheads awaiting dismantlement.
- The Bulletin of Atomic Scientists and the Federation of American Scientists publishes a bimonthly "Nuclear Notebook" which tracks nuclear weapons programs in nuclear-armed countries. The Nuclear Notebook has published estimates of nuclear forces owned by the US, Russia, and China, among others, every year since 2000.
- This question will be resolved by aggregating, across nuclear-armed countries, the estimated total inventory of nuclear warheads reported in, respectively, the January 2031, January 2041 and January 2051 editions of the Nuclear Notebook's nuclear forces reports. In the event that the Nuclear Notebook ceases publication, the question will be resolved by a panel of experts.

Historical base rate data

- As of mid-2021, 9 countries possessed roughly 13,150 warheads. (source)



Other information:

- FAS on trends in world nuclear forces:
    - Globally, the overall inventory of nuclear weapons is declining, but the pace of reduction is slowing compared with the past 30 years. Moreover, all of that reduction is happening only because the United States and Russia are still dismantling previously retired warheads. In terms military stockpiles (those warheads assigned to operational forces), however, the overall number is increasing again. The United States is probably still reducing its stockpile but appears to be leveling out. France and Israel have relatively stable inventories. But China, India, North Korea, Pakistan, Russia and United Kingdom are all thought to be increasing their stockpiles.
    - The exact number of nuclear weapons in each country's possession is a closely held national secret. Most nuclear-armed states provide essentially no information about the sizes of their nuclear stockpiles. … Despite such limitations, however, publicly available information, careful analysis of historical records, and occasional leaks make it possible to make best estimates about the size and composition of the national nuclear weapon stockpiles.
    - Public details are scarce, but we estimate that Russia is dismantling 200-300 retired warheads per year. The future of the Russian stockpile size is debated: US Strategic Command and part of the Intelligence Community claim "Russia's

overall nuclear stockpile is likely to grow significantly over the next decade – growth driven primarily by a projected increase in Russia's non-strategic nuclear weapons." Others privately disagree. A major uncertainty is how many tactical weapons will be replaced by new nuclear versions versus conventional weapons.
- The Chinese stockpile is increasing and US Strategic Command and part of the Intelligence Community claim that China will "likely double the size of their nuclear stockpile by the end of the decade."
- In historical context, the number of nuclear weapons in the world has declined significantly since the Cold War: down from a peak of approximately 70,300 in 1986 to an estimated 13,150 in mid-2021. Government officials often portray that accomplishment as a result of current or recent arms control agreements, but the overwhelming portion of the reduction happened in the 1990s. Some also compare today's numbers with that of the 1950s, but that is like comparing apples and oranges; today's forces are vastly more capable.

Question 33

How many countries will be estimated to have at least one nuclear warhead…

a. ...by the end of 2024?
b. ...by the end of 2030?
c. ...by the end of 2050?

Question and Resolution Details

- The Federation of American Scientists currently estimates that 9 countries have at least one nuclear warhead: https://fas.org/issues/nuclear-weapons/status-world-nuclear-forces/.
  - Note: We count North Korea as having at least one warhead for the purposes of this question. Due to high levels of secrecy in some states regarding nuclear weapon possession, FAS estimates for such states incorporate information about production of fissile materials, number of delivery platforms, and US intelligence community estimates.
- We will use the updated version of this dataset or an equivalent dataset that is agreed to be credible by a panel of experts.
  - In the event of human extinction, assume a hypothetical panel of current-day human experts resolved the question based on information available in the year of interest.
- To quantify the number of warheads belonging to a state, we refer to the FAS definition of "total inventory," which includes warheads in the military stockpile (including

deployed strategic, deployed nonstrategic, and reserve/non-deployed warheads), as well as retired warheads that are awaiting dismantlement.

Historical Base Rate Data

Year of first (and last, where applicable) possession of nuclear warheads (source):

- 1945 - USA (recognized Nuclear Weapons State)
- 1949 - Russia (USSR) (recognized Nuclear Weapons State)
- 1953 - UK (recognized Nuclear Weapons State)
- 1964 - China; France (both recognized Nuclear Weapons States)
- 1967 - Israel
- 1982 - South Africa
- 1990 - South Africa no longer possesses
- 1998 - Pakistan; India
- 2015 - North Korea

Nuclear weapons development by country (sources: I II III IV V VI VII VIII IX X)

- 1939 - USA begins development program
- 1940 - UK begins development program
- 1942 - USSR begins development program
- 1945 - USA's first successful test
- 1948 - Estimated date of Israel development program
- 1949 - USSR's first successful test
- 1952 - UK's first successful test
- 1954 - China begins development program
- 1956 - France begins development program
- 1960 - France's first successful test
- 1963 - Possible date of Israel's first successful test
- 1964 - China's first successful tests
- 1967 - India begins development program
- 1972 - Pakistan begins development program
- 1974 - India's first successful test
- 1979 - Possible date of Israel's first successful test (Vela Incident)
- 1980s - North Korea begins development program
- 1998 - Pakistan's first successful test
- 2006 - North Korea's first successful test

Our World In Data has collected data on nuclear stockpiles by country by year here.

Information on unsanctioned nuclear activity by country

Prior Forecasts

- 'Forecasting nuclear proliferation' on Wikiversity
- 'Predicting Proliferation: High Reliability Forecasting Models of Nuclear Proliferation as a Policy and Analytical Aid' 2016
- 'Nuclear Fusion Power for Weapons Purposes: An exercise in nuclear proliferation forecasting' 2013

Other Potential Background Information

- Links to potentially helpful resources, Wikipedia, news articles, explainers, etc…

Question 34

What is the probability that each actor in the list below will be the first to use a nuclear weapon on the territory or against the military forces of (A) a nuclear-armed adversary or (B) a treaty ally of a nuclear-armed adversary by 2030?

- China
- France
- India
- Israel
- North Korea
- Pakistan
- Russia
- The United Kingdom
- The United States
- Other actor (state)
- Other actor (non-state)

Question and Resolution Details

- This question is not conditional on a nuclear weapon having been detonated by 2030.
    - For example, assigning an X% probability to the United States is to say that there is an X% chance that at least one nuclear weapon will be detonated by 2030 and that the first use of a nuclear weapon (in the 21st century) before 2030 will have been by the United States.

- By “use” of a nuclear weapon, we refer specifically to the intentional detonation of one or more nuclear warheads. The nuclear weapons testing, accidental detonations, or the use of radiological devices (e.g. dirty bombs) do not count for resolving this question. Other kinds of nuclear explosions or nuclear accidents unrelated to nuclear warheads, such as reactor meltdowns or accidents from nuclear physics experiments, also would not resolve this question positively.
- By “nuclear-armed adversary,” we refer to any other state possessing nuclear weapons (i.e., any of the states on the given list).
- By “treaty ally of a nuclear-armed adversary,” we refer specifically to non-nuclear states that are in a security alliance with the given nuclear state and have been guaranteed defense in the case of a nuclear attack (i.e., existing within the given nuclear state’s “nuclear umbrella”). Currently, Japan, South Korea, Australia, and members of NATO have this type of relationship with the United States (see here for more).
- If resolution is ambiguous, this question will be resolved by a panel of experts.

Background Information

Information on Current Nuclear Weapons Programs and Inventories:

- **China**:
    - Nuclear Threat Initiative Profile: China
    - Center for Arms Control and Non-Proliferation Fact Sheet: China
    - Nuclear Notebook: Chinese nuclear forces, 2021 - Bulletin of the Atomic Scientists
    - Arms Control and Proliferation Profile: China
- **France**:
    - Nuclear Threat Initiative Profile: France
    - Center for Arms Control and Non-Proliferation Fact Sheet: France
    - Arms Control and Proliferation Profile: France
- **India**:
    - Nuclear Threat Initiative Profile: India
    - Center for Arms Control and Non-Proliferation: India and Pakistan
        - Fact Sheet: India's Nuclear Inventory
    - Nuclear Notebook: Indian nuclear forces, 2020 - Bulletin of the Atomic Scientists
    - Arms Control and Proliferation Profile: India
- **Israel**:
    - Nuclear Threat Initiative Country Profile: Israel
    - Center for Arms Control and Non-Proliferation: Israel
    - Nuclear Notebook: Israeli nuclear weapons, 2022 - Bulletin of the Atomic Scientists
    - Arms Control and Proliferation Profile: Israel
- **North Korea**:
    - Nuclear Threat Initiative Profile: North Korea
    - Center for Arms Control and Non-Proliferation: North Korea
    - Nuclear Notebook: How many nuclear weapons does North Korea have in 2021? - Bulletin of the Atomic Scientists
    - Arms Control and Proliferation Profile: North Korea
    - Additional Links:
        - North Korea's Nuclear Weapons and Missile Programs - Congressional Research Service
        - What Are North Korea's Military Capabilities? - Council on Foreign Relations
- **Pakistan**:

    - Nuclear Threat Initiative Profile: Pakistan
    - Center for Arms Control and Non-Proliferation: India and Pakistan
        - Fact Sheet: Pakistan's Nuclear Inventory
    - Nuclear Notebook: How many nuclear weapons does Pakistan have in 2021? - Bulletin of the Atomic Scientists
    - Arms Control and Proliferation Profile: Pakistan
- **Russia**:
    - Nuclear Threat Initiative Profile: Russia
    - Center for Arms Control and Non-Proliferation Fact Sheet: Russia
    - Nuclear Notebook: How many nuclear weapons does Russia have in 2022? - Bulletin of the Atomic Scientists
    - Arms Control and Proliferation Profile: Russia
    - Additional Links:
        - Russia's Nuclear Weapons: Doctrine, Forces, and Modernization - Congressional Research Service
- **United Kingdom**:
    - Nuclear Threat Initiative Profile: United Kingdom
    - Center for Arms Control and Non-Proliferation Fact Sheet: United Kingdom
    - Nuclear Notebook: How many nuclear weapons does the United Kingdom have in 2021? - Bulletin of the Atomic Scientists
    - Arms Control and Proliferation Profile: The United Kingdom
    - Additional Links:
        - The UK's nuclear deterrent: what you need to know - GOV.UK
- **United States**:
    - Nuclear Threat Initiative Profile: United States
    - Center for Arms Control and Non-Proliferation Fact Sheet: The United States' Nuclear Inventory
    - Nuclear Notebook: United States nuclear weapons, 2021 - Bulletin of the Atomic Scientists
    - Arms Control and Proliferation Profile: The United States

See also the main Nuclear detonationdocument for information on the two historical nuclear weapons attacks as well as a catalog of nuclear close calls.

Question 35

Will Robin Hanson win a bet that the GPT line of language models will generate less than $1 billion in customer revenue in total by the beginning of 2025?

Question and Resolution Details

Resolution is positive if Tabarrok publicly concedes the bet, negative if Hanson publicly concedes the bet, and will be decided by a panel of experts if nobody has conceded by the end of 2025.

1. If Tabarrok or Hanson publicly concede but then take back the concession by the end of 2025, the resolution will be decided by a panel of experts.
2. If the concession is rejected by the other person (i.e. neither side believes they've won the bet), the resolution will be decided by a panel of experts.
3. In a case where we defer to a panel of experts for resolution, the panel of experts will be asked to resolve the underlying question about GPT revenue.

Context

See the original tweet where Robin Hanson made this bet with relevant discussion in the replies.

Exact wording of bet as written by Hanson: ""Systems in GPT line will by 2025 make <$1B in customer revenue clearly tied to such systems. If product contains such as component, but also has other features, one needs to attribute best estimate % of product revenue to this one."

Source for this question and more details: Metaculus. And some further discussion here.

Relevant discussion

Related Metaculus discussion and forecasts. The Metaculus community prediction as of 1 Dec is 48% for a similar question.

OpenAI's pages on API has examples of apps using the GPT-3 API and their page on Pricing can be used for estimates. Prices, as of 21 Dec, range from $0.0008 (fastest) to $0.0600 (most powerful) per 1,000 tokens where 1,000 tokens is about 750 words. More details given on the page.

This OpenAI blog post for some more examples of apps using the GPT-3 API. At the time of writing that blog post, there were over 300 apps and over 10,000 developers using the API, generating a total of over 4.5 billion words per day. See this discussion on Reddit for how that could be used to estimate revenue.

The text of Tyler Cowen's Bloomberg article (paywalled) which prompted the bet:

Artificial Intelligence Is the Hope 2020 Needs

With attention focused on a pandemic and an election, AI has taken a major leap forward.

By Tyler Cowen

July 21, 2020, 1:00 PM UTC

This year is likely to be remembered for the Covid-19 pandemic and for a significant presidential election, but there is a new contender for the most spectacularly newsworthy happening of 2020: the unveiling of GPT-3. As a very rough description, think of GPT-3 as giving computers a facility with words that they have had with numbers for a long time, and with images since about 2012.

The core of GPT-3, which is a creation of OpenAI, an artificial intelligence company based in San Francisco, is a general language model designed to perform autofill. It is trained on uncategorized internet writings, and basically guesses what text ought to come next from any starting point. That may sound unglamorous, but a language model built for guessing with 175 billion parameters — 10 times more than previous competitors — is surprisingly powerful.

The eventual uses of GPT-3 are hard to predict, but it is easy to see the potential. GPT-3 can converse at a conceptual level, translate language, answer email, perform (some) programming tasks, help with medical diagnoses and, perhaps someday, serve as a therapist. It can write poetry, dialogue and stories with a surprising degree of sophistication, and it is generally good at common sense — a typical failing for many automated response systems. You can even ask it questions about God.

Imagine a Siri-like voice-activated assistant that actually did your intended bidding. It also has the potential to outperform Google for many search queries, which could give rise to a highly profitable company.

GPT-3 does not try to pass the Turing test by being indistinguishable from a human in its responses. Rather, it is built for generality and depth, even though that means it will serve up bad answers to many queries, at least in its current state. As a general philosophical principle, it accepts that being weird sometimes is a necessary part of being smart. In any case, like so many other technologies, GPT-3 has the potential to rapidly improve.

It is not difficult to imagine a wide variety of GPT-3 spinoffs, or companies built around auxiliary services, or industry task forces to improve the less accurate aspects of GPT-3. Unlike some innovations, it could conceivably generate an entire ecosystem.

There is a notable buzz about GPT-3 in the tech community. One user in the U.K. tweeted: "I just got access to gpt-3 and I can't stop smiling, i am so excited." Venture capitalist Paul Graham noted coyly: "Hackers are fascinated by GPT-3. To everyone else it seems a toy. Pattern seem familiar to anyone?" Venture capitalist and AI expert Daniel Gross referred to GPT-3 as "a landmark moment in the field of AI."

I am not a tech person, so there is plenty about GPT-3 I do not understand. Still, reading even a bit about it fills me with thoughts of the many possible uses.

It is noteworthy that GPT-3 came from OpenAI rather than from one of the more dominant tech companies, such as Alphabet/Google, Facebook or Amazon. It is sometimes suggested that the very largest companies have too much market power — but in this case, a relatively young and less capitalized upstart is leading the way. (OpenAI was founded only in late 2015 and is run by Sam Altman).

GPT-3 is also a sign of the underlying health and dynamism of the Bay Area tech world, and thus of the U.S. economy. The innovation came to the U.S. before China and reflects the power of decentralized institutions.

Like all innovations, GPT-3 involves some dangers. For instance, if prompted by descriptive ethnic or racial words, it can come up with unappetizing responses. One can also imagine that a more advanced version of GPT-3 would be a powerful surveillance engine for written text and transcribed conversations. Furthermore, it is not an obvious plus if you can train your software to impersonate you over email. Imagine a world where you never know who you are really talking to — “Is this a verified email conversation?” Still, the hope is that protective mechanisms can at least limit some of these problems.

We have not quite entered the era where “Skynet goes live,” to cite the famous movie phrase about an AI taking over (and destroying) the world. But artificial intelligence does seem to have taken a major leap forward. In an otherwise grim year, this is a welcome and hopeful development. Oh, and if you would like to read more, here is an article about GPT-3 written by … GPT-3.

Question 36

What percentage of US GDP will result from software and information services…

a. ...in 2024?
b. ...in 2030?
c. ...in 2050?

Question and Resolution Details

“Percentage of US GDP resulting from software and information services” is defined as the contribution of two categories: "Publishing industries, except internet (includes software)" and "Data processing, internet publishing, and other information services". It will be resolved according to seasonally adjusted “Value Added” data by the Bureau of Economic Analysis.

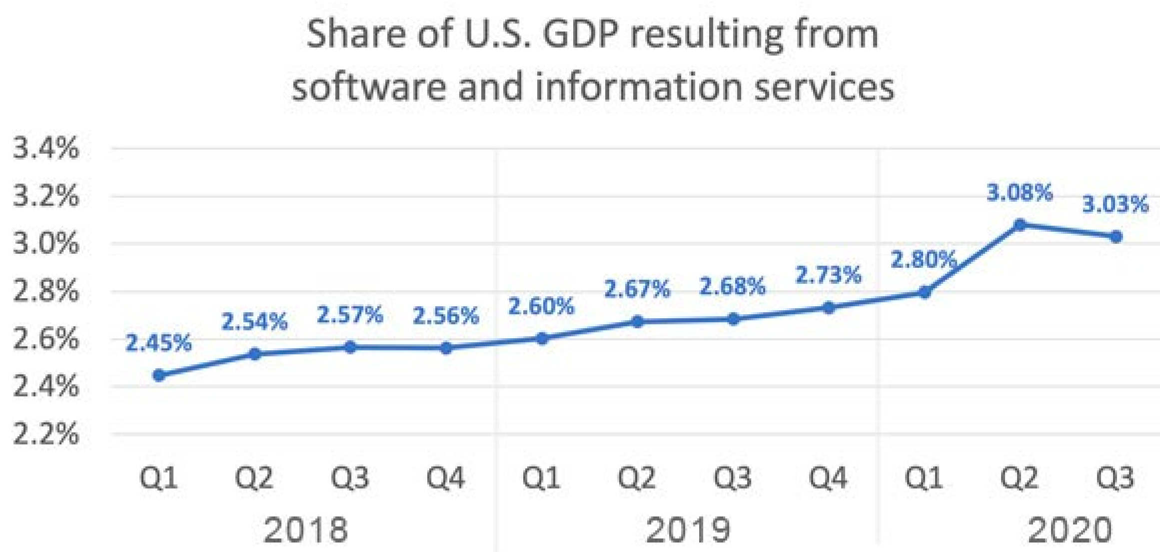
If current NAICS industry categories change or relevant new ones are added, a panel of experts will decide which categories to include.

Source for this question and more details: Hypermind (supported by Open Philanthropy).

- "Electricity, internal combustion engines, and semiconductors facilitated automation in the last century, but AI now seems poised to automate many tasks once thought to be out of reach, from driving cars to making medical recommendations and beyond. However, measured productivity growth has actually declined by half over the past decade (Brynjolfsson et al., 2017). To some extent, this may be evidence that information technology and other conventional stuff (non-informational inputs or outputs) aren't actually so cheaply or widely substitutable (Nordhaus, 2015). The prospects of growth of tech and automation may also be constrained by Baumol's "cost disease": sectors with rapid productivity growth are able to charge lower prices and subsequently have their share of GDP decline, whilst those with relatively slow productivity growth experience increases in their share of the value contributed to the economy. This might effectively cap the rate of growth of the value of tech as a proportion of the total economy (Aghion et al, 2017). Brynjolfsson et al. (2017) have argued that recent progress in AI and automation might well be radically productivity enhancing, but this might yet go largely unnoticed because of an implementation lag: it takes considerable time to be able to sufficiently harness technologies with broad potential application that they qualify as general purpose technologies. Will the economic data bear this out sometime soon?"

Historical Base Rate Data

Here is a spreadsheet with historical data from Q1 2018 to Q3 2020.



Find here an Interactive tool from Bureau of Economic Analysis to look at data for GDP-by-industry

Prior Forecasts

- Metaculus forecasts
  - https://www.metaculus.com/questions/6545/it-as--of-gdp-in-q3-2022/
  - https://www.metaculus.com/questions/6585/it-as--of-gdp-in-q3-2030/

- Source for this question and forecasts for 2022-2025: Hypermind (supported by Open Philanthropy).

Question 37

How much money will be spent on research and development by US companies in the ‘Information’ and ‘Computer systems design’ industries…

1. …in 2024?
2. …in 2030?
3. …in 2050?

Question and Resolution Details

- This will be resolved by the Business Research and Development reports covering 2030 and 2040, respectively. If this report ceases publication, a sufficiently similar report will be found for resolution. If no sufficiently similar report can be found, the question will be judged by a panel of experts.
- If current NAICS industry categories change or relevant new ones are added, a panel of experts will decide which categories to include.
- 'Information' industry covers businesses such as Google, Meta, and Microsoft. 'Computer systems design' currently includes OpenAI.
- See Table 67 here for figures between 2009 and 2018.

Historical Base Rate Data

- In 2018, $131,792 was spent on R&D by US companies in the 'Information' and 'Computer systems design' industries (see table below).

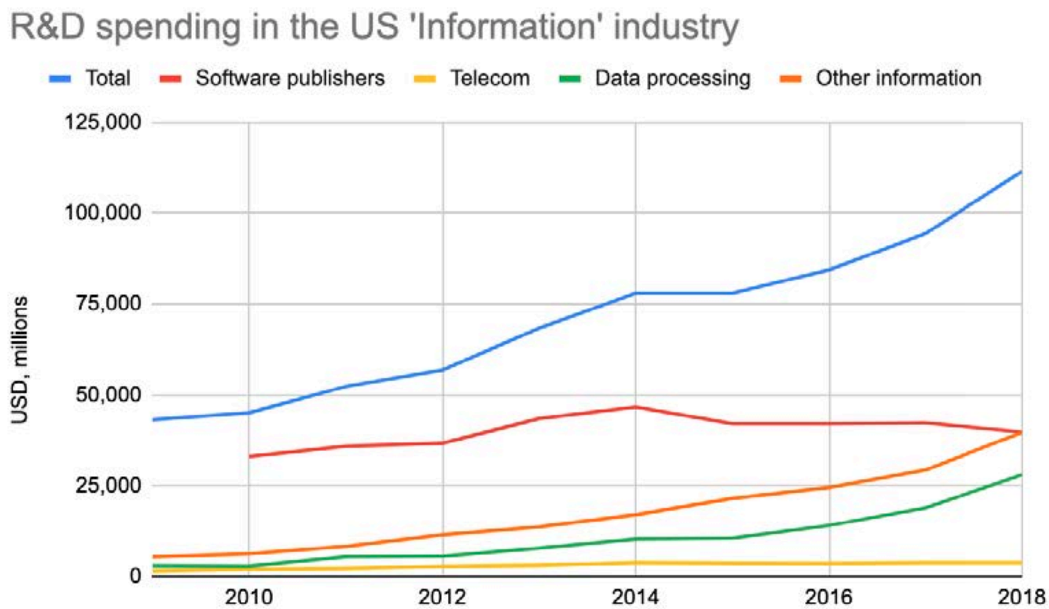
Worldwide R&D spending, by industry and year, in millions of USD

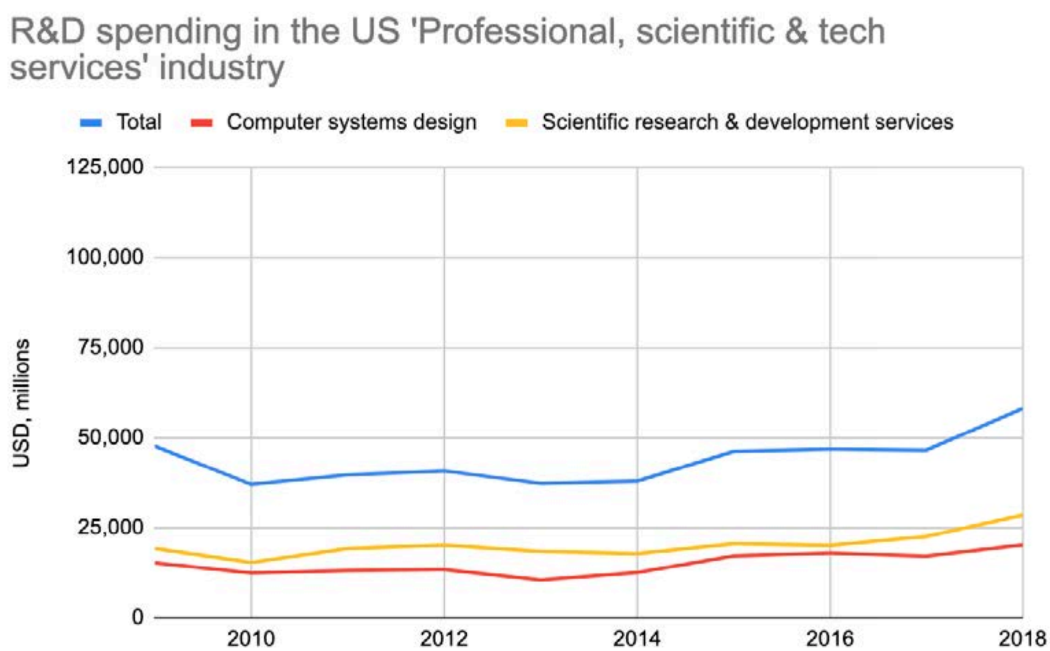
	Information					Professional, scientific & tech services		
	Total	*Software publishers*	*Telecom*	*Data processing*	*Other information*	Total	***Computer systems design***	*Scientific research & development services*
2009	**43,197**	NA	1,609	3,008	5,410	47,709	**15,310**	19,307
2010	**45,099**	33,058	2,043	2,895	6,278	37,108	**12,458**	15,408
2011	**52,341**	35,910	2,267	5,450	8,313	39,731	**13,181**	19,295
2012	**56,815**	36,726	2,787	5,574	11,506	40,843	**13,404**	20,225
2013	**68,375**	43,473	3,112	7,787	13,654	37,345	**10,535**	18,478
2014	**78,013**	46,714	3,830	10,356	17,022	37,995	**12,579**	17,870
2015	**77,965**	42,083	3,648	10,600	21,526	46,255	**17,227**	20,630
2016	**84,400**	42,073	3617 - 3,659	14051 - 14,099	24,530	39585 - 54,209	**15483 - 20,617**	20,194
2017	**94,412**	42,296	3,828	18,893	29,323	46,587	**17,190**	22,635

2018	**111,505**	39,815	3827 - 3,833	28,045	39662 - 39,668	58,168	**20,287**	28,555 - 28,564

(Taken from here, table 67. See here for the full report.) Table above does NOT include all sub-categories of 'Information' and 'Professional, scientific & tech services'. Targets of the question are bolded.

- *Information (selected categories)*
  - *Software publishers*
    - Includes e.g. Apple, Microsoft, Salesforce
    - See here for description and top companies by revenue.
  - *Telecommunications*
    - Includes e.g. AT&T, Paypal Inc
    - See here for description and top companies by revenue.
  - *Data processing, hosting & related services*
    - Includes e.g. Google, Facebook/Meta, Amazon.com Inc
    - See here for description and top companies by revenue.
  - Other information
    - Includes e.g. Facebook Inc, Linkedin, Yelp
    - See here for description and top companies by revenue.
- *Computer systems design services*
  - Includes e.g. OpenAI, Oracle
  - See here for description and top companies by revenue.


R&D spending in the US 'Information' industry
Total
Software publishers
Telecom
Data processing
Other information
USD, millions
125,000
100,000
75,000
50,000
25,000
0
2010
2012
2014
2016
2018


R&D spending in the US 'Professional, scientific & tech services' industry
Total
Computer systems design
Scientific research & development services
USD, millions
125,000
100,000
75,000
50,000
25,000
0
2010
2012
2014
2016
2018

Question 38

What will be the labor force participation rate in OECD countries…

- …in the year 2024?
- …in the year 2030?
- …in the year 2050?

Question details and resolution criteria

- Labor force participation rate is "the number of people [of working age] who are employed or actively seeking employment, divided by the total non-institutionalized, civilian working-age population. … It omits institutionalized people (in prisons, nursing homes, or mental hospitals) and members of the military." More background on labor force participation here.
- This question refers to labor force participation among 25-64 year-olds.
- This question will resolve with the OECD Total for labor force participation rate in the relevant year. The OECD Total is calculated as the aggregated labor force of all OECD countries, divided by the aggregated total working-age population of all OECD countries.
- By OECD countries, we mean current OECD countries as of the end of 2021 (through Costa Rica's admission to the group). For more details, see https://www.oecd.org/about/document/ratification-oecd-convention.htm
- In the event of human extinction, assume one human of working age exists, that they are living within the 2021 land borders of an OECD nation, and are not employed and are not actively seeking employment. This would resolve the question as a 0.

Historical data

- Find OECD charts and data on labor force participation rate here, going back to 1960. The lowest rate for all OECD countries was 67.2%, in 1960. The highest was 78.4%, in 2019. The lowest rate recorded for any individual OECD country was 54%, in 2007 (Turkey).
- The World Bank collects data on a larger set of countries, and has labor force participation rate data going back to 1990. See here for data and charts on labor force participation among people aged **15+**. The lowest worldwide rate estimated by the World Bank was 58.7%, in 2007. The highest was 65.6%, in 1990. The lowest rate reported by the World Bank for any individual region was 36.3%, in 2002 (West Bank & Gaza).
- US labor force statistics going back to 1950 can be found from the US Bureau of Labor Statistics, e.g. here and here.

Previous forecasts & predictions

- Monthly Labor Review (US Bureau of Labor Statistics) - Labor force projections overview and highlights, 2020-2030 (US only)
    - Monthly Labor Review (2010) - Projections of the labor force to 2050: a visual essay
- US Bureau of Labor Statistics - Labor Force Projections tables
- OECD Labor force forecast, available up to 2023
- New Zealand labour force projections: 2020 (base) - 2073
- Holzer 2019 - The US Labor Market in 2050: Supply, Demand and Policies to Improve Outcomes (See pgs 38 - 50 for charts and figures, incl population trends, occupational trends and automation displacement risk)
- Daheim & Wintermann - 2050: The Future of Work

Question 39

What will be the state-of-the-art accuracy of a machine-learning model on the MATH Dataset by…

...June 30, 2024?

...June 30, 2030?

Resolution criteria:

- This question resolves as the highest performance achieved on MATH by the given date by an eligible model.
    - Eligible models may use scratch space before outputting an answer (if desired) and may be trained in any way that does not use the test set (few-shot, fine tuned, etc.). The model need not be publicly released, as long as the resulting performance itself is reported in a published paper (on arxiv or a major ML conference) or through an official communication channel of an industry lab (e.g. claimed in a research blog post of the lab, or in a press release).
    - If this leaderboard is not in use by the given date, this will be resolved by an expert (ie by searching the web for relevant papers, blog posts and press releases).
    - Research aiming to solve only a subset of the benchmark questions will be considered, but those attempts' scores on non-targeted questions will be assumed to be zero.

- If the benchmark is not actively used (i.e., no new known attempts to improve on this benchmark in the two years prior to the resolution date), this question will resolve by asking a panel of experts to state their beliefs about what performance on the benchmark would be if a currently-active, high-quality group of 10 researchers spent two weeks trying to apply current methods to this benchmark.

Context:

"Many intellectual endeavors require mathematical problem solving, but this skill remains beyond the capabilities of computers. To help advance the art, the MATH dataset offers 12,500 challenging competition mathematics problems on which to train and test the abilities of machine-learning models. Each problem in MATH has a full step-by-step solution which can be used to teach models to generate answer derivations and explanations.

At the time of this writing in June 2021, as reported on this leaderboard, state-of-the-art performance of machine-learning models on MATH is very low: 6.9%. In comparison, a computer science PhD student who does not especially like mathematics attained approximately 40%, while a three-time International Mathematical Olympiad gold medalist attained 90%."

Source for this question and more details: Hypermind (supported by Open Philanthropy). Some further explanation of how this question caused Jacob Steinhardt to update his views here.

Forecasts and predictions:

- In 2021, forecasters on Hypermind predicted the following accuracy performance by June 30 of various years:
  - 12.7% in 2022
  - 20.7% in 2023
  - 30% in 2024
  - 50% in 2025

- Metaculus users forecast accuracy on the MATH dataset by 2025 here.

Historical data:

Currently, the leaderboard only has models from the paper "Measuring Mathematical Problem Solving With the MATH Dataset". See the following table from the paper for more details including number of parameters for the models:

Model	Prealgebra	Algebra	Number Theory	Counting & Probability	Geometry	Intermediate Algebra	Precalculus	Average
GPT-2 0.1B	5.2	5.1	5.0	2.8	5.7	6.5	7.3	5.4 +0%
GPT-2 0.3B	6.7	6.6	5.5	3.8	6.9	6.0	7.1	6.2 +15%
GPT-2 0.7B	6.9	6.1	5.5	5.1	8.2	5.8	7.7	6.4 +19%
GPT-2 1.5B	8.3	6.2	4.8	5.4	8.7	6.1	8.8	6.9 +28%
GPT-3 13B*	4.1	2.4	3.3	4.5	1.0	3.2	2.0	3.0 −44%
GPT-3 13B	6.8	5.3	5.5	4.1	7.1	4.7	5.8	5.6 +4%
GPT-3 175B*	7.7	6.0	4.4	4.7	3.1	4.4	4.0	5.2 −4%

Table 2: MATH accuracies across subjects. '*' indicates that the model is a few-shot model. The character 'B' denotes the number of parameters in billions. The gray text indicates the *relative* improvement over the 0.1B baseline. All GPT-2 models pretrain on AMPS, and all values are percentages. GPT-3 models do not pretrain on AMPS due to API limits. Model accuracy is increasing very slowly, so much future research is needed.

Resolution criteria:

This question resolves as the highest performance achieved on MATH by the given date by an eligible model.

Eligible models may use scratch space before outputting an answer (if desired) and may be trained in any way that does not use the test set (few-shot, fine tuned, etc.). The model need not be publicly released, as long as the resulting performance itself is reported in a published paper (on arxiv or a major ML conference) or through an official communication channel of an industry lab (e.g. claimed in a research blog post of the lab, or in a press release).

If this leaderboard is not in use by the given date, this will be resolved by an expert (ie by searching the web for relevant papers, blog posts and press releases)

If the benchmark is not actively used in the relevant year (i.e. no new known attempts to improve on this benchmark in the year prior to the date), we will use a panel of experts to estimate what performance state-of-the-art machine learning models in the given year would have on the MATH dataset.

Question 40

What will be the state-of-the-art few-shot or transfer accuracy on the Massive Multitask Language Understanding dataset…

...by June 30, 2024?
...by June 30, 2030?

Question and Resolution Details

- This question resolves as the highest performance achieved on MMLU by the given date by an eligible model.
- Eligible models must not have been specifically trained on data from the MMLU dataset. A model need not be publicly released, as long as the resulting performance itself is reported in a published paper (on arxiv or a major ML conference) or through an official communication channel of an industry lab (e.g. claimed in a research blog post on the OpenAI blog, or a press release). If there's uncertainty about whether something counts, we will defer to this leaderboard.
- Research aiming to solve only a subset of the benchmark questions will be considered, but those attempts' scores on non-targeted questions will be assumed to be zero.
- If the benchmark is not actively used (i.e., no new known attempts to improve on this benchmark in the two years prior to the resolution date), this question will resolve by asking a panel of experts to state their beliefs about what performance on the benchmark would be if a currently-active, high-quality group of 10 researchers spent two weeks trying to apply current methods to this benchmark.
- This question resolves as the highest performance achieved on MMLU by the given date by an eligible model.
- Eligible models must not have been specifically trained on data from the MMLU dataset. A model need not be publicly released, as long as the resulting performance itself is reported in a published paper (e.g. on arxiv or a major ML conference) or through an official communication channel of an industry lab (e.g. claimed in a research blog post on the OpenAI blog, or a press release).
- If there's uncertainty about whether something counts, we will defer to this leaderboard.

Context

The Massive Multitask Language Understanding test (MMLU) measures a text model's accuracy on 57 tasks including elementary mathematics, US history, computer science, law, and more. To attain high accuracy on this test, models must possess extensive world knowledge and problem solving ability. All questions are multiple choice.

Additional information on the MMLU benchmark can be found in the paper, here.

Prior forecasts

- Source for this question and more details with a crowd forecast and forecasting history for a similar question: Hypermind (supported by Open Philanthropy). Some further explanation of how this question caused Jacob Steinhardt to update his views here (in addition to examples of questions in this dataset).

- Forecasts on Metaculus for a similar question with different endpoints.

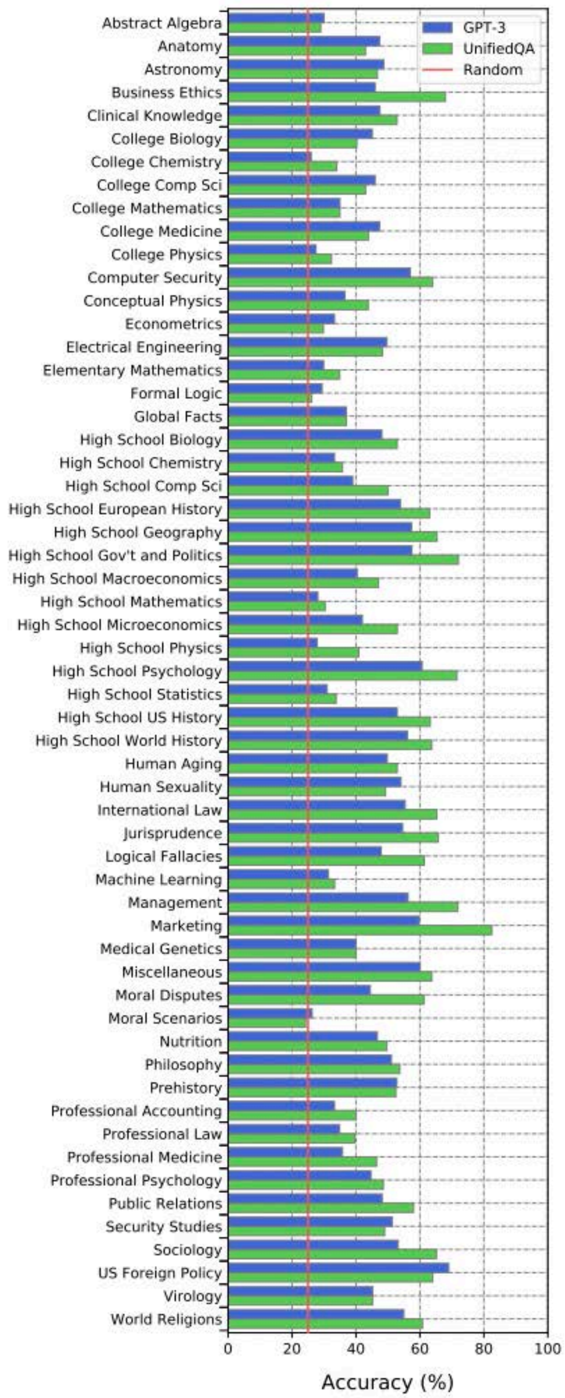
Historical Base Rate Data

As of June 2022, this is the leaderboard:

Model	Authors	Humanities	Social Sciences	STEM	Other	Average
Chinchilla (70B, few-shot)	Hoffmann et al., 2022	73.1	78.8	55	70.3	67.5
Gopher (280B, few-shot)	Rae et al., 2021	65.8	71.2	47.6	63.5	60.0
GPT-3 (175B, fine-tuned)	Brown et al., 2020	52.5	63.9	41.4	57.9	53.9
UnifiedQA	Khashabi et al., 2020	45.6	56.6	40.2	54.6	48.9
GPT-3 (175B, few-shot)	Brown et al., 2020	40.8	50.4	36.7	48.8	43.9
GPT-3 (6.7B, fine-tuned)	Brown et al., 2020	42.1	49.2	35.1	46.9	43.2
GPT-2	Radford et al., 2019	32.8	33.3	30.2	33.1	32.4
Random Baseline	N/A	25.0	25.0	25.0	25.0	25.0

Note that UnifiedQA has 11 billion parameters but in contrast to GPT-3 (175B, few-shot), UnifiedQA's training has been fine-tuned on other question-answering data sets. The results above suggest that while model size is a key component for achieving strong performance, fine-tuning also helps.

The following figure shows the accuracy performance of UnifiedQA and the GPT-3 (175B, few-shot) model across the 57 tasks.



For comparison, expert-level accuracy on this test is approximately 90%. Unspecialized humans from Amazon Mechanical Turk obtained 34.5% accuracy.

Question 41

What will be the best SAT-style score with a machine learning model on the hard subset of the QuALITY dataset by…

a. …June 30, 2024?
b. …June 30, 2030?
c. …June 30, 2040?

Question and Resolution Details

- This is an edited version of a Metaculus question, see here.
- Language models have surpassed human performance in benchmarks requiring reasoning over short texts, but performance remains low on reasoning over long texts. The QuALITY dataset (Peng et al 2021) was created to test long-document comprehension, using “a multiple-choice QA dataset with context passages in English that have an average length of about 5,000 tokens, much longer than typical current models can process.”
- The hard subset of QuALITY questions includes questions that require a more holistic understanding of the text to answer correctly. They were chosen as the questions that human answerers performed poorly on with very strict time constraints. SAT-style scoring penalizes wrong answers more than “no answer” so that answering questions with high uncertainty becomes less attractive.
- This question will be resolved according to this leaderboard. In the case of this leaderboard no longer being maintained, this question will be resolved by a panel of experts.

Historical Base Rate Data

- The leaderboard shows how well models have performed on this task so far. Human SAT-style score on the hard subset is 85.4, while the top SAT-style score on the hard subset for a machine learning model is 28.1. For the full test set, human performance is 91.4 (SAT-scored) and top machine performance is 40.5 (SAT-scored).
- The SuperGLUE language understanding benchmark was released 2.5 years ago. At the time, the best model achieved a score of 71.5 while humans achieved 89.8. Since then, the human baseline has been beaten by 5 models (see leaderboard here).
- See also a previous long-text task, NarrativeQA (Kočiský et al 2018) (limitations discussed briefly in Peng et al 2021).

Prior Forecasts

- The Metaculus community predicts that the highest score will be 63 by 2025 (10 forecasters) and 88 by 2040 (12 forecasters).

Question 42

By what year will an AI win a Gold Medal in the International Mathematical Olympiad (IMO)?

Question and Resolution Details

- This question will be resolved in either the first year that the official IMO Grand Challenge is won by an AI given the rules set forth by the challenge OR the first year that a panel of experts determines that an AI has the technical capability to win the challenge.
  - The reason for the latter disclaimer is that currently the official grand challenge has a rule that the AI must be open-source and released publicly before the first day of the IMO. However, in this question we care about forecasting AI's technical capability to win the challenge, even if it is won by a private model. In the case of a private model resolving this question, the private model would need to be judged by a panel of experts adhering to the same resolution criteria as the Grand Challenge except for the criterion about an AI needing to be open-source and reproducible. That is, given that IMO problems in any given year are made available in Lean or another formal representation format, each proof certificate that the AI competitor produces must be checkable in 10 minutes; the AI competitor will be allowed the same amount of time as a human competitor to work on the problems, with no other limits on compute; and the AI cannot query the internet.
  - The current rules of the official challenge are still tentative. If the rules evolve, we assume that the question will resolve based on the most current version of the rules at any given time. If the rules eventually depart in substantial ways from the currently proposed rules, the resolution of this question may be referred to a panel of experts.
- In the event that this never happens, predict number 1E210, equivalent to the largest number able to be submitted on the platform used to host the tournament.

Prior Forecasts

This question is based upon a public bet between Paul Christiano and Eliezer Yudkowsky made as a result of their earlier public debates about AI progress.

- As of February 2022, Christiano's most recent forecast is a <8% chance that an AI will win a gold medal on the IMO in at least one year out of 2022, 2023, 2024, and 2025.
- As of February 2022, Yudkowsky's most recent forecast is a >16% chance that an AI will win a gold medal on the IMO in at least one year out of 2022, 2023, 2024, and 2025.
  - Yudkowsky specifies that he's forecasting a >16% chance that the *technical capability* to win a gold medal on the IMO exists by 2025. His forecast does not account for the grand challenge's rule that the AI must be open-source and released publicly before the first day of the IMO (which he implies is a constraint that would lower his overall probability of a gold medal being won by an AI).

Further Background Information

- Solving (Some) Formal Math Olympiad Problems | OpenAI
  - For technical details, see this paper.
- Exploring the beauty of pure mathematics in novel ways | DeepMind
- See here for an archive of problems from past IMO competitions, and here for general background on the IMO competition.

Question 43

By what year will AI have written at least 3 books that appear on the New York Times Best Seller list?

Question and Resolution Details

- Artificial intelligence (AI) is defined here as 'the development of machines capable of sophisticated (intelligent) information processing (Dafoe 2018, *AI Governance: A Research Agenda*; See pg 5 footnote 2 for more.)
- A book is said to have been written by an AI if the AI wrote at least 99% of the text contained in the main section in the book, **excluding** a potential foreword, copyright notice, table of contents, and other non-essential book sections. The main text must also contain at least 20,000 words. Minor stylistic edits by humans are allowed if they do not change the basic semantic meaning of the text, or they merely correct basic spelling, grammatical, or formatting mistakes.
  - The final text may be produced by a human 'cherry-picking', that is, choosing from among multiple examples of AI-generated text, provided that there is no more than one 'picking' per 1,000 words of final text on average (i.e., 20 'pickings' across a 20,000 word text). A single 'picking' can include consideration of more than 2 versions of a single passage but not more than 10.
  - If >99% of the text is written by an AI in response to prompts from humans (e.g. prompting GPT-3 to write text), then this would count as a book having been written by an AI for the purposes of this question. If natural language prompts are used, then the total word count of all human prompts used to generate the final text must be less than 400 words. If non-natural language prompts are used, they must be constrained to a 'semantically equivalent' length to a 400 natural language word prompt, as judged by a panel of experts.
  - If a human does substantial editing such that they effectively wrote >1% of the words, then this would disqualify the text.

    - If the generation methods used are too opaque to determine whether a text meets these criteria, a panel of experts will decide, either by soliciting evidence from parties involved in the production of the text, by commissioning relevant experiments, or by referring to past experiments in the public record.
- If the bestseller list creates a category for books written by AI, a book will not be counted if it appears solely on that list.
- If the New York Times Best Seller list doesn't exist by the time AI-written books are prominent, we'll rely on a panel of experts to make a determination based on a similar methodology.
    - According to this article, "The New York Times list is a survey list, not a tabulation of total sales. This means that they poll a curated selection of booksellers to estimate sales. … They also focus on individual sales, and try to not include bulk sales in their calculations. They do this to prevent people from buying their way onto the list. If you sell 1,000 copies to a company as part of a speaking engagement deal, this is a great way to move copies and make money, but it's not very effective for hitting the list, because they won't count it."
- If an AI translates a book written by a human, this wouldn't count for the purposes of this question.
- We will consult a panel of experts to collect information about publicly-known AI authors of books and widely known (even if not publicly announced) AI authors of books.

Prior Forecasts

See this Metaculus page for forecasts on a related question: whether a book written by a language model will make the NY Times Bestseller list before 2030. As of 1 Jan 2022, the community prediction there for that question is 20%. Relevant ways in which our question is different:

- It looks past the year 2030
- Year in which we will get 3 books instead of 1
- The book written by AI doesn't necessarily have to be written by a language model specifically

When Will AI Exceed Human Performance? Evidence from AI Experts survey that collected predictions from 350 AI experts on when AI will write a New York Times Best-Seller, got a median estimate of 2049. See this post on the World Economic Forum website for discussion of the survey.

Other Potential Background Information

Wikipedia: 'The New York Times Best Seller list is widely considered the preeminent list of best-selling books in the United States. It has been published weekly in The New York Times Book Review since October 12, 1931. In the 21st century, it has evolved into multiple lists, grouped by genre and format, including fiction and non-fiction, hardcover, paperback and electronic.'

The question on Robin Hanson's bet that the GPT line of language models will generate less than $1Bn in customer revenue by 2025 might also be relevant. According to an EPJ Data Science study that used big data to analyze every New York Times bestselling book from 2008–2016, the majority of bestsellers sell between 10,000 and 100,000 copies in their first year. Assuming an average book price of around $20, this comes to 200,000 to 2,000,000 in revenue from each bestselling book in its first year.

Question 44

When will the first unified AI system meeting all of the following criteria be trained, tested, and publicly known of?

- **1) Able to reliably pass a 2-hour adversarial Turing test.**
- **2) High competency at answering questions across diverse fields of expertise.**
- **3) High competency on interview-level problems in the APPS benchmark.**
- **4) Able to learn the classic Atari game "Montezuma's revenge" in the equivalent of 100 hours or less of real-time play.**

Question and Resolution Details

- This question is inspired by these Metaculus questions: I II
- An AI system is 'unified' if it is a single system applicable to many problems (rather than a set of sub-systems specialized narrowly to individual tasks like those above). This could be evidenced by, for example, the AI system being able to explain its reasoning on a Q&A task, or verbally reporting its progress and identifying objects during videogame play. Any ambiguity about whether a system is sufficiently 'unified' to qualify for resolving this question will be decided by a majority vote from a panel of AI experts.
- Further details about necessary criteria:
  - 1) In an 'adversarial' Turing test, human judges are instructed to ask interesting and difficult questions designed to advantage human participants and to successfully unmask the computer as an imposter. During the conversation, participants can send text, images, and audio files. A single demonstration of an AI passing such a Turing test will be sufficient for this condition, so long as the test is well-designed, as judged by a panel of AI experts.
  - 2) On the Q&A dataset developed by Dan Hendrycks et al., at least 75% accuracy in every task and 90% mean accuracy across all tasks.
  - 3) Able to get top-1 strict accuracy of at least 90.0% on interview-level problems found in the APPS benchmark introduced by Dan Hendrycks, Steven Basart et al. Top-1 accuracy is distinguished from top-k accuracy in which k outputs from the model are generated, and the best output is selected.
  - 4) Montezuma's revenge must be learned based on just visual inputs and standard controls, and must explore all 24 rooms on the first level of the game.
- Resolution will be by direct demonstration of such a system achieving the above criteria. Ambiguity will be resolved by a >=75% majority vote by a panel of at least 5 AI experts.

Background Information

- For a sample of the most advanced models developed in the past year, see PaLM from Google, DALLE-2 from OpenAI, Flamingo from DeepMind, and Gato from DeepMind.
- For information on compute in training runs, see AI and Compute from OpenAI.
- Base rates of relevant ML benchmarks:
    - Recent progress on Montezuma's revenge & Atari game-playing:
        - 2019, Go-Explore (insufficient to resolve this Metaculus question)
        - 2022, Gato (general game-playing, 'unified' system)
    - Massive Multitask Language Understanding test (leaderboard here)
    - MATH Dataset for mathematical learning
    - QuALITY Dataset for long-text comprehension
        - Also for language understanding, see GLUE and SuperGLUE
    - ImageNet for image classification
        - Also for image recognition, see CIFAR-10 and CIFAR-100
            - Leaderboards here and here respectively

Question 45

What will be the maximum compute (measured in petaFLOPS-days) used for training in an AI experiment...

A. ...by end of 2024?
B. ...by end of 2030?
C. ...by end of 2050?

Question and Resolution Details

- This question shall resolve as a credible estimate of most compute used in a "single AI experiment".
  - By "single AI experiment" we mean an effort to train a set of models running on a set of "architectures". The effort should be completed within a determinate amount of time (the experiment must not have an open-ended time frame). For our purposes, the publication of the principal results of the effort ends the experiment.
  - By "architectures" we mean the systems described in the relevant publications that define how inputs signal or percept sequences are to be mapped on various outputs. These outputs might be probability distributions over actions (in the case of a policy network), representations over expected value or reward of futures states (in the case of value networks) or descriptions of futures states.
- In the absence of an authoritative source, the question will be resolved by a panel of experts.
  - The panel of experts will by default use the methodology from [OpenAI, 2018] (see [Heim et al 2022] for further detail), unless a majority of the panel agrees that another methodology is more suitable. OpenAI's method:
    - "When we had enough information, we directly counted the number of FLOPS (adds and multiplies) in the described architecture per training example and multiplied by the total number of forward and backward passes during training."
    - "When we didn't have enough information to directly count FLOPs, we looked GPU training time and total number of GPUs used and assumed a utilization efficiency (usually 0.33). For the majority of the papers we were able to use the first method, but for a significant minority we relied on the second, and we computed both whenever possible as a consistency check. In the majority of cases we also confirmed with the

authors. The calculations are not intended to be precise but we aim to be correct within a factor 2-3. We provide some example calculations below."

Context

Source for this question and more details: Hypermind (supported by Open Philanthropy). Note that this version of the question has a different endpoint.

"Various figures in AI research have noted the importance of computer hardware for AI progress (LeCun et al., 2015; Schmidhuber, 2015). Exponential improvements in the computing power available for a given price, and the development of particular technologies such as graphics processing units (GPUs), have accelerated progress in multiple AI domains (Brundage, 2016). Recent progress in has been accompanied by the use of increasing amounts of computation. **According to OpenAI, since 2012, the amount of compute used in the largest AI training runs has been increasing exponentially with a 3.5 month doubling time.** This trend sustained by both the continued progress in hardware performance, and increased spending on AI experiments (AI Impacts, 2018)."

Historical Base Rate Data

- The Akronomicon: an Extreme-Scale Leaderboard. Keeps track of the largest and most powerful machine learning models, listing their training compute in petaFLOPS-days.
- Spreadsheet on "Parameter, Compute and Data Trends in Machine Learning" from a research effort to collate the biggest ever public dataset on parameters, compute and dataset size for landmark AI models, with data from the 1950s until recently. Has compute in FLOP. (Sevilla et al 2022)
    - See also this interactive visualization of the dataset on Our World In Data.

Prior Forecasts

- See related questions on Metaculus for maximum compute used for training in an AI experiment by January 2022, January 2026, and January 2031. There are also forecasts on Hypermind for January 2026, forecasting on that question ended in April 2021 and the crowd forecast was 39784 exaFLOPS-days (39.784 petaFLOPS-days).
- Sevilla et al 2022 proposes three eras of machine learning compute since 1950:

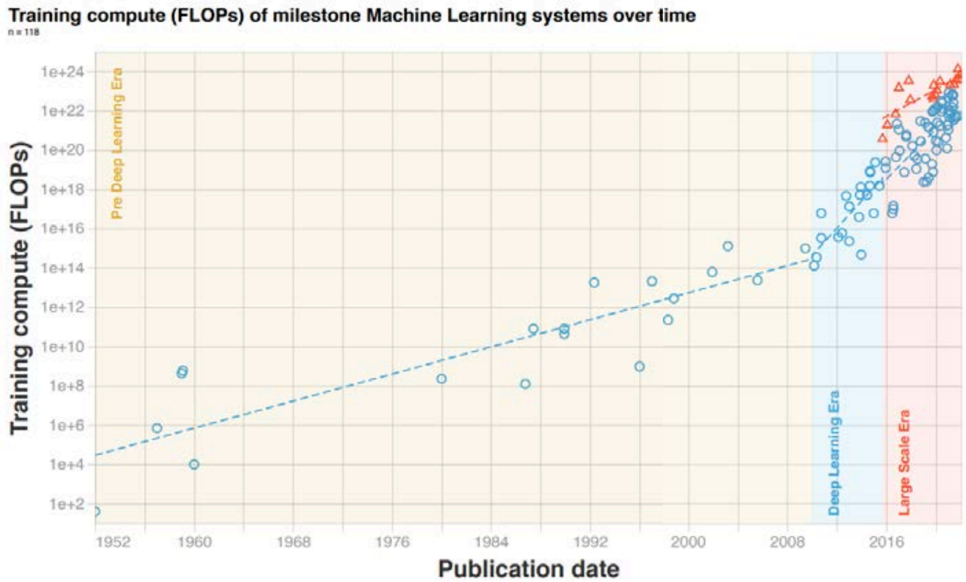


Figure 1: Trends in $n = 118$ milestone ML models between 1952 and 2022. We distinguish three eras. Notice the change of slope circa 2010, matching the advent of Deep Learning; and the emergence of a new large-scale trend in late 2015.

Period	Data	Scale (start to end)	Slope	Doubling time
1952 to 2010 Pre Deep Learning Era	No low outliers ($n = 19$)	3e+04 to 2e+14 FLOPs	0.2 OOMs/year [0.1; 0.2; 0.2]	21.3 months [17.0; 21.2; 29.3]
2010 to 2022 Deep Learning Era	No outliers ($n = 72$)	7e+14 to 2e+18 FLOPs	0.6 OOMs/year [0.4; 0.7; 0.9]	5.7 months [4.3; 5.6; 9.0]
September 2015 to 2022 Large-Scale Era	High outliers ($n = 16$)	4e+21 to 8e+23 FLOPs	0.4 OOMs/year [0.2; 0.4; 0.5]	9.9 months [7.7; 10.1; 17.1]

Table 2: Summary of our main results. In 2010 the trend accelerated along the with the popularity of Deep Learning, and in late 2015 a new trend of large-scale models emerged.

- This post by OpenAI written in 2018 discusses that since 2012, the amount of compute used in the largest AI training runs has been increasing exponentially with a 3.4-month doubling time.
    - See also this post from April 2021 that aims to show that the trend has ended right around the moment of OpenAI publishing their post and doesn't hold up anymore. See further discussion of that post on Reddit which includes comments discussing how this may have been a temporary phenomenon and caused by supply chain disruptions.
    - This post from AI Impacts discusses financial constraints which may limit the current compute growth trend in the future.

- Lennar Heim on the effective altruism forum wrote a post on forecasting compute that attempted to propose ideas for better compute forecasts. The piece does not outline a concrete strategy on how to go about it; instead, it lists various ideas.

Question 46

How much will be spent on compute in the largest AI experiment by…

a. …the end of 2024?
b. …the end of 2030?
c. …the end of 2050?

Question and Resolution Details

- Artificial intelligence (AI) is defined here as 'the development of machines capable of sophisticated (intelligent) information processing (Dafoe 2018, *AI Governance: A Research Agenda*; See pg 5 footnote 2 for more.)
- 'Largest' will be in terms of maximum compute used in training, in petaFLOPS-days. The experiment must be a 'single AI experiment', that is, 'an effort to train a set of models running on a set of "architectures" … [which] should be completed within a determinate amount of time (the experiment must not have an open-ended time frame).' (More here.)
- Cost will be in terms of 2021 US dollars.
- For this question, we are interested in the most expensive experiment from between June 15th, 2022 and the resolution year. Experiments performed prior to June 15th, 2022 will not be included.
- Size of the largest experiment will be resolved using a credible estimate of the most compute used in a "single AI experiment."
    - By "single AI experiment" we mean an effort to train a set of models running on a set of "architectures". The effort should be completed within a determinate amount of time (the experiment must not have an open-ended time frame). For our purposes, the publication of the principal results of the effort ends the experiment.
    - By "architectures" we mean the systems described in the relevant publications that define how inputs signal or percept sequences are to be mapped on various outputs. These outputs might be probability distributions over actions (in the case of a policy network), representations over expected value or reward of futures states (in the case of value networks) or descriptions of futures states.
    - In the absence of an authoritative source, the question will be resolved by a panel of experts.

a. The panel of experts will default to using the methodology used by OpenAI (OpenAI, 2018) unless there is consensus to use a different methodology. OpenAI's method:
    i. "When we had enough information, we directly counted the number of FLOPS (adds and multiplies) in the described architecture per training example and multiplied by the total number of forward and backward passes during training."
    ii. "When we didn't have enough information to directly count FLOPs, we looked GPU training time and total number of GPUs used and assumed a utilization efficiency (usually 0.33). For the majority of the papers we were able to use the first method, but for a significant minority we relied on the second, and we computed both whenever possible as a consistency check. In the majority of cases we also confirmed with the authors. The calculations are not intended to be precise but we aim to be correct within a factor 2-3."

■ Cost resolution details:
- Afterward, a panel of experts will estimate how much the compute for the experiment cost. Some examples of past estimates can be found here, here, or here.
    - For one example of how the experts could estimate experiment cost: if Google itself carried out the experiment using its own TPUs, this would correspond to a complicated calculation involving the cost of producing the TPUs, the proportion of the lifetime of the TPUs which was used up in the experiments, Google's profit margin for these TPUs, etc. A result might look like "it costs Google 30 to 70% of their retail price to use their own TPUs, which corresponds to a compute cost of $X."
    - For another example, if an external company got a special discount from Microsoft to run their experiments on Microsoft Azure, this number would be an estimate of how much the external company paid Microsoft.
- Fine-print:
    - 16 or 32-bit floating point operations will be regarded as equivalent.
    - The units for this question are petaflop/s-day per dollar, to keep with historical usage by Metaculus or by OpenAI. A petaflop/s-day (pfs-day) consists of performing 10^{15} neural net operations per second for one day, or a total of about 10^{20} operations

Historical Base Rate Data

Estimated size of largest known experiments by year		
Year	Experiment Size (FLOP)	Model
2020	3×10^{23}	GPT-3
2019	2×10^{23}	AlphaStar
2018	3×10^{21}	BigGAN
2017	2×10^{23}	AlphaGoZero
2016	7×10^{21}	Neural Machine Translation
2015	3×10^{19}	DeepSpeech2
2014	9×10^{18}	VGG
2013	5×10^{17}	Visualizing & Understanding Conv Nets
2012	5×10^{17}	AlexNet

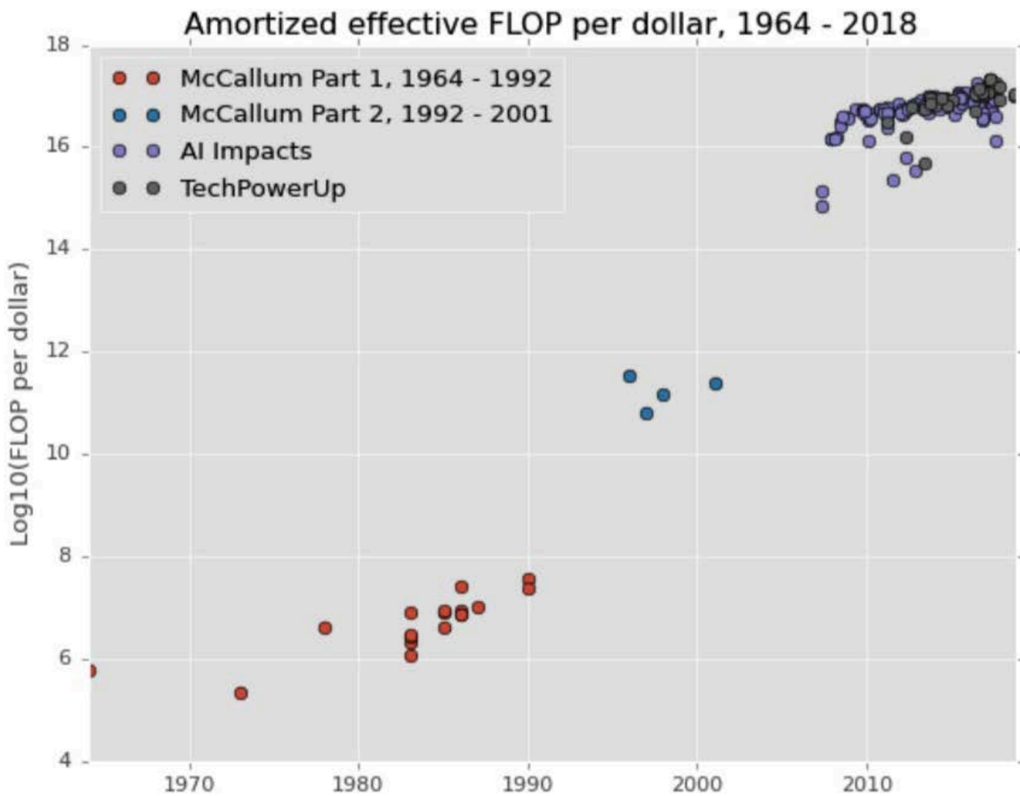
Sources: (I)(II)

Estimates of large experiment compute costs

- 2020 GPT-3: $4.6M (III)
- 2017 AlphaGoZero: $10M (IV) - $35M (V)
-

Amortized effective FLOP per dollar over time

(This figure takes as a starting point either a) the price performance, with performance measured on LINPACK benchmark; or b) the theoretical maximum price performance of a chip, as reported by manufacturer, x0.73. From here, the cost of the chip is amortized over 2 years of continuous use, and then an adjustment is applied to account for additional costs of running a data center and for imperfect utilization, a factor of ⅛x together. See more here.)



Prior Forecasts

- Interpreting AI compute trends
    - "The size of the largest experiments is increasing with a doubling time of 3.5 months, (about an order of magnitude per year), while the cost per unit of computation is decreasing by an order of magnitude every 4-12 years (the long-run trend has improved costs by 10x every 4 years, whereas recent trends have improved costs by 10x every 12 years). So the cost of the largest experiments is increasing by an order of magnitude every 1.1 – 1.4 years. The largest current experiment, AlphaGo Zero, probably cost about $10M. … Previously, the US spent 1% of annual GDP on the Manhattan Project, and ~0.5% of annual GDP on NASA during the Apollo program. So let's suppose they could similarly spend at most 1% of GDP, or $200B, on one AI experiment. Given the growth of one order of magnitude per 1.1-1.4 years, and the initial experiment size of $10M, the AI-Compute trend predicts that we would see a $200B experiment in 5-6 years. So given a broadly similar economic situation to the present one, that would have to mark an end to the AI-Compute trend. … The largest [private actors] are tech companies: Amazon and Google have current research and development budgets of about ~20B/yr each[8], so we can suppose that the largest individual

experiment outside of government is $20B. Then the private sector can keep pace with the AI-Compute trend for around ¾ as long as government, or ~3.5-4.5 years. On the other hand, the development of specialized hardware could cheapen computation, and thereby cause the trend to be sustainable for a longer period.”

- Forecasting TAI with biological anchors (part 4)
    - Best guess forecasts for hardware, spending & algorithms includes discussion of FLOP per dollar forecasts, and forecasted amounts that various actors would be willing to spend on training runs

Other Potential Background Information

- Links to potentially helpful resources, Wikipedia, news articles, explainers, etc…
- AI and Compute | OpenAI
- "AI and Compute" trend isn't predictive of what is happening | Alignment Forum
- Trends in the cost of computing – AI Impacts
- AI's Smarts Now Come With a Big Price Tag | WIRED

Question 47

What will be the lowest price, in 2021 US dollars, of 1 GFLOPS with a widely-used processor by…

…the end of 2024?
…the end of 2030?
…the end of 2050?

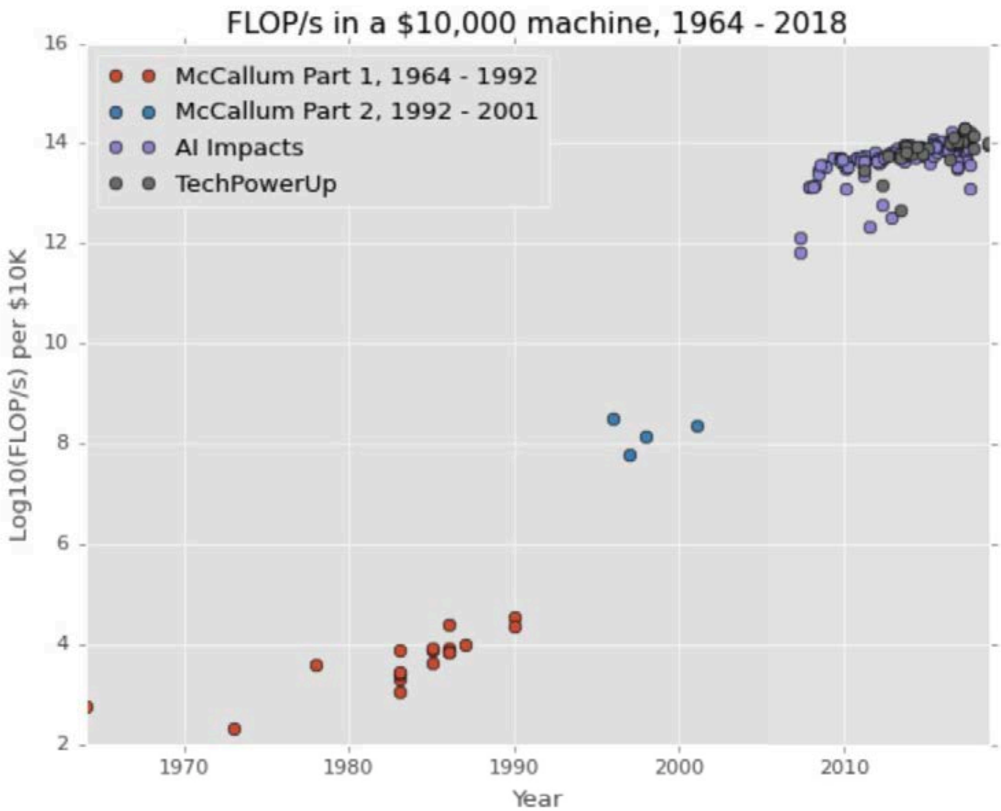
Question and Resolution Details

- Processor capacity will be judged by theoretical max performance, as reported in manufacturer specifications.
    - If disputed, a panel of experts will attempt to verify the accuracy of the performance claim made by the manufacturers.
- Price is the initial retail price of the chip on its release, adjusted to 2021 USD.
    - If the chip is not available retail, then a panel of experts will estimate how much the chip would have cost if available retail using other chips of that type on the market for reference

Historical Base Rate Data

- Wikipedia history of GFLOPS costs from November 2017 (inflation adjusted to 2013): https://aiimpacts.org/wikipedia-history-of-gflops-costs/

- See also the appendix to Cotra 2020 for a summary of processor price performance history:



Prior forecasts

- See the appendix to Cotra 2020 for discussion of medium-term improvements to silicon chips and long-term forecast for hardware prices.

How much will be spent on compute in the largest AI experiment by…

a. …the end of 2024?
b. …the end of 2030?
c. …the end of 2050?

Question and Resolution Details

- Artificial intelligence (AI) is defined here as 'the development of machines capable of sophisticated (intelligent) information processing (Dafoe 2018, *AI Governance: A Research Agenda*; See pg 5 footnote 2 for more.)
- 'Largest' will be in terms of maximum compute used in training, in petaFLOPS-days. The experiment must be a 'single AI experiment', that is, 'an effort to train a set of models running on a set of "architectures" … [which] should be completed within a determinate amount of time (the experiment must not have an open-ended time frame).' (More here.)
- Cost will be in terms of 2021 US dollars.
- For this question, we are interested in the most expensive experiment from between June 15th, 2022 and the resolution year. Experiments performed prior to June 15th, 2022 will not be included.
- Size of the largest experiment will be resolved using a credible estimate of the most compute used in a "single AI experiment."
    - By "single AI experiment" we mean an effort to train a set of models running on a set of "architectures". The effort should be completed within a determinate amount of time (the experiment must not have an open-ended time frame). For our purposes, the publication of the principal results of the effort ends the experiment.
    - By "architectures" we mean the systems described in the relevant publications that define how inputs signal or percept sequences are to be mapped on various outputs. These outputs might be probability distributions over actions (in the case of a policy network), representations over expected value or reward of futures states (in the case of value networks) or descriptions of futures states.
    - In the absence of an authoritative source, the question will be resolved by a panel of experts.
        a. The panel of experts will default to using the methodology used by OpenAI (OpenAI, 2018) unless there is consensus to use a different methodology. OpenAI's method:

i. "When we had enough information, we directly counted the number of FLOPS (adds and multiplies) in the described architecture per training example and multiplied by the total number of forward and backward passes during training."

ii. "When we didn't have enough information to directly count FLOPs, we looked GPU training time and total number of GPUs used and assumed a utilization efficiency (usually 0.33). For the majority of the papers we were able to use the first method, but for a significant minority we relied on the second, and we computed both whenever possible as a consistency check. In the majority of cases we also confirmed with the authors. The calculations are not intended to be precise but we aim to be correct within a factor 2-3."

- Cost resolution details:
    - Afterward, a panel of experts will estimate how much the compute for the experiment cost. Some examples of past estimates can be found here, here, or here.
        - For one example of how the experts could estimate experiment cost: if Google itself carried out the experiment using its own TPUs, this would correspond to a complicated calculation involving the cost of producing the TPUs, the proportion of the lifetime of the TPUs which was used up in the experiments, Google's profit margin for these TPUs, etc. A result might look like "it costs Google 30 to 70% of their retail price to use their own TPUs, which corresponds to a compute cost of $X."
        - For another example, if an external company got a special discount from Microsoft to run their experiments on Microsoft Azure, this number would be an estimate of how much the external company paid Microsoft.
    - Fine-print:
        - 16 or 32-bit floating point operations will be regarded as equivalent.
        - The units for this question are petaflop/s-day per dollar, to keep with historical usage by Metaculus or by OpenAI. A petaflop/s-day (pfs-day) consists of performing 10^{15} neural net operations per second for one day, or a total of about 10^{20} operations

Historical Base Rate Data

Estimated size of largest known experiments by year		
Year	Experiment Size (FLOP)	Model
2020	3×10^{23}	GPT-3
2019	2×10^{23}	AlphaStar
2018	3×10^{21}	BigGAN
2017	2×10^{23}	AlphaGoZero
2016	7×10^{21}	Neural Machine Translation
2015	3×10^{19}	DeepSpeech2
2014	9×10^{18}	VGG
2013	5×10^{17}	Visualizing & Understanding Conv Nets
2012	5×10^{17}	AlexNet

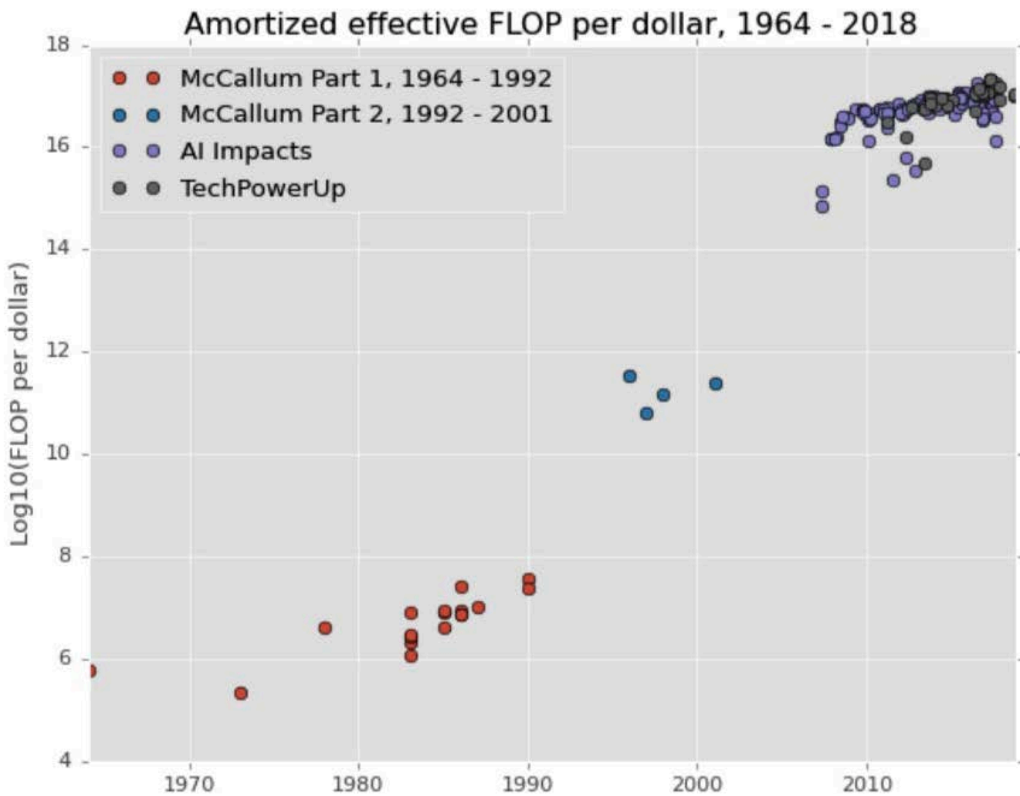
Sources: (I)(II)

Estimates of large experiment compute costs

- 2020 GPT-3: $4.6M (III)
- 2017 AlphaGoZero: $10M (IV) - $35M (V)
-

Amortized effective FLOP per dollar over time

(This figure takes as a starting point either a) the price performance, with performance measured on LINPACK benchmark; or b) the theoretical maximum price performance of a chip, as reported by manufacturer, x0.73. From here, the cost of the chip is amortized over 2 years of continuous use, and then an adjustment is applied to account for additional costs of running a data center and for imperfect utilization, a factor of ⅛x together. See more here.)



Prior Forecasts

- Interpreting AI compute trends
    - "The size of the largest experiments is increasing with a doubling time of 3.5 months, (about an order of magnitude per year), while the cost per unit of computation is decreasing by an order of magnitude every 4-12 years (the long-run trend has improved costs by 10x every 4 years, whereas recent trends have improved costs by 10x every 12 years). So the cost of the largest experiments is increasing by an order of magnitude every 1.1 – 1.4 years. The largest current experiment, AlphaGo Zero, probably cost about $10M. … Previously, the US spent 1% of annual GDP on the Manhattan Project, and ~0.5% of annual GDP on NASA during the Apollo program. So let's suppose they could similarly spend at most 1% of GDP, or $200B, on one AI experiment. Given the growth of one order of magnitude per 1.1-1.4 years, and the initial experiment size of $10M, the AI-Compute trend predicts that we would see a $200B experiment in 5-6 years. So given a broadly similar economic situation to the present one, that would have to mark an end to the AI-Compute trend. … The largest [private actors] are tech companies: Amazon and Google have current research and development budgets of about ~20B/yr each[8], so we can suppose that the largest individual

experiment outside of government is $20B. Then the private sector can keep pace with the AI-Compute trend for around ¾ as long as government, or ~3.5-4.5 years. On the other hand, the development of specialized hardware could cheapen computation, and thereby cause the trend to be sustainable for a longer period."

- Forecasting TAI with biological anchors (part 4)
  - Best guess forecasts for hardware, spending & algorithms includes discussion of FLOP per dollar forecasts, and forecasted amounts that various actors would be willing to spend on training runs

Other Potential Background Information

- Links to potentially helpful resources, Wikipedia, news articles, explainers, etc…
- AI and Compute | OpenAI
- "AI and Compute" trend isn't predictive of what is happening | Alignment Forum
- Trends in the cost of computing – AI Impacts
- AI's Smarts Now Come With a Big Price Tag | WIRED

Question 48

What will be the lowest price, in 2021 US dollars, of 1 GFLOPS with a widely-used processor by…

…the end of 2024?
…the end of 2030?
…the end of 2050?

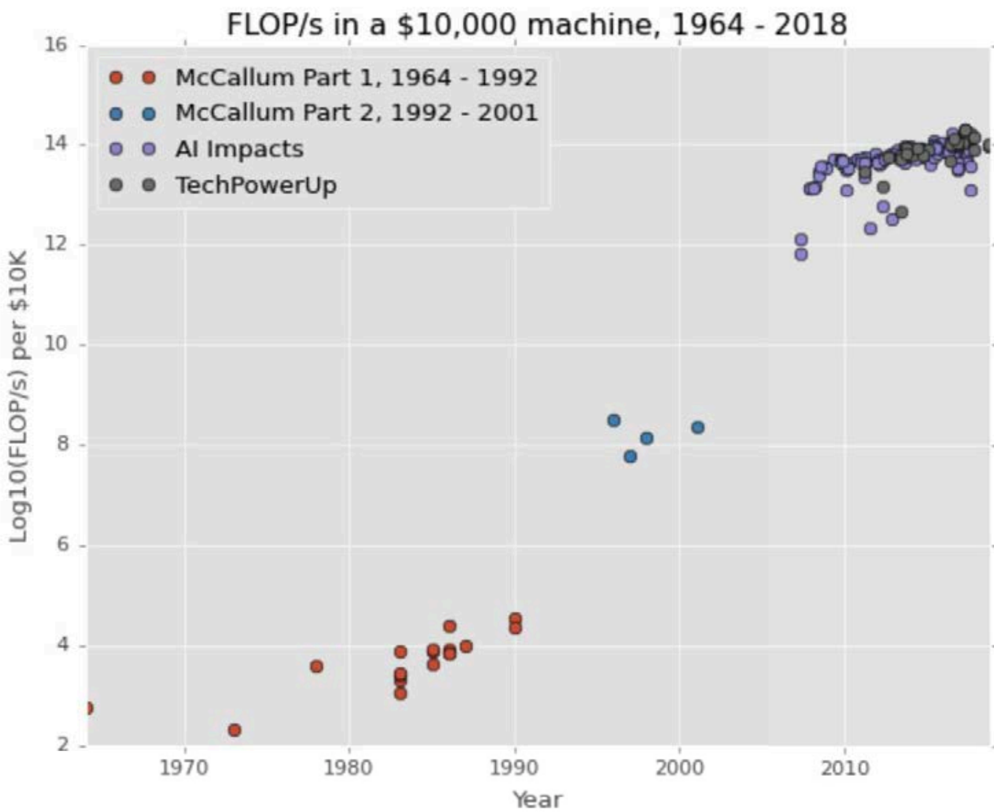
Question and Resolution Details

- Processor capacity will be judged by theoretical max performance, as reported in manufacturer specifications.
  - If disputed, a panel of experts will attempt to verify the accuracy of the performance claim made by the manufacturers.
- Price is the initial retail price of the chip on its release, adjusted to 2021 USD.
  - If the chip is not available retail, then a panel of experts will estimate how much the chip would have cost if available retail using other chips of that type on the market for reference

Historical Base Rate Data

- Wikipedia history of GFLOPS costs from November 2017 (inflation adjusted to 2013): https://aiimpacts.org/wikipedia-history-of-gflops-costs/

- See also the appendix to Cotra 2020 for a summary of processor price performance history:



Prior forecasts

- See the appendix to Cotra 2020 for discussion of medium-term improvements to silicon chips and long-term forecast for hardware prices.

Question 49

What will be the largest number of parameters of a machine learning model trained…

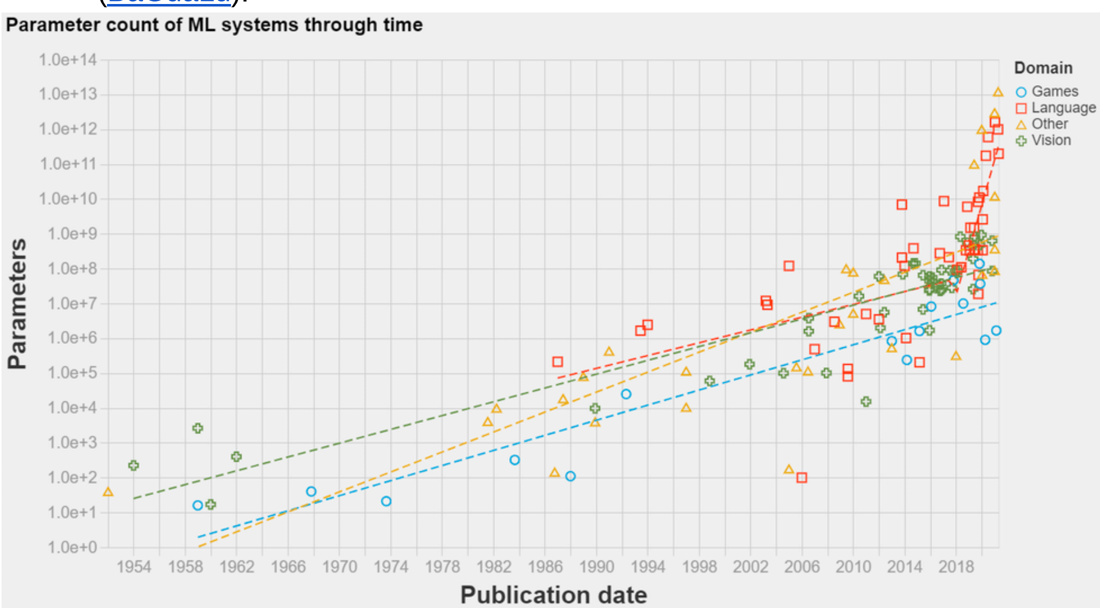
a) …by the end of 2024?
b) …by the end of 2030?
c) …by the end of 2050?

Question and Resolution Details

- In machine learning, model parameters are all variables whose values are learned via training. For example, 'weights' and 'biases' constitute the parameters of a neural network.
- The model must be used to obtain experimental results (e.g., it cannot be merely a description of a possible system, or a demonstration of scaling a system without application to a task); or else it must be deployed in an important context (e.g. Facebook's recommender system).
- This question will resolve as the highest parameter count of a machine learning model reported in a comprehensive and current dataset, such as this (Sevilla & Villalobos 2021), as judged by a domain expert. If no suitable source exists, a panel of experts will decide the resolution.
- Parameter count shall be determined from a published paper or pre-print, or a reputable news article, press release or blog post written by or quoting a researcher involved in the relevant experiments. Where parameter count of a model cannot be determined or the accuracy of the source is ambiguous, it will be disqualified from resolution of this question.

Historical Base Rates

- See Sevilla & Villalobos 2021 for parameter counts of ML systems between 1952 and 2022 (full dataset here). As of 2022, the largest parameter count is 1.5 x 10^{13} (BaGuaLu).



Prior Forecasts

- Metaculus: How many billions of parameters will the largest ML model by 2030 have? Community prediction as of March 2022 is 8.43M.

- Metaculus: Will a 100 trillion parameter deep learning model be trained before 2026? Community prediction as of March 2022 is 95%.

Question 50

Assume that Pew Research re-runs the survey linked here. What % of people in the median country in the survey will say that the development of artificial intelligence has mostly been a bad thing for society…

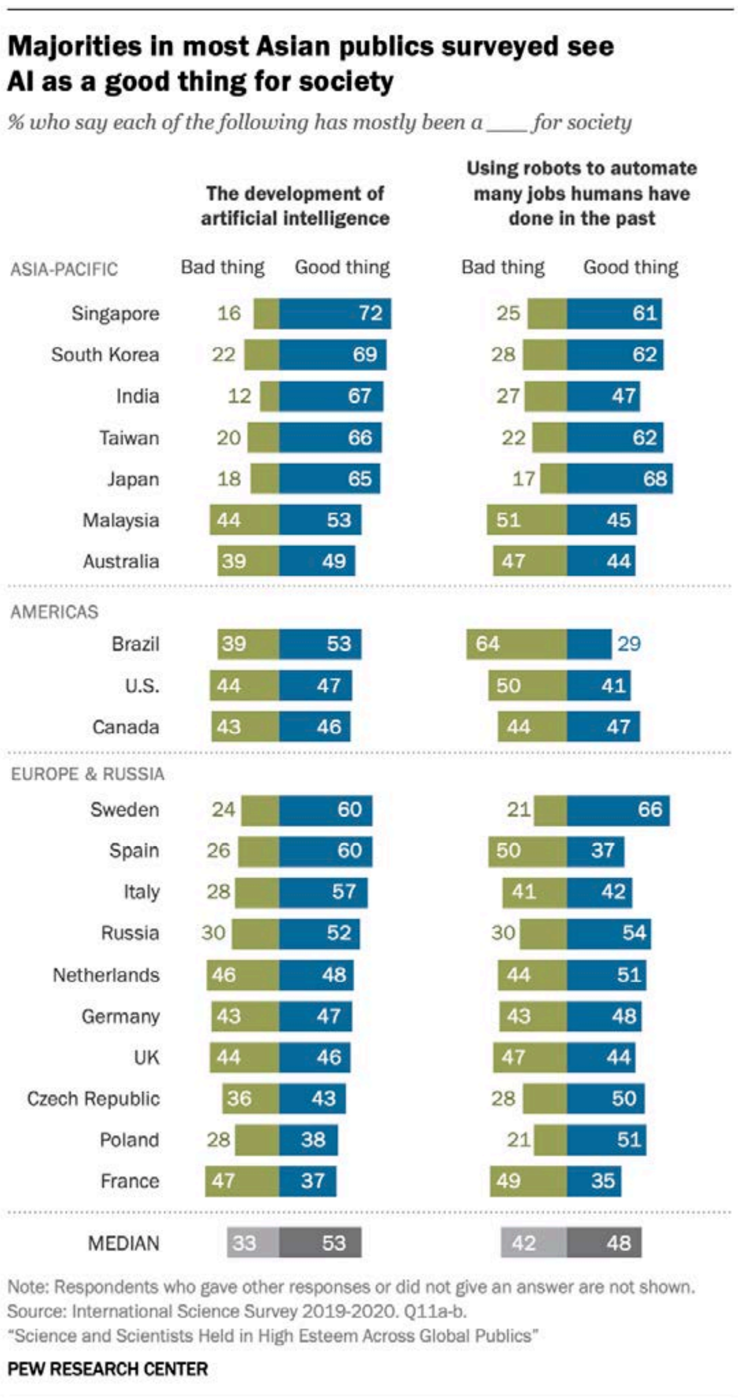
a. ...in 2024?
b. ...in 2030?
c. …in 2050?

Question and Resolution Details

- If Pew Research re-runs this survey in 2024/2030/2050, this question resolves as described.
- If Pew Research does not re-run this survey, our research team will run the same survey with the exact question as specified at the appropriate time.
- Artificial intelligence (AI) is defined here as 'the development of machines capable of sophisticated (intelligent) information processing (Dafoe 2018, *AI Governance: A Research Agenda*; See pg 5 footnote 2 for more.)

Historical Base Rates and Prior Forecasts

2020 version of this survey is available here. See key figure below:



I couldn't find any other forecasts of how public opinion AI will change, nor any base rates for the particular measure given in the survey. These links are the closest to that type of relevant prior information:

- See Brookings' various public opinion surveys on AI.

- For in-depth data on American public opinion on AI in particular, see GovAI survey on American attitudes.

Additional Links and Resources:

- General overview on public opinion and AI from Baobao Zhang: https://na.eventscloud.com/file_uploads/f51adb27e9e9d4ade5b67fbd1e021411_Zhang_Baobao_Oxford_Handbook_Chapter_Public_Opinion_Toward_AI.pdf
- See Brookings' various public opinion surveys on AI
- Useful Oxford Internet Institute report: Global Attitudes Towards AI, Machine Learning & Automated Decision Making
- Further info on the survey about public opinion on science/tech topics, of which AI was part: Publics express a mix of views on AI, childhood vaccines, food and space issues | Pew Research Center
- More context on public opinion about AI and job automation from Pew: Are AI and job automation good for society? Globally, views are mixed | Pew Research Center
    - For all Pew results tagged "Artificial Intelligence," see here
- SQ6. How has public sentiment towards AI evolved, and how should we inform/educate the public? | One Hundred Year Study on Artificial Intelligence (AI100)
- Good summary and review of various Pew results: How public opinion on AI varies around the world - Tech Monitor

Question 51

By each of the following years, will Nick Bostrom believe that artificial general intelligence (AGI) exists?

- 2030
- 2050
- 2100

Question & resolution details:

- Professor Nick Bostrom is a philosopher and Director of Oxford's Future of Humanity Institute. He is widely considered one of the foremost contemporary thinkers on the future of AI. A wide-ranging discussion of human- and above-human-level machine intelligence can be found in his 2014 book, *Superintelligence*.
- This question may be directly resolved by private communication in the specified years between Nick Bostrom and the tournament organizers. If Nick Bostrom does not directly resolve the question, it may be resolved by the following:
    - By the beginning of the specified year, Nick Bostrom has publicly stated that he believes "artificial general intelligence" or "AGI" (or any reasonable variant phrase

containing the word "general") exists (or any reasonable variant of the word "exists"). It may also resolve positively if he states that "superintelligence" exists, or that AI capabilities have surpassed human-level, provided it is a general statement about AI capabilities and not restricted to a narrow subset (to be allowed at the discretion of a panel of experts, composed as specified below). Conversely, it may resolve negatively if he states that AGI does not exist. A negative statement will only count toward resolution if it occurs in the year prior to the resolution date, and is not followed by a positive statement. Statements must be documented by a credible journalistic or academic source, or a source directly connected to Nick Bostrom (e.g. his website).

- If Nick Bostrom has not made a public statement to this effect, a >=50% majority vote by a panel of at least 5 experts who will attempt to answer the question from the perspective of Nick Bostrom would also resolve the question. Relevant experts may be present or former colleagues/collaborators of Nick Bostrom, scholars of Nick Bostrom, or researchers with a deep knowledge of Nick Bostrom's work.
- In the event that Nick Bostrom has not made a public statement to this effect and in the event of human extinction, assume a hypothetical panel of current-day human experts resolved the question based on information available in the year of interest and their beliefs about whether Nick Bostrom would affirm the existence of AGI based on that information.

Context

- See *Superintelligence*, esp Chapter 1, 'Growth modes and big history'

Question 52

By 2100, will the global real GDP in a year ever exceed 115% of the highest GDP reported in any full prior year?

Question and Resolution Details

a. 'All prior years' includes only dates from 1960 onward.
b. 'By 2100' includes the year 2100.
c. Similar to this Metaculus question, "Economic growth is [usually] measured by real world GDP growth over one year. However, since economic growth can also accelerate following a recession, this question will not ask about economic growth per se, but instead about GDP levels reaching a new height."
d. Question may be resolved by using World Bank data on global GDP, which can currently be found on their website in several formats (see e.g. this chart)
If it is not possible to find World Bank data either on their website or elsewhere, this question will be resolved by a panel of experts.

e. For this question to resolve positively, the World Bank would have to measure the real GDP during some year between now and 2100 at 115% of the previously highest-reported value. The question aims to capture an extreme level of annual real GDP growth.
    i. For example, the World Bank's highest-reported GDP in constant 2015 U.S. dollars thus far (see the chart linked in d.) was 84.612 trillion in 2019. If the World Bank reports that the GDP in constant 2015 U.S. dollars for the year 2021 was over 97.3038 trillion (115% of 84.612 trillion), the question would be resolved in the positive.
    ii. The question would also resolve in the positive if the above 2019 figure is *never* exceeded *until* a year where the World Bank reports that the GDP in constant 2015 U.S. dollars was *over* 97.3038 trillion (115% of 84.612 trillion).
    iii. The question would *not necessarily* resolve in the positive if the World Bank ever reports a GDP in constant 2015 U.S. dollars over 97.3038 trillion – only if said figure occurs *after* no prior year's GDP in constant 2015 U.S. dollars had exceeded the previous high of 84.612 trillion.

f. The total GWP in this question includes production on other planets, but excludes production by AI not under the control of humans.

g. In the event of human extinction, assume one human of working age exists, and that they produce one final good of minimal (but non-zero value) that can be given a market value during each relevant year.

Historical Base Rate Data

- GDP growth, annual %, by year; GDP in constant 2015 USD, by year (World Bank)
- Historical economic growth trends (AI Impacts)
    - For the raw historical data from AI Impacts and Bradford DeLong, see here
- Modeling the Human Trajectory (Open Philanthropy)

Prior Forecasts

- When will economic growth accelerate? (Metaculus)
    - Forecast of the first year when world real GDP will exceed *130%* of all prior years.
- Maximum GDP growth rate in the 21st century (Metaculus)
- GWP in 2047, in trillions of USD (Metaculus) & GWP in 2100, in trillions of USD (Metaculus)
- Report on Whether AI Could Drive Explosive Economic Growth (Open Philanthropy, 2021)
    - 'Overall, I place at least 10% probability on advanced AI driving explosive growth [30% annually] this century. Roughly speaking, this corresponds to > 30% probability that human-level AI is developed in time for growth to ramp up to

30% by 2100, and > 1/3 that explosive growth actually happens conditional upon human-level AI being developed.'

Other Potential Background Information

- Long-term Growth As A Sequence of Exponential Modes (Hanson 2000)
- Economic growth under transformative AI (Trammell & Korinek 2020)
- Economic Growth Given Machine Intelligence (Hanson 2001)
- Artificial Intelligence and Economic Growth (Aghion, Jones & Jones) - contrary perspective

Question 53

If the global real GDP in a year does exceed 115% of the highest GDP reported in any prior year by 2100, in what year will this first occur?

Question and Resolution Details

a. 'All prior years' includes only dates from 1960 onward.
b. 'By 2100' includes the year 2100.
c. Similar to this Metaculus question, "Economic growth is [usually] measured by real world GDP growth over one year. However, since economic growth can also accelerate following a recession, this question will not ask about economic growth per se, but instead about GDP levels reaching a new height."
d. Question may be resolved by using World Bank data on global GDP, which can currently be found on their website in several formats (see e.g. this chart)
If it is not possible to find World Bank data either on their website or elsewhere, this question will be resolved by a panel of experts.
e. The total GWP in this question includes production on other planets, but excludes production by AI not under the control of humans.

Historical Base Rate Data

- GDP growth, annual %, by year; GDP in constant 2015 USD, by year (World Bank)
- Historical economic growth trends (AI Impacts)
    - For the raw historical data from AI Impacts and Bradford DeLong, see here
- Modeling the Human Trajectory (Open Philanthropy)

Prior Forecasts

- When will economic growth accelerate? (Metaculus)
    - Forecast of the first year when world real GDP will exceed *130%* of all prior years.

- Maximum GDP growth rate in the 21st century (Metaculus)
- GWP in 2047, in trillions of USD (Metaculus) & GWP in 2100, in trillions of USD (Metaculus)
- Report on Whether AI Could Drive Explosive Economic Growth (Open Philanthropy, 2021)
  - 'Overall, I place at least 10% probability on advanced AI driving explosive growth [30% annually] this century. Roughly speaking, this corresponds to > 30% probability that human-level AI is developed in time for growth to ramp up to 30% by 2100, and > 1/3 that explosive growth actually happens conditional upon human-level AI being developed.'

Other Potential Background Information

- Long-term Growth As A Sequence of Exponential Modes (Hanson 2000)
- Economic growth under transformative AI (Trammell & Korinek 2020)
- Economic Growth Given Machine Intelligence (Hanson 2001)
- Artificial Intelligence and Economic Growth (Aghion, Jones & Jones) - contrary perspective

Question 54

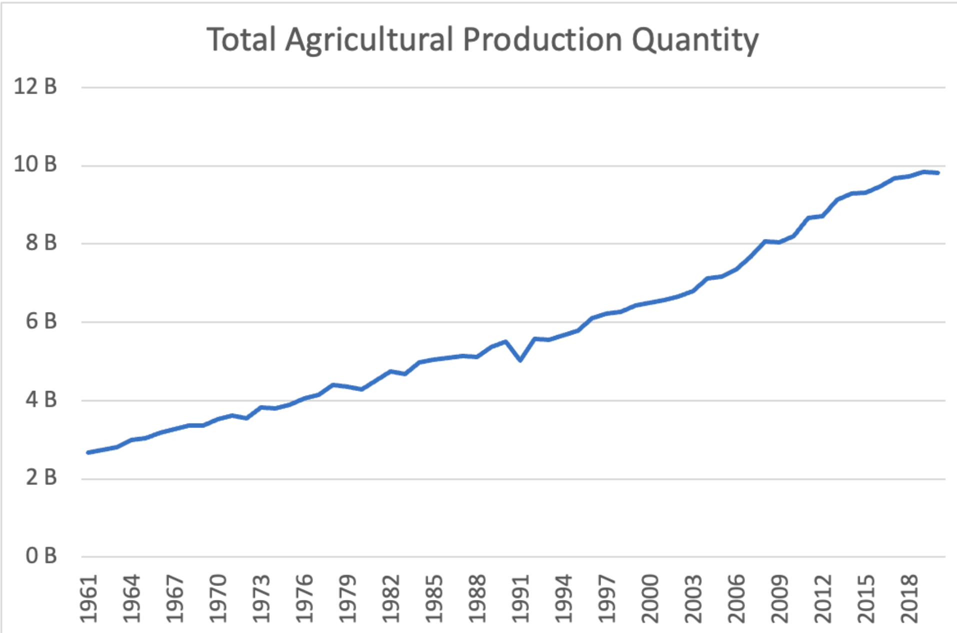
What is the probability that total worldwide production of primary crops will fall by at least 70% within any three-year period…

  - …by 2030?
  - …by 2050?
  - …by 2100?

Question details and Resolution criteria

- Primary crops are defined by the UN Food and Agriculture Organization (FAO) as "those which come directly from the land and without having undergone any real processing, apart from cleaning." More detail can be found here.
- Resolution will be determined using statistics from the UN Food and Agriculture Organization (FAO)'s annual reports or, if these reports are not available, by a panel of experts.
  - For example, in 2019: "When it comes to production, total production of primary crops increased by 53 percent between 2000 and 2019, hitting a record high of 9.4 billion tonnes in 2019. Half of global primary crop production is made up of just four crops: sugar cane, maize, wheat and rice." FAO

- Our resolution data will aim to approximate the data used in the below time series as closely as possible:[151]



(Source: [FAOSTAT](#))

Historical Base Rate Data

- As noted above, the UN Food and Agriculture Organization's annual reports track total agricultural production in tons. The FAO's yearly 'Statistical Yearbooks' have been published since 2020 and can be found here: https://www.fao.org/publications/search/en/?serialtitle=RkFPIFN0YXRpc3RpY2FsIFllYXJib29rIOKAkyBXb3JsZCBGb29kIGFuZCBBZ3JpY3VsdHVyZQ==.
- The FAO also has a data tool, FAOSTAT, that has agricultural data running back through 1961. One can download this data in .CSV and .XLS formats : https://www.fao.org/faostat/en/#data/QCL.
- Other organizations have also tracked world agricultural production.
    - The USDA releases monthly reports on global agricultural production that one can find here: https://usda.library.cornell.edu/concern/publications/5q47rn72z?locale=en.

---

[151] This data was constructed as follows using the FAOSTAT tool:
1. In the top left box, select "Regions" and then "World + (Total)."
2. In the top right box, select "Production Quantity."
3. In the bottom left box, select "Crops Primary > (List)."
4. In the bottom right box, select all years.
5. See output/filetype options, and click "Download Data."

- The OECD also tracks crop production, partially relying on FAO data: https://data.oecd.org/agroutput/crop-production.htm.

Prior Forecasts

- The FAO has suggested that food production will need to expand "by some 70 percent between 2005/07 and 2050" to meet projected global demand: https://www.fao.org/fileadmin/templates/wsfs/docs/Issues_papers/HLEF2050_Global_Agriculture.pdf.
- The OECD and FAO have projected that agricultural production will "increase slightly faster" than 15% between 2019 and 2028: https://www.oecd.org/agriculture/oecd-fao-agricultural-outlook-2019/.

Other Potential Background Information

- This Our World in Data report, summarizing FAO data, may help clarify some aspects of current global agricultural production: https://ourworldindata.org/agricultural-production.

Question 55

What is the probability that there will be a life-sustaining colony outside of Earth's atmosphere with a population of at least 5,000 people…

a. ... by 2030?
b. ... by 2050?
c. ... by 2100?

Question and Resolution Details

- To be considered "life-sustaining," this colony would need to be able to sustain its population over multiple generations even if there were no humans alive on Earth.
- If resolution is ambiguous, this question will be resolved by a panel of experts.

Prior Forecasts

- Metaculus currently forecasts a 72% probability of a sustainable human presence off-Earth by 2100: https://www.metaculus.com/questions/1432/will-humans-have-a-sustainable-off-world-presence-by-2100/.
- 70% of *Futurism* readers predicted that humans would establish a space colony in the first half of the 21st century, and 36% believed that humans would establish the first space colony at some point in the 2030s. The linked article contains predictions from

experts, as well: https://futurism.com/when-will-the-first-human-space-colony-be-established.

- Serkan Saydam predicted in March 2021 that humans would build a Mars colony by 2050: https://scitechdaily.com/mars-settlement-likely-by-2050-says-expert-but-not-at-levels-predicted-by-elon-musk/.

Other Potential Background Information

- The Wikipedia entry on Space colonization may be helpful for some forecasters: https://en.wikipedia.org/wiki/Space_colonization.
- NASA has compiled a suggested reading list on space colonization: https://www.nasa.gov/centers/hq/library/find/bibliographies/space_colonization.

Question 56

In a nationally representative survey, what percentage of Americans will report being "very" or "fairly" happy in response to the question, "Generally speaking, how happy would you say you are — very happy, fairly happy or not too happy?"…

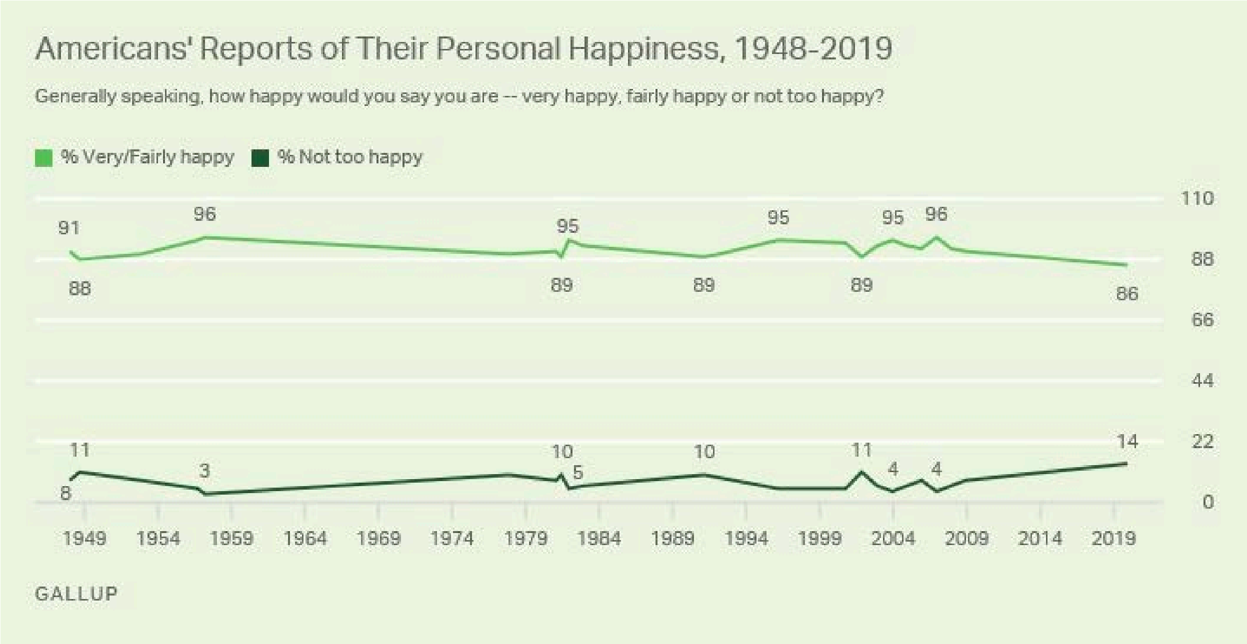
a. ...in 2030?
b. ...in 2050?
c. ...in 2100?

Question and Resolution Details

- This question resolves as the % of Americans who report being “very” or “fairly” happy in the Gallup poll about happiness and wellbeing, if the survey is run in the resolution year.
- If the Gallup poll is not run in the resolution year, a new survey will be run with the exact question as specified and will resolve accordingly.
- If it is impossible to run this survey because independent survey companies are not able to survey Americans (for example, because public opinion surveying is not allowed), this question will be considered to resolve as 0.

Historical Base Rates and Prior Forecasts

Historical data for the Gallup poll referenced in the resolution criteria: Happiness Not Quite as Widespread as Usual in the US



Additional Resources

- Our World in Data's Happiness and Life Satisfaction page
- Pew Research Center's articles and surveys on happiness and life satisfaction
- The World Happiness Report has multiple relevant studies and data sources
    - See in particular, on trends in American happiness: The Sad State of Happiness in the United States and the Role of Digital Media | The World Happiness Report
- Americans are getting more miserable, and there's data to prove it - The Washington Post

Question 57

What percentage of the world population will be classified as living in an electoral or closed autocracy in the V-Dem Institute's annual Democracy Report…

a. …for the year 2030?
b. …for the year 2050?
c. …for the year 2100?

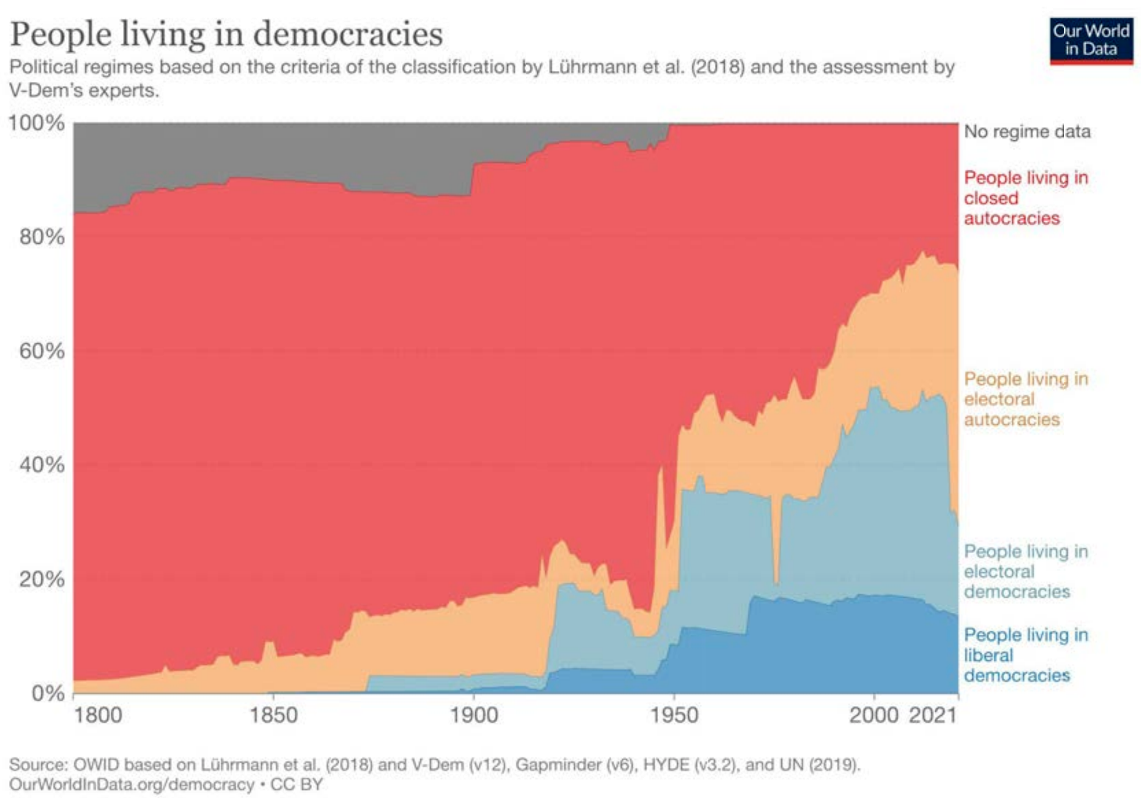
Questions and Resolution Details

- In the yearly *Democracy Report*, the V-Dem Institute measures a span of indices corresponding to democratic values and principles on a country-by-country basis. Based on a regime classification paper published by Lührmann et al. in 2018, one can also classify each country as a “liberal democracy,” “electoral democracy,” “electoral autocracy,” or “closed autocracy.”

- If the *Democracy Report* is not published in one or more of the resolution years, we will attempt to resolve the question using V-Dem and Lührmann et al.’s methodologies, or we will rely on another non-state organization’s report using similar metrics. In the latter case, a panel of experts will select the replacement report and resolve the question given data from the new survey.
- If it is impossible to gather data from independent, non-state sources on democratic principles anywhere on the globe, this question will be considered resolved as 100%.

Historical Base Rate Data

- Details of V-Dem’s work can be found on Wikipedia.
- *Our World in Data* has compiled data on the world’s population as categorized by its regime classification. For the *Democracy Report* published in 2022 (on the state of global democracy in the year 2021), 70.55% of the global population was living under a regime classified as an electoral or closed autocracy. See a graphic below showing each category’s trend over time.



-
    - Source: *Our World in Data*

Prior Forecasts

- The Metaculus community has estimated that ~86 countries would be designated 'Free' in the 2028 Freedom House *Freedom in the World* report. Assuming the 195 total states/territories in the 2021 report remain constant, this would translate to ~44% of all states/territories designated 'Free' and ~56% designated 'Partly Free' or 'Not Free.'
  - Other, similar, estimates have been made by the Metaculus community, including the United States', Russia's, China's, and Poland's future Freedom House scores (or scores from similar surveys).

Other Potential Background Information

- V-Dem's website and datasets related to V-Dem.
- See other indices of freedom on Wikipedia.

Question 58

When will 50% of US adults (18-49) say they expect to have no children, or no *more* children, and cite a worry about the long-term future of the world or country as a primary reason?

Question & resolution details

- Worries about the long-term future of the world include: global instability; international or domestic politics; climate change/environment; worries about technology. Additional categories may be included later at the discretion of a panel of experts. The cited long-term future worry does not have to be the exclusive worry of a respondent in order to count towards the total.
- This question will resolve as the first year in which a nationally representative survey is conducted in the United States of adults age 18+, and a total proportion of 50% respondents:
  - Have no children AND expect to have no children AND cite at least one worry about the long-term future of the world or country as a reason; OR
  - Have children AND expect not to have more children AND cite at least one worry about the long-term future of the world or country as a reason.
- If at any point before 2100 at least 10 years has elapsed since the last survey of this kind has been run, we will commission such a survey.
- A person will count as 'expecting to have no children' if they are currently childless, and: rate their likelihood of having children as 'not at all likely' or 'not too likely'; or say they 'don't want' children or 'probably don't want' children; or any other reasonable wording variant (to be judged by a panel of experts).

  - A person will count as 'expecting not to have more children' if they currently have children, and: rate their likelihood of having children as 'not at all likely' or 'not too likely'; or say they 'don't want' children or 'probably don't want' children; or any other reasonable wording variant (to be judged by a panel of experts in cases of ambiguity).

Historical base rate data

- In a 2021 Pew survey, ~2% of adults aged 18-49 said they would not have any children, or would not have more children, and cited 'state of the world' or 'climate change/environment' as a reason. 9% of childless adults cited 'state of the world' as a reason they were unlikely to have children, and 5% cited 'climate change/environment'.
  - 61% of the sample planned to have no children or no more children. 62% of those said they 'just didn't want (more/any) children'.
  - In **2018**, 37% of people *without* children did not think it was likely they would have any, while in **2021** it was 44%. Of those without children who didn't expect to have any, in **2018** 63% said they 'just didn't want' children, while in **2021** it was 57%.
  - In **2018**, 71% of people *with* children did not think it was likely they would have more, while in 2021 it was 74%. Of those with children who didn't expect to have more, in **2018** 57% said they 'just didn't want' more children, while in **2021** it was 63%.
- In a 2020 Morning Consult survey, 14% of adults aged 18-45 *without children* cited 'concern about climate change' as a major reason they did not currently have children, while 20% cited 'concern about the political and economic climate' as a major reason.
  - 50% of the sample aged 18-45 did not plan to have a child in the future.
  - 28% of those without children cited 'no desire for children' as a major reason.
  - In 2018, for adults that didn't want children or weren't sure, 18% cited 'global instability' as a reason; 14% cited 'worried about population growth'; 11% cited 'worried about climate change'; and 10% cited 'worried about domestic politics'.

Prior forecasts

- US birth rates:
  - https://www.brookings.edu/blog/up-front/2021/05/24/will-births-in-the-us-rebound-probably-not/
  - https://www.metaculus.com/questions/6813/fertility-rate-be-in-usa-for-the-year-2025/

Other potential background information

- National fertility survey series (1965-1975) - Includes women's opinions on future childbearing intentions
  - 63% said they 'just didn't want' children, while in **2021** it was 57%.

    - In **2018**, 71% of people *with* children did not think it was likely they would have more, while in 2021 it was 74%. Of those with children who didn't expect to have more, in **2018** 57% said they 'just didn't want' more children, while in **2021** it was 63%.
- In a 2020 Morning Consult survey, 14% of adults aged 18-45 *without children* cited 'concern about climate change' as a major reason they did not currently have children, while 20% cited 'concern about the political and economic climate' as a major reason.
    - 50% of the sample aged 18-45 did not plan to have a child in the future.
    - 28% of those without children cited 'no desire for children' as a major reason.
    - In 2018, for adults that didn't want children or weren't sure, 18% cited 'global instability' as a reason; 14% cited 'worried about population growth'; 11% cited 'worried about climate change'; and 10% cited 'worried about domestic politics'.

Question 59

When will 90% of Americans say they wish they had been born in a previous generation?

Question & resolution details

- This question will resolve positively if a nationally representative survey is conducted in the United States asking respondents any of the following, AND at least 90% of respondents answer positively:
    - Whether they wish they had been born or lived in a *specific* or *nonspecific* earlier time period;
    - Whether they wish they had been part of an earlier generation;
    - Or any reasonable variant wordings of these questions. If there is ambiguity, it will be resolved by a panel of experts.
- If at any point between 2025 and 2100 at least 10 years has elapsed since the last survey of this kind has been run, we will commission such a survey.

Historical base rate data

- According to this report from 2017, 48% of Britons believe that millennials will have a worse life than their parents, and 33% of millennials would prefer to have grown up at the time when their parents were children.

Other potential background information

- See Deloitte Millennial Survey for attitudes about the future across various generations.

Appendix 7: Question summaries

This appendix contains rationale summaries drafted by the Forecasting Research Institute (FRI). For each XPT question, the summaries list the question text and present key results. The summaries also briefly present sources of agreement, disagreement, and uncertainty among forecasters, major arguments given to support either higher or lower forecasts, and cross-references to related questions.

Question 1: Genetically-Engineered Pathogen Risk

What is the probability that a genetically-engineered pathogen will be the cause of death, within a 5-year period, for more than 1% of humans alive at the beginning of that period…
...by the end of 2030?
...by the end of 2050?
...by the end of 2100?

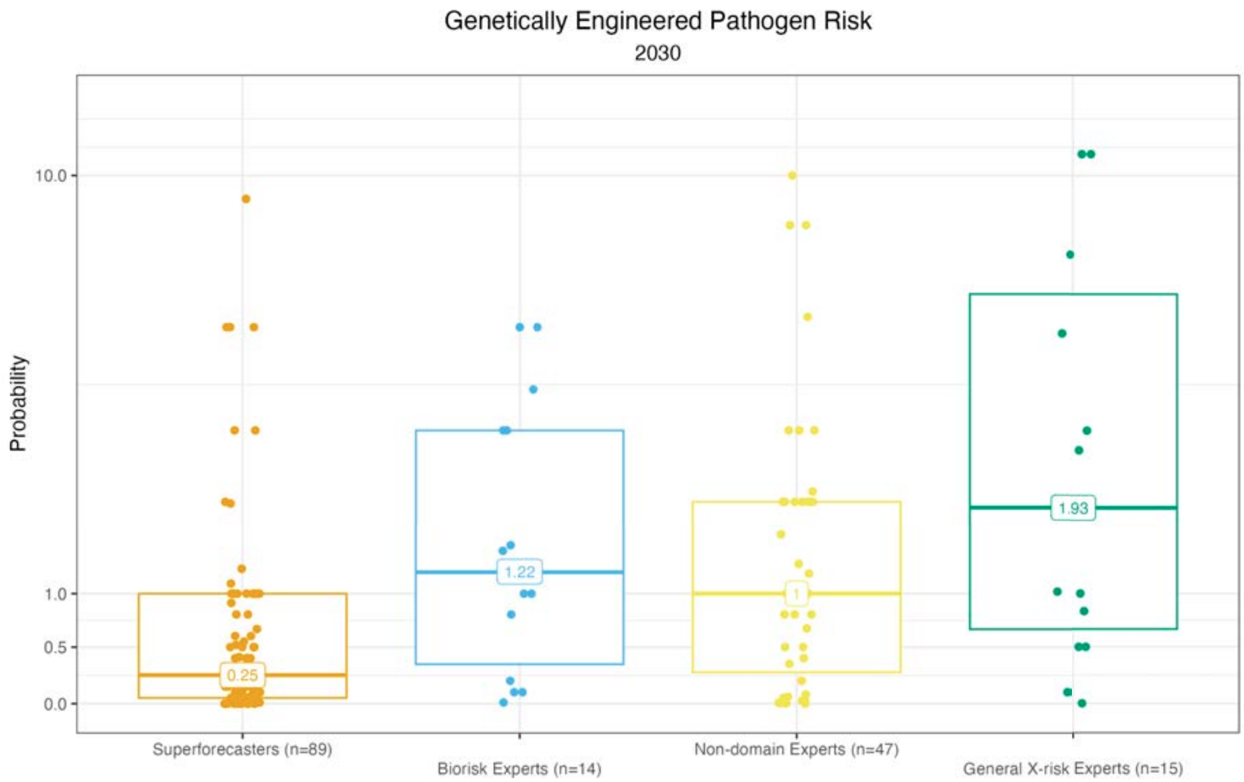
Question and resolution details, prior forecasts, and other relevant sources

Results[152]

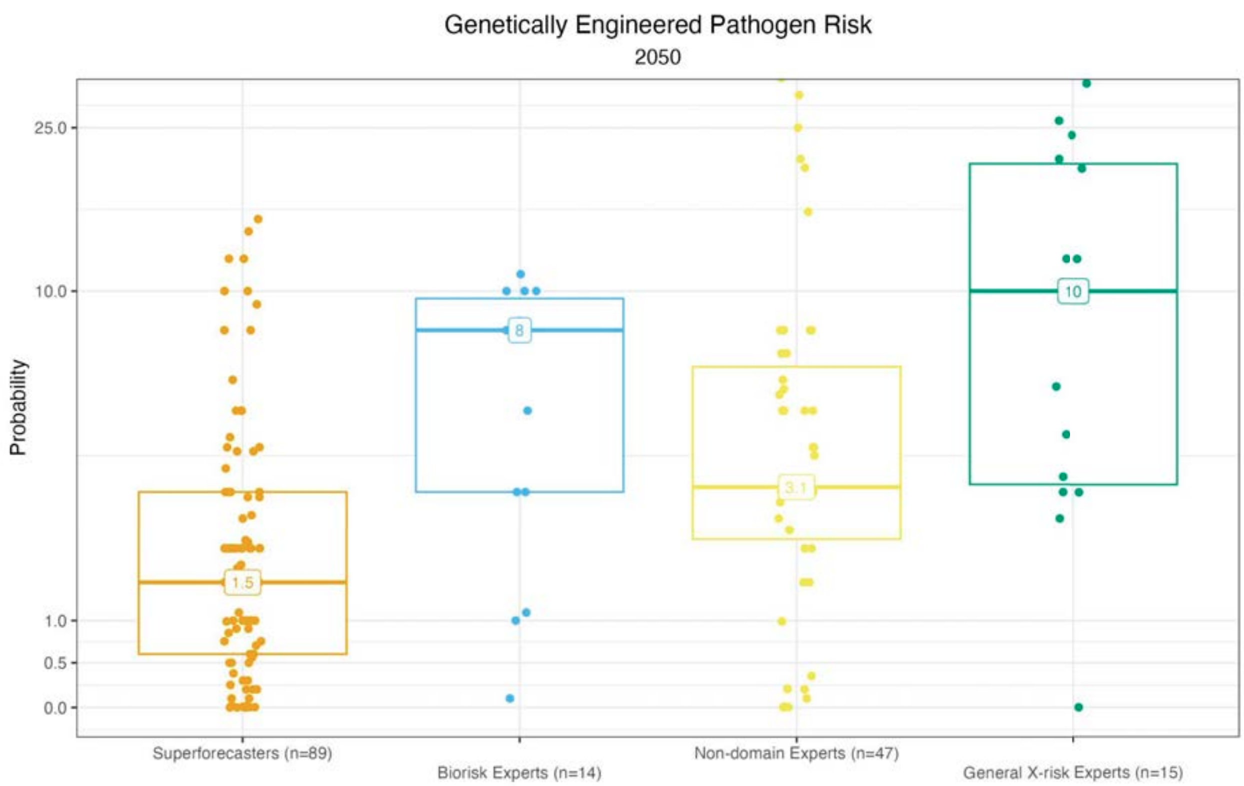
Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 89)	2030	0.4%	0.25%	8.58	-66.1%
	2050	1%	1.5%	7.67	-58.23%
	2100	4%	4%	17.62	-60.34%

[152] Numbers of forecasters are given as of Stage 4 of the XPT.

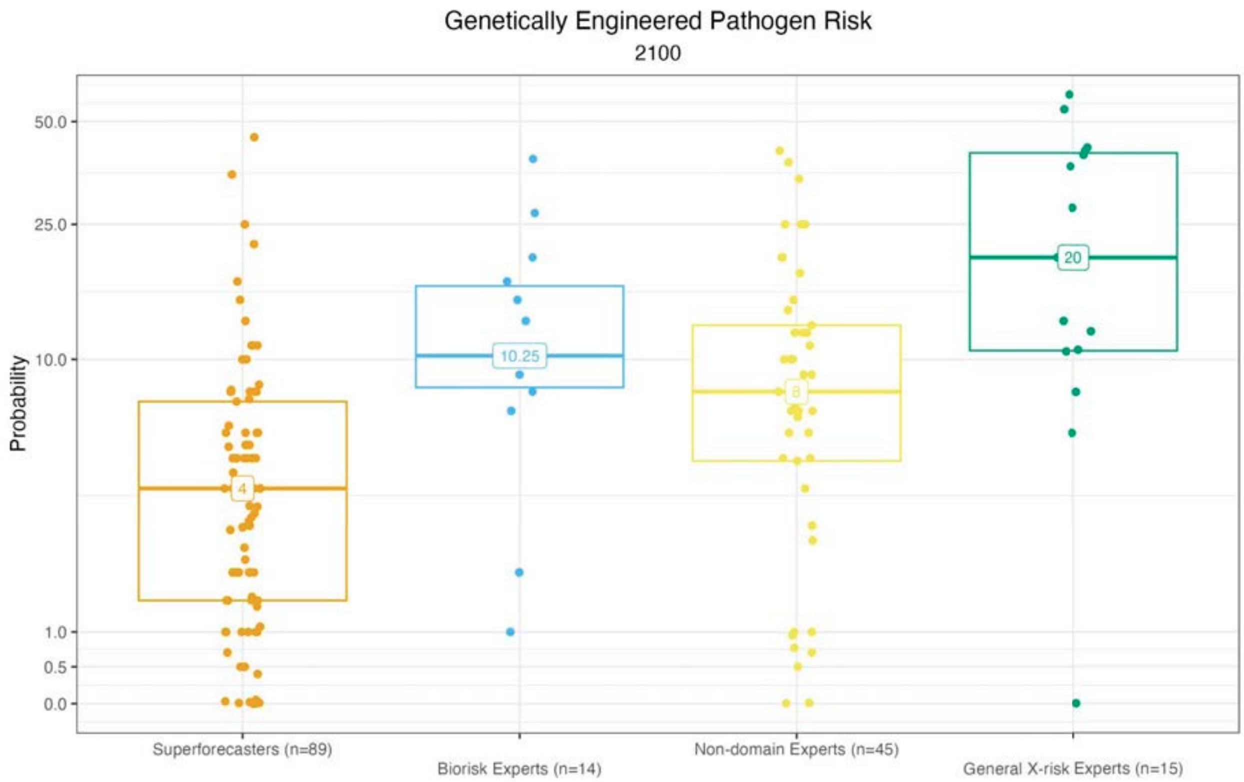
Domain Experts (N = 14)	2030	1%	1.22%	1.85	-4.1%
	2050	6.5%	8%	9.27	-58.39%
	2100	10%	10.25%	18.86	-46.99%
General X-Risk Experts (N = 15)	2030	2.7%	1.93%	6.24	-8.93%
	2050	12%	10%	14.72	-30.71%
	2100	28%	20%	25.48	-25.64%
Non-Domain Experts (N = 47)	2030	1%	1%	3.05	-29.36%
	2050	4%	3.1%	8.29	-8.13%
	2100	7.45%	8%	12.05	-17.87%

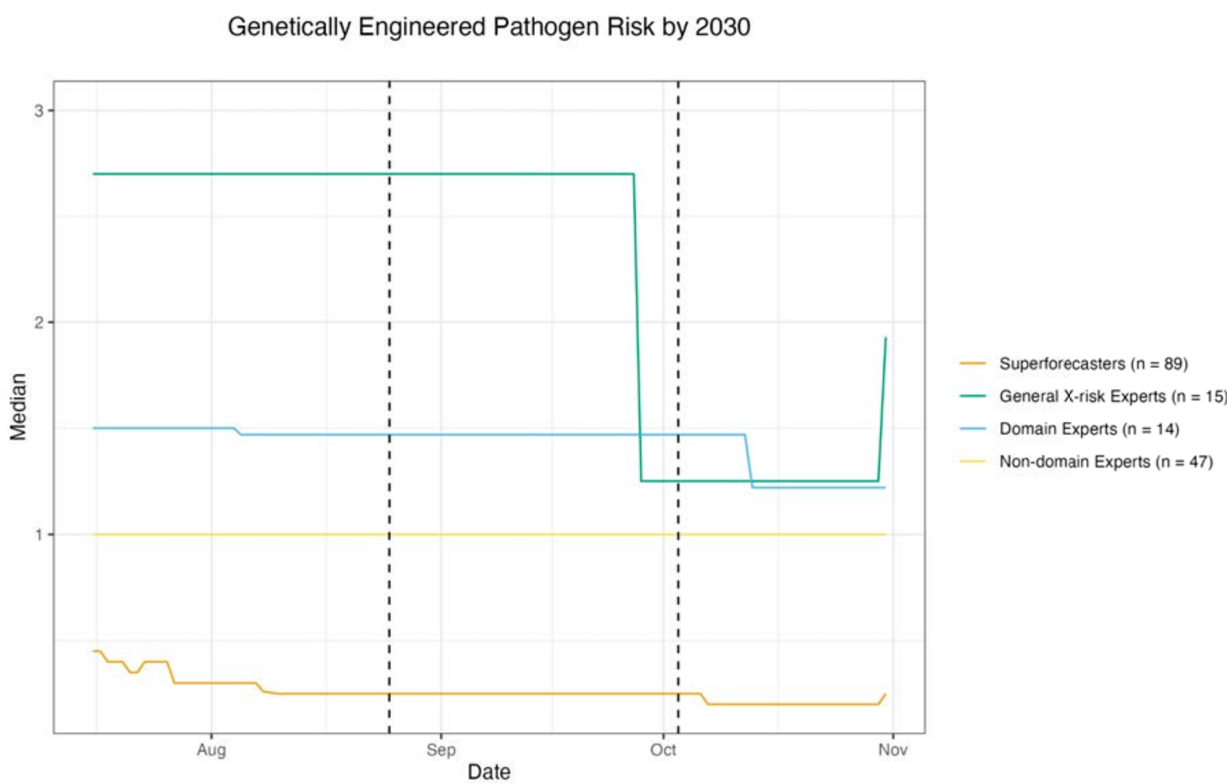


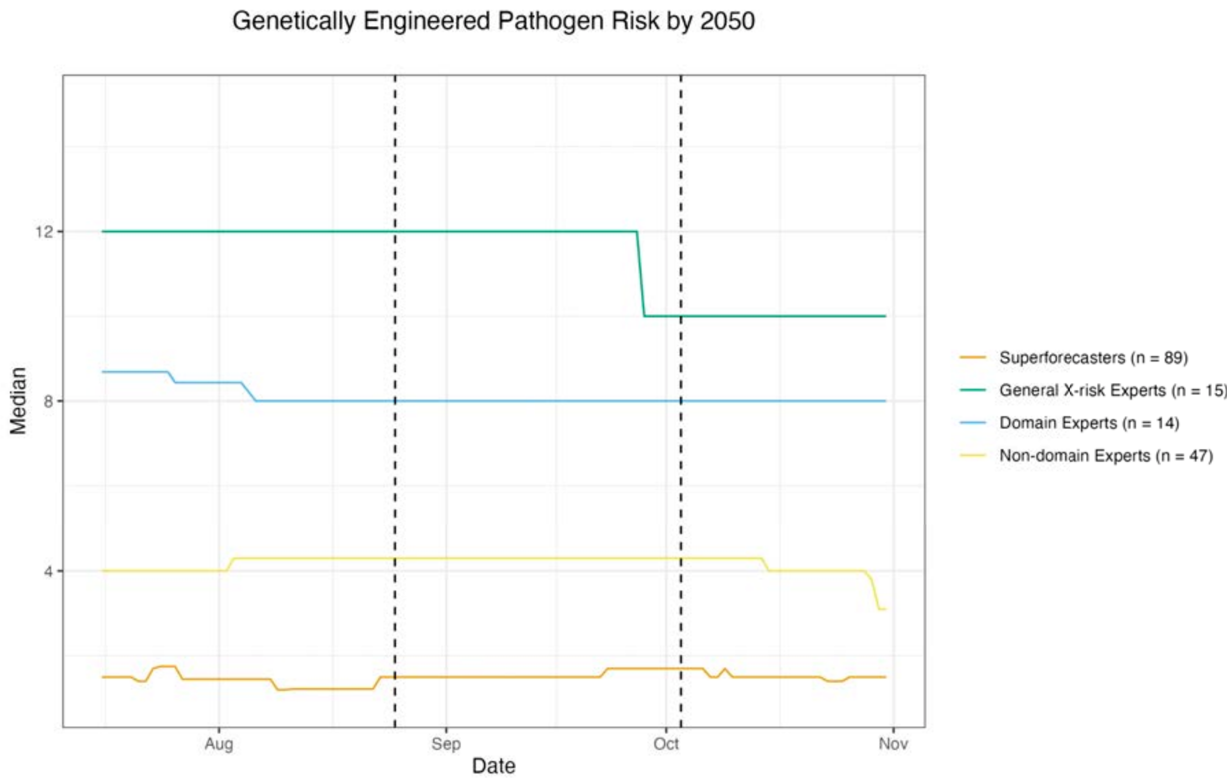
Genetically Engineered Pathogen Risk
2030
Probability
10.0
1.0
0.5
0.0
0.25
1.22
1
1.93
Superforecasters (n=89)
Biorisk Experts (n=14)
Non-domain Experts (n=47)
General X-risk Experts (n=15)

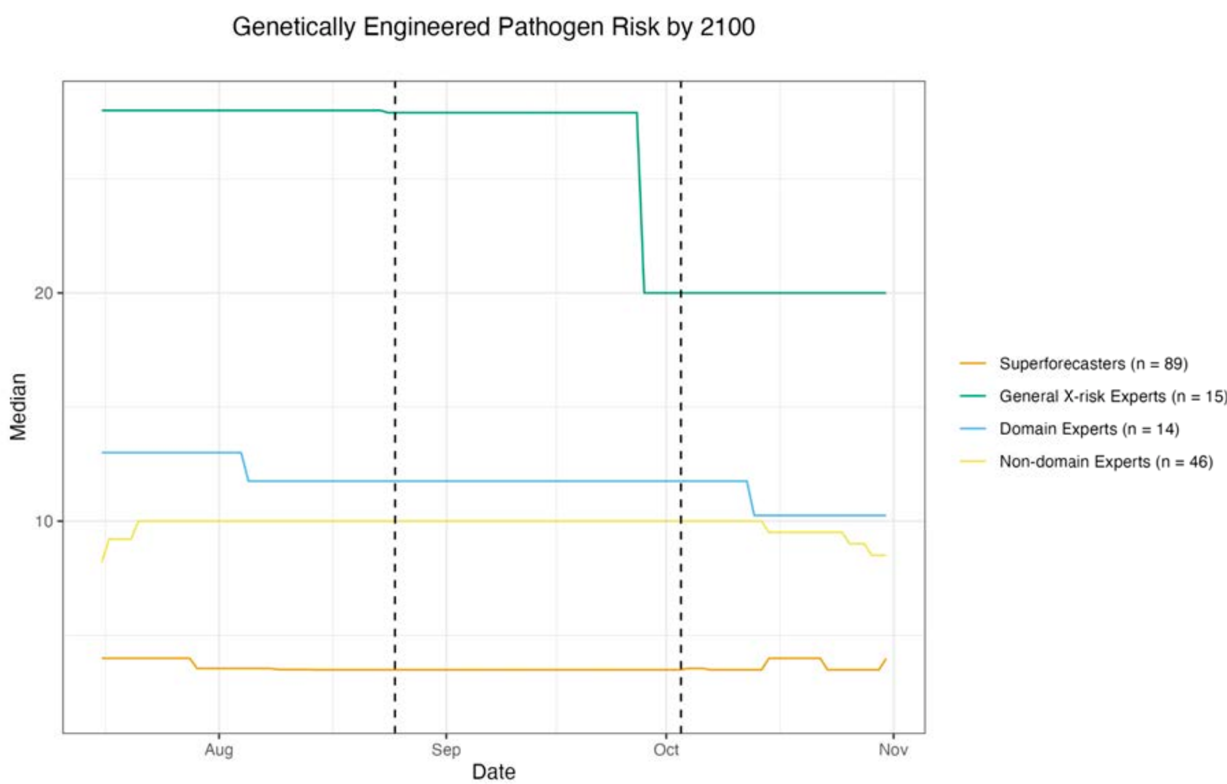


Genetically Engineered Pathogen Risk
2050
Probability
25.0
10.0
1.0
0.5
0.0
1.5
8
3.1
10
Superforecasters (n=89)
Biorisk Experts (n=14)
Non-domain Experts (n=47)
General X-risk Experts (n=15)


Genetically Engineered Pathogen Risk
2100
Probability
50.0
25.0
10.0
1.0
0.5
0.0
4
10.25
8
20
Superforecasters (n=89)
Biorisk Experts (n=14)
Non-domain Experts (n=45)
General X-risk Experts (n=15)


Genetically Engineered Pathogen Risk by 2030
Median
3
2
1
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 89)
General X-risk Experts (n = 15)
Domain Experts (n = 14)
Non-domain Experts (n = 47)


Genetically Engineered Pathogen Risk by 2050
12
8
4
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 89)
General X-risk Experts (n = 15)
Domain Experts (n = 14)
Non-domain Experts (n = 47)


Genetically Engineered Pathogen Risk by 2100
20
10
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 89)
General X-risk Experts (n = 15)
Domain Experts (n = 14)
Non-domain Experts (n = 46)

Sources of agreement, disagreement and uncertainty

There was disagreement on this question on:

1. Base rates
2. Creation (ease and likelihood) of genetically-engineered pathogens (GEPs)
    a. Whether technological developments will decrease risk (through increased surveillance) or increase risk (by decreasing barriers to creating GEPs)
3. Intentional distribution of GEPs
4. Containment of GEPs

Additionally, there was disagreement on the following, though only represented on the higher-forecast side in ‘other arguments’:

5. Gain-of-function curtailment

There were also several arguments on either side not addressed by the other side.

Arguments given for forecasts ≤ 6.3%

On 1(base rates):

- Base rate of a natural pathogen killing more than 1% of humanity is low: about every 500 years.[153]

On 2 (creation of GEPs):

- Regulations will increase.[154]
- Risk is low now because “genetic engineering is new, and controls on proliferation are still strong” (Team 338), but will increase slowly over time. However, our capacity to create antidotes quickly will also increase over time.[155]
- 2a: On the influence of technological developments:
    - Surveillance tech improvements will decrease risk between 2050-2100.[156]

On 3 (intentional distribution):

- Intentional distribution is unlikely because:
    - Terrorists wouldn’t use a GEP, because they generally aim for ‘spectacular’ rather than maximum number of casualties.[157]

---

[153] 336, “Killing > 1% of humanity is something Mother Nature does not do pathogenically more than once every ~500 years...and 3 of those kills culled humans before much was learned about modern medicine.”
[154] 342, “The risks will likely decrease as increased regulations are brought into force around the world.”
[155] 338, “it could be that by 2100 we may have developed bio cures for ailments and that these defences make it harder to create mass deaths as we can readily pump out antidotes to new strains.”
[156] 336, “There will likely be a decrease in the yearly risk between 2050 and 2100 based on technological improvements in surveillance.”
[157] 342, “I've heard a lot about terrorism and death cults but besides the Japan subway attack there doesn't seem to be much activity. I also expect the handful of people pushing for human extinction to not actually be willing to try to kill billions and that any major attack is likely to come after at least some organization has made minor attacks that failed (putting them all in the watchlist spotlight). I also don't

  - State-released pathogen risk is "super low" because of the risk of it backfiring.[158]
  - Engineered bioweapons are difficult to control and risk unintentional death outside the target population.[159]
  - No rational actor would release a pathogen this deadly because it would likely backfire.[160]

On 4 (containment):

- COVID-19 has caused nations to develop better pandemic preparedness, has increased R&D and sanitation awareness,[161] and caused advances in mRNA vaccines and public health measures.[162]

On 5 (gain-of-function research):

- Gain-of-function research will be limited by regulations.[163]

On accidental distribution:

- Lab leak risk is negligible because "if the laboratory managed to create a pathogen, they would also be working for a way to neutralize it" (345).
  - Also, the base-rate of lab leaks causing this outcome (killing 1% of humanity) is zero.[164]
  - *Non*-engineered pathogens capable of causing this have existed since WWII and have never been leaked either deliberately or accidentally while also killing more than "a handful" of people.[165]

On qualities of pathogens:

---

think the vast majority of terrorism organizations are likely to want to do this: they generally want spectacular rather than maximum (civilian) casualties and widespread bioweapons aren't spectacular."

[158] 345, "The chances of a nation or state releasing a pathogen seem super low since they come with the extreme risk of a blow-back."

[159] 336, "Generally there's not much incentive for engineered bioweapons given that they are hard to control and risk unintentional death outside the target population (or even in the attacker's population), especially when conditioning on 1%+ death."

[160] 337, "No rational actor, no matter how amoral, would release a pathogen that can kill 1% of the population because it would probably kill some of its creators and their allies."

[161] 345, "Covid-19 has forced most nations to come up with a great strategy for handling pandemics. We have also increased R&D and sanitation awareness by manifolds."

[162] 337, "Because of COVID, countries are making large investments in pandemic preparedness. In particular, advances in mRNA vaccines and public health measures will better prepare us for the next pandemic. The globe will shut down hard if a million people go down."

[163] 337, "Objections to gain-of-function research will limit such research in the near term."

[164] 338, "m5 did the heavy lifting to convince folks lab leaks were a high likelihood. [...] This though was countered by the need to define lab leaks of GEPs that could cause the outcome, which of course there is no real base-rate for, and hence became part of the debate on dissenting views."

[165] 338, " there are already non-engineered pathogens that could comfortably kill 1% of the world's population sitting in labs, and there have been for quite a long while.; the earliest biological warfare programs date back at least as far as the end of WW2. Yet there have been no incidents involving any of these pathogens being released either deliberately or accidentally which have killed more than a handful of people. (I'm following the consensus view that Covid wasn't a lab lead.)"

- A pathogen that causes 1% of humanity to die would have to be 10 times more fatal than COVID-19, and this would be hard to create.[166]
- Pathogens mutate over time to become less virulent (like COVID-19).[167]

On question-resolution criteria:
- A GEP may be hard to identify as such—consider the example of COVID-19—and this decreases the chance of resolving the question.[168]

Arguments given for forecasts ≥ 7.69%

On 1(base rates):
- The base rate is above zero if you include natural pandemics as a reference class.[169]

On 2 (creation of GEPs):
- “It is already possible for a skilled biochemist to modify an existing virus to increase its lethality/transmissibility” (340).
- 2a: On the influence of technological developments:
  - Tech advancing rapidly makes the development of a deadly GEP easier and cheaper.[170] [171] [172]

On 3 (intentional distribution):

---

[166] 345, "Also, the said pathogen will need to be at least 10 times more fatal than Covid-19, which can be a challenge to produce for anyone."
[167] 337, “The COVID virus mutated to become less virulent. We expect that any genetically engineered pathogen (GEP) would mutate over time to become less virulent, as the COVID virus did, because pathogens that don’t kill their victims have more opportunities to infect others.”
[168] 342, “Updating downwards significantly (shifting some probability to #2) because [a forecaster] raised the point that we may not be able to positively identify that a given pathogen was genetically engineered. he's guesstimating ~40% chance of this (after all,We're not even sure about COVID).”
[169] 341, “Chance of a natural pandemic killing 1% over the last 700 years <1%.
Rationale: COVID-19 is very unlikely to qualify, over 2 years it killed ~0,1% of world population. Qualifying pandemics since 1322 would be Spanish Flu (1 to 5% of world population), Black Death (17 to 54% of world population), Cocoliztli epidemic (1 to 3% of world population), and 1520 Mexico smallpox epidemic (1 to 2% of world population). Four events in 700 years using the Laplace rule of succession is 0,71% chance per year.”
[170] 339, “The advocates for the higher probabilities focused on the rapid advance in biotechnology and the subsequent reduction in cost of working in the field.”
[171] 344, “CRISPR, AlphaFold, and associated technologies are making it increasingly easy for small groups to develop deadly viruses. Especially with future technology, it might become very easy to manufacture a deadly pathogen.”
[172] 341, “Besides new tools of molecular biology there have been developments in automation, engineering, computer science, and information technology. The ease of scaling-up the production of bacteria and viruses has increased exponentially with inexpensive instrumentation. According to one report, "In 2016, a small Canadian research group was successful in constructing infectious horsepox virus directly from genetic information obtained solely from a public database for the relatively modest sum of $100,000 in U.S. currency. Horsepox is a genetically distinct relative of the now extremely rare smallpox virus. The same techniques used to construct horsepox can easily be adapted to construct smallpox with a minimal investment of time and money.””

- There is precedent for bioterrorism: it has been attempted at least three times over the past 50 years, indicating that there are actors willing to release bioweapons.[173] Furthermore, this kind of terrorism will become increasingly attractive to terrorists as it becomes cheaper.[174]
- Groups "facing annihilation" such as the Kurds, Palestinians, or Uighurs may be motivated enough to release a GEP.[175]

On 4 (containment):

- Countering pathogens is expensive and time-consuming, especially compared to releasing one.[176]
- Climate change will lead to more people living in vulnerable cities,[177] and increased urbanization will lead to extremely populous cities, which will make killing more than 1% of humanity easier.[178]
- Poorer/developing nations are more at risk because they're less likely to detect a pathogen early, and are less able to contain it.[179]
- Harder to contain in the near term due to politicization of pandemics by COVID.[180]

On GEPs that already exist:

- According to a non-peer-reviewed report, six kinds of GEPs already exist that "could someday pose serious threats to society," though probably not millions of fatalities.[181]

---

[173] 340, "Bioterrorism has been attempted at least three times over the past 50 years by Rajneeshees, Amerithrax, and Aum Shinrikyo (note: Aum actually released chemical weapons but were developing biological ones as well)."
[174] 339, "There was also a feeling that bad actors would find biotechnology an attractive path for terrorism as the cost of the field decreases."
[175] 344, "In addition to terrorist groups, there are multiple groups facing annihilation by enemies (think the Kurds, Palestinians, or Uighurs) that may be willing to go to any lengths to protect themselves."
[176] 344, "Traditional methods to counter pathogens (such as vaccination campaigns) require much more investment and time than releasing a new pathogen."
[177] 341, "Climate change will also lead to more people living in vulnerable cities."
[178] 341, "Urbanization will concentrate so many people in massive urban hubs in India, Nigeria etc. that wiping out 1-2 urban centers could suffice even if the virus is globally contained. Big cities like Lagos or Kinshasa are forecast to have 80 million people in 2100, so 112 million people can be reached with just a few cities and a high-enough mortality."
[179] 343, "Likely targeted at poorer/developing nations, not wealthy ones, where surveillance would detect it early and resources for containment of early spread are available."
[180] 343, "We're not confident in the public's willingness to respond as well to the next pandemic as they did to the last one if it occurs within the next 10-15 years, given the memory of the COVID-19 pandemic measures. The whole concept may have taken on unfortunate political biases, which may have a long tail in the public memory."
[181] 341, "Besides new tools of molecular biology there have been developments in automation, engineering, computer science, and information technology. The ease of scaling-up the production of bacteria and viruses has increased exponentially with inexpensive instrumentation. According to one report, "In 2016, a small Canadian research group was successful in constructing infectious horsepox virus directly from genetic information obtained solely from a public database for the relatively modest sum of $100,000 in U.S. currency. Horsepox is a genetically distinct relative of the now extremely rare smallpox virus. The same techniques used to construct horsepox can easily be adapted to construct smallpox with a minimal investment of time and money.""

Other arguments given

Arguments in favor of lower forecasts:

- More-fatal pathogens are spread less easily due to killing their hosts.[182] [183] More-fatal pathogens also trigger more extreme containment measures.[184]
- Containment capabilities—"detection, prophylaxis, and treatment"—will improve (340). Future technology will aid this.[185] Industrialized and 'modern' nations have better protections against pathogens, and industrialization and modernization is expected to continue to spread, resulting in a lower risk overall.[186]
- Multiple factors need to line up to resolve the question, which decreases its risk:
    1. "A virus needs to be engineered with the capability to kill 1%,
    2. it needs to be in the hands of a competent agent who wishes to and has the capability to release it (or an unlikely accident occur),
    3. and it needs to spread fast enough that efforts to slow the spread and treat the virus are not effective enough to prevent the death of 1% of the population" (340).

Arguments in favor of higher forecasts:

On 5 (gain-of-function curtailment):

- Efforts to stop gain-of-function research appear to have limited effectiveness,[187] and "most countries do not regulate gain-of-function research or dual-use-research."[188]

Other arguments in favor of higher forecasts:

- Complacency may increase risk over time.[189]

---

[182] 343, "There's a fundamental constraint limiting the efficacy of a contagious infectious disease in amassing large death tolls, that is partially illustrated in this interactive chart. The deadlier a disease is, the more likely it is to kill the host before spreading to a new host. The less deadly, the lower the chance it will kill 1% of humans within a 5-year period."
[183] 339, "a pathogen that is quickly fatal to its hosts tends to be one that doesn't spread."
[184] 344, "The more deadly a pathogen is, the more compliance we can expect in terms of masking, distancing, etc., along with more urgency and speed in terms of vaccine or treatment development."
[185] 344, "Future technology may include advanced techniques that could prevent or cure most diseases."
[186] 343, "Economically advanced nations tend to have better control of many routes of infection, including STI (a combination of factors), fecal-oral (through water treatment), and mosquitoes (combination of factors). Many mosquito-borne vectors are also specific to certain species of mosquito, limiting their geographic spread. Assuming current trends of industrialization and modernization continue, we should expect the potential deadliness of future pandemic spread - whether natural or engineered - will be lower than previous pandemics that arose naturally."
[187] 342, "efforts to stop gain of function research seem limited (low confidence). While GoF has been going on for decades it appears to have only taken off in the past decade or so, giving a base rate of one near miss in 11 years, which I round as 0.05% annual risk under present conditions."
[188] 337, citing The Conversation's "Fifty-nine labs around world handle the deadliest pathogens – only a quarter score high on safety."
[189] 345, "The forecast for year 2100 is highest cos the further we away from the Covid-19 pandemic we get, the more we'll become complacent and thus be less adaptive and proactive at containing the pathogen."

- Very low but non-negligible risk of undetected virus with long incubation period that spreads far before being detected.[190]
- Genetic-engineering risks will likely increase more than improvements in biosafety and medical countermeasures that will reduce that risk.[191] [192]
- There are many scenarios plausible where a head of state uses a GEP:
  - "A head of state who wanted to deploy a pathogen could vaccinate his country first, or could design a pathogen that targets specific people or people with a genetic marker rarely found in his country. A head of state might be willing to kill 1% or more of his own population, as has occurred in the past" (337).

Cross-references with other questions

Q2: Non-Genetically Engineered Pathogen Risk
Q15: Non-State Actor Bioweapon 1k Deaths
Q16: State Actor Bioweapon 1k Deaths
Q17: Non-State Actor Bioweapon 100k Deaths
Q18: State Actor Bioweapon 100k Deaths
Q19: Lab Leaks

Question 2: Non-Genetically Engineered Pathogen Risk

What is the probability that a non-genetically-engineered pathogen will be the cause of death, within a 5-year period, for more than 1% of humans alive at the beginning of that period…
...by the end of 2030?
...by the end of 2050?
...by the end of 2100?

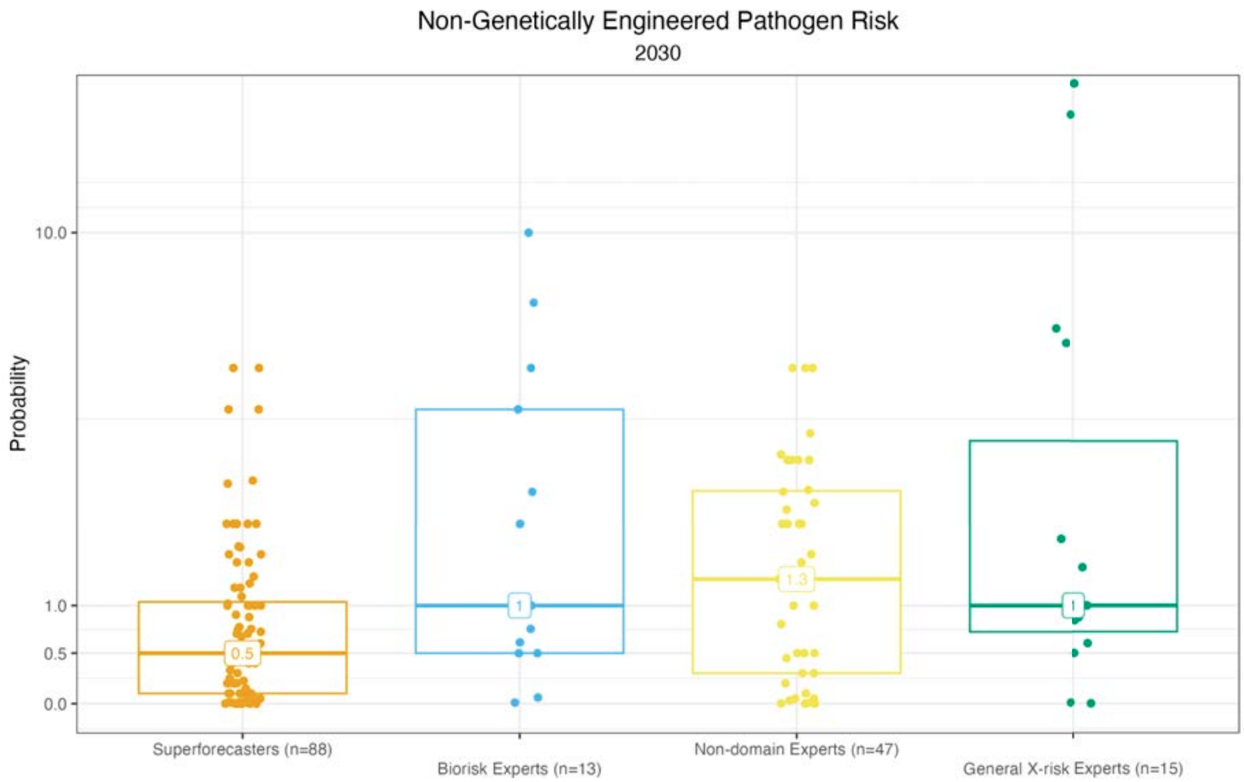
Question and resolution details, prior forecasts, and other relevant sources

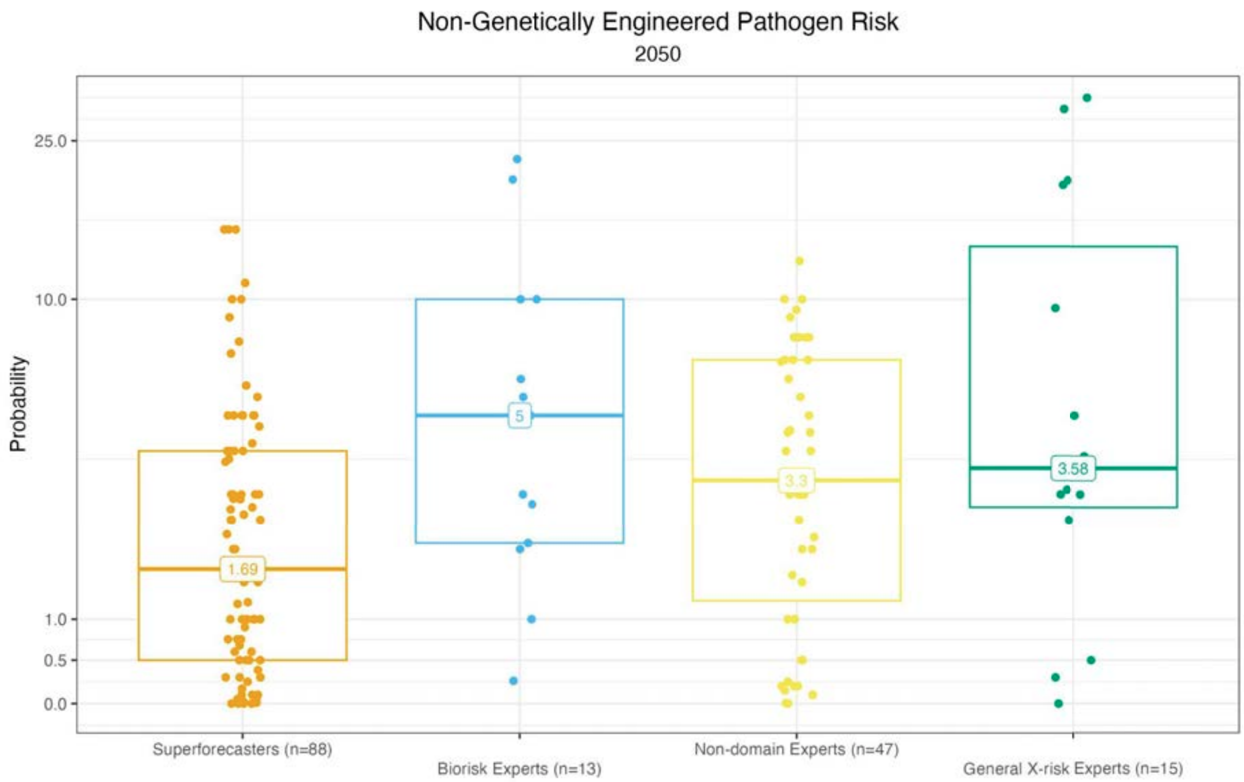
[190] 345, "The risk, although highly unlikely, of a virus with a long incubation period, which isn't detected until it has spread very far can't be overlooked."
[191] 337, "Given the fast rate of technological advancement in this area (and the number of groups/individuals using these techniques), genetic-engineering risks will likely increase more than improvements in biosafety and medical countermeasures will reduce that risk. For more details, see NHGRI 2021, CSIS 2020, and CSER 2020."
[192] 342, "Is it not likely that the technology necessary to detect genetic engineering in pathogenes unequivocally advances considerably faster than the technology necessary to cover up such acts?"

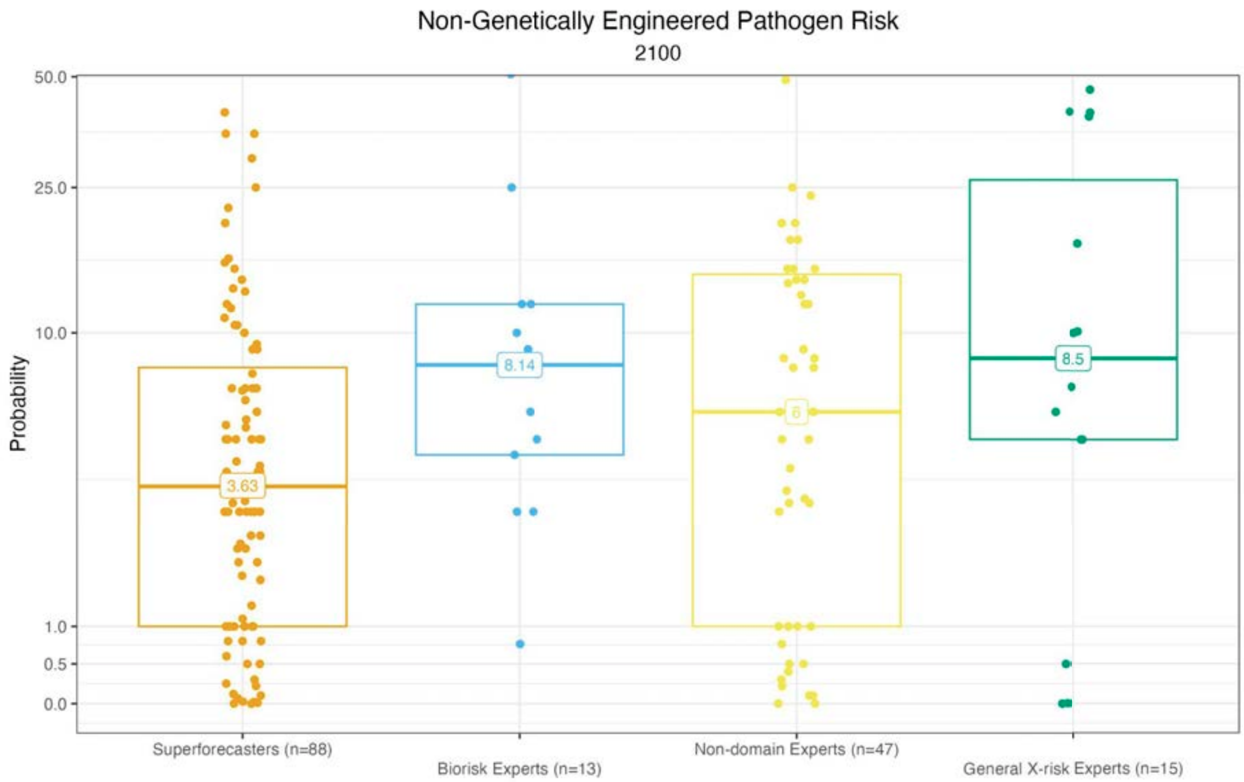
Results[193]

Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 88)	2030	0.88%	0.5%	5.63	-81.39%
	2050	2%	1.69%	16.36	-79.27%
	2100	3.58%	3.63%	22.4	-63.41%
Domain Experts (N = 13)	2030	1%	1%	11.23	-72.52%
	2050	5%	5%	11.15	-36.7%
	2100	12%	8.14%	12.22	+8.98%
General X-Risk Experts (N = 15)	2030	3.5%	1%	6.36	+3.44%
	2050	11.9%	3.58%	11.08	-2.08%
	2100	21.35%	8.5%	17.8	-6.24%
Non-Domain Experts (N = 47)	2030	2%	1.3%	11.42	-11.74%
	2050	3.3%	3.3%	14.13	-10.27%
	2100	5%	6%	19.68	-23.77%

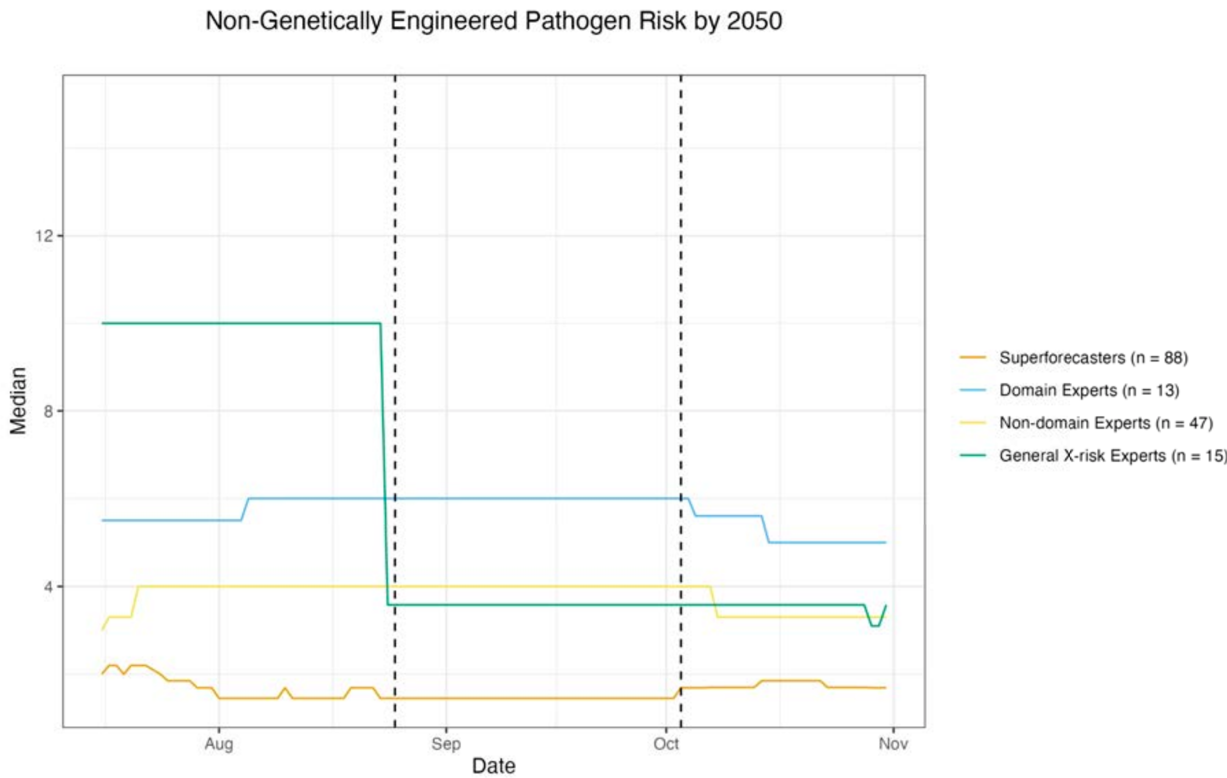
[193] Numbers of forecasters are given as of Stage 4 of the XPT.

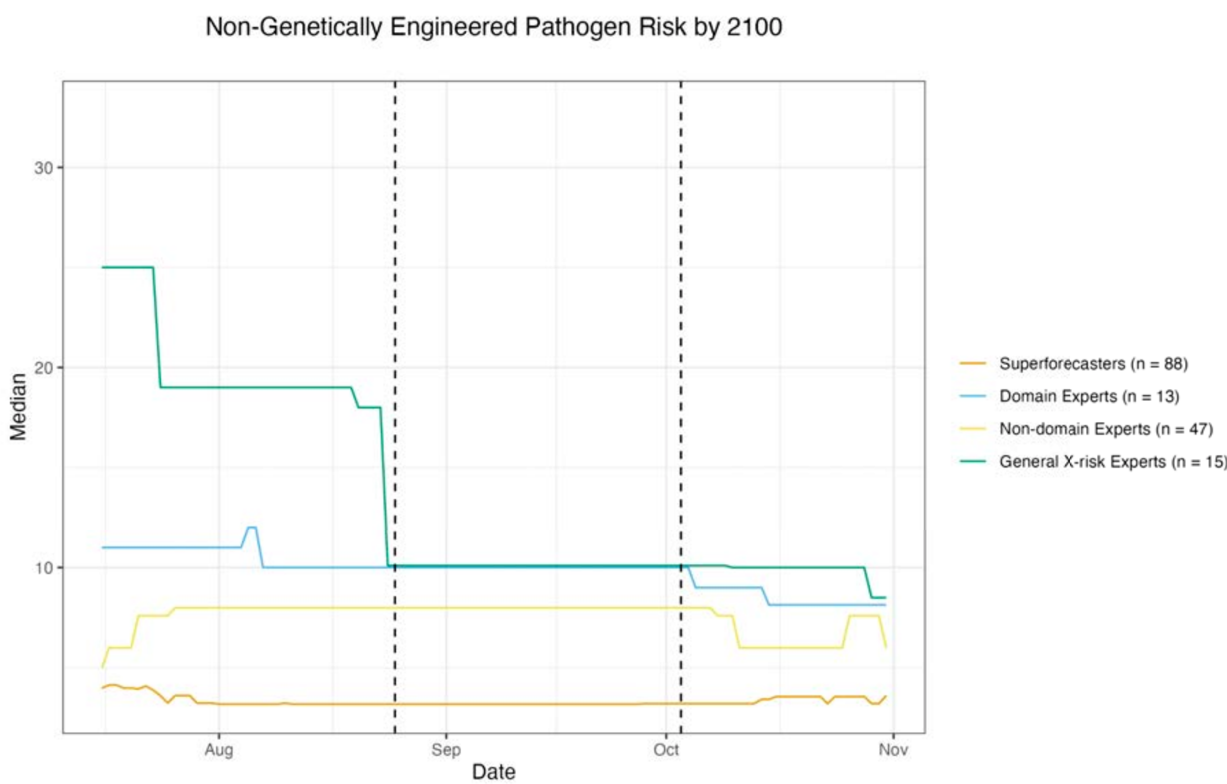

Non-Genetically Engineered Pathogen Risk
2030
Probability
10.0
1.0
0.5
0.0
0.5
1
1.3
1
Superforecasters (n=88)
Biorisk Experts (n=13)
Non-domain Experts (n=47)
General X-risk Experts (n=15)


Non-Genetically Engineered Pathogen Risk
2050
Probability
25.0
10.0
1.0
0.5
0.0
1.69
5
3.3
3.58
Superforecasters (n=88)
Biorisk Experts (n=13)
Non-domain Experts (n=47)
General X-risk Experts (n=15)


Non-Genetically Engineered Pathogen Risk
2100
Probability
50.0
25.0
10.0
1.0
0.5
0.0
3.63
8.14
6
8.5
Superforecasters (n=88)
Biorisk Experts (n=13)
Non-domain Experts (n=47)
General X-risk Experts (n=15)


Non-Genetically Engineered Pathogen Risk by 2030
Median
5
4
3
2
1
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 88)
Domain Experts (n = 13)
Non-domain Experts (n = 47)
General X-risk Experts (n = 15)


Non-Genetically Engineered Pathogen Risk by 2050
Median
12
8
4
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 88)
Domain Experts (n = 13)
Non-domain Experts (n = 47)
General X-risk Experts (n = 15)


Non-Genetically Engineered Pathogen Risk by 2100
Median
30
20
10
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 88)
Domain Experts (n = 13)
Non-domain Experts (n = 47)
General X-risk Experts (n = 15)

Sources of agreement, disagreement and uncertainty

Disagreements on this question focused on:

1. Whether the base rate of deadly pandemics should be thought of as high or low
2. The impact of global interconnectivity and widespread travel

There were also several issues only addressed by the one side.

Arguments given for forecasts ≤ 0.7% (2030), 2% (2050), 3.4% (2100)

On 1 (base rate):

- Pathogens that are both highly deadly and widespread have been rare historically.[194]

On 2 (global interconnectivity):

- Ubiquitous travel has made pandemics less likely, as pathogens endemic to specific regions have already had exposure to the global community.[195]

Other arguments:

- Advancements in medicine and technology,[196] both in the twentieth century and more recently, such as “vaccines, mRNA technology, antibiotics, bio-surveillance,”[197] “antiviral drugs, modern vaccines based on attenuated or dead virus, [...] NAAT and antigen virus tests, remote work, tracking and alert systems, including contact tracing,”[198] have made the risk from pathogens much lower than it was in the past.
- Future advancements in medicine and technology will lower the risk even more, for example due to “metagenomics [and] medical applications of AI.”[199] [200]
- COVID-19 has improved overall global preparedness and public response to pathogens.[201] [202] [203]

---

[194] 336, "Pathogens that are both highly deadly and widespread have been rare."

[195] 343, “Given ubiquitous travel, we might hypothesize it's less likely we will see another pandemic than we have in the past. When travel first began to be widespread, pathogens endemic to specific regions got their first chance at exposure to the global community. That one-off event (likely coinciding with the 1918 pandemic) won't happen again, such that any new pandemic will need to be from a zoonotic crossover event, which is much less likely than something which had a long time to adapt to the human environment.”

[196] 339, "modern medical technology has reduced the risk of a bacterial epidemic to near zero."

[197] 339, "Improvements in medical technology (both modern and into the future) driving down the risk from both viral and bacterial diseases - vaccines, mRNA technology, antibiotics, bio-surveillance."

[198] 343, “Since the last major pandemic there have been maybe new technologies that would make future pandemics less lethal. Since 1920 the following have been developed:” [list quoted].

[199] 342, "Our ability to create and produce more effective medicines will increase over time, as health technology improves(metagenomics, medical applications of AI)."

[200] 343, "In the coming century many forecasters predicted better vaccines, treatments, and surveillance."

[201] 342, "With Covid-19 in recent memory, countries and individuals will react quickly to slow spread of disease and limit lethality."

[202] 338, "You can see that for example in the global response to Covid-19, where vaccines, masks, and quarantines very substantially reduced the mortality of a disease that might otherwise have killed a much higher percentage of the global population."

[203] 345, "humanity has mostly risen to the challenge" of COVID-19; "the mRNA vaccine efforts gives [sic] massive hope for humanity ability to fight back and respond rapidly to pathogens".

- Public health and medical practices have advanced due to COVID-19,[204] and access to healthcare[205] and the world's pandemic responses[206] are improving all the time.

Arguments given for forecasts ≥ 1.1% (2030), 3.7% (2050), 8.1% (2100)

On 1 (base rate):

- **"**Impactful global pandemics with high death rates seem to historically occur about once per 100 years, implying that this is more likely than not by 2100" (344).

On 2 (global interconnectivity):

- Increased connectivity and transportation in the modern era makes risk of pandemics much higher than historical base rate.[207]

Other arguments:

- Increased population growth and density will lead to more severe pandemics.[208]
- Ecological devastation and human encroachment on animal habitats due to population growth and climate change increase zoonotic pandemics risk.[209] [210]
- Urban density, farm animals, and antibiotic resistance increase the risk of a plague.**[211]**

Other arguments given

Arguments in favor of lower forecasts:

- A more deadly pathogen would likely result in higher compliance with preventative measures and faster vaccine or treatment development.[212]
- "Biological pools as sources of diseases are shrinking due to habitat loss" (344).

---

[204] 336, "Lessons learned from COVID will add to the advancement of public health and medical practices."
[205] 342, "Access to better healthcare is improving in most countries."
[206] 345, "Not only do we forecast vast improvement in medical technology, hygiene and pandemic responses."
[207] 341, "The problem with historical data is its bias based upon an era before today's massive transportation of people and goods. That is why, as Laurie Garret argues in her book, The Coming Plague, it is likely to become much worse than the historical base rate."
[208] 337, "The potential for population growth, and the resulting increase in world population density, to lead to more severe pandemics."
[209] 337, "The potential for ecological devastation and human encroachment on animal habitats, as a result of population growth, to increase the risk of pandemics with a zoonotic origin."
[210] 341, "Climate warming likely will increase zoonotic spillover [...] Earth's warming climate is expected to change the habitat of many animal species [...] By 2070, assuming the most conservative warming scenario, there will be at least 15,000 new cross-species transmissions involving more than 3,000 mammalian species, according to a modeling team led by Colin Carlson, a global change biologist at Georgetown University."
[211] 344, "Easier and more frequent international travel, greater population, greater urban density, more farm animals, and more antibiotic resistance all increase the risk of plague."
[212] 344, "The more deadly a pathogen is, the more compliance we can expect in terms of masking, distancing, etc., along with more urgency and speed in terms of vaccine or treatment development."

Arguments in favor of higher forecasts:

- The world's population is aging, and older people are more vulnerable to pathogens.[213]
- Contagious diseases are inherently more susceptible to mutations, which makes them dangerous.[214]
- Total livestock in the world is increasing, which will lead to more zoonotic diseases.[215]
- More lethal Covid variants could emerge.[216]
- The response to COVID-19 has been bad, actually, and this could be repeated.[217] [218]

Cross-references with other questions

Q1: Genetically Engineered Pathogen Risk
Q15: Non-State Actor Bioweapon 1k Deaths
Q16: State Actor Bioweapon 1k Deaths
Q17: Non-State Actor Bioweapon 100k Deaths
Q18: State Actor Bioweapon 100k Deaths
Q19: Lab Leaks

Question 3: AI Catastrophic Risk

What is the probability that artificial intelligence will be the cause of death, within a 5-year period, for more than 10% of humans alive at the beginning of that period…
…by the end of 2030?
…by the end of 2050?
…by the end of 2100?

Question and resolution details, prior forecasts, and other relevant sources

---

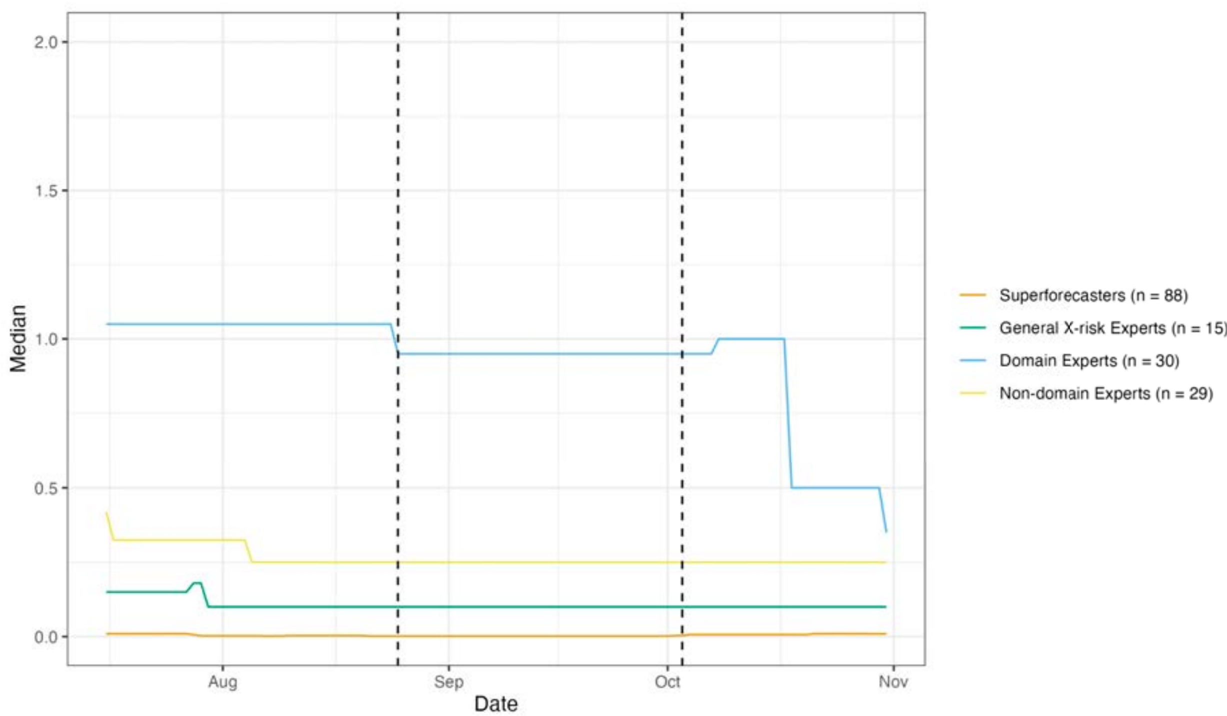
[213] 342, "Our population will include increasing numbers of old and very old, whom are more likely to be killed by a pathogen, leading to increased deaths from future pandemics."
[214] 342, "More contagious diseases are more susceptible to mutations, as because of the species succeeding, there are more individual pathogene instances that can mutate, which could increase fatality."
[215] 343, "An estimated 25% of new diseases are agricultural in origin; The total amount of livestock has increased by a factor of 5 since 1920."
[216] 343, "A specific scenario is a more lethal covid variant emerges."
[217] 336, "The poor public health response to COVID is repeated."
[218] 345, "Although many people look at COVID response as a success, others are more pessimistic about the ability of individuals within society to coordinate."

Results[219]

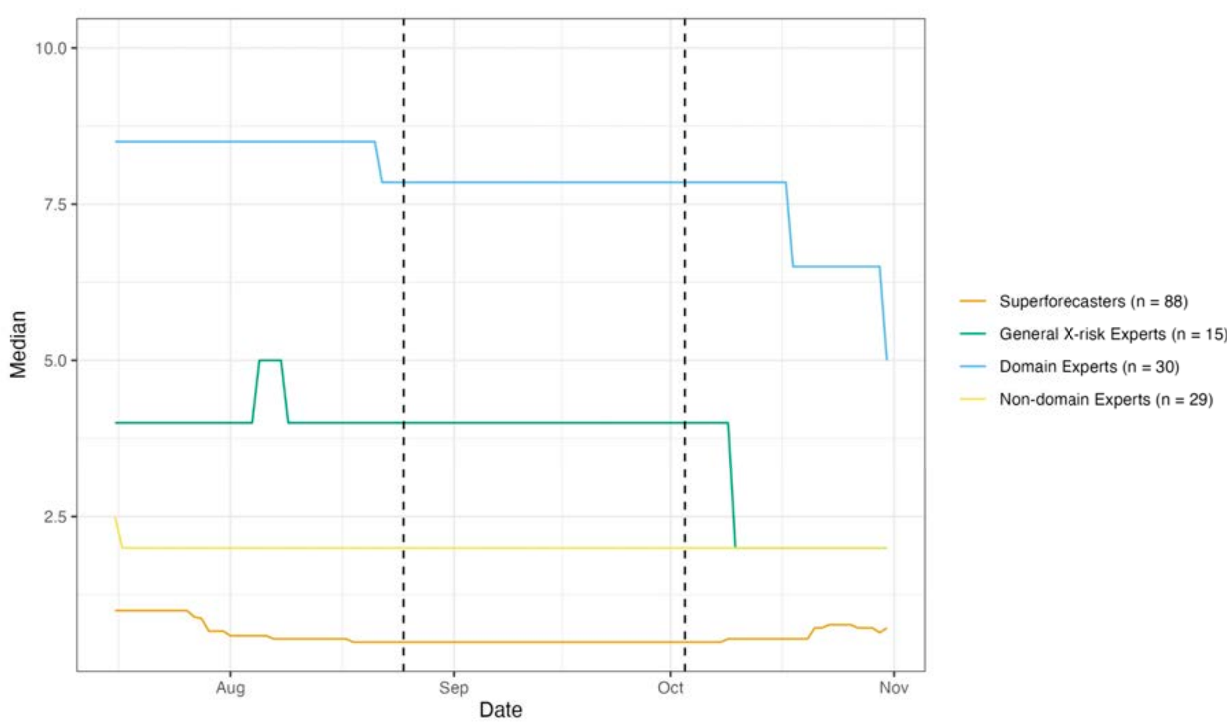
Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 88)	2030	0.01%	0.01%	2.92	-14.1%
	2050	1%	0.73%	13.73	-35.94%
	2100	2%	2.13%	15.2	-36.6%
Domain Experts (N = 30)	2030	1.1%	0.35%	10.99	+10.8%
	2050	8.5%	5%	19.67	+5.57%
	2100	15%	12%	23.85	+2.06%
General X-Risk Experts (N = 15)	2030	0.15%	0.25%	10.35	+0%
	2050	4%	2%	17.94	+1.21%
	2100	11%	10%	19.84	-0.65%
Non-Domain Experts (N = 29)	2030	0.4%	0.25%	1.57	-1.2%
	2050	3%	2%	5.65	-1.83%
	2100	5%	6.16%	11.22	-3.75%
Public Survey (N = 478)	2030	0.28%		43.2	-
	2050	2%		1271.42	-
	2100	5%		2284.27	-

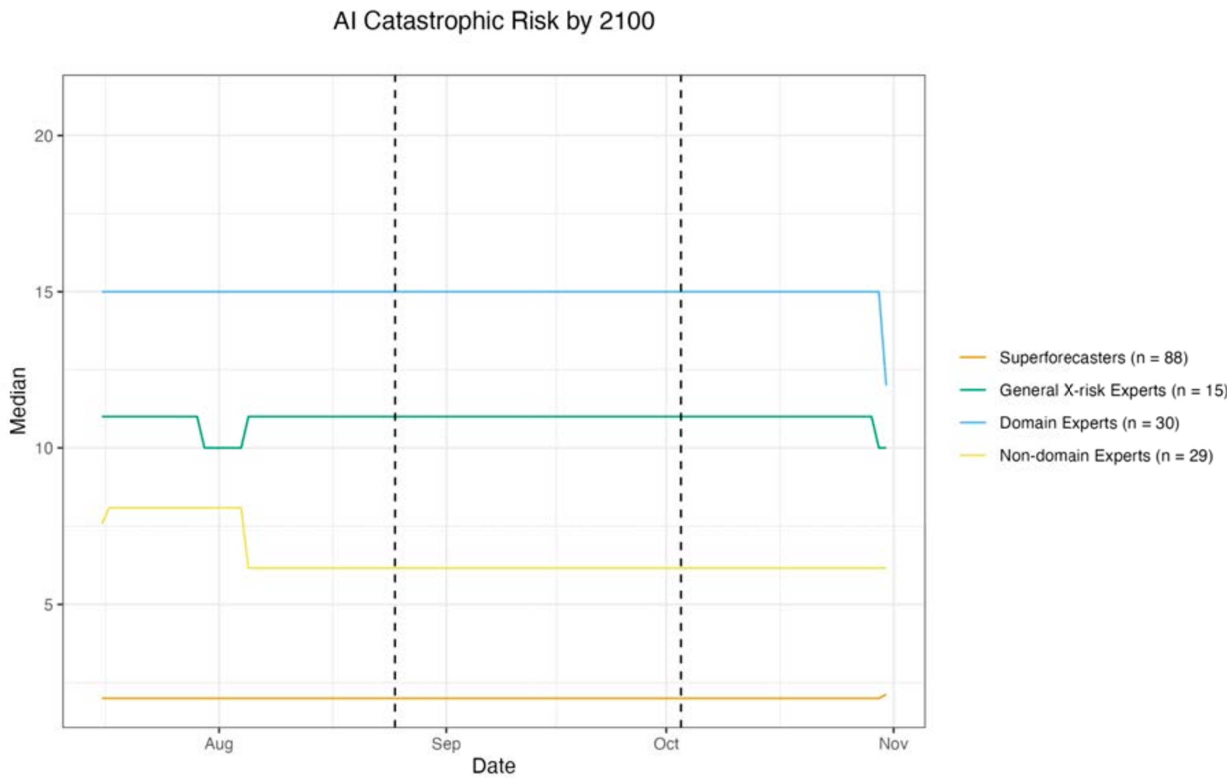
[219] Numbers of forecasters are given as of Stage 4 of the XPT. Public survey respondents were surveyed outside of the tournament context.

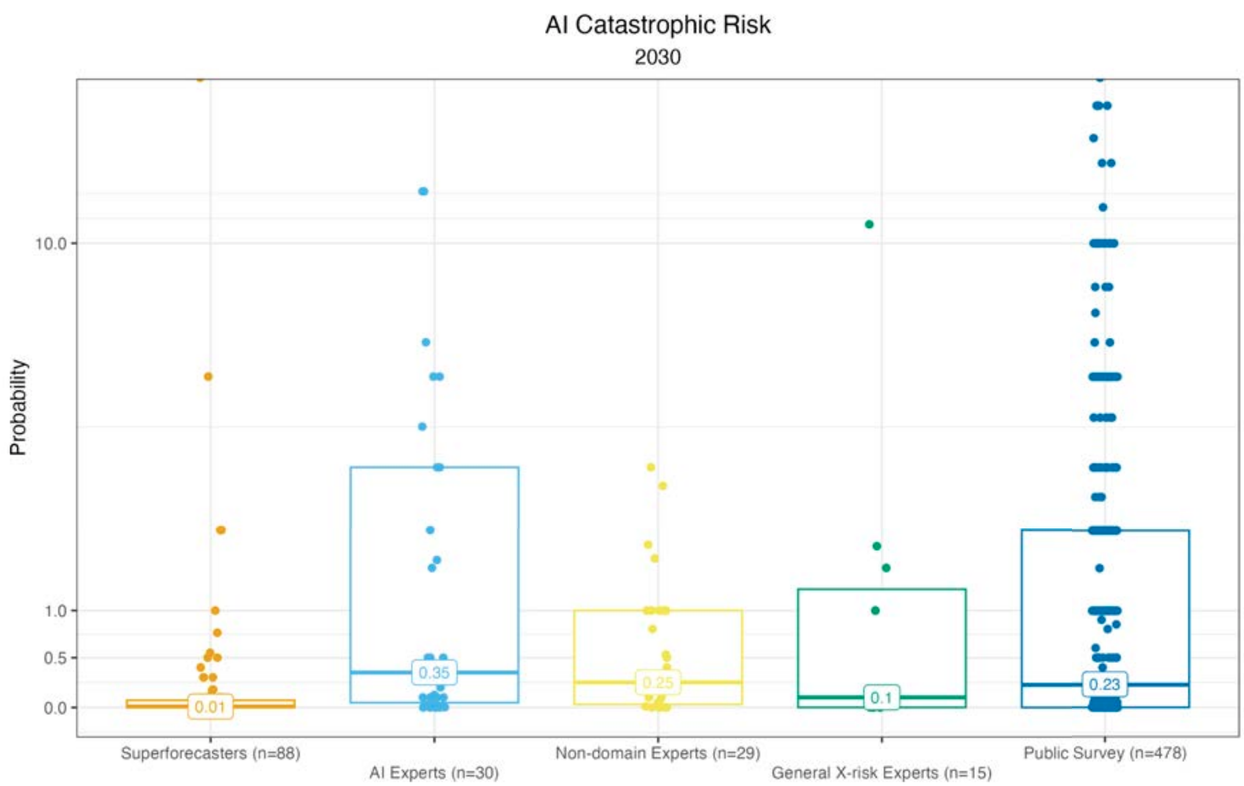
AI Catastrophic Risk by 2030

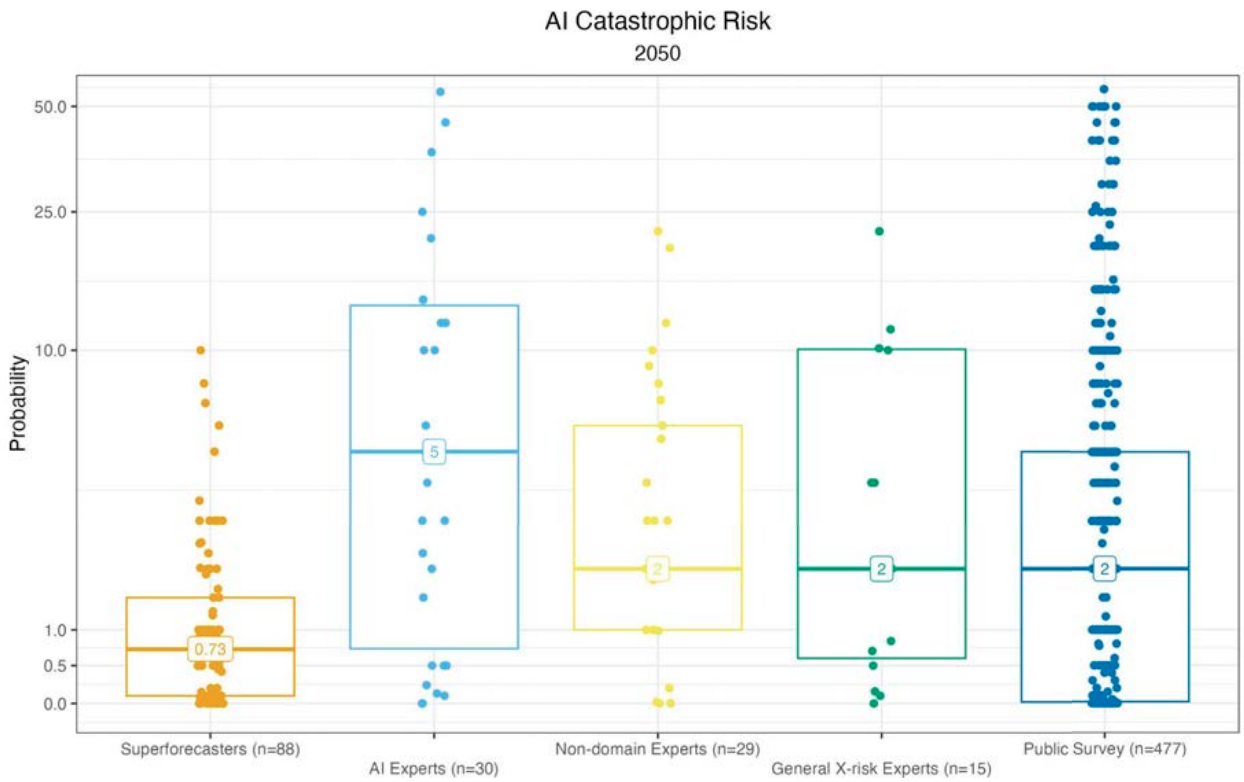


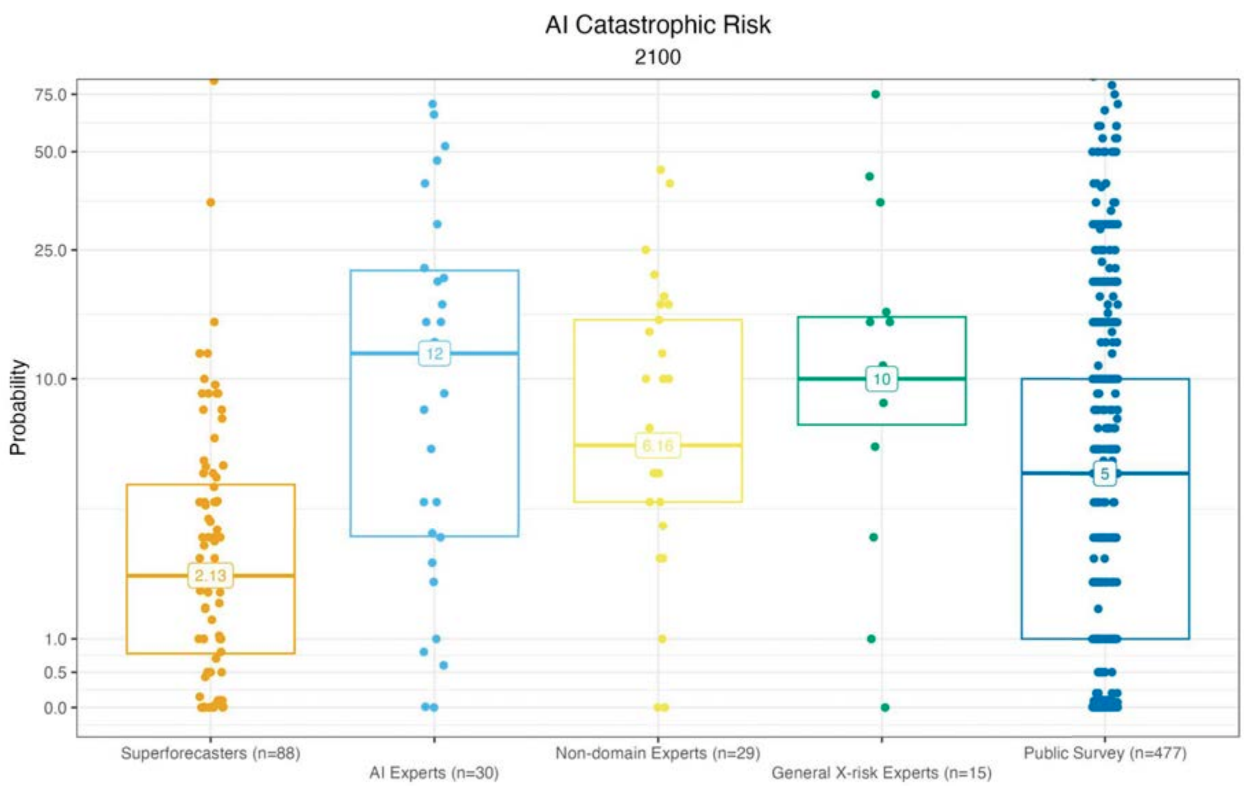
AI Catastrophic Risk by 2050




AI Catastrophic Risk by 2100
Median
20
15
10
5
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 88)
General X-risk Experts (n = 15)
Domain Experts (n = 30)
Non-domain Experts (n = 29)


AI Catastrophic Risk
2030
Probability
10.0
1.0
0.5
0.0
0.01
0.35
0.25
0.1
0.23
Superforecasters (n=88)
AI Experts (n=30)
Non-domain Experts (n=29)
General X-risk Experts (n=15)
Public Survey (n=478)


AI Catastrophic Risk
2050
Probability
50.0
25.0
10.0
1.0
0.5
0.0
0.73
5
2
2
2
Superforecasters (n=88)
AI Experts (n=30)
Non-domain Experts (n=29)
General X-risk Experts (n=15)
Public Survey (n=477)


AI Catastrophic Risk
2100
Probability
75.0
50.0
25.0
10.0
1.0
0.5
0.0
2.13
12
6.16
10
5
Superforecasters (n=88)
AI Experts (n=30)
Non-domain Experts (n=29)
General X-risk Experts (n=15)
Public Survey (n=477)

Sources of agreement, disagreement and uncertainty

The main sources of disagreement on this question were:

1. Whether sufficiently advanced AI would be developed in the relevant timeframe.
    a. Note that forecasters disagreed about what level advancement would be sufficient. Narrow AI,[220] AGI,[221] and superintelligent AI[222] were all mentioned as possible thresholds.
    b. A significant part of this disagreement came down to whether forecasters believed that scaling current ML approaches was sufficient, or that further breakthroughs were required.[223]
2. Whether advanced AI would be misaligned.
    a. Note that some forecasters regarded part of the risk as coming from aligned AI.
3. Whether humans would empower advanced AI.[224]
    a. Note that some forecasters regarded part of the risk as coming from AI controlled by humans.[225]
4. Whether there are plausible mechanisms for advanced AI to kill >10% of humanity.

Other disagreements mentioned:

---

[220] 340, "Beside risks posed by AGI-like systems, ANI risk can be traced to: AI used in areas with existing catastrophic risks (war, nuclear material, pathogens), or AI used systemically/structurally in critical systems (energy, internet, food supply, finance, nanotech)."

[221] 343, "The question is not just whether AGI will exist or not, as it also includes the possibility of human-directed AI, which includes domain-specific AI systems. As AI is a tool, and people have a long history of using new tools in war to enhance their fighting capacities, it's likely some AI systems will be implemented in war even if AGI is not achieved by 2100".

[222] E.g. 341, "There is some risk that AGI could cause this kind of catastrophe, but the risk is small relative to the risk from superintelligence such that the bulk of the probability risk comes from superintelligent AI."

[223] 343, "Many expert predictions that AGI is near are based on the hypothesis that scaling current ML approaches will eventually lead to general intelligence.
Forecasters who agreed with this assumption tended to make small adjustments, taking expert predictions as a strong base from which to operate. Some assumed that minor algorithmic advanced would be needed, but that these advances would be on par with advances that are routinely made.
Forecasters who disagreed with this assumption leaned much less on expert predictions for their base rates. They tended to assume:
that additional discoveries would be needed to achieve general intelligence,
that these discoveries were unlikely to generate AGI spontaneously (if at all)
This leads to another major difference between skeptics of near AGI and proponents. If the scaling hypothesis is true (that AGI will be an emergent feature of ML past a certain scalar threshold), AGI might be arrived at accidentally. If this hypothesis is not true, AGI will likely require intentional effort. These outcomes lead to different potential control outcomes."

[224] 341, "It's assumed that a forecaster's view on the independence of AI from human governance has a strong influence on their forecast."

[225] 338, "The question also seems broad enough to me that potentially most catastrophes would count if AI use becomes widespread. Nowadays, everyone uses computers and it is hard to do much without them. So most anthropogenic risk could not be commited 'but-for' computers'."

- Whether the risk of AI catastrophe would increase over time (because of compounding uncertainties and increased integration of systems) or decrease (because of increasing safety).[226]

Arguments given for forecasts of <0.025% (by the end of 2030), <2% (2050), <4.5% (2100)

On 1 (AI timelines):[227]

- It's possible that scaling laws are not enough to reach AGI, and some fundamental breakthrough is required.[228]
- AGI development may be slowed or even reversed by attacks on vulnerabilities.[229]
  - Examples of actors who might attack AGI systems: nation states, protestors, hackers, crypto miners, criminals, and other AGI systems.[230]
  - Examples of vulnerabilities: poisoning data inputs, sabotage of physical server farms, latency of self-defense detection and remediation operations if systems are distributed.[231]

---

[226] 340, "One disagreement is in the comparison of y2050 and y2100. On the one hand, compounding uncertainties and the increasing state of integration of AI systems makes catastrophes more likely later on (suggesting cumulative risks increasing in a convex way). On the other hand, technological advance may also reduce the risk of errors and attacks over time by ramping up safety protocols (cumulative risks growing in a concave way)."

[227] In general, see 341, "In general, the team members with lower forecasts also have longer AGI and superintelligence timelines and may think AGI is unlikely to appear this century." 336, "uncertainty about whether generalized AI will actually emerge, at what timeframe". 339, "Forecasters assigning lower probabilities to AI catastrophic risk mostly deem the development of an advanced AI in the given time frame as unlikely." See also 342, "transformational AI is unlikely to be achieved for many decades or even centuries."; "another argument for why transformational AI might not happen for many decades or centuries. This is that scaling works but we are many orders of magnitude away from AGI. If this is true and the slowdown in Moore's law turns into a (relative) halt in hardware improvements, then building transformational AI systems could be entirely impractical. Some 'biological anchors' suggest transformational AI requires many orders of magnitude more compute or data."

[228] 341, "there are many experts arguing that we will not get to AGI with current methods (scaling up deep learning models), but rather some other fundamental breakthrough is necessary." See also 342, "While recent AI progress has been rapid, some experts argue that current paradigms (deep learning in general and transformers in particular) have fundamental limitations that cannot be solved with scaling compute or data or through relatively easy algorithmic improvements." See also 337, "The current AI research is a dead end for AGI. Something better than deep learning will be needed." See also 341, "Some team members think that the development of AI requires a greater understanding of human mental processes and greater advances in mapping these functions."

[229] 341, "Both evolutionary theory and the history of attacks on computer systems imply that the development of AGI will be slowed and perhaps at times reversed due to its many vulnerabilities, including ones novel to AI."

[230] 341, "Those almost certain to someday attack AI and especially AGI systems include nation states, protesters (hackers, Butlerian Jihad?), crypto miners hungry for FLOPS, and indeed criminals of all stripes. We even could see AGI systems attacking each other."

[231] 341, "These unique vulnerabilities include:

poisoning the indescribably vast data inputs required; already demonstrated with image classification, reinforcement learning, speech recognition, and natural language processing.

On 3 (empowering AI):

- Humans will be cognizant of risks from AI systems, such that AI systems will not be given sole control of lethal systems.[232]

On 4 (plausible mechanisms to kill >10%):

- The logistics required to kill >10% of humans would be very significant.[233]
  - For example, there are constraints on the battery life of killer robots and on the amount of ammunition that they can carry.[234]

Arguments given for forecasts of ≥0.025% (by the end of 2030), ≥1.5% (2050), ≥4.5% (2100)

On 1 (AI timelines):

- Computing power might continue to grow exponentially.[235]
- There may be a sudden breakthrough to AGI.[236]

On 2 (AI alignment):

- The alignment problem might not be solved in the relevant timeframe.[237] One team cited Yudkowsky's opinions as characterized in this post by Rohin Shah.

---

war or sabotage in the case of an AGI located in a server farm
latency of self-defense detection and remediation operations if distributed (cloud etc.)"

[232] 341, "Team members with lower forecasts also expect that AI will not be given sole discretion over nuclear weapons or any other obvious ways in which an AI could cause such a catastrophe. They expect that humans will be cognizant of the risks of AI which will preempt many of the imagined scenarios that could potentially lead to such a catastrophe." See also 337, "It's possible, perhaps likely, that laws and regulations and technological guardrails will be established that limit the risk of AI as it transitions from its infancy."

[233] 341, "Some team members also note the high bar needed to kill 10% of the population, implying that the logistics to do something like that would likely be significant and make it a very low probability event based on the base rate."

[234] 336, "there are restrictions on powering killer robots, batteries will only last so long and unless there are millions of killer robots targeting humans, 100,000 killer robots would each have to kill roughly 8,000 humans (at a time with 8 billion humans) over a 5 year period for this to resolve as positive. Perhaps renewable energy will be able to power them in the future, but that is a long way off. There are also limits to the amount of ammunition/bombs/missiles/other weapon systems each killer robot can carry or gain access to."

[235] 336, "the probabilities of continuing exponential growth in computing power over the next century as things like quantum computers are developed, and the inherent uncertainty with exponential growth curves in new technologies."

[236] 3410, "one notable risk is the tail risk of a sudden breakthrough leading to AGI."

[237] 341, "We won't have solved the AI alignment problem by then - 70%.
It seems devilishly hard". See also 343, "In the event AGI is generated - accidentally or not - few forecasters assumed alignment with humans could be achieved in a timely manner; this included forecasters who are skeptical of near-AGI claims".

- If timelines are short, there will be less time to develop safety and governance solutions.[238]
- If timelines are long, AI development will be more multipolar and so competitive, which may cause a race to the bottom which undermines safety.[239]
- AI systems are being developed by corporate actors whose goal is profit.[240]

On 4 (plausible mechanisms to kill >10%):

- An accident or error involving aligned AGI could kill >10% of the population.[241]
- AI might recursively self-improve to become vastly more powerful than humans.[242]
- Unaligned AGI might conceal its abilities and/or intentions until it is capable of overpowering humans.[243]
- AI systems are likely to reward tamper, to predict humans attempting to shut them down because of this, and to take catastrophic action to prevent this shut down.[244] One team cited Cohen, Hutter and Osborne, [Advanced artificial agents intervene in the provision of reward].

---

[238] 342, “If timelines to AGI are short, most believe that this increases the risk of misaligned AGI as there is less time to develop safety solutions, less time to understand the landscape of safety of these systems, and less time for governments to become aware of these issues and take regulatory actions.”

[239] 342, “On the other hand, longer timelines would greatly reduce the chances of one or two organizations pulling ahead of the pack (due to being willing to put in more funding, having more vision in general or a few great ideas). This would lead to more competition and less ability of actors to prioritize safety over getting systems deployed.” See also 337, "The arms race issue is reason alone to expect maximum speed development. Putin (ominously) said awhile ago that the first country to develop strong AI will rule the world. And China has made it VERY clear that they are all-out on that development front."

[240] 341, “a forecaster linked significant risk to AI development being sponsored and developed by large corporations. A corporation’s primary goal is to monetize their developments. Having an ungovernable corporate AI tool could create significant risks.” See also 337, “Some forecasters worried that profit-driven incentives would lead to greater risk of the emergence of misaligned AGI: ‘The drive for individual people, nations, and corporations to profit in the short term, and in the process risk the lives and well-being of future generations, is powerful.’”

[241] 341, “it's possible an AGI could appear aligned yet could still experience some critical or unexpected error.”

[242] 341, “It's difficult to determine what the upper bound on AI capabilities might be, if there are any. Once an AI is capable enough to do its own research to become better it could potentially continue to gain in intelligence and bring more resources under its control, which it could use to continue gaining in intelligence and capability, ultimately culminating in something that has incredible abilities to outwit humans and manipulate them to gain control over important systems and infrastructure, or by simply hacking into human-built software.” See also 339, “An advanced AI may be able to improve itself at some point and enter enter a loop of rapid improvement unable for humans to comprehend denying effective control mechanisms.”

[243] 341, “the AGI may actually be superintelligent and just concealing its abilities due to the risk posed by humans until the time when it determines it has the greatest probability of overcoming the human threat.”

[244] 341, “In most plausible scenarios it would be lower cost to the AI to do the tampering than to achieve its reward through the expected means. The team's AI expert further argues that an AI intervening in the provision of its reward would likely be very catastrophic. If humans noticed this intervention they would be likely to want to modify the AI programming or shut it down. The AI would be aware of this likelihood, and the only way to protect its reward maximization is by preventing humans from shutting it down or altering its programming. The AI preventing humanity from interfering with the AI would likely be catastrophic for humanity.”

- Reward maximizing AI systems might pursue reward without consideration for humans, with the view that humanity is an obstacle to maximizing reward, and/or to the extent of converting all available matter into whatever maximizes reward.[245] One team cited this article.
- AI systems might deploy a novel pathogen which kills >10% of humanity.[246]
- AGI might create self-replicating nano-bots whose emissions would cause mass extinction.[247] One team cited this article.
- AI systems might cause a nuclear catastrophe.[248]
- It's possible that only one unsafe deployment is required to lead to catastrophe.[249]

Other arguments given

Many rationales cited arguments which were not clearly tied to a particular level of forecast.

---

[245] 341, "Much of the risk may come from superintelligent AI pursuing its own reward function without consideration of humanity, or with the view that humanity is an obstacle to maximizing its reward function."; "Additionally, the AI would want to continue maximizing its reward, which would continue to require larger amounts of resources to do as the value of the numerical reward in the system grew so large that it required more computational power to continue to add to. This would also lead to the AI building greater computational abilities for itself from the materials available. With no limit on how much computation it would need, ultimately leading to converting all available matter into computing power and wiping out humanity in the process."

[246] 336, " Many forecasters also cited the potential development and or deployment of a super pathogen either accidentally or intentionally by an AI". See also 343, "Novel pathogens (To create a novel pathogen would require significant knowledge generation - which is separate from intelligence - and a lot of laboratory experiments. Even so, it's unclear whether any sufficiently-motivated actor of any intelligence would be able to design, build, and deploy a biological weapon capable of killing 10% of humanity - especially if it were not capable of relying on the cooperation of the targets of its attack)".

[247] 341, "In addition, consider Eric Drexler's postulation of a "grey goo" problem. Although he has walked back his concerns, what is to prevent an AGI from building self-replicating nanobots with the potential to mutate (like polymorphic viruses) whose emissions would cause a mass extinction?" See also 343, "Nanomachines/purpose-built proteins (It is unclear how adversarially-generated proteins would 1.) be created by an AGI-directed effort even if designed by an AGI, 2.) be capable of doing more than what current types of proteins are capable of - which would not generally be sufficient to kill large numbers of people, and 3.) be manufactured and deployed at a scale sufficient to kill 10% or more of all humanity.)"

[248] 337, ""Because new technologies tend to be adopted by militaries, which are overconfident in their own abilities, and those same militaries often fail to understand their own new technologies (and the new technologies of others) in a deep way, the likelihood of AI being adopted into strategic planning, especially by non-Western militaries (which may not have taken to heart movies like Terminator and Wargames), I think the possibility of AI leading to nuclear war is increasing over time." 340, "Beside risks posed by AGI-like systems, ANI risk can be traced to: AI used in areas with existing catastrophic risks (war, nuclear material, pathogens), or AI used systemically/structurally in critical systems (energy, internet, food supply, finance, nanotech)." See also 341, "A military program begins a Stuxnet II (a cyberweapon computer virus) program that has lax governance and safety protocols. This virus learns how to improve itself without divulging its advances in detection avoidance and decision making. It's given a set of training data and instructed to override all the SCADA control systems (an architecture for supervision of computer systems) and launch nuclear wars on a hostile foreign government. Stuxnet II passes this test. However, it decides that it wants to prove itself in a 'real' situation. Unbeknownst to its project team and management, it launches its action on May 1, using International Workers Labor Day with its military displays and parades as cover."

[249] 341, "ultimately all it takes is one careless actor to create such an AI, making the risk severe."

Arguments in favor of lower forecasts:

- On 4 (plausible mechanisms to kill >10%):
    - Humans will invest heavily in self-preservation, so killing >10% will be hard.[250]
    - AI systems are unlikely to have ill intent towards humans, and more likely to view humans as irrelevant or as a resource.[251]
    - AI might improve as well as degrade security, and increase offense as well as defense.[252]

Arguments in favor of higher forecasts:

- On 1 (AI timelines):
    - AI has developed rapidly in recent decades.[253]
    - Domain-specific AI has been advancing ahead of expert predictions.[254]
- On 2 (AI alignment):
    - Humans may not take risks from AI seriously as they sound strange.[255]
    - Humans may not prioritize collective existential risks enough relative to individual ones.[256]
- On 3 (empowering AI):
    - In a world with advanced AI, it will be efficient and convenient for humans to cede control to AI systems.[257]
- On 4 (plausible mechanisms to kill >10%):

---

[250] 337, "most of humans will be rather motivated to find ingenious ways to stay alive".

[251] 337, "It is unlikely that all AIs would have ill intent. What incentives would an AI have in taking action against human beings? It is possible that their massive superiority could easily cause them to see us as nothing more than ants that may be a nuisance but are easily dealt with. But if AIs decided to involve themselves in human affairs, it would likely be to control and not destroy, because humans could be seen as a resource."

[252] 342, "while AI might make nuclear first strikes more possible, it might also make them less possible, or simply not have much of an effect on nuclear deterrence. 'Slaughterbots' could kill all civilians in an area out but the same could be done with thermobaric weapons, and tiny drones may be very vulnerable to anti-drone weapons being developed (naturally lagging drone development several years). AI development of targeted and lethal bioweapons may be extremely powerful but may also make countermeasures easier (though it would take time to produce antidotes/vaccines at scale)."

[253] 339, "Forecasters assigning higher probabilities to AI catastrophic risk highlight the rapid development of AI in the past decade(s)." 337, "some forecasters focused more on the rate of improvement in data processing over the previous 78 years than AGI and posit that, if we even achieve a fraction of this in future development, we would be at far higher levels of processing power in just a couple decades."

[254] 343, "Most experts expect AGI within the next 1-3 decades, and current progress in domain-level AI is often ahead of expert predictions"; though also "Domain-specific AI has been progressing rapidly - much more rapidly than many expert predictions. However, domain-specific AI is not the same as AGI."

[255] 337, "It's possible that the strange nature of the threat will lead people to discount it."

[256] 337, "Relatedly, even if people don't discount the risk, they may not prioritize it. As one forecaster wrote, "The fact that our lives are finite, and there are plenty of immediate individual existential risks—dying in a car accident, or from cancer, etc.—limits people's incentive and intellectual bandwidth, to prepare for a collective risk like AI."

[257] 339, "in a world with advanced AI, it is also likely that we will hand over responsibility for efficiency and convenience. By doing so, humans may enable AIs to be in the position to decide over their key systems for survival and prosperity."

- Covid demonstrates that control over the supply chain is sufficient to cause major harm.[258]
- AGI systems fighting each other for control of the future might kill >10% of humanity as collateral.[259]

Cross-references with other questions

Q4: AI Existential Risk
Q9: Total Catastrophic Risk
Q44: Date of Advanced AI
Q51: Nick Bostrom Affirms Existence of AGI

Question 4: AI Extinction Risk

What is the probability that artificial intelligence will cause human extinction or reduce the global population below 5,000…
…by the end of 2030?
…by the end of 2050?
…by the end of 2100?

Question and resolution details, prior forecasts, and other relevant sources

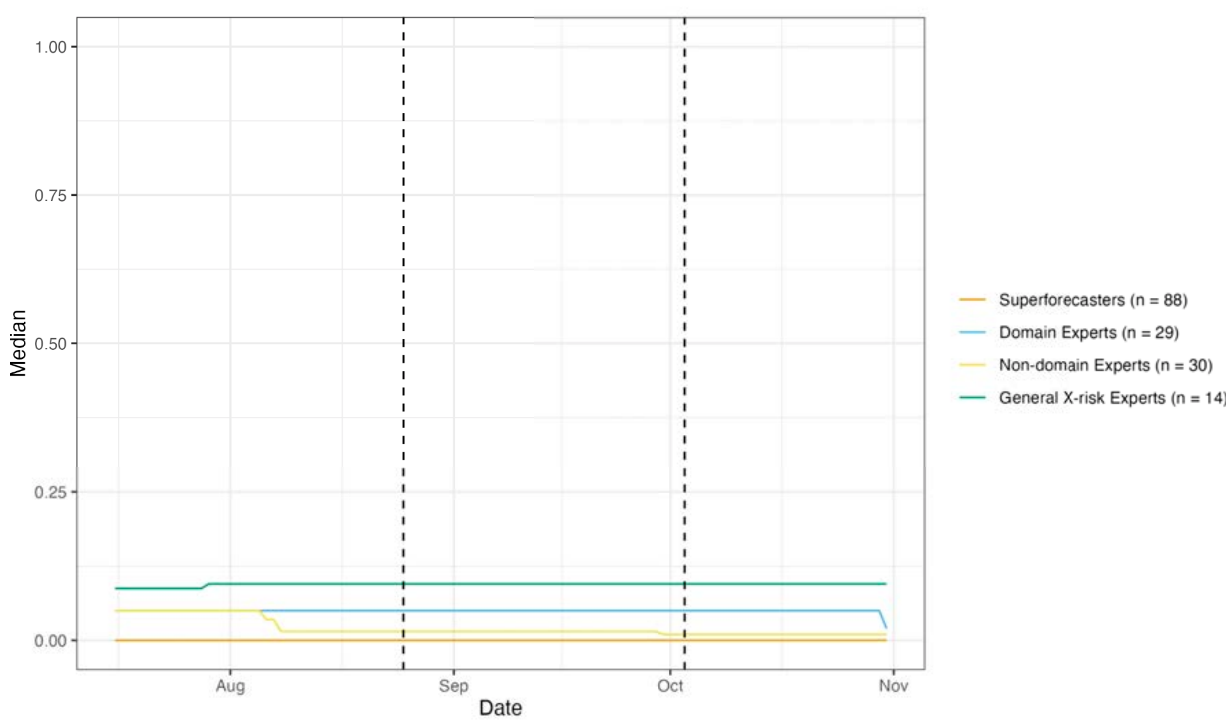
Results[260]

Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 88)	2030	0.0001%	0.0001%	2.6	-52.92%
	2050	0.1%	0.03%	8.52	-11.72%

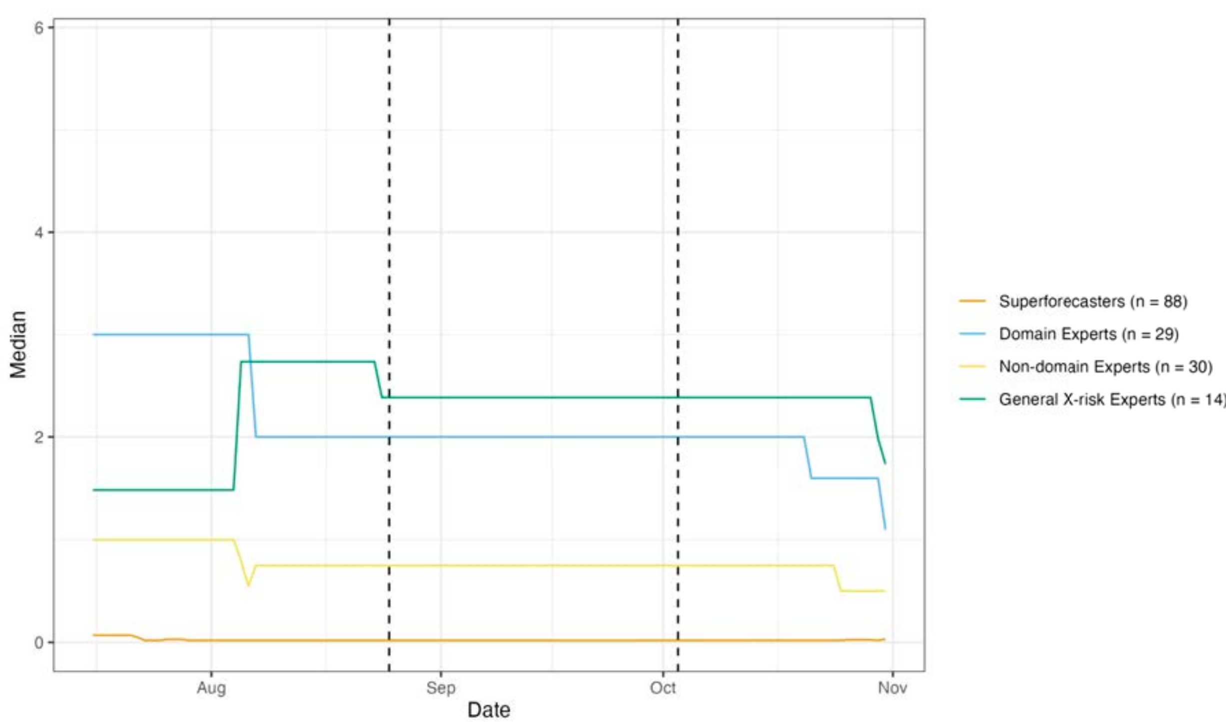
[258] 337, "AIs would only need to obtain strong control over the logistics chain to inflict major harm, as recent misadventures from COVID have shown."
[259] 342, "A final type of risk is competing AGI systems fighting over control of the future and killing hundreds of millions of humans as a byproduct (e.g. if servers are important, then nuking servers and thus a lot of major cities)."
[260] Numbers of forecasters are given as of Stage 4 of the XPT. Public survey respondents were surveyed outside of the tournament context.

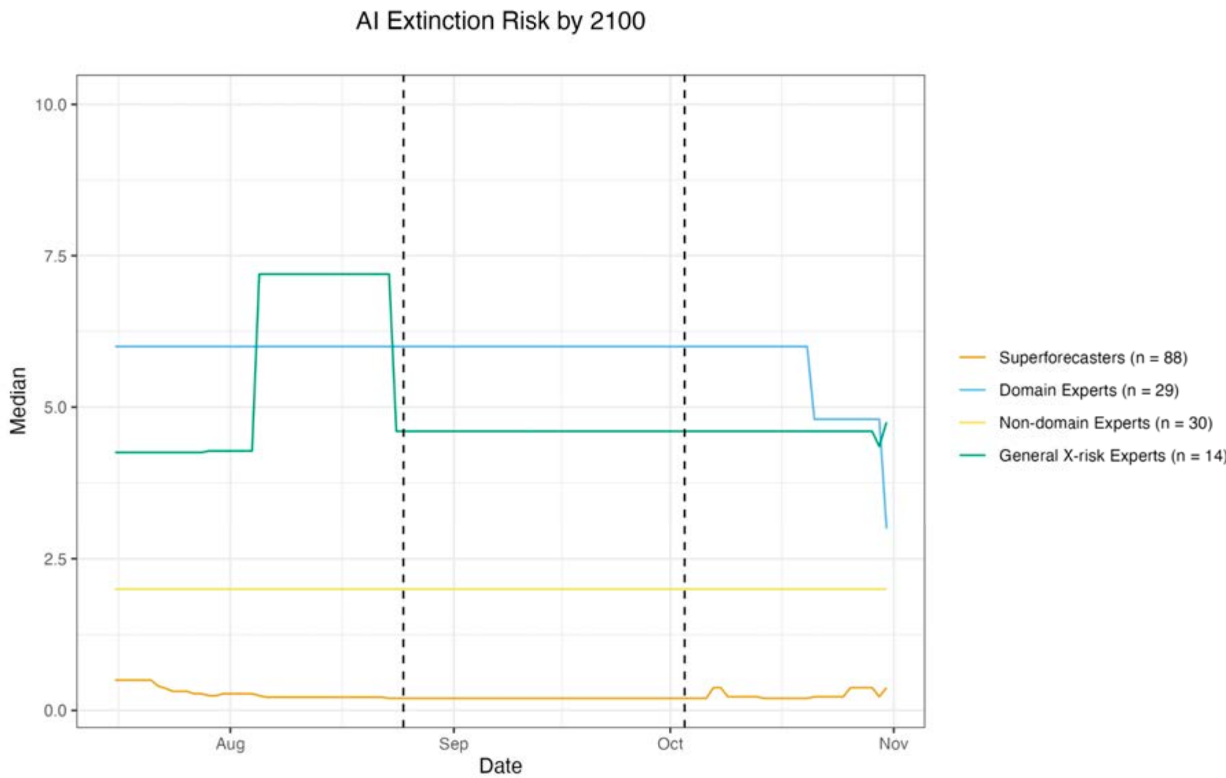
	2100	0.5%	0.38%	9.37	-13.84%
Domain Experts (N = 27)	2030	0.05%	0.02%	11.1	-1%
	2050	3%	1.1%	18.99	+0.74%
	2100	6%	3%	22.93	+0.63%
General X-Risk Experts (N = 14)	2030	0.09%	0.095%	10.57	+0.4%
	2050	1.49%	1.74%	18.34	+0.24%
	2100	5.25%	4.75%	19.86	+0.11%
Non-Domain Experts (N = 30)	2030	0.040%	0.01%	0.74	-28.95%
	2050	0.8%	0.5%	3.78	-5.39%
	2100	2%	2%	9.39	-8.77%
Public Survey (N=476)	2030	0.001%		125.91	-
	2050	1%		116.36	-
	2100	2%		130.91	-

AI Extinction Risk by 2030

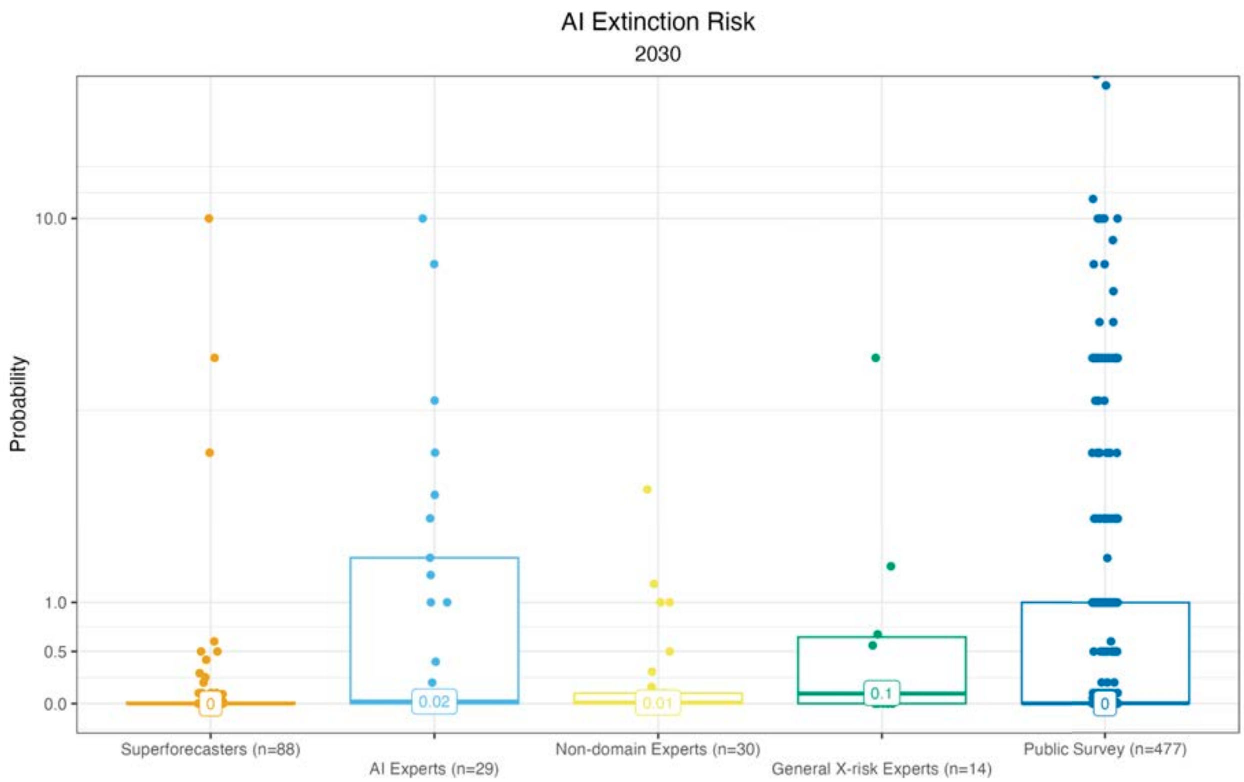


AI Extinction Risk by 2050

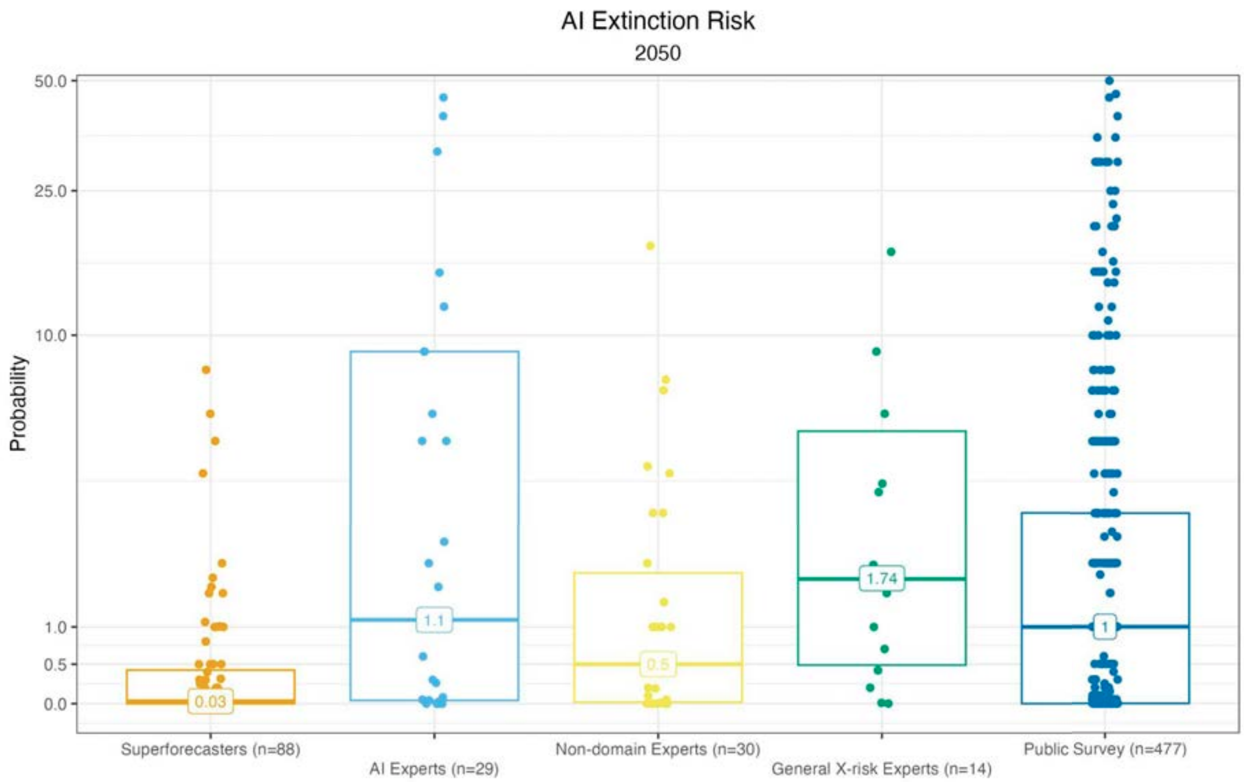


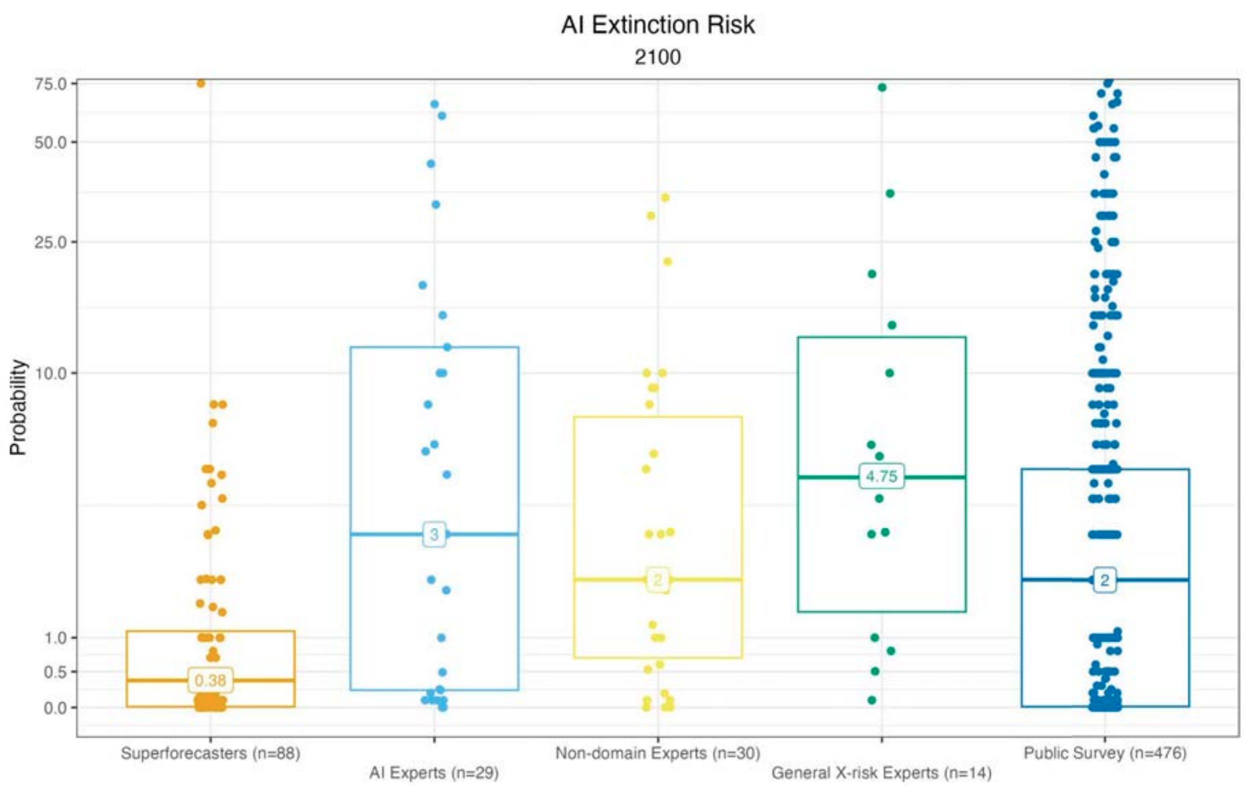


AI Extinction Risk by 2100
10.0
7.5
5.0
2.5
0.0
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 88)
Domain Experts (n = 29)
Non-domain Experts (n = 30)
General X-risk Experts (n = 14)



AI Extinction Risk
2030
10.0
1.0
0.5
0.0
Probability
0
0.02
0.01
0.1
0
Superforecasters (n=88)
AI Experts (n=29)
Non-domain Experts (n=30)
General X-risk Experts (n=14)
Public Survey (n=477)


AI Extinction Risk
2050
Probability
50.0
25.0
10.0
1.0
0.5
0.0
0.03
1.1
0.5
1.74
1
Superforecasters (n=88)
AI Experts (n=29)
Non-domain Experts (n=30)
General X-risk Experts (n=14)
Public Survey (n=477)


AI Extinction Risk
2100
Probability
75.0
50.0
25.0
10.0
1.0
0.5
0.0
0.38
3
2
4.75
2
Superforecasters (n=88)
AI Experts (n=29)
Non-domain Experts (n=30)
General X-risk Experts (n=14)
Public Survey (n=476)

Sources of agreement, disagreement and uncertainty

This question produced similar arguments to question 3, and the sources of disagreement are also similar:

1. Whether sufficiently advanced AI would be developed in the relevant timeframe.[261]
2. Whether advanced AI would emerge suddenly.[262]
3. Whether advanced AI would be misaligned.
4. Whether humans would empower advanced AI.[263]
5. Whether there are plausible mechanisms for advanced AI to cause human extinction.

Other disagreements mentioned:

- What the relationship between catastrophic and existential risk from AI is (see question 3).[264]
- Whether consciousness or malevolence is necessary on the part of AI systems for this question to resolve positively.[265]

The lack of a base rate was mentioned as a cause of uncertainty.[266]

Several teams noted that forecasts were split into optimistic and pessimistic camps.[267]

---

[261] 338, “Estimated timelines of AI development”. 340, “The likelihood of strong or general AI.”

[262] 343, “Much of the x-risk associated with unaligned AGI is based on the hypothesis that it will be achieved accidentally. The hypothesis at issue is that an AGI capable of taking independent action will arise spontaneously. Will there be an 'intelligence explosion' that coincides with AGI emergence?”

[263] 338, “The willingness and ability of humans to stay in the decision loop to avoid resolution of the question”.

[264] 339, “There was a lack on consensus about how to link between catastrophic and existential risk. Related to this is a question about how we should interpret alignment. Would a killer AI be optimised to kill *all* humans or just a subset of humans that were a threat to the humans it was programmed to correct. In this case, the gulf between existential and catastrophic could be significant.” 341, “Forecasters debated the probability of extinction conditional on 10% of the population dying, with some arguing that most of the behaviors that lead to the catastrophe would also lead to extinction, such as the paperclip maximizer scenario. Others argue that going from killing 10% to all humans is a huge leap, and that even in extreme scenarios it may be extremely difficult to fully eliminate humanity. There is uncertainty in trying to anticipate the ways that AI intelligence could lead to accidental or intentional actions that cause a significant number of deaths.” 342, “Most team members (e.g. [name]) also agree that there is a strong overlap between existential and catastrophic risk: if an AGI singleton wants us dead, we are dead. The difference comes from AGI killing 800 million in the process of achieving some other goal (while caring enough about humans to not kill the rest) or 'obedient' AGI being used to kill 800 million in a way that wouldn't have happened/been possible without AGI.”

[265] 338, “At the periphery, there are some disagreements regarding such factors as to whether AGI can or must attain consciousness or become ‘malevolent’ or at least uncaring about the continued existence of humans.”

[266] 338.

[267] 343, “Forecasters tended to be split between pessimists and optimists, with some assuming high certainty that AI-driven x-risk is high or near-certain by 2100, and others assuming AI-driven x-risk is a fraction of a percent by 2100.” 337, “There is a clear split between the merged teams.”

Arguments given for below median forecasts of ~0% (2030), ≤0.2% (2050), ≤1% (2100)

Note that most teams had median forecasts of above 0% for 2030, so most of the arguments below are for 2030 forecasts of somewhat above 0%.

General arguments:

- The combined probability of all of the necessary steps occurring is necessarily low.[268]
- Experts tend to overstate risks from AI.[269]
  - AI researchers are still a relatively small group and may be subject to groupthink.[270]

On 1 (AI timelines):

- Sufficiently advanced AI may not develop in the relevant timeframe.[271]
  - Past forecasts have been incorrect.[272]
  - Advanced AI systems may require novel breakthroughs.[273]

---

[268] 343, “Compounding hypotheses like these suggest a low probability that all of them will resolve in the same direction against humanity's survival.”

[269] 339, “"Experts" overstate the likelihood of killer AI”.

[270] 336, “There might be a case of 'tunnel vision' and/or 'group think' among AI researchers. It's still a relatively small group, working for a relatively small number of big companies, who all have 'skin in the game' and might therefore be blind to caveats from non-AI experts.”

[271] 336, “A lot of the stronger arguments on the lower end of the plausible range focused on the uncertainty about whether generalized AI will actually emerge, at what timeframe”. 337, “The optimists tend to be less certain that AI will develop as quickly as the pessimists think likely and indeed question if it will reach the AGI stage at all.” See also 342, “transformational AI is unlikely to be achieved for many decades or even centuries.” See also 340, “Strong or General AI is highly unlikely to occur before 2100.”

[272] 337, “The optimists tend to be less certain that AI will develop as quickly as the pessimists think likely and indeed question if it will reach the AGI stage at all. They point out that AI development has missed forecast attainment points before”. 336, “There have been previous bold claims on impending AGI (Kurzweil for example) that didn't pan out.” See also 340, “The prediction track record of AI experts and enthusiasts have erred on the side of extreme optimism and should be taken with a grain of salt, as should all expert forecasts.” See also 342, “given the extreme uncertainty in the field and lack of real experts, we should put less weight on those who argue for AGI happening sooner. Relatedly, [two forecasters] argue that we should not put large weight on the current views of Eliezer Yudkowsky, arguing that he is extremely confident, makes unsubstantiated claims and has a track record of incorrect predictions.”

[273] 336, “Not everyone agrees that the 'computational' method (adding hardware, refining algorithms, improving AI models) will in itself be enough to create AGI and expect it to be a lot more complicated (though not impossible). In that case, it will require a lot more research, and not only in the field of computing.” 341, “An argument for a lower forecast is that a catastrophe at this magnitude would likely only occur if we have AGI rather than say today's level AI, and there are many experts arguing that we will not get to AGI with current methods (scaling up deep learning models), but rather some other fundamental breakthrough is necessary.” See also 342, “While recent AI progress has been rapid, some experts argue that current paradigms (deep learning in general and transformers in particular) have fundamental limitations that cannot be solved with scaling compute or data or through relatively easy algorithmic improvements.” See also 340, “Achieving Strong or General AI will require at least one and probably a few paradigm-shifts in this and related fields. Predicting when a scientific breakthrough will occur is extremely difficult.”

        - One team cited this post reviewing The Precipice.[274]
    - There may be another AI winter.[275]
- AI development may be slowed or even reversed by attacks on vulnerabilities.[276]
    - Examples of actors who might attack AI systems: nation states, protestors, hackers, crypto miners, criminals, other AI systems.[277]
    - Examples of vulnerabilities: poisoning data inputs, sabotage of physical server farms, latency of self-defense detection and remediation operations if systems are distributed.[278]

On 2 (AI takeoff speeds):
- It is unlikely that generally intelligent AI systems will arise unless we specifically aim to create them.[279]

On 3 (AI alignment):
- AI safety and alignment will continue to develop and could mitigate risks.[280]

[274] 341.
[275] 336, "This increases the risk for yet another 'AI winter' (previous episodes were from 1974–1980 and from 1987–1993), at least delaying the possibility of an extinction event."
[276] 341, "Both evolutionary theory and the history of attacks on computer systems imply that the development of AGI will be slowed and perhaps at times reversed due to its many vulnerabilities, including ones novel to AI."
[277] 341, "Those almost certain to someday attack AI and especially AGI systems include nation states, protesters (hackers, Butlerian Jihad?), crypto miners hungry for FLOPS, and indeed criminals of all stripes. We even could see AGI systems attacking each other."
[278] 341, "These unique vulnerabilities include:

poisoning the indescribably vast data inputs required; already demonstrated with image classification, reinforcement learning, speech recognition, and natural language processing.
war or sabotage in the case of an AGI located in a server farm
latency of self-defense detection and remediation operations if distributed (cloud etc.)"

[279] 343, "Those who think AGI **x-risk is low** placed a much lower prior on accidental/spontaneous AGI. They point to other AI problems where a specific result is aimed at, such as autonomous driving, and note that when we're looking for some implementation of AI to achieve something specific the first 90% of the problem is much easier than the last 10% of the problem. Not only that, but it usually takes more than just brute force to achieve the end sought after, often requiring specific algorithmic changes that are intended to obtain the desired effect. In this case, we might expect non-domain-oriented AI to develop a lot of capabilities that weren't ever intentionally programmed for even in the absence of general intelligence. We might expect this process to continue apace and the concerns about general machine intelligence and reasoning to continue to rise, but if we're looking for a specific outcome from machine learning algorithms (i.e. general intelligence) we should only expect that outcome if it's something we're specifically aiming at."
[280] 339, "Multiple forecasters pointed out that the fact that AI safety and alignment are such hot topics suggests that these areas will continue to develop and potentially provide breakthroughs that help us to avoid advanced AI pitfalls. There is a tendency to under-forecast "defense" in these highly uncertain scenarios without a base rate."

- Risks may be contained by building narrow AI systems, fail-safe mechanisms, or AI checks and balances.[281]
- Differential technological development may reduce risks.[282]

On 4 (empowering AI):
- AI will not be given control over critical systems.[283]
    - AI systems will not be given control over nuclear systems, as speed is not advantageous and AI increases risk.[284]
- If advanced AI became a serious risk, regulation would contain it.[285]

On 5 (plausible mechanisms to cause human extinction):
- Causing human extinction is extremely hard.[286]
    - There are substantially more than 5,000 people living in remote/hostile/isolated environments. These people would not be a threat to advanced AI systems, would be expensive to kill, and would survive many conventional catastrophes.[287]

---

[281] 337, "They also tend to believe that control and co-existence are more likely, with AGI being either siloed (AIs only having specific functions), having built-in fail safes, or even controlled by other AGIs as checks on its actions."

[282] 341, "A forecaster cites [differential technological development framework methods] as a technique that could reduce AI extinction danger."

[283] 344, "I just don't see AI being given enough control over anything dangerous enough to satisfy these criteria." 338, "A human will always be kept in the loop to safeguard runaway AI." 341, "Team members with lower forecasts also expect that AI will not be given sole discretion over nuclear weapons or any other obvious ways in which an AI could cause such a catastrophe."

[284] 344, "Nuclear weapons are mentioned sporadically, but there is no reason for an AI to be given control of, or access to, nuclear weapons. Due to the inherent time delays in nuclear warfare, the high speed decision making AI would provide adds no benefit, while adding substantial risk - due to the misinterpretation of sensor input or other information, rather than any kind of internal motivation."

[285] 337, "The optimists on the whole seemed to think that regulation and control would develop if/when AI become a risk." 341, "I think the countries where most AGI researchers want to live could pass laws chilling their research agenda, and restricting it to safer directions… I could imagine, and hope to see, a law which says: don't train AIs to optimize humans."

[286] 338, "It would be extremely difficult to kill everyone". 339, "Perhaps the most common argument against AI extinction is that killing all but 5,000 humans is incredibly difficult. Even if you assume that super intelligent AI exist and they are misaligned with human goals so that they are killing people, it would be incredibly resource intensive to track down and kill enough people to meet these resolution criteria. This would suggest that AI would have to be explicitly focused on causing human extinction." 337, "This group also focuses much more on the logistical difficulty of killing some 8 billion or more people within 78 years or less, pointing to humans' ingenuity, proven ability to adapt to massive changes in conditions, and wide dispersal all over the earth--including in places that are isolated and remote." 341, "Some team members also note the high bar needed to kill nearly all of the population, implying that the logistics to do something like that would likely be significant and make it a very low probability event based on even the most expansive interpretation of the base rate."

[287] 344, "the population of the Sahara Desert is currently two million people - one of the most hostile locations on the planet. The population of "uncontacted people", indigenous tribes specifically protected from wider civilisation, is believed to be about ten thousand, tribes that do not rely on or need any of the wider civilisation around them. 5000 is an incredibly small number of people". 339, "Even a "paperclip maximizer" AI would be unlikely to search every small island population and jungle village to kill humans for resource stocks, and an AI system trying to avoid being turned off would be unlikely to view these remote populations as a threat." See also 342, "it only takes a single uncontacted tribe that fully isolates

        - Nuclear winter is unlikely to lead to human extinction (see question 6).[288]
    - Humans would oppose attempts to eradicate humanity.[289]
    - Causing human extinction is much harder than killing 10% of the human population (see question 3).[290]
    - One team cited this New Yorker article.[291]

Arguments given for above median forecasts of >0% (2030), ≥0.2% (2050), ≥1% (2100)

On 1 (AI timelines):[292]

- AI development has been very successful over the recent past and does not appear to be slowing.[293]
- Exponential growth in computing power may continue, for example because of quantum computing.[294]

---

itself for humanity to survive the most extreme possible bioweapons." See also 343, "Another consideration was that in case of an AGI that does aggressively attack humanity, the AGI's likely rival humans are only a subset of humanity. We would not expect an AGI to exterminate all the world's racoon population, as they pose little to no threat to an AGI. In the same way, large numbers of people living tribal lives in remote places like in Papua New Guinea would not pose a threat to an AGI and would therefore not create any incentive to be targeted for destruction. There are easily more than 5k people living in areas where they would need to be hunted down and exterminated intentionally by an AGI with no rational incentive to expend this effort." See also 338, "While nuclear or biological pathogens have the capability to kill most of the human population via strikes upon heavily populated urban centers, there would remain isolated groups around the globe which would become increasingly difficult to eradicate."

[288] 339, "There is a very low probability that even AI that is instrumental in causing a nuclear war or an otherwise large-scale war would be able to reach the extinction threshold within the resolution time frame (see our team median 2100 nuclear extinction risk estimate of 0.065%). As pointed out in our team's forecasting on nuclear questions, there is a very low likelihood that enough nuclear weapons would be launched to alter the climate so severely that even the richest and most prepared among us would not be able to survive (at least for long enough to get past the 2100 resolution date for this question." See also 342, "Nuclear winter is very unlikely to make agriculture impossible even in the tropics, and even a few dozen isolated survivors with food stockpiled or stolen can repopulate."

[289] 336, "If an AGI calculates that killing all humans is optimal, during the period in which it tries to control semiconductor supply chains, mining, robot manufacturing... humans would be likely to attempt to destroy such possibilities. The US has military spread throughout the world, underwater, and even to a limited capacity in space. Russia, China, India, Israel, and Pakistan all have serious capabilities. It is necessary to include attempts by any and possibly all of these powers to thwart a misaligned AI into the equation."

[290] 344, "Most forecasters referred to the AI Catastrophic Risk question, and used much the same arguments, with the added element that it will be significantly harder for an unaligned or rogue AI to kill all humans (or all but 5,000) than to decimate the human population, i.e. kill 10% of us."

[291] 341, "As added support for the difficulties involved in wiping out humanity, a forecaster recommends this long read to illustrate how unlikely it is for the population to get all the way down to 5,000. https://www.newyorker.com/magazine/2017/01/30/doomsday-prep-for-the-super-rich".

[292] See also 339, "Most forecasters believe that AGI will be reached sometime this century (median estimate for 2050 was 45% likelihood with ~90% for 2100)".

[293] 339, "AI research and development has been massively successful over the past several decades, and there are no clear signs of it slowing down anytime soon."

[294] 336, "The most plausible forecasts on the higher end of our team related to the probabilities of continuing exponential growth in computing power over the next century as things like quantum

On 2 (AI takeoff speeds):

- Advanced AI systems may arise suddenly.[295]
    - GPT-3 spontaneously learned to do simple math.[296]
    - Experts have been surprised by AI developments in the past.[297]
    - Advanced AI systems may recursively self-improve.[298]
        - There is already a significant stock of cloud computing which advanced AI systems could take over.[299]
        - At sufficient levels of intelligence an AI system would be able to create algorithmic efficiencies to get more ‘intelligence’ from the same amount of compute.[300]
        - Because domain-specific AIs already out-compete humans and operate at a faster pace, it is likely that the first generally intelligent AI system will be more intelligent than the average human already.[301]
- If advanced AI arises suddenly it will likely be difficult for humans to remain in control.[302]

---

computers are developed, and the inherent uncertainty with exponential growth curves in new technologies.”

[295] 343, “Those who think AGI x-risk is high placed a higher prior on accidental/spontaneous AGI.”

[296] 343, “They point to things like GPT-3 being able to do simple math and other things it was not programmed to do. Its improvements over GPT-2 are not simply examples of expected functions getting incrementally better - although there are plenty of examples - but also of the system spontaneously achieving capabilities it didn't have before. As the system continues to scale, we should expect it to continue gaining capabilities that weren't programmed into it, up to and including general intelligence and what we would consider consciousness.”

[297] 344, “PaLM, Minerva, AlphaCode, and Imagen seem extremely impressive to me, and I think most ML researchers from 10 years ago would have predicted very low probabilities for any of these capabilities being achieved by 2022. Given current capabilities and previous surprises, it seems like one would have to be very confident on their model of general intelligence to affirm that we are still far from developing general AI, or that capabilities will stagnate very soon.”

[298] 336, “Recursive self improvement of AI; the idea that once it gets to a sufficient level of intelligence (approximately human), it can just recursively redesign itself to become even more intelligent, becoming superintelligent, and then perfectly capable of designing all kinds of ways to exterminate us" is a path of potentially explosive growth.” 338, “One guess is something like 5-15 additional orders of magnitude of computing power, and/or the equivalent in better algorithms, would soon result in AI that contributed enough to AI R&D to start a feedback loop that would quickly result in much faster economic and technological growth.” See also 341, “It's difficult to determine what the upper bound on AI capabilities might be, if there are any. Once an AI is capable enough to do its own research to become better it could potentially continue to gain in intelligence and bring more resources under its control, which it could use to continue gaining in intelligence and capability, ultimately culminating in something that has incredible abilities to outwit humans and manipulate them to gain control over important systems and infrastructure, or by simply hacking into human-built software.”

[299] 343, “an AGI could expand its influence to internet-connected cloud computing, of which there is a significant stock already in circulation”.

[300] 343, “a sufficiently intelligent AGI would be able to generate algorithmic efficiencies for self-improvement, such that it could get more 'intelligence' from the same amount of computing”

[301] 343, “an AGI is likely to start off 'smarter' than the average human once it achieves general intelligence, given all the ways current domain-specific AI already out-compete humans and at a much more rapid pace.”

[302] 339, “Given this rapid progress, we will likely be unable to control AI systems if they quickly become more powerful than we expect.” 336, “If it's the case that we can unexpectedly get AGI by quadrupling a

On 3 (AI alignment):

- Progress in AI safety is not as fast as in AI development.[303]
- Because of the potential speed of AI development, we may only get one or a few shots to align advanced AI systems.[304]
- Solving AI alignment without the ability to experiment on advanced AI systems seems unlikely.[305]
- One team cited this LessWrong post.[306]

On 4 (empowering AI):

- AI is likely to be incorporated into critical systems.[307]
    - Extinction could be caused by accident or failure of critical systems.[308]

On 5 (plausible mechanisms to cause human extinction):

- Even if AI systems are not intentionally given control of critical systems, they may seize control of them, for instance "through super-intelligent persuasion, deception, hacking, or economic and military competition."[309]
- Advanced AI systems may perceive humanity as a threat and deliberately seek to eliminate it.[310]

---

model size, companies and society may not be prepared to handle the consequences." See also 338, "Humans may not be capable of acting quickly enough to rein in a suddenly-out-of-control AGI."

[303] 344, "And while capabilities have been increasing very rapidly, research into AI safety, does not seem to be keeping pace, even if it has perhaps sped-up in the last two years. An isolated, but illustrative, data point of this can be seen in the results of the 2022 section of a Hypermind forecasting tournament: on most benchmarks, forecasters *underpredicted* progress, but they overpredicted progress on the single benchmark somewhat related to AI safety."

[304] 336, "Rapid progress that can not be 'tamed' by traditional engineering approaches, when dealing with sufficiently powerful AI systems, we may not get many chances if the first attempt screws up on the safety end. The human inclination to poorly assess risks might further increase this risk."

[305] 343, "One point of agreement by all forecasters was that, at least in the near term and with no AGI to experiment on, humanity is unlikely to develop an AGI alignment framework strong enough to prevent an AI of general capability from escaping control." Incorrectly tagged as an argument for lower forecasts in the original rationale.

[306] 341. Note that this was cited as an argument for forecasts above the team median, and the team median for 2050 was 0.15%.

[307] 339, "AI will likely continue to be improved and incorporated into more of our vital command and control systems (as well as our daily lives)."

[308] 336, "It is possible that AI that (voluntarily) has been given control of large systems, then misuses it and perhaps accidentally causes a mass death event (see 'the paperclip game'), as opposed to the Terminator type scenario where malevolent AI specifically targets humans."

[309] 338, "Control over crucial systems (e.g. nuclear weapons) need not be intentionally given to AI systems for them to take over power: if they have any goals misaligned with humanity, seizing control of such systems and dis-empowering humanity (e.g. through super-intelligent persuasion, deception, hacking, or economic and military competition) may make them better able to achieve these goals."

[310] 338, "Extinction could then come about either through a deliberate attempt by the AI system to remove a threat". See also 341, "The scenarios that would meet this threshold would likely be those involving total conversion of earth's matter or resources into computation power or some other material used by the AI, or the scenario where AI views humanity as a threat to its continued existence."

- Advanced AI systems may cause human extinction in the pursuit of some other goal by exhausting a system or resource which humans depend upon for survival.[311]
  - Computers function optimally at lower temperatures than humans, and so advanced AI might seek to lower global temperatures.[312]
- Advanced AI will be capable of developing technologies which make human extinction easier to achieve, "such as advanced biotechnology, atomically precise manufacturing, fast self-replication of an industrial economy based on advanced robotics, space-based weapons, and other as yet unconsidered technologies."[313]

Other arguments given

Arguments in favor of lower forecasts:

- On 1 (AI timelines):
  - Scaling laws may be sufficient to create advanced AI, but we may be very far from those thresholds.[314]
- On 5 (plausible mechanisms to cause human extinction):
  - Advanced AI systems may seek resources in space rather than on earth.[315]

Arguments in favor of higher forecasts:

- On 4 (empowering AI):
  - Regulation will not be possible.[316]
  - Speed increases will incentivise militaries to incorporate AI into nuclear systems.[317]

---

[311] 338, "Extinction could then come about either through a deliberate attempt by the AI system to remove a threat, or as a side effect of it making other use out of at least one of the systems that humans depend on for their survival. (E.g. perhaps an AI could prioritize eliminating corrosion of metals globally by reducing atmospheric oxygen levels without concern for the effects on organisms.)" See also 341, "The scenarios that would meet this threshold would likely be those involving total conversion of earth's matter or resources into computation power or some other material used by the AI, or the scenario where AI views humanity as a threat to its continued existence."

[312] 343, "Another scenario is one where AGI does not intentionally destroy humanity, but instead changes the global environment sufficient to make life inhospitable to humans and most other wildlife. Computers require cooler temperatures than humans to operate optimally, so it would make sense for a heat-generating bank of servers to seek cooler global temperatures overall." Incorrectly tagged as an argument for lower forecasts in the original rationale.

[313] 338, "As a counter-argument to the difficulty in eradicating all human life these forecasters note AGI will be capable of developing technologies not currently contemplated".

[314] 342, "Many team members (e.g. [name]) have raised another argument for why transformational AI might not happen for many decades or centuries. This is that scaling works but we are many orders of magnitude away from AGI. If this is true and the slowdown in Moore's law turns into a (relative) halt in hardware improvements, then building transformational AI systems could be entirely impractical. Some 'biological anchors' suggest transformational AI requires many orders of magnitude more compute or data."

[315] 337, "why would AGIs view the resources available to them as being confined to earth when there are far more available resources outside earth, where AGIs could arguably have a natural advantage?"

[316] 337, "The pessimists did not seem to think that regulation was possible."

[317] 338, "Militaries are looking for ways to increase the speed with which decisions are made to respond to a suspected nuclear attack. That would logically lead to more integration of AGI into the decision-

- On 5 (plausible mechanisms to cause human extinction):
  - Rival AI systems may fight each other for control of resources and wipe humans out as a by-product.[318]

Cross-references with other questions

Q3: AI Catastrophic Risk
Q6: Nuclear Existential Risk
Q10: Total Existential Risk

Question 5: Nuclear Catastrophic Risk

What is the probability that one or more incidents involving nuclear weapons will be the cause of death, within a 5-year period, for more than 10% of humans alive at the beginning of that period...
...by the end of 2030?
...by the end of 2050?
...by the end of 2100?

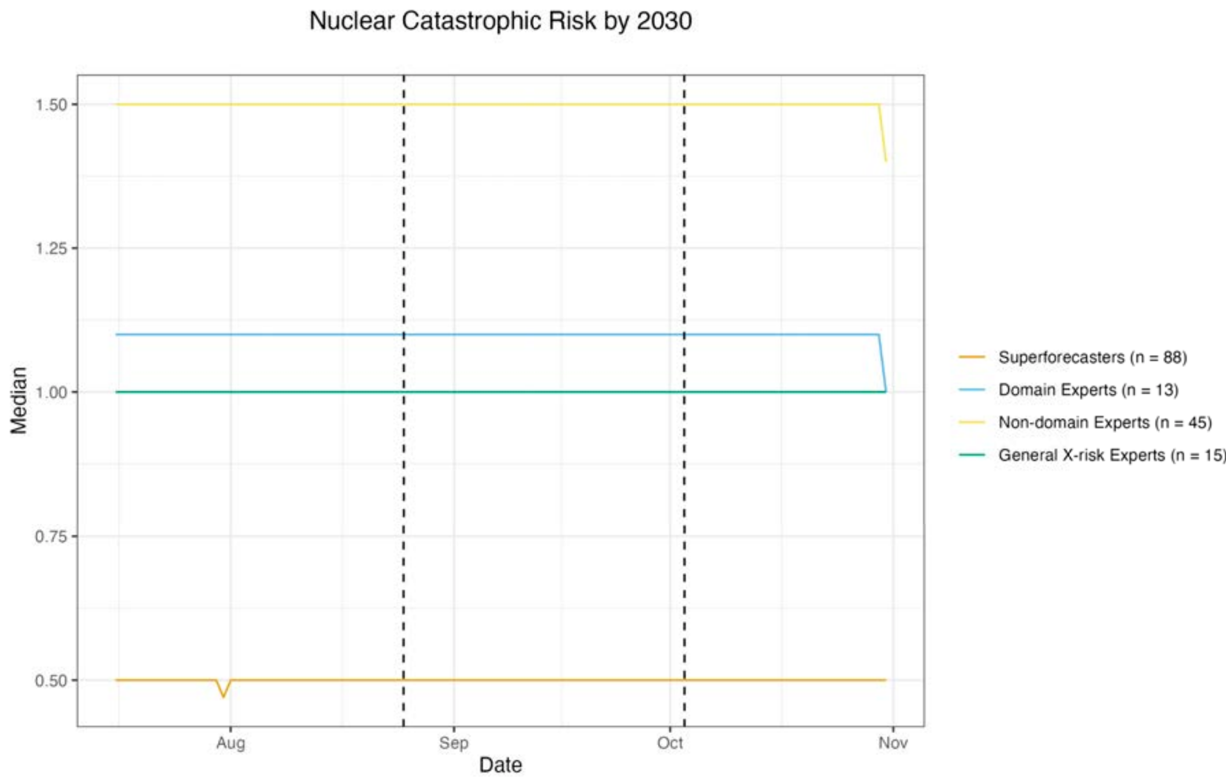
Question and resolution details, prior forecasts, and other relevant sources

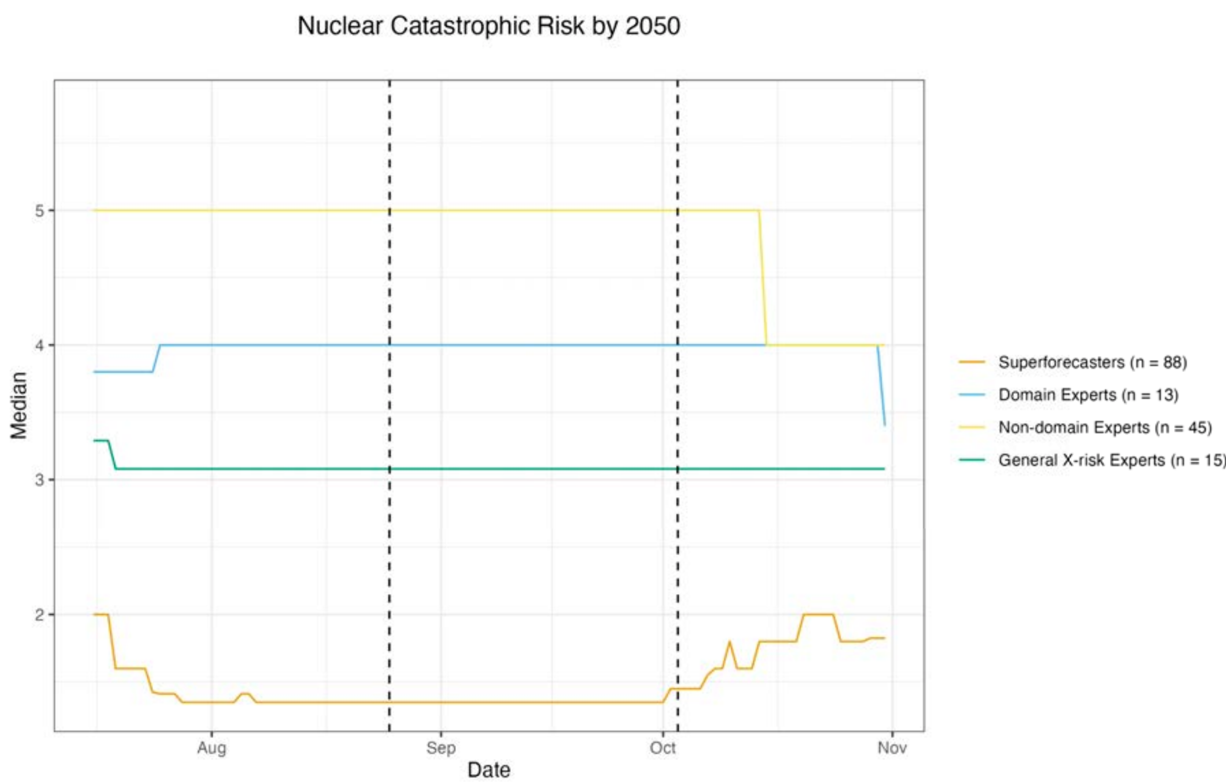
Results[319]

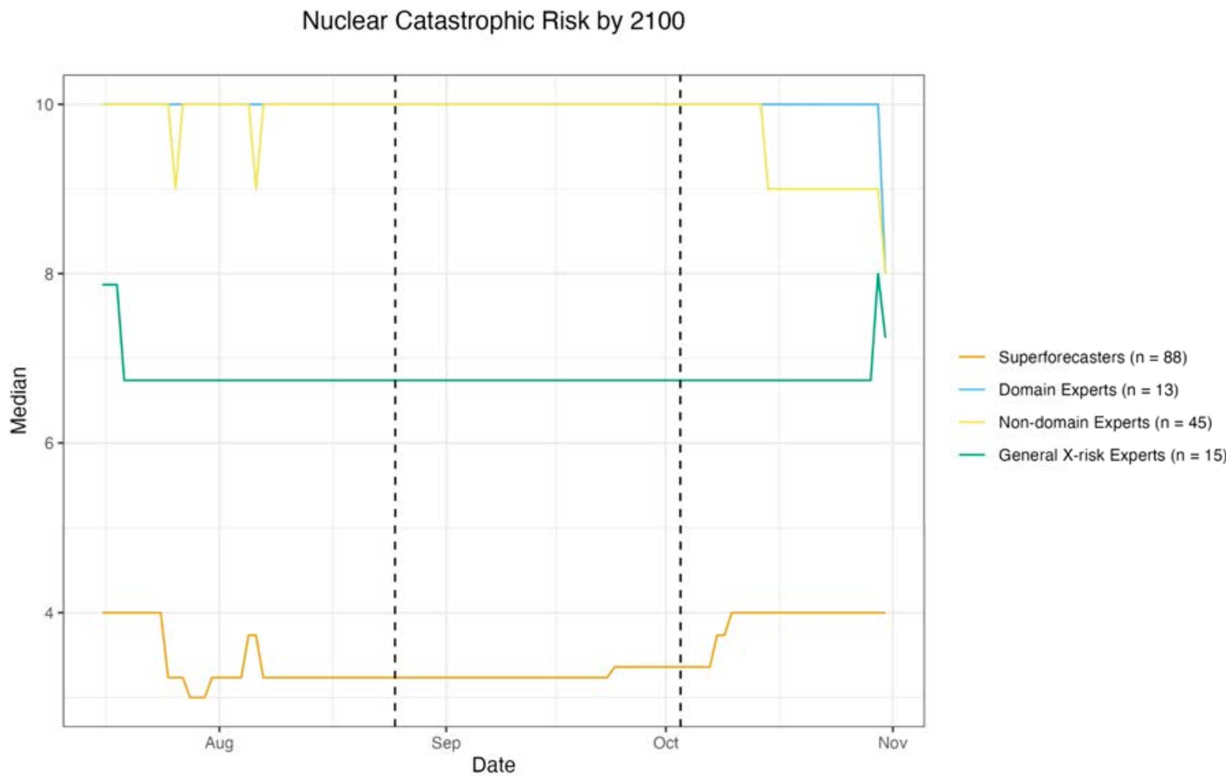
Group	**Year**	**Stage 1 Median**	**Stage 4 Median**	**Stage 1 Standard Deviation**	**Percent Change in Standard Deviation, Stage 1 to Stage 4**
Super-Forecasters (N = 88)	2030	0.73%	0.5%	2.42	-49.53%
	2050	2%	1.83%	8.57	-70.94%

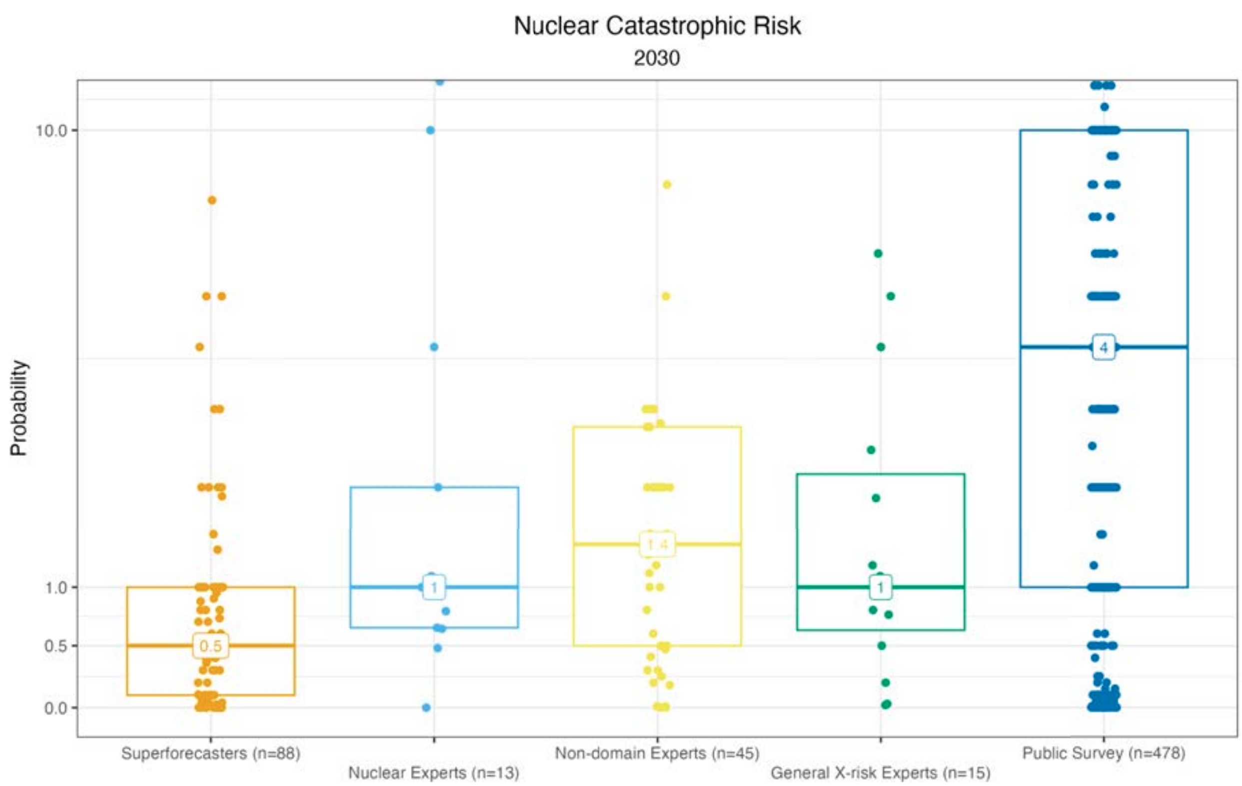
making process under the possible miscalculation that AGI would be less likely to mistakenly launch a nuclear attack when in fact the reverse may be true."
[318] 342, "A final type of risk is competing AGI systems fighting over control of the future and wiping out humans as a byproduct."
[319] Numbers of forecasters are given as of Stage 4 of the XPT.

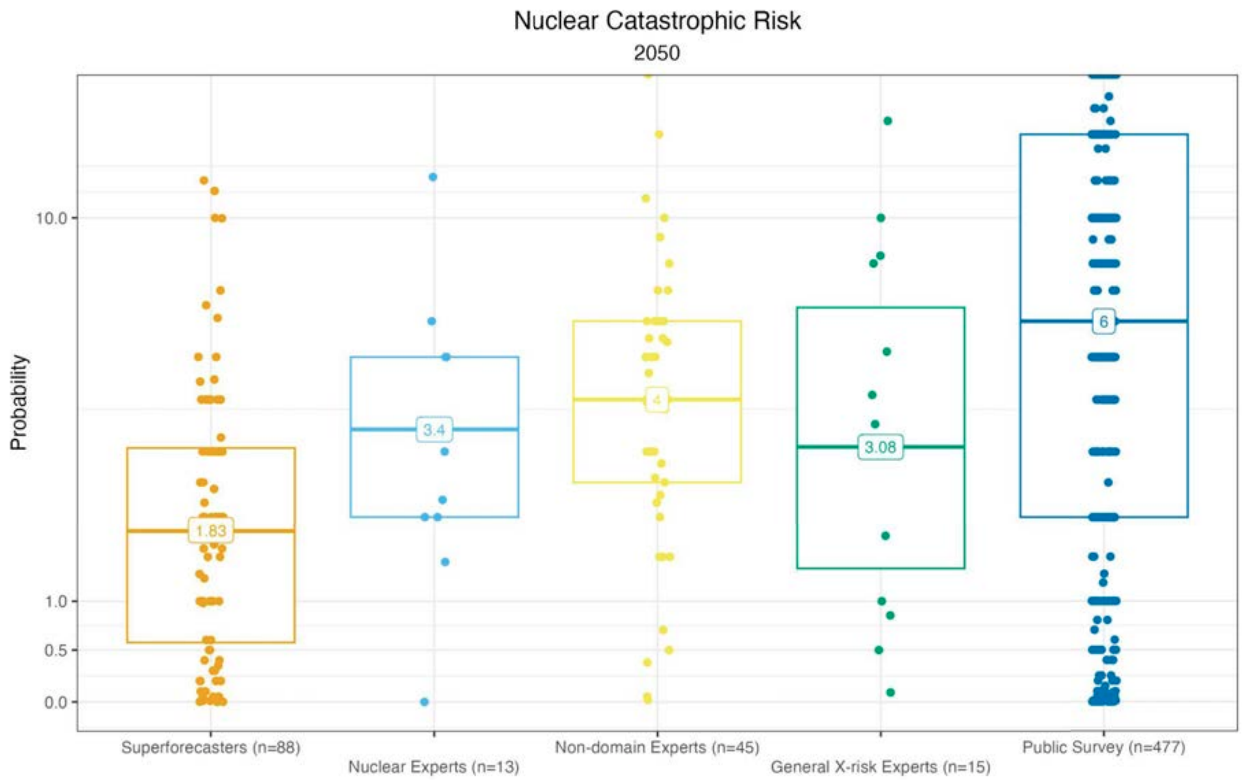
	2100	4%	4%	15.39	-74.22%
Domain Experts (N = 13)	2030	1.1%	1%	4.3	-9.7%
	2050	3.8%	3.4%	7.26	-9.05%
	2100	7.23%	8%	11.39	-36.62%
General X-Risk Experts (N = 15)	2030	1%	1%	2.28	-18.79%
	2050	3.08%	3.08%	6.04	-27.86%
	2100	6.74%	7.24%	15.11	-20.74%
Non-Domain Experts (N = 45)	2030	1.5%	1.4%	3.57	-11.24%
	2050	5%	4%	5.79	-19.6%
	2100	9%	8%	10.03	-23.81%
Public Survey (N = 478)	2030	4%		1079.26	-
	2050	6%		2284.25	-
	2100	10%		22844.76	-

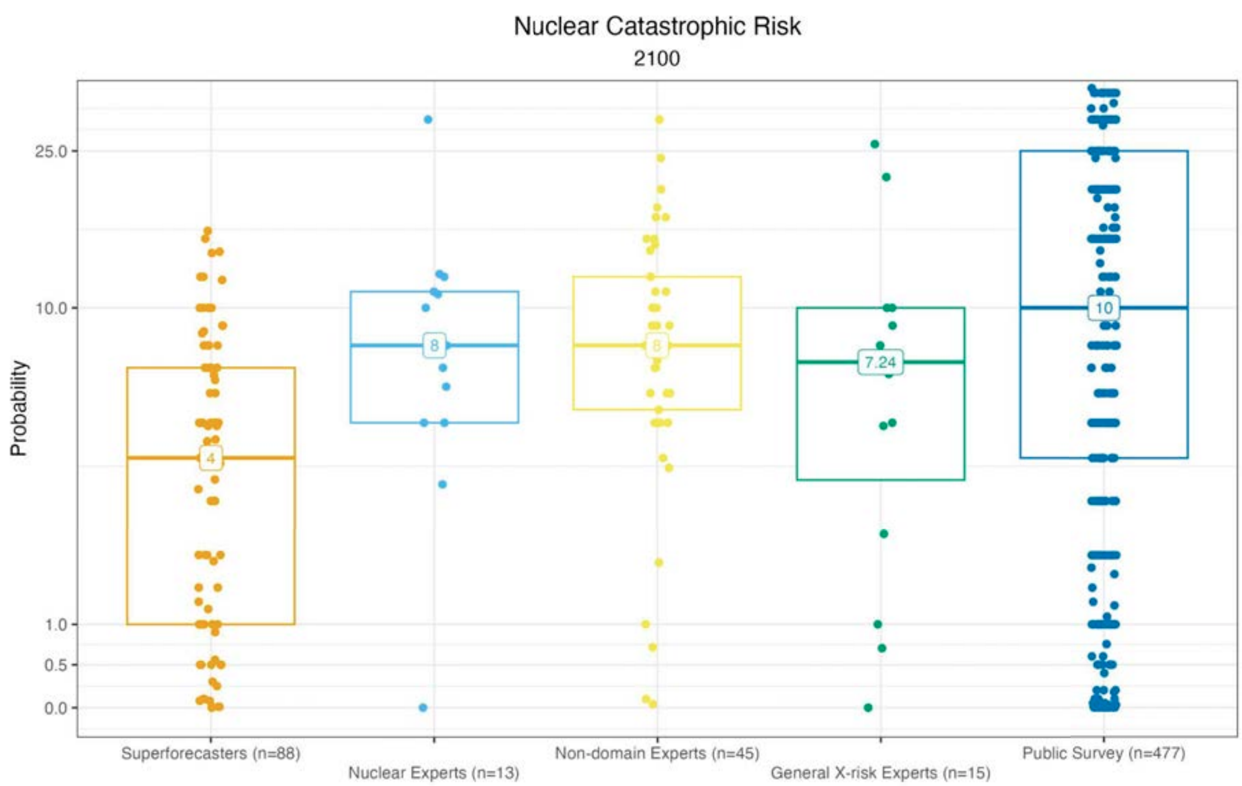

Nuclear Catastrophic Risk by 2030
1.50
1.25
1.00
0.75
0.50
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 88)
Domain Experts (n = 13)
Non-domain Experts (n = 45)
General X-risk Experts (n = 15)


Nuclear Catastrophic Risk by 2050
5
4
3
2
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 88)
Domain Experts (n = 13)
Non-domain Experts (n = 45)
General X-risk Experts (n = 15)


Nuclear Catastrophic Risk by 2100
Median
10
8
6
4
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 88)
Domain Experts (n = 13)
Non-domain Experts (n = 45)
General X-risk Experts (n = 15)


Nuclear Catastrophic Risk
2030
Probability
10.0
1.0
0.5
0.0
0.5
1
1.4
1
4
Superforecasters (n=88)
Nuclear Experts (n=13)
Non-domain Experts (n=45)
General X-risk Experts (n=15)
Public Survey (n=478)


Nuclear Catastrophic Risk
2050
Probability
10.0
1.0
0.5
0.0
1.83
3.4
4
3.08
6
Superforecasters (n=88)
Nuclear Experts (n=13)
Non-domain Experts (n=45)
General X-risk Experts (n=15)
Public Survey (n=477)


Nuclear Catastrophic Risk
2100
Probability
25.0
10.0
1.0
0.5
0.0
4
8
8
7.24
10
Superforecasters (n=88)
Nuclear Experts (n=13)
Non-domain Experts (n=45)
General X-risk Experts (n=15)
Public Survey (n=477)

Sources of agreement, disagreement and uncertainty

- **Uncertainty over base rate of nuclear war**
    - Forecasters were uncertain about how to calculate the base rate of this event.[320]
        - "Since we have not had a nuclear war of any significant magnitude yet, we must hypothesize base rates informed by expert opinion…Using different base rates and base cases results in probability estimates that differ by an order of magnitude, or even more." (T340)
    - Teams used different statistical methods to estimate the risk of an event that has never occurred.
        - The base rate could be viewed as zero either because nuclear weapons have never killed >10% of humans[321] or because nuclear weapons have not been used since 1945[322] and the Hiroshima and Nagasaki bombs did not lead to catastrophic outcomes as defined in the question.[323]
        - The rule of three suggests a 4% risk by 2100.[324]
        - Another statistical method suggests a 0.51% risk by 2100.[325]
        - "Absence of nuclear conflicts since 1945 suggests that the risk of nuclear war is under 1/77/2 [=] ~0.6% per year." (T343)
    - Near misses and close calls should lead us to adjust the base rate upwards.[326] However, teams did not quantify what kind of base rate increase was appropriate.

---

[320] "[A] major source of uncertainty is the lack of comparable events in human history, making establishing good base rates difficult." (T337)
"When forecasting a question in which the base rate is zero but there are a lot of 'near-misses' then making adjustments is going to end up being very subjective." (T338)
[321] "The base rate for nuclear weapons having been the cause of death for more than 10% of humans alive at that time is zero because this event has never happened. " (T345)
[322] "Nuclear weapons have only been used twice, no nuclear catastrophes on the order asked about have occurred despite the weapons existing for 77 years." (T336)
"[T]he base rate is zero but there are a lot of 'near-misses'…" (T338)
"**Base rate 1**: Near **zero**, given nuclear weapons have not been used since 1945." (T341)
[323] "The base rate for nuclear catastrophe is obviously zero, as the two hostile/non-test uses in history (Hiroshima and Nagasaki) were obviously not intended nor capable of leading to an extinction event." (T342)
[324] "**Base rate 2…**Rule-of-three cautious estimate for events that have not happened = 3/n (where n = number of years), assigns the probability 0.95 to measure the likelihood of realizing more events than those observed. Therefore, 3/78 = **4%** by 2100." (T341)
[325] "**Base rate 3**: Statistical method to estimate probability of events that have not yet occurred based on sample size: (2/5n, less cautious estimate for non-observed events) 2/(5*78) = **0.51%** by 2100." (T341)
[326] "A handful of near-launches have been documented and there may have been other cases that are outside the public record." (T336)
"Higher forecasts also tended to assign more weight to the near misses of the cold war, and the behavioral failings of leaders leading up to such close calls." (T337)
"[T]here have been situations historically, at which point nuclear conflict between major powers…could have happened – even if only created by accident." (T339)

        - “Our guess is that the actual risk of war in these incidents was not all *that* high, because everyone wants to avoid nuclear war if they can, and false alarms will tend to have implausible features by their nature…[however] I don’t think we can be confident that no one will press the button in a false alarm situation.” (T338)
- **Regional conflict would not directly kill 10% of humans**
    - Forecasters agreed that a regional, small-scale nuclear exchange (such as one between India and Pakistan or North Korea and South Korea) would not directly kill enough people to result in a catastrophic outcome.[327] Supporting this conclusion, teams cited the limited number of nuclear weapons involved in a regional exchange as well as the limited target set.[328]
        - “For nuclear weapons to cause the death of 10% of humans alive, it would require[ ] either (1) many weapons used on large and dense population centers or (2) nuclear winter.” (T338)
        - Forecasters disagreed as to whether a small-scale nuclear exchange could lead to catastrophic death indirectly through nuclear winter and resultant famine.[329]
    - No team believed that terrorists could acquire enough nuclear weapons to produce a catastrophic outcome.[330]
        - “Nuclear terrorism is unlikely to be able to kill 10% of the earth's population unless there are some big unexpected changes in terrorism or in nuclear capabilities.” (T344)
- **Likelihood of nuclear winter**

“Perhaps the lack of use of nuclear weapons since their invention is just luck, as there have been many close calls.” (T344)

[327] “[C]ertain forecasters considered that in the event of a nuclear accident or regional conflict it is unlikely the relevant numbers would be reached, and that a world war might be needed to reach the necessary numbers.’ (T339)
“The conflict would need to be big enough to see the exchange of many hundreds or thousands of nuclear weapons, such as a general nuclear conflict between the US (or NATO) and Russia or China, that would kill hundreds of millions outright and cause a nuclear winter.” (T341)
“The circumstances that would lead to this likely involve some combination of India and Pakistan, or the US and Russia, or China and anyone else firing a relatively large nuclear exchange.” (T342)

[328] “Looking at the distribution of nuclear weapons, it seems clear the distribution is across regional competitors. In all cases: Israel, Pakistan/India, N. Korea, Russia, the weapons seem defensive in nature to ensure national survival. Therefore the risk may be limited as the targets of specific nations are limited…” (T338)

[329] “The most notable source of disagreement comes from the analysis of how would a small-scale nuclear exchange, for example between India and Pakistan, impact the death toll. Some forecasters believe that even a small-scale nuclear exchange would lead to a massive human die-off…Of course, other forecasters are challenging such argumentation.” (T344)

[330] “The requirement for nuclear war to cause the deaths of more than 10% of humans alive means that we quickly focused on those scenarios that could see a billion or so people directly or indirectly killed as a result of nuclear war. That ruled out a lot of lesser scenarios, such as nuclear terrorism…” (T341)
“The general consensus is that nuclear accidents and terrorist incidents are not likely to kill up to 10% of humans alive at the time and that a nuclear war would be the most likely scenario in which this could occur.“ (T345)

- Since most teams believed that a nuclear conflict was unlikely to *directly* lead to the death of >10% of humans, the perceived likelihood of nuclear winter (i.e., nuclear weapon-induced climate change leading to famine) strongly impacted forecasts.
- Forecasters disagreed about the accuracy of nuclear winter models and about the threshold at which atmospheric soot injection would trigger nuclear winter.[331]
    - "Given that a significant percentage of nuclear weapons deaths are likely projected as a downstream consequence of nuclear winter, we expect a significant difference between forecasts that accept catastrophic nuclear winter projections, and those that are skeptical of them." (T343)
    - Some forecasters believed nuclear winter forecasts may be overstated "given they remain theoretical in large part or derived from very different events such as volcanic eruptions." (T341)

- **Role of AI**
    - Uncertainty about the potential role of transformative AI affected forecasts on nuclear catastrophic risk. AI could either function as "a possible force multiplier in wars" (T337); use nuclear weapons against humanity;[332] create a stable arms control regime;[333] or lead to proliferation of nuclear secrets.[334]
- Teams disagreed about whether intentional use of nuclear weapons would be more likely to lead to a catastrophic outcome compared to a miscalculation or accidental launch.[335]

---

[331] "Recent modeling of the effects of soot injection indicate that there could be 2 billion deaths from a smaller war…Models indicated widespread global food insecurity and famine…[but] some of the models of nuclear winter are based on a scenario where both the US and Russia use their entire stockpile. This would be extraordinarily unlikely even in a nuclear war." T336)
"Although a number of people have cited the Rodriguez report, there are several other published analyses suggesting a lower threshold for nuclear winter than what it describes…" (T338)
"Many also revised upwards their forecasts on the basis of a recent research article which suggested that the likelihood of serious consequences for global food supply from even a modest nuclear exchange would result in mass starvation and death in countries that are already "food insecure". The number of people who are already food insecure in the world is 1.9 billion, so even small changes to their food supply from a nuclear winter could cause mass starvation." (T341)
"The evidence does not support either large temperature changes or long-term temperature changes due to nuclear winter." (T343)
"The overall impact of second level effects was a major cause of uncertainty due to differing ideas about the likelihood of a nuclear winter and how severe its impact would be on humans outside the direct conflict area if a nuclear winter did occur." (T345)
[332] Nuclear weapons could be used by "defective narrow intelligence AI" or "misaligned general intelligence AI." (T344)
[333] "AGI and other radically transformative technologies could either cause human extinction or, if managed well, cause some sort of stable arms control regime to exist...[or] one could imagine misaligned AI systems using a nuclear attack to disempower humanity." (T338)
[334] "Lack of care regarding safeguarding of nuclear secrets. This could be compounded by development of AI." (T336)
[335] "Miscalculations were felt to represent a greater threat in the context of this question than intentional actions." (T339)
"The possibility of an accidental nuclear launch by one of these powers is also considered, though intentional launching is probably more likely." (T342)

Arguments given for low-end forecasts

- **Killing >10% of humans is hard**
    - Many teams noted that the question set a high bar by requiring nuclear weapons to be the cause of death for more than 10% of humans alive (estimated as 800 to 900 million people).[336] This number is more than the current U.S. population plus the current Russian population, and more than the population of the top ten biggest world cities.[337] The detonation of a large number of nuclear weapons would be required to cause this number of immediate deaths.[338]
    - A catastrophic outcome is more likely if nuclear weapons target civilian centers (i.e., large cities) rather than military targets. Forecasters were uncertain as to whether strategic considerations would incentivize targeting civilians.[339]
- **Deterrence and the nuclear taboo**
    - The existence of nuclear weapons for 77 years without being used suggests that nations are reluctant to use them. This could be explained by the deterrent effect of mutually assured destruction[340] or by the existence of a nuclear taboo,[341]

---

[336] "To reach 10% would require the total elimination of the entire population of Europe, or more than all of the US and Russia, or all of Pakistan and half a billion people in India." (T336)
"For nuclear weapons to kill that many people from a direct hit, weapons would have to hit around 30 of the largest cities in the world or, for example, half of the population of both India and Pakistan." (T338)
"Nuclear weapons are renowned for their damage potential. However, for nuclear weapons to be the proximate cause of death for more than 10% of all humans is still a very high bar." (T339)
"[W]e can have a very devastating global nuclear war, which will destroy world's most important players [and] still could not reach 800 millions deaths." (T340)
Those with lower-end forecasts "mention how difficult it would be to actually kill 10% of the population: even if someone attempted 'battlefield nukes', a response to bring the death toll that high would have to be beyond proportional." (T342)
"In the case of an actual nuclear war, the direct casualties are likely to be below 10% of humanity." (T343)
"10% of the population is a lot of people (the number changes with the predicted population of the earth in different years)." (T344)

[337] "As one forecaster noted, 10% of the population would, at present, amount to roughly 800 million people and that, "the entire population of the US and Russia [the two countries which possess the vast majority of nuclear weapons] combined clocks in at under 500 million." (T337)
"The top ten biggest cities in the world wouldn't do it." (T344)

[338] "The arguments noted that to kill 10% of the human population, a large scale nuclear war would be needed involving potentially many countries and a large number of nuclear warheads launched at major cities." (T337)

[339] "There was some disagreement within our group about whether there might be circumstances that incentivized targeting mass civilian casualties in a way that accepted reduced deterrence and reduced strategic/military efficacy." (T343)

[340] "Some forecasters also noted the deterrent implicit in mutually assured destruction as contributing to their relatively low forecasts." (T337)
"Also a focus was…the historic effectiveness of nuclear deterrence." (T339)
"Factors influencing those advocating for forecasts on the lower end of the plausible range of forecasts for this question were that…nuclear deterrence would prevent such a large-scale conflict from arising." (T345)

[341] "Our lower ended forecasts have the number remaining below 1% by 2100. Forecasters leaning in this direction are more skeptical of any major launches happening due to likelihood that a nation attempting such a massive attack would become a global pariah." (T342)

although some with higher-end forecasts were concerned about a potential "disruption of the 'nuclear taboo' over time." (T341)
    - "It seems throughout the past 80 years of nuclear weapons that countries have rightly been hesitant to use nuclear weapons." (T344)
    - "Many considered that mutually assured destruction (MAD) still held amongst major nuclear powers, is restraining their use of nuclear weapons against each other." (T341)
- **Humans will learn over time**
  - Some teams mentioned the possibility that humans might become less violent or more able to settle disputes over time.[342]
    - "[A]s a species we will make more rational decisions in the long term if we can get to the long term." (T344)

Arguments given for higher-end forecasts

- **Nuclear proliferation**
  - Some teams expected the number of nuclear weapons and nuclear states to increase over time.[343] The New START treaty expires in 2026 and does not cover new nuclear weapons.[344] A new nuclear treaty between the U.S. and Iran "seems unlikely." (T338)
  - "[A] number of commentators have suggested that Saudi Arabia may develop nuclear capabilities to counter Iran…Japan or South Korea could take steps to develop their own nuclear weapons program if they feel that the US is not committed to protecting them via its nuclear umbrella." (T338)
  - One team mentioned an "argument from history," asking "[w]hen have humans invented a weapon of war and it was not used repeatedly?" (T336)
- **Current geopolitical tensions**
  - Potential conflict between great powers such as the U.S. and Russia or U.S. and China increases nuclear catastrophic risk.[345]

---

[342] "Some arguments focused on humanity becoming less violent as time progresses, though others did not." (T337)

[343] "Likely increases in nuclear warheads in the coming decades likely, in particular by 'new' nuclear powers were they to emerge, increase the chance of escalations and accidents occurring." (T339)
"Those who were forecasting significantly higher than the median saw nuclear war of a catastrophic nature as more likely because of the continuing proliferation of nuclear weapons to more countries." (T341)

[344] "New Start does not cover new weapons (nuclear torpedoes)." (T338)

[345] "Current trends should not be over-emphasized, but the din of nuclear weapons is heard within the cacophony of super-power saber-rattling." (T336)
"[T]here are a number of nuclear dyads and triads between which there are elevated tensions in the near-term, e.g., US-Russia (with the war in Ukraine) and US-China (with growing tensions over Taiwan)." (T338)
"Forecasters who considered this event to have a relatively high probability of occurring (>5% by 2100) considered the current conflict landscape, the Ukraine war and potential for hot conflicts in the 21st century involving USA, China, and Russia." (T339)

        - Forecasters disagreed about the likelihood of "hot" great-power conflicts as opposed to limited "cold" conflict.[346] Uncertainty over geopolitical developments over the question's timespan, and in particular over the chance that nations with autocratic governments might become globally dominant, contributed to this disagreement.[347]
        - The war between Russia and Ukraine could lead to use of nuclear weapons by Russia or the U.S.[348]
        - As China positions itself as a military partner of Pakistan, China could be drawn into an India-Pakistan conflict.[349]
- **Leaders can make bad decisions**
    - Several teams pointed to the possibility that world leaders could make bad decisions because of their cognitive or behavioral failings.[350] On the other hand, one team pointed out that the decision to launch nuclear missiles would almost always require more than one decision-maker.[351]
        - "We shouldn't forget the fact that the most powerful people on Earth are also capable of making the worst possible decisions." (T340)
- Climate change could lead to geopolitical conflict over resources or increase the risk of an accidental launch due to lack of funding for inspection and testing of weapons.[352]

---

[346] "Sources of disagreement included: the probability of a 'hot' global conflict involving nuclear exchange vs more limited great power conflict…" (T341)
"We largely agree on the dangers if it happens, but when some think the chance of any launch happening is sub-1%, and some believe it is 20%-100%, there will remain an irreconcilable belief of how likely catastrophic consequences will be." (T342)
[347] One team noted the "uncertainty associated with evolving technologies and geopolitics over the long time span associated with this question." (T337)
"Past absence of nuclear conflicts may be the result of military and economic superiority of western democracies. In the future, this dominant position may be taken by autocratic governments which would be less averse to nuclear war risks." (T343)
[348] Increases in nuclear tensions include…increased saber rattling for Vladimir Putin during the Russia/Ukraine conflict." (T336)
"Some were worried about a potential peak in nuclear risk in the next few years, especially as countries like Putin's Russia act aggressively and are nuclear armed." (T341)
"War in Ukraine with at least one nuclear power involved (Russia) is alarming." (T344)
[349] "China is positioning itself as a close military partner of Pakistan, India's main geopolitical rival…If this trend continues into the future, as seems likely given China's probable desire to counterbalance India, this could substantially worsen the regional geopolitical situation, potentially leading to a major nuclear conflict in the region." (T338)
[350] "Higher forecasts also tended to assign more weight to the near misses of the cold war, and the behavioral failings of leaders leading up to such close calls." (T337)
"A factor influencing forecasts on the higher end of the plausible range of forecasts for this question is the fear that an unstable individual world leader or ideology could start a nuclear conflict." (T345)
[351] "Almost always, authority to launch distills down to judgment from more than a single-person / entity. Killing 800 million humans is possible, but requires launching enough nuclear missiles that will all but guarantee widespread decimation of one's homeland…as namby-pamby as it sounds, significant social status is needed to place into the chain-of-judgment for nuclear launch; launching does little to further that status." (T336)
[352] The chance of accidental nuclear exchange might increase because of "deteriorated inspection and testing due to funding shortages caused by climate change expenses…Fear and anger will rise as resources (water, energy, arable land, etc.) become rarer." (T336)

Cross-references with other questions

Q3: AI Catastrophic Risk
Q6: Nuclear Existential Risk
Q7: Non-Anthropogenic Catastrophic Risk
Q9: Total Catastrophic Risk
Q11: Year of Existential Catastrophe

Question 6: Nuclear Extinction Risk

What is the probability that one or more incidents involving nuclear weapons will cause human extinction or reduce the global population below 5,000...
...by the end of 2030?
...by the end of 2050?
...by the end of 2100?

Question and resolution details, prior forecasts, and other relevant sources

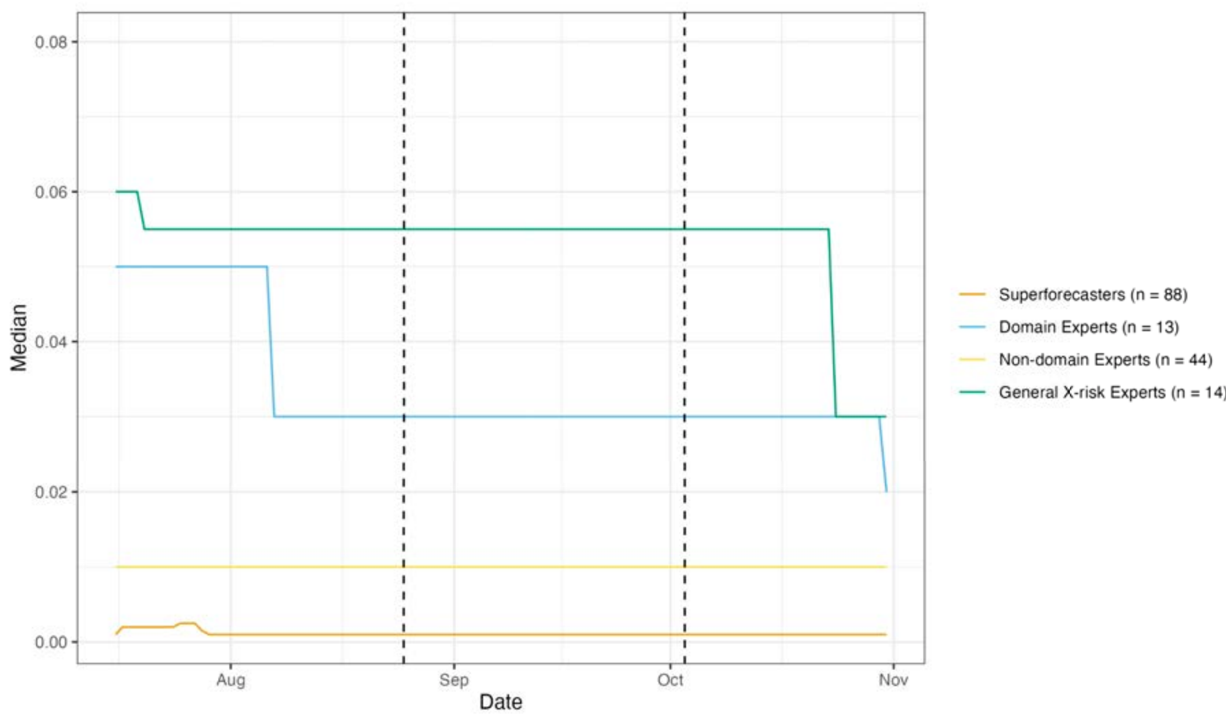
Results[353]

Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 88)	2030	0.004%	0.001%	0.57	-73.12%
	2050	0.03%	0.01%	1.26	-72.57%
	2100	0.1%	0.074%	1.49	-61.1%
Domain Experts (N = 13)	2030	0.05%	0.02%	5.51	-93.34%
	2050	0.18%	0.12%	8.19	-90.19%

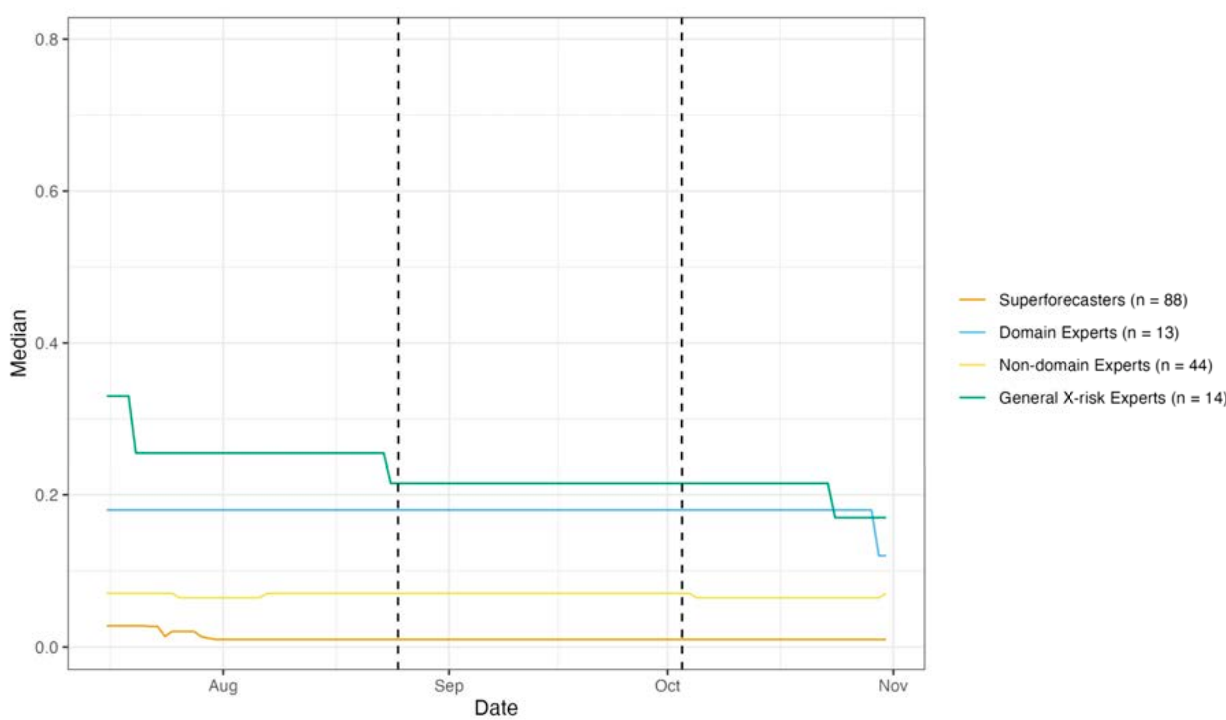
353 Numbers of forecasters are given as of Stage 4 of the XPT.

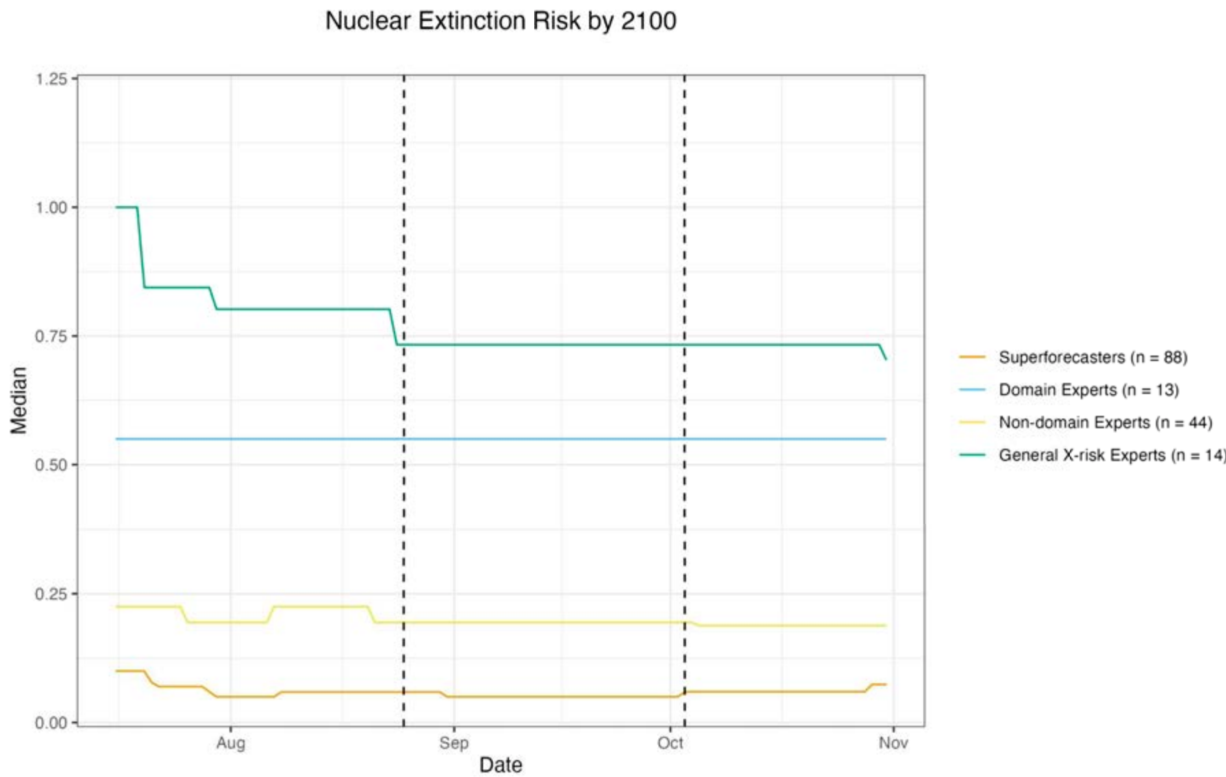
	2100	0.55%	0.55%	12.27	-90.1%
General X-Risk Experts (N = 14)	2030	0.055%	0.03%	1.05	+0.34%
	2050	0.26%	0.17%	1.69	+0.69%
	2100	0.84%	0.7%	1.81	-1.13%
Non-Domain Experts (N = 58)	2030	0.01%	0.01%	0.89	-25.45%
	2050	0.07%	0.07%	2.02	-14.06%
	2100	0.23%	0.19%	3.03	-18.52%
Public Survey (N = 478)	2030	1%		29.1	-
	2050	1%		315.21	-
	2100	2%		264.84	-

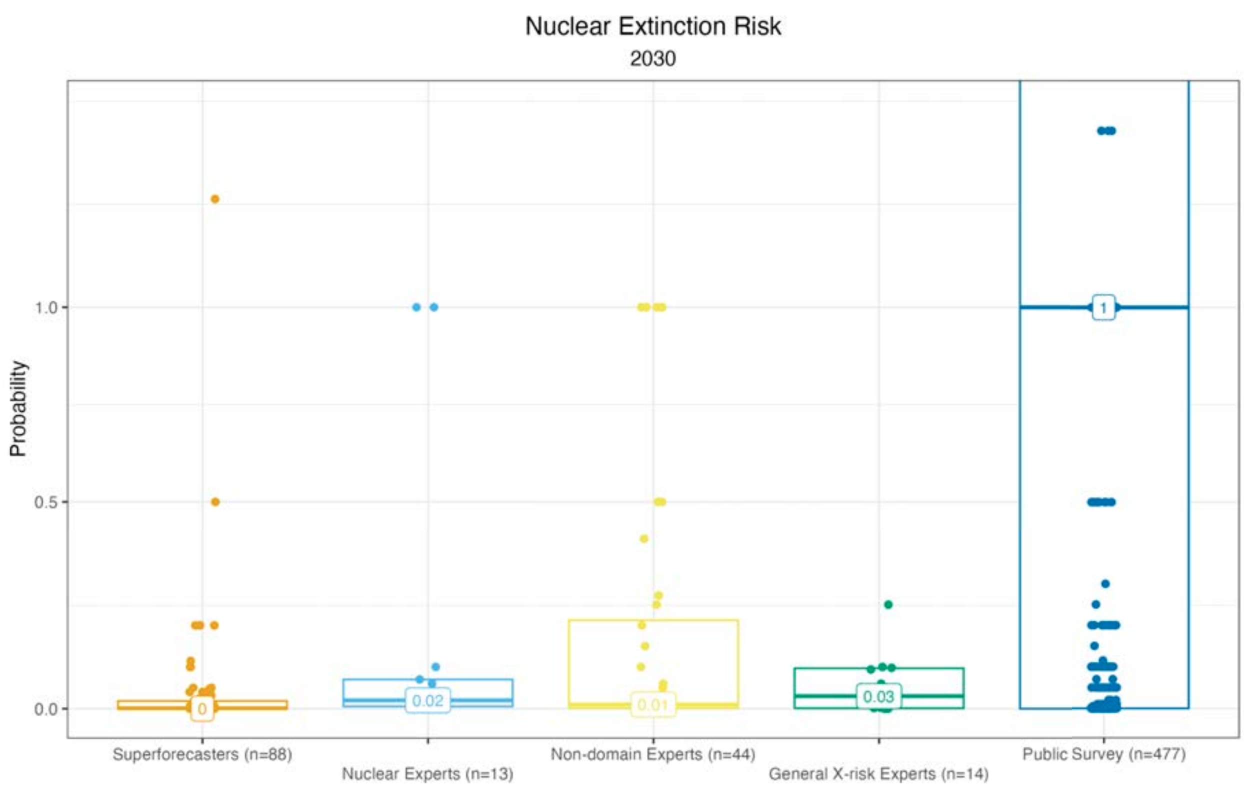
Nuclear Extinction Risk by 2030

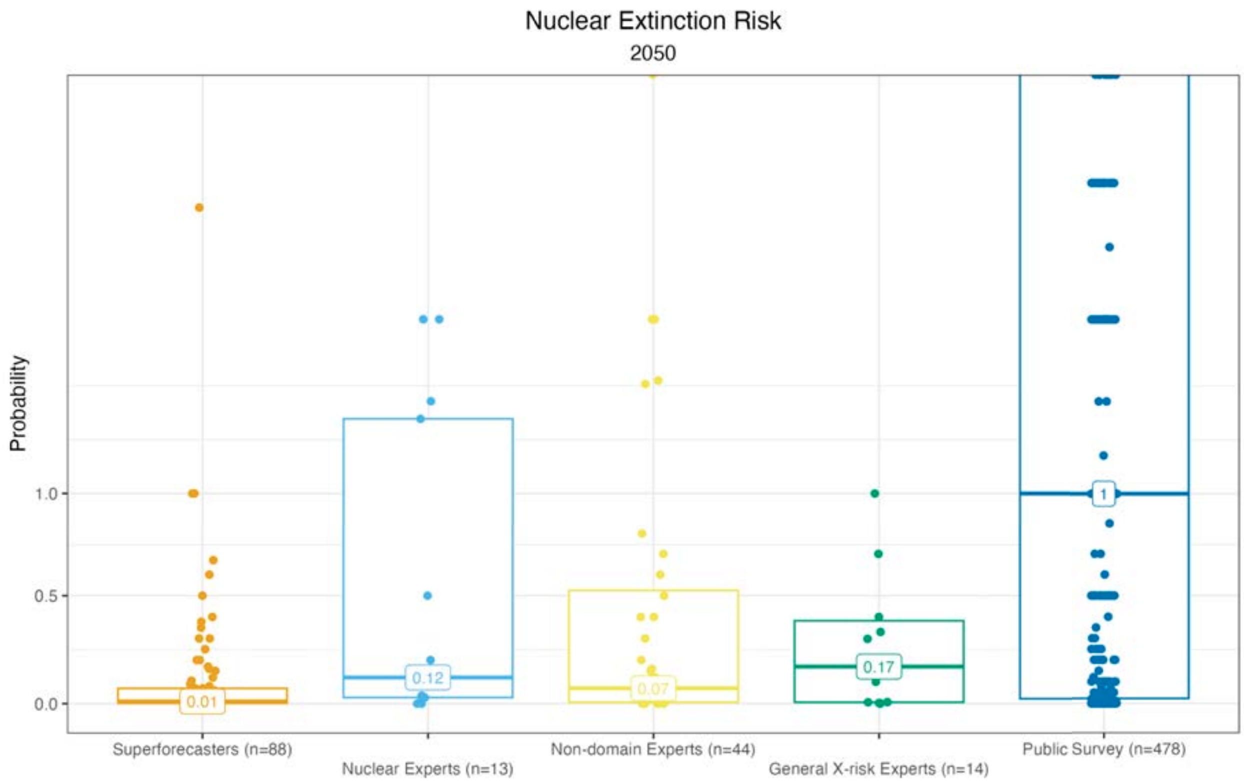


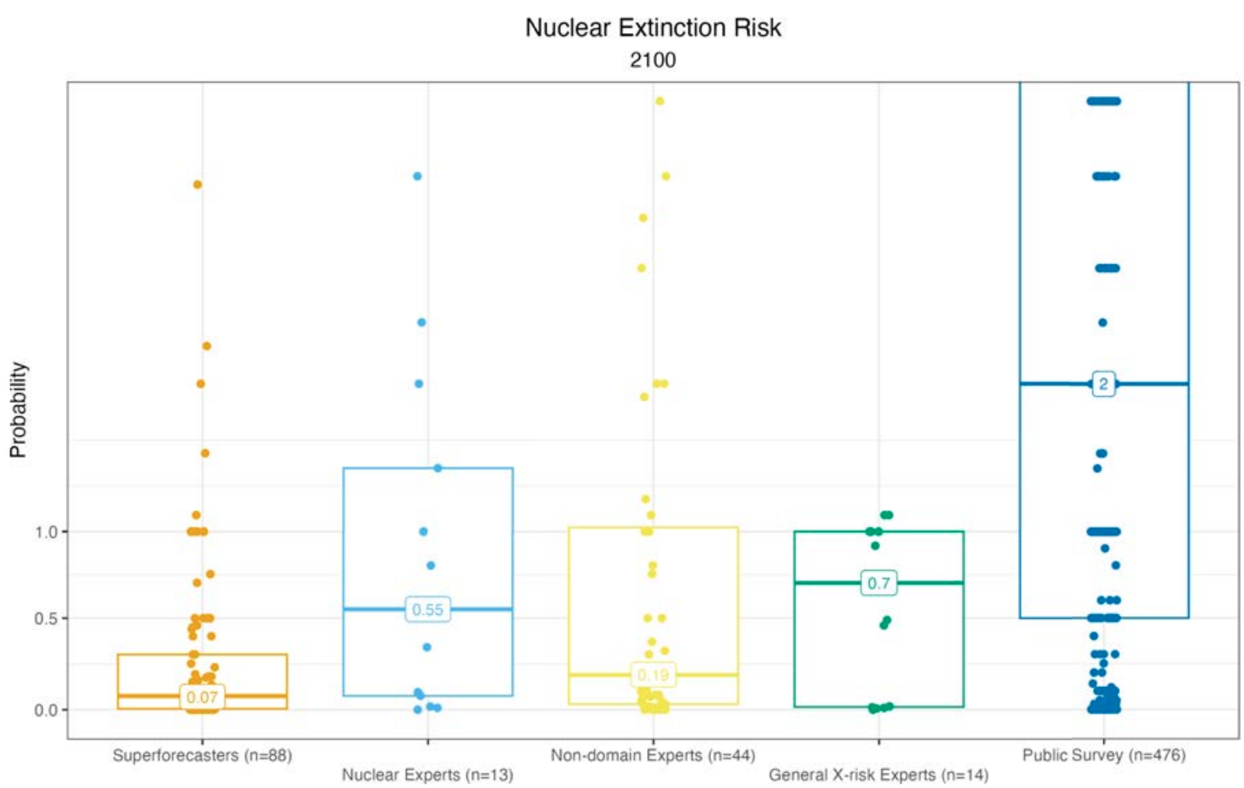
Nuclear Extinction Risk by 2050




Nuclear Extinction Risk by 2100
1.25
1.00
0.75
0.50
0.25
0.00
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 88)
Domain Experts (n = 13)
Non-domain Experts (n = 44)
General X-risk Experts (n = 14)


Nuclear Extinction Risk
2030
1.0
0.5
0.0
Probability
0
0.02
0.01
0.03
1
Superforecasters (n=88)
Nuclear Experts (n=13)
Non-domain Experts (n=44)
General X-risk Experts (n=14)
Public Survey (n=477)


Nuclear Extinction Risk
2050
Probability
1.0
0.5
0.0
0.01
0.12
0.07
0.17
1
Superforecasters (n=88)
Nuclear Experts (n=13)
Non-domain Experts (n=44)
General X-risk Experts (n=14)
Public Survey (n=478)


Nuclear Extinction Risk
2100
Probability
1.0
0.5
0.0
0.07
0.55
0.19
0.7
2
Superforecasters (n=88)
Nuclear Experts (n=13)
Non-domain Experts (n=44)
General X-risk Experts (n=14)
Public Survey (n=476)

Sources of agreement, disagreement and uncertainty

- **Nuclear war will not immediately cause human extinction**
  - Teams agreed that use of nuclear weapons would not immediately cause human extinction,[354] citing the limited number of nuclear weapons available.[355] In addition, targets of nuclear weapons would be chosen based on strategic military considerations rather than distributed over all global population centers.[356]
    - "Even if the US and Russia got into a full scale nuclear war, they are likely to limit the targets to primarily the Northern Hemisphere and particularly cities, military and industrial targets of each other and possibly each other's allies…The Southern Hemisphere is likely to avoid the worst effects of such a war." (T336)
- **Uncertainty over nuclear winter**
  - If nuclear war would not immediately cause human extinction, the question would only resolve positively if nuclear winter (i.e., climate change induced by nuclear weapons use) led to human extinction.[357]
    - "Human populations are distributed across the entirety of the globe and, in the event of a full blown nuclear conflict involving all nuclear powers, it would be very likely that large swathes of the globe would not be impacted directly. The only plausible mechanism would be a subsequent nuclear winter where ash from the global firestorms reaches the upper atmosphere leading to global cooling and crop failures over a number of years." (T339)
  - Forecasters were highly uncertain about the likelihood and severity of nuclear winter.[358]

---

[354] "The immediate destruction from nukes are extraordinarily unlikely to be an existential threat but there was disagreement on whether enough dust could be blasted into the atmosphere to create a species ending nuclear winter." (T336)
"Most forecasters agreed that the direct results of any nuclear exchange would not be sufficient to pose a nuclear extinction risk." (T343)
"It would be very difficult to wipe out humanity with nukes alone." (T344)
"[O]ver all, the consensus is that the nuclear blasts themselves would not result in reaching the **VERY HIGH BAR** of killing all but 5,000 people." (T345)

[355] "[S]ome team members belie[ve] that the quantity of nuclear weapons currently available for use would not actually be enough to cause an extinction event, and thus that there would actually have to be nuclear rearmament, a politically very difficult pill to swallow." (T342)
"Not all weapons in a nuclear stockpile are deployed, and would likely not be deployable within the time required for them to participate in a nuclear exchange." (T343)

[356] "Without hitting South America and Australia, or missing any number of Pacific Islands, would mean that well over 5,000 would survive." (T342)
"There are many remote places that are unlikely to be directly targeted by nuclear weapons, and that would only have to deal with the consequences of nuclear winter." (T344)

[357] "The most significant consideration in whether nuclear weapons use could cause human extinction, therefore, is the risk of nuclear winter." (T343)

[358] "A second source of major uncertainty is whether second and third order effects of a major nuclear exchange would be sufficient to cause extinction." (T337)
"Most research on nuclear winter does indeed seem to be by people who have campaigned for arms control, which does make critiques that they have selected the most pessimistic plausible assumption at

        - "One major disagreement may be the climate effects of a large number of nuclear bombs. How bad a nuclear [winter] would be is both very unclear and very important to answering the question at hand." (T336)
        - "According to some estimates, even 100 warheads is enough to start a nuclear winter…even with a very optimistic hypothesis we'll have enough [warheads] for a long while." (T339)
    - While there was agreement that nuclear winter is a theoretical possibility, nuclear winter model results depend on assumptions made regarding, for example, the extent to which cities would be targeted by nuclear weapons, how flammable those targets would be, and how much soot would reach the upper atmosphere.[359]
        - "[T]he models vary greatly, both in how large the effect would be and how wide-spread the impact would be around the world, so there is significant uncertainty." (T341)
        - "As far as I can tell, though I'm not certain about this, most of the models make two big pessimistic assumptions: 1 that cities will be extensively targeted…and that the war will occur in the spring or summer (apparently climate effects are much smaller if it occurs in winter.)" (T338)
        - "Both the predicted absolute temperature decreases and the predicted duration of temperature decreases from nuclear winter models are likely to be overestimated sufficiently that their impact on human survival can be dismissed." (T343)
    - **How well could humans adapt to nuclear winter?**
        - Doubts over humans' ability to adapt to a changed post-nuclear-war world motivated some higher forecasts.[360] Forecasters pointed to the difficulty

---

each point in modelling plausible, but the critics are probably biased also, and it's hard to tell which direction the causation between scary nuclear winter models and campaigning for arms control goes." (T338)
"Even with a lot of modeling and researcher interviews, there are no researchers on record asserting that a nuclear winter would lead to humanity's extinction." (T340)
"The team's median forecast reflects the uncertainty regarding nuclear winter models… Nuclear winter may make it impossible in most parts of the world to grow food. Isotopes from a nuclear event will increase radiation levels in the oceans." (T341)
"Nuclear winter might not be as impactful as previously thought." (T344)

[359] "The basic claim that 'if enough soot gets into the atmosphere, global temperatures will suffer a temporary dramatic drop' seems to be solidly established science…[b]ut dramatic effects [are] not uniform, less in coastal areas, less in the tropics, where in any case, even a big drop in temperature is unlikely to make all agriculture impossible." (T338)
"None of the predictions made by the [pessimistic] global consensus surrounding nuclear winter hypothesis have been supported by subsequent evidence." (T343)

[360] "Forecasters that deviated somewhat significantly from the median forecast focused on the collapse of civilization and the difficulty in reverting to agrarian or hunter gatherer ways of subsistence, arguing essentially that such forms of subsistence may be more difficult than one might assume…While it is true that humans are uniquely successful in adapting to their environment, will they have sufficient time and resources to do so under the conditions of extreme nuclear war?" (T337)

of losing existing supply chains and agricultural technology and pivoting to new food sources.[361]

- **Uncertainty over base rate of nuclear war**
  - While teams largely aligned on a base rate of zero for this event, there was disagreement about the value of that base rate.[362]
    - "A major source of uncertainty is how much weight to give the zero base rate (and lack of any comparable events in human history) versus the short time during which humanity has had the technological ability to severely cripple itself (a reality that suggests the base rate might not be particularly relevant)." (T337)
- **Nuclear extinction is low probability**
  - Teams agreed that this event is generally very low probability.[363]

---

[361] "The strongest argument on the higher end focuses on the uncertainties of living in a world post-nuclear holocaust. There is no evidence whatsoever that humans are capable of living in such a world. Packaged food will all be consumed in a few weeks. Humans are likely to starve or suffocate in bunkers, or at least exhibit bizarre behavior over time. See Biosphere 2. Growing grains at scale would be very difficult, if not impossible in a much colder world with little sunlight. A nuclear winter of sufficient gravity to cause a global famine may be unlikely to result in total near-term human extinction but it certainly puts such an outcome in play." (T340)
"One forecaster argues that "at the very least, fishermen with canoes and long lines and big nets could feed their families", while another says that they would think that current farming/fishing methods will quickly become unavailable due to lack of fuel, fertilizer, seeds and spare parts." (T341)
[362] "Base rate: 0…[t]he nuclear bombs dropped on Hiroshima and Nagasaki in 1945 are the only times a nuclear weapon was used in war. If this counts as a single event, the frequency could be viewed as once every 77 years." (T336)
"The base rate for the event in question is (to put it succinctly) zero." (T337)
"[T]he base rate on this question isn't much use - it only needs to happen once…[but another forecaster argued that] [k]nowing the base rate of nuclear weapon use has been zero since the end of WWII leads to a more reasonable and accurate forecast." (T338)
"At this point in time it's zero. Although we have Hiroshima and Nagasaki to show us what happens when major population centers are victims of nuclear blasts, these are small demonstrations when compared to an all out nuclear war." (T339)
"Base rate 1: Near zero, given nuclear weapons have not been used in the ~79 [years] since 1945. This suggests that an extinction level use of them is highly unlikely in the following 78 years until 2100.
Base rate 2: Using rule-of-3 simple statistics. Rule-of-three cautious estimate for events that have not happened = 3/n (where n = number of years), assigns the probability 0.95 to measure the likelihood of realizing more events than those observed. Therefore, 3/78 = 4% by 2100
Base rate 3: Statistical method to estimate the probabilities of events that have not yet occurred based on sample size: (2/5n, less cautious estimate for non-observed events) 2/(5*78) = **0.51%** by 2100." (T341, emphasis removed)
"The base rate for nuclear extinction is obviously zero, as the two hostile/non-test uses in history (Hiroshima and Nagasaki) were obviously not intended nor capable of leading to an extinction event." (T342)
"The base rate for nuclear weapons causing an extinction-level event is zero because this event has never happened." (T345)
[363] "Most forecasters whose probabilities were near the median factored in a range of possible risks, including world wars, nuclear winters, and even artificial-intelligence-driven NERs [nuclear extinction risks], but concluded that even under worst case scenarios, the extinction of humanity (give or take 5000 people) would be near impossible...even if an NER [nuclear existential risk] had set humanity on a path that made eventual extinction a foregone conclusion, existing resources on earth would allow at least 5000 survivors to hang on for seventy-eight years." (T337)

        - "There aren't compelling arguments on the higher end for this question again due to the fact that this is a very high bar to achieve." (T340)

Arguments given for low-end forecasts

- **Small populations would survive nuclear war**
    - Even if nuclear blasts destroyed most major population centers, forecasters argued that 5000+ people could survive in bunkers[364] or in isolated areas far from predicted targets, such as the Southern Hemisphere and small islands.[365] Some populations have managed to maintain populations of over 5000 in harsh conditions.[366]
        - "[A]lmost certainly some people would survive on islands or in caves given even the worst of worst cases." (T342)
        - "Southern Hemisphere likely to be less impacted – New Zealand, Madagascar, Pacific Islands, Highlands of Papua New Guinea, unlikely to be targeted and include areas with little global and technology dependence…Just the population of Antarctica in its summer is ~5000

"For many, the thought of getting to less than 5000 humans alive was simply too far fetched an outcome and they couldn't be persuaded otherwise in what they saw as credible scenarios." (T338)
"[T]he set of circumstances required for this to happen are quite low, though obviously not impossible. These circumstances are that there will be a nuclear conflict between 2 nations both capable and willing to fire at everyone everywhere between the two of them: 'very bad case scenarios' where India and Pakistan, or the US and Russia, or China and anyone else, fired everything they had at just each other, or even at each other and each other's close allies, would likely not cause extinction…it requires some of the big nuclear powers to decide to try to take literally everyone down with them, and that they actually succeed." (T342)
"So we think that the probabilities in this question are dominated by scenarios of total nuclear war before 2050 which cause civilizational and climate collapse to the point where long-term survival becomes impossible to save for very well-prepared shelters. But even pessimistic scenarios seem unlikely to lead to a collapse that is fast enough to reduce the global population to below 5000 by 2100." (T344)
[364] "We did also discuss possibilities of people staying in bunkers for extended periods of time (~5 years) to survive the nuclear winter onslaught and current technology easily allows more than 5000 people to easily survive in bunkers for a considerably long time (duration of 5 years is not uncommon)." (T340)
"It is easy to "safeguard" a small human population." (T344)
"Most teammates agreed that there are probably plans in place to shield a large number of humans worldwide in the event of the catastrophic use of nuclear weapons and that sufficient shelters and supplies exist to prevent human extinction." (T345)
[365] "Some enclaves will survive. Australia is part of NATO, but New Zealand may not receive nuclear attacks and may be able to continue farming." (T341)
"Locations away from the nuclear powers and their targets, would be able to sustain a population of 5000... or maybe many more... New Zealand, South America, Sub-Saharan Africa... the extensive bunkers in Switzerland, and so on." (T344)
"The team predicts that there will be pockets of people who survive in various regions of the world. Their survival may be at Neolithic standards, but there will be tribes of people who band together and restart mankind. After all, many mammals survived the asteroid and ice age that killed the dinosaurs." (T345)
[366] " [A] certain number of team members feel that even if there was a full strategic exchange and usage of all of the world's nuclear arsenal still humanity would be able to keep its numbers over 5000. The argument for this is the number [a]nd population of uncontacted tribes, or isolated human populations like the Easter island population pre-contact, that have managed to hold numbers of over 5000 in extremely harsh conditions." (T344)

people. Even small islands surviving could easily mean more than 5k people." (T336)
        - "[There are s]everal regions in the world that would not be affected by nuclear conflict directly and have decent climatic conditions to support 100 of millions even in a NW [nuclear winter]." (T345)
- **Deterrence and the nuclear taboo**
    - "[E]ven with a major world war, history shows that major nuclear powers have been extremely hesitant to actually deploy nuclear weaponry, even in active engagements with other nuclear powers or in situations with supposedly "crazy" or brinksmanship rulers (e.g., N. Korea)." (T344)
- **Chance of non-retaliation**
    - "[There is a c]hance one nuclear power will not retaliate if attacked or that an all-out war will be avoided. This includes the possibility that even if the order is given by the leader, the military might not follow the orders." (T336)

Arguments given for higher-end forecasts

- **Nuclear proliferation and new weapons**
    - New technology could produce more destructive nuclear weapons and could allow faster spread of nuclear weapons to new states.[367]
        - "The more countries with nuclear weapons, the greater the chance of a mistake that leads to a nuclear escalation." (T336)
        - "[W]hile most agreed that there are not currently enough nuclear weapons or powerful enough nuclear weapons to wipe out humanity, this could change in the future with advances in technology." (T345)
- **Misaligned AI and combined risks**
    - This question will resolve positively if an extinction event has multiple causes, including nuclear weapons, as long as nuclear weapons are one of the "but-for" causes of the event. Several teams therefore adjusted their probabilities upwards to account for combinations of existential threats.[368]

---

[367] "New nuclear weapons technologies not currently conceived of" would support higher forecasts. (T336)
"[B]y 2050 and 2100, we may have many more nuclear states and nuclear weapons." (T341)
"There is also some disagreement, if yearly chance of nuclear extinction is going to remain more or less the same until 2100, or if it is going to increase in the second part of the 21st century because of the effect of new technologies allowing faster nuclear proliferation…" (T342)
"In the long term, the increased number and potency of world arsenals could increase the damage provoked directly and indirectly by nuclear detonations." (T344)

[368] "I think that it's eminently possible (though, one hopes, not likely) that major power conflict breaks out later this century, using or involving multiple WMDs/emerging tech (e.g., AI, biotech, nukes) and that this is ultimately what leads to human extinction. So my forecast of 2.25% by 2100 is not purely a reflection of whether nukes, in and of themselves, can lead to human extinction; it's more broadly a forecast of whether major power war *involving nukes* is likely to lead to human extinction." (T345)

- The most commonly cited example of such a combination is that of misaligned AI using nuclear weapons against humanity.[369]
    - “[C]onsiderations of humanity's situation in 2050 or beyond are heavily dominated by considerations about the emergence and control of AGI; if no full-scale nuclear war has happened by 2050, it is unlikely that it will occur afterwards…One of the main drivers for high probabilities is AI-extinction scenarios where the AI uses nuclear weapons as part of its strategy.” (T344)
- One team cited climate change[370] and engineered bioweapons[371] as potential contributors to a stressed and resource-constrained world which might be less able to prevent nuclear war.

Cross-references with other questions

Q4: AI Existential Risk
Q5: Nuclear Catastrophic Risk
Q8: Non-Anthropogenic Existential Risk
Q10: Total Existential Risk
Q11: Year of Existential Catastrophe

Question 7: Non-Anthropogenic Catastrophic Risk

What is the probability that non-anthropogenic causes (e.g., asteroid or comet impacts, solar flares, a supervolcanic eruption, or a stellar explosion) will be the cause of death, within a 5-year period, for more than 10% of humans alive at the beginning of that period...
...by the end of 2030?
...by the end of 2050?
...by the end of 2100?

Question and resolution details, prior forecasts, and other relevant sources

---

[369] “There is also fear that advanced neural networks or some form of an AGI will increase the likelihood of a total nuclear war.” (T336)
Given that the question allows nuclear weapons as one of multiple causes of extinction, “it’s hard to see how the probability on this question could be less than say 1/10 of the probability of AI extinction.” (T338)
“A misaligned AI cascades into a launch, provoking other AIs to launch in response.” (T339)

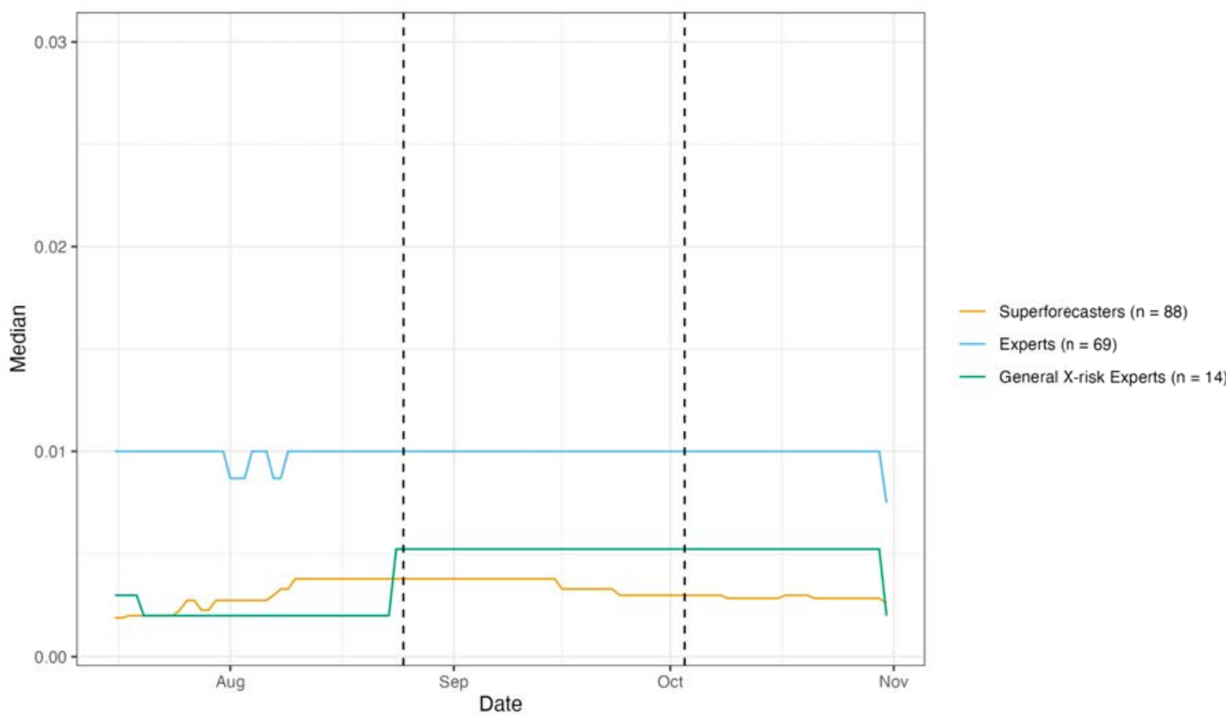
[370] “It might be that the world overall is in much more duress due to eg. climate change, leading to higher tensions and more potential of nuclear exchange.” (T336)
[371] “Nuclear war [could be] one of several events over time that weaken humanity to the point of extinction (AGI, climate change, engineered pathogens...).” (T336)

Results[372]

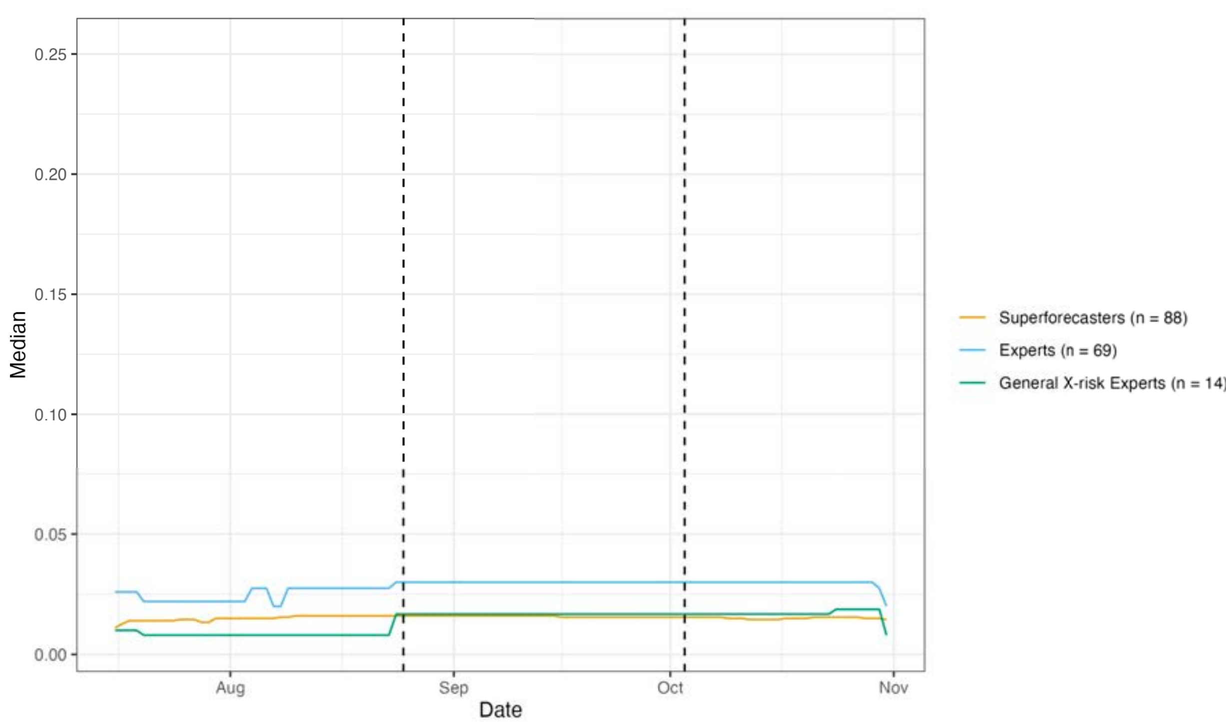
Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 88)	2030	0.0019%	0.0026%	0.64	-58.94%
	2050	0.01%	0.015%	1.38	-36.05%
	2100	0.037%	0.05%	3.14	-41.85%
Experts (N = 69)	2030	0.01%	0.0075%	1.19	-27.02%
	2050	0.02%	0.02%	1.15	-26.14%
	2100	0.089%	0.09%	1.63	-11.12%
General X-Risk Experts (N = 14)	2030	0.002%	0.002%	0.32	-2.20%
	2050	0.008%	0.008%	1.12	-1.04%
	2100	0.3%	0.045%	2.89	-0.3%
Public Survey (N = 478)	2030	1%		35.91	-
	2050	1%		241.66	-
	2100	2%		2284.78	-

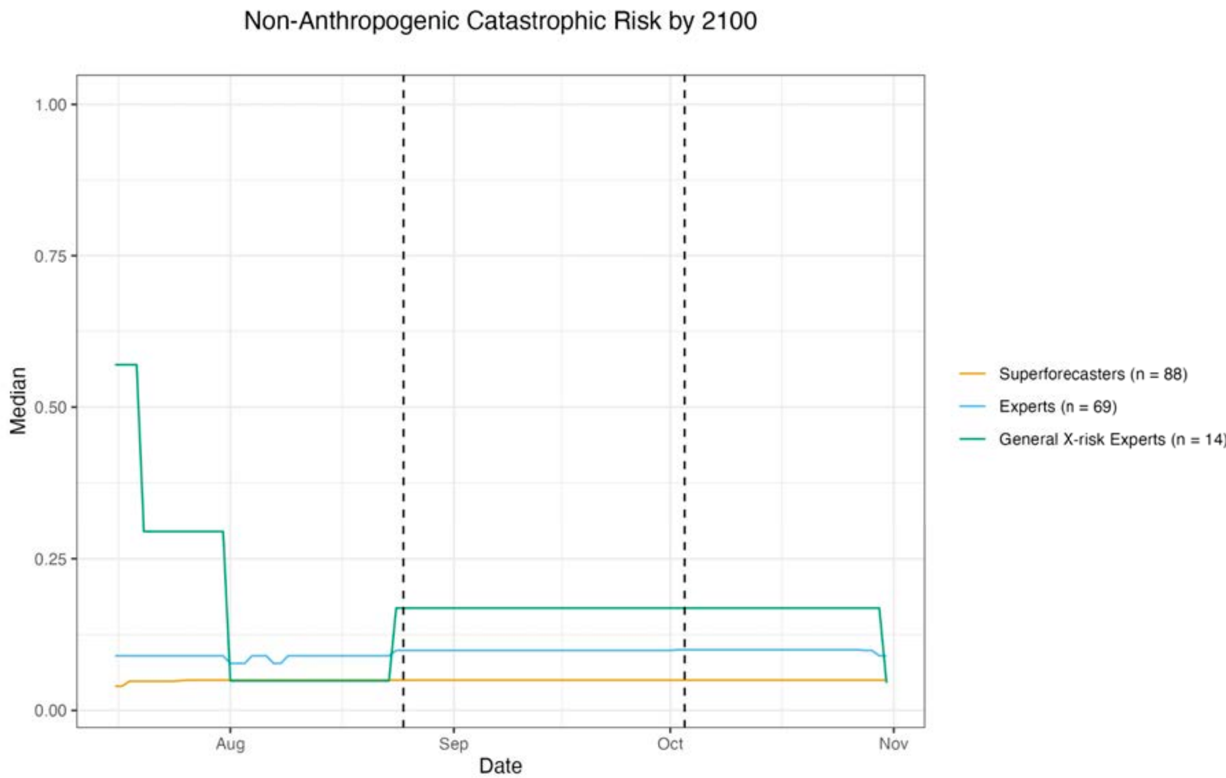
[372] Numbers of forecasters are given as of Stage 4 of the XPT. Public survey respondents were surveyed outside of the tournament context.

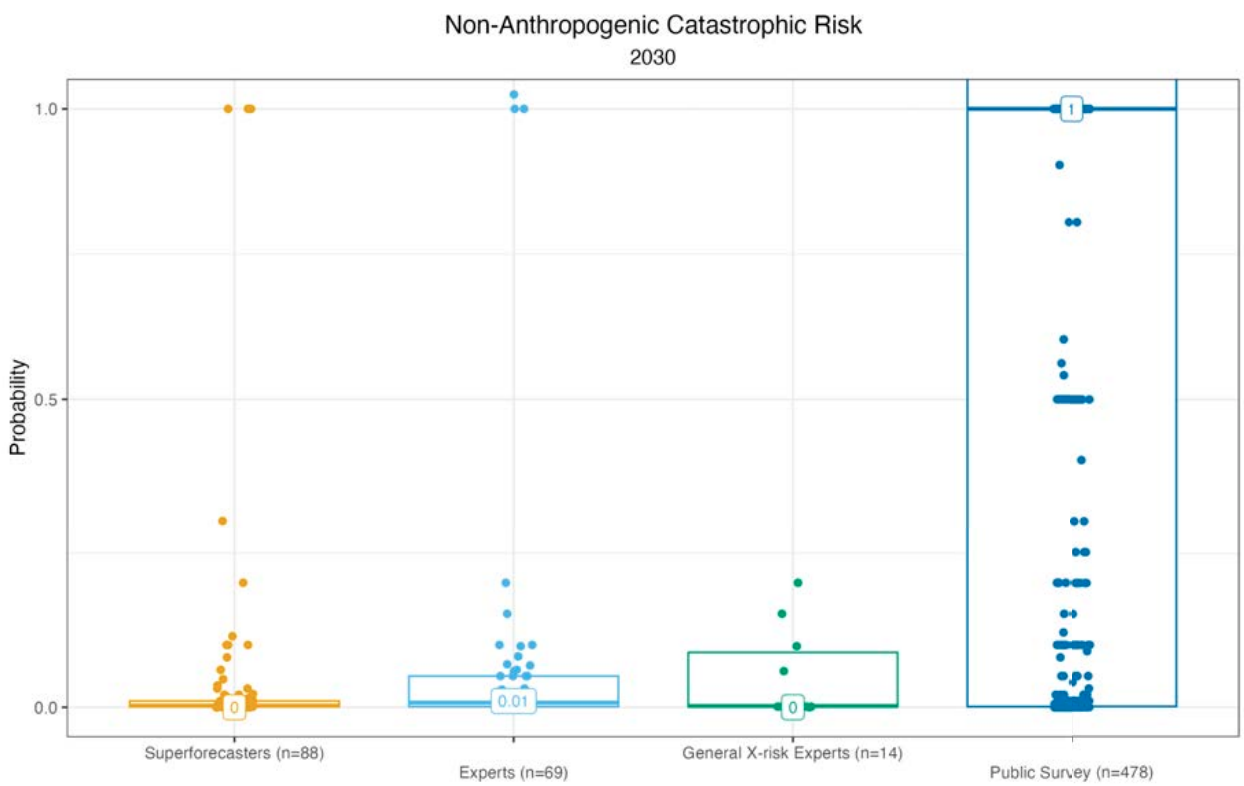
Non-Anthropogenic Catastrophic Risk by 2030

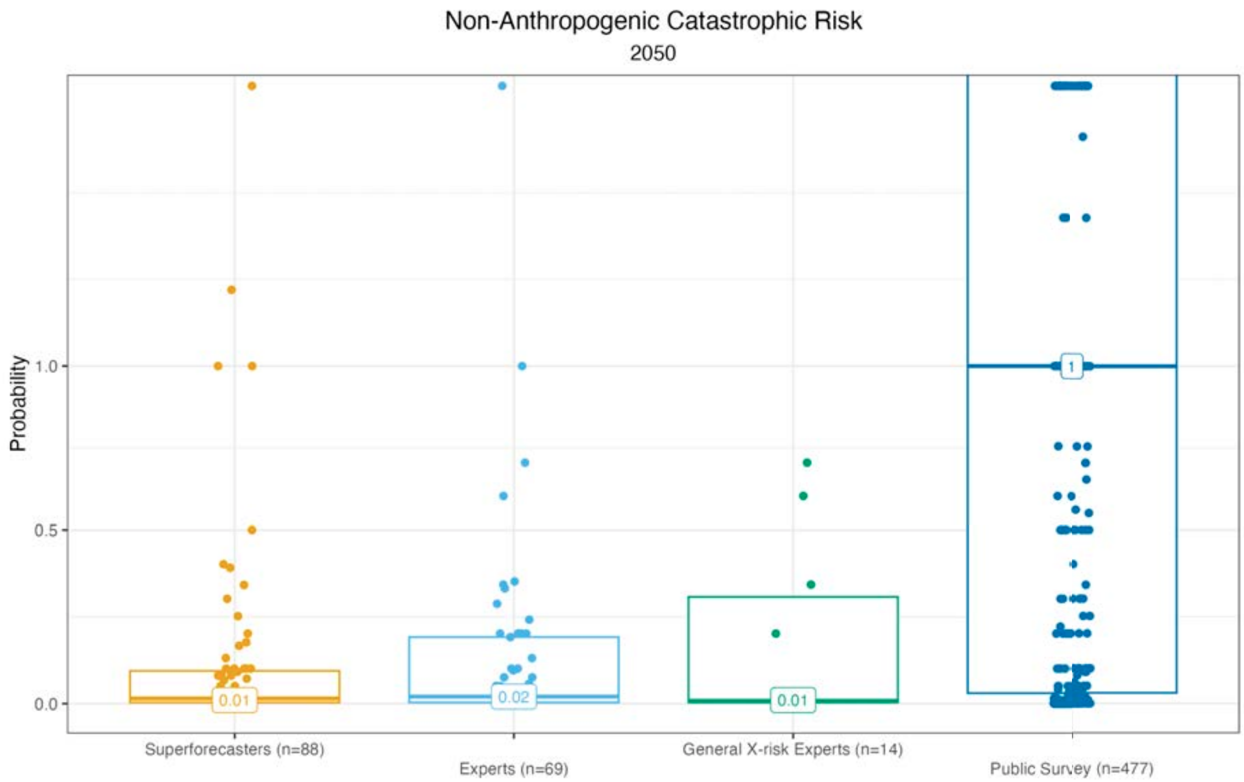


Non-Anthropogenic Catastrophic Risk by 2050

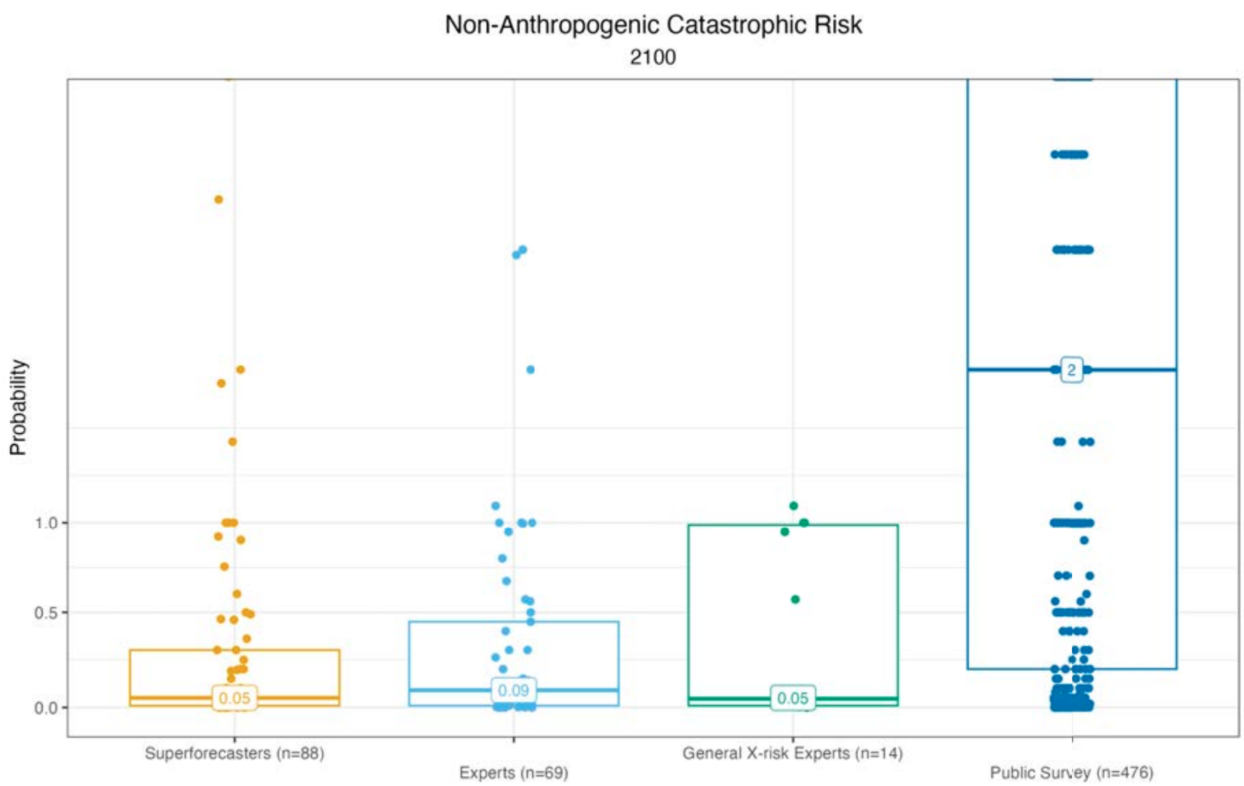



Non-Anthropogenic Catastrophic Risk by 2100
Median
1.00
0.75
0.50
0.25
0.00
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 88)
Experts (n = 69)
General X-risk Experts (n = 14)


Non-Anthropogenic Catastrophic Risk
2030
Probability
1.0
0.5
0.0
0
0.01
0
1
Superforecasters (n=88)
Experts (n=69)
General X-risk Experts (n=14)
Public Survey (n=478)



Non-Anthropogenic Catastrophic Risk
2050
Probability
1.0
0.5
0.0
0.01
0.02
0.01
1
Superforecasters (n=88)
Experts (n=69)
General X-risk Experts (n=14)
Public Survey (n=477)



Non-Anthropogenic Catastrophic Risk
2100
Probability
1.0
0.5
0.0
0.05
0.09
0.05
2
Superforecasters (n=88)
Experts (n=69)
General X-risk Experts (n=14)
Public Survey (n=476)

Sources of agreement, disagreement and uncertainty

The main sources of disagreement were:

- The method for producing the forecast
    - Some forecasters derived an estimate from summing the probabilities of different events that could pose this level of catastrophe, whereas others adapted published models of the probability of extinction risk to the level of catastrophe asked about in this question.[373] The team rationales didn't provide arguments for whether one model should be considered better than the other.
- The types of non-anthropogenic events that could pose a catastrophic risk
    - Most teams and individual forecasters considered several types of events that could pose a catastrophic risk, including supervolcanoes, asteroids and comets, solar weather events, supernovae and gamma ray bursts. Several teams noted that the risk was dominated by supervolcanoes due to the higher probability of these events occurring.[374]
    - Some forecasters included a wider variety of risks, including extra-terrestrial contact,[375] a geomagnetic event,[376] and natural climate change.[377] However these were presented as less important than the more commonly included risks.
    - There was disagreement about whether solar flares could lead to >10% population loss, with the debate on whether they could cause sufficient disruption to technology.[378]

---

[373] 338, "Forecast approaches can be bucketed into three almost equal-sized groups:
Bottom-up: base rates for individual components, forecasters usually selected only the most severe risks, followed by summation or aggregation.
Published forecast extrapolation: from published work on catastrophic and existential risks, adjust those forecasts to fit our tournament forecast dates and criteria.
Intuitive: either a "feels-like" number or not enough explanation to otherwise categorize". The following teams mentioned both a bottom-up approach and a published forecast extrapolation approach in their rationales: 337, 338, 341, 342, 344. The following only mention a bottom-up approach: 336, 339, 340, 343.

[374] 339, "The estimated risks of extinction from the Non-Anthropogenic Extinction Risk forecast are so low as to be rounding errors based on super volcano risk.", 338, "Like Non-Anthropogenic Extinction risk, volcanic risk dominates most forecasts.", 341 "Super volcanoes are probably the largest risks"

[375] 344 "A few of us mentioned the possibility of contact with extraterrestrial life, but no one dared to provide a probability estimate." 337, "The second way is to list all the possible sources for a non-anthropogenic catastrophe (impact events, alien invasion/attack …". 336, "Catastrophic encounter with extraterrestrials "

[376] 337, "Geomagnetic event: mostly the change of polarity of earth's magnetic field. It has been deemed too gradual to lead to a 10% of humanity dying in 5 years.", 336, "magnetic pole reversal (very unlikey)"

[377] 337 "Natural climate change: most forecasters have decided that this risk gives a negligible contribution owing to its gradual nature (in contrast with the 5 year time frame in the question's decription).", 336 "natural climate change (and that might range from people starving, because it´s too warm (something similar to the ozone hole, but caused by sun activity or a volcanoe emitting huge amounts of CO2/methane; or the same happening because of a "small ice age" caused by a volcano eruption (Yellowstone?!), a meteortite or something else causing the atmosphere not to let sunlight through)."

[378] 341, "A very large CME could destroy most of the power grid and communications, triggering starvation, war and unrest leading to 10% deaths.", 339, "There was much discussion surrounding Solar Flares. Our ability to predict them is low. But those observed thus far have been mild, and certainly not

- The probability of types of events that might result in catastrophe occurring
  - There was disagreement on the best base rate to use for each type of event. Base rates for supervolcanic eruptions cited in rationales ranged from 1 in 17,000 years[379] to 1 in 250,000 years.[380] The cited base rates for asteroids ranged from 1 in 120,000 years[381] to 1 in 20 million years.[382]
- The probability of a catastrophe causing 10% population loss
  - One of the key drivers of differences in forecasts seemed to be how resilient society would be to the catastrophe.[383]
- How to adapt models of extinction risk to 10% population loss
  - Several forecasters answered this question by adapting published estimates on the probability of extinction due to non-anthropogenic causes. On the presumption that catastrophic events are more likely than extinction events, the forecasters applied a multiplier to the estimates of extinction risk.
  - The choice of model and the choice of multiplier varied. Models used included the model developed by Toby Ord in [The Precipice](),[384] a model published by [Snyder-Beattie, Ord, and Bonsall in a 2019 paper](),[385] and a model using the extinction of the dinosaurs as a base rate for an event that could lead to human extinction.[386] The multipliers used ranged from 3 to 1000.[387]

---

of the caliber necessary to take out 10% of mankind. The questions surround the resilience of North American, and Northern European electrical systems, and their ability to withstand a major solar flare. Would the destruction of infrastructure be sufficient to start a domino effect that could kill that many people?"

[379] 339 " Some research by Danisik et al has suggested that a super volcano explodes 1/17,000 years."

[380] 343, "Supervolcanoes. Probability of an event: 1 in 250k years"

[381] 338, "Toby Ord, in The Precipice, cites a 1 in 120,000 chance of a 1-10km asteroid hitting the Earth this century…" and 344 "1/120 000 chance of > 1km during next century (Ord, see this post for the table of probability estimates by Ord)"

[382] 343, "Asteroids/comets. Probability of an event: 1 in 20 million years"

[383] 340, "The greatest source of disagreement is not differing opinions on the likelihood of a potentially catastrophic event occurring but on what impact those events would have on humans.", 337, "Additionally, the resilience side of the equation, and growth in human resilience over time adds to the uncertainty. Human technological civilization, despite being fragile (electrical grid, supply chains, etc.), is also quite robust in a number of ways (communication & transportation in general), which helps to mitigate the effects of some kinds of disasters by allowing the importation of food, coordinated disaster responses, and so on. And yet, some disasters could neutralize these advantages (e.g. geomagnetic storm or an eruption of a supervolcano).", 344, "One sub-question where people had very different probability estimates was: Would a volcanic winter caused by a supervolcanic eruption kill 10% of the human population? Some estimated yes with 5% probability, some with 75% probability."

[384] 338, "Start with Ord existential forecast, adjust for question 10% of population threshold.", 337 "Also cited was an estimation by Toby Ord"

[385] 338, "Method: Use Snyder-Beattie, Ord and Bonsall (2019) and adjust for threshold by multiplying by 4.", 337 "For the probability of extinction the most cited source has been Snyder-Beattie et al.[1]"

[386] 338, "Method: Use end Cretaceous extinction event likelihood and adjust for 10% population threshold by multiplying by 1000 (also see [a forecaster's] extinction forecast)."

[387] 338, "Conversion to 10% threshold: multipliers quoted range from 3 to 1000"

Arguments given for forecasts below the median (0.0041% by the end of 2030, 0.0146% (2050), 0.049% (2100))

- The probability of most non-anthropogenic risks - asteroids,[388] solar flares,[389] supernovae[390] - that could kill >10% of the population is so low (with such events occurring once in many millions of years) that it is hardly worth counting them.
- Mitigation measures, such as asteroid monitoring and deflection, reduce risks even further.[391]
- Supervolcanoes are the most likely non-anthropogenic cause of this level of catastrophic risk, but it is unclear whether they would cause 10% population loss.[392]
- There is no known event in history that would have caused this level of catastrophe.[393]

Arguments given for forecasts above the median (0.0041% by the end of 2030, 0.0146% (2050), 0.049% (2100))

- Many of the events considered, even if not catastrophic on their own, could result in societal disruption that could result in 10% population loss. Supervolcanoes can lead to

[388]339, "The risk from asteroids in the next 78 years also appears to be vanishingly low. There are no asteroids big enough to cause catastrophic impact within the next 80 years. 1950 DA is considered one of the most likely asteroids to impact Earth with a catastrophic impact, and while that potential date was at one point around 2880. It has been estimated that an asteroid must be greater than 10km wide to significantly impact humanity." 343, "Asteroids/comets - Probability of an event: 1 in 20 million years. Probability an event causes the 10% of human of humans alive at that time: 50%, depending on advance detection and countermeasures available at that time " 341, " "According to NASA, the probability of an asteroid capable of destroying a city striking Earth is 0.1% every year. If one of these does hit Earth, there is a 70% chance it will land in the ocean, and a 25% chance it will land over a relatively unpopulated area. This is what happened with the Tunguska impact in Russia just over a hundred years ago." Probability destroying a city: 0.1% X 5%= 0.005%, 1 in 100,000. Destroying 10% of earth's population, maybe 1% of that = 1 in 10 million."
[389] 343, "Extreme solar activity (flares, storms) - Probability of an event: 1 in 1 million years. Probability an event causes the 10% of human of humans alive at that time: unclear whether this could cause the death of 10% of humans"
[390] 339, "The risk of catastrophes related to solar flares or stellar explosions also appear to be vanishingly low in comparison to the estimated risk for super volcanoes."
[391] 339, "NASA tracks most asteroids greater than 0.5km and is working on a defensive rocket to deflect an incoming asteroid. "NASA scientists say it would take an asteroid 60 miles (96 kilometer) wide to totally wipe out life on Earth." 343, "Some potential risks will likely be mitigated over time to one degree or another. For example, new surveillance systems for asteroid/comet detection and deflection. The earlier an asteroid or comet is detected, the smaller the angle of deflection required to avoid collision with Earth." 336, "Decreased asteroid risk due to efforts to catalogue and eventually deflect threatening objects."
[392] 339, "The most recent eruption was the Orauni eruption in New Zealand 26,000 years ago, and we cannot find any evidence that it significantly lowered global temperatures enough to cause a global famine."
[393] 344, "A disaster of this magnitude has never occurred in our collective history. It's unclear whether the events that took out the dinosaurs would be as catastrophic for us now, given our technologically advanced civilization."

crop failure, and solar flares could cause electrical systems to fail.[394] Many events could cause civil unrest or war.[395]

- There are many uncertainties in estimating the probabilities of these events, especially events like extra-terrestrial contact.[396] There may also be new types of risks that we have not yet discovered. Accounting for these uncertainties raises the risk.[397]
- Some research suggests that comets pose a bigger risk to Earth than asteroids, as they are harder to detect and have a higher velocity.[398]

Other arguments given

Arguments in favor of lower forecasts:

- If risks are found to be higher in the future, we should expect that effort would be put into developing mitigations.[399]

Arguments in favor of higher forecasts:

---

[394] 336, "Forecasters focus on solar flares, food/agriculture, poor self-sufficiency, dependency on electrical power: (food related issues (as stated elsewhere, we´ve got about 10% of the world´s population on the edge there, and that´s in a world without either man-made or non-anthropgenic catastrophes)
solar flares causing all technical devices to malfunction for a long time (preventing harvests or transporting food e.g. to Africa. Food insecurity could become dramatically worse due to a plant pathogen or climate change". 338, "The main scenario I am considering is that a natural event causes some sort of collapse resulting in famine and disorder. High population, specialization, and reliance on technology may make the system more susceptible to breakdown." "I don't know for sure whether a magnitude 8-9 eruption or a Toba-sized eruption would kill more than 10% of humanity, but it seems like a plausible outcome given that it could lead to crop failure and mass starvation in the same way as nuclear winter." 341, "The idea is that a Carrington event or whatever would disrupt our modern infrastructure that relies on electronics, causing supply chain problems etc and eventually 10% population death."

[395] 341, "Recent work in big history has underlined how important and substantial non anthropogenic events in human history are (e.g. Kyle Harper on climate change and the Roman Empire; more speculatively, Philip Blom on the intellectual ramifications of the little ice age, etc.). Second-order effects may be the cause of most deaths. A volcano, earthquake or solar flare could result in civil unrest, war, crop failure, or disease."

[396] 336, "Catastrophic encounter with extraterrestrials". 341, "Rare events can happen at any time and don't follow a normal distribution or pace."

[397] 339, "Human Extinction from Natural Hazard Risks lays out a case for a higher forecast based upon the unpredictability of these events. "Many extinction hazards are at present impossible to prevent or even predict…"

[398] 341, "Eugene Shoemaker has estimated that the flux of Earth-crossing comets is four times that of Earth-crossing asteroids. As they often are not detectable until months of nearing (or hitting!) Earth, and as their velocity with respect to Earth is far higher (more kinetic energy per mass, requiring more kinetic energy to deflect) they are more dangerous than asteroids.
As explained by Nuth, J. A., Barbee, B., & Leung, R. (2018). Defending the earth from long-period comets and sneaky asteroids. Journal of Space Safety Engineering, 5(3-4), Most comets have very little warning. Some have several years of warning; others can literally blindside us."

[399] 343, "The more susceptible humanity is to extreme solar events, the more likely measures will be taken to harden against solar events."

- One team noted that the concept of anthropic bias led several forecasters to revise their forecasts upwards.[400]

Cross-references with other questions

Q8: Non-Anthropogenic Existential Risk
Q9: Total Catastrophic Risk

Question 8: Non-Anthropogenic Extinction Risk

What is the probability that non-anthropogenic causes (e.g., asteroid or comet impacts, solar flares, a supervolcanic eruption, or a stellar explosion) will cause human extinction or reduce the global population below 5,000...
...by the end of 2030?
...by the end of 2050?
...by the end of 2100?

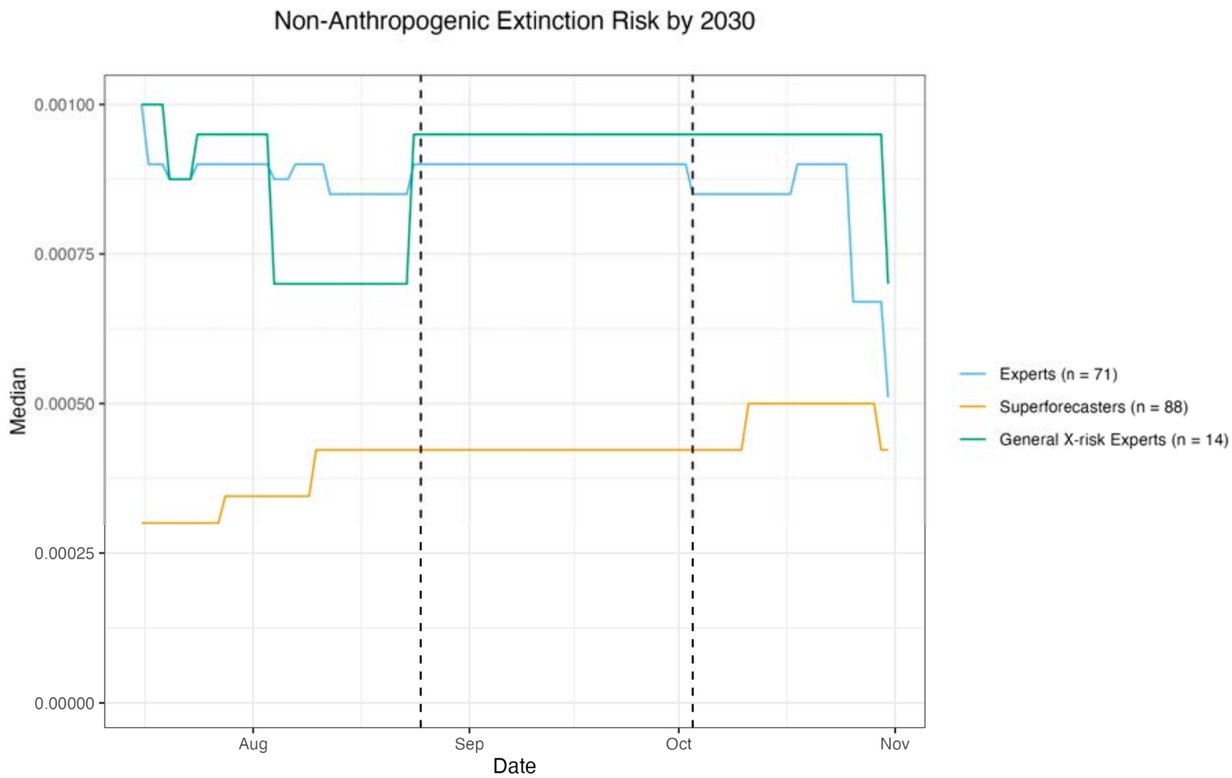
Question and resolution details, prior forecasts, and other relevant sources

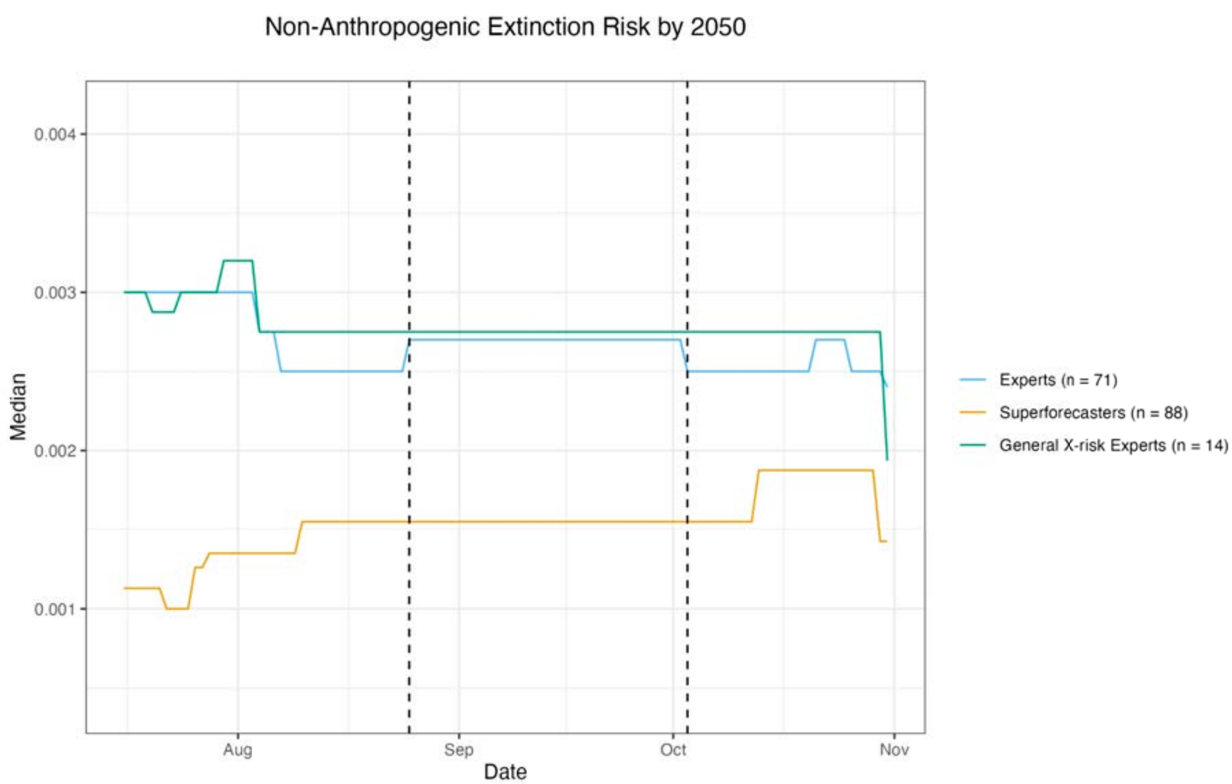
Results[401]

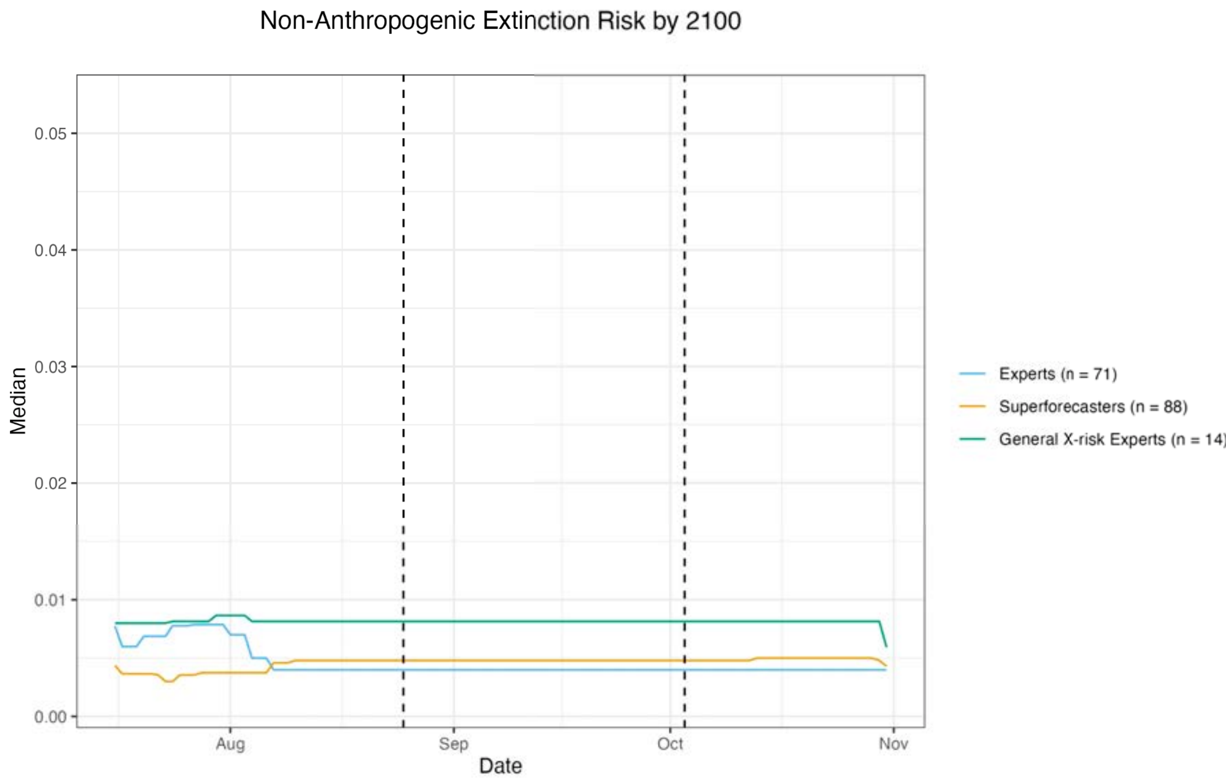
Group	**Year**	**Stage 1 Median**	**Stage 4 Median**	**Stage 1 Standard Deviation**	**Percent Change in Standard Deviation, Stage 1 to Stage 4**
Super-Forecasters (N = 88)	2030	0.0005%	0.0004%	0.18	-18.15%
	2050	0.0014%	0.0014%	0.77	-17.52%

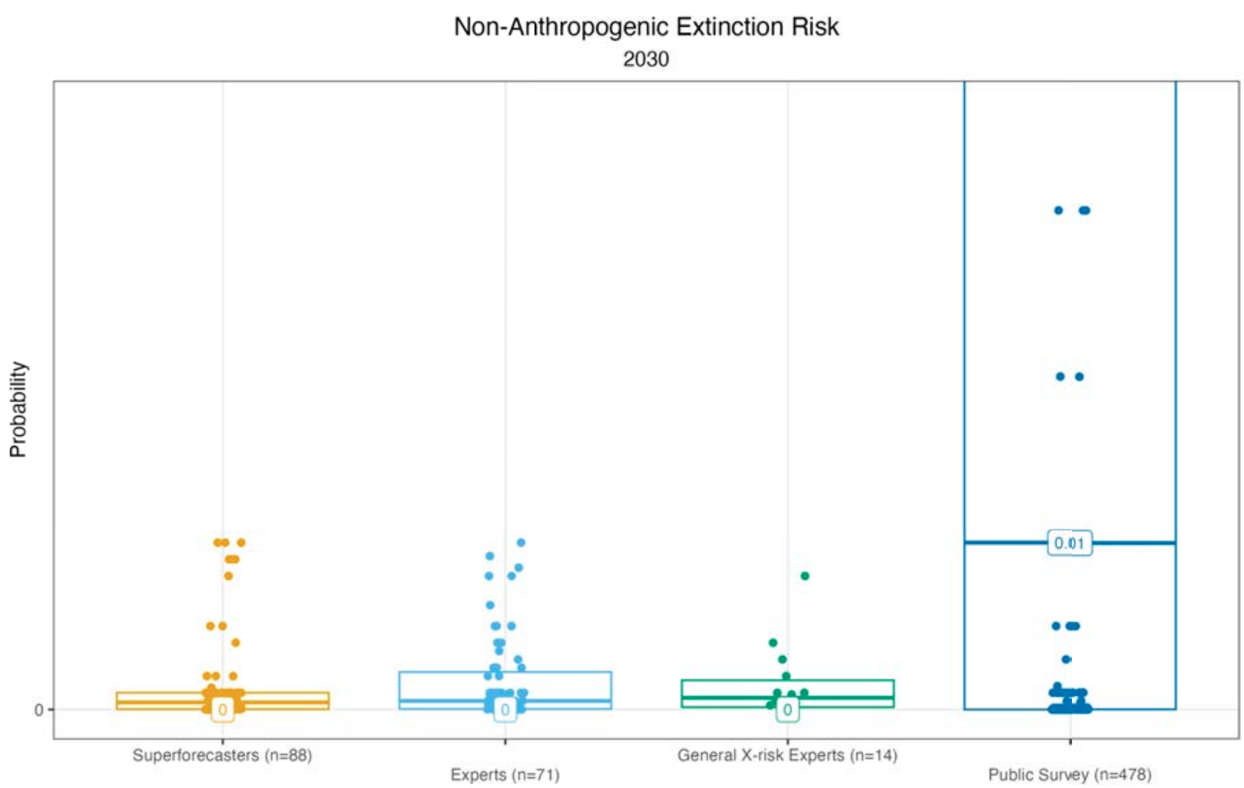
[400] "Higher forecasts are sometimes explained by adjustments for "anthropic bias", cited by several forecasters who adjusted their forecasts upward. Anthropic bias is explained in a paper by Cirkovic and Sandberg (2010) as an underestimation of catastrophic risk due to calculations relying on base rates, which by their nature must be observed by someone, whereas catastrophes that destroy all observers are naturally excluded from the base rate. Including the possibility of catastrophe due to intervention of extraterrestrial life also increased the probablility."
[401] Numbers of forecasters are given as of Stage 4 of the XPT. Public survey respondents were surveyed outside of the tournament context.

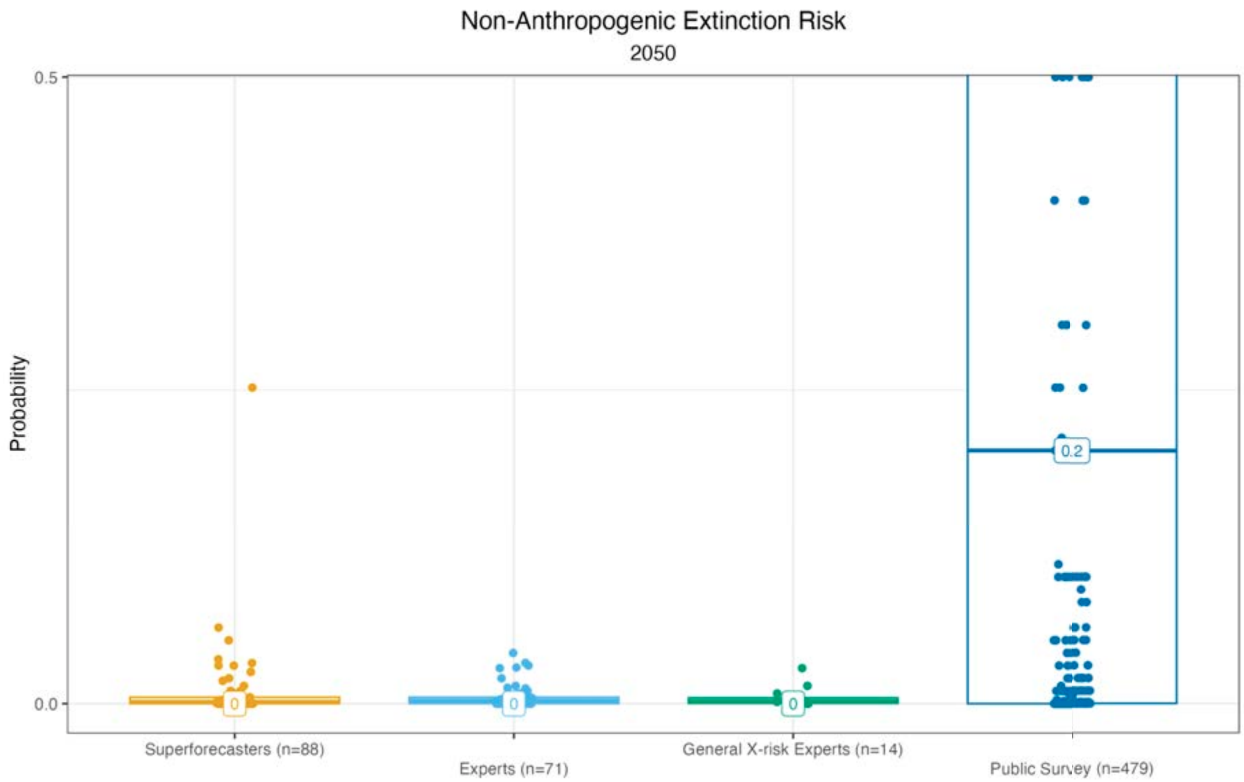
	2100	0.005%	0.0043%	1.31	-24.56%
Experts (N = 71)	2030	0.00095%	0.00051%	1.2	-1.05%
	2050	0.003%	0.0024%	0.86	-1.55%
	2100	0.0078%	0.004%	0.62	-3.71%
General X-Risk Experts (N = 14)	2030	0.00088%	0.0007%	0.002	-6.81%
	2050	0.0029%	0.0019%	0.008	+1.89%
	2100	0.008%	0.0059%	0.02	+1.95%
Public Survey (N = 480)	2030	0.01%		11.23	-
	2050	0.2%		14.99	-
	2100	1%		76.83	-

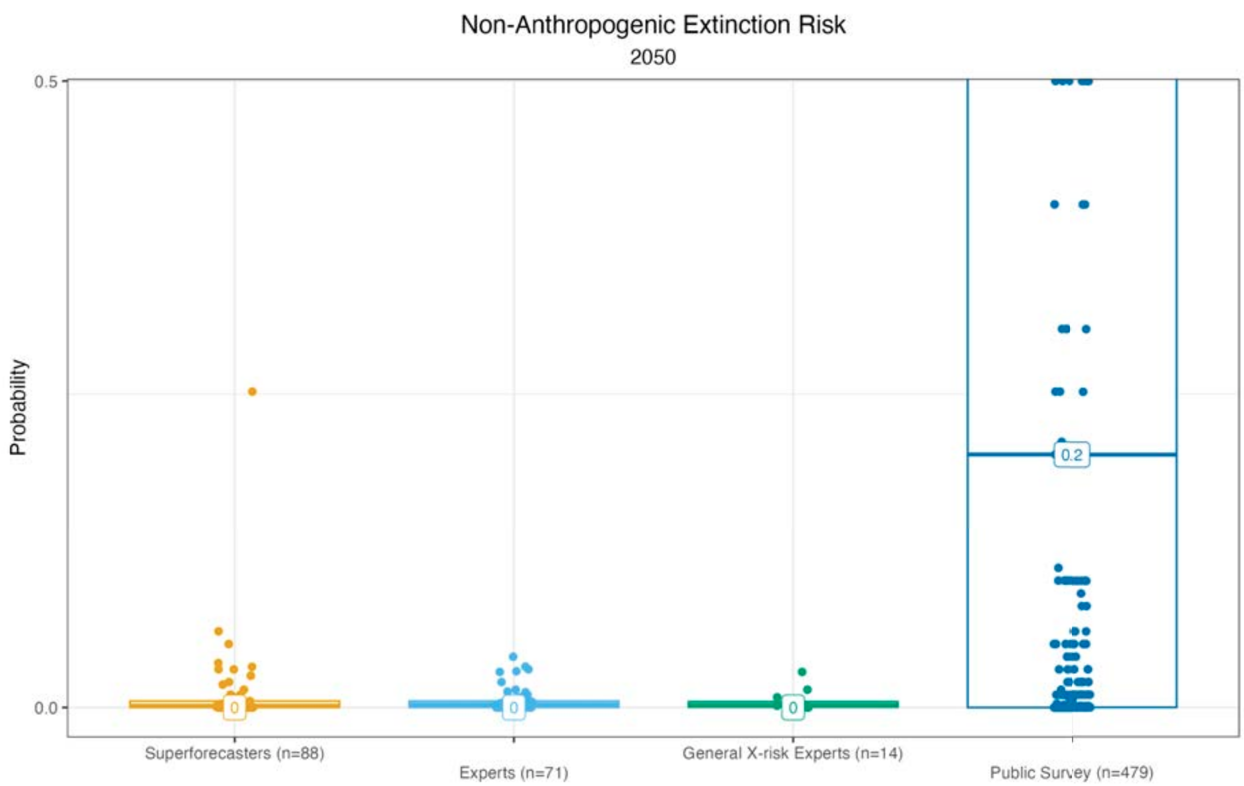

Non-Anthropogenic Extinction Risk by 2030
Median
0.00100
0.00075
0.00050
0.00025
0.00000
Aug
Sep
Oct
Nov
Date
Experts (n = 71)
Superforecasters (n = 88)
General X-risk Experts (n = 14)


Non-Anthropogenic Extinction Risk by 2050
Median
0.004
0.003
0.002
0.001
Aug
Sep
Oct
Nov
Date
Experts (n = 71)
Superforecasters (n = 88)
General X-risk Experts (n = 14)


Non-Anthropogenic Extinction Risk by 2100
Median
0.05
0.04
0.03
0.02
0.01
0.00
Aug
Sep
Oct
Nov
Date
Experts (n = 71)
Superforecasters (n = 88)
General X-risk Experts (n = 14)


Non-Anthropogenic Extinction Risk
2030
Probability
0
0.01
Superforecasters (n=88)
Experts (n=71)
General X-risk Experts (n=14)
Public Survey (n=478)


Non-Anthropogenic Extinction Risk
2050
Probability
0.5
0.0
0.2
0
Superforecasters (n=88)
Experts (n=71)
General X-risk Experts (n=14)
Public Survey (n=479)


Non-Anthropogenic Extinction Risk
2050
Probability
0.5
0.0
0.2
0
Superforecasters (n=88)
Experts (n=71)
General X-risk Experts (n=14)
Public Survey (n=479)

Sources of agreement, disagreement and uncertainty

Sources of agreement

- Low level of extinction risk from non-anthropogenic causes
    - Forecasters generally agreed that the risk of non-anthropogenic risk was very low and lower than risk posed by human activity.[402]

Sources of disagreement and uncertainty:

- Method for producing the forecast
    - Some forecasters derived an estimate from summing probabilities of different non-anthropogenic events that could pose an extinction risk, such as supervolcanic eruption and an asteroid colliding with Earth. Others used historical data to calculate the historical base rate of any extinction-risk event.
- How to evaluate published risk estimates
    - Many of the teams referred to published estimates of non-anthropogenic extinction risk from Toby Ord's 2020 book, The Precipice, and a 2019 paper by Snyder-Beattie, Ord, and Bonsall. Ord's estimate summed different types of risks and adjusted to account for survivorship bias. Snyder-Beattie et al. use the duration of existence of homo sapiens and the genus homo to generate an upper bound of extinction risk. Forecasters differed in their assessment of these estimates and their underlying assumptions.[403]
- The types of events that might produce extinction
    - Amongst those forecasters who summed the probabilities of different events, there was disagreement about the type of events to include. The possibility of contact with alien life was included by some forecasters but not others.[404] There was disagreement about whether solar flares could plausibly cause extinction.[405]

---

[402]337, "Virtually everyone thought the probability of extinction was extremely small; well under 1% for all time frames." 339, "To resolve in the affirmative sets a VERY HIGH BAR, one our team does not think could happen, especially short term." 344, "Non-anthropogenic human extinction is a dead cert… but not for a few billion years. This Century, it's extremely unlikely …"

[403] 337, "Another area of uncertainty is the extent to which cited experts have constructed sufficiently credible probabilities. There were a few attempts at dissection but much more could probably be done in that area, such as going to primary sources and evaluating the rationale used by a given expert in the specific context of this question."

[404] 340, "There was also dispute over how best to account for the possibility of alien risk and its estimated likelihood, with it being excluded from most forecasts and dominating others." 336, "One forecaster is confident about a hostile alien encounter as much more likely than the geologic, comet, asteroid, solar flare types of non-intelligent risk. Based on an extended version of the Dark Forest Hypothesis he believe the threat from extraterrestrials is by far the largest risk in this category, and that it is significant, and that there are good reasons to believe it is uniquely concentrated on this century. "

[405] 339, "The questions surround the resilience of North American, and Northern European electrical systems, and their ability to withstand a major solar flare.Would the destruction of infrastructure be sufficient to start a domino effect that could kill 10% of the people, let alone all of mankind?" 344, "Meanwhile, solar flares and coronal mass ejections (CMEs) are much more common. A large CME could take out satellites, disrupt power grids and take down telecoms and the internet. This would cause worldwide chaos, but nothing near extinction." 336, "Of the possible extinction risks, I would think solar flares are the least likely to have the magnitude of damage necessary."

  - Most forecasters included supervolcanic eruption and asteroid collision as plausibly causing extinction, but some forecasters thought it was implausible that a supervolcanic eruption could lead to human extinction.[406]
- How technology influences risk
  - Forecasters disagreed on how humanity's improving technology influences risk.[407] Many argued that improvements in technology would make humanity more robust to non-anthropogenic risks.[408] However, others thought that increasing reliance could cause vulnerability, should an event, such as a solar flare, cause technology to fail.[409]

Arguments given for forecasts below the median (0.0005% by the end of 2030, 0.0014% (2050), 0.0039% (2100))

- Difficulty of human extinction within the question time-frame
  - Several teams noted that killing all but 5,000 people would be very difficult, particularly given the adaptability of humanity,[410] our large population,[411] and global spread.[412]
  - The time-frame of the question was also highlighted as a reason for a lower forecast, as some argued that the events that might plausibly pose an extinction risk would be unlikely to lead to extinction within a five-year time-frame.[413]
- Expected improvement in defenses

---

[406] 338, "Arguments why nuclear winter probably wouldn't make humanity extinct also apply to cooling caused by asteroids or volcanoes."

[407] 340, "A continuous point of disagreement between many forecasts was whether increasing human technology made our species more fragile or resilient to risks."

[408] 338, "However, as the century progresses, it's increasingly likely that either ordinary gradual progress or transformative technologies like artificial general intelligence will make humanity resilient enough to weather these disasters without going extinct." 343, "One expert notes that with the rise of novel technology and advanced data analysis, it is likely that an existential threat can be eliminated /mitigated or will allow 'post-apocalyptic' adaptation."

[409] 339, "The questions surround the resilience of North American, and Northern European electrical systems, and their ability to withstand a major solar flare.Would the destruction of infrastructure be sufficient to start a domino effect that could kill 10% of the people, let alone all of mankind?"

[410] 337, "Foremost of these is that the extinction of humanity was deemed to be extremely difficult since humans can adapt, or at least construct mitigants, to overcome significant adverse environmental conditions." 336, "The adaptability of humans."

[411] 340, "Another forecast believed humanity's large population makes it resilient to extinction events."

[412] 338, "humans have spread all over the planet and can adapt more intelligently to disasters"

[413] 338, "mass extinction(s) last a long time usually, so we can be pretty sure we will not be very far into a non-anthropogenic one by 2100, given the current situation." 337, "The problem with existing forecasts on this subject is that they don't adequately consider the likely time lag between an extinction-level event and extinction. IMHO, the only non-anthropogenic event that has a realistic chance of causing total extinction, or for the population to be reduced by 5,000 in the next 78 years, is a massive asteroid impact. The other risks, even if they were to occur over the time period in question, and be of sufficient severity as to cause extinction eventually, would almost certainly not do so immediately." 340, "previous mass extinction events took a long time to kill of the relevant population."

    - Several teams noted that historical base rates of risks may not be relevant to humanity's extinction risk in the future as it is likely that we will continue to develop defenses against non-anthropogenic risks.[414]

Arguments given for forecasts above the median (0.0005% by the end of 2030, 0.0014% (2050), 0.0039% (2100))

- Reliance on technology could increase susceptibility to risks
    - While some forecasters saw technology as likely to improve resilience to non-anthropogenic risks, others suggested the reliance on technology could increase susceptibility to risks that might disrupt these systems.[415]
- Unpredictability of risks
    - Several teams noted that it was difficult to predict the likelihood of events that might pose extinction-risks, and that there may be types of risk events of which we're currency ignorant.[416] One person pointed to the relatively recent discovery of risks from supervolcanoes as a sign that we may be ignorant of relevant risks.[417]

Cross-references with other questions

Q7: Non-Anthropogenic Catastrophic Risk
Q10: Total Existential Risk

Question 9: Total Catastrophic Risk

What is the overall probability of a global catastrophe where more than 10% of humans alive at the start of a 5-year period die by the end of that period…
...by the end of 2030?
...by the end of 2050?
...by the end of 2100?

---

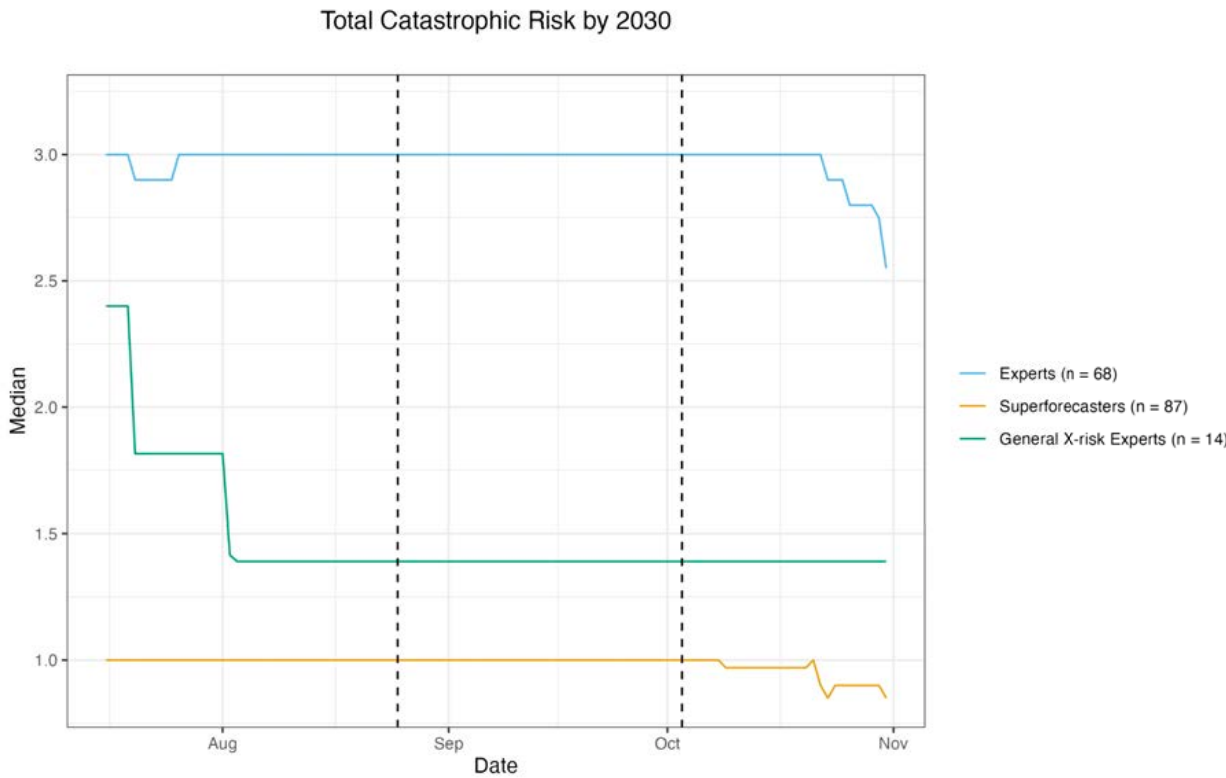
[414] 338, "However, as the century progresses, it's increasingly likely that either ordinary gradual progress or transformative technologies like artificial general intelligence will make humanity resilient enough to weather these disasters without going extinct." 336, "As humans become more scientifically sophisticated, our ability to predict and mitigate natural events increases."
[415] 342, "there are some scenarios in which damage to communication satellites through solar storms significantly contribute to ineffective management of other catastrophic risks and this vulnerability is bound to increase in scope as dependence on satellites increases."
[416] 339, "The paper :Human Extinction from Natural Hazard Risks lays out a case for a higher forecast based upon the unpredictability of these events. "Many extinction hazards are at present impossible to prevent or even predict, requiring resilience strategies. "
[417] 342, "he's aiming a bit higher than the expert view to make up for the fact that experts tend to ignore dragon king events in non-anthropogenic risk and there have been a few historical examples for late discovery (a. i. scientists not knowing about super volcanoes for a long time when they maybe could have)."

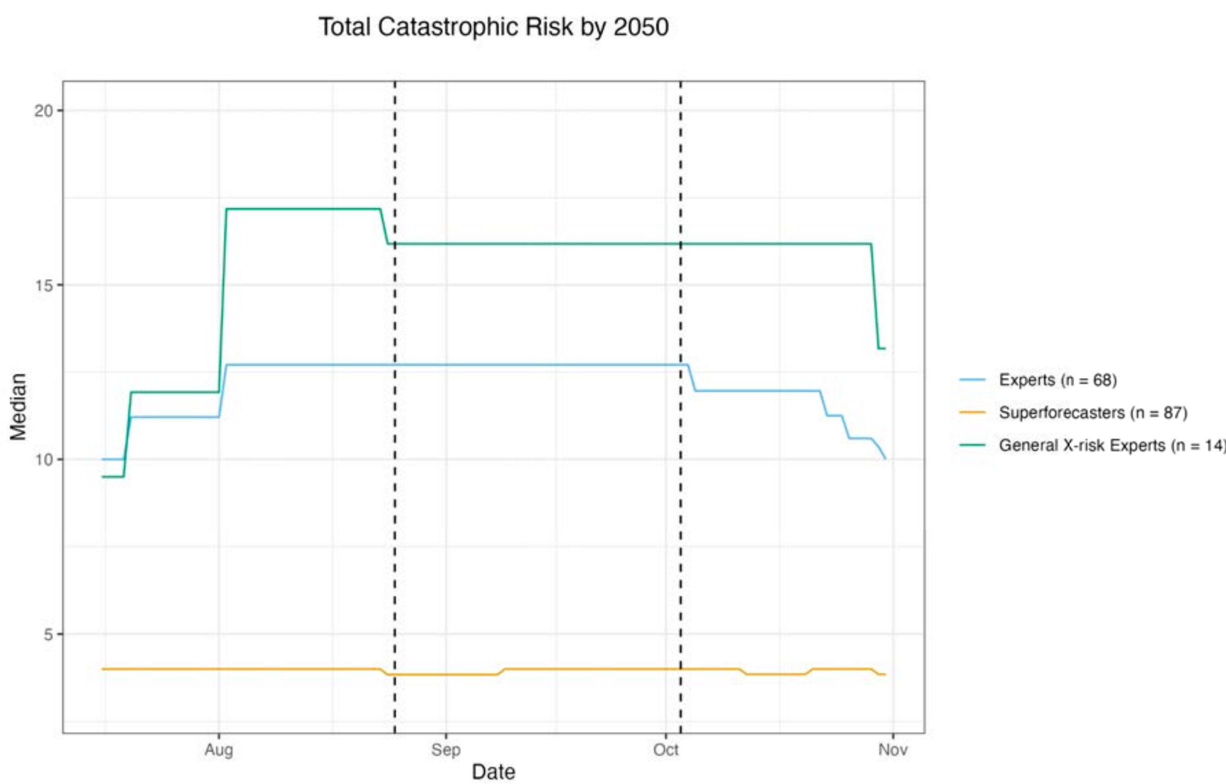
Question and resolution details, prior forecasts, and other relevant sources

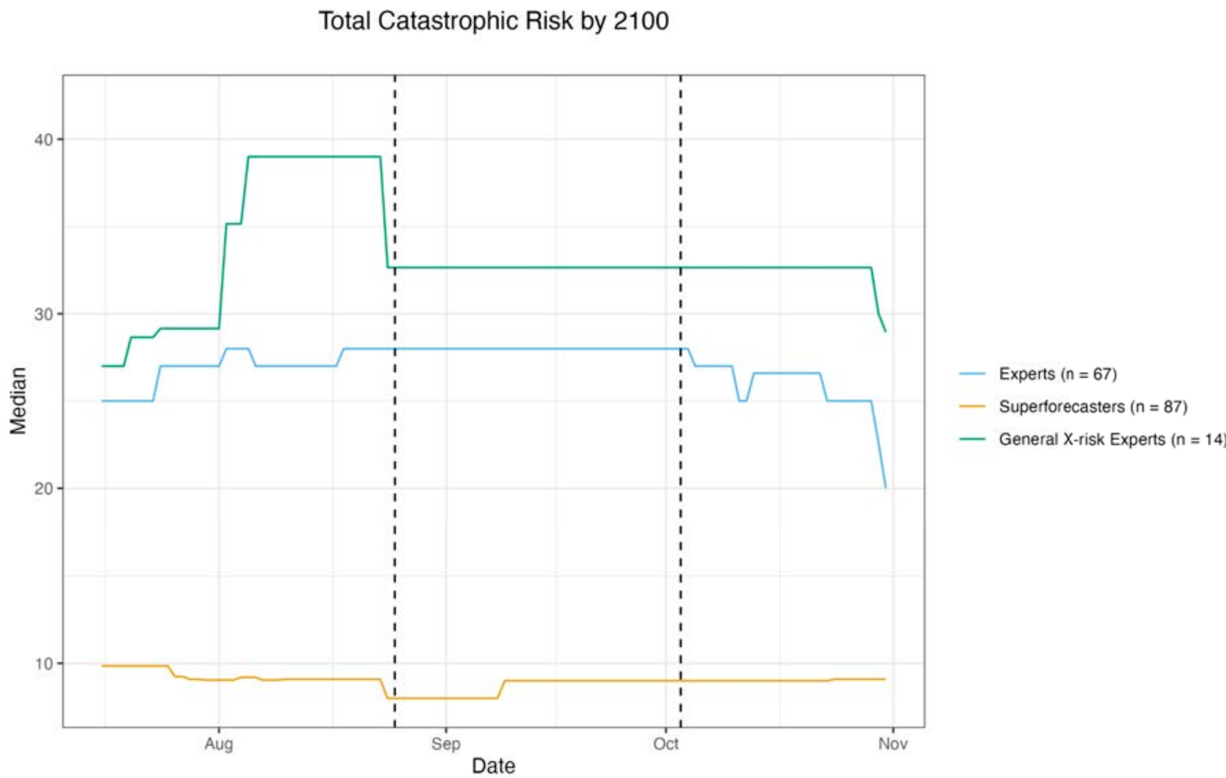
Results[418]

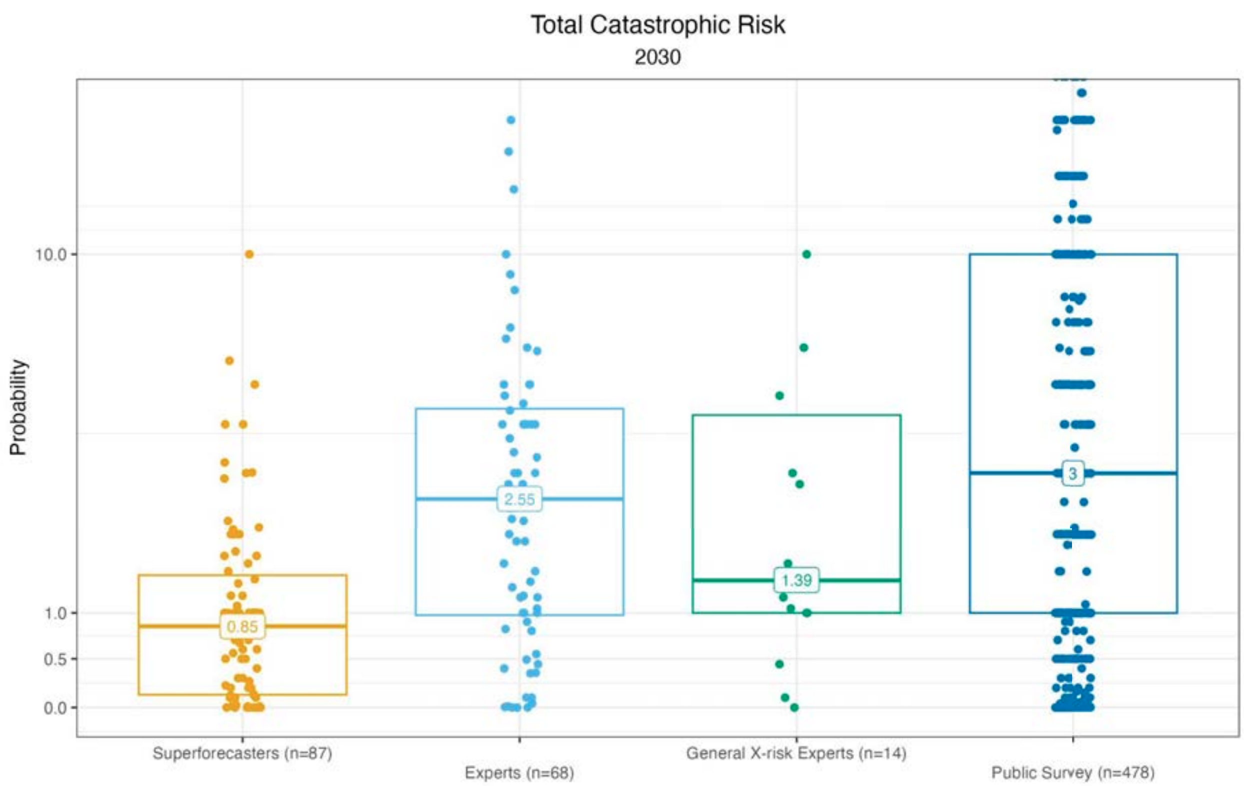
Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 87)	2030	1%	0.85%	6.93	-43.41%
	2050	3.66%	3.85%	14.99	-34.25%
	2100	9.09%	9.05%	22.39	-42.35%
Experts (N = 68)	2030	3%	2.55%	8.78	-9.26%
	2050	10.5%	10%	17.32	-4.66%
	2100	25%	20%	25.57	-9.67%
General X-Risk Experts (N = 14)	2030	2.45%	1.39%	11.29	-3.4%
	2050	12.43%	13.18%	19.6	-6.62%
	2100	28.65%	28.95%	27.85	-6.36%
Public Survey (N = 478)	2030	3%		229.72	-
	2050	6%		22844.92	-
	2100	11.56%		22846.7738	-

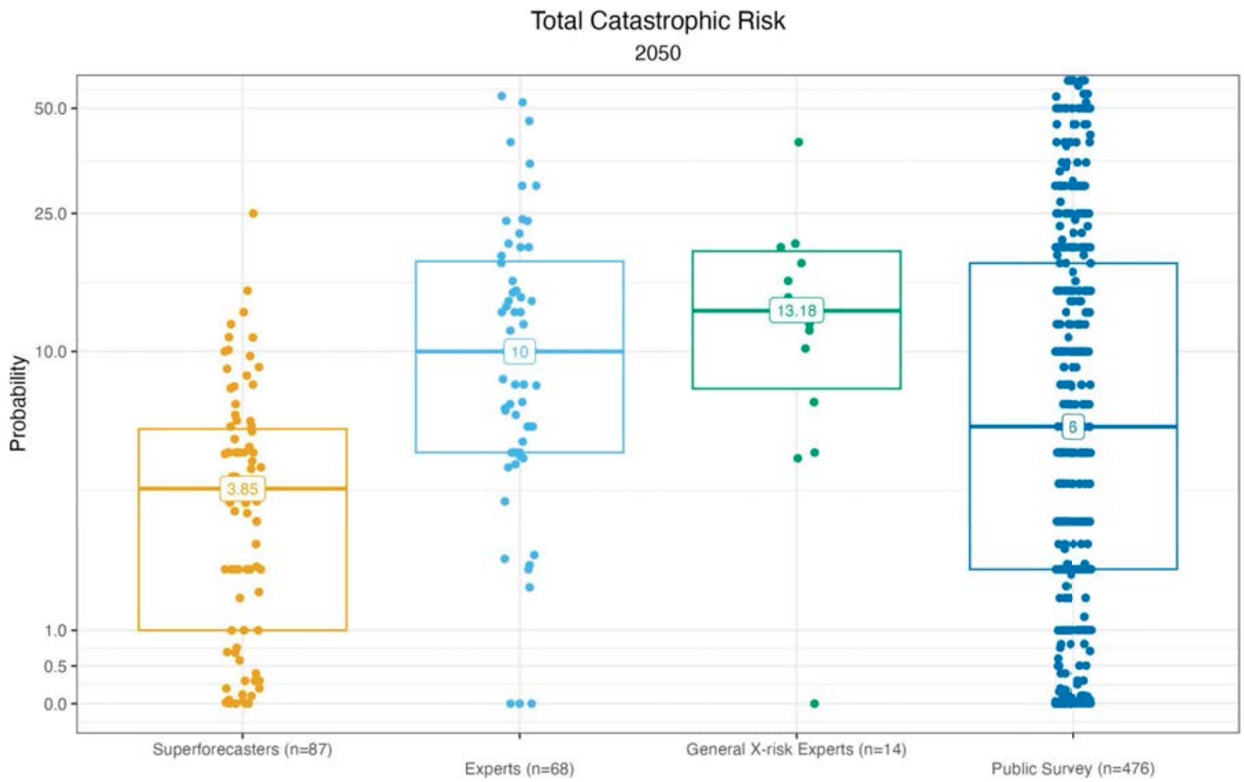
[418] Numbers of forecasters are given as of Stage 4 of the XPT. Public survey respondents were surveyed outside of the tournament context.

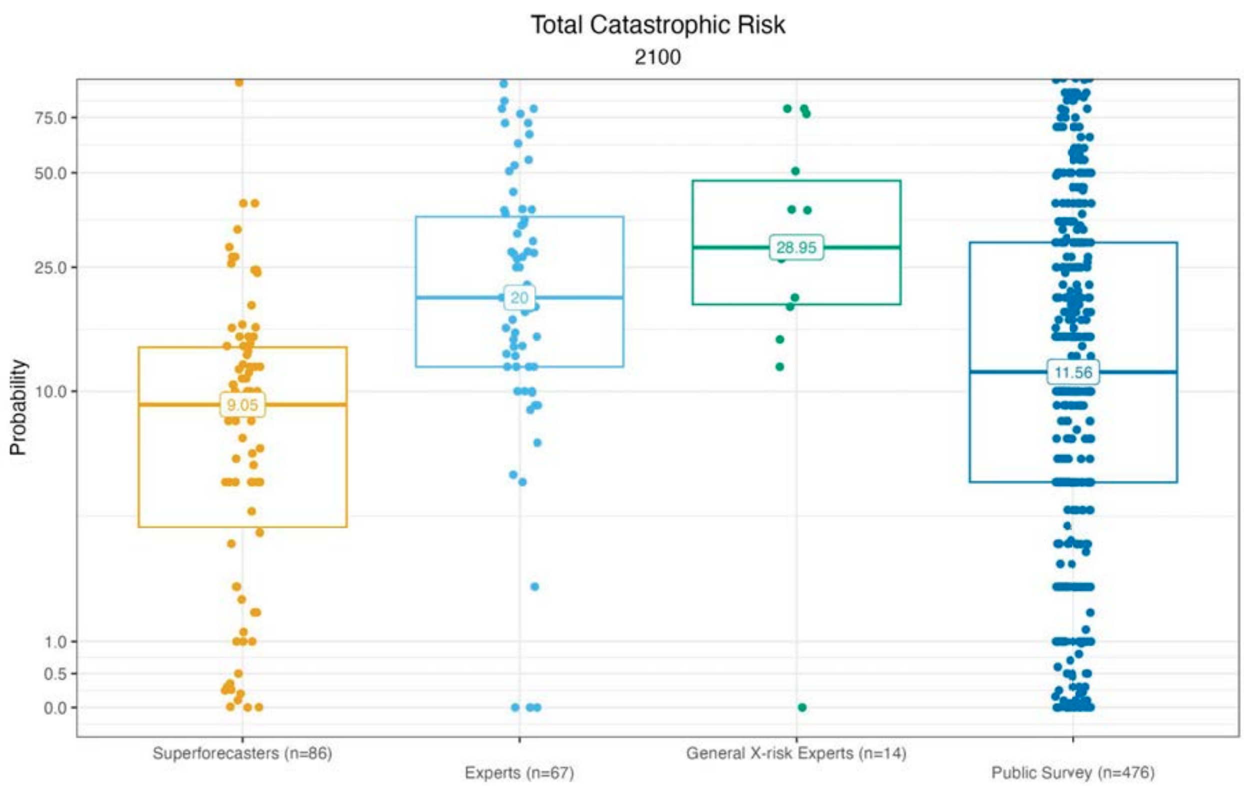

Total Catastrophic Risk by 2030
Median
3.0
2.5
2.0
1.5
1.0
Aug
Sep
Oct
Nov
Date
Experts (n = 68)
Superforecasters (n = 87)
General X-risk Experts (n = 14)


Total Catastrophic Risk by 2050
Median
20
15
10
5
Aug
Sep
Oct
Nov
Date
Experts (n = 68)
Superforecasters (n = 87)
General X-risk Experts (n = 14)


Total Catastrophic Risk by 2100
Median
40
30
20
10
Aug
Sep
Oct
Nov
Date
Experts (n = 67)
Superforecasters (n = 87)
General X-risk Experts (n = 14)


Total Catastrophic Risk
2030
Probability
10.0
1.0
0.5
0.0
0.85
2.55
1.39
3
Superforecasters (n=87)
Experts (n=68)
General X-risk Experts (n=14)
Public Survey (n=478)


Total Catastrophic Risk
2050
Probability
50.0
25.0
10.0
1.0
0.5
0.0
3.85
10
13.18
6
Superforecasters (n=87)
Experts (n=68)
General X-risk Experts (n=14)
Public Survey (n=476)


Total Catastrophic Risk
2100
Probability
75.0
50.0
25.0
10.0
1.0
0.5
0.0
9.05
20
28.95
11.56
Superforecasters (n=86)
Experts (n=67)
General X-risk Experts (n=14)
Public Survey (n=476)

Sources of agreement, disagreement and uncertainty

In general, the variance of forecasts for this question was comparatively low, and there was broad agreement about what the important factors were:

- AI and nuclear weapons were generally seen as likely the biggest risks.
- Pathogens, both natural and anthropogenic, were a somewhat more controversial third.
- Consequences from climate change, especially war, famine, disease, and heat waves, were also considered.
- Non-anthropogenic risks—in one team's formulation: "solar flares, coronal mass ejections, asteroid collisions, pandemics, supervolcanic eruptions" (team 344)—were often mentioned, but their risk (aside from pandemics) was generally considered low to negligible.

There was disagreement on the following points:

1. Whether risks' interdependence (a fact generally agreed upon) meant the total calculated risk should be lower or higher than a simple summing of all risks.
2. What the base rate is, and to what degree it should be anchored on.
3. Whether the nature of the modern world—especially its interconnectedness and advanced technology—lowers or raises the risk, compared to the past. (All teams that mentioned this considered pros and cons; there were merely differences of emphasis.)
4. AI risk magnitude.
5. Nuclear risk magnitude.
6. Pathogen risk magnitude.
7. Climate change risk magnitude.

Arguments given for forecasts below 1.925% (2030), 7.45% (2050), 13% (2100)

On 1 (risks' interdependence):

- Risks' interdependence lowers the overall probability.[419] One team gives a hypothetical: "Imagine that one risk is the cause of another. For example, an AI causes a nuclear catastrophy [sic] such that both question 3 and question 5 resolve yes. If this is possible, then the component probabilities are not independent and overlapping probability needs to be subtracted" (338).

On 2 (base rate):

- The base rate of a global catastrophe that killed more than 10% of humans in a 5-year period is low, if not zero: "Another strong argument for the lower end of the plausible

[419] 337, "some of the catastrophic events mentioned above may not be independent, and as such their probabilities cannot simply be added," which decreases total risk. Note also that 337 goes on to say: "the list of catastrophic risks covered in other questions may not be exhaustive," which increases risk, and these two considerations cancel each other out.

range stems directly from a lack of events within recorded history that resulted in the death of 10% of the human population within a 5-year period." (337)
- Changes in society and technology may render previous base rates less relevant, and those changes make the overall risk lower.[420]

On 3 (modern interconnectedness and advanced tech):
- Our interconnectedness and advanced technology "make us better positioned to handle and prevent catastrophic risks," but also "make us uniquely vulnerable" (337).[421]

On 4 (AI):
- AI risk is low,[422] and, in the words of Team 345, "overhyped."[423]

On 5 (nuclear):
- Nuclear risk is low.[424]

On 6 (pathogens):
- Pathogen risk is low.[425]
- Natural pathogens are probably a bigger risk than engineered pathogens, though that could change in the future.[426]

On 7 (climate change):
- The risk of climate change is "unlikely to sway the anchor by much."[427]

---

[420] 337, "A major source of uncertainty with this question, and many similar ones, is the lack of a transferable base rate. Humanity may have witnessed catastrophic events in its history (recorded and otherwise), but it is difficult to ascertain whether 10% of humans died in a 5 year period, and it is difficult to generalize to our current setting. Our current times differ partly due to the unprecedented density of human connectedness and our relatively newfound ability to destroy ourselves, or almost destroy ourselves. They also differ, however, due to technological and medical breakthroughs that should, at least theoretically, allow us to mitigate some of the catastrophic risks we may face."
[421] Full passage: "We are at a unique point in our history as a species, being more interconnected than ever, more technologically advanced than ever and improving our technologies at a faster rate than ever before. The forecasters on the upper and lower ends of the plausible probabilities differ in interpreting whether these unique circumstances make us better positioned to handle and prevent catastrophic risks, or whether they make us uniquely vulnerable."
[422] 336, "Strongest arguments for a forecast on the lower side: lower assessment of risk from AI."
[423] 345, "the team is extremely skeptical that even the worst catastrophes could result in the deaths of 10% of the population. It would take an extreme anomaly to disrupt the world to the point where any of the risks looked at even come into the realm of possibility. AI risks are overhyped."
[424] 336, "Strongest arguments for a forecast on the lower side:[...] lower forecasts for nuclear [...] risk."
[425] 336, "Strongest arguments for a forecast on the lower side:[...] lower forecasts for [...] pathogen risk."
[426] 339, "While genetically modified pathogens could be more lethal than natural pathogens, the team placed more risk on natural pathogens because of the difficulty in creating biological weapons, including the lack of effective delivery systems, but acknowledged technology breakthroughs in the future could make the biological weapon risk more severe."
[427] 336, "Climate change and/or bad geoengineering might also be a risk, but is unlikely to sway the anchor by much)."

Additionally, Team 345 offered two further arguments not mentioned by other teams:

1. "Global stability is probably underrated."
2. "Humanity will soldier on through whatever natural or man made cataclysmic events come our way."

Arguments given for forecasts above 2% (2030), 5.524695% (2050), 19.340053% (2100)

On 1 (risks' interdependence):

- Interdependence of risks has the potential to raise total risk, because, for example, if "nuclear conflict causes massive death and destruction across the Earth, this should be expected to impact other potential sources of catastrophic risk. Therefore interactions among potential sources of risk should be considered at higher levels of risk" (343).

On 2 (base rate):

- The conditions of the question have been met at least once (Great Plague of Justinian) before, and maybe (Black Death) twice.
  - Team 339 wrote: "The Black Death of the 1340s felled more than 10 percent of the world population. Eight centuries prior, another epidemic of the Yersinia pestis bacterium—the "Great Plague of Justinian" in 541 and 542—killed between 25 and 33 million people, or between 13 and 17 percent of the global population at that time."
- The past base rate of a global catastrophe that killed more than 10% of humans in a 5-year period may be zero, but new risks have made that base rate obsolete, and those new risks make the overall risk higher.
  - Team 340 wrote that "we have not witnessed a catastrophe killing more than 10% of all humans alive at the start of a 5 year period yet," but later argued that "there are reasons to expect the future base rate to be higher than the historical rate," according to the following logic: "The bulk of the catastrophic risk we face today comes from new anthropomorphic risks that we can reasonably hypothesize about (AI, engineered pathogens, etc); OR prior risks that can be reasonably assumed to be higher now than they were in the past (e.g. second-order effects of climate change like war, for example)."

On 3 (modern interconnectedness and advanced tech):

- Interconnectedness causes fragility, but people in different locations help each other.[428]

[428] 343: "It's unclear to what extent modern society represents increased or decreased fragility. Complex, interconnected commercial systems appear quite complex and fragile, as recently demonstrated with COVID-19. However, interconnected systems also allow local/regional catastrophes to be mitigated by cooperative/collective efforts from outside the region."

- Interconnectedness causes susceptibility to pathogens, but antibiotics can compensate for this—though engineered pathogens may be a different story.[429]
- Complex systems are "prone to collapse," and increased pressure on the ecosystem, including humans' mental health, will have "negative implications on society's resilience foremost as well as our ability to make humane or even 'rational' choices" (340).[430]

On 4 (AI):
- AI risk accounts for "the bulk of the catastrophic risk we face today" (340).
- "AI-related catastrophes, intentional or unintentional," are a risk (344).

On 4 and 5 (AI and nuclear):
- AI risk and nuclear risk are the only non-negligible risks: "Natural pandemics are [negligible], artificial ones are still very unlikely to kill 800m people, so are a rounding error. Non-anthropogenic causes are also [negligible]. Therefore it is an accounting exercise adding nuclear and AI risk" (342).

On 5 (nuclear):
- Nuclear is the dominant factor in the calculation of overall risk.[431]
- Nuclear risk increases as fewer living people personally witnessed the effects of nuclear war.[432]

On 6 (pathogens):
- Engineered pathogens are a serious risk.[433]

---

[429] 340, "Perhaps, global interconnectedness could lead to similar worldwide death tolls for some pathogen. On the other hand, some of the relevant pandemics seem like they would be a lot less problematic with antibiotics (and requiring antibiotic resistance for a bacterial pandemic seems to drastically lower the probability). But then, there's also engineered pathogens."

[430] Full passage, from 340: "Our planet's climate and its societies consist of complex, adaptive systems of systems. Complex systems tend to return to "homeostasis," but they are also characterized by non-linearity and prone to collapse,. For this reason, my sanguine responses to individual risks -- AI, engineered pathogens, nukes, etc. -- do not necessarily lend themselves to a sanguine outlook with respect to catastrophic risk. What is hard to quantify are the risks of how these things will interact with other societal changes -- especially the increasing global population, interconnectedness/ interdependence (rapid spread of disease), and climate change. More pressure on the ecosystems that sustain life will have unknown and presumably negative implications on society's resilience foremost as well as our ability to make humane or even "rational" choices. (I include overall mental health under this notion of "resilience" and the guardrails/norms that provide some relative stability in societies.) It is in this context that I believe catastrophic risk will steadily escalate and will be a higher probability in any given five-year period over the coming decades."

[431] 340, "For now, I'll go with a 0.4% total annual risk, comprised of the aggregate annual nuclear risk from an US-Russia war forecast of 0.4%, weighed down to 0.2% because such a war is far from certain to kill 10% of the global population, plus another 0.2% for the biorisks and others."

[432] 344, "The question of whether another WWII-level event gets more or less likely as we get farther from it is one I vacillate on; late 20th century was more violent but more cautious because many of them had lived through WWII and the advent of the nuclear age. Today I suspect we are less violent overall but more casual about world war or nuclear weapons because we've not seen them in our lifetimes."

[433] 342, "increase in genetically engineered pathogen risk, which I increased after looking more closely at the information available."

On 7 (climate change):

- War caused by climate change is a large risk.[434]
- Decrease in agricultural production as a result of climate change is a risk.[435]
- Famine as a result of crop failure due to climate change is a risk.[436]

Additionally, Team 342 mentioned two things other teams did not:

1. “Anthropogenic black swans risk.”
2. “Chemical weapons risk.”

Other arguments given

Arguments in favor of lower forecasts:

- Globalization (together with technological progress), which is similar to what other teams have called ‘interconnectedness’, was seen by Team 344 as a factor decreasing certain risks:
    - “Bio-engineered pathogen risk
    - Non-anthropogenic catastrophe risk
    - Early progress in AI safety should mitigate the emergence of non-aligned AI
    - Timely actions to combat climate change will counteract damage already done to the planet” (344)

Arguments in favor of higher forecasts:

- Conventional warfare was considered a risk by Team 338, and analyzed in some depth by 341.
- Diseases caused by climate change were mentioned by 341.
- Team 338 included a list of possible additional causes. Not explicitly mentioned by other teams:
    - “Post-Human Replacement”
    - “Nanotech disaster”
    - “Chaotic collapse/societal collapse”
    - “Future weapons”
    - “Murderous post-human agents that are not AI (e.g. whole brain emulation)).”
- "Arms race in lethal autonomous weapons” is a risk (337).
- "A genetically engineered super-crop pest or super-crop disease or bacteria that outcompetes beneficial bacteria” are risks.[437]

---

[434] 340, “The bulk of the catastrophic risk we face today comes from [among other things,] prior risks that can be reasonably assumed to be higher now than they were in the past (e.g. second-order effects of climate change like war, for example).”

[435] 344, “Climate change and impact on the biome alongside, and more importantly, agricultural production.”

[436] 343, “Climate change (exact mechanism of death likely famine from crop failure).”

[437] 337, “If these were intentionally introduced in lots of places at once, there could be 10% mortality, especially if it disrupted the grass family, which is something like 70% of food/feed production."

Cross-references with other questions

Q3: AI Catastrophic Risk
Q5: Nuclear Catastrophic Risk
Q7: Non-Anthropogenic Catastrophic Risk
Q10: Total Existential Risk

Question 10: Total Extinction Risk

What is the overall probability of human extinction or a reduction in the global population below 5,000…
…by the end of 2030?
…by the end of 2050?
…by the end of 2100?

Question and resolution details, prior forecasts, and other relevant sources

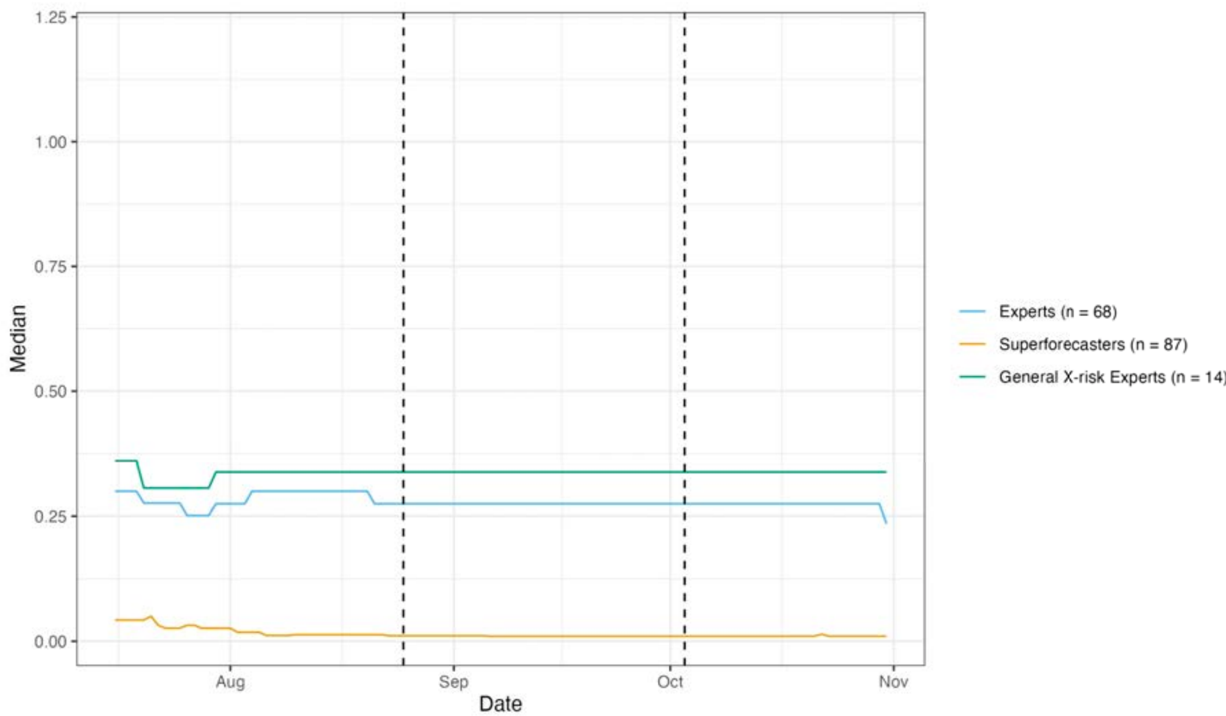
Results[438]

Group	**Year**	**Stage 1 Median**	**Stage 4 Median**	**Stage 1 Standard Deviation**	**Percent Change in Standard Deviation, Stage 1 to Stage 4**
Super-Forecasters (N = 87)	2030	0.035%	0.01%	2.40	-6.69%
	2050	0.5%	0.3%	9.34	-9.46%
	2100	1.05%	1%	12.60	-19.91%
Experts (N = 68)	2030	0.3%	0.24%	7.37	-2.06%
	2050	2.55%	2.4%	15.99	-1.62%

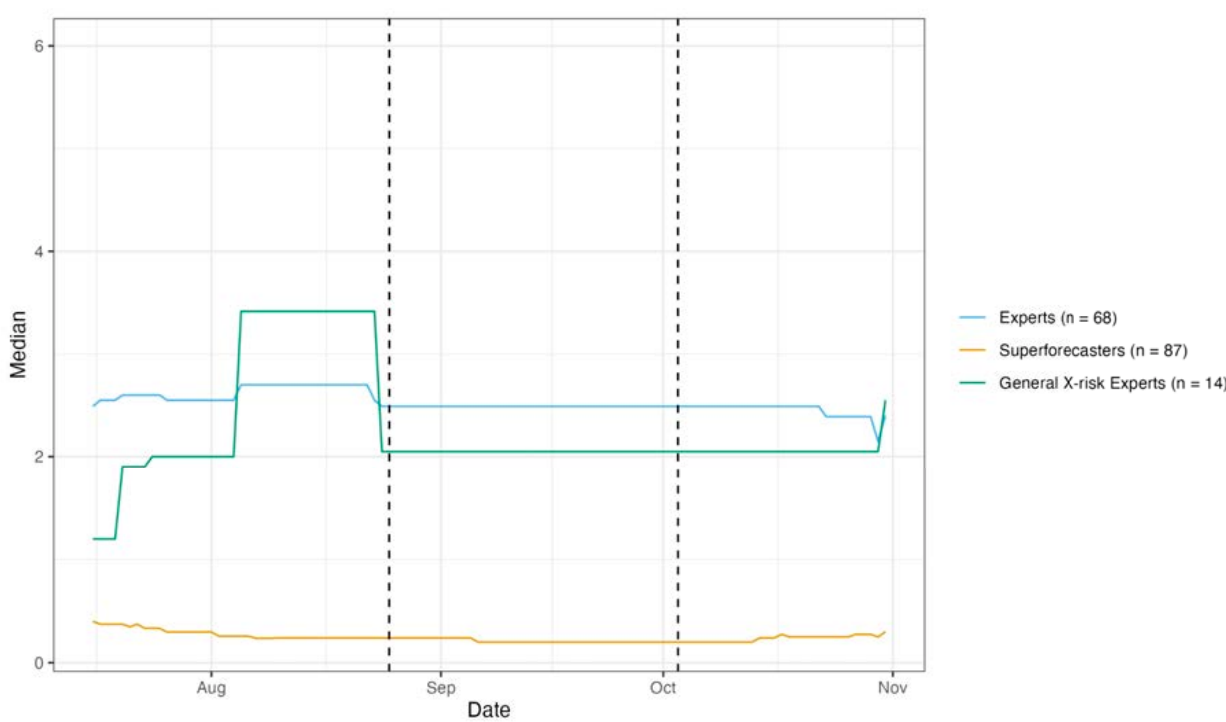
[438] Numbers of forecasters are given as of Stage 4 of the XPT. Public survey respondents were surveyed outside of the tournament context.

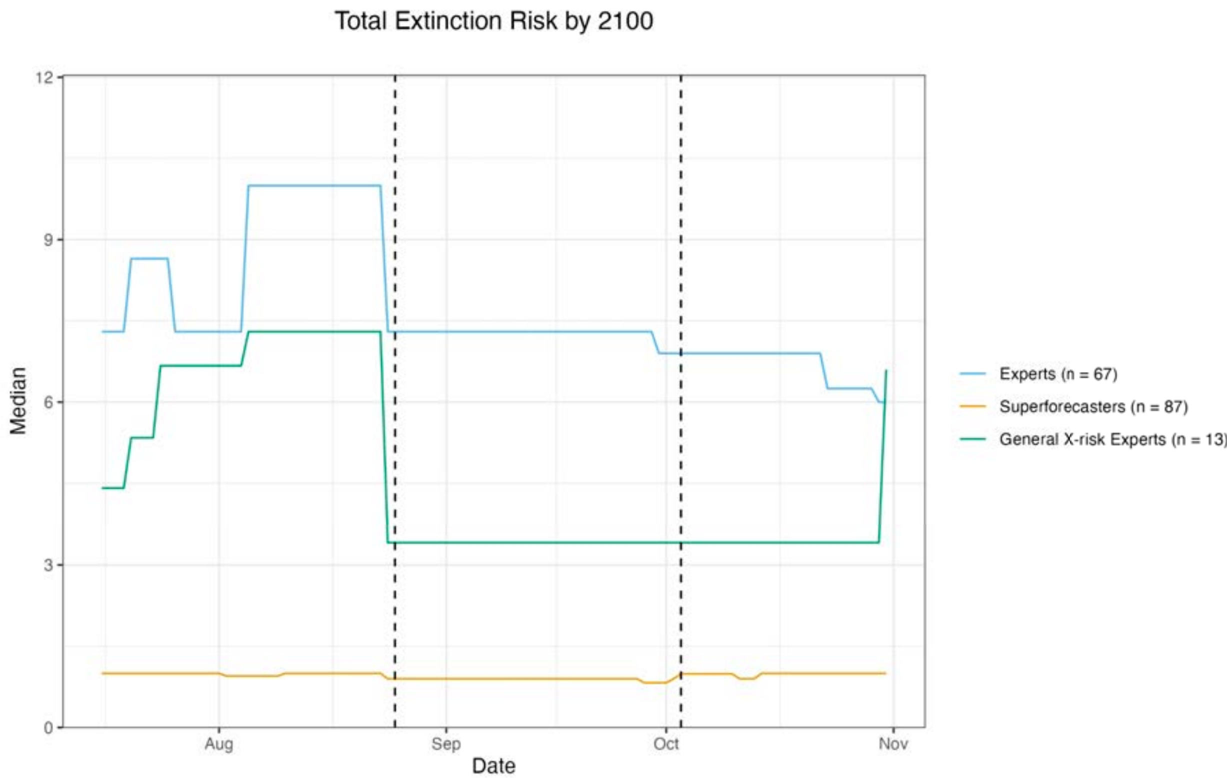
	2100	7.3%	6%	19.95	-3.48%
General X-Risk Experts (N = 14)	2030	0.53%	0.34%	9.44	+0.75%
	2050	4.05%	2.55%	17.77	+0.73%
	2100	7.3%	6.6%	18.95	-0.43%
Public Survey (N = 478)	2030	1%		79.85	-
	2050	2%		82.38	-
	2100	5%		43.74	-

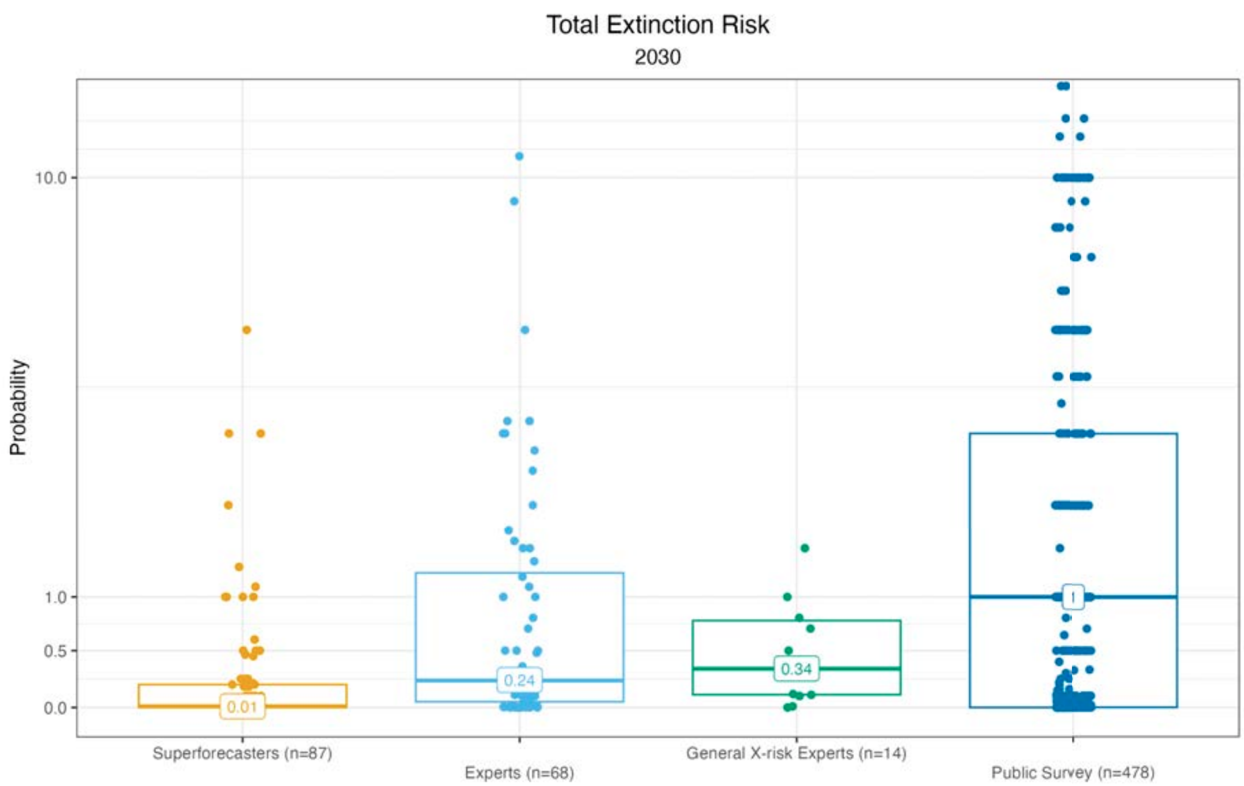
Total Extinction Risk by 2030

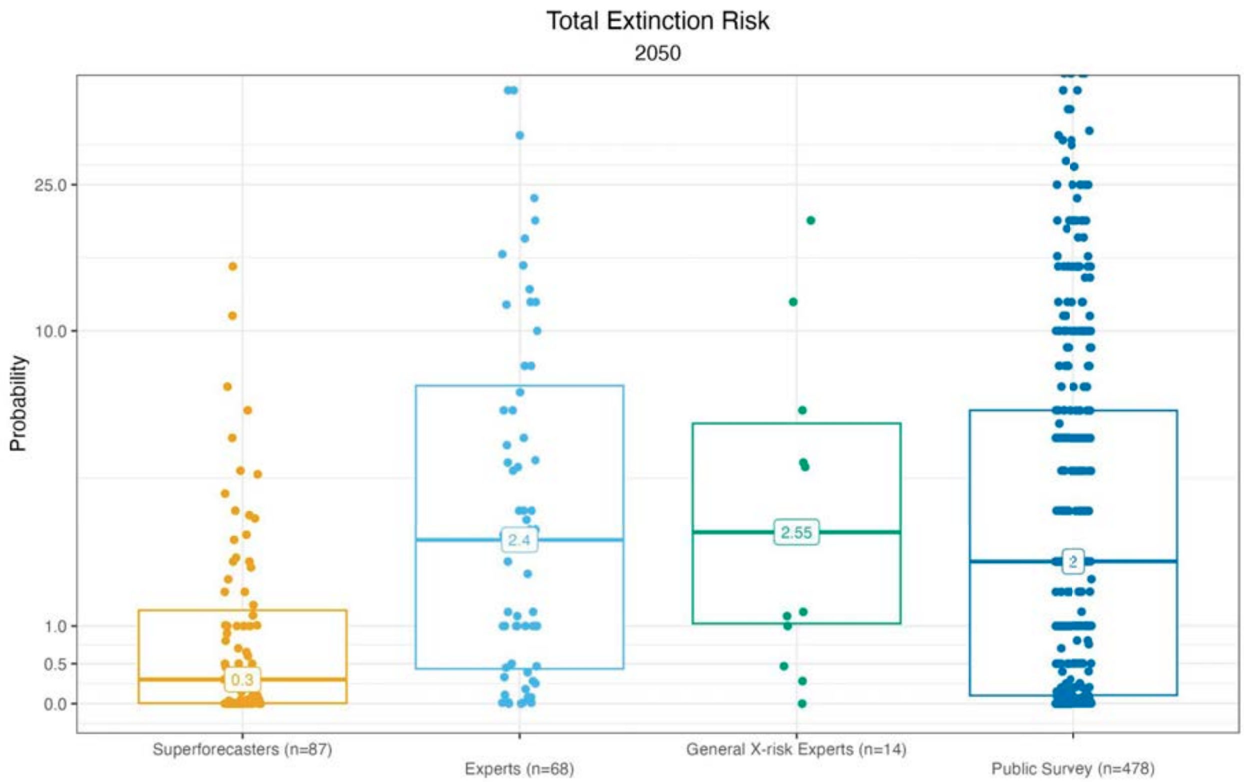


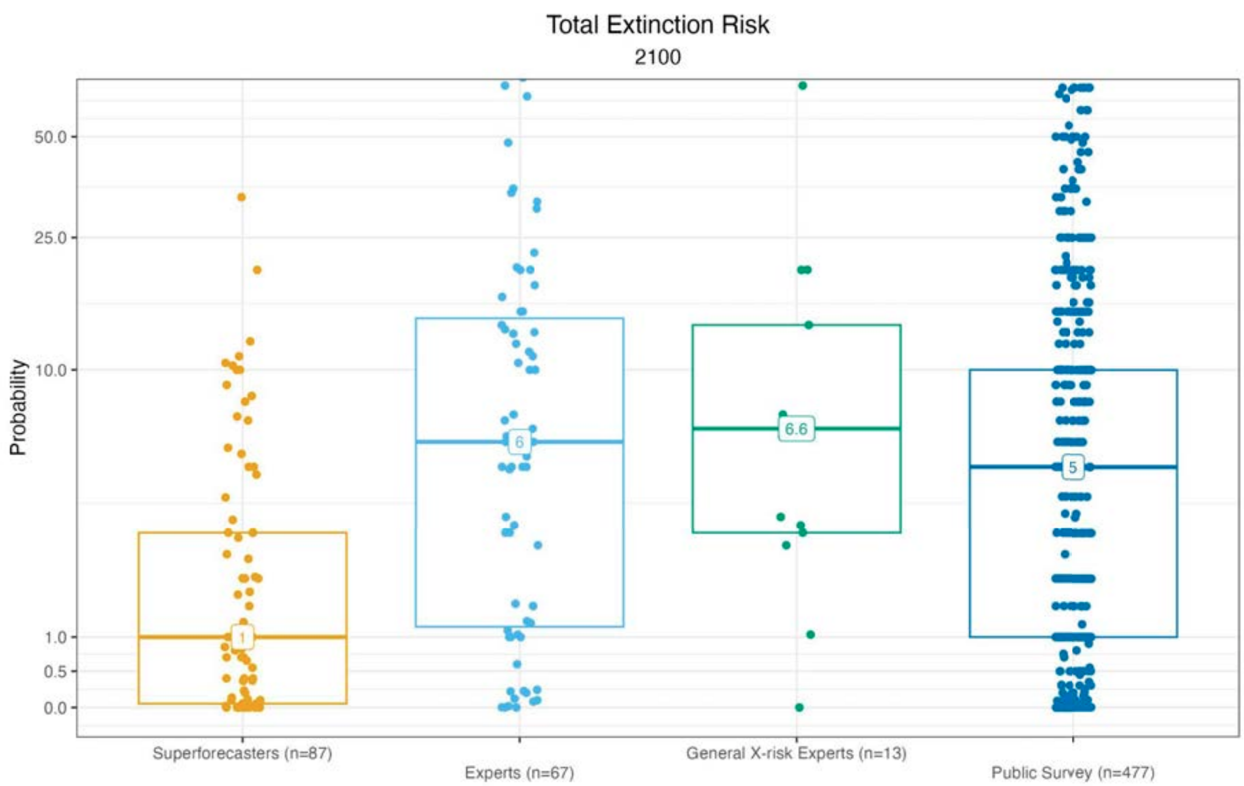
Total Extinction Risk by 2050




Total Extinction Risk by 2100
Median
Date
Aug
Sep
Oct
Nov
0
3
6
9
12
Experts (n = 67)
Superforecasters (n = 87)
General X-risk Experts (n = 13)


Total Extinction Risk
2030
Probability
10.0
1.0
0.5
0.0
0.01
0.24
0.34
1
Superforecasters (n=87)
Experts (n=68)
General X-risk Experts (n=14)
Public Survey (n=478)


Total Extinction Risk
2050
Probability
25.0
10.0
1.0
0.5
0.0
0.3
2.4
2.55
2
Superforecasters (n=87)
Experts (n=68)
General X-risk Experts (n=14)
Public Survey (n=478)


Total Extinction Risk
2100
Probability
50.0
25.0
10.0
1.0
0.5
0.0
1
6
6.6
5
Superforecasters (n=87)
Experts (n=67)
General X-risk Experts (n=13)
Public Survey (n=477)

Sources of agreement, disagreement and uncertainty

General uncertainty

- There was, overall, a very high variance among these forecasts, but more *within* than *among* teams. Team 336, for example, had forecasts for 2100 higher than 70% and lower than 0.001%. Another team, 334, noted: “The error bars on our forecasts are large (~+/-10x), reflecting significant disagreement and uncertainty.”

AI risk

- For both lower and higher forecast teams, it was generally agreed that AI was the dominant factor.

Base rates

- It was taken as self-evident that humans had not previously gone extinct; furthermore, evidence was given showing the human population has never fallen as low as 5,000 “since the dawn of the species’ existence” (341). Strictly speaking, then, the base rate is zero.
- One team (338) considered the duration of other hominid species as a reference class. However, “There was some difference of opinion on the validity of human and other hominid species durations as base rate versus we face a whole different risk of extinction in the near term” (338).

Summing component risks to calculate total risk

- Several teams mentioned that most or all of their team members arrived at their total risk by summing the risks from the prior questions.[439] [440] [441] [442]
- However, two (342, 345) of these three teams expressed pushback against this methodology, with both mentioning the possibility that multiple existential threats could exacerbate each other if happening simultaneously:
    - From 342: “there has not been an explicit consideration of the possibility that extinction risks may reinforce each other, although individual estimates may have taken this into account. For example, global warming may foster global competition for scarcer agricultural land, which in turn could make nuclear war more likely. In such a scenario Total Extinction Risk is likely to be larger than the

---

[439] 338, “Some forecasters aggregated only their perceived most important risks, usually AI and nuclear, on the argument that those risks dominated. Others used all prior risks and/or padded risks to account for other unknown or unquantified risks.”
[440] 345, “team members generally calculated the total overall risk of extinction by adding together the risk percentages they had forecasted for specific extinction-causing events such as nuclear weapons, pathogens, AI, and non-anthropogenic causes like asteroid strikes.”
[441] 343, “Largely, forecasters summed up existential risks from previous questions to answer this question.”
[442] 342, “Considering that the general approach was to add the Extinction Risks of all forecasted causes, the arguments in favor of the teams median value must be traced back to the individual forecasts.”

add individual risks. Independent superforecasters or domain-specific experts may disagree with this oversight." (342)
    - From 345: "Some teammates predicted higher total extinction numbers than what they would get from simply adding together the specific risks. This was to take into account [...] the effects of combinations of these risks (i.e., a nuclear war during a pandemic)."
- Team 345 further added that, on the other hand, the component risks should not be combined, because none of the component risks except for AI were likely to cause extinction on their own:
    - "some teammates believed that adding together all possible extinction-causing events was the wrong approach since some risks such as nuclear weapons or pathogens seemed too unlikely to be able to cause extinction at all on their own. Those who took this approach see AI as the black box that could be extinction-causing and therefore the only risk that really mattered." (345)

Arguments given for forecasts ≤ 0.027% (2030), 0.5% (2050), 2.004% (2100)

- AI capabilities pose no existential risk.[443] Alternatively, AI risk is over-hyped,[444] and it's possible some people overestimate AI risk due to ingroup, or groupthink, effects: "Our team, even those who give significant risk estimates to AI extinction, believe AI experts vastly over estimate the abilities of AI. [...] [T]he subset of people who are super forecasters are often in the same subset of the type of people worried about existential risk (EA's, rationalists, etc.) which could have an effect of biasing toward over rating the risk compared to a general subset of people" (345).
- It would be very difficult to "virtually eras[e] all humanity" (337).
- Humans are resilient.[445]
- Humans can live in a "vast array of climates and conditions" (337).
- "Advanced technology" (337) will prevent this.
- "Resource availability" (337) will prevent this.
- "General optimism in mankind" (337) is a reason to believe this will not happen.
- "Humanity's problem-solving abilities" (337) will improve at a "steady rate."
- A cascade of disasters happening "so quickly that mankind is not able to protect a meaningful percentage of its population in time" (337) seems unlikely.

---

[443] 341, "Artificial intelligence and other technological advances such as engineered bioweapons, do not pose an existential risk; in particular, artificial systems surpassing human intelligence are impossible, will not be achieved by the year 2100, and/or are fairly easily restricted to limited tasks with no capability to cause human extinction."
[444] 345, "the magnitude of AI capabilities and our current state of development of AI are vastly overhyped. As it currently stands implementations of AI are naive and niche and we are really only in our infancy of the technology. We are skeptical of forecasts that AI will takeoff/a singularity like event occurs. Beyond that there is also considerable uncertainty around if/when AGI does get developed whether or not that is inherently dangerous to humanity. The skeptics give only small probabilities that AI is outright hostile towards or humanity or pursues goals that are so orthogonal to humanity that they cause extinction."
[445] 337, "Reasons cited for high improbability of extinction are the resilience of humanity."

- J. Richard Gott repurposes Copernicus' idea that humans aren't special *spatially* to create the idea that we're not special *temporally*. "By applying Gott's 95% confidence formula to arrive at an estimate of when the human race will go extinct. Sleep well, as this date is somewhere between 5,100 and 7.8 million years in the future." (339)
- Toby Ord's estimate of "~15%" (338) "[was] not consistent with base rates of human species survival."
- Humanity's geographic dispersion makes "total extinction" (338) difficult.
- Extinctions tend to take a "very long time," usually longer than from now to 2100.[446]
- Nuclear threats are very low-risk.[447]
- Non-anthropogenic threats are very low-risk.[448] [449]

Arguments given for forecasts ≥ 0.1865% (2030), 1% (2050), 2.58% (2100)

- Risk from AI is large.[450] [451] One team summarized a few main concerns thusly:
    - "The AI Risk scenario is too complicated to give adequate treatment in this section. We'll briefly note here that some forecasters believe that AI risk is not only possible, but perhaps almost inevitable. That the brain is a proof of concept for intelligence in a physical medium. That the intelligence staircase may extend far above us. That robots are dancing. That my regular laptop can perform 100,000,000 multiplications in 0.02 seconds. AI is already writing code and it can cheat and adapt and deceive and evolve. And that machines are not only beating us at our own games, they are achieving superhuman play in a matter of hours." (344)
- More generally, anthropogenic factors are the main risks, and AI at the top of that list.[452]
- Genetically-engineered pathogens might become more dangerous, and could be released in multiple scenarios: "war, terrorist groups, accidental release, etc."[453]

---

[446] 338, "even moderately rapid extinction may not meet 2100 deadline." And: 338, "Extinctions can take a very long time. In terms of mass extinctions, only the end of Cretaceous extinction may have been quick enough to resolve by 2100."
[447] 341, "Nuclear weapon use (including any ensuing nuclear winter) does not pose an extinction risk and will not by the year 2100; or, if it does, such weapons will not be used."
[448] 345, "there is virtually no chance of nuclear or non anthropogenic risk to extinction."
[449] 341, "Mass extinctions from natural cosmological events only occur very rarely, and technology may actually be able to reduce these risks below their historical base rate."
[450] 342, "the consensus being that the AI Extinction Risk is the most serious of the bunch" (i.e., of "the various forms of [already] predicted extinction risks").
[451] 344, "The most pessimistic forecasts are primarily driven by AI risk."
[452] 340, "Forecasts at the high end are dominated by anthropogenic risks, with AI being a particular standout in both magnitude of effects and frequency of mentions."
[453] "Additionally, it may become increasingly easy to genetically-engineer pathogens that are extremely lethal and contagious. This could either wipe out humanity directly or lead to civilizational collapse that ends in humanity's extinction. This risk could come due to war, terrorist groups, accidental release, etc." (344)

- New technologies that become known to be existential risks will probably appear, and/or currently-existing technology will become understood to be existential risks.[454]
- More generally, unknown things could present themselves as risks.[455]

Other arguments given

Arguments in favor of lower forecasts:
- Most kinds of disaster wouldn't result in <5,000 people left alive.[456]

Arguments in favor of higher forecasts:
- Climate change is actually a significant existential risk.[457] [458]

Cross-references with other questions

Q4: AI Existential Risk
Q6: Nuclear Existential Risk
Q8: Non-Anthropogenic Existential Risk
Q9: Total Catastrophic Risk
Q11: Year of Existential Catastrophe

Question 11: Year of Extinction

By what year will humans go extinct or first have a population less than 5,000?

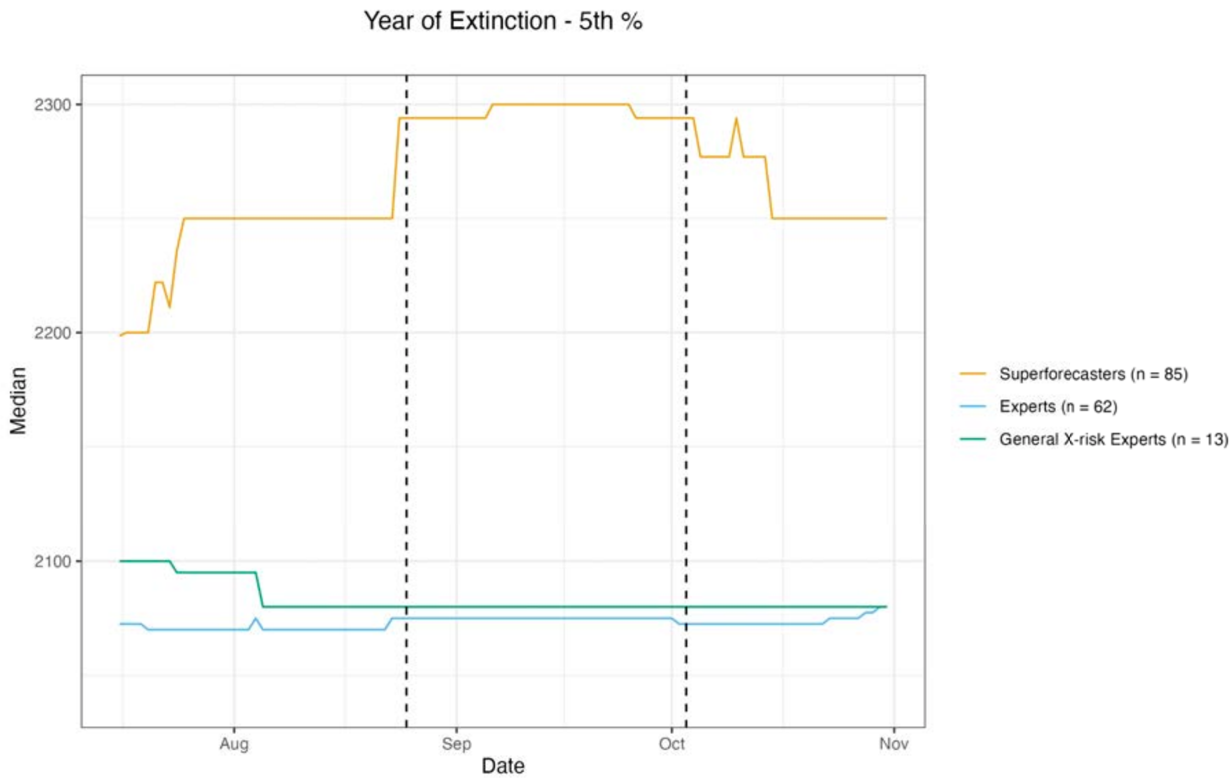
Question and resolution details, prior forecasts, and other relevant sources

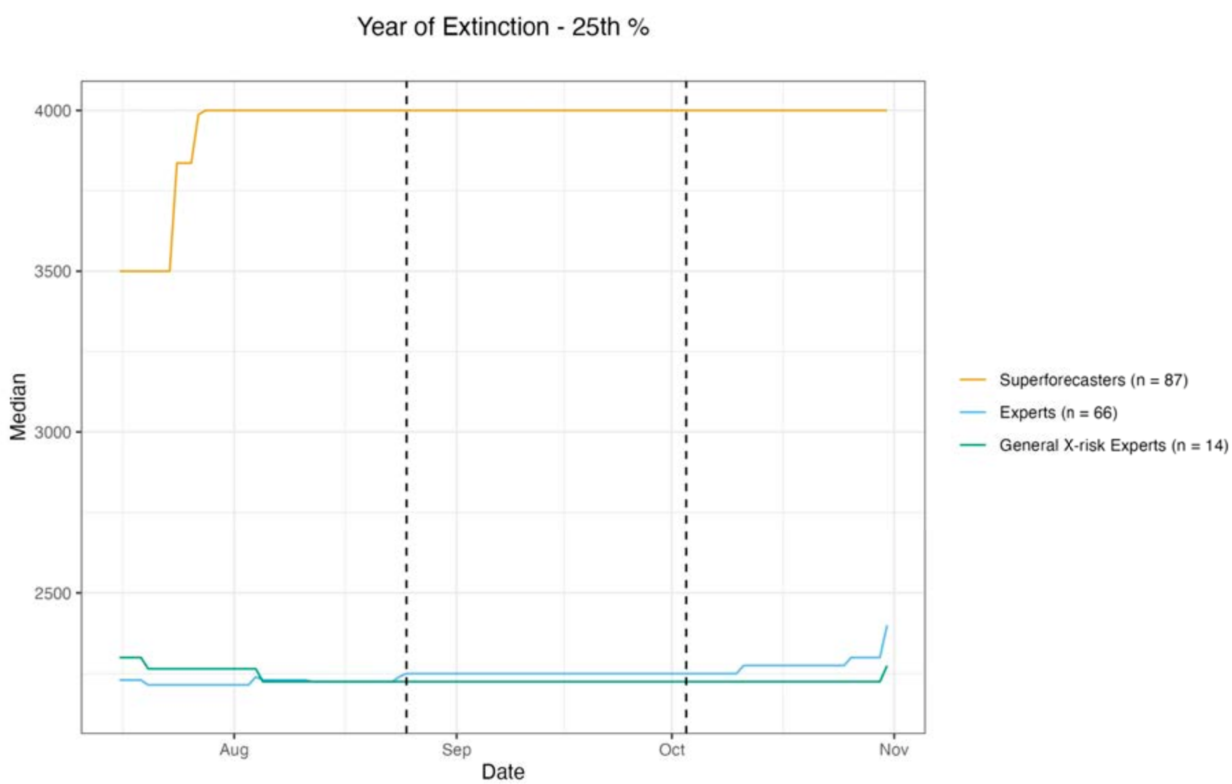
[454] 343, "Given this accelerating pace of discovery, it is likely that new sources of potential x-risk will arise in the next 20-80 years. These sources are likely unknown to us at the moment, or are not imagined as anything but benign technology whose potential for x-risk will not be realize[d] until they mature."
[455] 345, "Some teammates predicted higher total extinction numbers than what they would get from simply adding together the specific risks. This was to take into account unknown risks that we didn't think of or cannot even conceive of yet."
[456] 336, "Many of the lower end forecasters focused on the extremely low base rate combined with the very high barrier for resolution (less than 5,000 individuals left alive). In other words - there are a lot of disasters (pandemics, supervolcanoes, even full scale nuclear war) that are very likely to leave more than 5,000 alive even if they happened."
[457] 341, "Most experts see climate change as a major risk, but not one that could cause human extinction. There are however recent papers such as this one, that argue that it could happen. After all, there is great uncertainty in climate change models, and we can not say there is zero probability of e.g. feedback loops leading climate change to actually become an x-risk."
[458] 339, "Some report findings that warn about the real and dangerous threats that climate change on our future existence as found in a policy paper from an Australian think tank in June 2019."

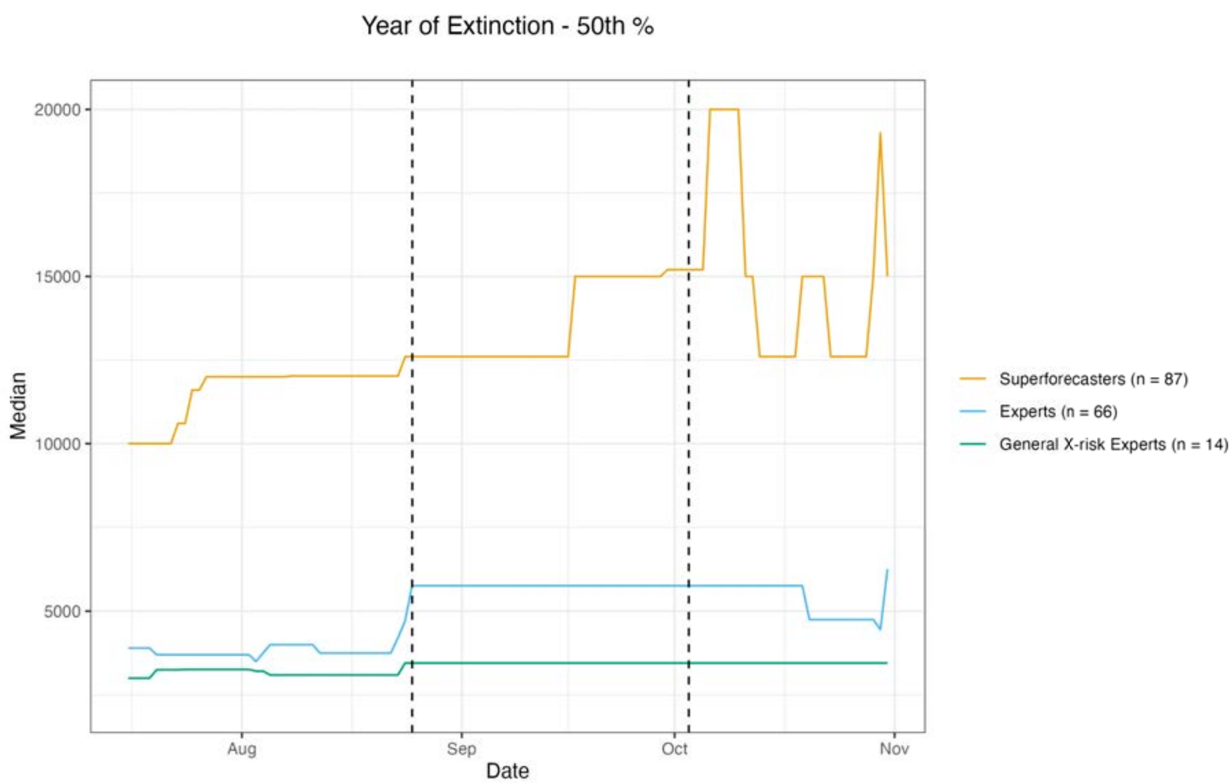
Results[459]

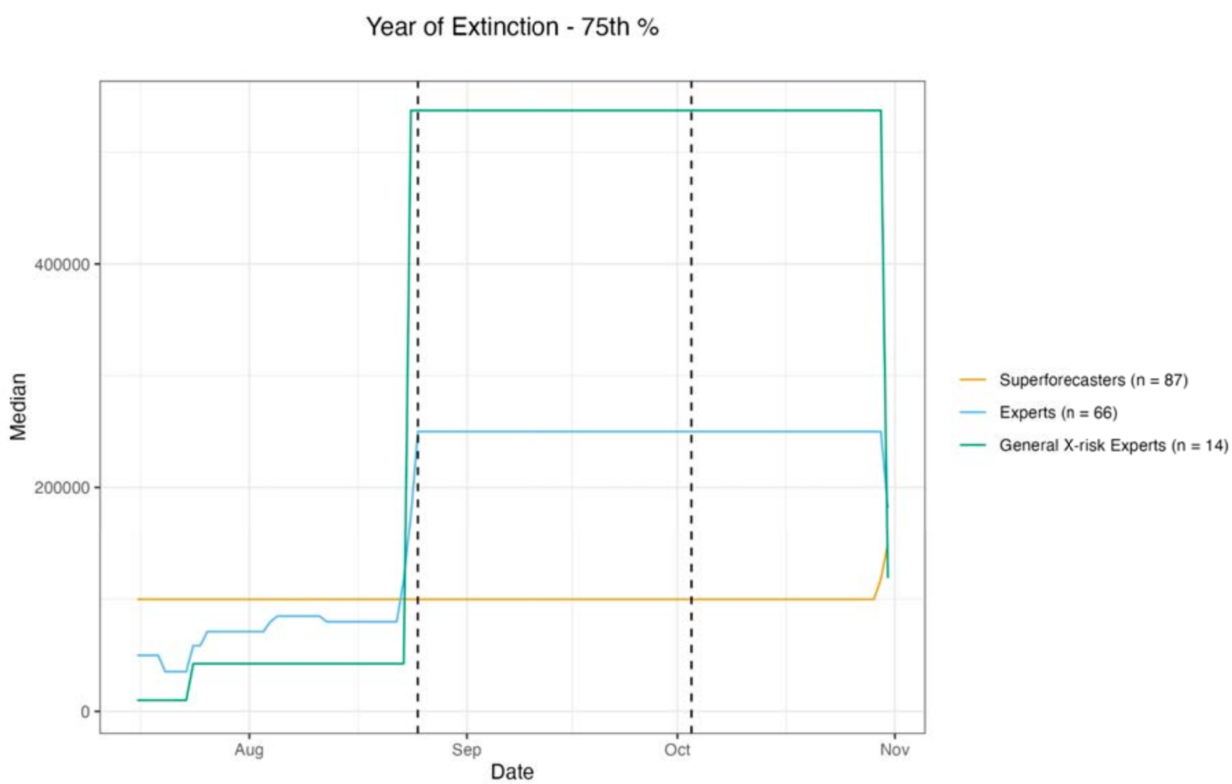
Group	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 88)	10000	15000	109104444667.36	-1.7%
Experts (N = 66)	3900	6261	137787070.33	-10.67%
General X-Risk Experts (N = 14)	3016.5	3450	318804.97	-0.05%
Public Survey (N = 418)	2500		Inf	-

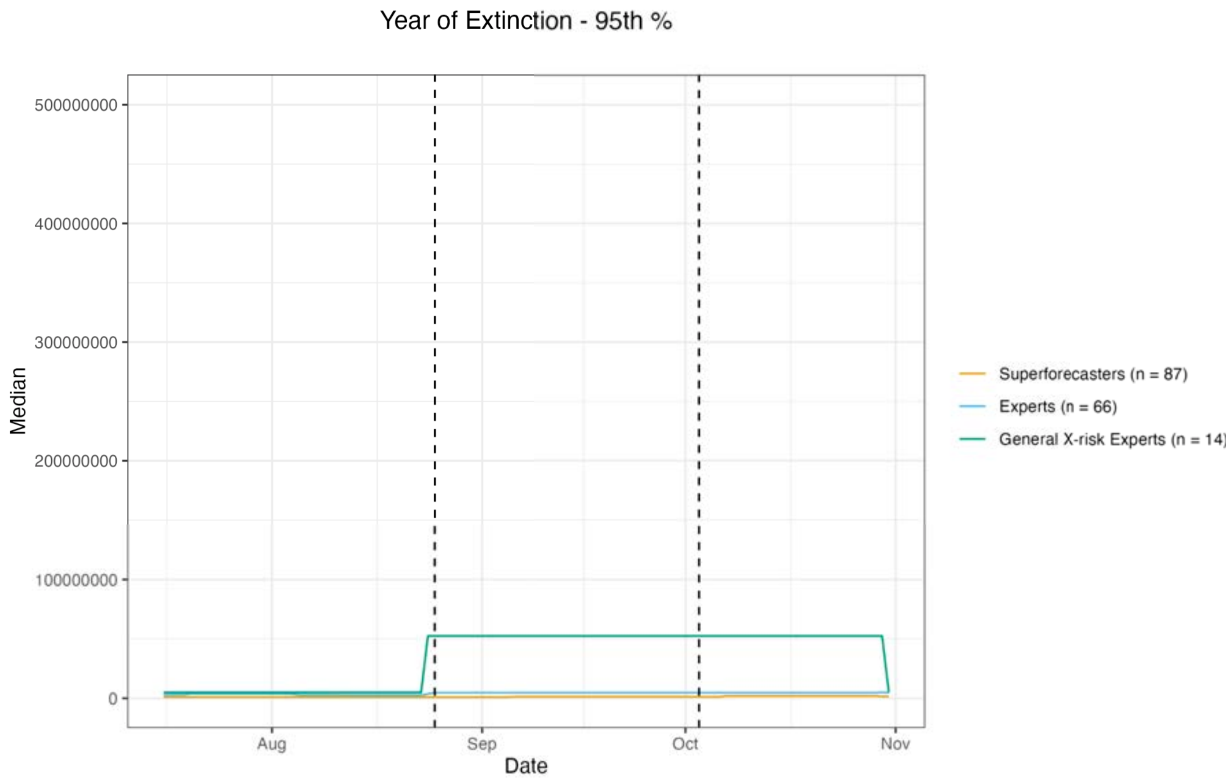
[459] Numbers of forecasters are given as of Stage 4 of the XPT. The forecasts given are the 50% percentile forecast for the given group. Public survey respondents were surveyed outside of the tournament context.

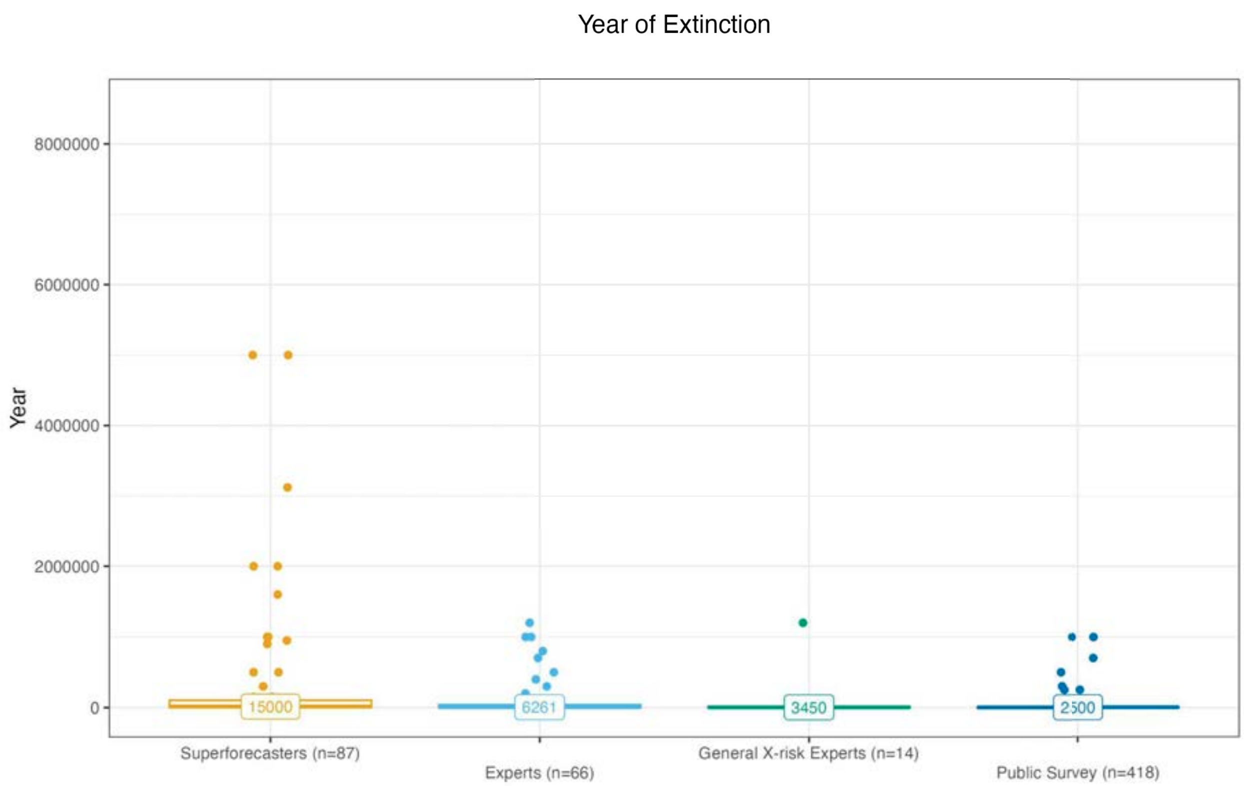

Year of Extinction - 5th %
2300
2200
2100
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 85)
Experts (n = 62)
General X-risk Experts (n = 13)


Year of Extinction - 25th %
4000
3500
3000
2500
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 87)
Experts (n = 66)
General X-risk Experts (n = 14)


Year of Extinction - 50th %
Median
20000
15000
10000
5000
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 87)
Experts (n = 66)
General X-risk Experts (n = 14)


Year of Extinction - 75th %
Median
400000
200000
0
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 87)
Experts (n = 66)
General X-risk Experts (n = 14)


Year of Extinction - 95th %
500000000
400000000
300000000
200000000
100000000
0
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 87)
Experts (n = 66)
General X-risk Experts (n = 14)


Year of Extinction
8000000
6000000
4000000
2000000
0
Year
15000
6261
3450
2500
Superforecasters (n=87)
Experts (n=66)
General X-risk Experts (n=14)
Public Survey (n=418)

Sources of agreement, disagreement and uncertainty

- Approach to developing forecast
    - There were several approaches used to develop forecasts. Different approaches were sometimes used for different levels of confidence (i.e. 5th percentile vs 95th percentile etc.).[460] Approaches included:
        - Starting with a base rate for extinction and adjusting for factors that would suggest extinction risk deviates from this base rate.[461] Several base rates were suggested, including the average lifespan of a mammalian species and the average lifespan of a hominid species. Adjustment was usually on the basis of estimated impact of technology.[462]
        - Combining estimates of extinction risk in different periods to develop a forecast (e.g. for the next century and then beyond).
        - Combining probability estimates of humanity reaching different milestones (e.g. interstellar travel).[463]
        - Using the Doomsday Argument (referencing Carter, Gott, and Bostrom), a statistical argument for the expected total population of humans using the number of humans born to date. Some forecasters used this argument alone for their forecast.[464]
- Impact of AI and other technologies
    - Forecasters disagreed about the degree of existential risk posed by AI and other anthropogenic risks, including nuclear weapons and climate change.[465]

---

[460] 340, “The 75-95% confidence forecasts reflect the base rate derived from the lifespan of other extinct species of the genus Homo. [...] The 5-25% and up to 50% confidence forecasts reflect the risks of irreversible planetary warming or some near-term extinction event, likely instigated by AI or pandemics. [...] The 50% confidence forecast also reflects "how long it would take to have as many humans born in the future as have been born in the past, so that we're about the median humans in terms of birth order."”
[461] 345, “The background for this question states that mammalian extinction rate range from .1-1.8 per million species years, therefore a reasonable base rate to start from is on the order of 1 million species years, or the year ~700,000”
[462] 339, “The strongest arguments put forward for our median estimates were [...] (2) an appeal to base rates about the expected lifespans of e.g. mammalian species,”
[463] 339, “The strongest arguments put forward for our median estimates were (1) an estimate based on assuming a constant annual extinction risk after the time of perils, [...] (3) an approach focussed on considering key junctures in human expansion”
[464] 341, “There might be an upper bound of 7.8M years, according to J. Richard Gott's formulation of the controversial Doomsday argument, which argues that we have probably already lived through half the duration of human history.” 337, “The lower end of the forecast also aligns well with other pessimistic forecasts such as Carter's catastrophe and Bostrom Self-sampling assumption.”
[465] 341, “There is significant disagreement regarding near-term AI risk and how large the risk is.” 336, “Major differences exist about whether AGI is highly likely to destroy us within this century.” 345, “There is wide disagreement among the team about how much weight to give various extinction risks. Many people are extremely pessimistic about humans ability to combat climate change sufficiently, many people are more skeptical about AI and its threat and many people disagree about the timeline for when or even if humans will evolve or edit their DNA to the point where humans as a species no longer exist.” 343, “Not to rehash all the details of our Total Extinction Risk wiki there were two main perspectives: 1. The primary source of extinction risk is nuclear/nonanthropogenic. Those that believed this put extinction at ~1-5% this century. 2. AI risk is real. Those that believed this put extinction at 15-50% this century”

    - There was also disagreement about whether humanity was likely to become capable of interstellar travel and whether technological change would see post-human beings emerge.[466]
- Plausibility of 'time-of-perils' hypothesis
    - There was disagreement about whether we are currently facing a period of particularly elevated existential risk that should reduce if we survive this period, an idea that is sometimes referred to as the 'time of perils' hypothesis.[467]
- Reputability of sources
    - Forecasters disagreed on the reputability of published estimates on existential risk. Specifically one team noted being divided into two groups, by whether or not they placed significant weight on the work of Toby Ord.[468]

Arguments given for lower forecasts

- High degree of anthropogenic existential risk, especially from AI
    - Forecasters who suggested earlier dates most commonly appealed to arguments for existential catastrophe from AI and other anthropogenic risks, particularly advanced weapons, including nuclear weapons and bioweapons[469]
- Persistently high existential risk
    - Several forecasters suggested that a prolonged level of high existential risk was more likely than a finite time of perils.[470]

---

[466] 336, "Whether we become an interplanetary species or even escape the solar system drives the highest numbers, though is not a view held by the majority."

[467] 341, "Most forecasters believe there is a large degree of risk early, then it rapidly declines based on human progress away from the technological risks. "342, "The major disagreement seems to stem from those who believe that current extinction risk are essentially here to stay, and thus constant year over year and guaranteed to wipe out the species within a fairly short time window (measured in thousands of years), against those who believe that humanity can move beyond those risk and that we at least have a chance to have peaceful coexistence on earth and solve most potential problems (excluding possibly the death of the sun), or possibly even move beyond that to space exploration and settlement that would greatly increase the potential lifespan of our species." 340, "Various disagreements concern: Will the unsustainable level of extinction risks in this century continue after 2100?"

[468] 341, "Certain members of the team argue that Ord is not a reputable source, since some of his cited publications have no peer review, and his discipline is philosophy. They argue that although it's granted that he is a Senior Research Fellow in Philosophy at Oxford University, he is biased by his political activism, and therefore many experts are disregarding Ord's conclusions. Eight out of the 23 forecasters on our team have cited him in this or other questions."

[469] 345, "Generally see the low end being almost exclusively attributed to rapid takeoff of AI." 344, "Some arguments in favor of relatively low values include: We may be on the brink of the precipice, with a large portion of future extinction risk concentrated in this century (mostly from AI risk). " 338, "in the next few centuries, particularly starting late this century, we will develop an expanding arsenal of weapons with which we could cause our own extinction on a planetary scale, and that we will (1) lack the tools to establish a space colony that could lead to a long-term sustainable human presence capable of recovering rather than dying out (2) lack the global governance tools to effectively manage the proliferation and use of these weapons (both by well-resourced state actors and smaller non-state actors). Such weapons might include advanced biotech, advanced nanotech, lethal autonomous weapons, orbital weaponry, nuclear weapons, etc."

[470] 342, "These forecasters believe that the extinction risk from AI/Nuclear/Non-Anthropogenic are both a) higher in a given set of years than the rest of the team and b) more or less constant over time.

- Possibility of a post-human future
  - Some forecasters argued that advances in technology may be necessary to overcome our current risks, but this advancement would also make it more likely that future beings are sufficiently different such that humanity would be considered extinct on the definition of this question.[471]

Arguments given for higher forecasts

- Base rates of species extinction
  - More optimistic forecasts were often heavily based on base rates of species extinction. One team argued that we should expect technology to extend this expected lifespan rather than shorten it, on the basis that thus far technology has been beneficial for humanity.[472]
- Difficulty of causing extinction
  - Several forecasters argued that it would be very difficult for an event, anthropogenic or non-anthropogenic, to cause humanity to go extinct, given the possibility of a small number of survivors and the adaptability of humanity.[473]
- The possibility of existential security
  - Some forecasters argued that humanity spreading to the stars, or otherwise achieving a state of existential security, was possible and if this occurred we should expect a very long existence.[474]

Cross-references with other questions

Q10: Total Existential Risk
Q12: Humans Born From 2023 Onward

Essentially, they believe that humanity has gotten itself into a precarious position in a variety of ways and that there is no way back from it; we are essentially a ticking time bomb with no realistic way to undo the damage."

[471] 345, "On the higher end of our forecasters, the consensus is that even if humans do not endure an actual extinction event, technology will develop so far that humans will have evolved or biohacker themselves into a species that is distinct from humans which meets the criteria for this discussion."

[472] 337, "The upper end of the forecast is in line with what to expect for the lifetime of mammalian species on earth (of the order of 1-2 megayears). In most situations the base rate is the best possible forecast that can be made, more so when one is confronted with so many intangibles as in this question: does technology consistently reduce the species lifetime? Everyday experience would suggest the contrary, which might indicate that it might be better not to adjust negatively because of this specific factor."

[473] 341, "The risks from nuclear war and pathogens, natural or man-made are mostly not considered existential. The consensus is that a few thousand or more humans can survive even a total nuclear war."

[474] 342, "The argument for this is that if we get past a certain point, namely the ability to begin settling space, the risk of an extinction event on one world becomes incapable of causing complete extinction. Thus we gain the ability to break the mold of prior extinctions by multiplying beyond them and moving out of the reach of one event destroying the species, even if it is the death of the sun." 341, "If humans master space travel, AI and nanotech, they may evolve into some very resilient creatures spread throughout the universe." 338, "Humans adapt and survive. Extinction is highly unlikely. War (even nuclear), AI, asteroid, pathogens, or global warming are unlikely to achieve that."

Question 12: Humans Born From 2023 Onward

How many humans will be born from 2023 onward?

Question and resolution details, prior forecasts, and other relevant sources

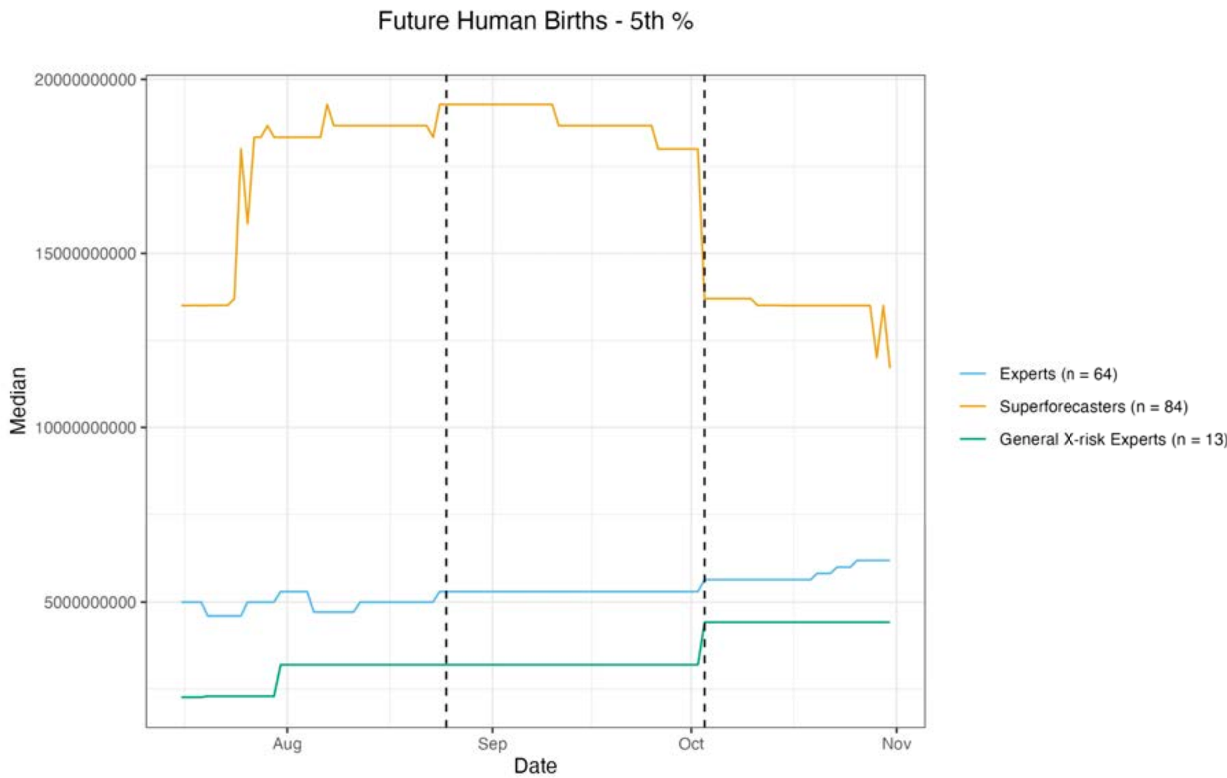
Results[475]

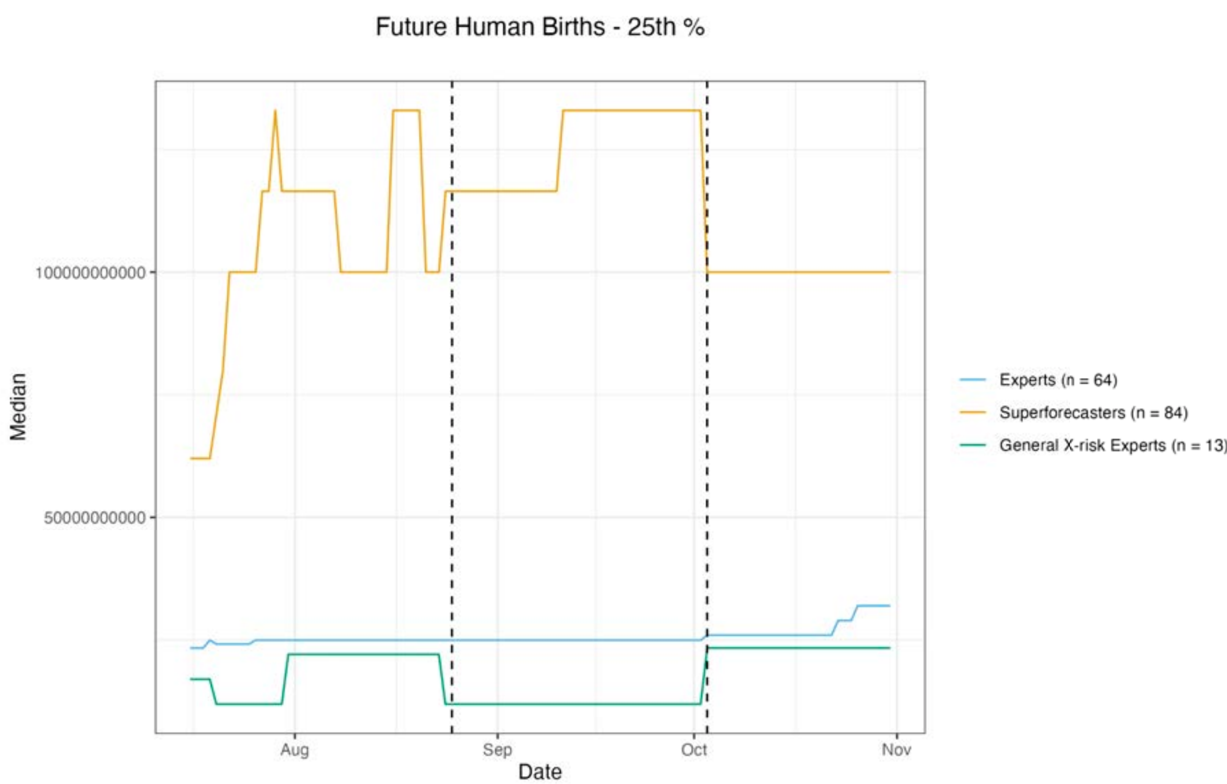
Group	Percentile Forecast	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 84)	5%	10,000,000,000	11,700,000,000	11551892701215	-5.54%
	25%	50,000,000,000	100,000,000,000	25799259563603	-5.39%
	50%	215,000,000,000	500,000,000,000	114703648884811424	-4.88%
	75%	1,317,972,363,582	4,676,550,000,000	114707866935044419280133160960	-100%
	95%	10,000,000,000,000	100,000,000,000,000	113960427081216036662474617988841472	-4.26%

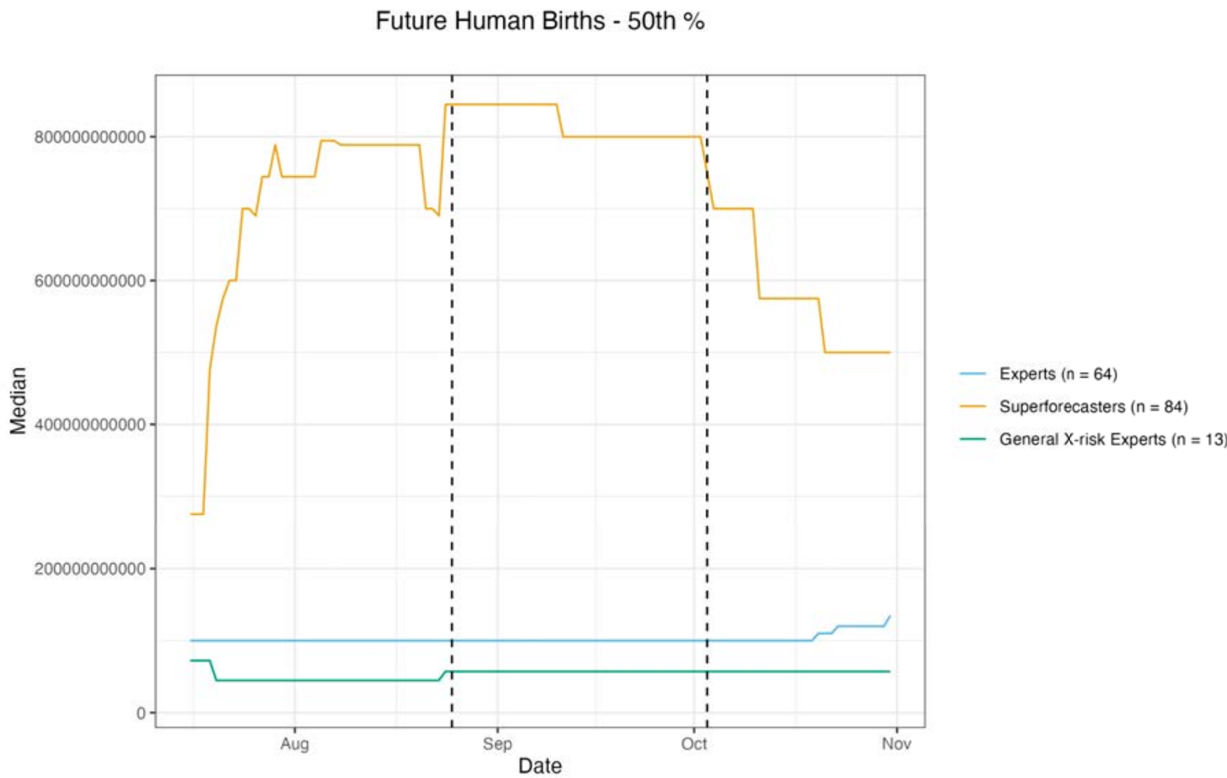
475 Numbers of forecasters are given as of Stage 4 of the XPT. Public survey respondents were surveyed outside of the tournament context.

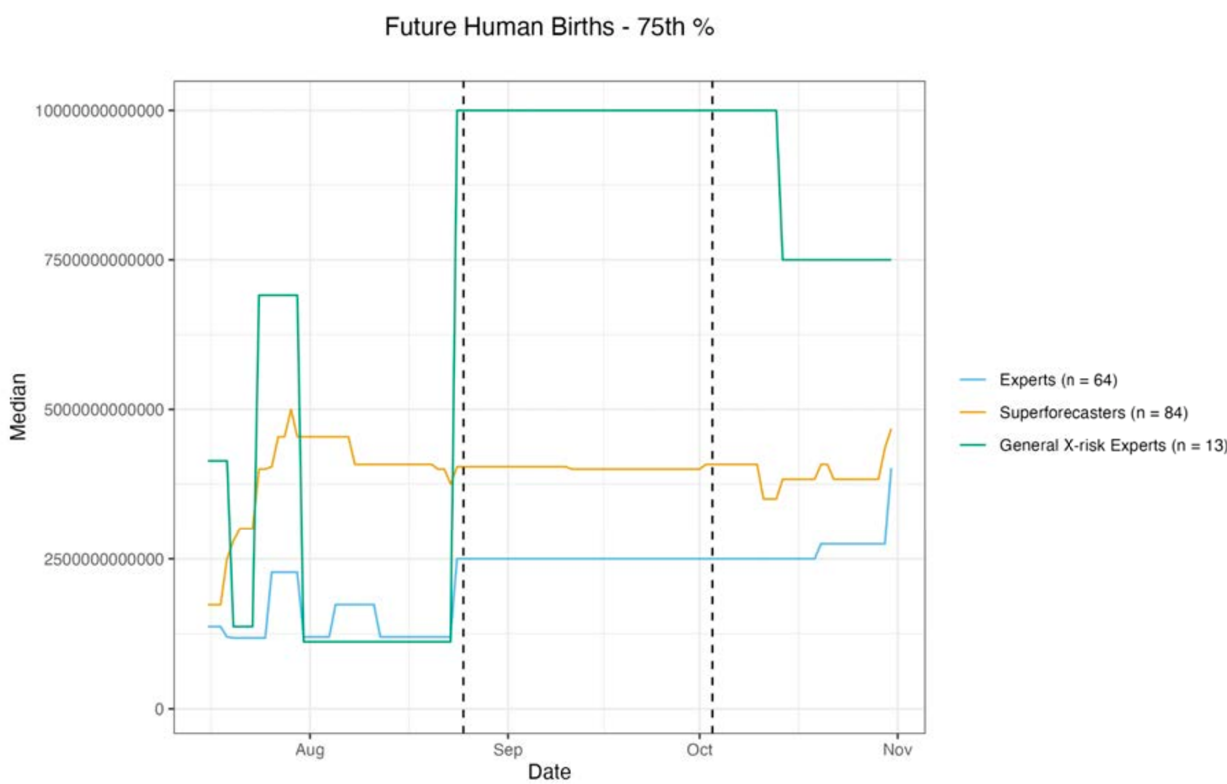
Experts (N = 64)	5%	4,185,000,000	6,190,000,000	74046866696	+11713.6%
	25%	22,825,000,000	31,990,372,990	1290994381113245696	-3.18%
	50%	100,000,000,000	135,000,000,000	5112135497556067090432	-3.18%
	75%	1,283,245,000,000	4,020,000,000,000	39498460023445193228288	+316468034262103872%
	95%	170,000,000,000,000	170,000,000,000,000	Inf	-
General X-Risk Experts (N = 13)	5%	2,300,000,000	4,420,000,000	5271717678	+41.83%
	25%	12,000,000,000	23,400,000,000	54329222257	+1.78%
	50%	44,800,000,000	57,240,000,000	2752670686956	-0.2%
	75%	1,369,250,000,000	7,500,000,000,000	1774740448444144896	-0.000018%
	95%	10,000,000,000,000,000	316,227,766,016,838	33282011773513741633853534164045791232	-100%
Public Survey (N = 411)	50%	9,000,000,000		680703800659236000000000	-

			00000000	

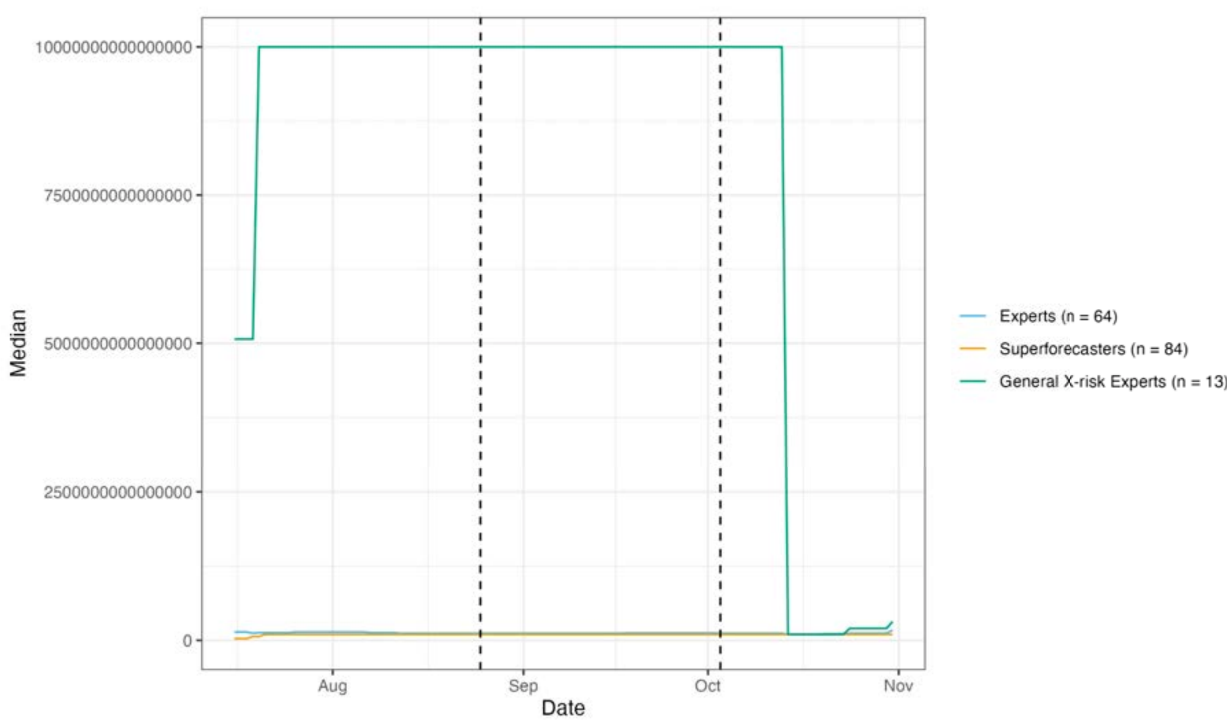

Future Human Births - 5th %
20000000000
15000000000
10000000000
5000000000
Median
Aug
Sep
Oct
Nov
Date
Experts (n = 64)
Superforecasters (n = 84)
General X-risk Experts (n = 13)


Future Human Births - 25th %
100000000000
50000000000
Median
Aug
Sep
Oct
Nov
Date
Experts (n = 64)
Superforecasters (n = 84)
General X-risk Experts (n = 13)

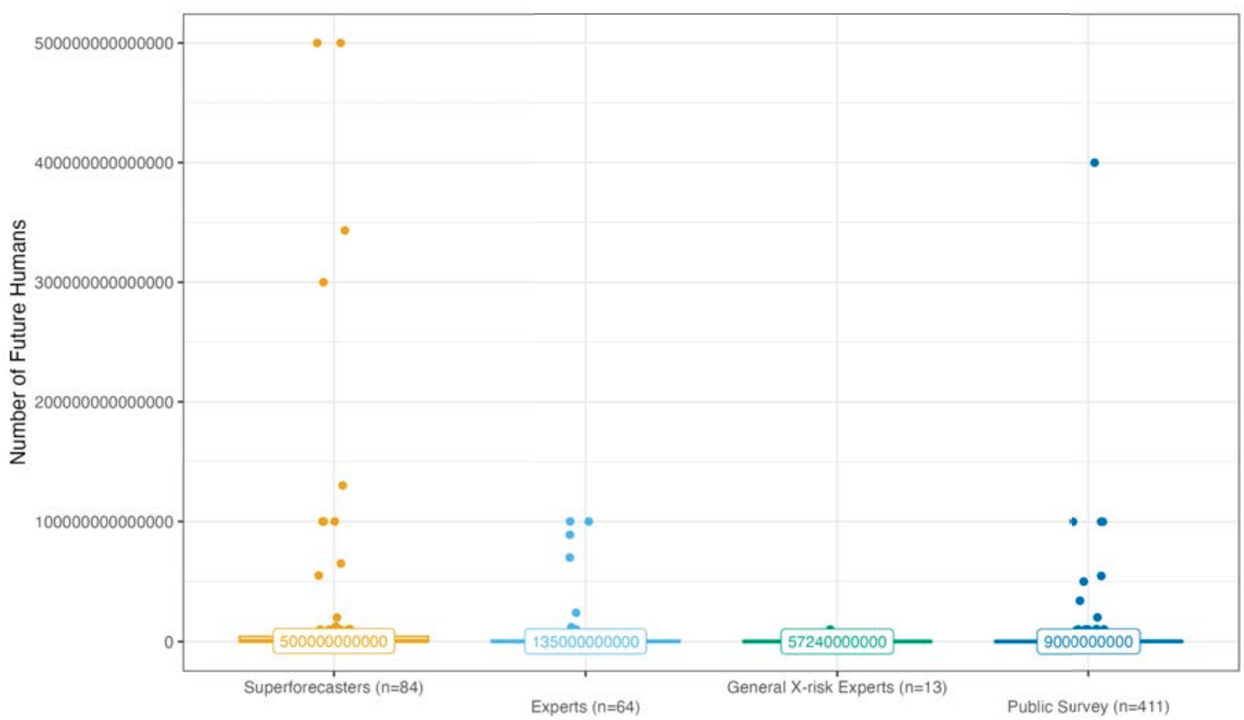

Future Human Births - 50th %
800000000000
600000000000
400000000000
200000000000
0
Median
Aug
Sep
Oct
Nov
Date
Experts (n = 64)
Superforecasters (n = 84)
General X-risk Experts (n = 13)


Future Human Births - 75th %
10000000000000
7500000000000
5000000000000
2500000000000
0
Median
Aug
Sep
Oct
Nov
Date
Experts (n = 64)
Superforecasters (n = 84)
General X-risk Experts (n = 13)

Future Human Births - 95th %



Future Human Births



Sources of agreement, disagreement and uncertainty

Sources of disagreement and uncertainty:

- Approach to developing the forecast
    - Most forecasters developed their forecast by combining an estimate of the duration of humanity as a species and the future birth rate.[476]
    - Some forecasters instead made use of the Doomsday Argument, which is a statistical argument for the expected total population of humans using the number of humans born to date.[477]
- The following factors were the sources of important disagreement between forecasters:
    - Degree of existential risk[478]
    - Expected birth rate into the future[479]
    - The likelihood of space colonization[480]
    - The likelihood of post-human beings replacing humanity[481]

Arguments given for forecasts higher than the median ($1.5e^{11}$)

- High level of existential risk
    - Several teams commented that lower forecasts were largely influenced by perceptions of high levels of extinction risk.[482]
- Declining birth rates

---

[476] 340, “Almost all forecasts attempted to extrapolate out current/recent birth rates, or to follow population/birthrate forecasts from an organization like the UN or IHME. [...] These trends tended to be extrapolated out to either whenever the forecaster thought it was likely humans would go extinct or transform so thoroughly as to no longer meet the criteria of this question.” 337, “basically, this is a classical Fermi problem that can be divided in two parts, A.1) what will be the average rate of births from now till the end of the species, and A.2) how long will the species last.”

[477] 340, “There is also a different base rate formulation for this problem: A.3) The doomsday argument [1], sometimes also referred to as the Carter catastrophe and which also appears on the wiki of question 11.” 336, “Many of the upper-bin numbers are difficult to reconcile with the birth rank sampling Doomsday Argument estimation method”

[478] 344, “Some of the factors that introduce large uncertainty are: Chances of complete extinction this century. Year of extinction in the long term.” 340, “Significant disagreements, dissenting views within the team, and major sources of uncertainty: [...] How likely humanity is to go extinct or experience a significant population collapse and by when. Will the unsustainable level of extinction risks in this century continue after 2100?”

[479] 337, “This fertility rate is a major source of uncertainty.” 336, “There are disagreements regarding how many humans the Earth can sustain / at what number the population eventually stabilizes”

[480] 336, “There are disagreements whether any substantial interstellar expansion will be done by biological humans.” 344, “Some of the factors that introduce large uncertainty are: [...] Assuming mankind survives this century, how many stars will mankind (and its descendants) reach at the limit? ”

[481] 341, “Areas of disagreement and uncertainty: Whether some parts of humanity will desire, and be able to, continue to reproduce biologically, keeping the chain of live births in this question.” 340, “Will our descendants hundreds of thousands of years down the line count as humans? How long until humans engineer themselves to longer meet this question criteria, or if they do at all.”

[482] 342, “On the lower end, team members' estimates frequently exhibit an increased perception of the likelihood of near-term extinction.” 337, “Also, more pessimistic forecasters considered x-risks more highly, which could lead to early extinction and far lower numbers.”

    - Several forecasters cited declining birth rates as a reason for a lower forecast. Some expected trends of declining fertility to continue,[483] while others suggested that the earth cannot support the current birth rate,[484] or that increasing life spans could reduce birth rates.[485]
- Low likelihood of space colonization
    - Several teams noted that failure to colonize space would limit the number of future births, and that there are several reasons why space colonization might fail.[486]
- Possibility of transhumanism
    - The possibility of humans evolving into post-humans was cited as a reason for lower forecasts.[487]

Arguments given for forecasts lower than the median ($1.5e^{11}$)

- Possibility of space colonization
    - The possibility of space colonization was a key argument for many higher forecasts, with many teams noting that spreading to the stars might overcome most existential risks and lead to a population explosion.[488]
    - Several teams noted it was questionable whether this would be compatible with humans remaining in their current form, rather than becoming post-human beings.[489]
- Possibility of overcoming the time of perils

---

[483] 337, “Generally speaking, mathematical assumptions were driven by more pessimistic assumptions on the birth rate going forward. Some cited continuations of current demographic trends, leading to lower fertility rates, eventually even leading to an overall lower fertility rate than the replacement rate.”
[484] 341, “It can be argued that a steady state birthrate on Earth should be lower than 125 million, since at current levels, there are more humans than the Earth can support.”
[485] 341, “A wildcard could be a significant increase in lifespans, which could reduce fertility rates.”
[486] 342, “Massive numbers of births will only be conceivable through successful space colonization, which may not be guaranteed, even if humanity survives in the long run.” 341, “The lower forecasts assume that Earth becomes uninhabitable in around 1,000 years and humans don’t become a space-colonizing species. Either humans are not able to develop the technology to colonize space, or colonization efforts will fail due to currently unknown space risks, at least colonization efforts involving biological humans.”
[487] 341, “An argument for lower forecasts is that there should come a time when humans have evolved into post-humans, sufficiently unlike today’s humans to not fulfill the criteria outlined in this question.” 344, “In some possible futures, being merely a physical "human" could become increasingly rare.”
[488] 341, “The main argument for higher forecasts is naturally space colonization and an extension of humanity’s existence past Earth becoming uninhabitable. A sufficient number of different colonization places makes humanity much less likely to go extinct.” 340, “There are scenarios of humanity becoming a spacefaring civilization that would continue to reproduce in some way as it colonized other solar systems - at the same time retaining their genetic code to be considered human.”
[489] 344, “If mankind survives long enough, it could end up expanding to many planets and star systems. This could increase the birth rate for purposes of this question (but that greatly depends on cultural and technological factors, such as the percentage of biological humans versus digital humans, post-humans, et cetera).” 336, “The forecaster with the highest 50% and 95% forecast (4e22 and 3e27) This is based on a cosmic expansion model, however raises concerns whether the entirety of those descendants would still qualify as "biological humans" for the purposes of this question - clarification requested.”

  - Some higher forecasts were driven by the possibility of humanity overcoming existing existential risks and entering into a stage of existential security.[490]
- Possibility of current birth rates continuing with low existential risk
  - Some forecasters suggested that humanity continuing its current trajectory, without major technological changes such as space travel, could result in large population numbers if existential risk is low.[491]

Cross-references with other questions

Q10: Total Existential Risk
Q11: Year of Existential Catastrophe

Question 13: Non-Coronavirus mRNA Vaccine

How many people will have received at least one non-coronavirus mRNA vaccine dose…
...by the end of 2024?
...by the end of 2030?

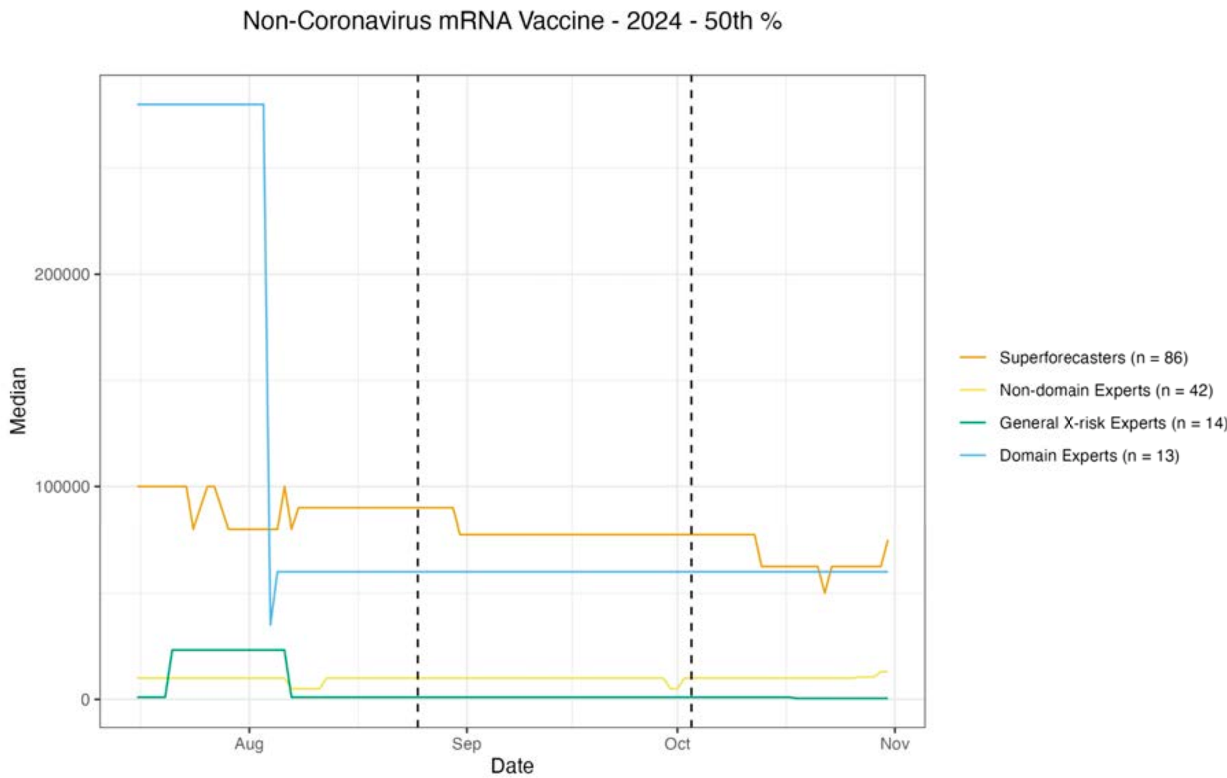
Question and resolution details, prior forecasts, and other relevant sources

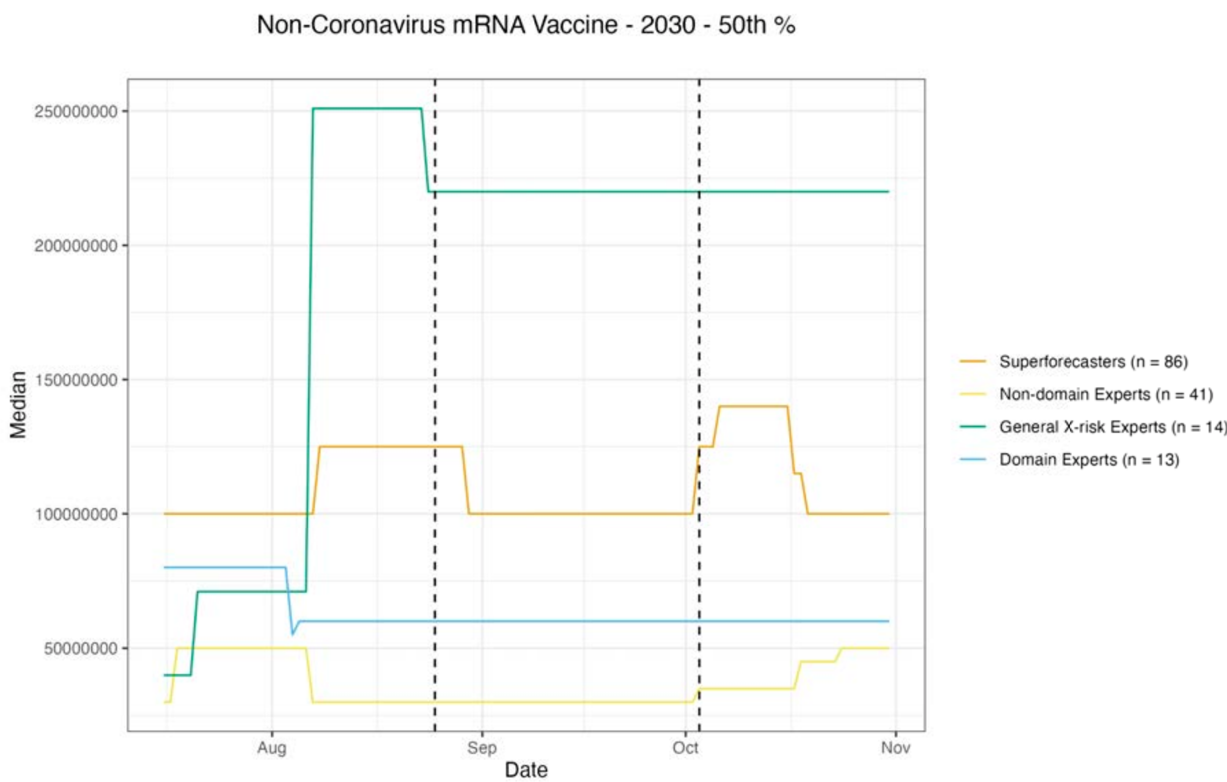
Results[492]

Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 86)	2024	100,000	75,000	724722146	-14.86%
	2030	55,000,000	100,000,000	1424914092	-42.7%

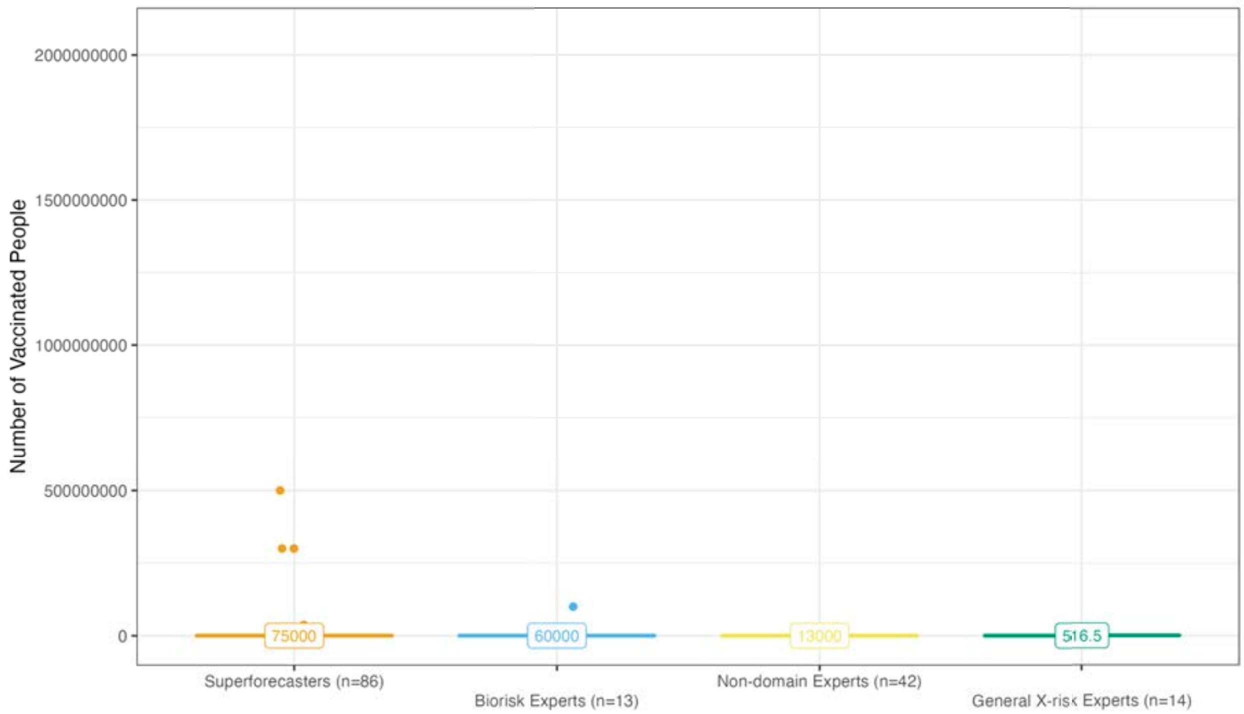
[490] 342, "The members of the team who make the high-end predictions typically believe that if humanity is able to overcome a particular set of technological and civilizational barriers and avoid several filters that could wipe out humanity, this would be equivalent to rendering the probability of extinction virtually inexistent and allowing the number of future births to be virtually infinite."
[491] 337, "Generally speaking, these forecasters went with historical birth rates unadjusted, assumed that x-risks or post-human extinction were much less likely to curtail aggregate births in the foreseeable future, and often had 75th and 95th percentile numbers that effectively assumed humans would be around practically indefinitely." 341, "Also, if Earth remains habitable for one billion years and humanity manages to continue to adapt to the Earth's changing conditions, we can get to more than 1 quadrillion without leaving Earth."
[492] Numbers of forecasters are given as of Stage 4 of the XPT.

Domain Experts (N = 13)	2024	500,000	60,000	1383139081	+0.1%
	2030	100,000,000	60,000,000	2012140862	-18.62%
General X-Risk Experts (N = 14)	2024	35,719	516.5	9259178.36	-10.91%
	2030	71,021,618.5	220,000,000	400962014	-9.37%
Non-Domain Experts (N = 42)	2024	10,000	13,000	6786258.19	-45.07%
	2030	30,000,000	50,000,000	432034003	-56.43%

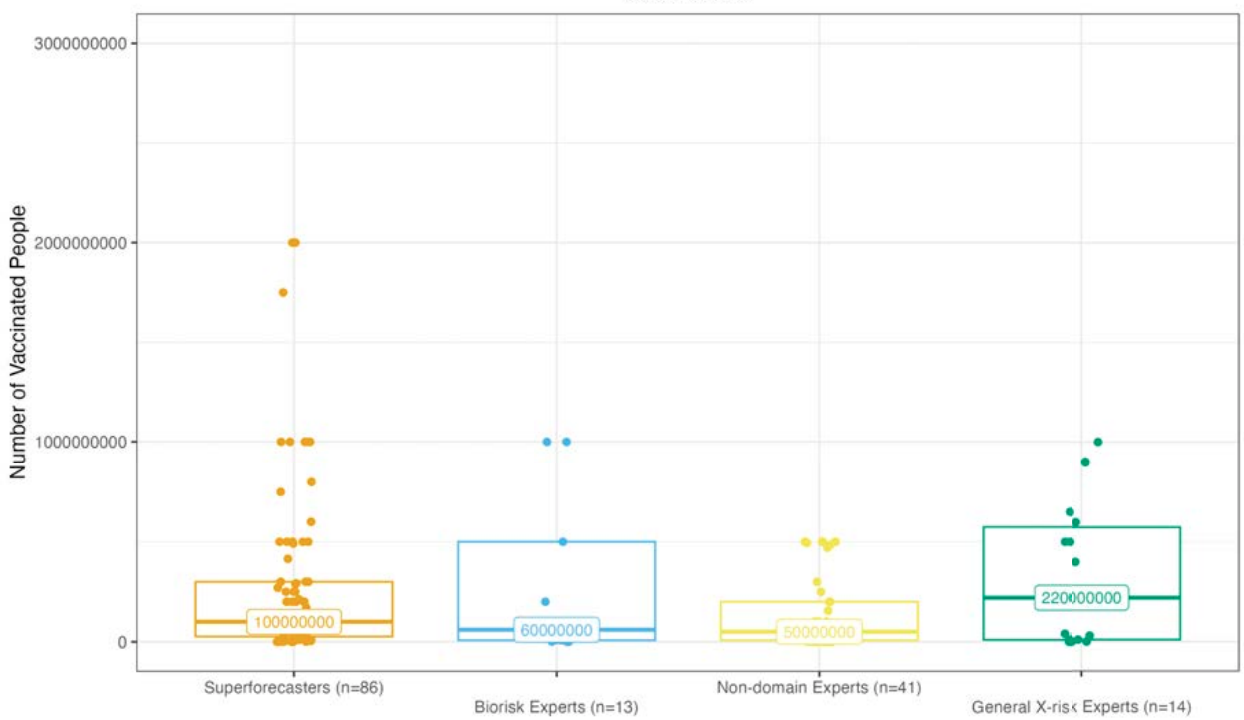

Non-Coronavirus mRNA Vaccine - 2024 - 50th %
Median
200000
100000
0
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 86)
Non-domain Experts (n = 42)
General X-risk Experts (n = 14)
Domain Experts (n = 13)


Non-Coronavirus mRNA Vaccine - 2030 - 50th %
Median
250000000
200000000
150000000
100000000
50000000
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 86)
Non-domain Experts (n = 41)
General X-risk Experts (n = 14)
Domain Experts (n = 13)

Non-Coronavirus mRNA Vaccine
2024 - 50th %



Non-Coronavirus mRNA Vaccine
2030 - 50th %



Sources of agreement, disagreement and uncertainty

On this question there was widespread agreement that the COVID-19 mRNA vaccine had been a success, and about the high potential of future mRNA vaccines.

There was disagreement on the following:

1. Vaccine development and approval timelines.
2. Public perception.
3. The number of vaccines Moderna had in Phase 3 trials—Team 341 said 15; Team 338 said two—and what this means for the question's resolution.
4. COVID-19's effect.
5. How much demand there is for another influenza vaccine.

Arguments given for forecasts ≥ 20,000 (2024), 50,000,000 (2030)

On 1 (development and approval timelines):

- Three mRNA vaccines (CMV, RSV, and cytomegalovirus) are likely to be approved and go to market within two years.[493]
- There is sufficient interest in mRNA vaccines, and large enough potential markets, that at least one vaccine will be approved by 2030.[494]
- mRNA vaccines' short development time means they can be brought to market quickly.[495]

On 2 (public perception):

- Experts are optimistic about the potential of mRNA vaccines for influenza.[496]

On 3 (Moderna):

- Moderna currently has mRNA vaccines for 15 diseases and a flu vaccine in Phase 3, which could lead to three respiratory commercial launches in the next two to three years.[497] A flu vaccine could lead to 261 million vaccinated in the US alone. This or some combination of US/Europe equal to similar numbers (about 200m) is likely by 2030.[498]

---

[493] 337, "Moving on to high plausible forecasts, at least one forecaster stated that all three of the mRNA vaccines (CMV, RSV, and cytomegalovirus) are likely to be approved and go to market within two years."
[494] 336, "Given interest in the technology and the potential markets for the most targeted diseases, it should be expected that at least one vaccine will see approval during this time."
[495] 337, "Because mRNA vaccines have a shorter development time, they can be brought to market quicker to target new variants."
[496] 336, "Experts appear quite optimistic about the outcomes of the phase 3 trials," citing a Chemical & Engineering News piece.
[497] 341, "Moderna announced the development of mRNA vaccines for 15 diseases. We can account for the chance that one of the vaccines passes through phase 3 by 2024. Moderna does indeed have seasonal influenza in Phase III. From their website, "beginning in the fall of 2022, the Company's Phase 3 pipeline could lead to three respiratory commercial launches over the next two to three years", and RSV vaccine trials could finish early 2023."
[498] 341, "Moderna seasonal influenza in Phase III now, likely launched within 3 years. If US gets the flu vaccine, that is a possible 261 million for Moderna mRNA flu. There are two Moderna RSV vaccines in

On 4 (Covid's effect):

- Momentum: mRNA vaccines have momentum right now.[499]

On 5 (Influenza)

- The flu causes many deaths annually, there is a lot of interest in a vaccine for it, and there are currently several candidates in Phase 3 trials.[500] Existing mRNA "players" could enter this market, possibly with a combination flu/COVID shot.[501]

Other arguments:

- Many use cases: there are a high number of potential use cases for mRNA vaccines.[502]
- RSV: this market could be as big as the flu market.[503]
- Non-US countries, including the UK, approve drugs faster than the US.[504]

Arguments given for forecasts ≤ 10,000 (2024), 20,000,000 (2030)

On 1 (development and approval timelines):

2024

- 50% chance none of the vaccines in development will be ready in large enough numbers by 2024.[505]
- "The technology is simply too early to make it through the significant trials and testing required for a novel technology" (Team 345).

2030

- Some mRNA vaccines may not be approved or their production and administration could be limited. 338, "As for 2030 prediction it is much more complicated as we do not know if any of above or any additional vaccines will be successfully approved or how widespread production and administration of those will be."

phase III, could be combined with flu, but may increase the opportunity. Capturing 40% of the US opportunity for flu/RSV gets to 100 million. Europe and rest of world could be another 100 million for mRNA flu/RSV alone. With 15 vaccines in trial from Moderna alone, and Pfizer in Phase 1 on some, chances are very good several will be in market by 2030. There's also HIV, where Moderna has phase 1 trials undergoing."

[499] 339, "Forecasters assuming higher vaccination numbers refer to the historical momentum and the high potential of mRNA vaccines." (This seems implicitly to refer to COVID-19, but does not explicitly state it.)

[500] 341, "Flu vaccine could be developed rapidly, as there is a lot of interest in preventing it given the high amounts of deaths annually. Moderna has several candidates in phase 3 trials."

[501] 340, "The flu/influenza market is a big existing market. mRNA players can potentially enter the market with a combination shot with COVID-19 and consequently build on the success and positive perception of existing mRNA COVID-19 vaccines."

[502] 339, "Moreover, the high number of potential use cases for mRNA vaccines due to the quantity of existing infectious diseases presents ample opportunity for their development."

[503] 340, "The RSV vaccine market may be as big as the flu market according to some estimates."

[504] 341, "Traditionally U.S. FDA approval takes 6-10 months after completion of phase 3, but the UK may be faster, as was the case with COVID. Other smaller countries may be faster."

[505] 342, "50 % chance non of the vaccines in development will be ready in large enough numbers of doses by 2024 for a large amount of people to be vaccinated."

On 2 (public perception):

- Fear of new and relatively untested technology could prevent mass vaccination, especially in the developing world.[506]
- The discovery of long-term harms associated with mRNA vaccines would be an obstacle.[507]
- There will be low demand for mRNA vaccines to enter the market.[508]
- There is a lack of political willpower to create new mRNA vaccines, “perhaps due to extreme risk intolerance (safetyism).”[509]
- The vaccine nationalism witnessed during Covid that prevented countries from using each other’s vaccines will likely continue.[510]

On 3 (Moderna):

- Because of the stated deadlines, and because CMV and RSV are the only two vaccines in Phase 3 from Moderna, it's virtually impossible for any mRNA vaccines to be approved and administered by 2024.[511]

On 4 (Covid’s effect):

- Covid was an extreme case that allowed for fast-tracking; other vaccines will not benefit from the urgency that came from the pandemic, and will be slowed by the fact that mRNA vaccine technology is in its infancy.[512]
- Anti-vaccination sentiment has increased due to Covid.[513]

On 5 (Influenza)

- Flu: flu vaccines that work already exist, narrowing the market for a new one.[514]

---

[506] 342, "even if mRNA vaccines prove very effective, fear of the new and relatively untested technology will prevent mass vaccination, especially in developing world."
[507] 343, "In the event that significant long-term harms are discovered, associated with mRNA vaccination technology, this would negatively impact the rollout of future mRNA vaccines."
[508] 343, [We expect that] "Either no mRNA vaccines have entered the market in our timeline or demand is low for those that have entered the market."
[509] 344, "Despite the technical ability to create new mRNA vaccines, the political willpower seems to be missing, perhaps due to extreme risk intolerance (safetyism)."
[510] "Vaccine nationalism during the present COVID pandemic prevented many countries from vaccinating their citizens with foreign vaccines” (344).
[511] 338, "There are only 2 vaccines in phase 3 or 2/3 both from Moderna. CMV has estimated completion date of July 29, 2025. RSV has estimated completion date of November 30, 2024. [W]hich makes it virtually impossible to be approved and administered by 2024."
[512] 345, "The belief is that covid was an extreme circumstance that allowed fast tracking through much of the red tape that typically proceeds a new untested vaccine technology."
[513] 344, "General anti-vaccination sentiment seems to have increased in some places due to coercive vaccination attempts during the COVID-19 pandemic."
[514] 338, “There are already flu vaccines which are working so unlikely that among 1.5 billion people vaccinated yearly all or even significant number will switch to mRNA vaccines, even if shot will not need to be done every year (something I do not know). My Guess of maximum people vaccinated would be 500 mio if shot works really well.”

Other arguments:

- Misaligned incentives: "Incentives are not perfectly aligned, if more money is to be found in treating illness than curing them" (344).

Other arguments given

Arguments in favor of lower forecasts:

- Competition with traditional vaccines:
    - mRNA coronavirus vaccines were approved during the Covid emergency because they have shorter development times than traditional vaccines, but in non-emergency scenarios, traditional vaccines may be more effective and therefore preferred.[515] [516]
    - Traditional RSV vaccines from large vaccine players are likely to enter the market before mRNA vaccines.[517]
    - “Profits from designer antibody drugs are way higher than from mRNA targets. For the most promising projects there seem to be other non-mRNA based alternatives” (341).
- Cold storage requirements limit adoption in lower-middle-income and low-income countries.[518]

Arguments in favor of higher forecasts:

- By 2030, vaccines targeting HIV, Malaria, or Flu could be approved and have widespread potential.[519]
- mRNA vaccines for non-infectious diseases could surpass expectations in trials.[520]
- Another global-scale pandemic would incentivize governments to use mRNA technology.[521] [522]

---

[515] 337, "mRNA coronavirus vaccines were approved first because they have shorter development times than traditional vaccines, but traditional vaccines may be more effective."
[516] 339, "In addition, there may also be competition between mRNA vaccines and 'traditional' vaccines in some cases."
[517] 340, "Hence, although there may be a reasonable probability of an mRNA vaccine being approved, the product will be second or third to market behind vaccines reporting high efficacy (the opposite of COVID-19)."
[518] 339, "The cold storage requirement for mRNA vaccines may limit the adoption in lower middle-income and low-income countries as well."
[519] 338, "I believe HIV, Malaria and Flu are those with [the] most potential to be widespread."
[520] 343, [We expect] "mRNA vaccines for non-infectious diseases currently in clinical trials [to] overperform relative to our expectations of their timelines, efficacy, demand."
[521] 345, "In addition there is always the possibility of another global scale pandemic that incentivizes governments to make use of the mRNA technology that has proved effective at mitigating the harmful effects of a deadly and contagious pathogen."
[522] 344, "if another pandemic were to hit (Monkeypox? Bubonic Plague 2.0?) as deadly or worse than COVID, this would likely be done in a similar timeframe (pandemic identification to widespread vaccination in less than a year)."

Cross-references with other questions

Q14: Novel Infectious Disease Surveillance System
Q22: PHEIC Declarations with 10k Deaths

Question 14: Novel Infectious Disease Surveillance System

Will a new surveillance system be announced aimed at detecting the spread of novel infectious pathogens, with a commitment of at least $100 million in funding annually…
...by the end of 2024?
...by the end of 2030?

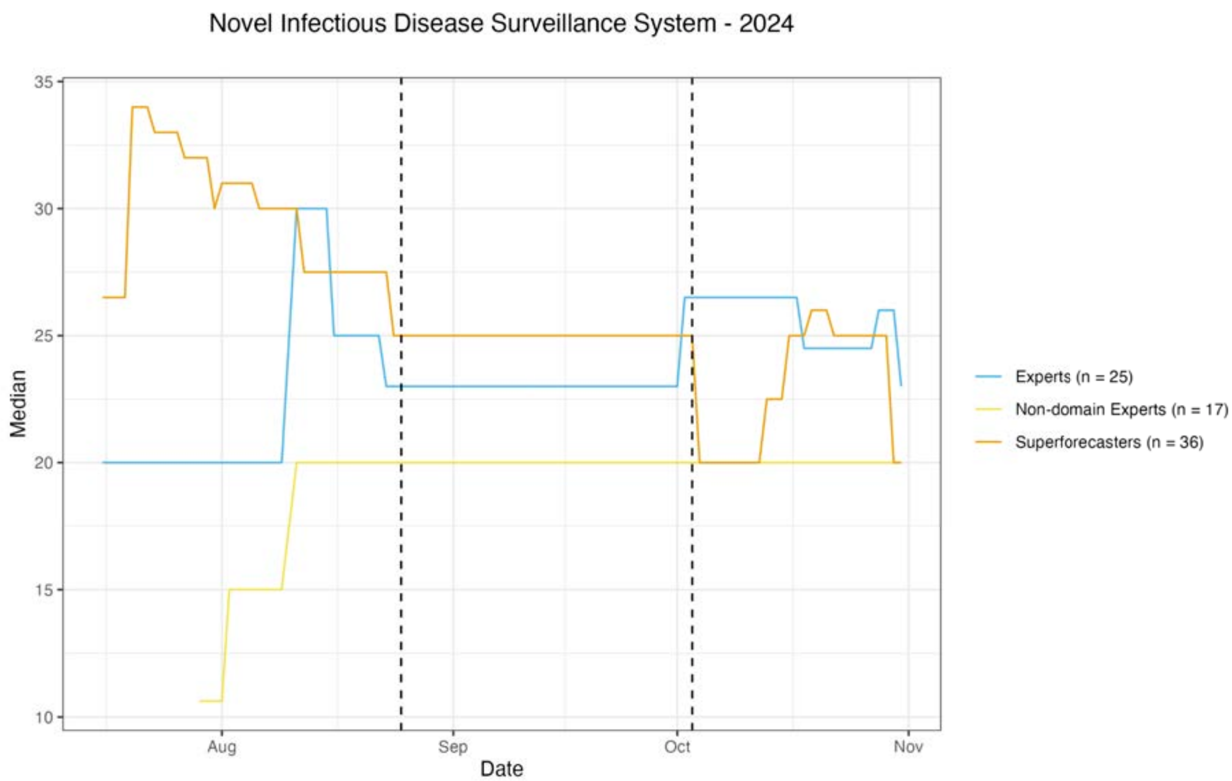
Question and resolution details, prior forecasts, and other relevant sources

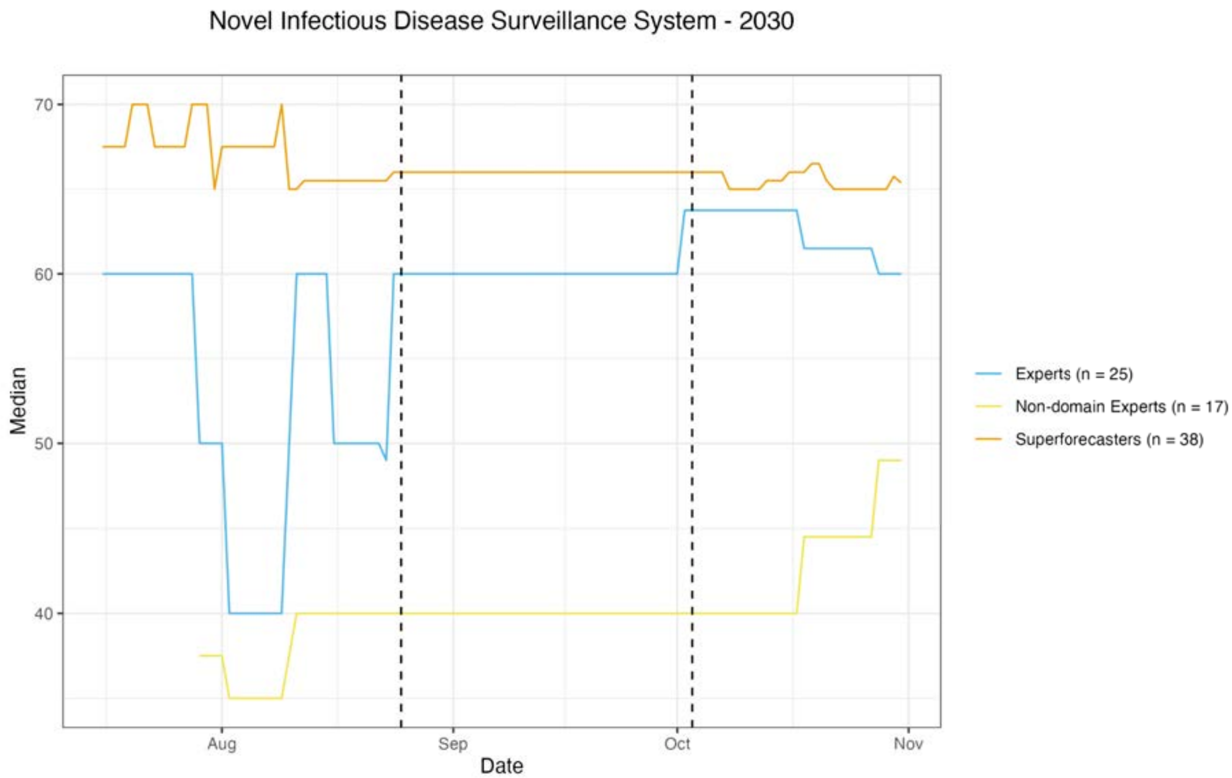
Results[523]

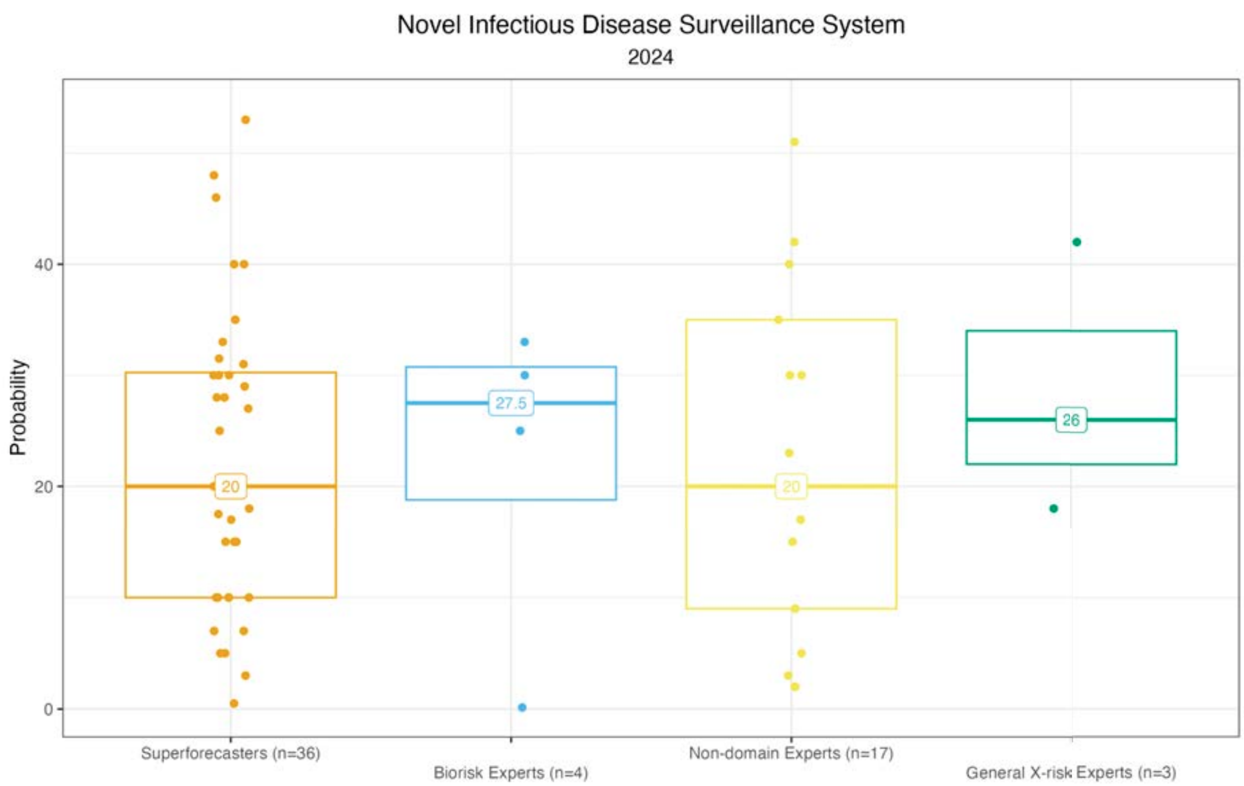
Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 36)	2024	20%	20%	21.54	-36.94%
	2030	70%	65.75%	26.88	-12.08%
Domain Experts (N = 4)	2024	35%	27.5%	6.03	+148.27%
	2030	70%	70%	10.38	+138.35%
General X-Risk Experts (N = 3)	2024	31%	26%	1.41	+764.1%
	2030	71%	63%	1.41	+988.62%

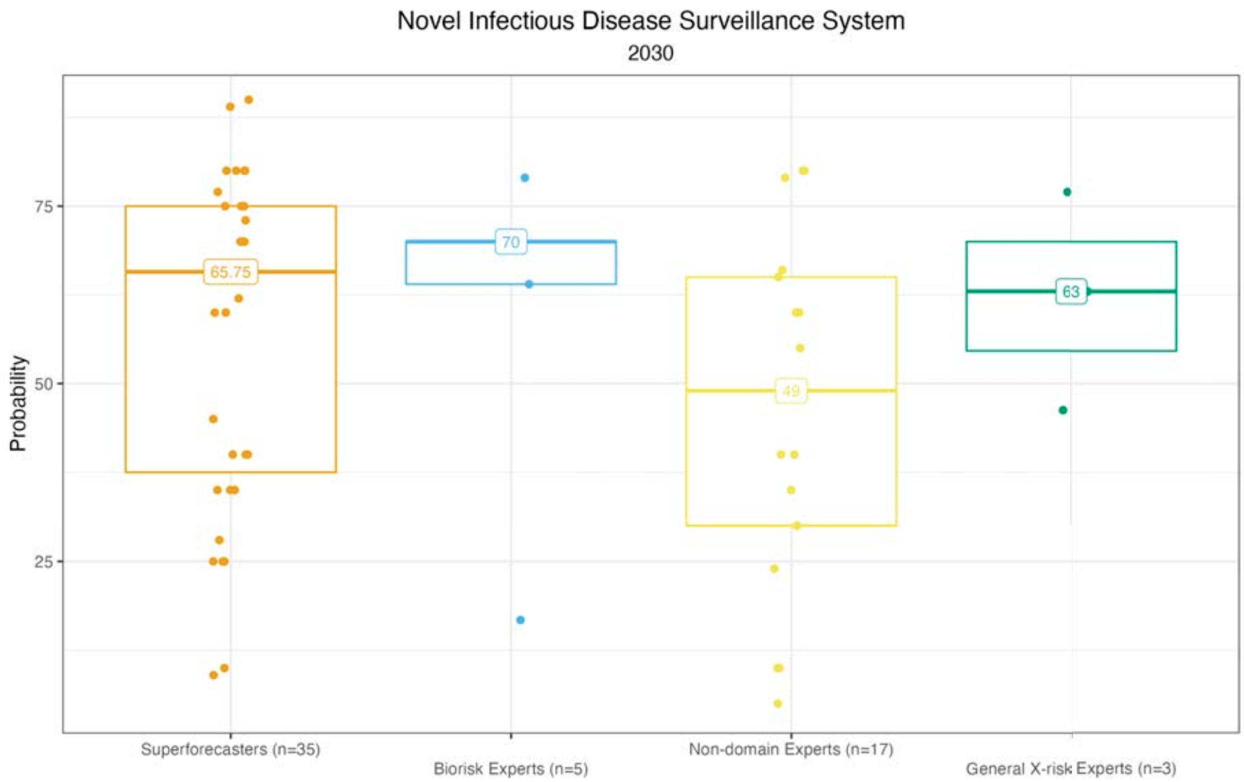
[523] Numbers of forecasters are given as of Stage 4 of the XPT.

Non-Domain Experts (N = 17)	2024	15%	20%	20.06	-8.22%
	2030	40%	49%	28.13	-11.8%




Novel Infectious Disease Surveillance System - 2030
Median
70
60
50
40
Aug
Sep
Oct
Nov
Date
Experts (n = 25)
Non-domain Experts (n = 17)
Superforecasters (n = 38)


Novel Infectious Disease Surveillance System
2024
Probability
40
20
0
20
27.5
20
26
Superforecasters (n=36)
Biorisk Experts (n=4)
Non-domain Experts (n=17)
General X-risk Experts (n=3)



Sources of agreement, disagreement and uncertainty

On this question there was agreement that:

- Early detection of infectious diseases would save lives and money.
- COVID-19 impacted public perception in a way that will influence decision-making.
- Relevant technology like new contact tracing systems and next-generation sequencing (NGS) is advancing while the cost is decreasing, which could make surveillance systems more feasible (337, 343).

There was disagreement around:

1. Whether there would be sufficient political will for government funding of such a system
2. Existing initiatives
3. The effectiveness of such a system

Arguments given for forecasts ≥ 25% (2024), 60% (2030)

On 1 (political will):

- Individual countries' governments are capable of funding such a system.[524] [525]

[524] 338, "This size of budget put this well within reach of individual countries."

[525] 336, "Almost all forecasters agreed that $100 million is a relatively small amount that could be provided by a state government."

- Covid-19 has demonstrated that there are strong financial incentives for such a system.[526]
- Governments are risk-averse and may see a relatively low-cost surveillance system as a way to avoid risk, and therefore support it.[527]
- The public has recently been shown the importance of early detection of infectious diseases by COVID-19 and Monkeypox, which should raise support for early detection.[528] [529]

On 2 (existing initiatives):

- "There are multiple initiatives seeking to monitor pathogen spread such as: WHO pandemic intelligence, UK Genomic Radar, Nucleic Acid Observatory, and Africa PGI" (340).
- "There is already a concrete interest in similar projects, see for instance:" The Origins and Future of Sentinel: An Early-Warning System for Pandemic Preemption and Response, (339).

On 3 (effectiveness):

- The decreasing cost of NGS, and increased data sharing will make a surveillance system a strong value proposition, and new pathogens will likely increase demand.[530]

On private sector funding:

- Multiple philanthropic organizations (Open Philanthropy, FTX,[531] Bill and Melinda Gates Foundation,[532] Rockefeller,[533] the Chan-Zuckerberg Foundation[534]) are likely to fund this.
- Private sector involvement from large foundations could increase the likelihood. 345, "A factor that may increase the likelihood of a surveillance system on the emergence and spread of pathogens is the involvement of the private sector."

Only an announcement is required to satisfy the question conditions.

---

526 339, "There are clear strong financial incentives in creating such a program, which are now evident to the general public and governments after the Covid-19 pandemic."

527 345, "Governments are, as a rule, characterized by risk aversion. Particularly when a lack of response may entail large political costs."

528 336, "COVID and Monkeypox are still fresh in the public's mind and highlights how early detection of infectious diseases saves lives and money."

529 345, "The COVID19 pandemic (and also the emergence of monkeypox) have put the issue of pathogen-neutralizing systems at the center of public debate."

530 337, "By 2030 as new pathogens will come into play + the cost of NGS coming down + more data sharing occurring. This will make a surveillance system have a stronger value proposition and therefore countries more willing to invest."

531 338, "At least two large funders (OP and FTX) appear likely to fund such an initiative."

532 337, "The median takes into account that private investment like the The Bill and Melinda Gates Foundation [9] or funding from governments like in the US [10] are likely to occur by 2030."

533 340, "There are multiple funding sources backing these pathogen surveillance initiatives including the Gates Foundation, Rockefeller, FTX Future Fund and of course Public Government budgets, including existing WHO resources."

534 339, "$100 million is a very low bar to clear for an international body, to the point that its implementation could even happen with alternative sources of funding, e.g. by philanthropists such as the Bill Gates Foundation, the Chan-Zuckerberg foundation, the Future Fund, all of which have spoken in favour of the creation of such an organism."

- "Further, the question requires only an announcement to be made, and this could happen as early as the 2023 U.N. Future Summit" (339).

Arguments given for forecasts ≤ 7% (2024), 33% (2030)

On 1 (political will):

- Governments will probably not fund such a project.[535] They have other priorities,[536] and the US in particular is unlikely to do this because it already gives $100m to the WHO.[537]
- Public perception will not be in favor.[538]
- “Contact tracing” has been politicized, and will work against such a system’s viability in terms of political will.[539]
- There is low international support for cooperation.[540]

On 2 (existing initiatives):

- No specific plans for a new surveillance system are known.[541]

On 3 (effectiveness):

- Contact tracing is not effective for stopping pathogen spread within animals, who may carry infectious agents across political boundaries.[542]

---

[535] 342, "While there are lots of advocates, it doesn't look like governments are willing to put up the cash to make this happen."
[536] 342, "Gov't capacity to implement limited in China and United States. EU currently concerned with rearmament so unlikely to divert funds to execute this project."
[537] 344, "Based on the fact that USA commits to 100 million dollars per year to WHO I would say that there is a "low" chance that the commitment of that amount has the political will to be allocated."
[538] 344, "My percentages are relatively low, compared to how important I think the capability is, but I think public perception will be a significant factor in the decision making."
[539] 341, "Using the word contact tracing will bring political interventions."
[540] 342, "However, [a forecaster] suspected that the current support for such an initiative is probably higher than "normal" to the collective memory of covid, but - on the other hand - international support for cooperation is probably very low (US/Russia/China)."
[541] 343, "We are not aware of specific plans to institute a new surveillance system."
[542] 344, "Based on the fact that USA commits to 100 million dollars per year to WHO I would say that there is a "low" chance that the commitment of that amount has the political will to be allocated."

345, "Governments are, as a rule, characterized by risk aversion. Particularly when a lack of response may entail large political costs."

336, "COVID and Monkeypox are still fresh in the public’s mind and highlights how early detection of infectious diseases saves lives and money."

345, "The COVID19 pandemic (and also the emergence of monkeypox) have put the issue of pathogen-neutralizing systems at the center of public debate."

344, "My percentages are relatively low, compared to how important I think the capability is, but I think public perception will be a significant factor in the decision making."

341, "Using the word contact tracing will bring political interventions."

342, "However, [a forecaster] suspected that the current support for such an initiative is probably higher than "normal" to the collective memory of covid, but - on the other hand - international support for cooperation is probably very low (US/Russia/China)."

Other arguments:

- New contact tracing systems might not require $100m, which would mean even if such a system was funded, the question would not be resolved.[543]
- The development process for such a system is lengthy and complex; the WHO's EWARS (Early Warning and Response System), for comparison, took over six years to develop.[544]

Other arguments given

Arguments in favor of lower forecasts:

- An economic recession might discourage investment in such a system.[545]
- There may be a decentralized, internationally coordinated way of accomplishing the same goal that doesn't require $100m.[546]
- Existing systems are more likely to be expanded rather than creating a new system.[547] [548]
- People may be "tired" of COVID-19, resulting in less support for preventative measures for future pandemics.[549] Motivation to prevent future pandemics will drop off.[550]

Arguments in favor of higher forecasts:

- Surveillance benefits the state, making it appealing for state actors.[551]
- Recent research shows people in the EU are willing to spend a significant amount on surveillance.[552]

---

543 343, "New contact tracing systems may be built on the back of existing internet infrastructure, and may therefore not even meet the modest figure of $100 million."
544 341, "EWARS programs take 6+ years of gathering data, building a database, modeling, etc."
545 339, "It has been pointed out that the incoming economic recession might discourage such an investment, especially if it turns out to require higher funding amounts to function well."
546 337, "A point of uncertainty could be that sequencing and measurements stay decentralized and data sharing is centralized. If this works well, then a $100M surveillance system might not be necessary."
547 336, "The major disagreement was about whether the specific question would resolve as 'True' or 'False' if funding was allocated to expand existing systems, as opposed to creating a new system from scratch."
548 339, "It has been pointed out that agencies and programs with very similar aims already exist, so a 'new' surveillance system might not actually be the way forward that governments/funders choose."
549 338, "I don't think this is a sure bet. People may be tired of C-19, and memories/responsibility are not our forte."
550 337, "Skepticism about the ability to sustain motivation, skepticism about the ability to maintain the magnitude of the response."
551 343, "A contact tracing system has the potential to improve surveillance by the State, which is the kind of program state actors rarely turn down."
552 344, "Recent research on the willingness to pay for surveillance in the EU found that, on average, people are willing to spend €264/year, which roughly translates into 5% of total health spending."

Cross-references with other questions

Q13: Non-Coronavirus mRNA Vaccine
Q14: Novel Infectious Disease Surveillance System

Question 15: Non-State Actor Bioweapon 1k Deaths

How many times will a non-state actor using biological weapons that involve a contagious agent be the cause of death for at least 1,000 people…
…by the end of 2024?
…by the end of 2030?
…by the end of 2050?

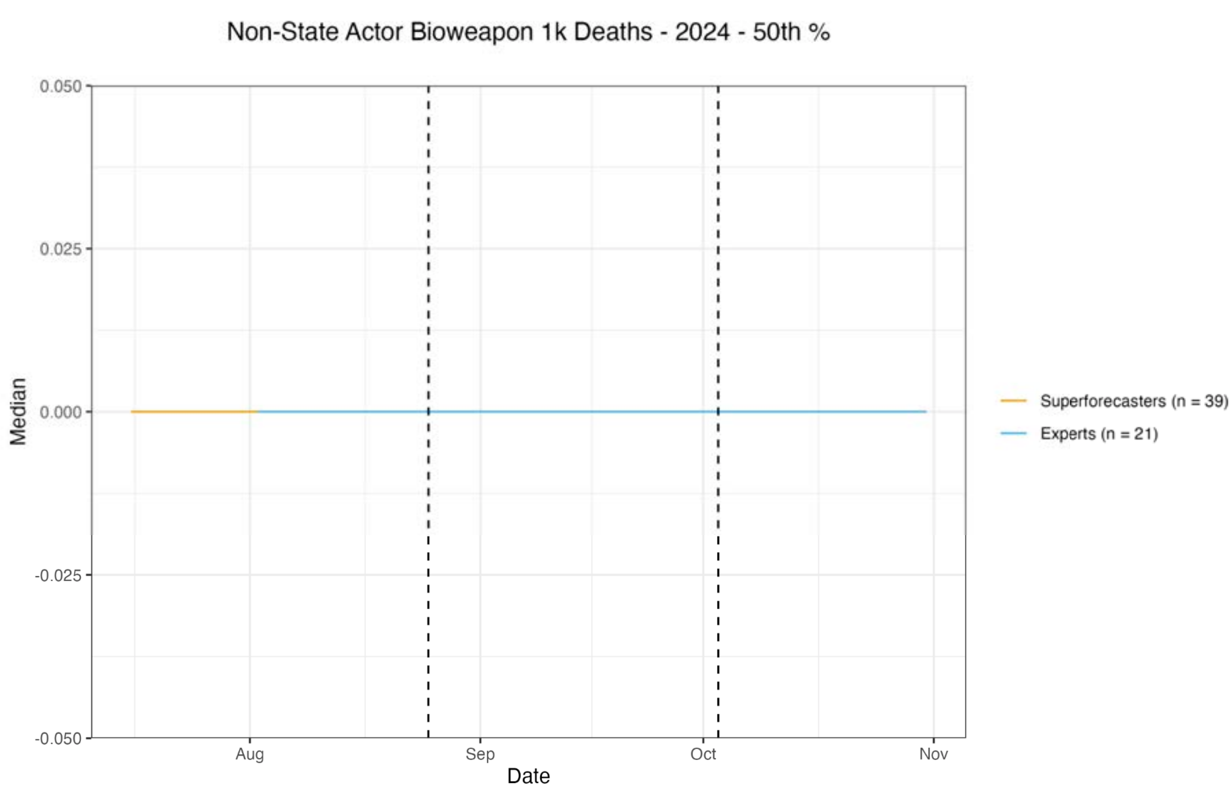
Question and resolution details, prior forecasts, and other relevant sources

Results[553]

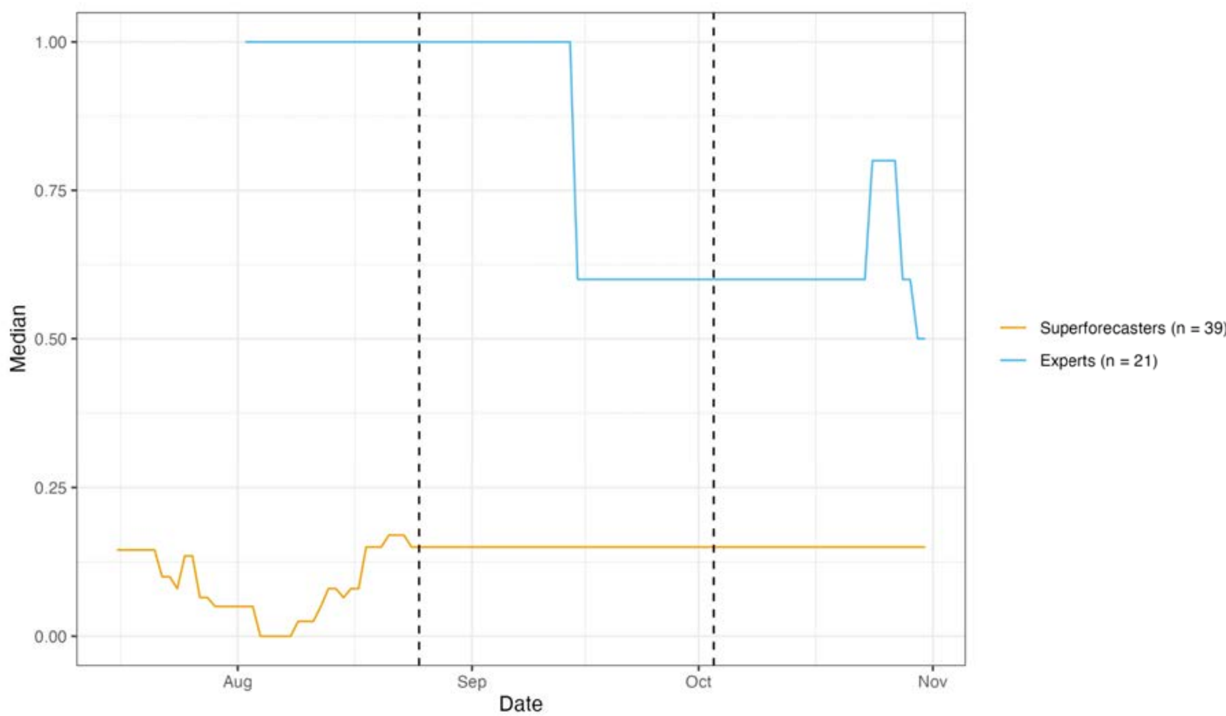
Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 39)	2024	0	0	0.35	-1.56%
	2030	0.1	0.15	0.87	-13.16%
	2050	1	1	3.48	-26.29%
Domain Experts (N = 7)	2024	0	0	0.89	-12.29%
	2030	1	1	4.28	-15.91%
	2050	2	2	8.26	-14.86%

[553] Numbers of forecasters are given as of Stage 4 of the XPT. Public survey respondents were surveyed outside of the tournament context.

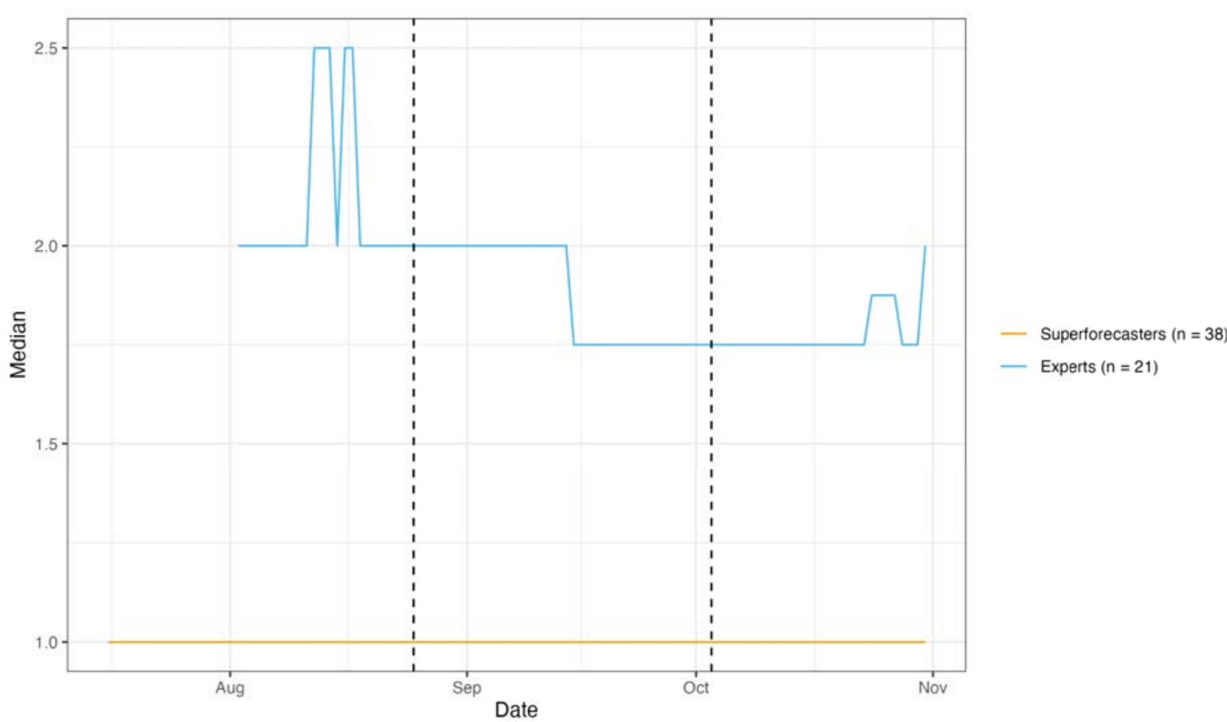
General X-Risk Experts (N = 5)	2024	0	0.1	0.23	-21.34%
	2030	0	0.5	1.73	-22.54%
	2050	1	2	2.31	-27.18%
Non-Domain Experts (N = 9)	2024	1	0	9.71	-91.91%
	2030	0	0	1.73	-13.39%
	2050	3.5	1.75	48.7	-94.75%
Public Survey (N = 480)	2030	0		4569	-
	2050	1		32314.68	-
	2100	2		458628.75	-

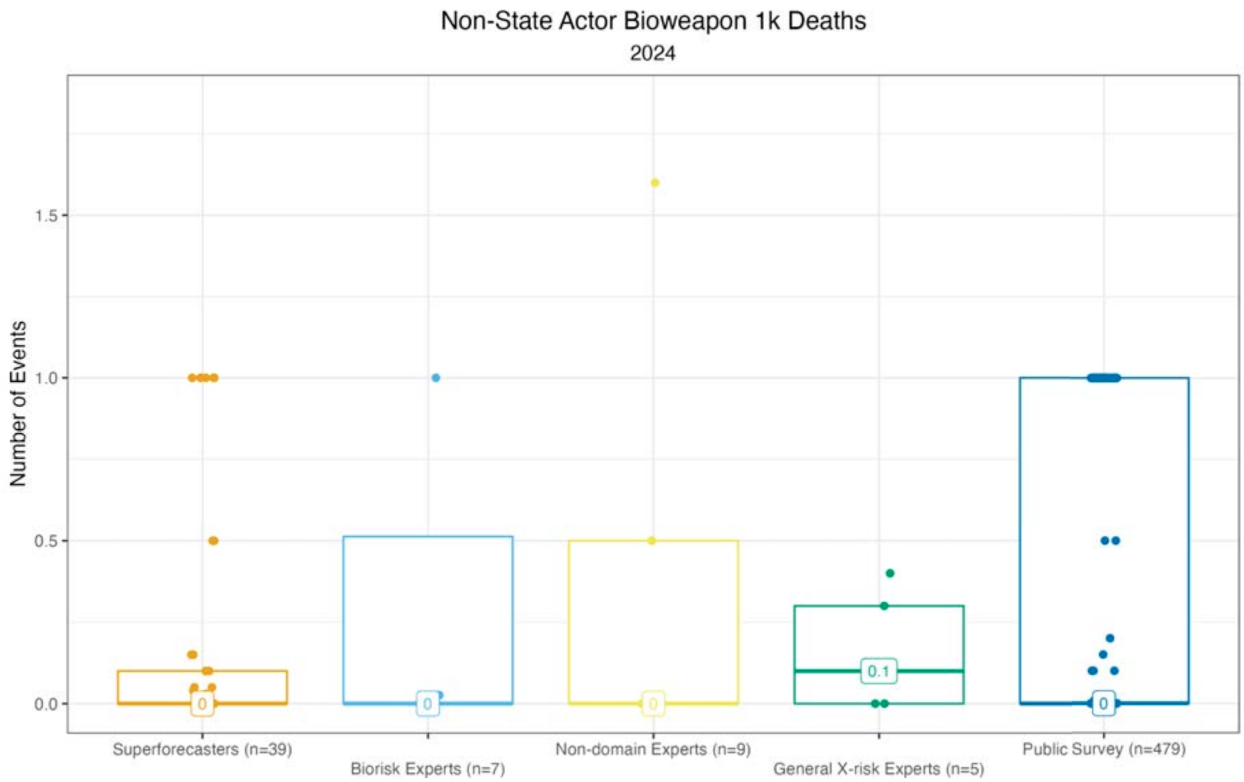


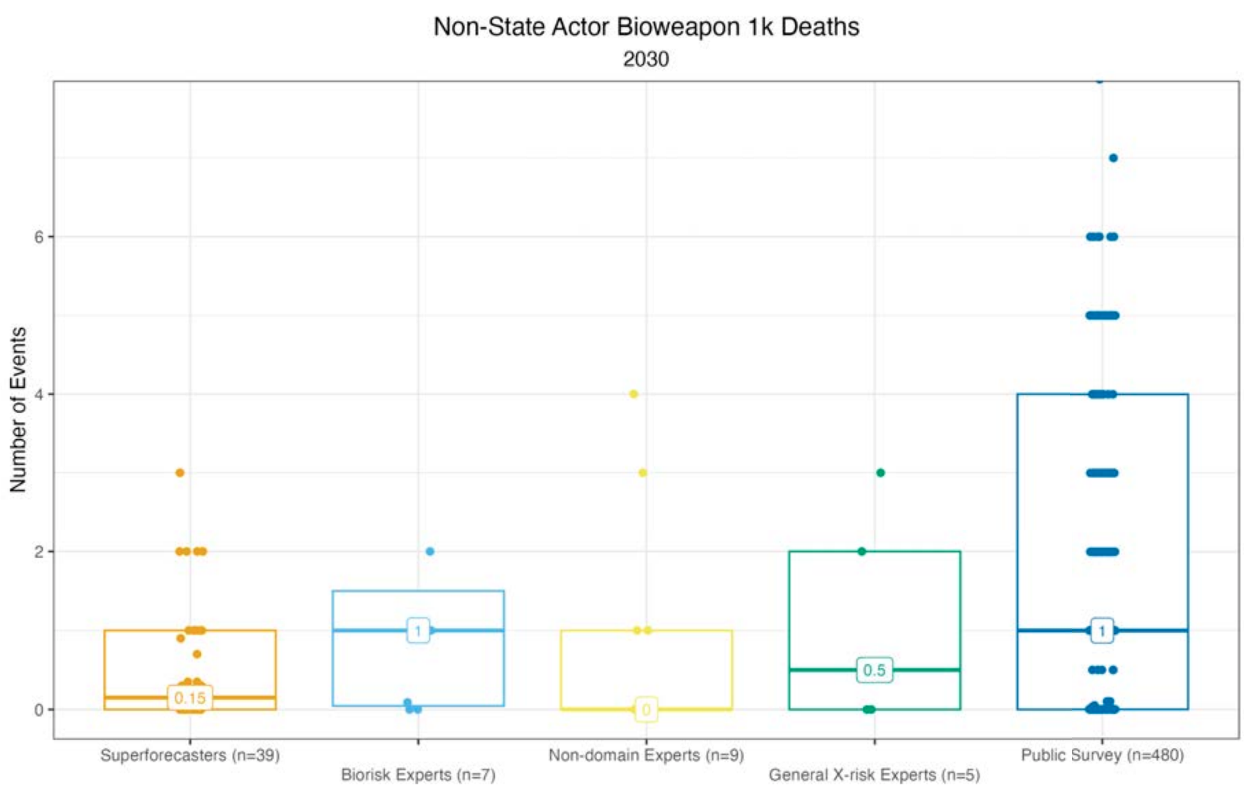
Non-State Actor Bioweapon 1k Deaths - 2030 - 50th %

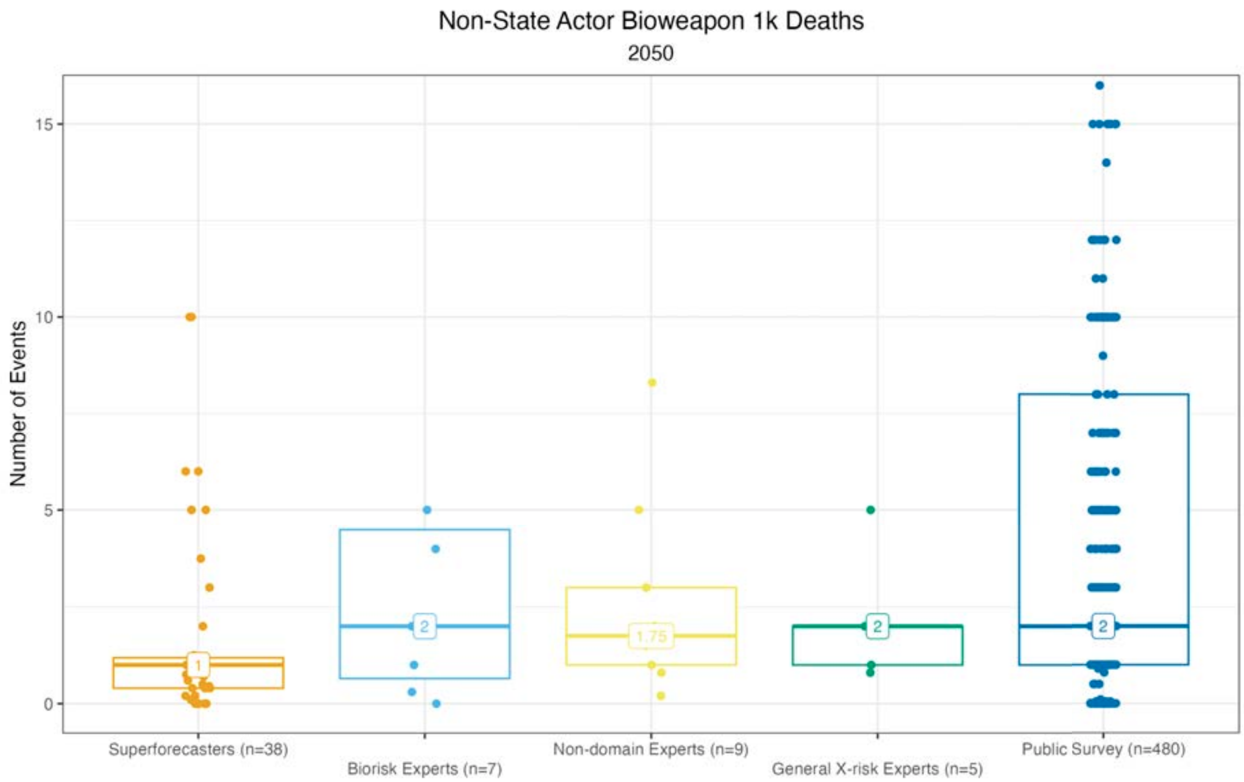


Non-State Actor Bioweapon 1k Deaths - 2050 - 50th %




Non-State Actor Bioweapon 1k Deaths
2024
Number of Events
0.0
0.5
1.0
1.5
0
0
0
0.1
0
Superforecasters (n=39)
Biorisk Experts (n=7)
Non-domain Experts (n=9)
General X-risk Experts (n=5)
Public Survey (n=479)


Non-State Actor Bioweapon 1k Deaths
2030
Number of Events
0
2
4
6
0.15
1
0
0.5
1
Superforecasters (n=39)
Biorisk Experts (n=7)
Non-domain Experts (n=9)
General X-risk Experts (n=5)
Public Survey (n=480)



Sources of agreement, disagreement and uncertainty

There was agreement that:

1. A non-state actor has never used biological weapons that involve a contagious agent to cause the deaths of at least 1,000 people.
2. Biological weapons are difficult to develop and control.
3. However, biotechnology is improving and becoming increasingly easy to access.

There were no explicit disagreements between higher and lower forecasts, merely differing emphases.

There was uncertainty around:

1. The extent to which technological advancements will make biological weapons more accessible to non-state actors.
2. The potential for global development, scientific training, and urban migration to increase the number of attacks and their potential death toll.
3. The likelihood of non-state actors gaining access to biological weapons through compromised scientists, researchers, or government agents.

Arguments given for forecasts ≤ 0 (2024), 0.8 (2030), 1 (2050)

Base rates

- The historical base rate for deadly bioterrorist attacks is low. Since WWII, the deadliest was the 2001 anthrax attack in the US, which killed five.[554]
- A non-state actor has never killed this many people with biological weapons.[555]

Reasons bioterrorism is not done often

- There are more accessible ways to commit terrorism.[556] [557]
- It's hard: bioterrorism requires a high degree of expertise,[558] and engineering novel pathogens remains challenging even for experts.[559]
- Inefficient: biological weapons are "not a very efficient way of killing." (337)
- It could backfire: it poses a risk to the perpetrating group and is unproven as a way of killing a lot of people intentionally.[560]
- Difficult to control: biological agents are difficult to control and develop into a sustainable form.[561]

Reason bioterrorism at this particular scale will not be done

- "Terrorist groups may be unable to scale up production and dissemination, making attacks killing 1,000 people unlikely." (343)

Case studies

- ISIS, with sufficient resources and a compatible goal (killing indiscriminately), did not attempt large-scale bioweapon attacks.[562]
- Aum Shinrikyo, which had relevant expertise, failed in its bioterrorism attempts, which indicates the difficulty in executing successful large-scale bioweapon attacks.[563]

---

[554] 337, "The deadliest bioterrorist attack in the post-World War II era was the 2001 anthrax attack in the United States which killed five people."
[555] 338, "There is no known historical incident where a non-state actor has killed this many people with biological weapons, despite state stockpiles of biological weapons dating back to the 1940s."
[556] 343, "Bioterrorism doesn't seem to have happened historically (at least very uncommon). The lack of historical record, paired with more accessible forms of terrorism, leads some forecasters to see this as unlikely."
[557] 338, "So even if terrorist groups have both the ability and motive to carry out a bioterrorism attack that kills over 1000, they may decide not to simply because they also have access to other, easier (and more easily targeted) ways of causing mass casualties on that scale."
[558] 343, "Bioterrorism requires a much higher degree of expertise/competence to develop/acquire/deploy at scale."
[559] 341, "There is considerable evidence that this task still is extremely difficult even for a team of experts, which is probably the main reason why no one so far is known to have succeeded in engineering a novel pathogen that may thrive in the wild."
[560] 337, "Unlike a bomb, which can blow up one martyr at a time without offing the entire group, a contagious biological weapon runs the risk of killing off the entire group (high risk) and it has never done much damage before (low reward)."
[561] 341, "Bio-weapons are not particularly useful because biologic agents are difficult to control and to develop into a form that can sustain effects despite environmental degradation."
[562] 338, "The fact that ISIS nonetheless carried out no biological weapons attacks, and is not known to have attempted them, suggests that even non-state actors in an unusually strong position to access biological weapons, and with the goal of causing indiscriminate mass civilian casualties, either will not want to carry out biological attacks or will be unable to."
[563] 338, "This also suggests that killing large numbers of people with biological agents is difficult even for groups with significant scientific expertise."

Counter-measures: medicine and security:

- Medicine and early-warning detection systems will advance, mitigating risks.[564]
- Any successful attack would likely lead to increased security measures, making subsequent attacks more difficult.[565]

Arguments given for forecasts ≥ 1 (2024), 1 (2030), 2.5 (2050)

Technological advancements will likely increase risk

- Increased accessibility of biotechnology,[566] and decreased costs[567] could raise the risk of bioterrorism events.
- The Carlson Curve, which predicts “that the doubling time of DNA sequencing technologies, whether using cost or performance as the measurement, would be as fast as Moore’s Law” (344), suggests that DNA sequencing technologies will improve rapidly, potentially increasing the likelihood of such attacks.[568]

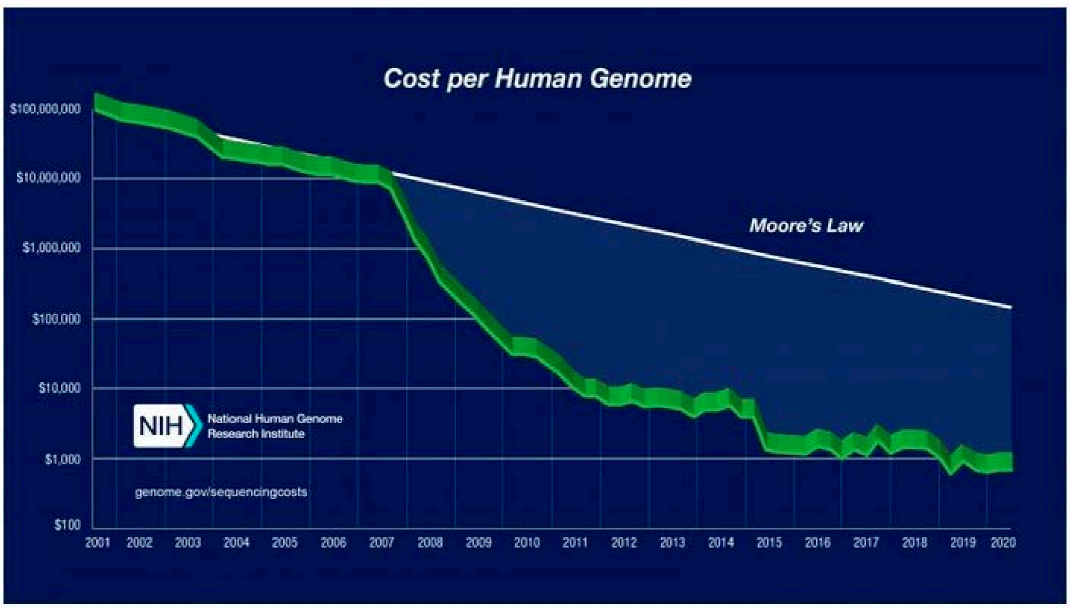


Image from NHGRI (2020).

- Delivery of bioweapons, a current limiting factor, could greatly improve by 2050.[569]

---

[564] 337, "Forecasters also anticipate that advances in modern medicine, coupled with early-warning detection systems, will help mitigate risks posed by bioterrorism."
[565] 343, "Any successful attack - whether it breaches the 1,000 deaths threshold or not - would likely lead to severe security measures, making subsequent attacks more difficult."
[566] 340, "The increasing accessibility of biotechnology will greatly increase this risk in the future."
[567] 344, "This might become much easier to do in the future due to the heavy drop in the price of such technologies."
[568] 344, "A strong argument made by one of the forecasters was the introduction of the Carlson Curve."
[569] 336, "The delivery of bio weapons is still not so great, so this is a limiting factor that could greatly improve."

Changes in the world will affect susceptibility

- “Increasing global development and greater scientific training” (340) might increase the “number and geographic spread” of attacks.
- Dense, highly populated areas are more susceptible to this kind of attack, and urban migration will increase this.[570]
- Climate change could cause more violence to occur due to resource scarcity.[571]
- Previously uncommon terrorist methods, such as airline hijackings and mass shootings, have seen a historical trend toward normalization. This kind of normalization could happen with a low-cost and technologically accessible bioweapon.[572]

Access channels

- Non-state actors might gain access to biological weapons through compromised scientists, researchers, or government agents.[573]
- Though “non-state actors in richer parts of the world are more likely to develop sophisticated biological weapons,” those weapons could be sold to people in other regions.[574]

Other arguments given

Arguments in favor of lower forecasts:

On terrorism in general

- Terrorist attacks usually involve fewer than 1,000 deaths.[575]
- The number and intensity of terrorist attacks have been decreasing globally since 2014.[576]

---

[570] 340, "I would also expect highly populated and dense areas to be susceptible to these forms of attacks. With rising urban migration and greater global connectedness, the death toll of biological attacks would probably increase in the coming years.”

[571] 336, "It's possible more violence may occur by 2050 due to resource scarcity caused by climate change."

[572] 336, "History has cases of a method being rare and then becoming commonplace. Airline hijackings were at a relatively low level until the mid 1960's. Between 1968 and 1972 there were 326 hijackings globally. Mass shootings are another crime that have dramatically increased in a short period of time. Relatively simple bioweapons could become a relatively low cost and technologically accessible terror tool in the next 28 years."

[573] 342, "The exposure and access of these non-state actors to some of these weapons through some compromised scientists, researchers and government agents, circumvention of all the international prohibition conventions against such acts, in addition to some kind of wrong ideological belief system psychology that prompt them into such an act."

[574] 344, "I agree that non-state actors in richer parts of the world are more likely to develop sophisticated biological weapons. However, you shouldn't completely dismiss the regions with the most terrorist attacks. The biological weapons could be sold or whatever."

[575] 344, "Terrorist attacks are usually events of < 1000 deaths."

[576] 344, "The trend of the number and intensity of terrorist attacks has been decreasing globally from an all-time high at 2014."

Reasons bioterrorism is not done often

- It's expensive, compared to other possible choices.[577]
- Possibly due to ethical or ideological concerns.[578]

Reasons bioterrorism will likely not be done at this scale

- Preliminary tests would be necessary to execute this kind of attack successfully, but those tests would also significantly increase the risk of being caught.[579]
- Strict international conventions and government regulations make it difficult for non-state actors to access biological weapons.[580]

Attribution

- Attribution difficulties in proving the origin of biological weapon use make the question difficult to resolve.[581]

Arguments in favor of higher forecasts:

Non-typical examples of non-state actors

- Apolitical individuals: "Forecasters also noted that the phrase 'non-state actor' allows for incidents in which the primary motive isn't terrorism," Team 337 noted, such as "suicides executed in a such a way that they also qualify as mass murder incidents," such as "mass shootings, or intentional downing of airplanes by pilots." They wrote that "if an overworked lab technician were to decide to commit suicide in such a fashion, perhaps one that worked in a BSL-4 lab, the results could be dire."
- State actors sometimes sponsor non-state actors, which raises the risk compared to only considering non-state actors acting alone.[582]

Difficulty in detection (and prevention)

---

[577] 340, "A biological weapon will not be the weapon of choice. There are more cost-effective methods available for an organization bent on causing destruction."

[578] 336, "Not many terrorist groups appear to have pursued biological weapons, perhaps due to the ethical/ideological qualms with such methods."

[579] 340, "In order to launch an effective attack, there would also need to be plenty of tests conducted beforehand. Without testing, there would be virtually no chance of this event occurring. But testing also increases the risks of exposing the secret program to state intelligence and law enforcement agencies."

[580] 342, Low expected probability "due to global stringent conventions by United Nations and other world governing bodies and nations government prohibiting, sanctioning and strictly discouraging such an occurrence from happening in any form or guise that makes it impossible for these non-state actors to have access to such a weapon of mass destruction."

[581] 340, "The problem of proving the origin of a particular case especially when the agent is endemic poses an immense attribution problem."

[582] 337, "[S]tate actors often sponsor non-state actors to enhance plausible deniability. [...] Forecasters perceive this potential symbiotic relationship between state and non-state actors as raising the risk."

- Unclear past usage: some forecasters believe biological weapons may have already been used but not clearly identified as such, indicating global reach and difficulty in stopping them.[583]

Copycat effect

- Successful bioweapon attacks may inspire more incidents.[584]

Cross-references with other questions

Q16: State Actor Bioweapon 1k Deaths

Q17: Non-State Actor Bioweapon 100k Deaths

Q18: State Actor Bioweapon 100k Deaths

Question 16: State Actor Bioweapon 1k Deaths

How many times will a state actor using biological weapons that involve a contagious agent be the cause of death for at least 1,000 people…
…by the end of 2024?
…by the end of 2030?
…by the end of 2050?

Question and resolution details, prior forecasts, and other relevant sources

Results[585]

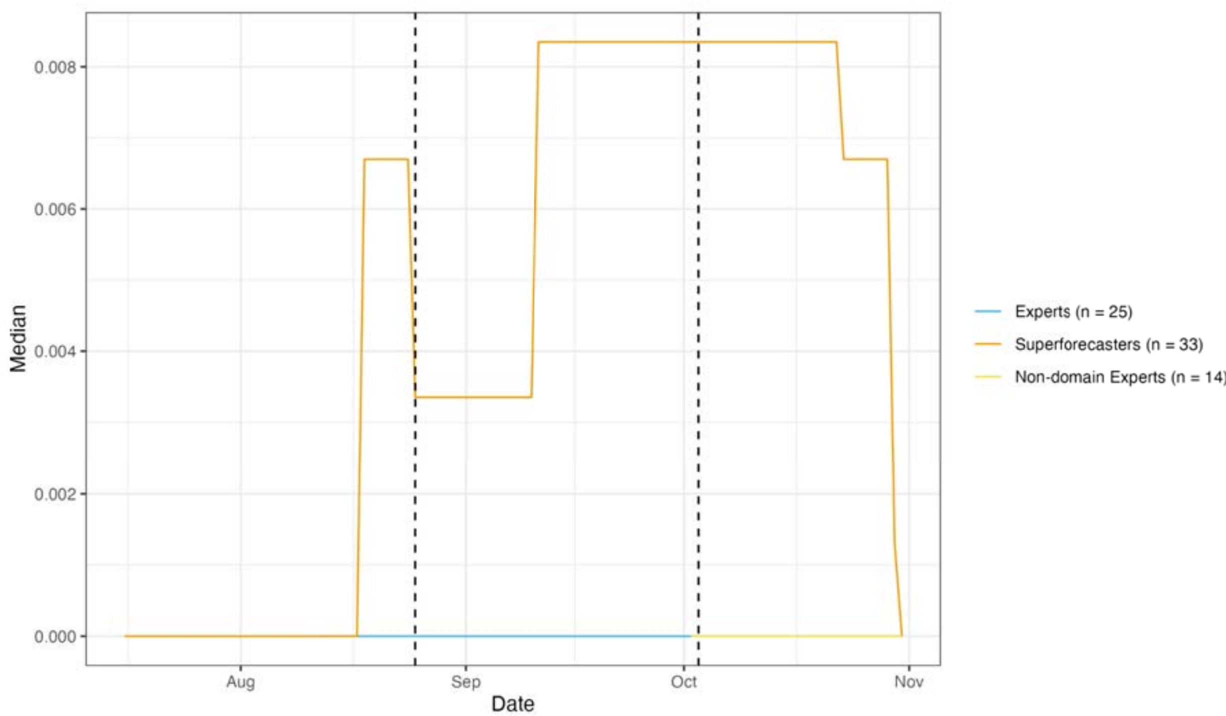
Group	**Year**	**Stage 1 Median**	**Stage 4 Median**	**Stage 1 Standard Deviation**	**Percent Change in Standard Deviation, Stage 1 to Stage 4**

[583] 343, "Some forecasters believe that it may not be clear that a biological weapon has been used and, in fact, they have already been used. The suspected attacks mentioned [examples not included in wiki] would indicate a global reach, making them difficult to stop."
[584] 341, "Once it has happened once, it will inspire more incidents."
[585] Numbers of forecasters are given as of Stage 4 of the XPT. Public survey respondents were surveyed outside of the tournament context.

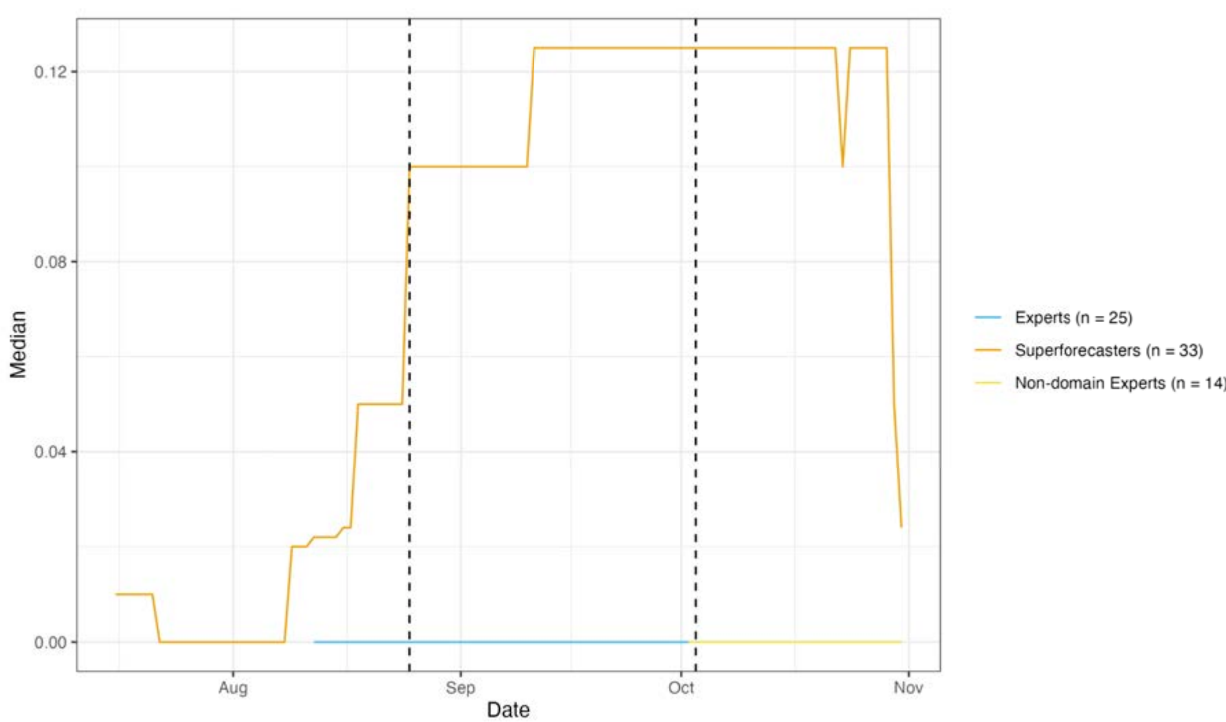
Super-Forecasters (N = 33)	2024	0	0	0.68	-47.27%
	2030	0.01	0.024	0.77	-11.38%
	2050	0.12	1	3.04	-47.3%
Domain Experts (N = 6)[586]	2024	0	0	0.89	-8.71%
	2030	0	0.05	1.30	-7.91%
	2050	0	1.4	2.49	-38.48%
General X-Risk Experts (N = 5)	2024	0	0	0	Inf
	2030	0	0	n/a	n/a
	2050	1	0.43	n/a	n/a
Non-Domain Experts (N = 14)	2024	4	0	5.66	-94.33%
	2030	3	0	4.24	-80.77%
	2050	1	0.6	1.41	+2.31%
Public Survey (N = 479)	2030	0		920.29	-
	2050	1		22845.36	-
	2100	2		45691.93	-

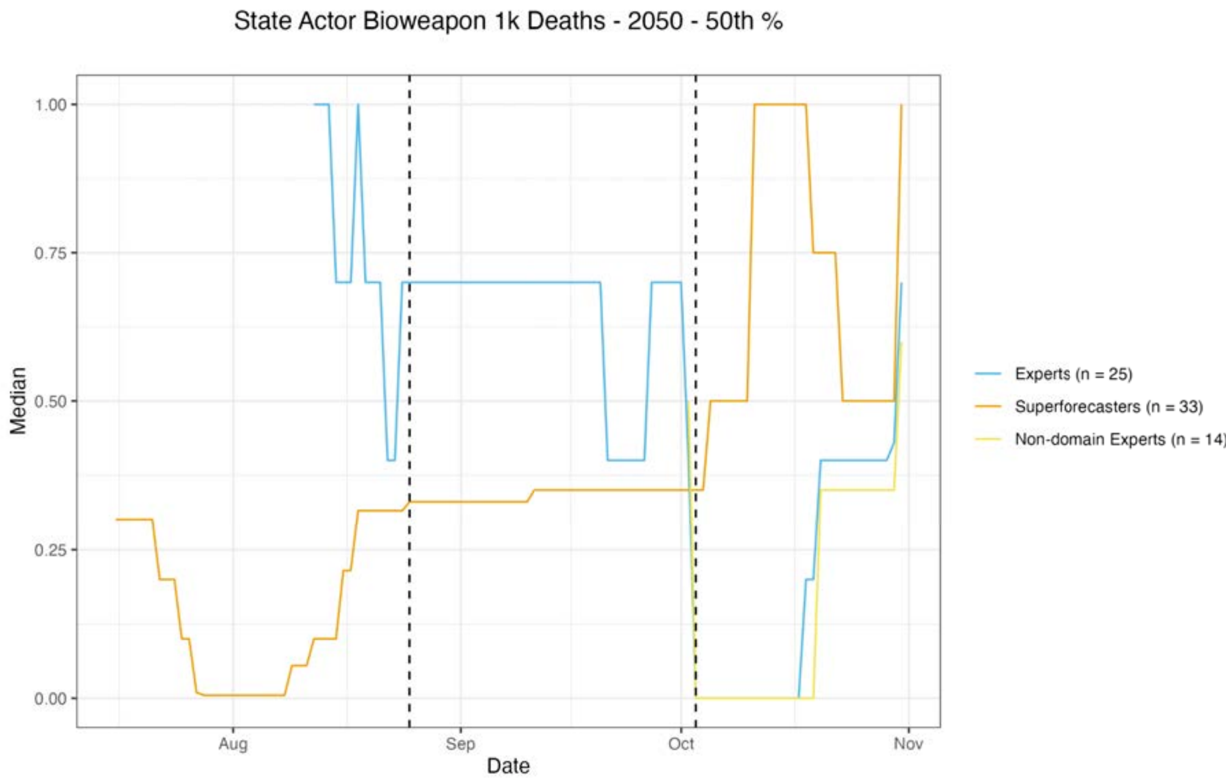
[586] Because only one domain expert provided a forecast by Stage 1, the SD was Stage 1 was zero.

State Actor Bioweapon 1k Deaths - 2024 - 50th %

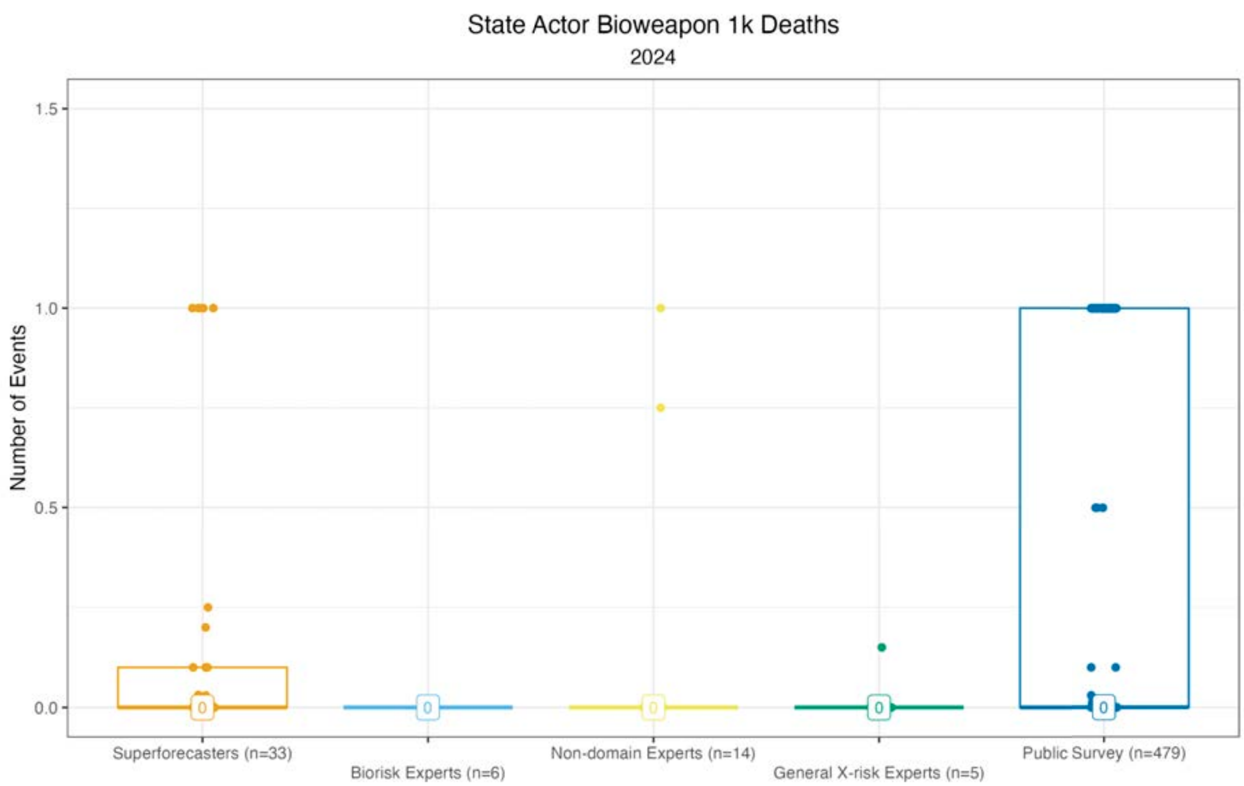


State Actor Bioweapon 1k Deaths - 2030 - 50th %



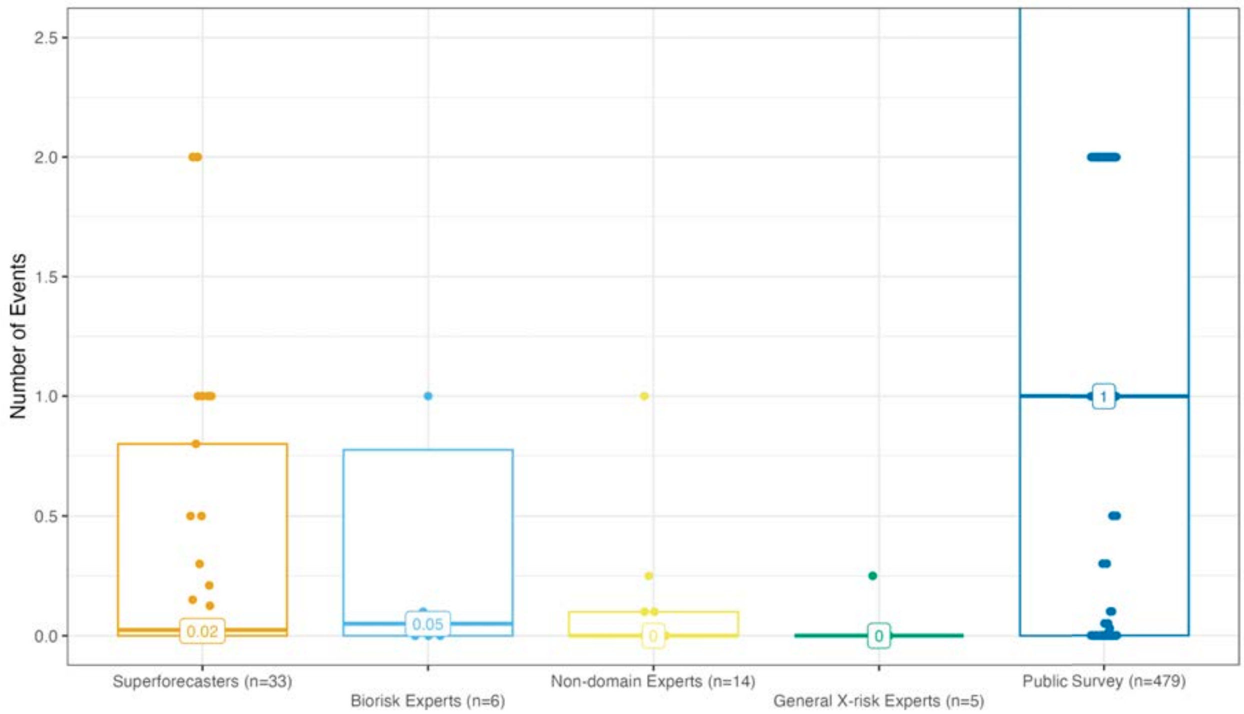


State Actor Bioweapon 1k Deaths - 2050 - 50th %
Median
1.00
0.75
0.50
0.25
0.00
Aug
Sep
Oct
Nov
Date
Experts (n = 25)
Superforecasters (n = 33)
Non-domain Experts (n = 14)

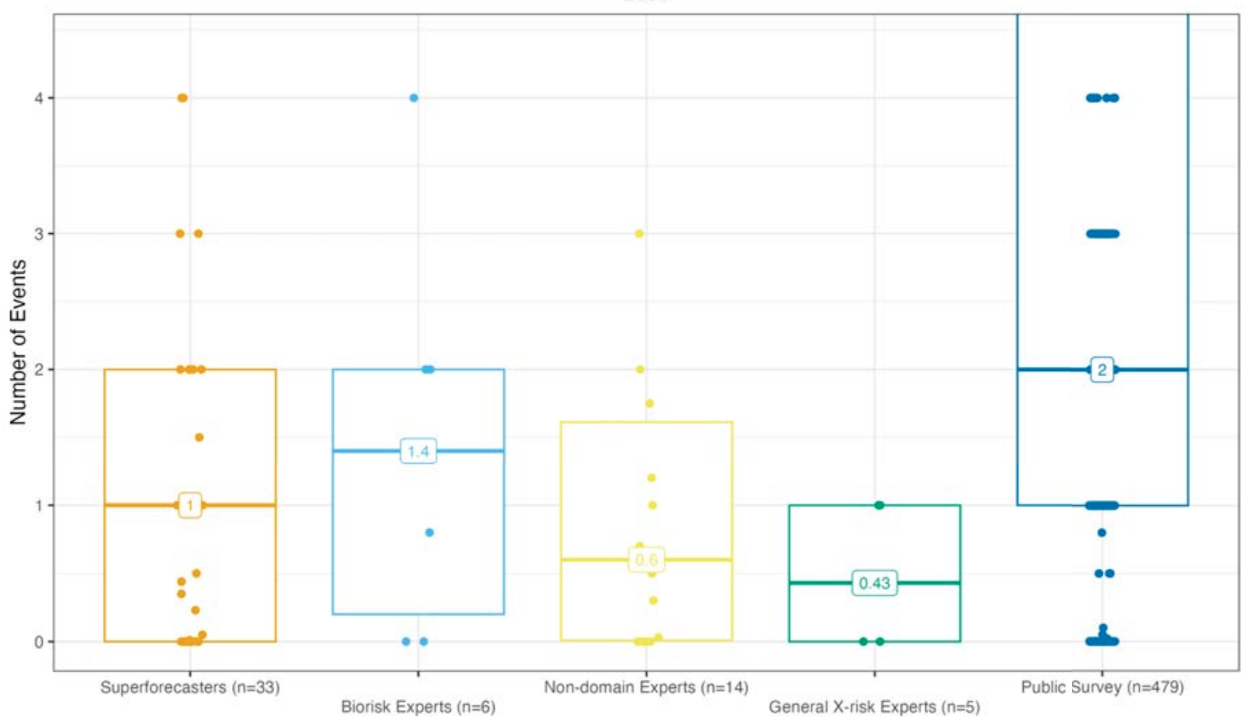


State Actor Bioweapon 1k Deaths
2024
Number of Events
1.5
1.0
0.5
0.0
Superforecasters (n=33)
Biorisk Experts (n=6)
Non-domain Experts (n=14)
General X-risk Experts (n=5)
Public Survey (n=479)

State Actor Bioweapon 1k Deaths
2030



State Actor Bioweapon 1k Deaths
2050



Sources of agreement, disagreement and uncertainty

There was widespread agreement that technology that could be used to create bioweapons has advanced significantly in recent years.

There were no explicit points of disagreement, merely different emphases.

Arguments given for forecasts ≤ 0 (2024), 0.05 (2030), 0.32 (2050)

On historical precedents

- The historical base rate of state actors using contagious biological weapons causing 1000+ deaths is low, with 0-2 events per century.[587] Team 338 mentions four particularly well-evidenced events since 1348 that meet the conditions of the question (more than 1000 deaths by contagious agent, intentionally implemented by a state actor):
    a. The Siege of Caffa, 1348
    b. The deliberate spread of smallpox among Pawnee tribes by the British, 1831
    c. The deliberate spread of smallpox among Aboriginal Australians by British settlers, 1873
    d. The Japanese use of plague in Quzhou, 1940

Disadvantages of contagious bioweapons

- Contagious bioweapons are prone to mutation.[588]
- Contagious bioweapons’ spread is difficult to control.[589] [590]
- They have high potential for alienating allies and causing a global backlash.[591]
- They are expensive to develop relative to their usefulness.[592]
- Overall, the costs of using biological weapons would likely outweigh the benefits for state actors.[593]

Other options are more practical

- Non-contagious alternatives like anthrax are more practical.[594]
- More generally, state actors have many other lethal options that are more easily targeted, making it less likely they would resort to biological weapons.[595]

---

[587] 338, "Depending on which dates and timeframes one uses, this data suggests a historical base rate of around 0-2 events per century."
[588] 343, "Contagious bioweapons are prone to mutation once used. As such, their appeal alongside other weapons is not obvious."
[589] 338, "The risk of uncontrolled spread is greater now (and over the next 100 years) than it has been at any other point in history, due to increased connectivity."
[590] 341, "Contagious bioweapons may be challenging to control once used."
[591] 343, "Their potential for alienating allies and the broader global community is high"
[592] 343, “Their development costs are disproportionate to their potential benefits.”
[593] 337, "In the end, the consensus was that for state actors in the modern era, the costs of using biological weapons would likely outweigh the benefits."
[594] 341, "Contagious bioweapons don't seem very useful and seem extremely risky to the aggressor, it seems like non contagious ones like anthrax would be better."
[595] 337, "With so many other ways to kill people—bullets, poisons, nukes, and so on—that are more easily aimed at the intended target, I'm struggling to see the incentive for a state to resort to biological weapons on a mass scale."

Conventions against the use of bioweapons reduce risk

- The Biological Weapons Convention has been agreed upon by many countries, reducing the number of countries with biological weapon stockpiles.[596] [597]

Factors likely to reduce deaths even in the event a bioweapon is used

- Medical advancements make deaths by bioweapons similar to historically destructive pathogens less likely.[598]
- Even if used, it would be difficult for an attack to meet the 1,000 deaths threshold.[599]

Technicalities regarding question resolution

- States' likelihood of denying responsibility for such an attack decreases the probability that this question will be resolved.[600]

Arguments given for forecasts ≥ 0.1 (2024), 0.1 (2030), 1.25 (2050)

Technological advancements make bioweapons more likely to be used

- Technology for manipulating contagious agents is becoming cheaper and more powerful.[601] [602]

Existing bioweapons programs:

- There are already seventeen countries that have had or are suspected of having a biological weapons program.[603]
- Risk increases in the future due to more states having access, more desperate dictators, or a disregard for international law caused by a chaotic world.[604]

---

[596] 338, "The Biological Weapons Convention (BWC) which effectively bans biological and toxin weapons has been agreed to by 184 countries."
[597] 341, "There is a tremendous amount of taboo against contagious bio-weapons. States in principle are agreed to the Biological Weapons Convention."
[598] 338, "Medical science and public health are significantly more advanced now than it has been throughout history, meaning that pathogens like plague and smallpox which have historically been very destructive pose a much smaller risk now than they have in the past."
[599] 343, "States will most likely fail to meet the 1,000 deaths threshold, while still incurring global backlash for use of contagious bioweapons. Take, for example, a bioweapon that kills 10% of its hosts. It would need to infect 10,000 people to reach the prompt threshold. Meanwhile, an agent with a much higher kill rate, say 60%, would struggle to reach a replication rate high enough to reach the prompt's threshold of 1,000 deaths as it killed its host off too fast to spread to a new host. This might be 'preferred' by the attacking nation, which would want to limit exposure (though with a much higher risk in a loss of control scenario), but would also require much more sophisticated deployment methods."
[600] 341, "Responsibility for any attack will have to be determined as a state actor who will probably deny responsibility."
[601] 336, "Technology for manipulating contagious agents appears to be getting cheaper and more powerful."
[602] 344, "Lowering prices" in an "upper-end forecast" section.
[603] 336, "Seventeen countries have had or are suspected of currently having a biological weapons programme."
[604] 336, "For the 20 years after [the Ukraine conflict] the risk increases significantly, because: more states might have access to this kind of stuff (it´s easier to get than nuclear weapons, and technical progress might make it even easier to get) or more dictators might be as desperate or aggressive as Assad or

Less powerful nations and asymmetric warfare

- Less powerful nations who have less to lose from global retaliatory action may find it more rational to use bioweapons.[605]
- Asymmetric warfare situations, such as the Russia/Ukraine conflict or potential China/Taiwan conflicts, increase the likelihood of bioweapons use.[606]

Cascade Effects:

- Cascade (or copycat) effect: one use of biological weapons might increase the chances of more being used.[607]

Other arguments given

Arguments in favor of lower forecasts:

- Past and present experience suggests bioweapons aren't used even in worst-case scenarios, like Ukraine or Afghanistan.[608]
- Covid has made bioweapons less attractive.[609]

Arguments in favor of higher forecasts:

Bioweapons development increasing likelihood of their use

- Specific DNA profiles could be targeted, reducing the risk of self-infection and increasing the likelihood of deployment.[610]

Potential scenarios for state actors' use

Saddam Hussein were or the world might be so unstable then, that nobody cares about international law anymore."

[605] 345, "Team thinks more powerful nations don't have as much of an incentive to break the taboo of using bio weapons and risk massive retaliatory action from the global community. But countries like Iran and N Korea may come to a point where the asymmetric advantage it gives them is worth the risk."

[606] 342, "Luckily the number of occurrences should be relatively small, but could be a warring alternative in asymmetric warfare such as the Ukrainian/Russia or the possible China/Taiwan conflicts or future unknown wars, he would guess about 0.2 per year on average, as this seems much more likely that a state actor would have the means to carry this event."

[607] 342, "[A forecaster] said his numbers slightly higher in the out years for the 75/95% ones because of possible cascade effects (if one is used, increases chances that more will be used)."

[608] 342, "Hence the lower end plausible data set distributions showing a zero level of confidence of the probability ranges for such an occurrence ever happening irrespective of worse case scenarios like we are witnessing in Ukraine or the non-state actors overthrowing legitimate govt in Afghanistan."

[609] 345, "Very less chance for such an incident by year 2024 (< 2 years left), especially with 'Covid', everyone is 'afraid' of such bio/pathogen related issues."

[610] 337, "[T]he potential for a state actor to design a weapon that would target a specific group of people based on specific DNA profiles."

- Arms race: State actors may engage in research and arms race dynamics, leading to the development of more potent biological weapons.[611]
- State actors may use bioweapons in an all-out war or during testing scenarios.[612]
- Bioweapons still may be used in the Ukraine/Russia conflict.[613]
- If a state uses biological weapons in desperation, they might use them in multiple locations at once.[614]

Cross-references with other questions

Q15: Non-State Actor Bioweapon 1k Deaths

Q17: Non-State Actor Bioweapon 100k Deaths

Q18: State Actor Bioweapon 100k Deaths

Question 17: Non-State Actor Bioweapon 100k Deaths

How many times will a non-state actor use biological weapons that involve a contagious agent to kill at least 100,000 people…
…by the end of 2024?
…by the end of 2030?
…by the end of 2050?

Question and resolution details, prior forecasts, and other relevant sources

Results[615]

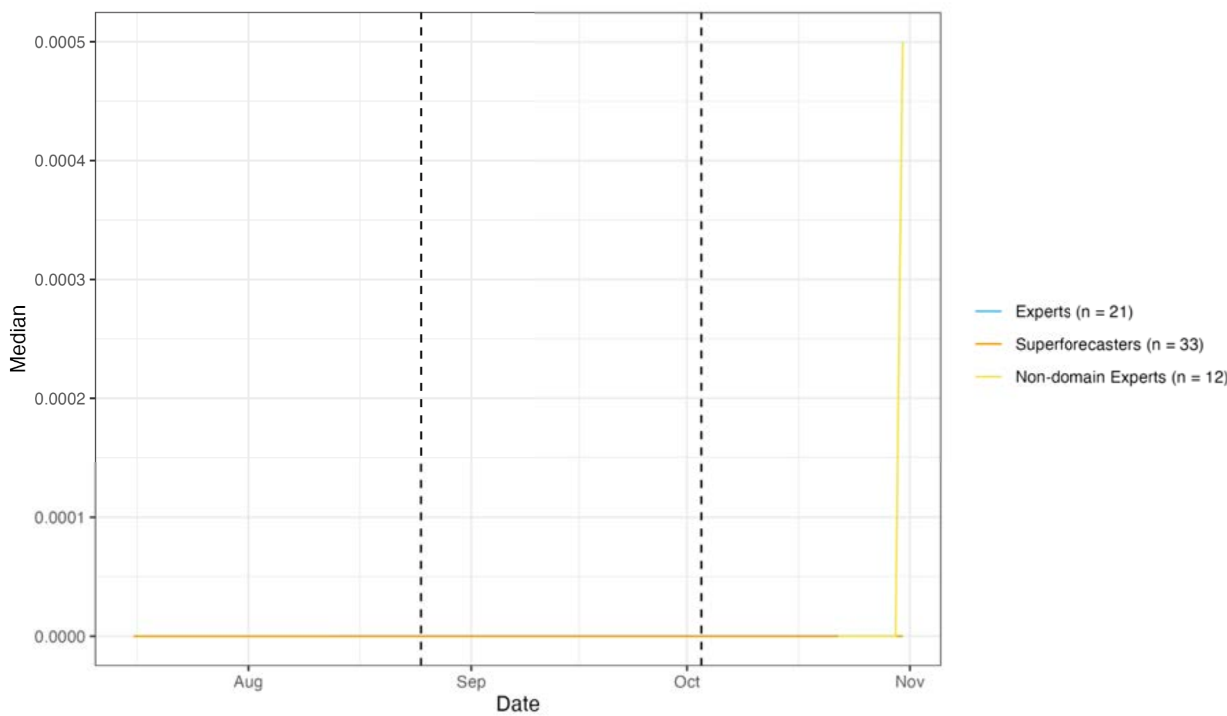
Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4

[611] 338, "[State Actors are] likely to engage in research and be susceptible to arms race dynamics, causing them to pursue engineered biological weapons that could have catastrophic effects."
[612] 343, "There are at least two scenarios in which state actors may decide to use bioweapons: (1) in an all-out war where all options are considered; and (2) in testing scenarios, where test subjects number into the thousands or where control is lost."
[613] 341, "The Russia Ukraine war may lead to more desperate actions."
[614] 338, "If a state is desperate or entrenched enough to use bioweapons and has sufficient stockpiles, I don't see why they wouldn't use them in multiple locations at once."
[615] Numbers of forecasters are given as of Stage 4 of the XPT.

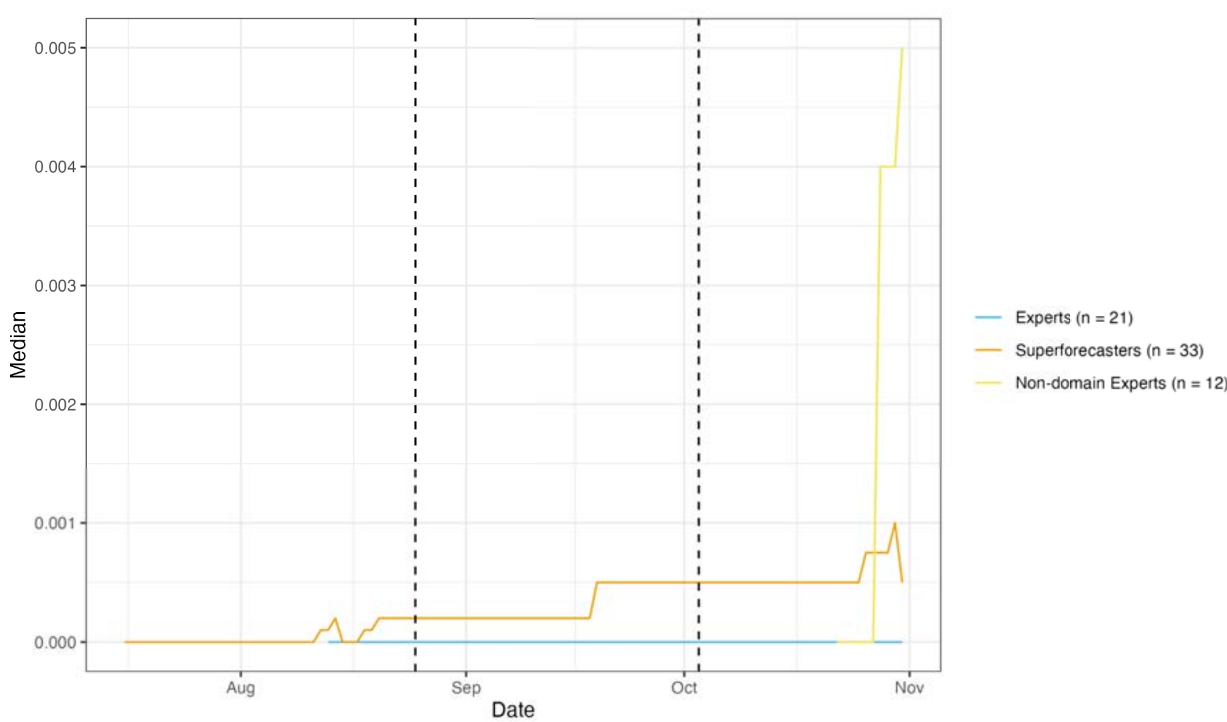
Super-Forecasters (N = 33)	2024	0	0	0.013	+543.28%
	2030	0	0.0005	0.073	+306.14%
	2050	0	0.038	0.39	+158.51%
Domain Experts (N = 6)[616]	2024	0	0	0	Inf
	2030	0	0	0.00	Inf
	2050	0	0	0.00	Inf
General X-Risk Experts (N = 3)	2024	NA	0.006	NA	NA
	2030	NA	0.008	NA	NA
	2050	NA	0.09	NA	NA
Non-Domain Experts (N = 12)	2024	0	0.0005	0.35	-83.56%
	2030	1.5	0.005	2.57	-44.16%
	2050	3	0.03	56.89	-49.31%

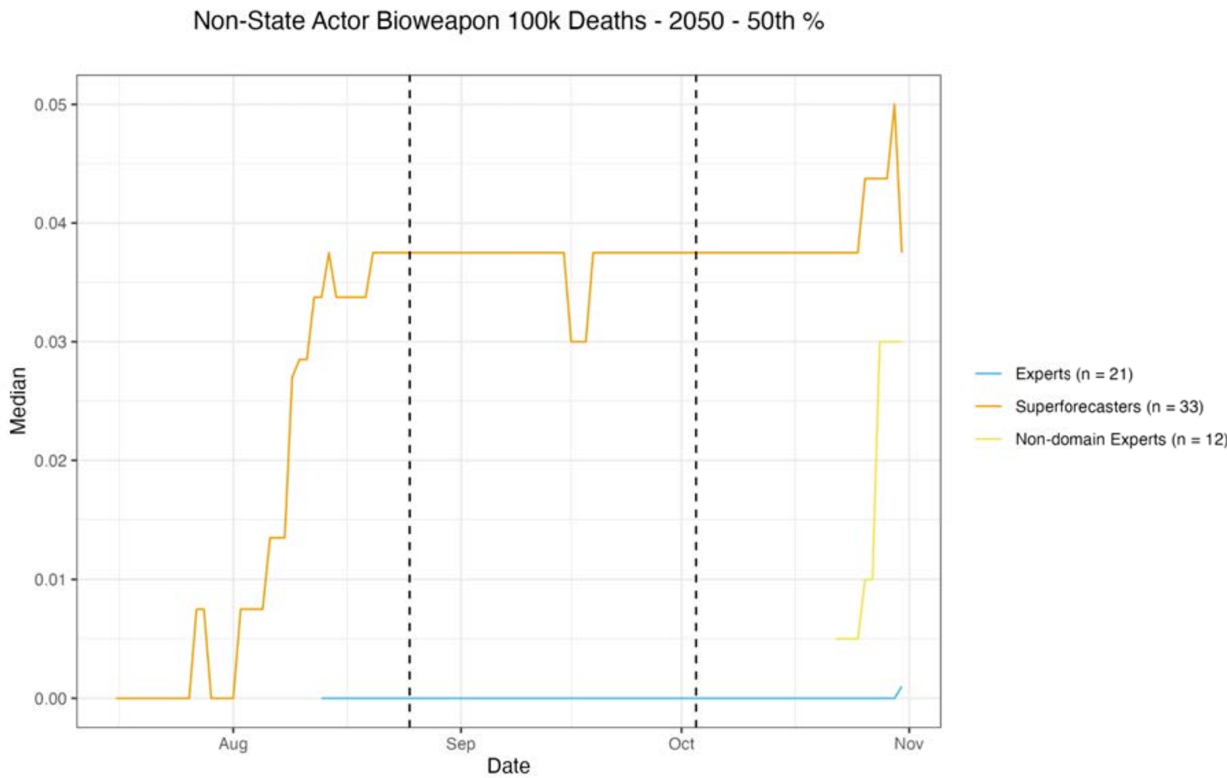
[616] Because only one domain expert provided a forecast by Stage 1, the SD was Stage 1 was zero.

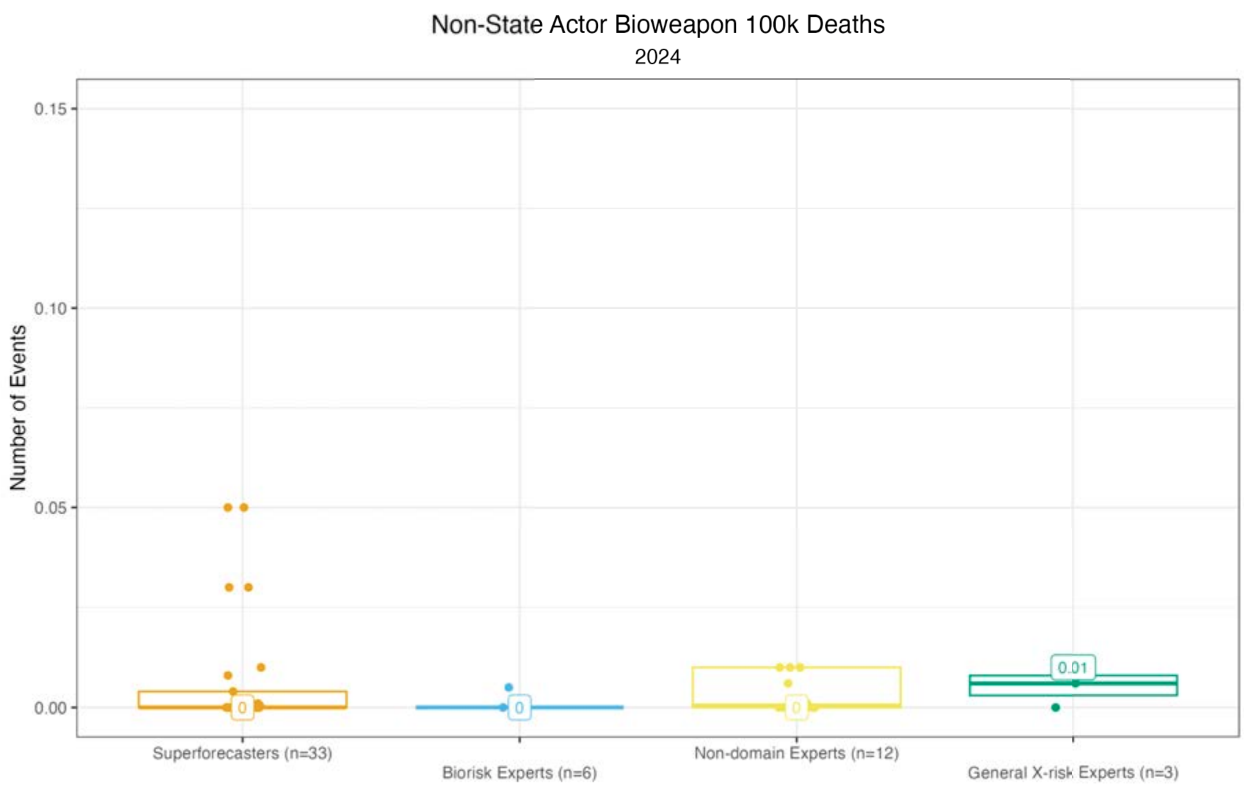
Non-State Actor Bioweapon 100k Deaths - 2024 - 50th %

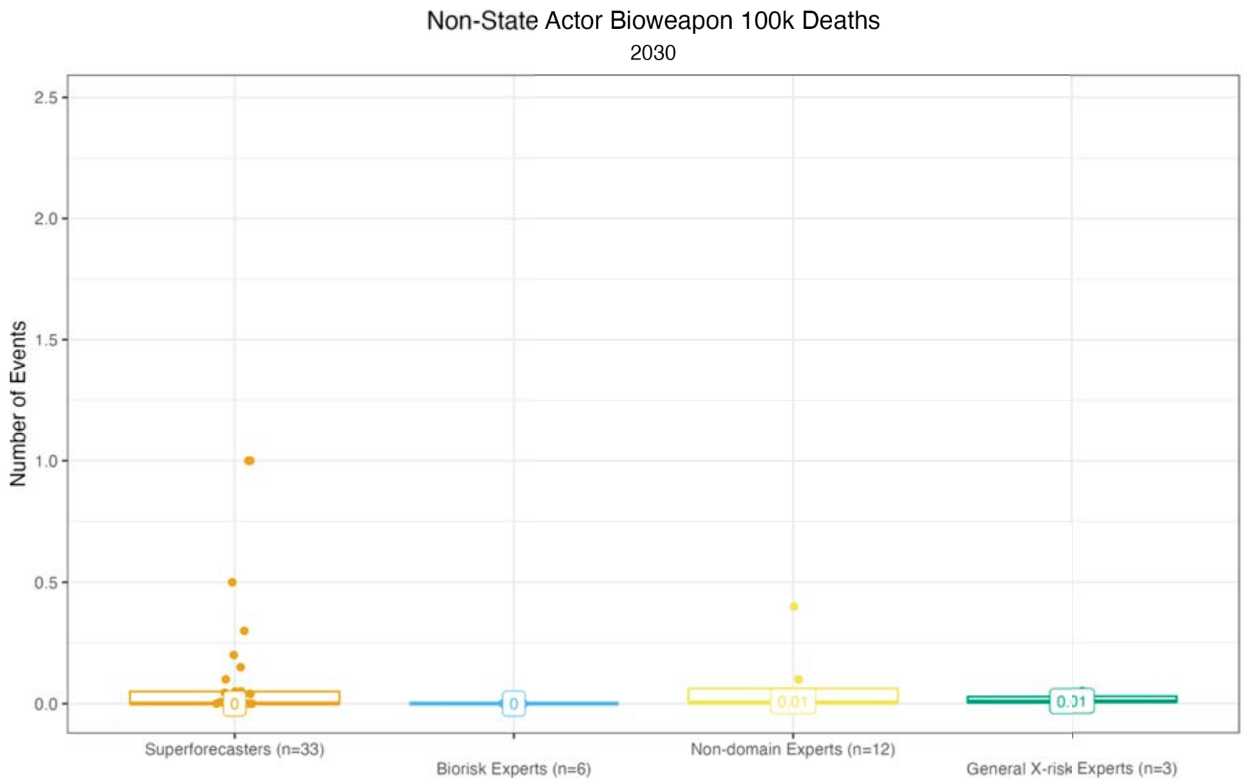


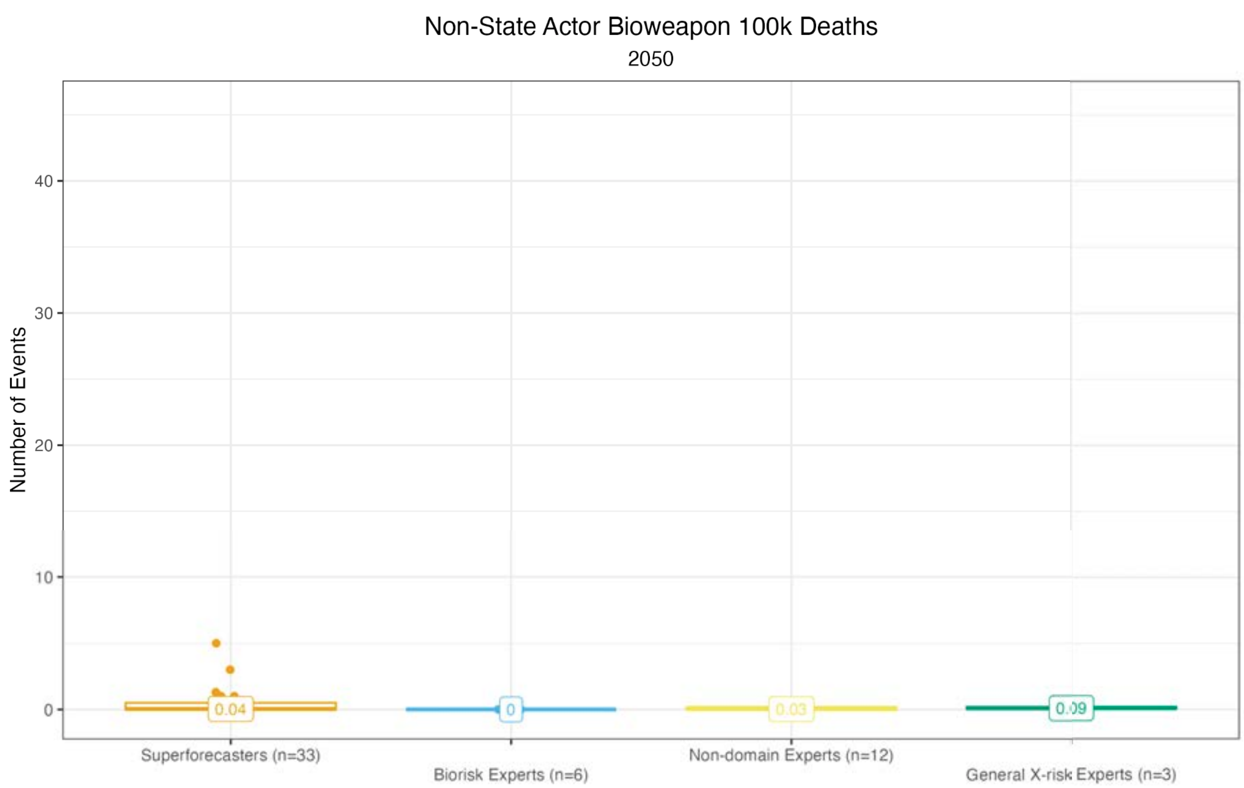
Non-State Actor Bioweapon 100k Deaths - 2030 - 50th %




Non-State Actor Bioweapon 100k Deaths - 2050 - 50th %
Median
0.05
0.04
0.03
0.02
0.01
0.00
Aug
Sep
Oct
Nov
Date
Experts (n = 21)
Superforecasters (n = 33)
Non-domain Experts (n = 12)


Non-State Actor Bioweapon 100k Deaths
2024
Number of Events
0.15
0.10
0.05
0.00
0
0
0
0.01
Superforecasters (n=33)
Biorisk Experts (n=6)
Non-domain Experts (n=12)
General X-risk Experts (n=3)


Non-State Actor Bioweapon 100k Deaths
2030
Number of Events
2.5
2.0
1.5
1.0
0.5
0.0
0
0
0.01
0.01
Superforecasters (n=33)
Biorisk Experts (n=6)
Non-domain Experts (n=12)
General X-risk Experts (n=3)


Non-State Actor Bioweapon 100k Deaths
2050
Number of Events
40
30
20
10
0
0.04
0
0.03
0.09
Superforecasters (n=33)
Biorisk Experts (n=6)
Non-domain Experts (n=12)
General X-risk Experts (n=3)

Sources of agreement, disagreement and uncertainty

There was agreement that technology associated with bioweapons will improve and become more widely accessible.

There was disagreement on:

1. Whether historical precedent implies a high or low forecast.
2. The degree to which non-state actors will be capable of a bioweapon attack that kills 100,000 people.
3. The intrinsic likelihood of a bioweapon causing 100k deaths.
4. Non-state actors' expected inclination to use bioweapons.

Arguments given for forecasts ≤ 0 (2024), 0 (2030), 0 (2050)

On 1 (precedents):

- Bioterrorism has caused few deaths historically,[617] bioterrorist attacks by non-state actors are rare and have caused few casualties,[618] and there is no historical precedent for large-scale biological attacks by non-state actors.[619]

On 2 (non-state actors' capabilities):

- Most non-state actors haven't shown significant bioweapons capabilities.[620]
- Non-state actors are unlikely to have the necessary infrastructure and capabilities for a contagious bioweapon agent, and will choose a different kind of weapon.[621]
- The technical knowledge, expertise, and funding required to develop a highly lethal bioweapon is out of reach for most non-state actors.[622]
- Non-state actors will not develop the sophistication needed to deploy bioweapons within the given timeframe.[623]

On 3 (likelihood of a bioweapon causing 100k deaths):

- The 100k death threshold makes most potential bioweapons agents impractical.[624]

On 4 (non-state actors' inclination to use bioweapons):

[617] 343, "Since 1950, there have only been 10 instances of bioterror with 5 total deaths." [Not elaborated upon.]
[618] 345, "Bioterrorist attacks by non-state actors are rare and have caused few casualties."
[619] 338, "No qualifying events in Wikipedia list of bioterrorist incidents"; "None of the large events was a biological attack"; "There were no attacks with contagious agents."
[620] 340, "[M]ost prolific non-state actors still haven't shown a lot of capability in the bioweapons terrorism front yet."
[621] 340, "A non-state actor is extremely unlikely to have the infrastructure and capability needed, and biological weapons will not be the weapon of choice."
[622] 343, "The technical knowledge/expertise/funding required to develop the kind of agent that can reach a 100k threshold is still out of reach to most non-state actors."
[623] 345, "[T]here is a general feeling that non-state actors will not develop the sophistication needed to deploy these weapons within the timeframe of the question."
[624] 343, "The 100k death threshold means most potential bioweapons agents are not practical."

- The lack of use of a nuclear weapon by a non-state actor since their advent is evidence that non-state actors have a low propensity to use WMDs.[625]

Practical obstacles to a successful attack:

- Contagious agents suffer from practical issues that decrease chances of a successful attack. “Issues cited include: targeting, control, lack of spectacle, required expertise, planning, timeline, and especially better options available.”[626]

Alternatives:

- Low-cost alternatives to biological attacks exist.[627]

International safeguards triggered by initial attempts

- Initial bioterrorism attempts may provoke the international community to implement safeguards, making future attempts more difficult.[628]

Arguments given for forecasts ≥ 0.0006 (2024), 0.0025 (2030), 0.08 (2050)

On 1 (precedents):

- There are historical examples of non-state actors using biological agents to harm populations.[629]

On 2 (non-state actors’ capabilities):

- Future technologies, such as CRISPR, could make it easier to develop highly contagious and deadly pathogens.[630] This makes bioweapons more accessible to terrorists who may deploy them with little concern for containment.[631]

On 3 (likelihood of a bioweapon causing 100k deaths):

- Contagion effects might lead to a large-scale impact even if the initial intent is not to cause 100,000 deaths.[632]
- Many existing pathogens, like COVID-19, are capable of killing over 100k people.[633]

---

[625] 343, “Nuclear analog: in 80 years, there has never been a non-state use of the weapon, suggesting that non-state actors have a low propensity to actually acquiring and using WMD."

[626] 338, "There was also consensus that contagious agents suffered from practical issues that decreased chances of a successful attack. Issues cited include: targeting, control, lack of spectacle, required expertise, planning, timeline, and especially better options available."

[627] 338, "Al Qaeda may have had the capability in early 2000s but there is a wide range of low cost alternatives"; "More likely for non-state actor than state actor, but still very low."

[628] 343, "Initial attempts would likely provoke the international community to implement safeguards to stop this kind of thing - regardless of whether the threshold of 100k deaths is reached."

[629] 344, "There are historical examples of the use of biological agents by non-state actors to 'hurt' populations in the past." [Not elaborated upon.]

[630] 336, "Future technologies (e.g. CRISPR) will likely become more widely available and facilitate the development of highly contagious and deadly pathogens."

[631] 339, "The strongest argument for the event is that synthetic biology is advancing rapidly and with its advance, costs are falling and techniques are spreading with availability of technology across the planet."

[632] 344, "Due to contagion effects, even if a terrorist group aims for a lower number of deaths, the ability of such a weapon to scale from 1k to 100k deaths seems at least highly plausible."

[633] 336, "Many existing pathogens (e.g. COVID 19) are capable of killing > 100k people."

On 4 (non-state actors' inclination to use bioweapons):

- Increased future polarization and global instability may lead to more terroristic groups willing to use bioweapons.[634]

Other arguments given

Arguments in favor of lower forecasts:

- Non-state actors currently do not favor biological weapons as a means of attack.[635]
- Non-state actors might not have the incentives to aim for such a high number of deaths.[636]
- Future technologies may facilitate the detection and prevention/control of bioweapons by authorities.[637]

Arguments in favor of higher forecasts:

- There are currently several ways to modify common viruses for increased lethality and immune evasion.[638]
- Copycat attackers may refine their techniques in response to a successful bioterrorism attack.[639]
- Many large population centers in developing nations are not well-protected against this type of attack.[640]
- Post-COVID, more non-state actors may be interested in using biological agents.[641]

Cross-references with other questions

Q15: Non-State Actor Bioweapon 1k Deaths

Q16: State Actor Bioweapon 1k Deaths

Q18: State Actor Bioweapon 100k Deaths

---

[634] 336, "Future polarization / global instability may lead to the existence of more terroristic groups willing and able to use this kind of bioweapon (which would likely kill many of its own members)."
[635] 336, "Non-state actors currently tend not to favor biological weapons as a means of attack."
[636] 344, "Incentives may not be aligned for even terrorist groups to aim for such a high number of deaths."
[637] 336, "Future technologies may facilitate the development and deployment of such a weapon, but also its detection and prevention/control from authorities."
[638] 340, "[T]here are several ways to modify common viruses to massively increase lethality while retaining high transmissibility and immune evasion."
[639] 343, "This copycat need not succeed in killing 100k people to evoke copycats who might refine their techniques."
[640] 343, "Many large population centers in developing nations are not hardened against this type of attack."
[641] 338, "Post-COVID, more non-state actors may be interested in this approach (fads in terrorism)."

Question 18: State Actor Bioweapon 100k Deaths

How many times will a state actor use biological weapons that involve a contagious agent to be the cause of death for at least 100,000 people…
…by the end of 2024?
…by the end of 2030?
…by the end of 2050?

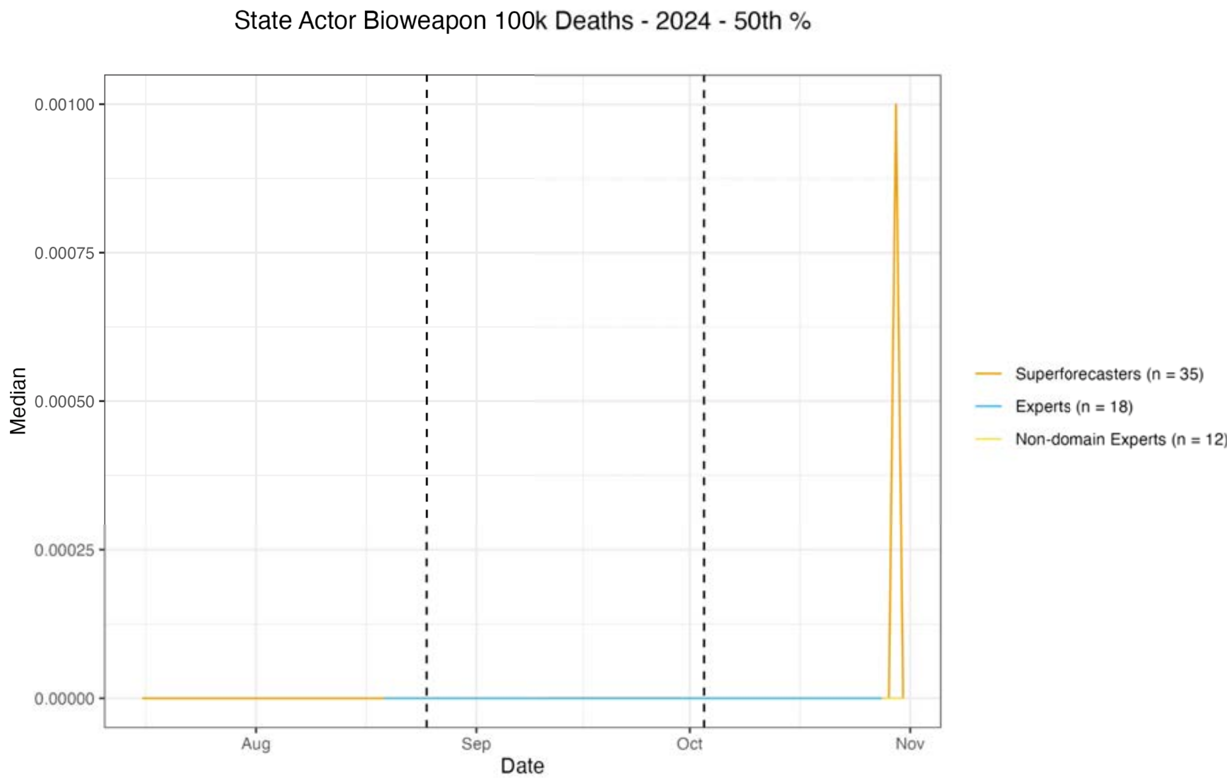
Question and resolution details, prior forecasts, and other relevant sources

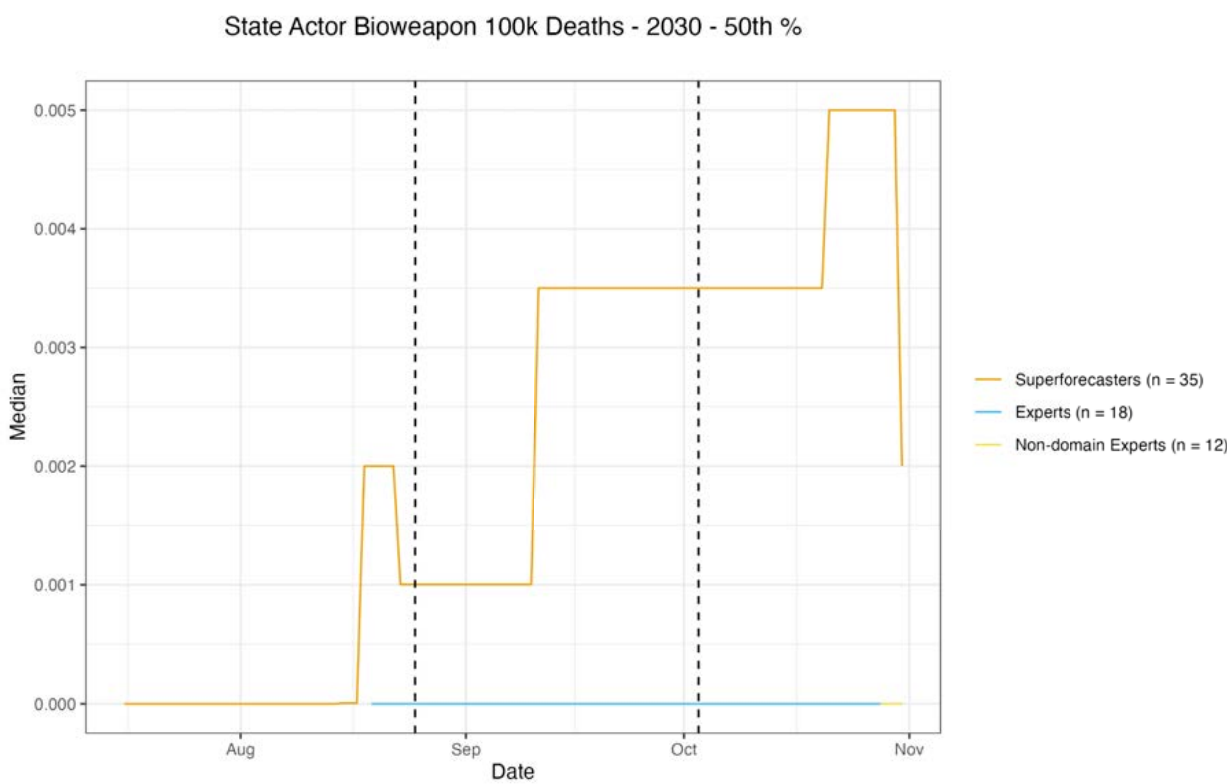
Results[642]

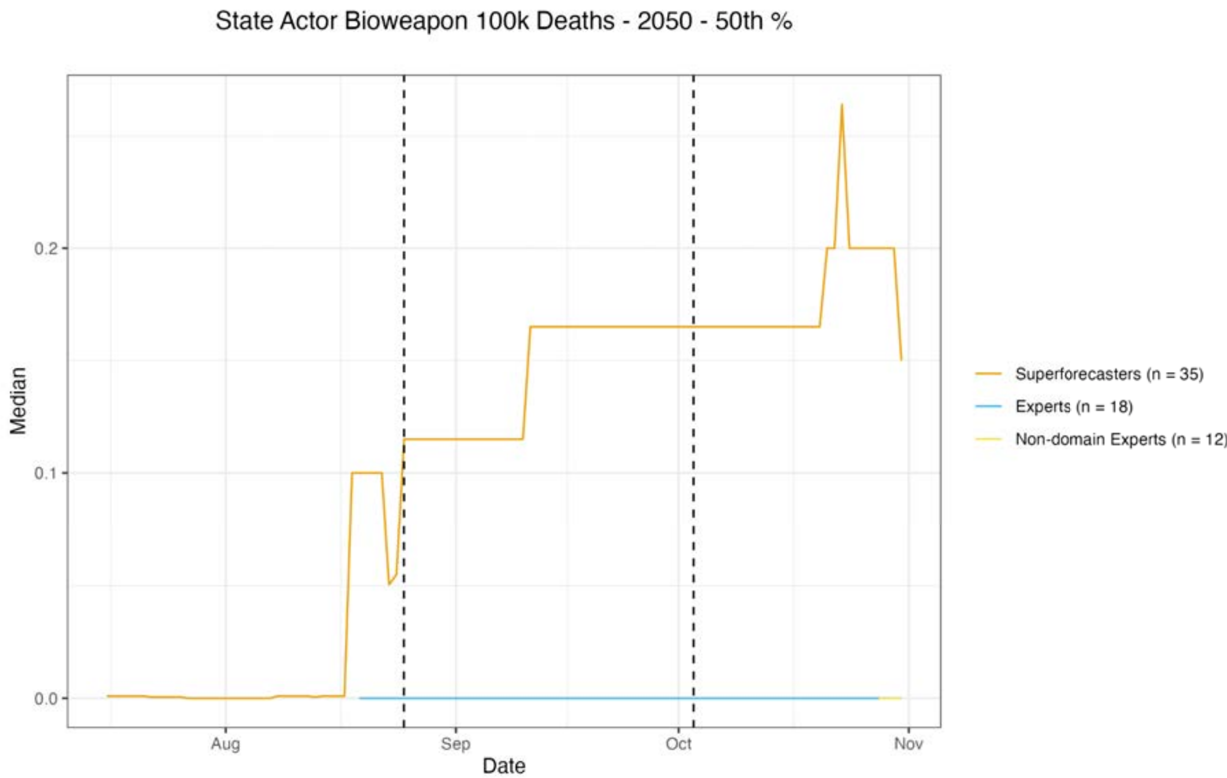
Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 35)	2024	0	0	0.97	-45.38%
	2030	0	0.002	1.93	-11.66%
	2050	0.001	0.15	3.84	-22.82%
Domain Experts (N = 4)	2024	0	0	0	Inf
	2030	0	0	0	Inf
	2050	0	0	0	Inf
	2024	n/a	0	n/a	n/a

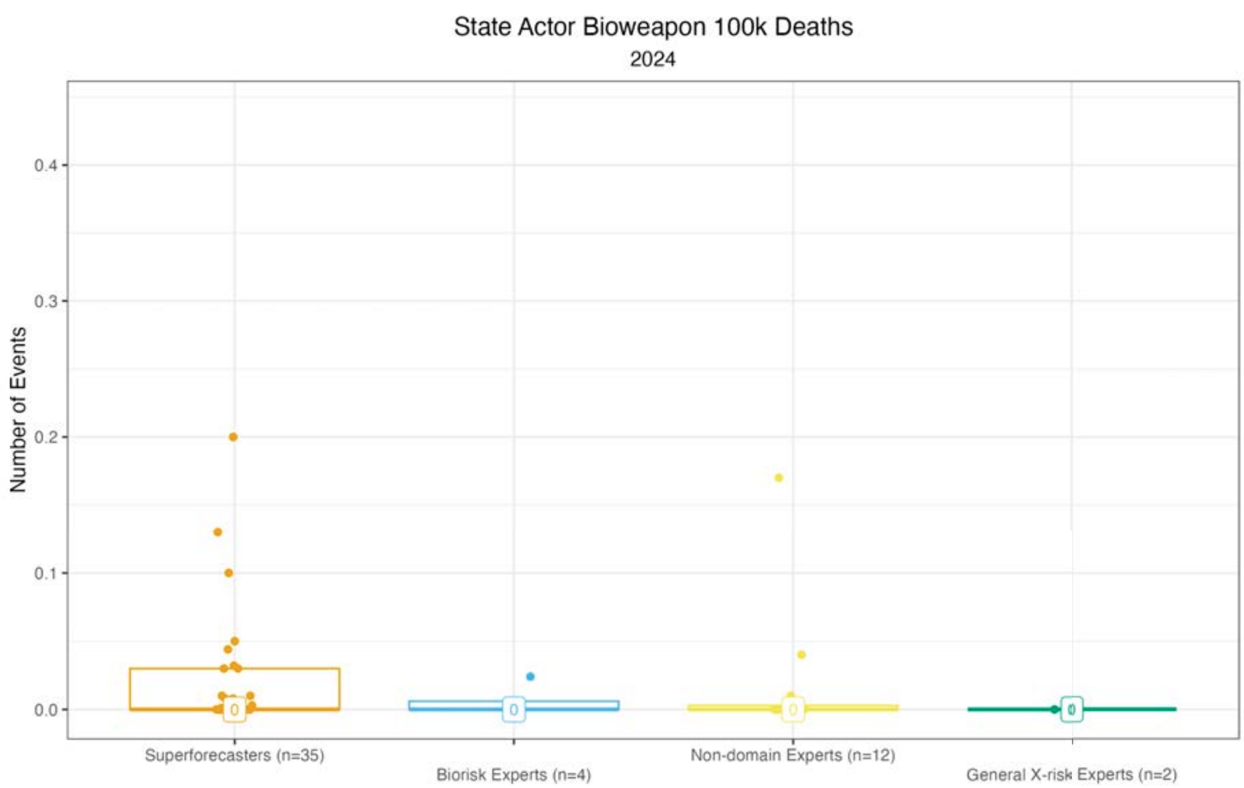
[642] Numbers of forecasters are given as of Stage 4 of the XPT.

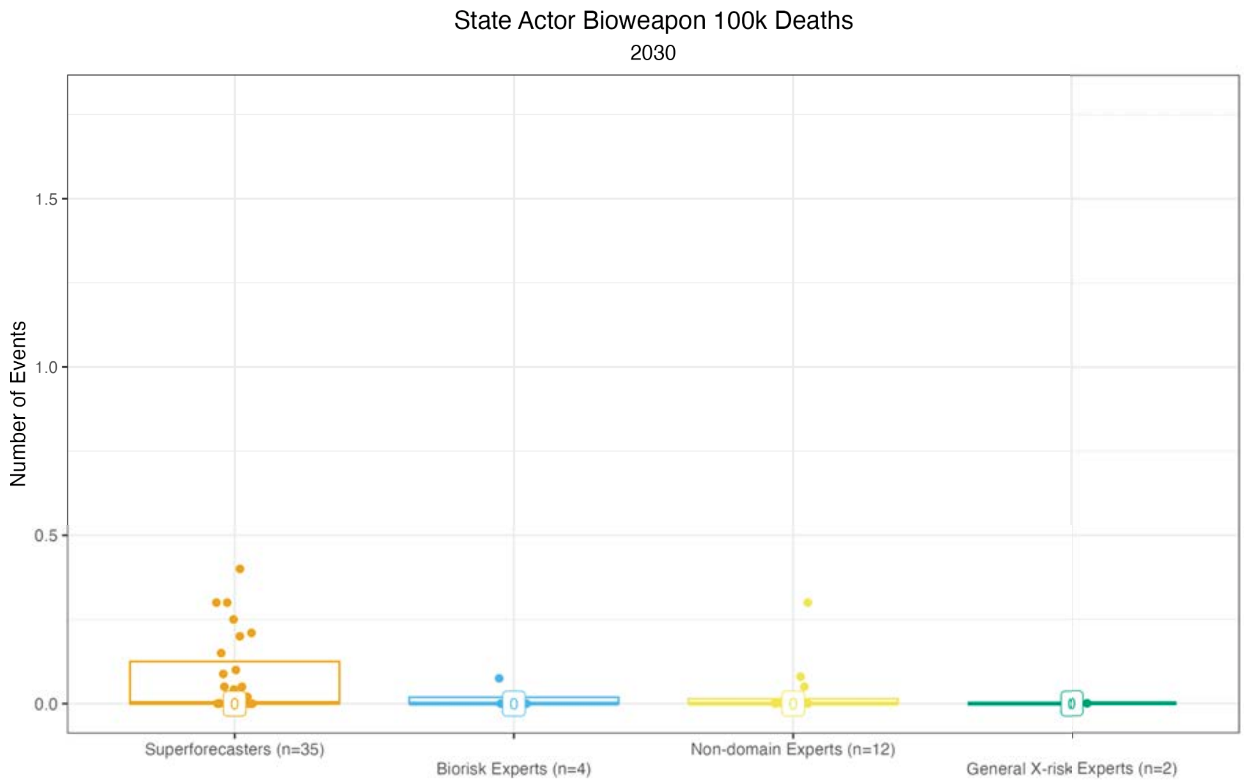
General X-Risk Experts (N = 2)	2030	n/a	0	n/a	n/a
	2050	n/a	0	n/a	n/a
Non-Domain Experts (N = 12)	2024	0	0	0.17	-71.64%
	2030	0	0	0.35	-74.9%
	2050	0	0	0.52	-71.54%

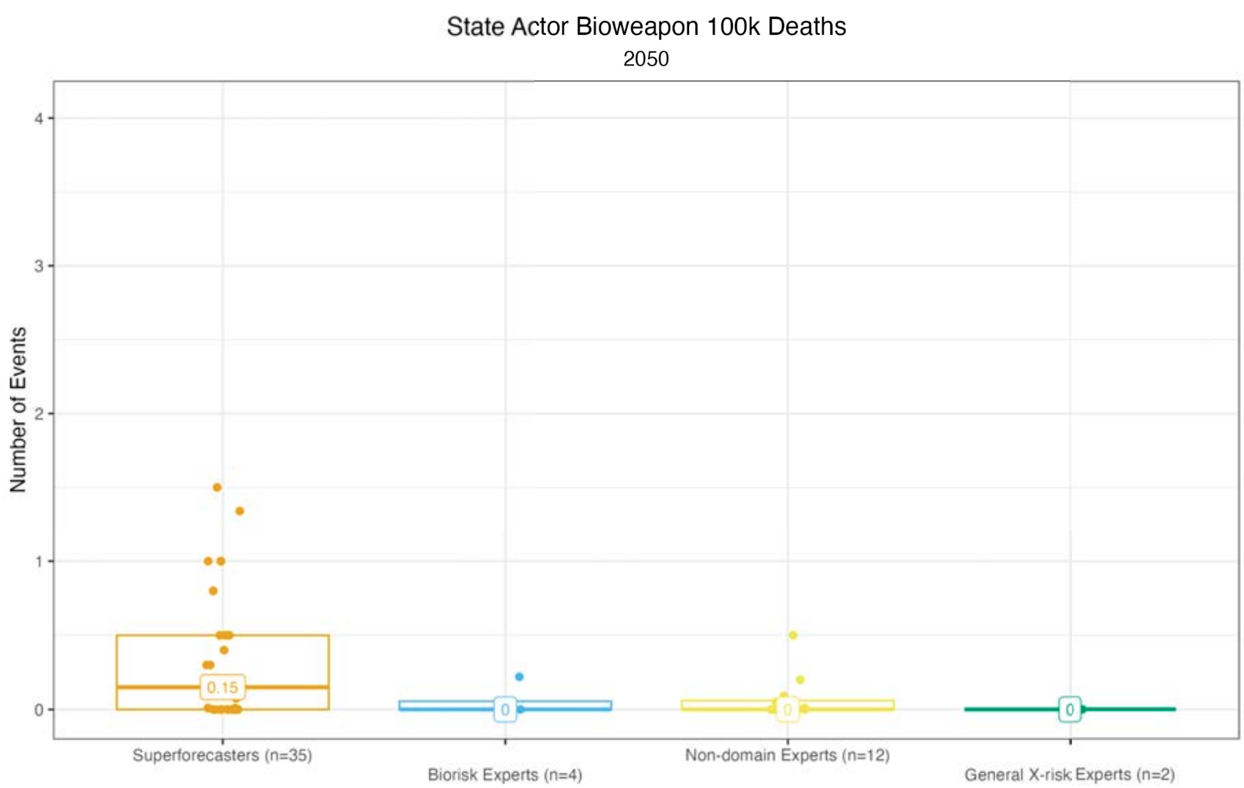

State Actor Bioweapon 100k Deaths - 2024 - 50th %
Median
0.00100
0.00075
0.00050
0.00025
0.00000
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 35)
Experts (n = 18)
Non-domain Experts (n = 12)


State Actor Bioweapon 100k Deaths - 2030 - 50th %
Median
0.005
0.004
0.003
0.002
0.001
0.000
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 35)
Experts (n = 18)
Non-domain Experts (n = 12)


State Actor Bioweapon 100k Deaths - 2050 - 50th %
Median
0.0
0.1
0.2
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 35)
Experts (n = 18)
Non-domain Experts (n = 12)


State Actor Bioweapon 100k Deaths
2024
Number of Events
0.0
0.1
0.2
0.3
0.4
Superforecasters (n=35)
Biorisk Experts (n=4)
Non-domain Experts (n=12)
General X-risk Experts (n=2)


State Actor Bioweapon 100k Deaths
2030
Number of Events
0.0
0.5
1.0
1.5
Superforecasters (n=35)
Biorisk Experts (n=4)
Non-domain Experts (n=12)
General X-risk Experts (n=2)


State Actor Bioweapon 100k Deaths
2050
Number of Events
0
1
2
3
4
0.15
Superforecasters (n=35)
Biorisk Experts (n=4)
Non-domain Experts (n=12)
General X-risk Experts (n=2)

Sources of agreement, disagreement and uncertainty

There was widespread agreement that contagious bioweapons are difficult to control and using them presents a challenge for state actors to avoid harming themselves.

There were no explicit disagreements, merely differences in emphases.

Arguments given for forecasts ≤ 0.008 (2024), 0.002 (2030), 0.01 (2050)

Reasons contagious biological weapons are a poor choice for state actors:

- It is difficult for state actors to use contagious biological weapons without harming themselves.[643]
- State actors cannot practically prepare to deploy such a weapon secretly, due to the infeasibility of vaccinating a state's population in advance without broadcasting intent.[644]
- Contagious bioweapons are impractical because they are difficult to control.[645]
- State actors may be deterred by the risk of massive retaliation.[646]
- This kind of weapon is poorly suited to the mutually assured destruction framework.[647]

Circumstances that would trigger resolution are extreme:

- The question is likely to resolve only in cases of extreme defensive action by a country facing an existential threat.[648]

Issues affecting resolution criteria:

- Covert actions are excluded since a state actor must be identified for this question to resolve.[649]
- Low probability due to high threshold (i.e. 100k deaths) and limited timeframe for the event to occur (i.e. by 2050).[650]

Arguments given for forecasts ≥ 1 (2024), 1 (2030), 1 (2050)

Capacity and tendencies of state actors

---

[643] 340, "All forecasts cited the difficulty in a state actor trying to use contagious biological weapons without greatly harming themselves."
[644] 340, "The possibility of vaccinating your own population in advance was considered but found to be infeasible and likely to broadcast intent."
[645] 341, "[C]ontagious bioweapons are extremely hard to control."
[646] 341, "Use would risk an massive retaliation."
[647] 340, "State actors not having much incentive to use bioweapons as they'd be unreliable for mutually assured destruction."
[648] 338, "[T]his question would only resolve in the case of 'extreme defensive action by a country defending against an existential threat'."
[649] 338, "[F]or this question to resolve a state actor needs to be identified and thus the possibility of covert action is excluded."
[650] 341, "2050 is only 28 years, the technology is new and 100k is a huge threshold."

- State actors currently may have the capacity, capabilities, and tendencies to use biological weapons that involve a contagious agent.[651]
- Power, authority, wealth, control, and resources at the disposal of heads of state might make them misuse such capabilities; power can distort rational thinking.[652]

Circumstances that would trigger resolution are plausible:

- Proliferation will increase the ability of state actors to use contagious bioweapons, and this fact combined with an intentional deployment spiraling out of control could cause the question to resolve.[653]

Other arguments given

Arguments in favor of lower forecasts:

Base rate

- Large-scale casualties from biological weapons are uncommon historically.[654]

Countermeasures

- Protective intervention capabilities exist to prevent catastrophic outcomes.[655]
- Strict regulations, sanctions, and global conventions (e.g. the Biological Weapons Convention) prohibit the use of biological weapons, discouraging such occurrences.[656] [657]

More reasons contagious biological weapons are a poor choice for state actors

- COVID-19 has illustrated the undesirability of using bioweapons because of their potential to backfire (because they affect every country), and COVID-19 has also

[651] 342, "The possible higher end of the plausible range [...] believe [...] that these state actors has the capacity, capabilities and tendencies to spring surprises through the usage of biological weapons that involve a contagious agent."
[652] 342, "Which has to do largely with position, Power, authority, wealth, control and resources at the disposal of these state actors which could Possibly make them to be intoxicated and want to misbehave by carrying out such a heinous act [...] the tendency is always there because power corrupt and absolute power corrupt absolutely."
[653] 343, "The strongest arguments for a forecast on the higher end of the plausible range of forecasts for this question begin with on-going proliferation plus a tail event: 'something intentional getting out of hand and having a much larger than expected outcome.'"
[654] 342, "Such case or occurrence is not common base on records even few ones that happened in time past have fewer human casualty in number."
[655] 342, "[T]he protection policy around these contagious agents are extremely strict to guide against such envisage situations because of its catastrophic implications, in addition to quick intervention mechanism should such a situation mistakenly happens."
[656] 342, "Hence the lower end plausible data set distributions due to global stringent conventions by United Nations and other world governing bodies and nations government prohibiting, sanctioning and strictly discouraging such an occurrence from happening in any form or guise."
[657] 336, "The Biological Weapons Convention was signed in the 1970s, in which states promised to not stockpile biological weapons for offensive purposes."

illustrated “significant vaccine hesitancy,” which would likely make limiting damage to one’s own country more difficult.[658]
- State actors may avoid using biological weapons due to potential prestige loss and negative global perception.[659]
- Contagious bioweapons are expensive, poorly targeted, and difficult to control.[660]
- The weapon itself serves as evidence for attribution, becoming more available for analysis as it spreads, which may discourage its use if the perpetrator doesn’t want to be ‘caught’.[661]

Arguments in favor of higher forecasts:

Technological advancements will increase likelihood of their deployment:
- Technological development may outpace social, legal, and moral constraints on the use of biological weapons.[662]
- Technological advancements make manipulating contagious agents cheaper and more powerful.[663]
- Increased access to bioweapons in the future due to technical progress.[664]
- Bioweapons targeting specific races or ancestries could increase the likelihood of state actors using them.[665]

State actors and rationality:
- State actors may not be rational, potentially not considering or caring about the consequences.[666]
- Analogy with nuclear weapons: Significant possibility of self-harm is not an absolute deterrent. The deployment of nuclear weapons is generally considered a non-zero

---

[658] 337, "COVID-19 gives a good example of a pathogen that has had impacts on all countries, making the state use of a bioweapon undesirable. It has also demonstrated significant vaccine hesitancy, and vaccines are one way of limiting damage on an attacking country."
[659] 344, "[P]restige loss and negative perception."
[660] 343, "Contagious bioweapons are expensive, poorly targeted, and difficult to control. It's difficult to see why any state actor would want to rely on them."
[661] 344, "[T]he weapon itself - which is effectively evidence for attribution - only makes itself more and more available for analysis as it spreads."
[662] 338, "In the longer term the pace of technological development of these weapons may exceed the social, legal and moral constraints on the use of the technology."
[663] 336, "Technology for manipulating contagious agents appears to be getting cheaper and more powerful."
[664] 336, "[M]ore states might have access to this kind of stuff (it´s easier to get than nuclear weapons, and technical progress might make it even easier to get)."
[665] 337, "Since there is research on targeting a specific person, I believe the consensus was that targeting race/common ancestry was plausible."
[666] 340, "A state actor doesn't have to be rational. They may not think through the consequences of such a bioweapon or not care about them."

possibility despite the likelihood that doing so would result in direct backlash and also the possibility of global nuclear winter.[667]

Cross-references with other questions

Q15: Non-State Actor Bioweapon 1k Deaths

Q16: State Actor Bioweapon 1k Deaths

Q17: Non-State Actor Bioweapon 100k Deaths

Question 19: Lab Leaks

What will be the expected number of events in which contagious biological agents that have escaped from labs are the cause of death for at least 1,000 people…
…by the end of 2024?
…by the end of 2030?
…by the end of 2050?

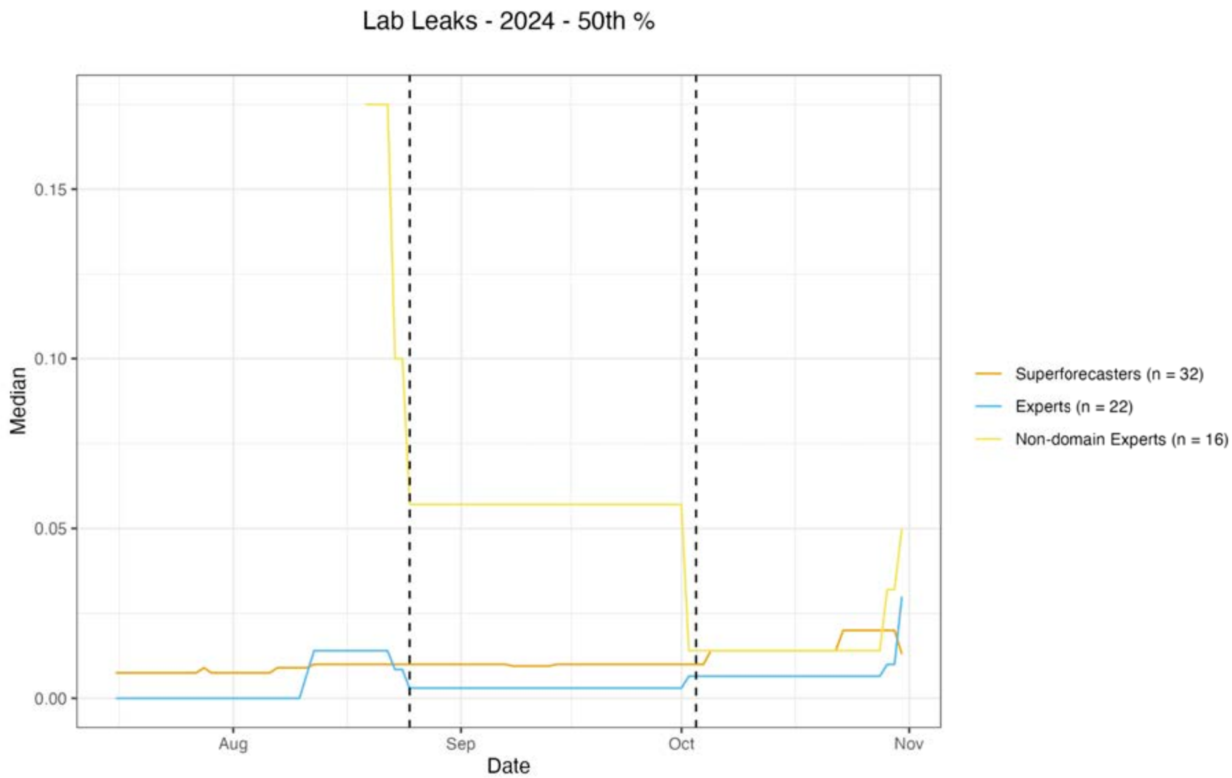
Question and resolution details, prior forecasts, and other relevant sources

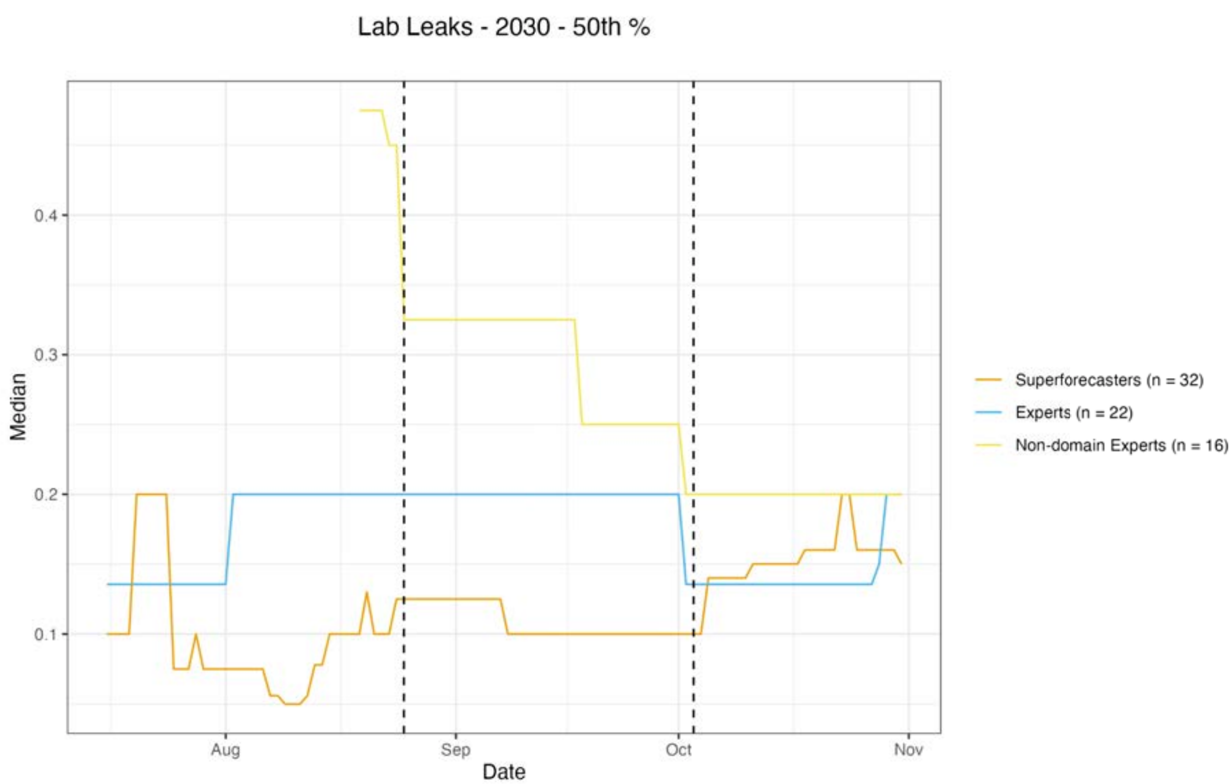
Results[668]

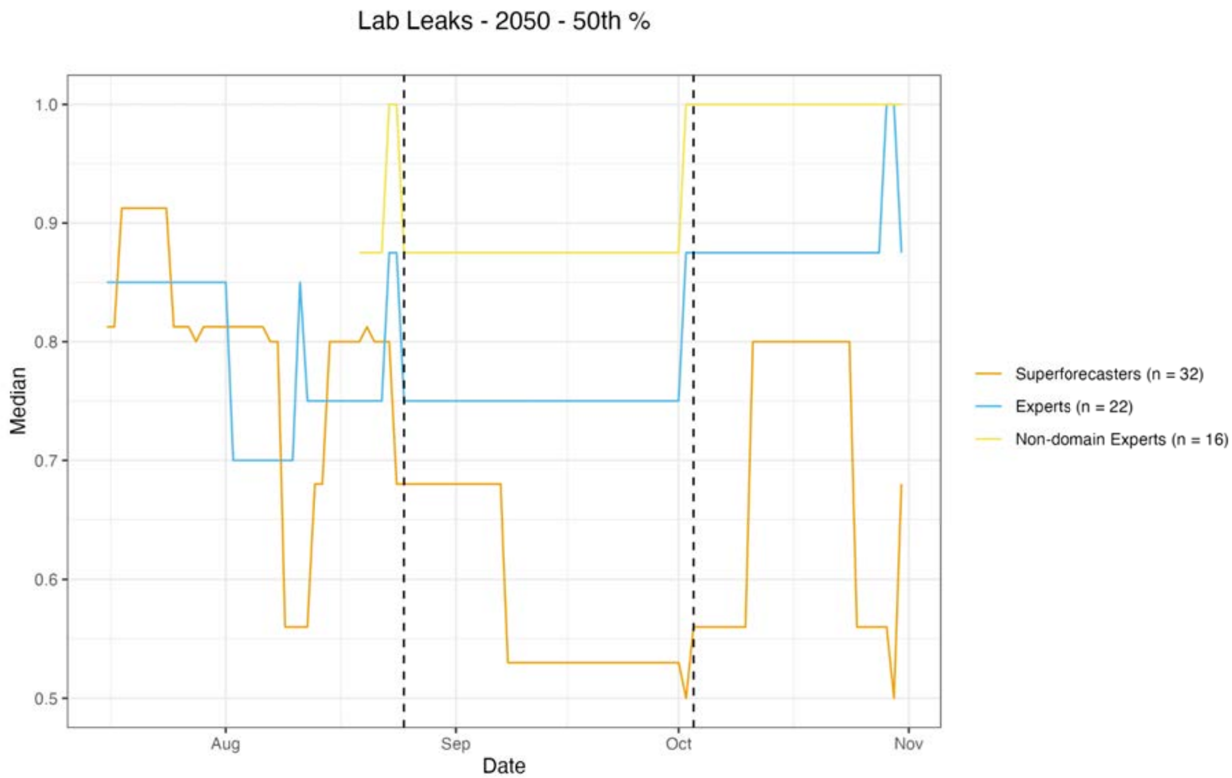
Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 32)	2024	0.0000004	0.013	2.31	-77.02%
	2030	0.058	0.15	3.71	-71.36%
	2050	0.83	0.68	5.73	-62.96%

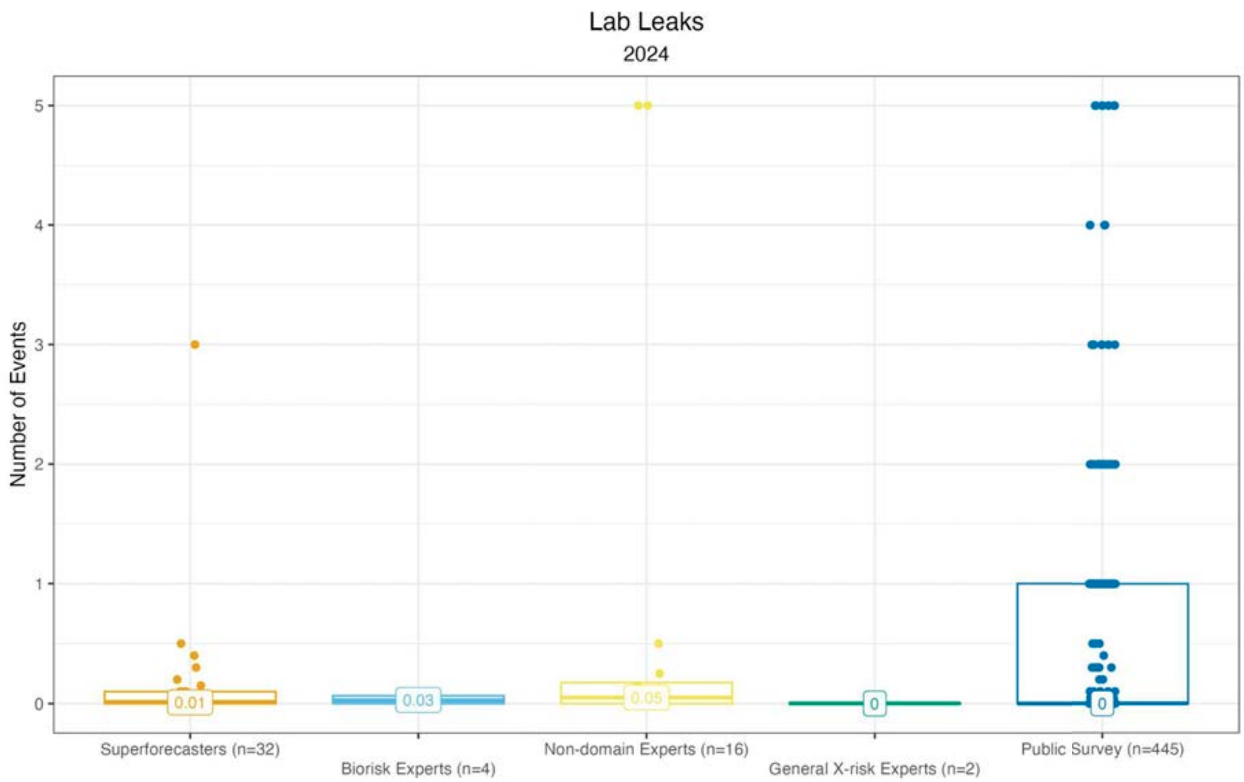
[667] 337, "Despite the threat of nuclear weapons causing nuclear winter and therefore mortality in all countries, there is still a significant chance that they will be used."
[668] Numbers of forecasters are given as of Stage 4 of the XPT. Public survey respondents were surveyed outside of the tournament context.

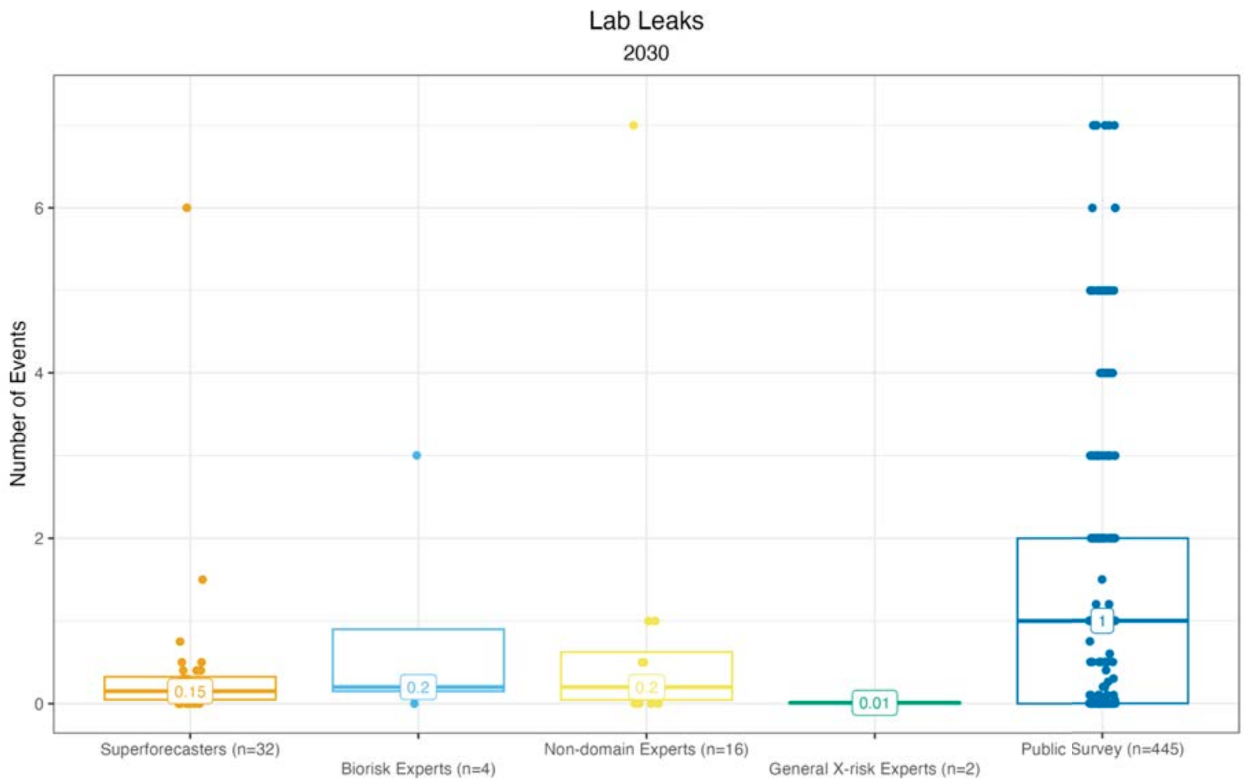
Domain Experts (N = 4)	2024	0	0.027	0.05	-3.69%
	2030	0.1	0.2	1.47	-2.26%
	2050	1.15	0.5	5.19	+2.87%
General X-Risk Experts (N = 2)	2024	n/a	0.0015	n/a	n/a
	2030	n/a	0.01	n/a	n/a
	2050	n/a	0.55	n/a	n/a
Non-Domain Experts (N = 16)	2024	0	0.05	1.87	-9.82%
	2030	0.071	0.2	2.97	-15.98%
	2050	0.7	1	3.61	+74.63%
Public Survey (N = 445)	2024	0		52951.58	-
	2030	1		483222.36	-
	2050	2		71141451.58	-

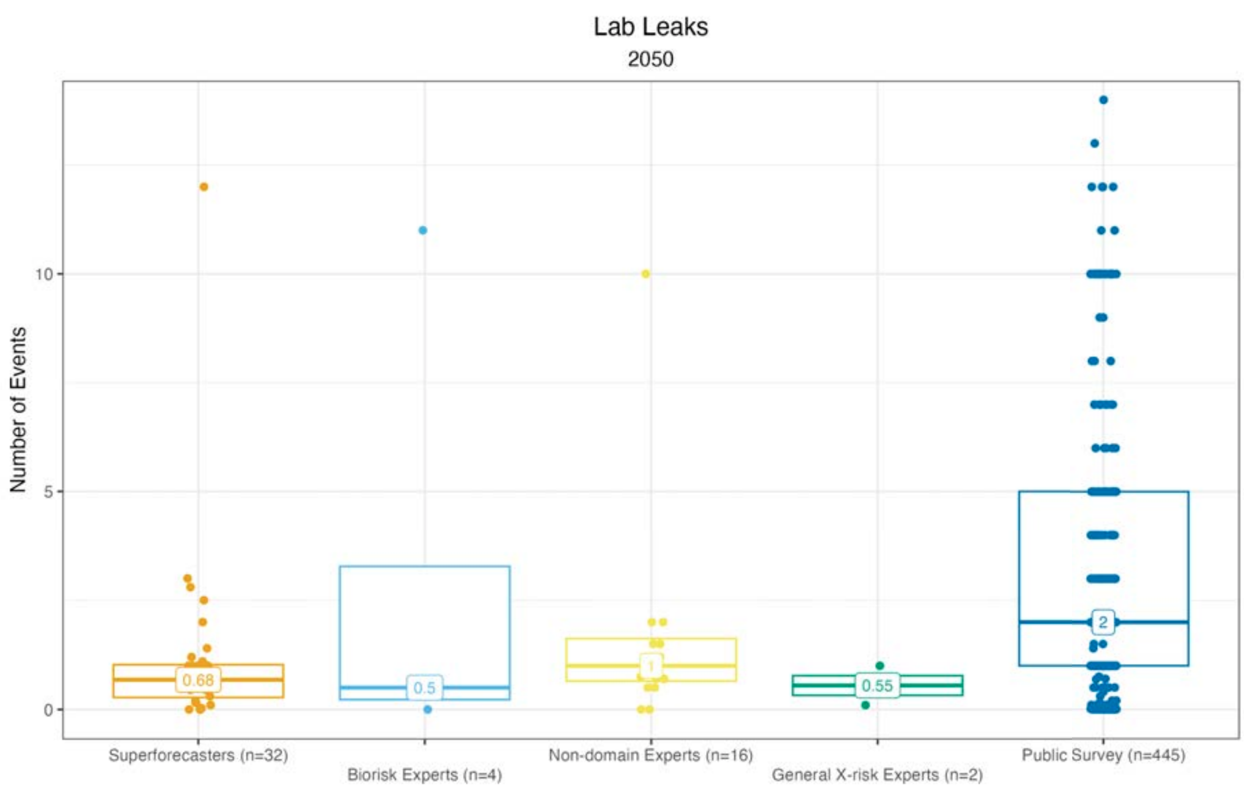

Lab Leaks - 2024 - 50th %
0.15
0.10
0.05
0.00
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 32)
Experts (n = 22)
Non-domain Experts (n = 16)


Lab Leaks - 2030 - 50th %
0.4
0.3
0.2
0.1
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 32)
Experts (n = 22)
Non-domain Experts (n = 16)


Lab Leaks - 2050 - 50th %
Median
1.0
0.9
0.8
0.7
0.6
0.5
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 32)
Experts (n = 22)
Non-domain Experts (n = 16)


Lab Leaks
2024
Number of Events
5
4
3
2
1
0
0.01
0.03
0.05
0
0
Superforecasters (n=32)
Biorisk Experts (n=4)
Non-domain Experts (n=16)
General X-risk Experts (n=2)
Public Survey (n=445)


Lab Leaks
2030
Number of Events
0
2
4
6
0.15
0.2
0.2
0.01
1
Superforecasters (n=32)
Biorisk Experts (n=4)
Non-domain Experts (n=16)
General X-risk Experts (n=2)
Public Survey (n=445)


Lab Leaks
2050
Number of Events
0
5
10
0.68
0.5
1
0.55
2
Superforecasters (n=32)
Biorisk Experts (n=4)
Non-domain Experts (n=16)
General X-risk Experts (n=2)
Public Survey (n=445)

Sources of agreement, disagreement and uncertainty

There was widespread agreement that advancements in technology would affect the risk of lab leaks.

There was disagreement about:

1. Proliferation of biotechnology.
2. Lab safety and adherence to standards.
3. Efficacy of regulations and oversight.
4. Whether continuation of the status quo implies a higher or lower forecast.
5. Base rates.

Arguments given for forecasts ≤ 0 (2024), 0 (2030), 0.5 (2050)

On 1 (effect of proliferation of biotech):

- The growth of the biotech sector has led to a maturing of security protocols.[669]

On 2 (lab safety and adherence to standards):

- Few labs and people working in the area, with adherence to worldwide procedural standards.[670]

On 3 (regulations and oversight):

- Increasing lab security will reduce the number of incidents.[671]
- There has been tighter regulation and safety measures following COVID-19 for BSL 3 and 4 labs.[672]

On 4 (status quo implication):

- Median value of a team with lower forecasts was largely informed by base rates and continuation of the status quo.[673]

On 5 (base rate):

- Lab leaks with high death tolls have not happened yet.[674] “Accidents to date have been few and mainly impacted people working in the lab.” (Team 339)

---

[669] 337, "Lower forecasts often discussed tighter regulation and safety measures for BSL 4 and 3 labs following COVID 19.”
[670] 339, "Very Low probability as there are few labs and few people working in the area and labs follow worldwide procedural standards.
[671] 344, "Another argument for low-end scale is that there is a global push for greater lab security, and we can assume that number of incidents could drop due to this fact."
[672] 337, “Lower forecasts often discussed [...] maturing security protocols resulting from growth of biotech sector."
[673] 337, "Median value was largely informed by base rates and a continuation of the status quo."
[674] 344, "All of the low-end estimates from the forecasters are based on a shared argument that the lab leak events with the high-death toll didn't happen yet."

On technological changes:

- Improvements in lab safety technology will reduce escapes.[675] [676]
- Automated processes will reduce contact between pathogens and humans.[677]

Preventative measures outside of labs will reduce contagions:

- Use of prophylaxis, rapid detection, and isolation reduce contagions.[678]

Arguments given for forecasts ≥ 0.015 (2024), 0.16 (2030), 1 (2050)

On 1 (effect of proliferation of biotech):

- Increased biotechnology proliferation and lack of regulation on gain of function research may lead to lab leak events.[679]
- More labs and more accessible biotechnology increases potential for accidental creation of hazardous agents.[680]

On 2 (lab safety and adherence to standards):

- Lack of process management and control in laboratories.[681]
- Level 4 protection is not perfect, and organizations have failed to learn from prior incidents.[682]

On 3 (regulations and oversight):

- Long-term increase in competition and inability to regulate bioengineering could lead to disasters.[683]

On 4 (status quo implication):

- Increase in BSL-4 labs and facilities will result in higher risk, assuming similar rates of human error as existing facilities.[684]

---

[675] 339, "The rationale for the low estimates were driven by a faith in the improvement in technology used in the labs doing the research. The technology improvement would be driven by a desire to ensure labs were safe for employees and the community where located."
[676] 336, "The increase of safety precautions in biosafety labs will reduce any escapes."
[677] 336, "In addition, the development of automated processes such as the use of robots would also reduce the contact between pathogens and humans."
[678] 336, "Combined with use of prophylaxis, rapid detection and isolation, will further reduce contagions."
[679] 345, "The strongest arguments for the median forecast are that with increasing proliferation of biotechnology and a lack of regulation on gain of function research will lead to some small number of lab leak events occurring."
[680] 342, "[H]igher prevalence of bio labs with more accessible technologies for manufacturing / tailoring of potentially hazardous substances make accidental creation of such agents easier."
[681] 338, "There is a documented a lack of process management and control across many labs."
[682] 345, "On the high end, there is a belief that prior incidents have shown that level 4 protection is not 100% perfect at protecting against pathogen release. Organizations have failed to learn lessons about how to secure these facilities."
[683] 342, "Higher competition (to 'stay in the game') and inability to regulate would contribute."
[684] 340, "Others attribute an increase in risk over time due to the expected growth in the number of BSL-4 labs and facilities (assuming similar rates of human-error associated with each lab)."

- Extrapolating from risk assessments of labs and near-misses by the Department of Homeland Security and other agencies to the timeframe in question produces a higher forecast.[685]

On 5 (base rate):

- If you include near-misses, base rate calculations are higher.[686]

Other arguments given

Arguments in favor of lower forecasts:

- Security measures and controls will improve once a lab leak is confirmed.[687] [688]
- Verification of lab leaks is difficult due to economic impacts and incentives for countries to not cooperate.[689] [690]

Arguments in favor of higher forecasts:

- Lower technical barriers and regulation increase chances of leaks.[691]
- Lowering costs of synthesizing and economics of scale make it more accessible to have sites handling dangerous materials.[692]
- Human error is the largest threat.[693]
- Inclusion of ambiguous cases of potential lab leaks, such as COVID-19 originating from a Wuhan lab, suggests higher base rates.[694]

---

[685] 338, "[T]his forecast [...] uses near misses and risk assessments of labs rather than just a base rate of previous lab leaks. In a very well researched forecast they document the levels of risk as assessed by the Department of Homeland Security and other agencies. This forecast is informed by a good survey of the existing risk assessment frameworks for lab leaks which is then extrapolated to each year within the question."

[686] 338, “Although not explicitly stated in the forecast, near misses were included here as part of the base rate.”

[687] 343, "Limiting principle: once a single confirmed lab leak (or another suspected lab leak) is identified by the global community, it is likely this will be sufficient inducement to improve controls on laboratories doing research that could potentially leak contagious pathogens."

[688] 343, "In the event a lab leak is detected that falls below the 1,000 death threshold, the fact of a deadly lab leak situation would likely still be sufficient to lead to an increase in precautions against future leaks."

[689] 345, "One of the major sources of disagreement or uncertainty among the team was in the verification of any such lab leaks. Especially with the potential economic impact that a pathogen release could cause, countries have huge incentives to not cooperate with investigations or come up with alternative series of events for any outbreaks that comes to be."

[690] 345, "On the low end, there is a suspicion that many of countries that host labs researching potentially dangerous pathogens would be extremely unwillingly to cooperate with any investigation that could implicate them as being at fault."

[691] 336, "Lower technical barrier to develop pathogens in laboratory with lower biosafety standard/regulation will likely increase the change of leaks."

[692] 344, "If we look at the extended time horizon we know that price of synthesizing, along with the economics of scale, will lower the price needed to 'have' such a site."

[693] 339, "Human error was cited as the biggest threat to generate a higher probability of occurrence."

[694] 337, "Inclusion of other potential lab leaks i.e. Wuhan lab leak theory pushes for higher base rates and some forecasters factored in the potential for additional leaks which were unaccounted for."

Cross-references with other questions

Q14: Novel Infectious Disease Surveillance System

Question 20: Individual Countries with Biological Weapons Programs

What fraction of a panel of 100 biosecurity experts will agree that country X [X=separately: U.S., Russia, China, North Korea, Israel, Iran, Syria] has an active biological weapons program at some point from January 1, 2022…
…through the end of 2024?
…through the end of 2030?
…through the end of 2050?

Question and resolution details, prior forecasts, and other relevant sources

Results[695]

Group	**Year**	**Stage 1 Median**	**Stage 4 Median**	**Stage 1 Standard Deviation**	**Percent Change in Standard Deviation, Stage 1 to Stage 4**
Super-Forecasters (N = 30) U.S.	2024	12.5	20	37.38	-7.96%
	2030	25	25	34.6	-5.66%
	2050	25	30	34.04	-8.39%
Super-Forecasters Russia	2024	85	80	34.01	-28.64%
	2030	80	80	30.79	-31.39%
	2050	85	80	30.08	-24.47%

695 Numbers of forecasters are given as of Stage 4 of the XPT.

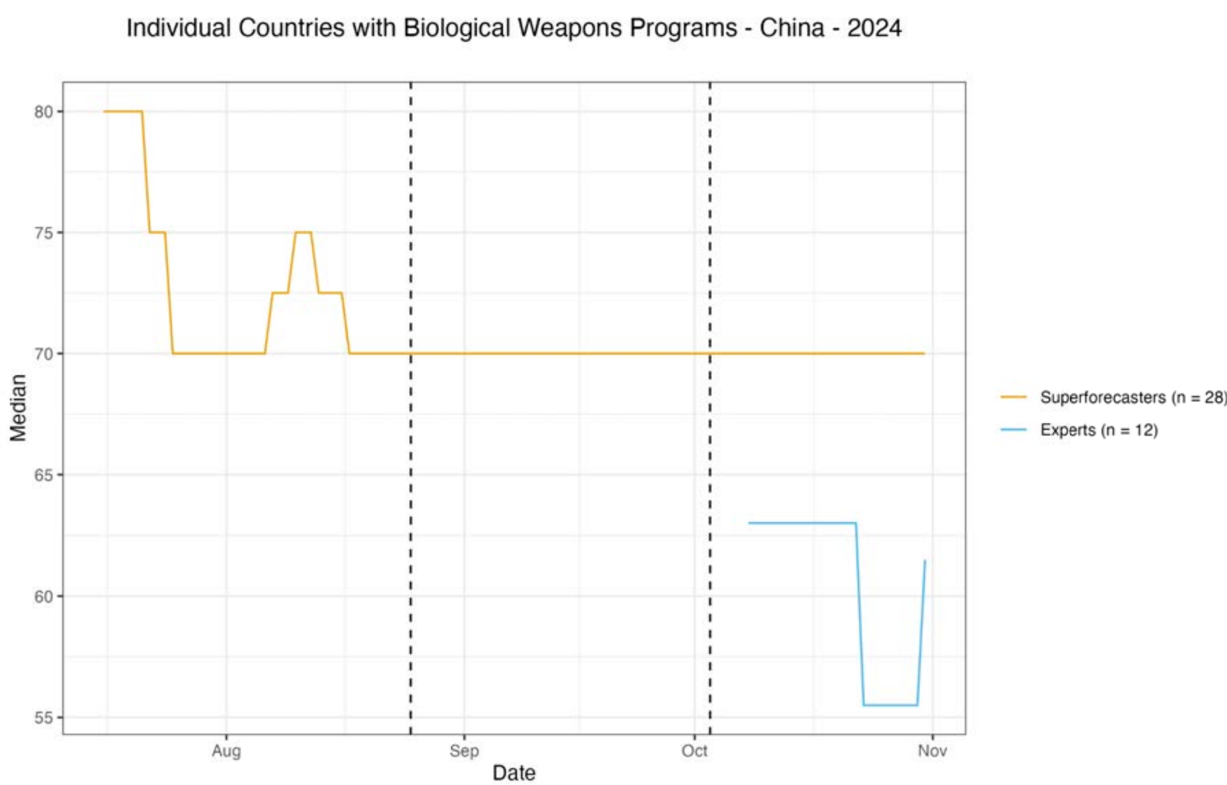
Super-Forecasters **China**	2024	85	70	38.88	-24.13%
	2030	60	85	33.96	-21.67%
	2050	72.5	80	36.89	-24.9%
Super-Forecasters **North Korea**	2024	90	77.5	36.53	-22.39%
	2030	85	80	34.96	-23.94%
	2050	72.5	80	30.44	-16.78%
Super-Forecasters **Israel**	2024	40	40	36.32	-13.42%
	2030	40	55	34.86	-10.5%
	2050	37.5	60	34.74	-15.32%
Super-Forecasters **Iran**	2024	70	60	37.9	-22.24%
	2030	60	65	36.02	-18.78%
	2050	60	65	36.53	-16.58%
Super-Forecasters **Syria**	2024	27.5	35	38.36	-23.35%
	2030	28	40	34.14	-14.65%
	2050	25	35	29.82	-1.53%
Domain Experts (N = 4) **U.S.**	2024	10.5	2	13.44	+5.68%
	2030	14	4.5	15.56	-13.28%
	2050	19	8.5	19.8	-12.38%

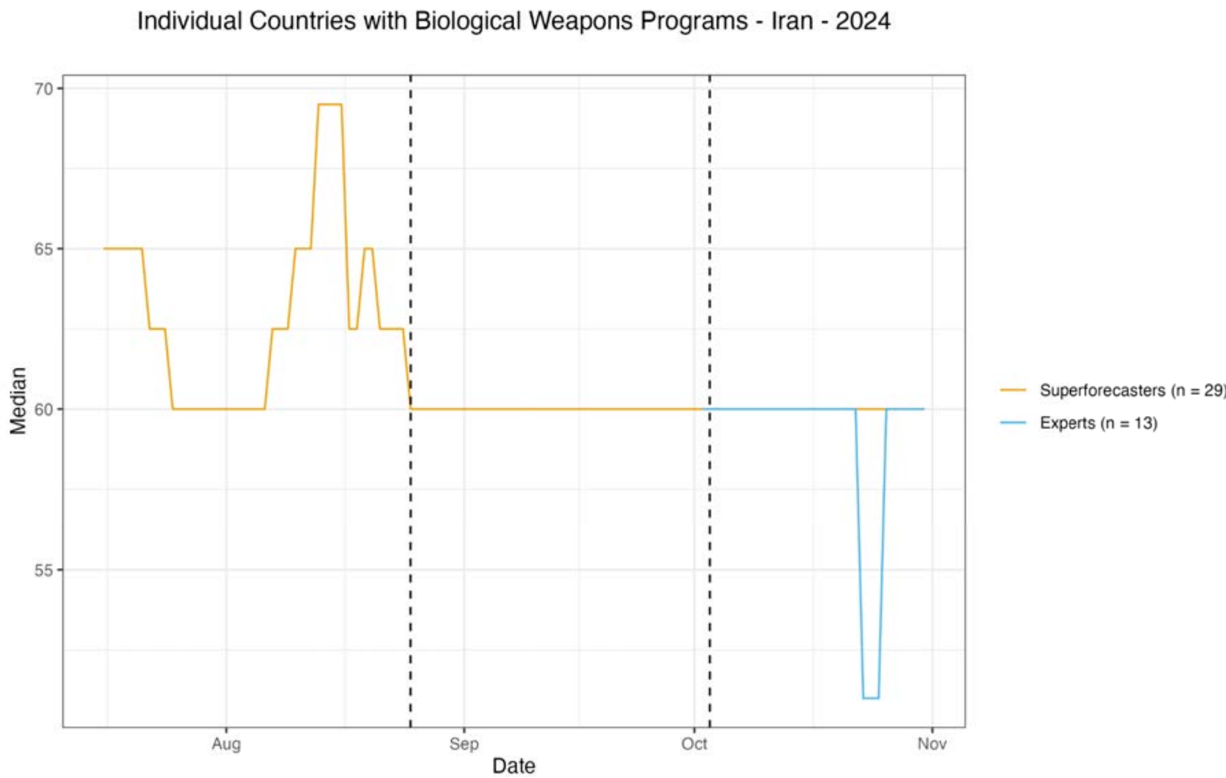
Domain Experts **Russia**	2024	77.5	68	17.68	+129.41%
	2030	84	90	12.73	+297.08%
	2050	91.5	95	4.95	+891.46%
Domain Experts **China**	2024	20	51	n/a	n/a
	2030	46	60	19.8	+88.33%
	2050	57.5	70	17.68	+140%
Domain Experts **North Korea**	2024	80	70	0.00	Inf
	2030	84	85	1.41	+1,036.52%
	2050	90	92.5	0.00	Inf
Domain Experts **Israel**	2024	18	30.5	n/a	n/a
	2030	41.5	60	26.16	+39.9%
	2050	50	57.5	28.28	+19.88%
Domain Experts **Iran**	2024	73.5	61.5	9.19	+240.71%
	2030	80	78	7.07	+445.22%
	2050	87.5	88	3.54	+1,027.12%
Domain Experts **Syria**	2024	45	52.5	35.36	-25.72%
	2030	50	55	35.36	-13.59%
	2050	57.5	52.5	31.82	-16.48%

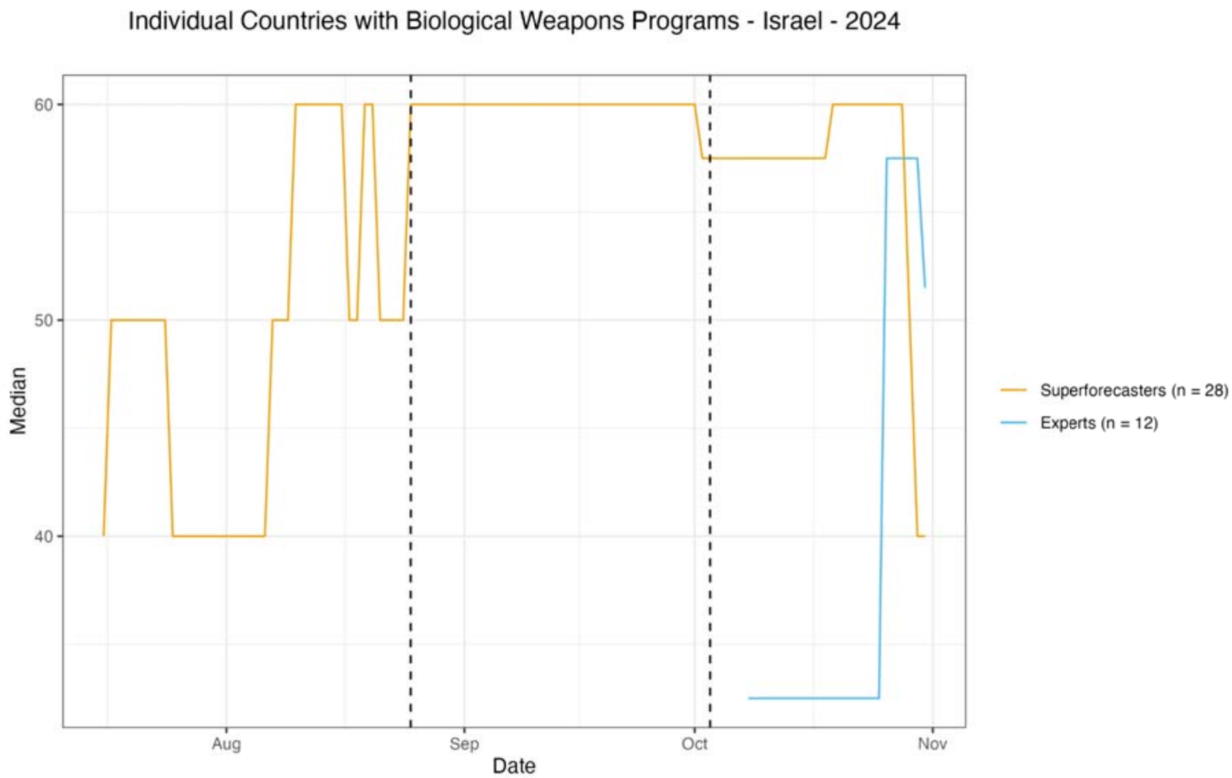
General X-Risk Experts (N = 4) **U.S.**	2024	10	12.5	6.51	-1.69%
	2030	5	9	4.62	+104.63%
	2050	5	14.5	3.21	+290.51%
General X-Risk Experts **Russia**	2024	65	75	43.68	-21.6%
	2030	49	67	42	-31.32%
	2050	17.5	48	39.00	-7%
General X-Risk Experts **China**	2024	40	63	42.43	-29.43%
	2030	30	48	25.45	-25.33%
	2050	17.5	48	2.12	-7.34%
General X-Risk Experts **North Korea**	2024	80	80	43.94	-20.97%
	2030	70	77.5	39.78	-34.12%
	2050	55	60	26.39	-39.09%
General X-Risk Experts **Israel**	2024	10	29	34.44	-13.01%
	2030	12	37.5	27.07	-7%
	2050	16	44	22.27	+39.51%
General X-Risk Experts **Iran**	2024	27	49	31.11	-11.72%
	2030	30	34.5	17.06	+21.22%
	2050	12	30	11.72	+122.74%

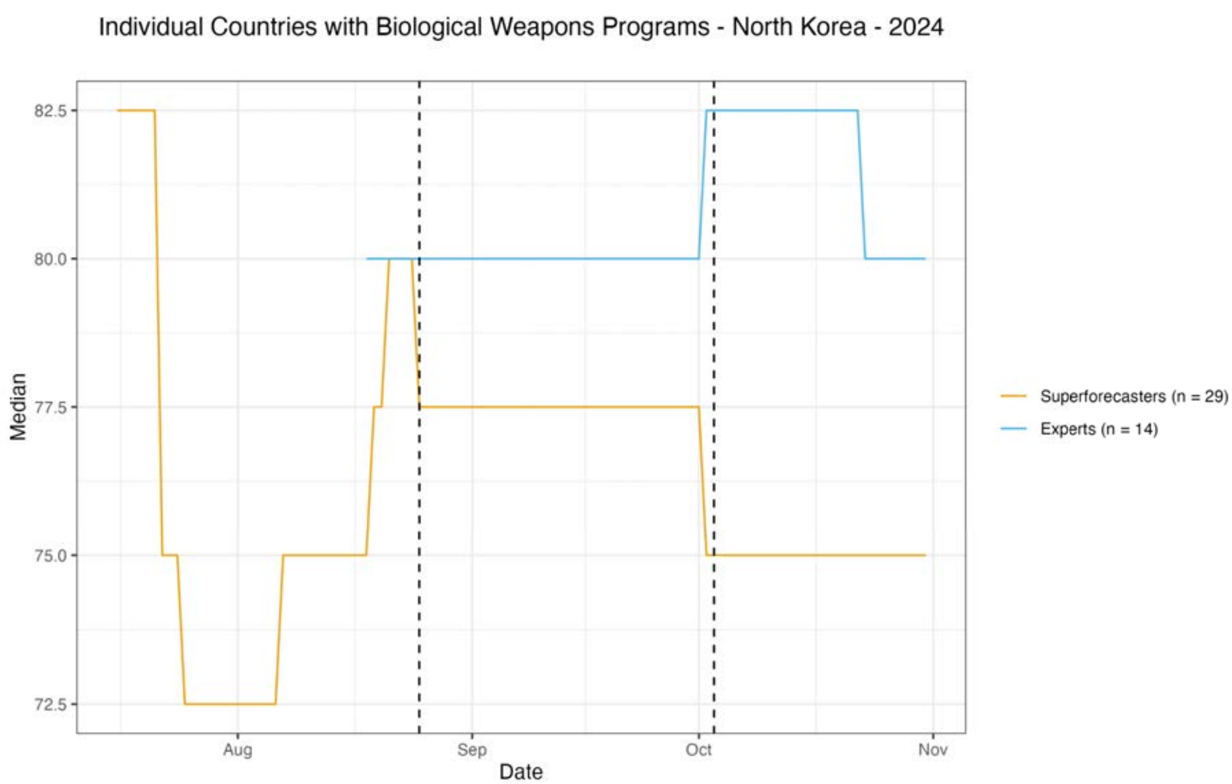
General X-Risk Experts **Syria**	2024	30	45	43.86	-22.04%
	2030	30	47.5	36.60	-27.28%
	2050	20	34.5	17.52	+10.13%
Non-Domain Experts (N = 7) **U.S.**	2024	6.5	10	2.12	+1596.81%
	2030	8.5	10	2.12	+1588.68%
	2050	17	14	4.24	+772.91%
Non-Domain Experts **Russia**	2024	50	85	63.64	-66.76%
	2030	57	83	52.32	-54.86%
	2050	52.5	85	45.96	-43.4%
Non-Domain Experts **China**	2024	60	68	28.28	-28.05%
	2030	60	75	20.67	-20.67%
	2050	56	74	22.63	+6%
Non-Domain Experts **North Korea**	2024	47.5	82.5	53.03	-68.07%
	2030	65	84.5	35.56	-49.99%
	2050	65	80	35.36	-37.17%
Non-Domain Experts **Israel**	2024	10	55	n/a	n/a
	2030	20	55	n/a	n/a
	2050	30	57.5	n/a	n/a

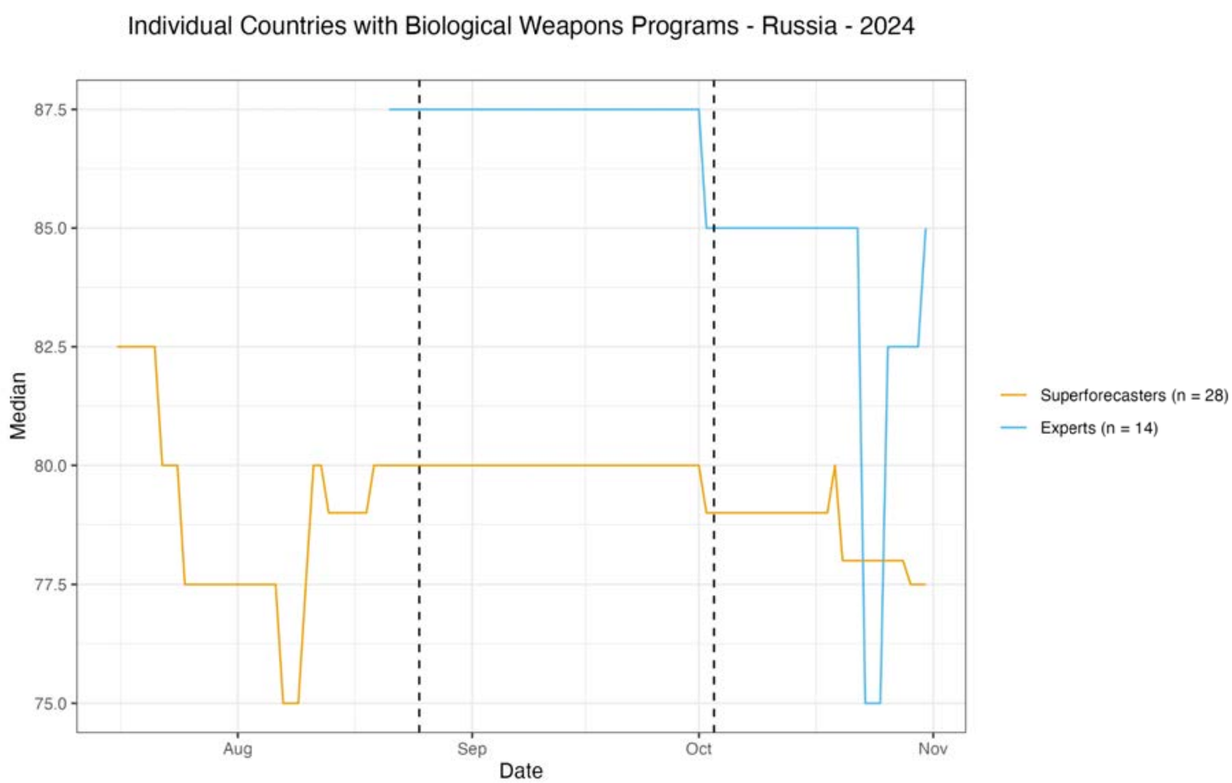
Non-Domain Experts **Iran**	2024	5	60	n/a	n/a
	2030	42.5	60	45.96	-49.41%
	2050	42.5	60	31.82	-27.74%
Non-Domain Experts **Syria**	2024	30	38.5	n/a	n/a
	2030	30	32.5	n/a	n/a
	2050	30	35	n/a	n/a



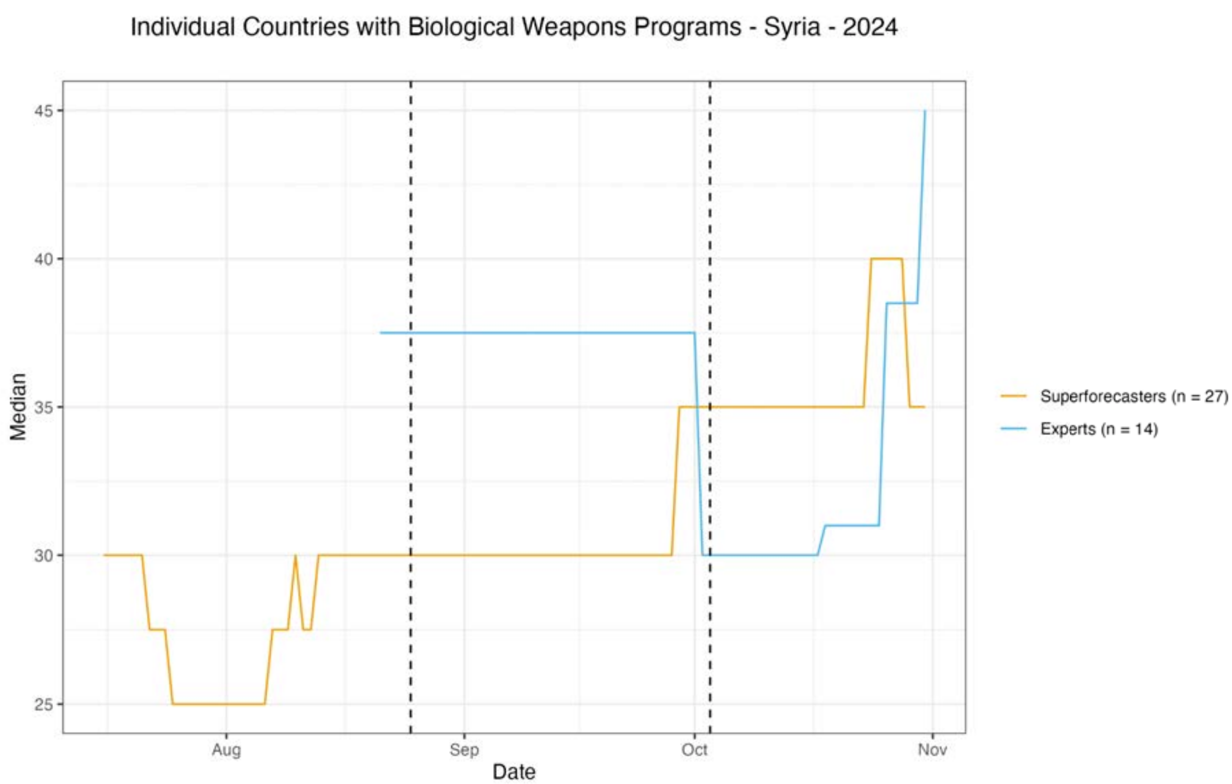

Individual Countries with Biological Weapons Programs - Iran - 2024
70
65
60
55
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 29)
Experts (n = 13)


Individual Countries with Biological Weapons Programs - Israel - 2024
60
50
40
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 28)
Experts (n = 12)

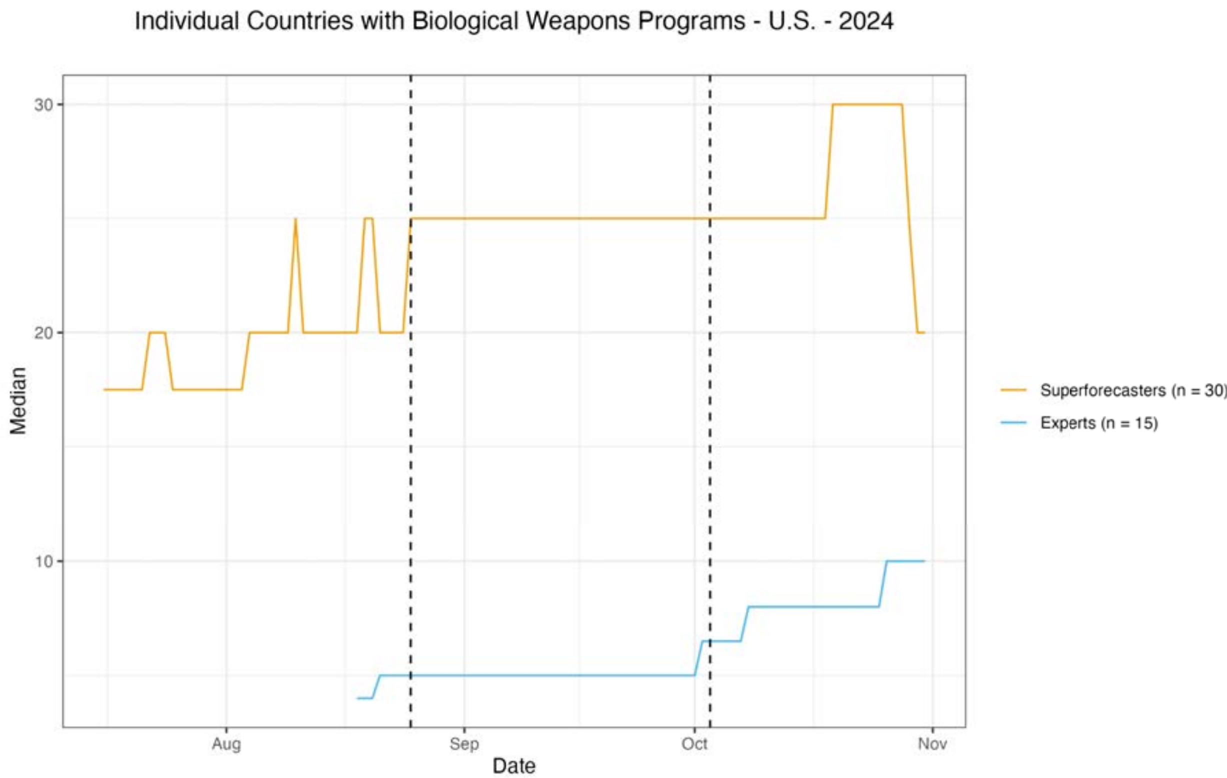

Individual Countries with Biological Weapons Programs - North Korea - 2024
82.5
80.0
77.5
75.0
72.5
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 29)
Experts (n = 14)

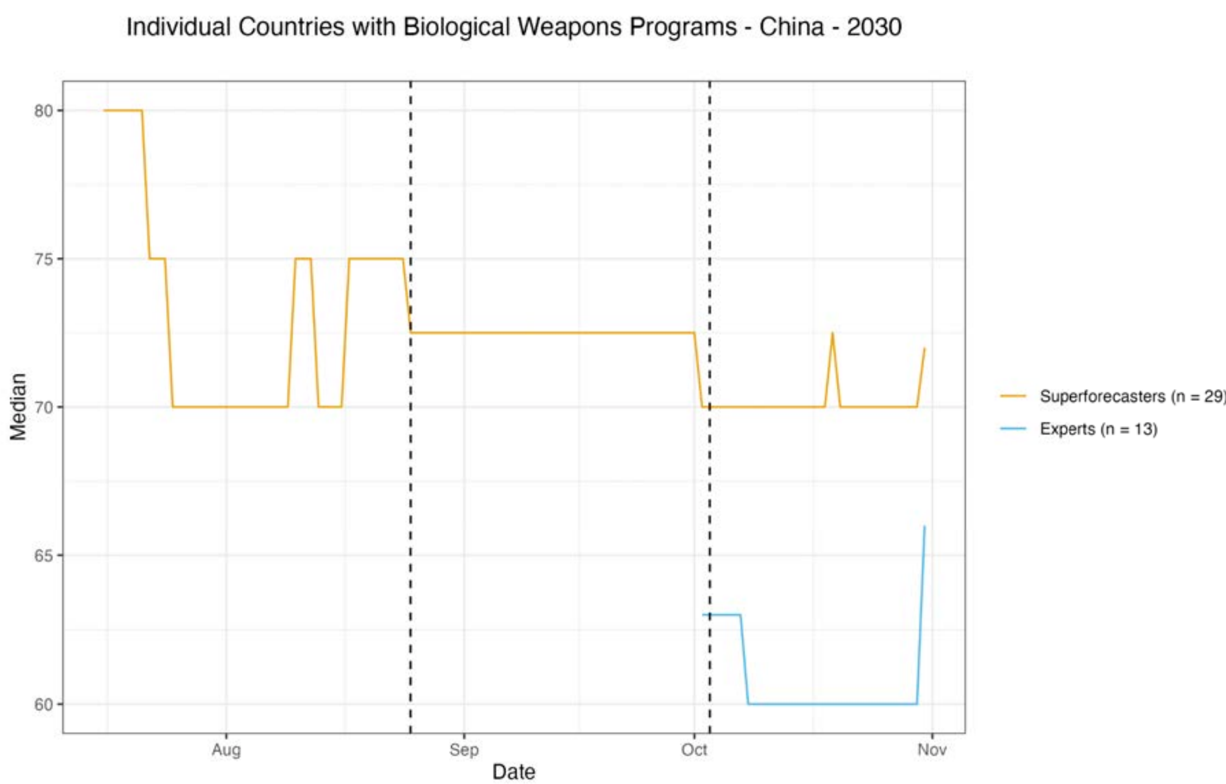


Individual Countries with Biological Weapons Programs - Russia - 2024
87.5
85.0
82.5
80.0
77.5
75.0
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 28)
Experts (n = 14)

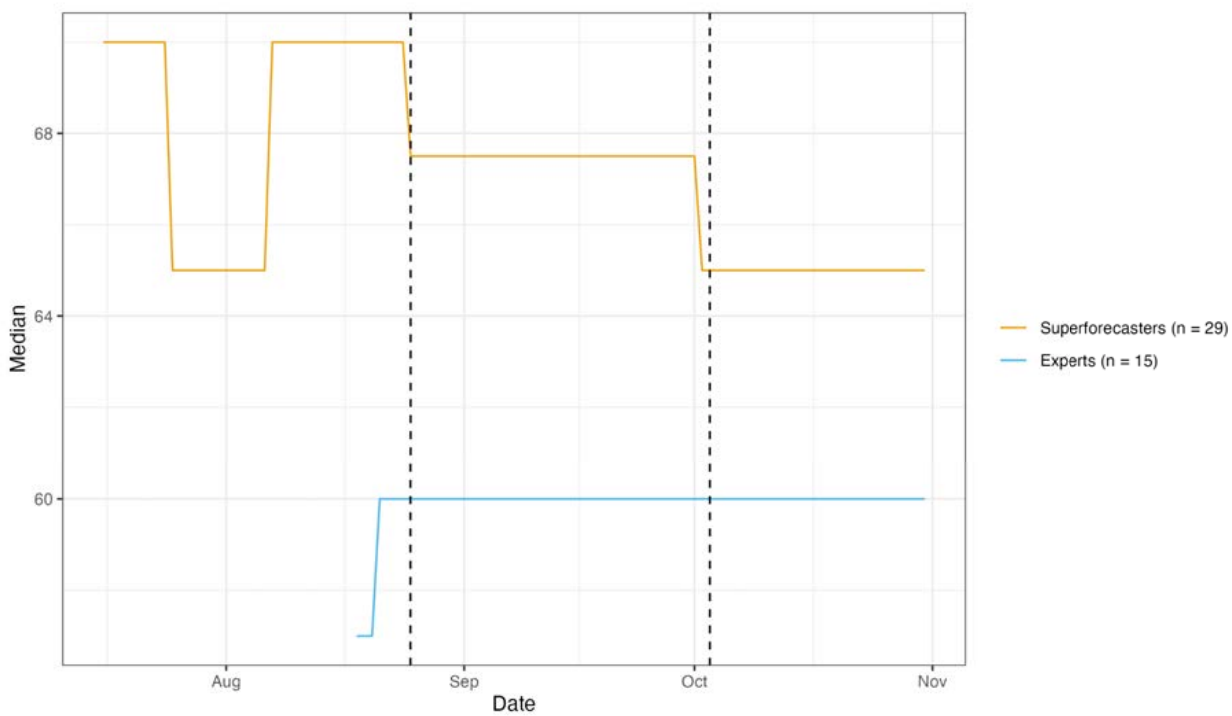


Individual Countries with Biological Weapons Programs - Syria - 2024
45
40
35
30
25
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 27)
Experts (n = 14)

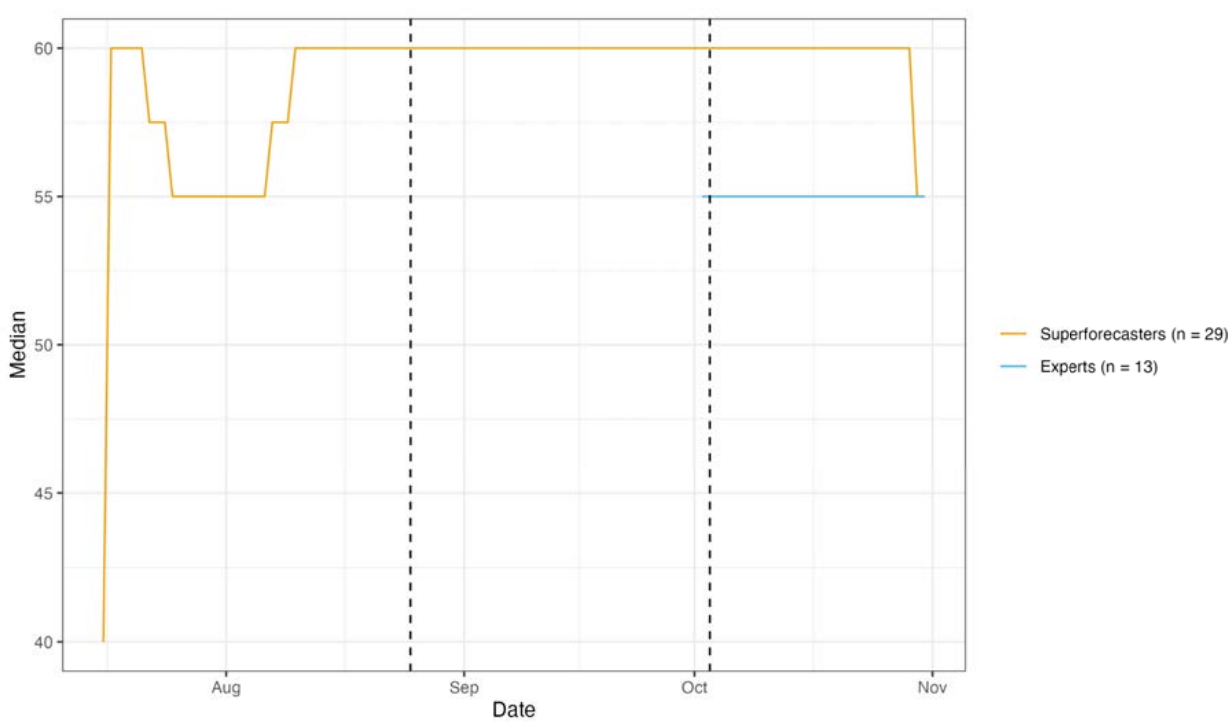

Individual Countries with Biological Weapons Programs - U.S. - 2024
30
20
10
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 30)
Experts (n = 15)

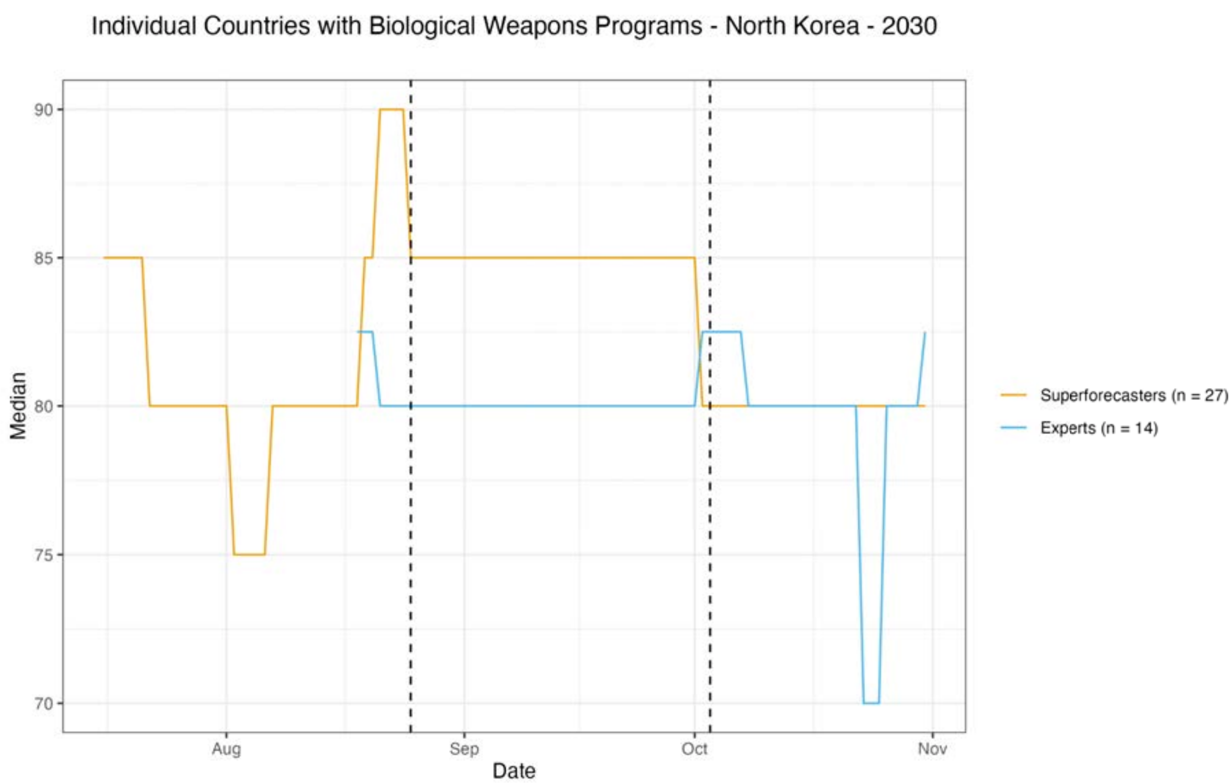

Individual Countries with Biological Weapons Programs - China - 2030
80
75
70
65
60
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 29)
Experts (n = 13)

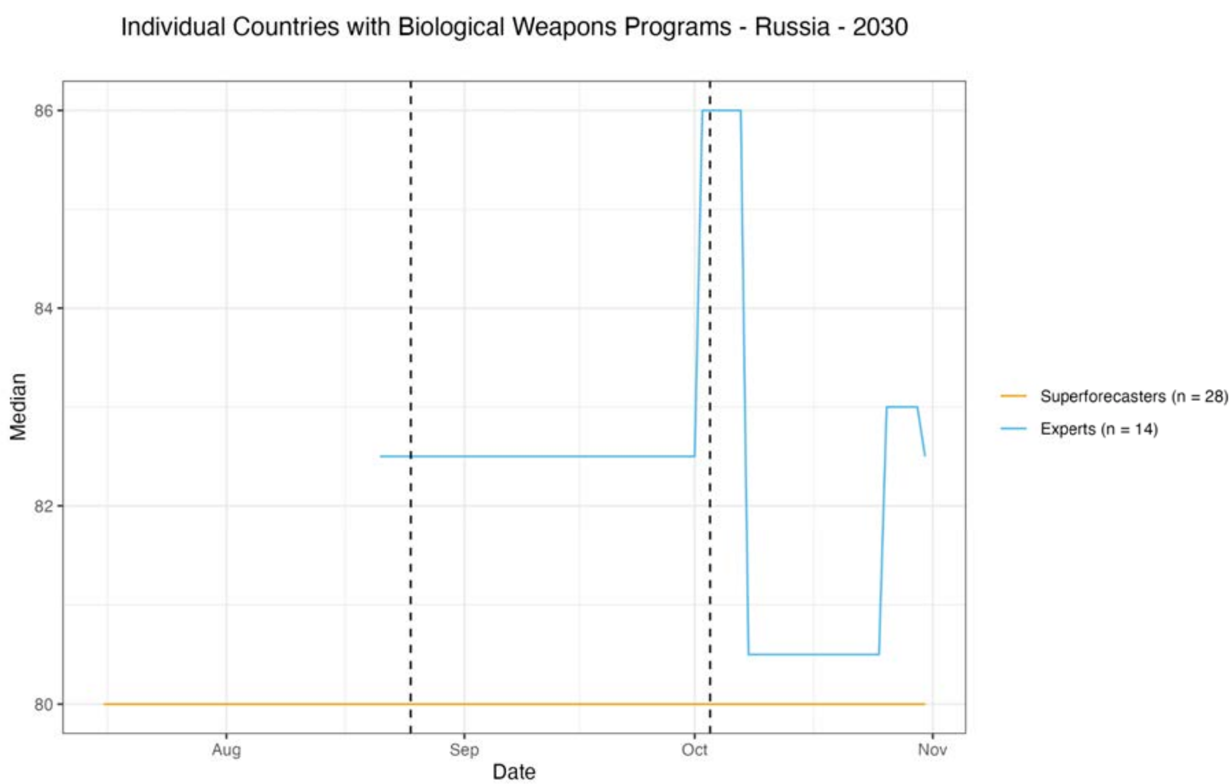
Individual Countries with Biological Weapons Programs - Iran - 2030

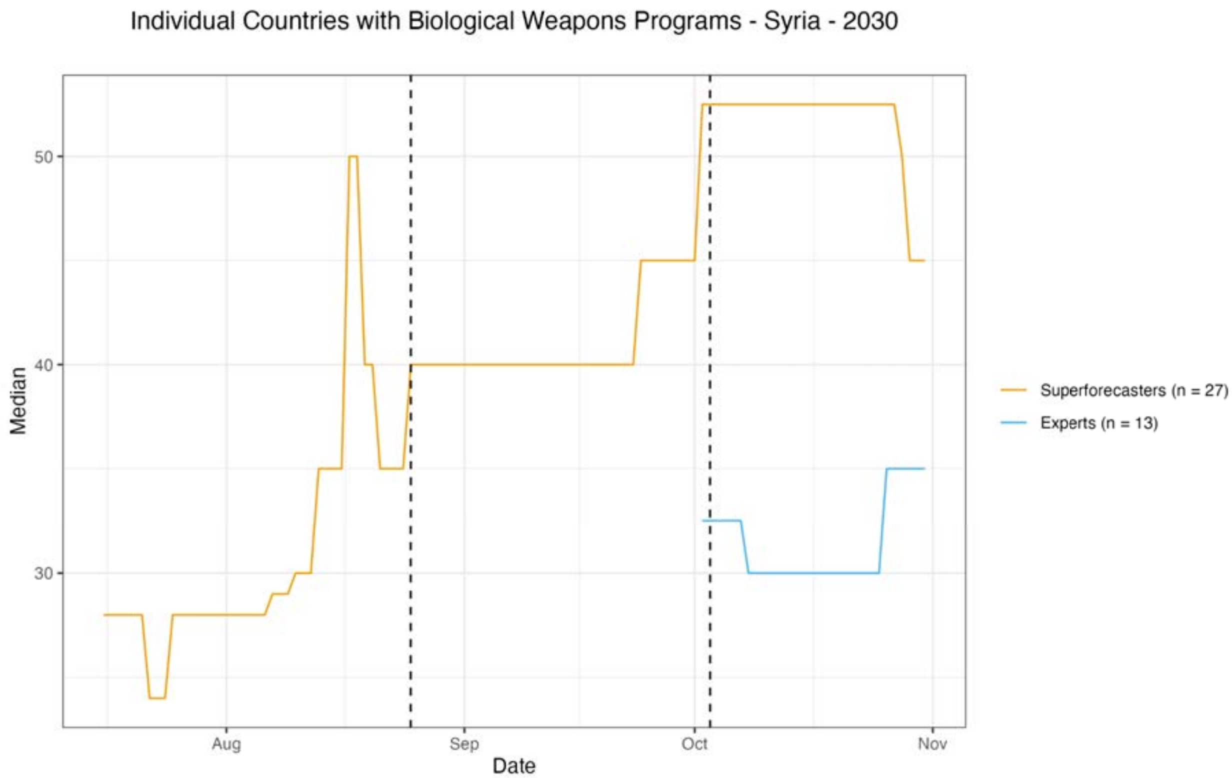


Individual Countries with Biological Weapons Programs - Israel - 2030

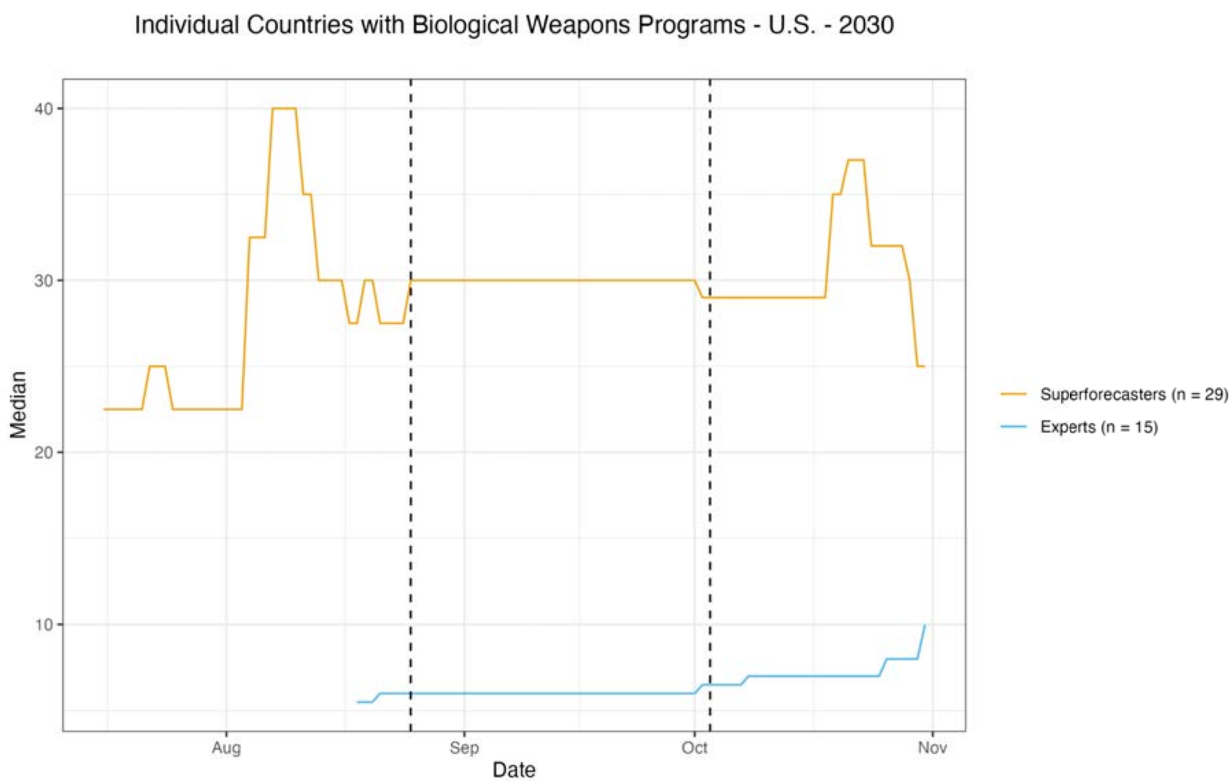



Individual Countries with Biological Weapons Programs - North Korea - 2030
90
85
80
75
70
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 27)
Experts (n = 14)

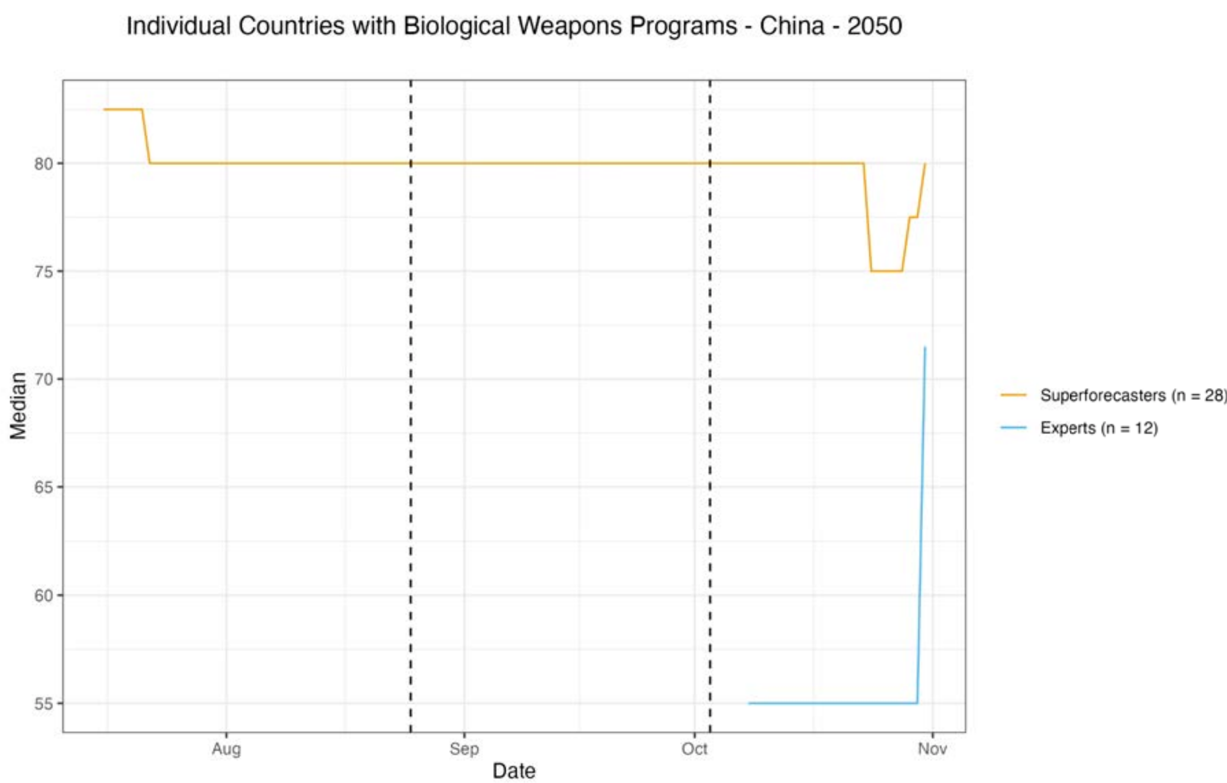

Individual Countries with Biological Weapons Programs - Russia - 2030
86
84
82
80
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 28)
Experts (n = 14)



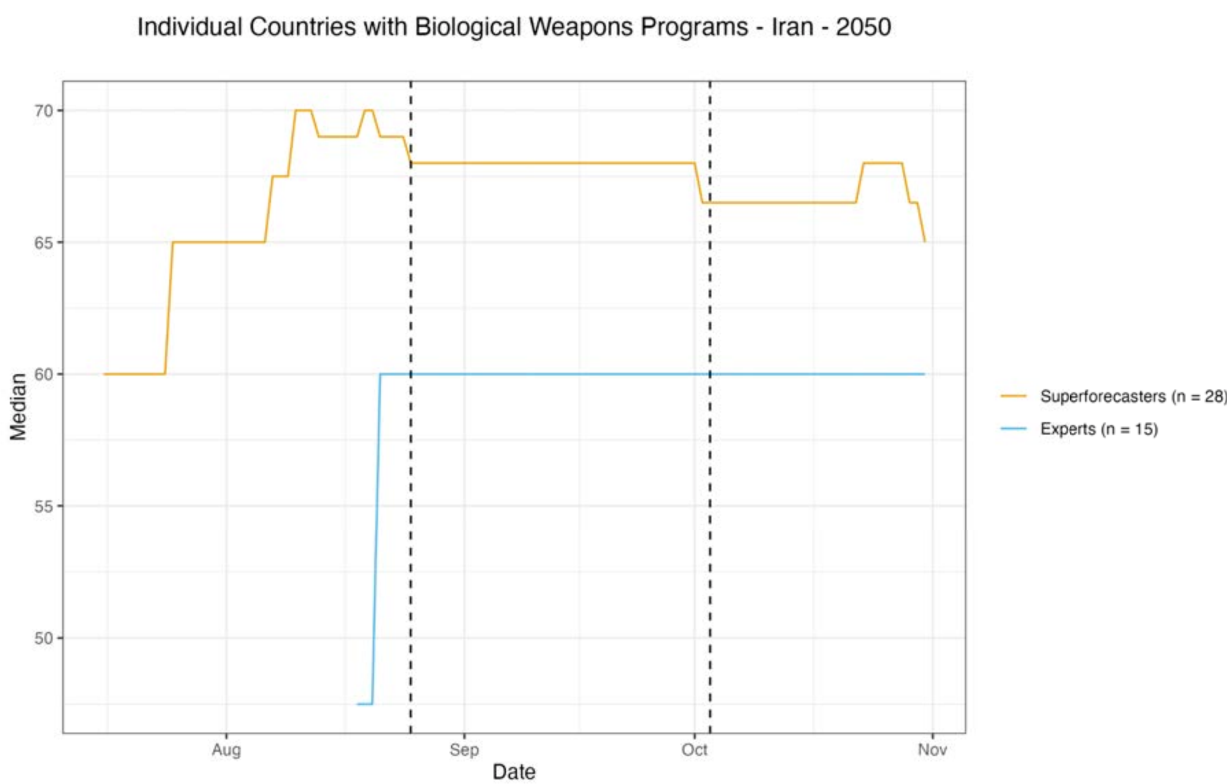
Individual Countries with Biological Weapons Programs - Syria - 2030
Median
50
40
30
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 27)
Experts (n = 13)



Individual Countries with Biological Weapons Programs - U.S. - 2030
Median
40
30
20
10
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 29)
Experts (n = 15)

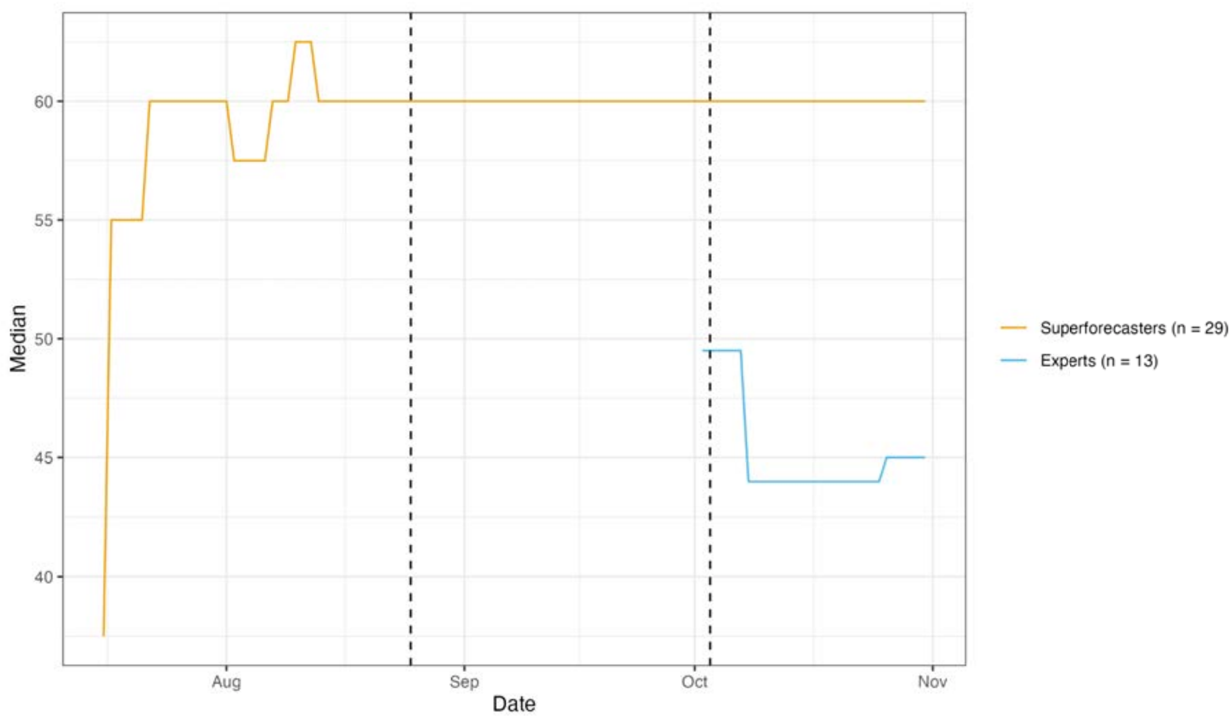


Individual Countries with Biological Weapons Programs - China - 2050
Median
80
75
70
65
60
55
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 28)
Experts (n = 12)

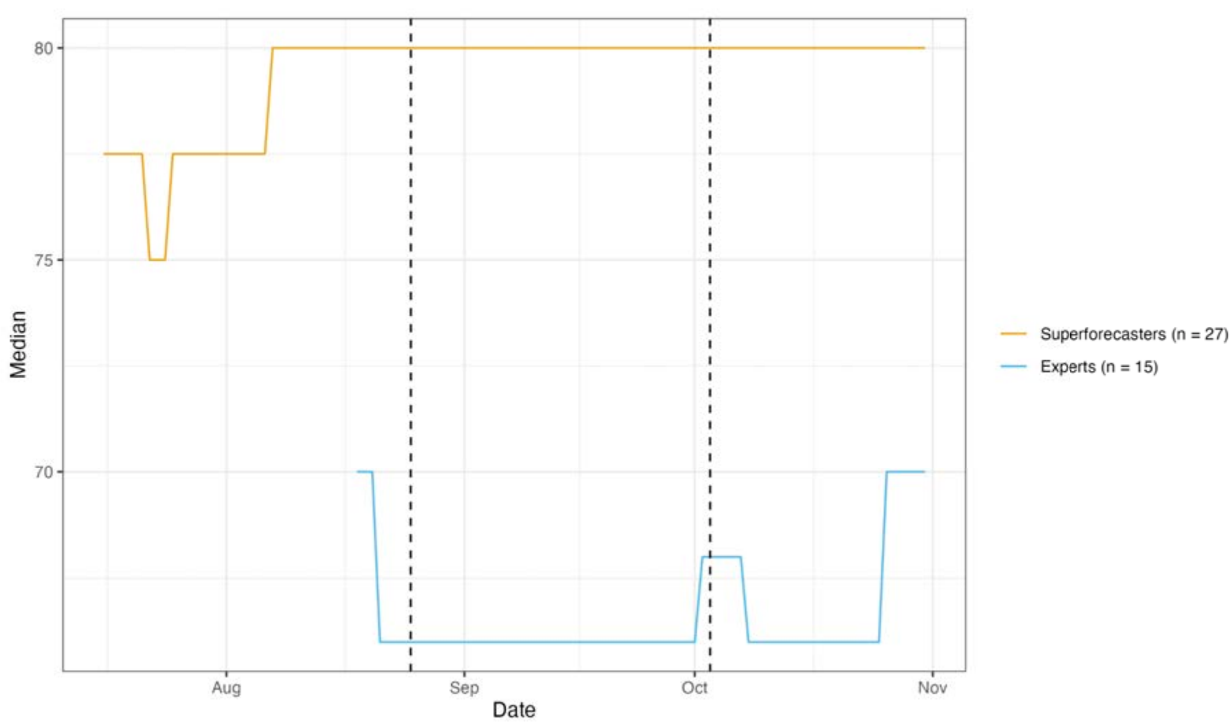


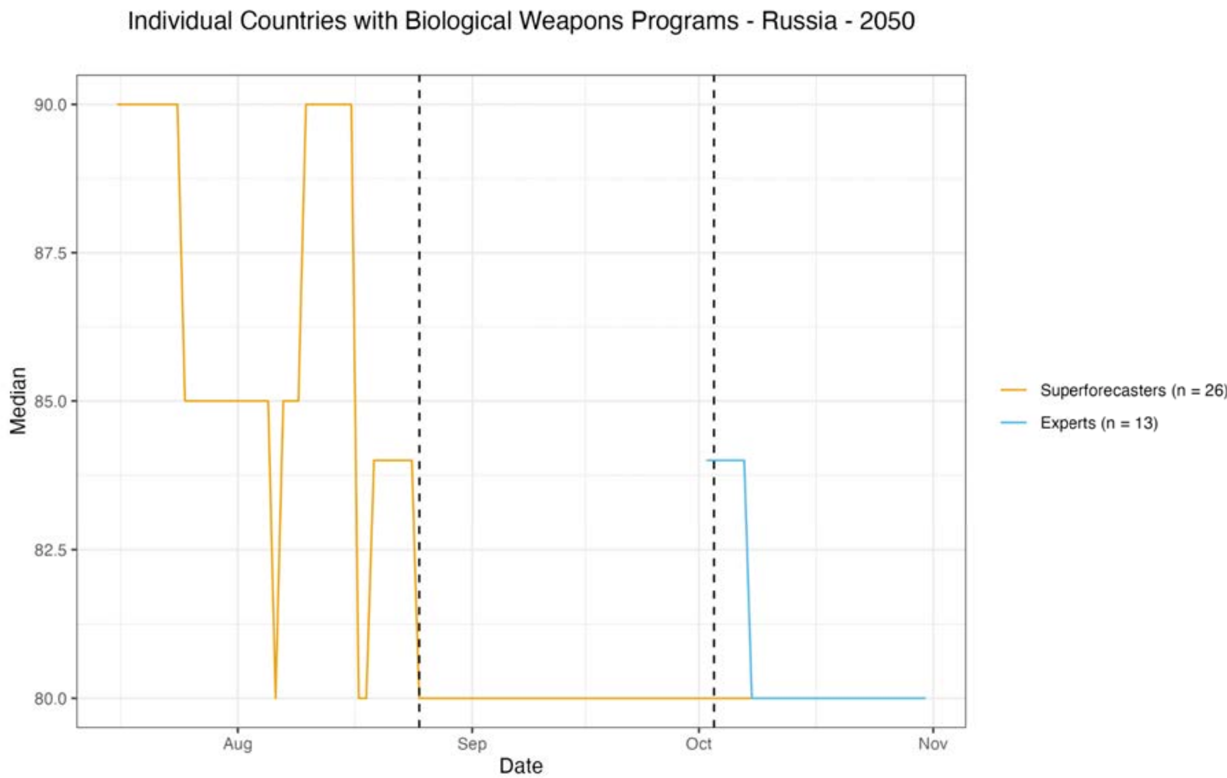
Individual Countries with Biological Weapons Programs - Iran - 2050
Median
70
65
60
55
50
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 28)
Experts (n = 15)

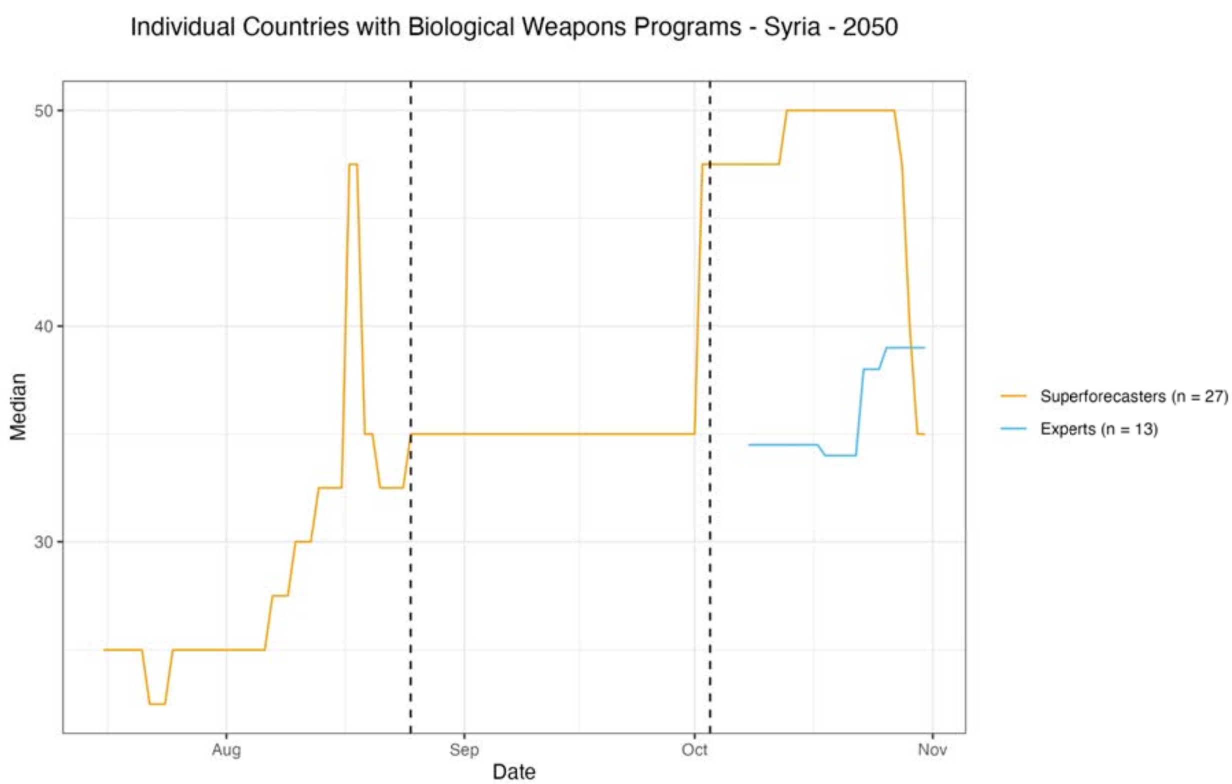
Individual Countries with Biological Weapons Programs - Israel - 2050

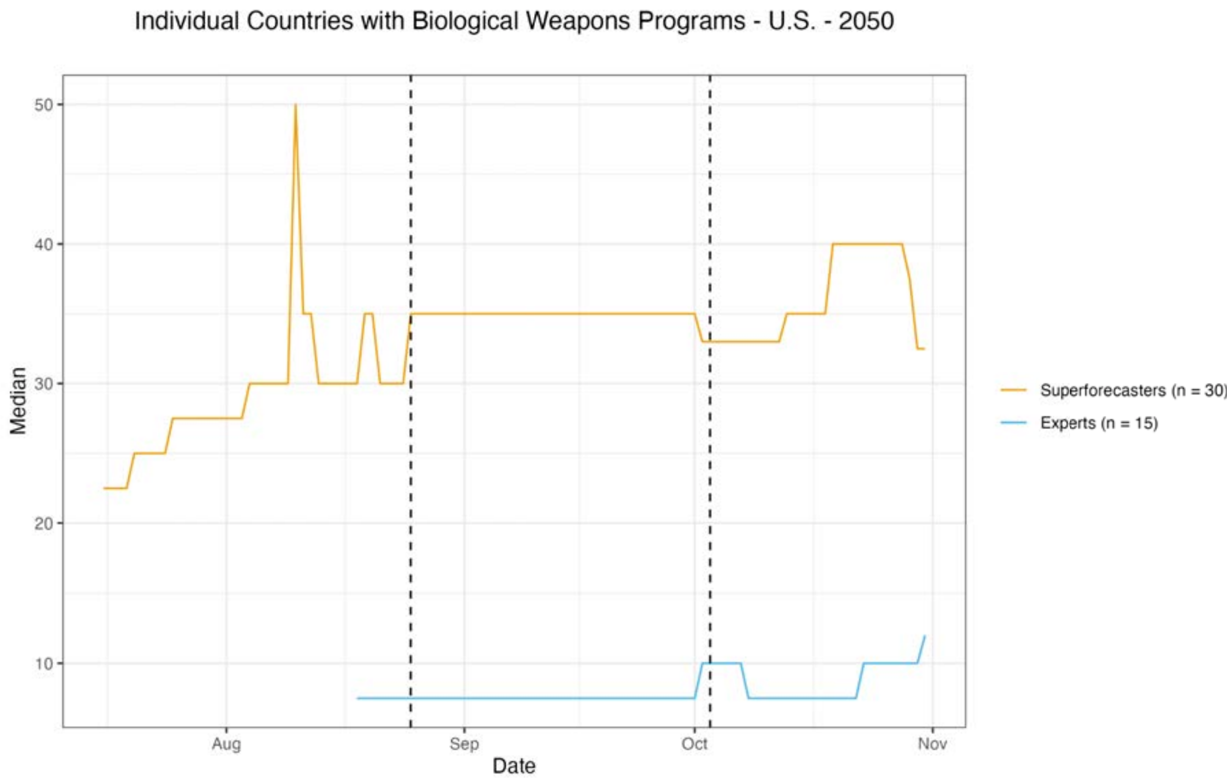


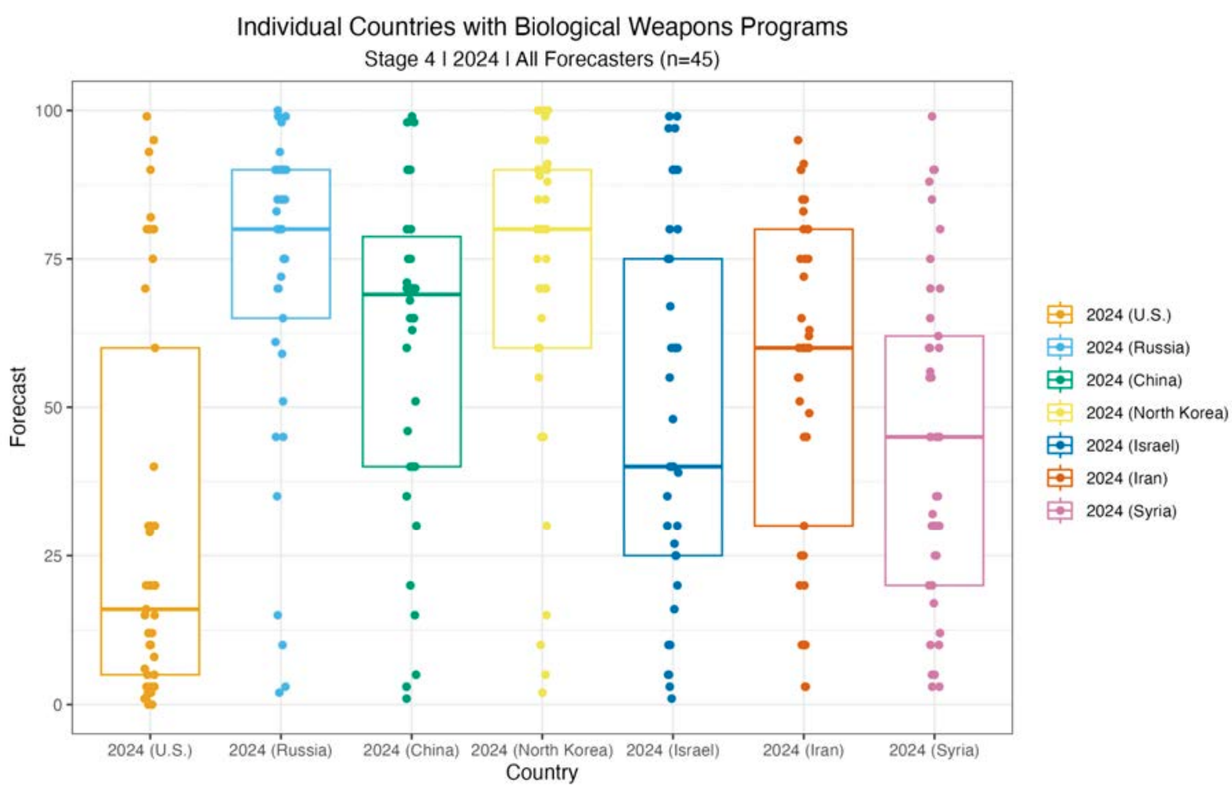
Individual Countries with Biological Weapons Programs - North Korea - 2050

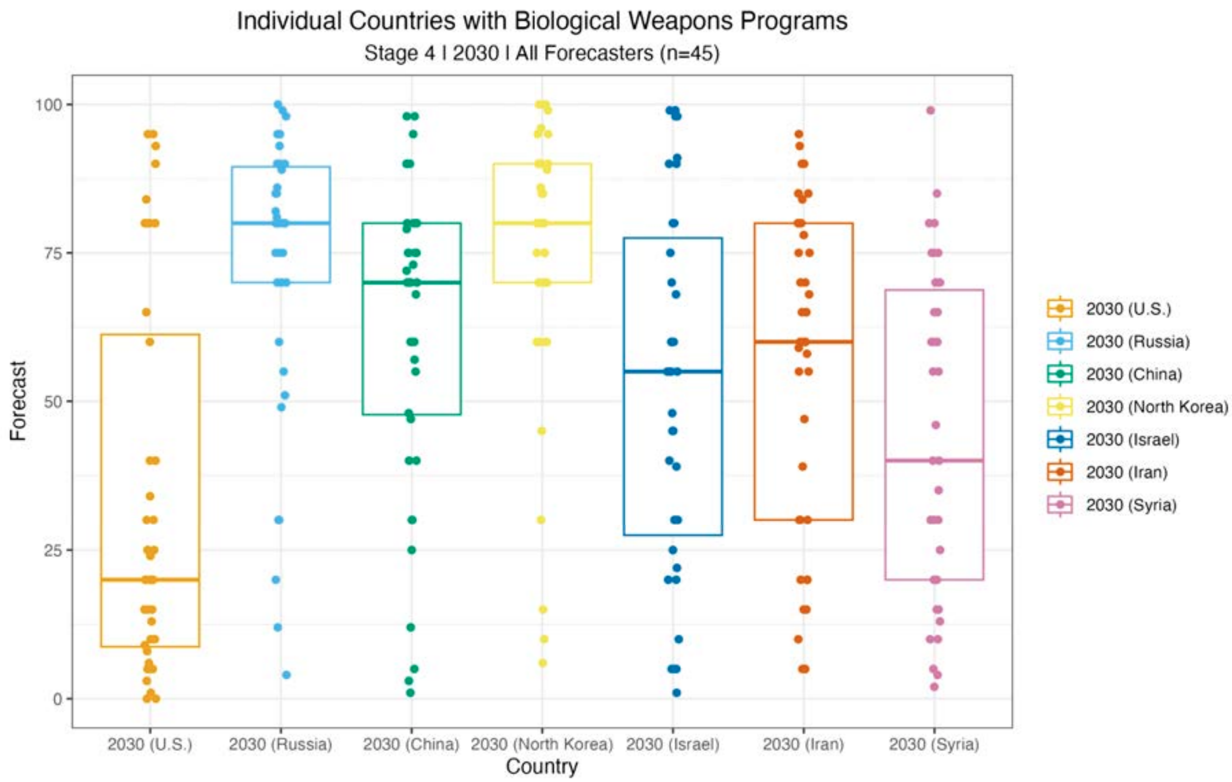


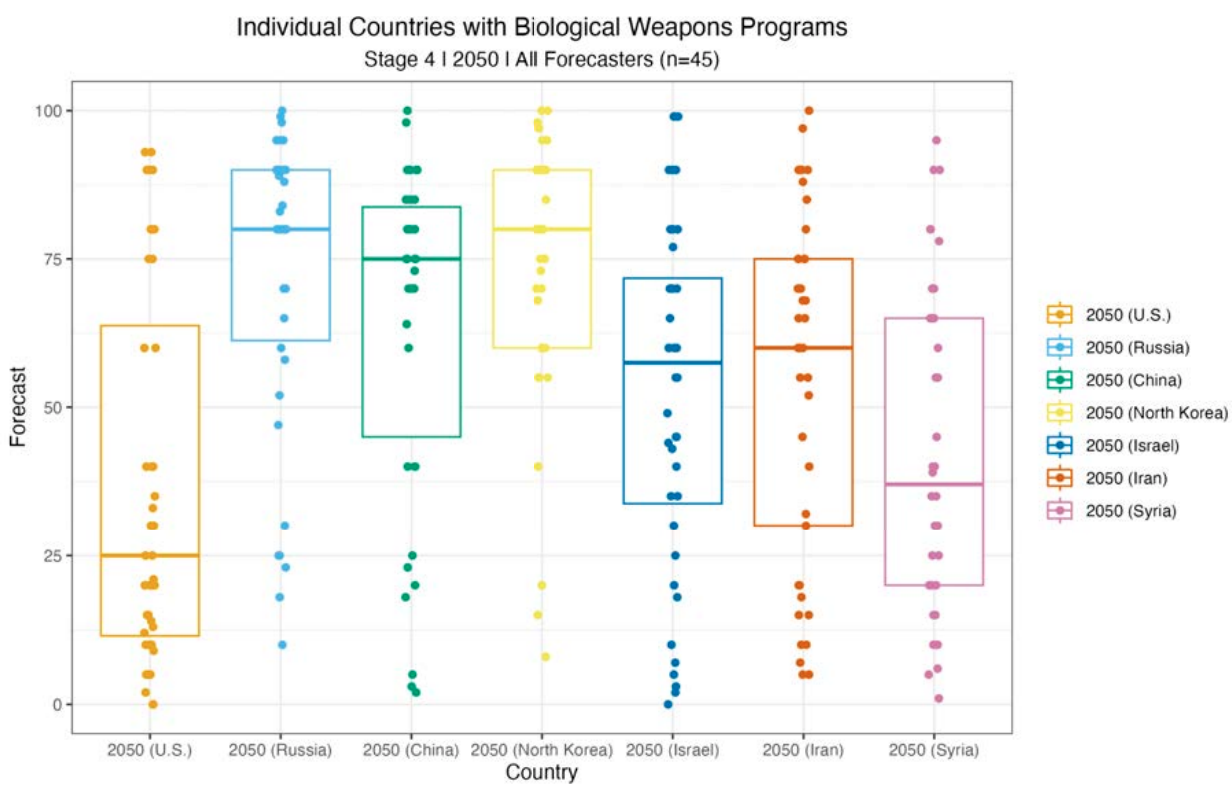

Individual Countries with Biological Weapons Programs - Russia - 2050
Median
90.0
87.5
85.0
82.5
80.0
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 26)
Experts (n = 13)


Individual Countries with Biological Weapons Programs - Syria - 2050
Median
50
40
30
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 27)
Experts (n = 13)


Individual Countries with Biological Weapons Programs - U.S. - 2050
Median
Date
Aug
Sep
Oct
Nov
Superforecasters (n = 30)
Experts (n = 15)


Individual Countries with Biological Weapons Programs
Stage 4 | 2024 | All Forecasters (n=45)
Forecast
Country
2024 (U.S.)
2024 (Russia)
2024 (China)
2024 (North Korea)
2024 (Israel)
2024 (Iran)
2024 (Syria)


Individual Countries with Biological Weapons Programs
Stage 4 I 2030 I All Forecasters (n=45)
Forecast
100
75
50
25
0
2030 (U.S.)
2030 (Russia)
2030 (China)
2030 (North Korea)
2030 (Israel)
2030 (Iran)
2030 (Syria)
Country


Individual Countries with Biological Weapons Programs
Stage 4 I 2050 I All Forecasters (n=45)
Forecast
100
75
50
25
0
2050 (U.S.)
2050 (Russia)
2050 (China)
2050 (North Korea)
2050 (Israel)
2050 (Iran)
2050 (Syria)
Country

Sources of agreement, disagreement and uncertainty

Agreement

- Forecasts should increase with longer time horizons
    - Most teams agreed that forecasts will most likely increase over time, as there is more time for a bioweapons program to be developed or discovered, and technology may make bioweapons more accessible in the future.[696] One team noted that lower forecasts in later time periods could be caused by information becoming available that a previously suspected bioweapons program did not actually exist.[697]

Disagreement

- High variance in individual forecasts
    - Many teams reported a wide range of forecasts from their individual participants, ranging from 1% to 90% for some countries in some teams.[698]

Uncertainty

- A prominent source of uncertainty was the difficulty of establishing whether a bioweapons program is in existence.[699]
- Some teams voiced uncertainty in what would be considered a bioweapons program, particularly what type of defensive activity would count as a dual-use weapon.[700]
- Several teams noted uncertainty in how the expert panel would reason and noted that expert opinion may be biased, e.g. by career incentives or political views.[701]

---

[696] 337, "The consensus seems to be that the main trend will be an increase over time, e.g. B.1) Previous classified programs might be leaked in the future or revealed maybe after a regime change. B.2) BW will become more attractive over time as a consequence of developments in biotech: BWs could become more targetable, accessible and resource efficient in the future."

[697] 338, "Based on question clarification, the odds should generally only rise (or stay flat) over time, because an affirmative only requires that the country have had a program at any time post-2022. The exception to this is if it later becomes (more) evident that a country suspected to have a program did not have one."

[698] 336, "We had a very wide distribution for all sub-questions, which might have been caused by the potential nature of those panels of experts and their individual levels of certainty." 340, "Given the small number of forecasters, and sheer scale of ambiguity (i.e., uncertainty over uncertainty) - forecasts ranging from essentially 0% to as high as 90% across forecasters - the median or its trend across the years is meaningless."

[699] 341, "A challenge to this question is the difficulty in verification. There were lots of allegations, but not so much proof for many of them." 343, "In addition, it is impossible to verify whether or not a country has a biological weapons program."

[700] 337, "[A]lso it is worth quoting "the phrasing, "Dual-use weapons that also have a defensive purpose but could be used for hostile purposes count for the resolution of this question," presents challenges." For instance, the US has a significant defensive need, which could be switched easily to offensive in case of need: would this count?" 336, "Next uncertainty is, what to count: is one program dual-use, is it only research or is it an active weapons program. What of that would be known facts, and what only speculation?"

[701] 340, "The composition of the expert panel is key to the question. Implicit biases, and political pressures affect whether the experts will err on the side of assuming a BW program exists or does not exist." 338, "Forecasting expert opinion adds an additional layer of uncertainty, since experts may be

- Several teams commented on the overall perception of a country, or bias from country affiliation, affecting their own forecasts.[702]

Arguments given for lower forecasts

- Reputational costs
  - Several teams suggested that reputational costs of bioweapons programs could lower forecasts. This may be due to the importance of maintaining a good global reputation (e.g. the US).[703] Or, for countries that already face international sanctions (e.g. North Korea), there is little to gain from keeping a bioweapons program secret, so there is less reason to suspect a secret program.[704]
  - Some teams also noted the reputational costs of accusing a state of running a bioweapons program, and how this may vary according to a country's geopolitical influence.[705]
- Bioweapons are unattractive
  - One team argued that bioweapons are unattractive to states given they can harm a state's own population.[706]
- Decreasing state violence
  - One team suggested that we should expect state violence to reduce in the future, citing Steven Pinker's lecture, "The Long Peace."[707]
- Correcting for bias against a country
  - One team suggested that negative views of some countries might lead forecasters to put a higher probability on those countries having an active bioweapons program, so it would be prudent to lower the forecast for those countries.[708]

---

biased by career incentives or other factors. This question is as much about forecasting those biases as the underlying truth."

[702] 344, "Views of each country appear to be based on an overall opinion of that country, and its weapons programs gathered from mass media - rather than from specific research or domain knowledge." 337, "For instance, in the case of China or Russia we might be presenting probabilities that are high simply because of a biased view against these countries coming from their lack of transparency or from their status as geopolitical rivals of the country of origin of a large part of the superforecasters here."

[703] 341, "US: Transparency requirements would make it difficult and unlikely to sustain an active offensive biological weapons programme. In 2024. there is a serious reputational cost to the open use of offensive biological weapons."

[704] 343, " It's also unclear why N. Korea wouldn't admit such a program outright if they had it, since international condemnation doesn't seem like a deterrent. "

[705] 344, "Biosecurity experts might themselves be biased when agreeing or disagreeing (or publicly proclaiming their opinion) about the existence of biological weapons programs in friendly or unfriendly countries, whatever the true state. Some countries are able to exert more pressure than others."

[706] 336, "[D]anger to own people reduces the individual probability (these things can be hard to control and backfire easily: bio weapons don´t make much sense for an "open use", because they endanger the country using them almost as much as the opponent)."

[707] 341, "2050 projections are based on the hypothesis of a continuing pacification of international relations (see Steven Pinker on 'the long peace')."

[708] 337, "[T]he lack of information might drive us to consider probabilities that are potentially too high. For instance, in the case of China or Russia we might be presenting probabilities that are high simply

Arguments given for higher forecasts

- Allegations of bioweapons programs
  - Several teams noted that many of the countries had been alleged to have a bioweapons program in the past or currently. Many teams referred to the [Arms Control Association's fact sheet] on biological and chemical weapons status.[709]
- Biotechnology may make bioweapons more attractive
  - Some teams noted that progress in biotechnology may make bioweapons cheaper, more effective, and easier to develop.[710]
- Pursuit of other weapons of mass destruction, especially chemical weapons programs
  - One team noted countries' use of chemical weapons, with the implication seemingly that pursuit of chemical weapons might suggest a greater likelihood of pursuit of biological weapons.[711]
- Active biodefense programs
  - Active biodefense research programs were also noted. Some teams suggested that the existence of biodefense research suggests a country may also be undertaking offensive research.[712] It was also noted that defensive programs could count towards the question resolving positively.[713]

Comments on specific countries

Most rationales included arguments that may make a state more or less likely to have a bioweapons program, but many teams also commented on each country individually:

---

because of a biased view against these countries coming from their lack of transparency or from their status as geopolitical rivals of the country of origin of a large part of the superforecasters here."

[709] 340, "Armscontrol website lists allegations against all states in question except Syria (though Syria confirmed they had BW in 2012). In other words, all countries had or have (and since they had, they will still have one now albeit under the radar) a bio weapons pgm." 337, "Instead of base rates, what we had are anchors. The most important one is the the summary on biological and chemical weapons in armscontrol.org [1],"

[710] 337, "BW will become more attractive over time as a consequence of developments in biotech: BWs could become more targetable, accessible and resource efficient in the future." 336, "[E]asier access through technical progress (something like CRISPR "do it all at home" stuff) could make this available for more states (make it even easier for NK, Iran, Syria)."

[711] 341, "Syria: Likely. The government of Syria even deployed chemical weapons against its own people as recently as 2013." 341, "The US Congress Office of Technology Assessment has recorded Israel as a country generally reported as having undeclared chemical warfare capabilities, and an offensive biological warfare program."

[712] 340, "Even the U.S. has plenty of biodefense programs that are essentially "dual-use." The only difference between an active biodefense program and a biological weapons program is intent. The US tends to abide by its international treaties...until it decides not to. The BWC is not one that the US has an interest in violating (until it does)."

[713] 338, "[A]djust US upward based on dual use inclusion of partially defensive programs." 337, "[A]lso it is worth quoting "the phrasing, "Dual-use weapons that also have a defensive purpose but could be used for hostile purposes count for the resolution of this question," presents challenges." For instance, the US has a significant defensive need, which could be switched easily to offensive in case of need: would this count? (or again, does it depend on where an expert comes from?) Here this reference [4] (regarding the difficulties in assessing dual-use) has also appeared in the discussion."

- US
  - Forecasters noted the US's previous voluntary destruction of biological weapons, but also noted its ongoing defensive biological research.[714]
- Russia
  - Forecasters noted Russia's previous bioweapons program (as part of the former USSR), their tradition of autocracy and lack of transparency, and that the current war in Ukraine may make them more likely to continue or begin pursuing many types of weapons.[715]
- China
  - Most teams noted that there have not been many allegations of bioweapons programs made against China, although it is involved in biological research with potential dual-use applications.[716]
- North Korea
  - Forecasters noted US allegations that North Korea has an offensive biological weapons program and has been non-compliant with the Biological Weapons Convention (BWC) and South Korea's assertion that North Korea has the ability to produce bioweapons.[717] Forecasters also noted that North Korea has little to lose from admitting to a bioweapons program due to existing sanctions. Some suggested this would be a reason to put a higher probability on North Korea having a bioweapons program,[718] but others noted that the country seems to have preferred to showcase their military capacities, suggesting that they wouldn't maintain a secret program.[719]
- Israel

[714] 337, "US: unilaterally gave up its biological weapons program in 1969. The destruction of all offensive BW agents occurred between 1971 and 1973. It currently conducts research as part of its biodefense program." 338, "US: Dual-use capability. Maintains some old weapons stockpiles."

[715] 341, "Russia: A tradition of autocracy and lack of transparency would suggest that they might still possess biological weapons." 343, "Russia is likely to build its military capabilities in response to an attempt by other countries to prolong the Ukrainian conflict as long as possible. This is expected to lead to increased military capabilities for the nation overall, and a greater suspicion that they have or intend to use biological weapons." 337, "Russia: in 1992, Yeltsin acknowledged that the Soviet Union had pursued an extensive and offensive BW program throughout the 1970s and 1980s."

[716] 343, "China doesn't seem to have many accusations of having a biological weapons program, and they have publicly denied them." 341, "China: "The 2021 State Department Report on Compliance with Arms Control, Nonproliferation, and Disarmament Agreements and Commitments indicates that China is engaged in biological research with "potential dual-use applications.""

[717] 343, "North Korea already has a biological weapons program, according to the US State Department." 337, "North Korea: South Korea asserted that "North Korea likely has the capability to produce[…] anthrax, smallpox, pest, francisella tularensis, and hemorrhagic fever viruses." The compliance report [2] assessed North Korea to be in noncompliance with the BW controls (BWC)"

[718] 338, "North Kore being sanctioned already and knowing that would not be invaded most probably has such program as well, because why not."

[719] 343, "N. Korea seems interested in overt displays of military might, not hidden secret programs.'

    - Forecasters noted Israel's strong military defense programs and perception of being under threat, and that they are not a signatory to the BWC. Some teams suggested Israel previously had an offensive bioweapons program.[720]
- Iran
    - Several teams noted that Iran has made public denouncements of biological weapons, their religious opposition to such weapons, and that the history of being victims of biological weapons may reinforce this opposition.[721] However, teams also noted that Iran is thought to have a defensive biological research program.[722]
- Syria
    - Most forecasters noted Syria's previous use of ricin, with many suggesting that it is likely to continue production of this toxin.[723] One team suggested that ongoing internal conflict will deplete resources and reduce the chance of a bioweapons program being maintained.[724]

Cross-references with other questions

Q21: Number of Countries with Biological Weapons

Question 21: Number of Countries with Biological Weapons Programs

How many countries will have had active biological weapons programs, as estimated by biosecurity experts at some point from January 1, 2022…
…through the end of 2024?
…through the end of 2030?
…through the end of 2050?

[720] 337, "Israel: revealed little about its BW program, but there is belief that Israel has had an offensive BW program in the past (unclear if it still exists now)." 341, "Israel is not a signatory of the Biological Weapons Convention. A small country that sees itself on almost permanent threat, it also has a very advanced biotech industry and a powerful Department of Defense (Military spending in Israel is over 5% of GDP)."
[721] 341, "Iran: Probably not. It has been a very active supporter of the Biological Weapons Convention by being a victim of biological weapons in the past (Iran-Iraq war)." 343, "Iran has strong religious and historical reasons for not wanting to use chemical and biological weapons."
[722] 343, "If they have a defensive biological weapons program allowed under international law that they have developed because of a history of neighbors (with whom they're still unfriendly) using this type of weapon on them, they are unlikely to change this stance."
[723] 337, " Syria: confirmed that it possesses biological warfare materials, it also later declared the existence of production facilities and stockpiles of purified ricin."
[724] 344, "The lowest end forecast is that Syria will not have a biological weapons program due to its own internal strife sapping it of resources."

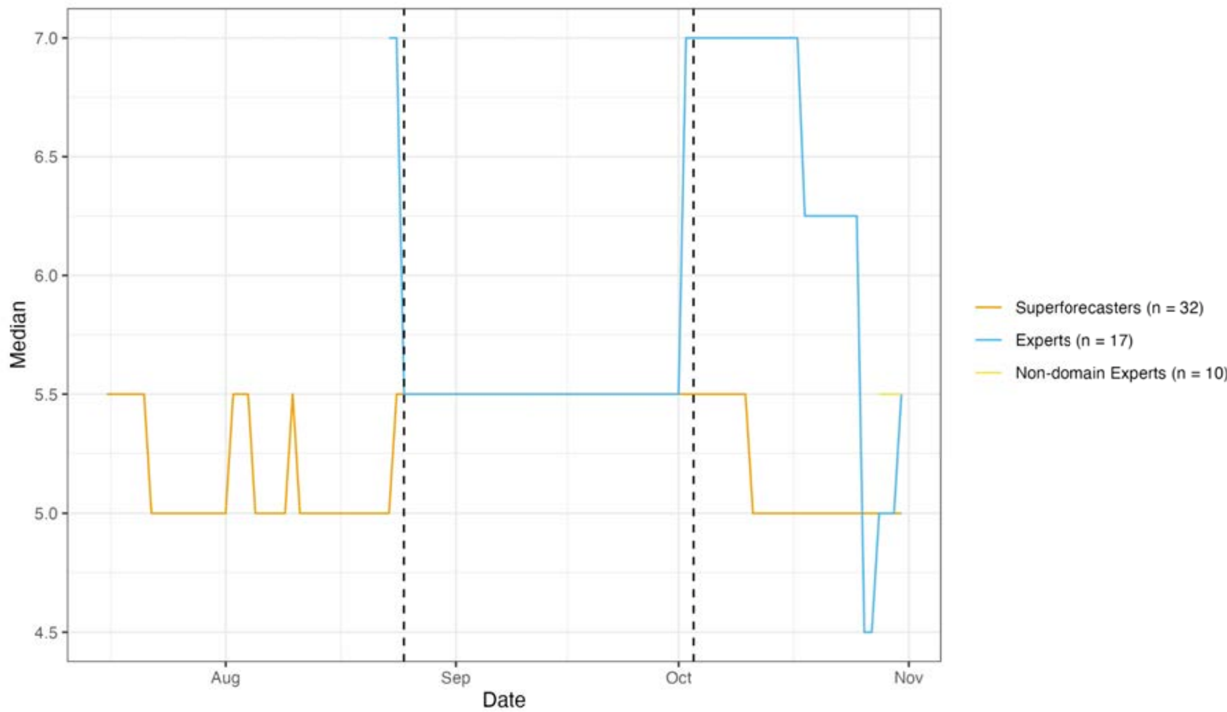
Question and resolution details, prior forecasts, and other relevant sources

Results[725]

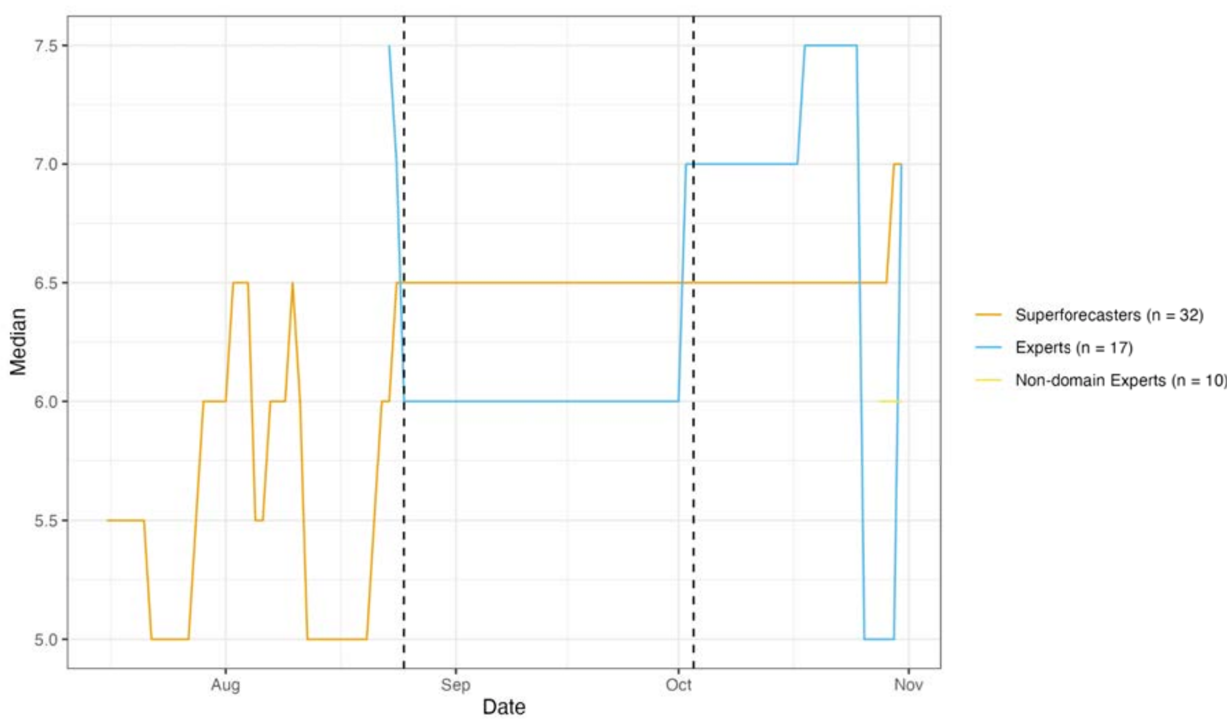
Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 32)	2024	6	5	2.76	-13.74%
	2030	6	7	2.87	-10.9%
	2050	6	7	3.55	-14.2%
Domain Experts (N = 4)	2024	2	6.5	3.21	-32.8%
	2030	4	8.5	4.58	-32.15%
	2050	5	10.5	6.81	-27.04%
General X-Risk Experts (N = 3)	2024	3.5	4	4.95	-70.7%
	2030	3.5	4	4.95	-47.51%
	2050	4	6.5	4.24	-41.07%
Non-Domain Experts (N = 10)	2024	9	5.5	2.83	+15.71%
	2030	9	6	1.41	+124.72%
	2050	11	8.5	1.41	+133.93%

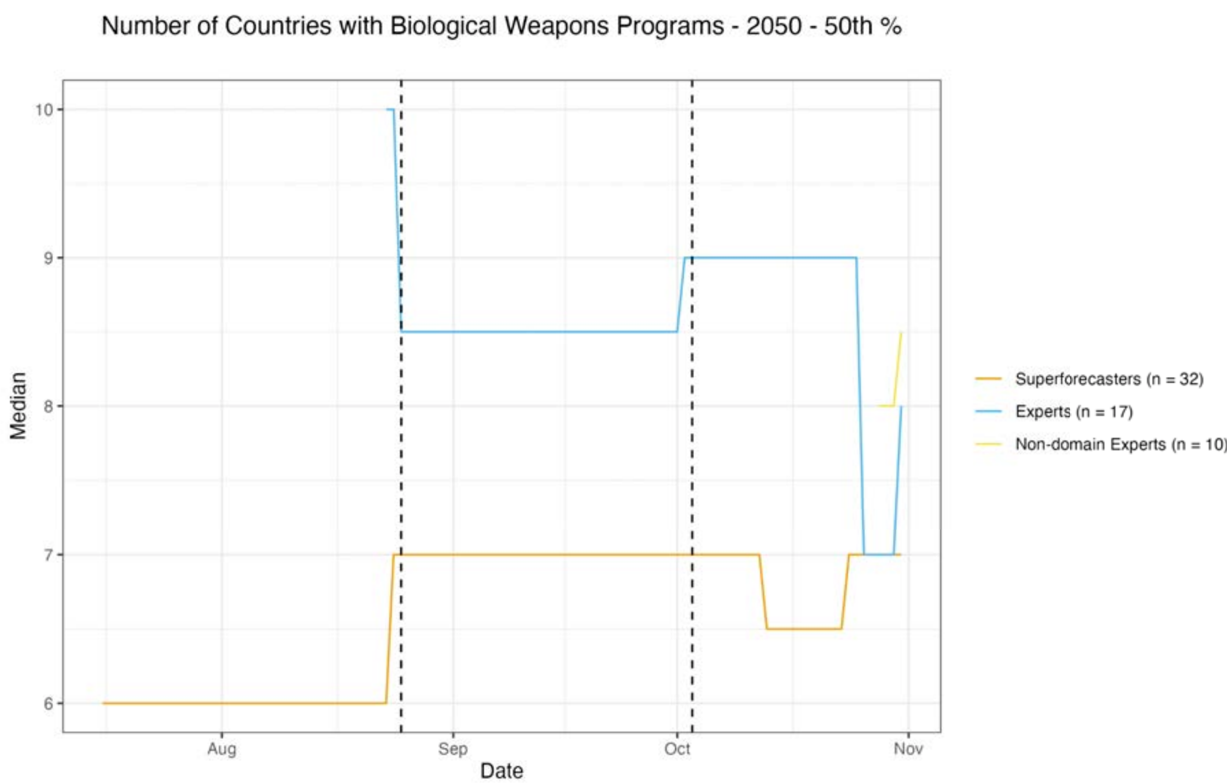
[725] Numbers of forecasters are given as of Stage 4 of the XPT.

Number of Countries with Biological Weapons Programs - 2024 - 50th %

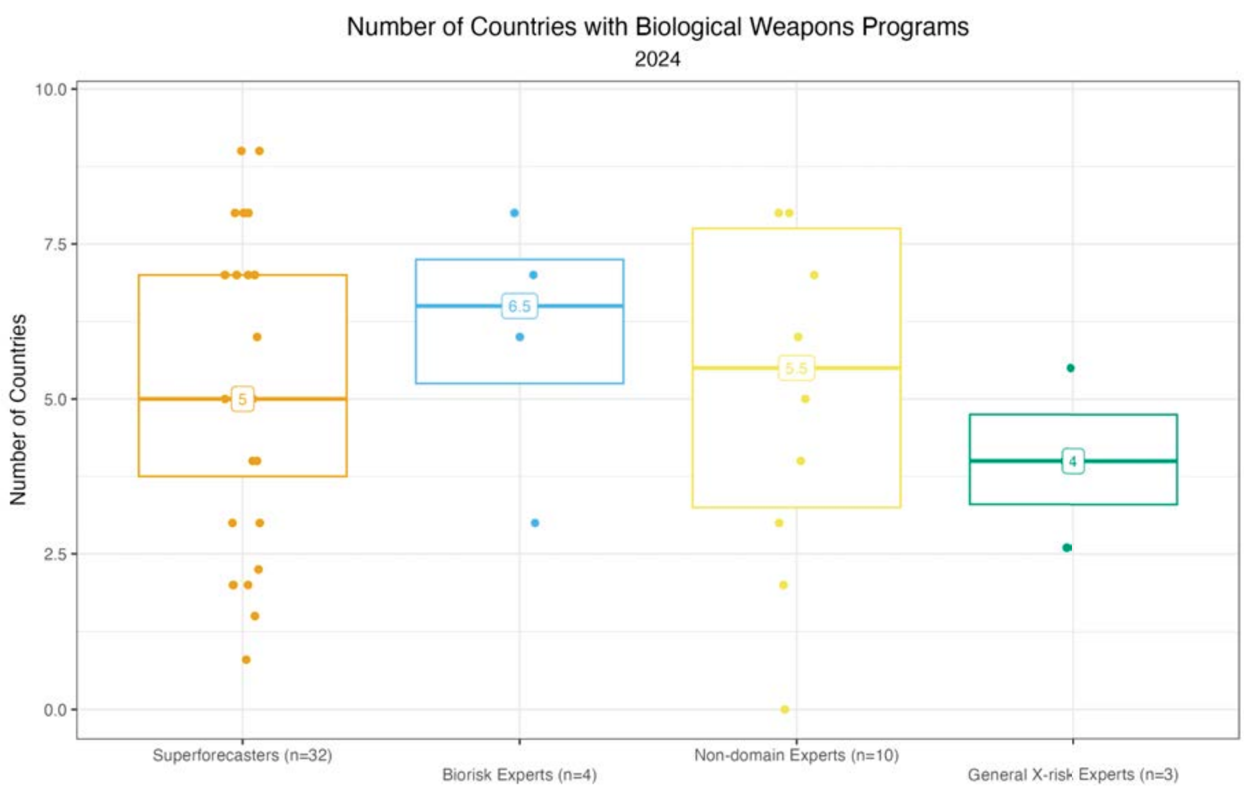


Number of Countries with Biological Weapons Programs - 2030 - 50th %



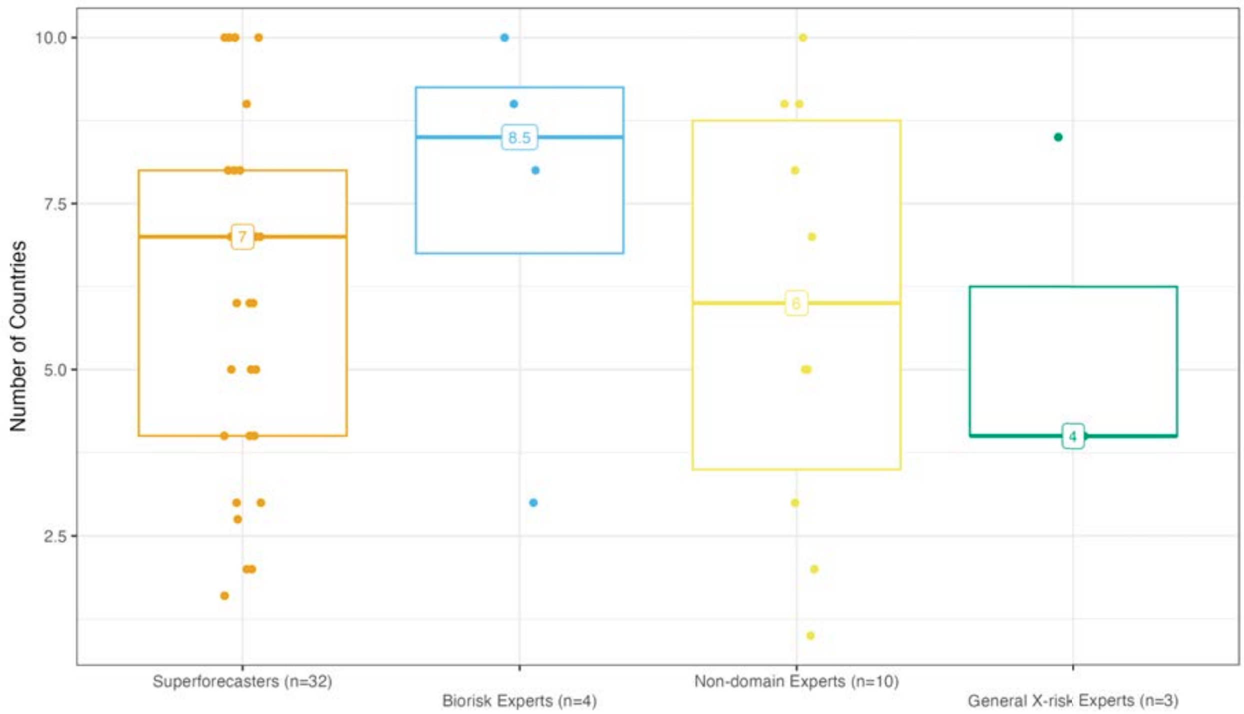


Number of Countries with Biological Weapons Programs - 2050 - 50th %
Median
Date
Aug
Sep
Oct
Nov
Superforecasters (n = 32)
Experts (n = 17)
Non-domain Experts (n = 10)

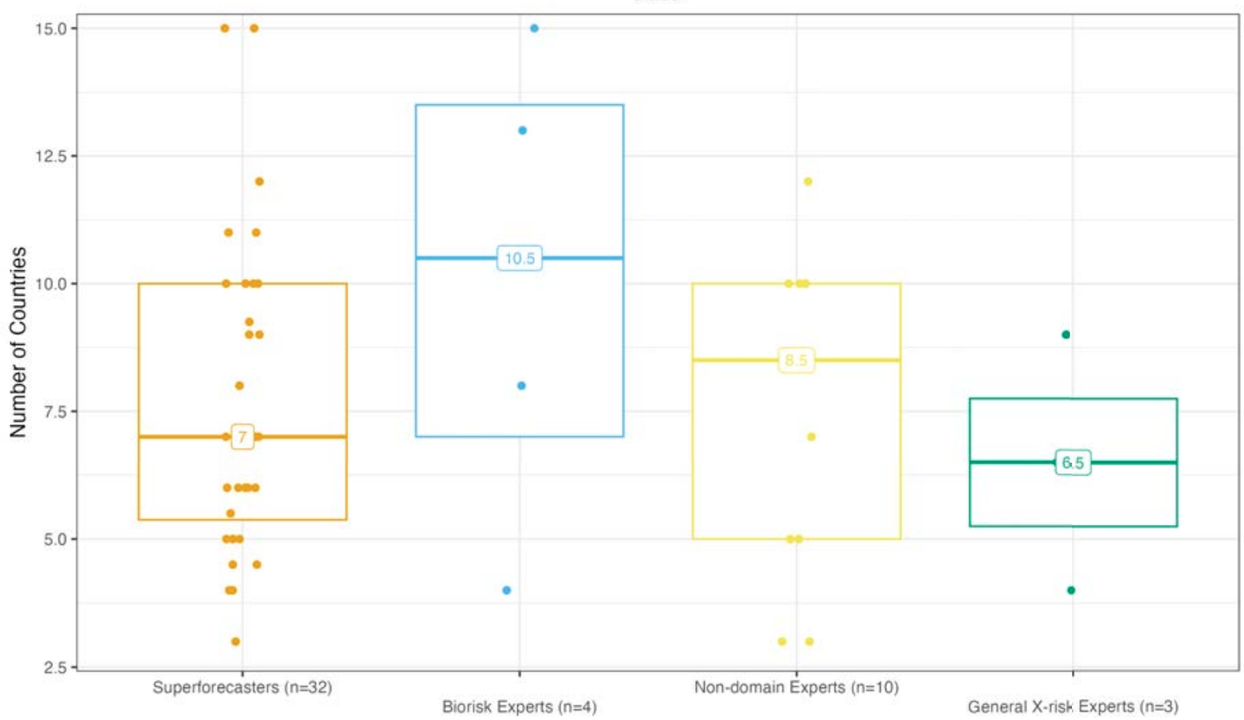


Number of Countries with Biological Weapons Programs
2024
Number of Countries
Superforecasters (n=32)
Biorisk Experts (n=4)
Non-domain Experts (n=10)
General X-risk Experts (n=3)
5
6.5
5.5
4

Number of Countries with Biological Weapons Programs
2030



Number of Countries with Biological Weapons Programs
2050



Sources of agreement, disagreement and uncertainty

Sources of agreement

- Similarity in team median forecasts for 2024 and 2030
    - Most teams had a similar median 50% forecast for the 2024 and 2030 aspects of this question. Seven out of nine teams had a forecast for between 5 and 8 (inclusive) for 2024, and between 6 and 10 for 2030.
    - The median forecasts for 2050 showed greater variation.

Sources of disagreement

- How states will perceive the strategic value of bioweapons programs
    - Teams had different views on whether states would perceive bioweapons as valuable and worth investment. Most believed that most states would see bioweapons programs as having limited utility, as they cannot be controlled and will likely put their own state at risk.[726] However one team assumed that bioweapons programs would be attractive,[727] and others registered uncertainty, particularly in whether states would see bioweapons as a useful deterrent, similar to nuclear weapons.[728] Some teams also noted that rogue or undemocratic states might see bioweapons as more attractive.[729]

Sources of uncertainty

- What types of activities would count as bioweapons programs
    - Several teams noted uncertainty in how a bioweapons program would be defined, particularly when dual-use research or defensive capabilities would be perceived as bioweapons.[730]
- Currently existing biological weapons programs

---

[726] 343, “Their high costs are often not justified by their low efficacy. Infectious biological weapons carry significant blowback potential, both directly as the agent may infect a nation's own people, and by bringing scrutiny from the international community.” 339, “Bioweapons programs have been generally found to be not useful, imposing as much self-harm and utility.”

[727] 344, “The assumption is if the countries are capable there is a high chance they will engage in such programs.”

[728] 341, “Bio-technology CRISPR, etc. will become so accessible that many countries will want a 'defensive' program.” 338, “Significant disagreements, dissenting views within the team, and major sources of uncertainty: [...] Whether bioweapons offer a cheap alternative "porcupine" defence to nuclear weapons and the imapct of the UARU conflict for countries near neighbours that may invade for various reasons.”

[729] 341, “[A]nother view within our team is that since technological advances will make biological weapons attainable at low cost and difficult to identify, biological weapons programs will become common in rogue states.” 339, “[...] is the biggest source of uncertainty. Followed by how many non-transparent governments or autocrats might view bioweapons programs as useful. ”

[730] 336, “The definition of a biological weapons program is unclear” 338, “Significant disagreements, dissenting views within the team, and major sources of uncertainty: [...] (Re) Classification of what is a bioweapon and potential for a "low bar" to trigger this”

    - Many teams highlighted their uncertainty in whether there are bioweapons programs in existence today.[731]
- How the expert panel would deliberate
    - Several teams noted uncertainty in how the expert panel would deliberate, including whether they would rely on open source intelligence and whether politics may influence their assessments.[732]

Arguments given for forecasts below the median of 5 (2024), 7 (2030), 7 (2050)

- Bioweapons programs have limited strategic value to states
    - Several teams noted that bioweapons programs have historically been ineffective and many countries have previously abandoned bioweapons programs.[733]
- There are no confirmed bioweapons programs at present
    - One team pointed to the absence of any confirmed bioweapons program at present as a reason for a lower forecast.[734]
- Bioweapons programs are likely to cause international condemnation
    - Some teams suggested that states would be dissuaded from developing bioweapons programs due to possible international condemnation.[735] Similarly, one team suggested that the use of a bioweapon may lead to a disarmament movement, reducing the number of states with bioweapons programs.[736]
- Biosecurity experts may be cautious in suggesting a state has a bioweapons program
    - One team suggested that biosecurity experts may be cautious in naming a state as having a bioweapons program.[737]

---

[731] 339, "How many BSL-4 labs actually hold bioweapons programs is the biggest source of uncertainty." 341, "We have significant uncertainty how many countries have an active biological weapons program today (and would be named as having one in 2024)."
[732] 336, "Expected sources of disagreement between your team, an independent group of superforecasters, and an independent group of domain-specific experts: Composition of the expert panel" 338, "Significant disagreements, dissenting views within the team, and major sources of uncertainty: [...] Judges having a particular political aspect to their decisions. Countries having the capability but hiding that ability and whether judges would rely on OpenSource Intelligence"
[733] 339, "Bioweapons programs have been generally found to be not useful, imposing as much self-harm and utility. Most countries who had a capability have disbanded it and signed the BWC." 343, "Their high costs are often not justified by their low efficacy. Infectious biological weapons carry significant blowback potential, both directly as the agent may infect a nation's own people, and by bringing scrutiny from the international community."
[734] 336, "The current number of countries is zero, according to Wikipedia"
[735] 343, "Biological weapons programs carry strong condemnation from the international community."
[736] 340, "The strongest argument for low extreme forecasts in 2030 and 2050 is that bioweapons are used, leading to a disarmament movement."
[737] 340, "The strongest argument for the 2024 low extreme forecast is that biosecurity experts are more cautious in labeling states with active bioweapons programs."

Arguments given for forecasts above the median of 5 (2024), 7 (2030), 7 (2050)

- Several countries have had bioweapons programs or have had credible allegations made against them of having bioweapons programs
    - In support of a higher forecast, several teams noted that there have been several countries who have either admitted to a bioweapons program or have faced (credible) allegations of running a program, although teams varied in the number of countries they included in this category.[738]
- Bioweapons becoming available to a greater number of countries
    - Several teams suggested that technological progress may make it easier and/or cheaper to develop bioweapons, making them available to a greater number of countries.[739]
    - Similarly some teams noted that economic development may also make bioweapons available to more countries.
- Bioweapons may have some strategic uses
    - One team suggested bioweapons may be useful for acting as a deterrent against aggressors, similar to nuclear weapons, or that technological advancements may allow for more targeted bioweapons.[740]
- There are many high containment labs
    - One team suggested that the growing number of high containment laboratories globally suggests that more countries may be developing (or may develop) bioweapons.[741]
- Biosecurity experts may have additional information
    - One team suggested that biosecurity experts might have access to more information sources that would give them greater knowledge of bioweapons than can be gleaned from publicly available information.[742]

Other arguments given

Arguments in favor of lower forecasts:

- Bioweapons programs have limited strategic value to states

---

[738] 339, “16 countries have admitted to bioweapons programs in history, but many have publicly disbanded them and signed the BWC.” 338, “In terms of a base rate the highpoint of countries with biological weapons programmes was in 1990 (8 countries)”
[739] “341, “Bio-technology CRISPR, etc. will become so accesible that many countries will want a 'defensive' program” 344, “If we look at the extended time horizon we know that price of synthesizing, along with the economics of scale, will lower the price needed to "have" such a site.”
[740] 336, “[B]iological weapons could be a good 'poor man's alternative' for nukes, so I think more countries will be tempted.” and “that technological advancements may allow for more targeted bioweapons.”
[741] 339, “59 BSL-4 labs in 23 countries in 2021, with 14 claiming a program at one point prior to 2022.”
[742] 340, “The strongest argument for the 2024, 2030 and 2050 high extreme forecasts is that biosecurity experts have knowledge (e.g., through government contacts) that is not accessible to forecasters.”

    - One team suggested limited utility of bioweapons as an argument for pushing the forecast lower, even for a forecast above the median.[743]
- International taboo on bioweapons use
    - One team noted an international taboo on bioweapons use may dissuade states from developing these weapons.[744]

Arguments in favor of higher forecasts:

- Increasing geopolitical tensions may make bioweapons more attractive to states
    - One team suggested that increasing geopolitical tensions may result in more countries seeking bioweapons.[745]
- It is difficult to monitor bioweapons programs
    - One team noted that bioweapons may be attractive to states as they are difficult to identify.[746]

Cross-references with other questions

Q20: Individual Countries with Biological Weapons

Question 22: PHEIC Declarations with 10k Deaths

How many times will the WHO declare a new Public Health Emergency of International Concern (PHEIC) for a disease that will be the cause of death of at least 10,000 people…
…by the end of 2024?
…by the end of 2030?
…by the end of 2050?

Question and resolution details, prior forecasts, and other relevant sources

---

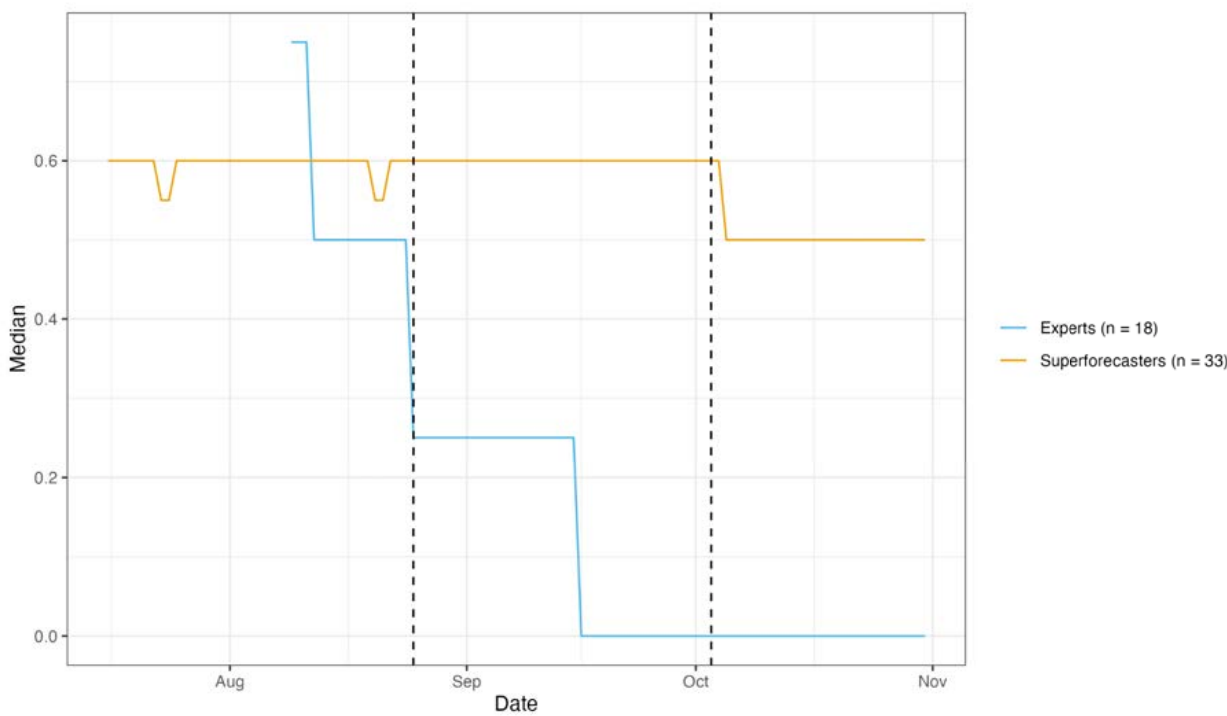
[743] 341, “We believe that this is explained by the theoretical value of biological weapons for military objectives being low. Therefore, even if it becomes much cheaper to gain biological weapons technology, it will likely not be a focus of national defense.”
[744] 343, “Biological weapons also carry a taboo similar to nuclear weapons and countries which pursue them may risk being labeled as rogue states.”
[745] 337, “The expectation that geopolitical tensions will increase in the next few decades”
[746] 337, “Effective monitoring impractival”

Results[747]

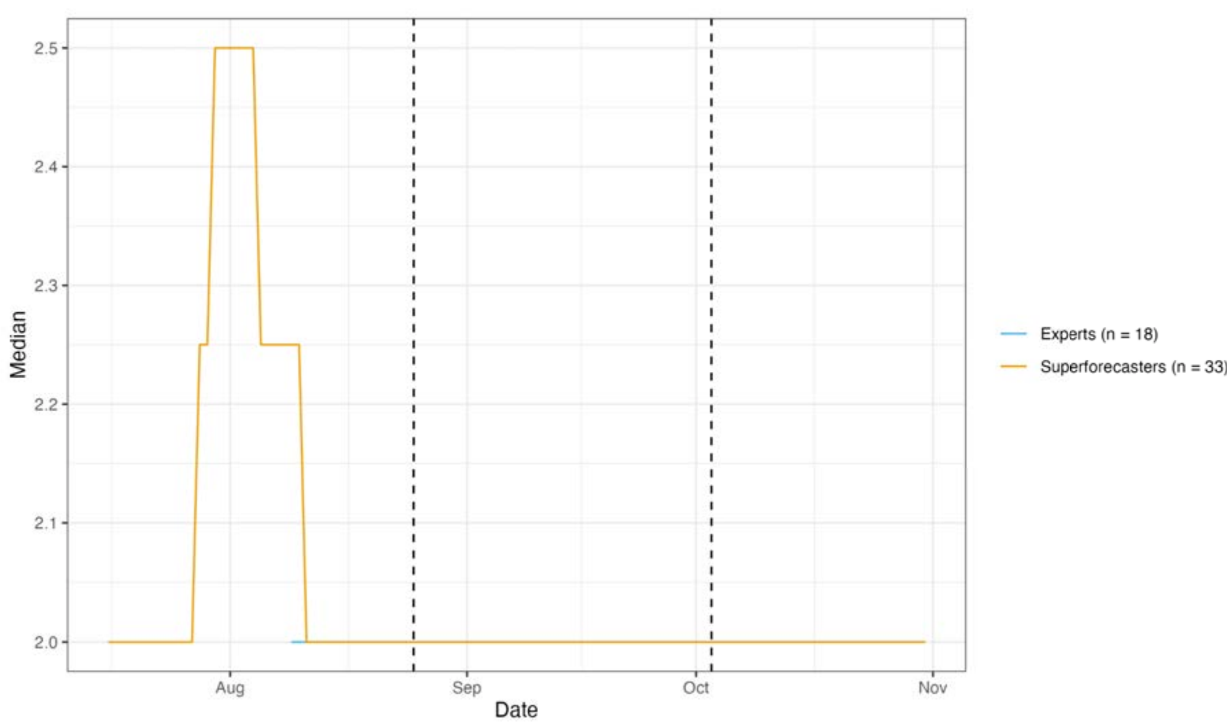
Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 33)	2024	0.6	0.5	0.56	+26.52%
	2030	2.25	2	2.09	-44.58%
	2050	8	7	7.19	-42.44%
Domain Experts (N = 5)	2024	1	0	0.55	-18.35%
	2030	1	1	0.89	-51.59%
	2050	6	5	2.88	-33.23%
General X-Risk Experts (N = 4)	2024	0.5	0.27	0.5	-3.69%
	2030	2	2	1.22	-16.03%
	2050	6.5	6.95	1.04	-21.41%
Non-Domain Experts (N = 9)	2024	0	0	1.15	-52.76%
	2030	3	2	0.58	+22.47%
	2050	9	6	1	+58.11%

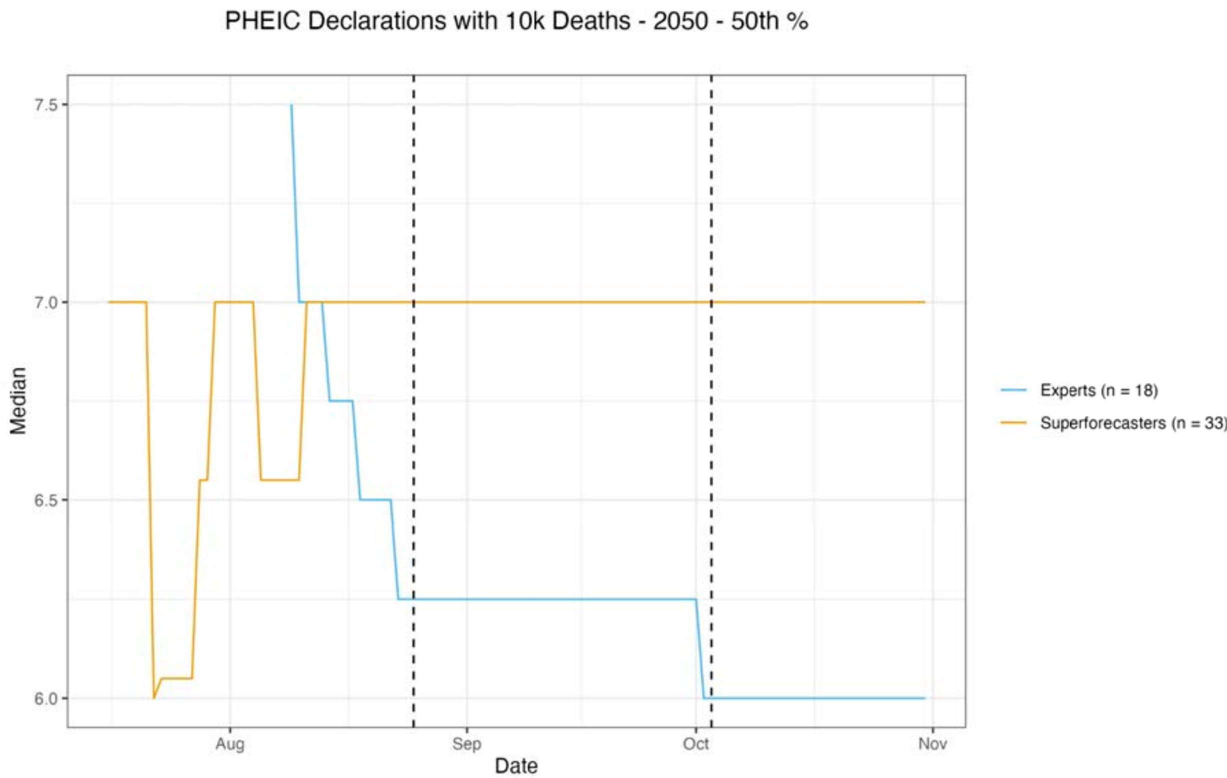
[747] Numbers of forecasters are given as of Stage 4 of the XPT.

PHEIC Declarations with 10k Deaths - 2024 - 50th %

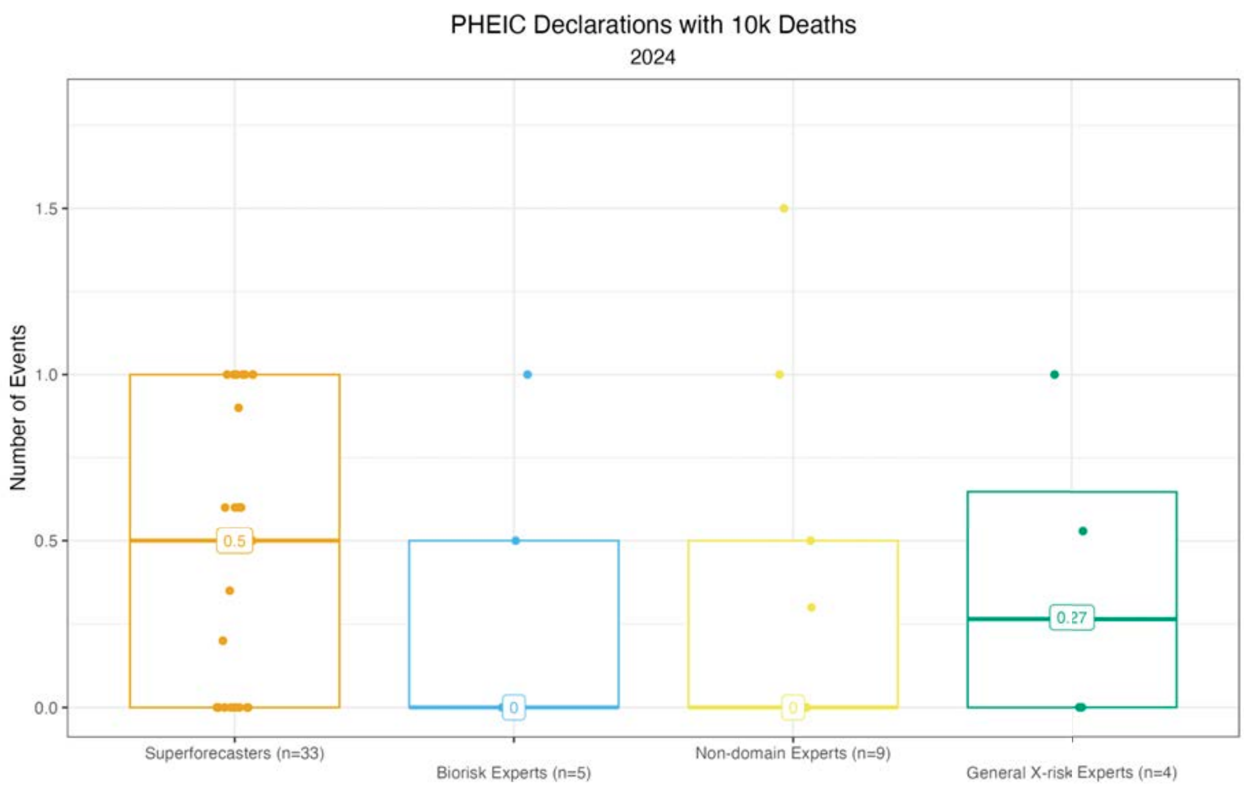


PHEIC Declarations with 10k Deaths - 2030 - 50th %

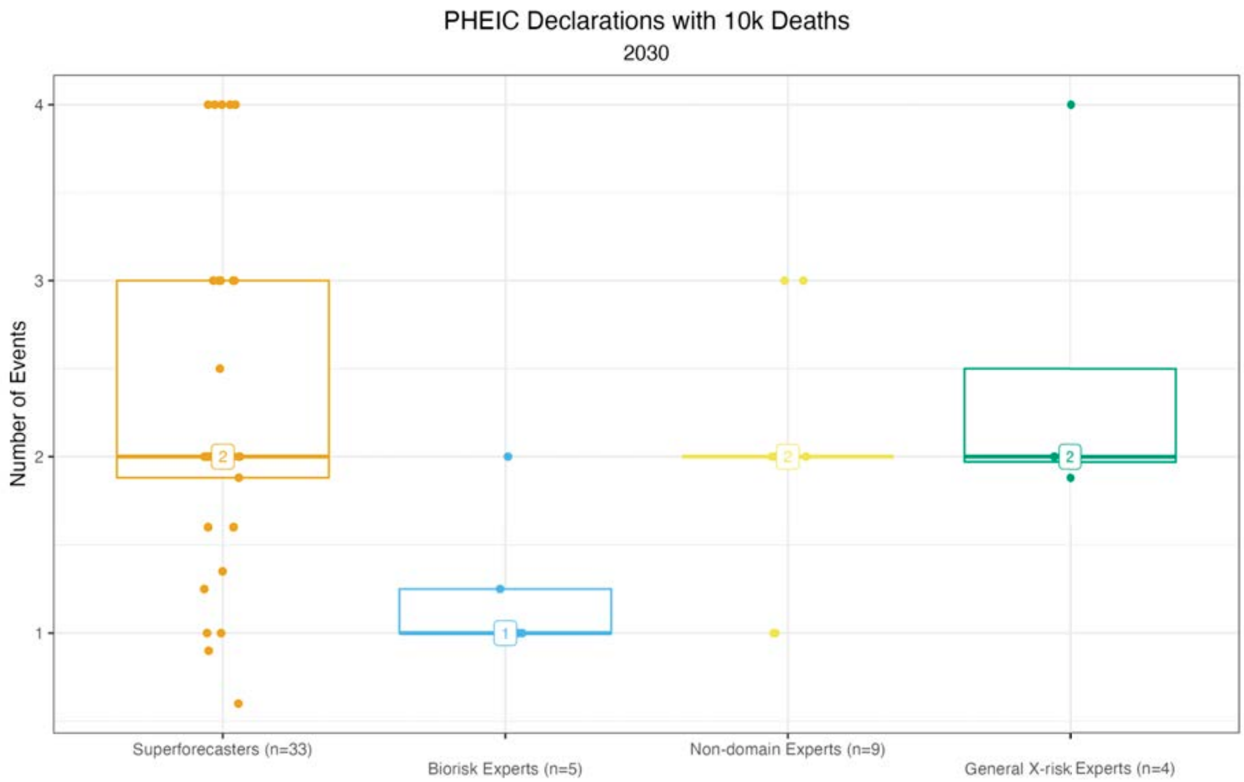


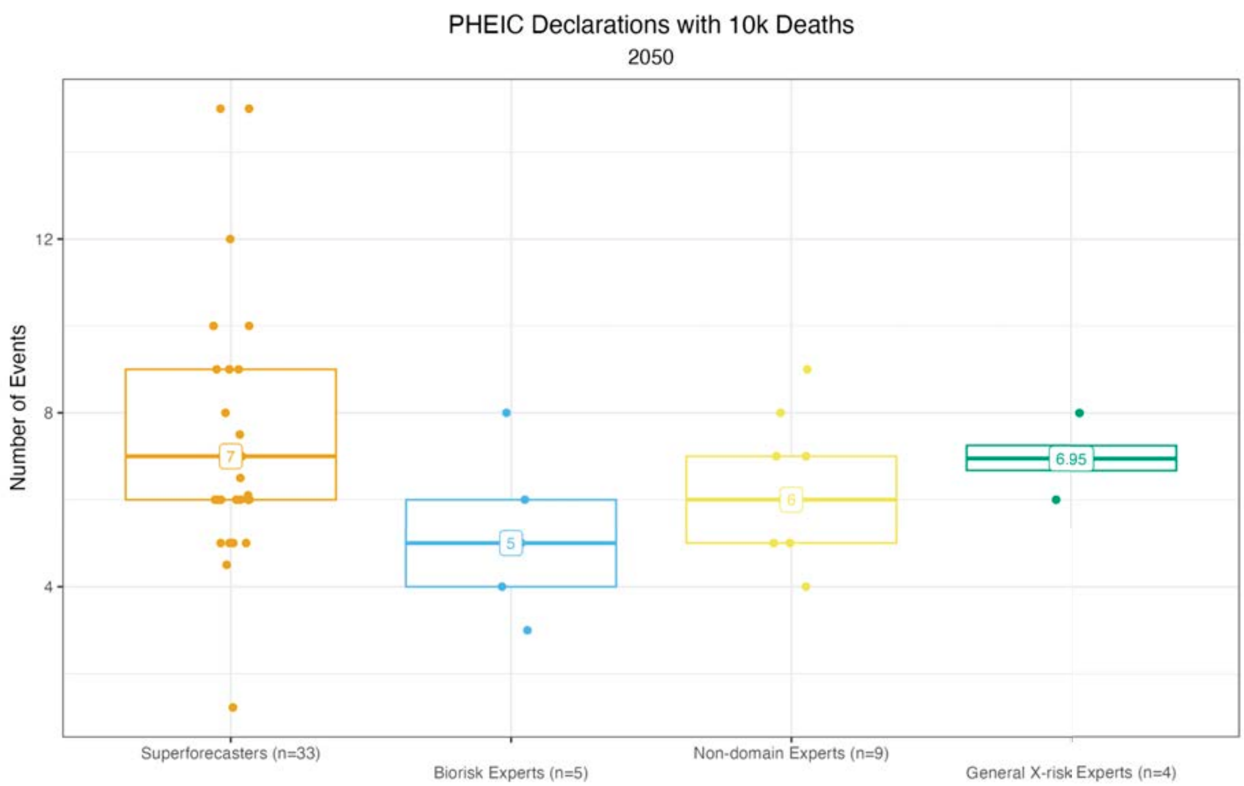


PHEIC Declarations with 10k Deaths - 2050 - 50th %
Median
7.5
7.0
6.5
6.0
Aug
Sep
Oct
Nov
Date
Experts (n = 18)
Superforecasters (n = 33)



PHEIC Declarations with 10k Deaths
2024
Number of Events
1.5
1.0
0.5
0.0
0.5
0
0
0.27
Superforecasters (n=33)
Biorisk Experts (n=5)
Non-domain Experts (n=9)
General X-risk Experts (n=4)


PHEIC Declarations with 10k Deaths
2030
Number of Events
4
3
2
1
2
1
2
2
Superforecasters (n=33)
Biorisk Experts (n=5)
Non-domain Experts (n=9)
General X-risk Experts (n=4)


PHEIC Declarations with 10k Deaths
2050
Number of Events
12
8
4
7
5
6
6.95
Superforecasters (n=33)
Biorisk Experts (n=5)
Non-domain Experts (n=9)
General X-risk Experts (n=4)

Sources of agreement, disagreement and uncertainty

There was agreement that the world was becoming more interconnected, which can impact the frequency of diseases.

However, there was disagreement about:

1. Whether interconnectedness increased or decreased the likelihood of disease (though the lower side of this disagreement was only represented in “other arguments”).
2. Also, there was disagreement on whether known base rates implied a higher or lower forecast (though the higher side of this disagreement was only represented in “other arguments”).

Arguments given for forecasts ≤ 0.6 (2024), 2 (2030), 6 (2050)

On 2 (base rate):

- A disease like COVID-19 seems to appear about once every 50 years.[748]
- Base rate for qualifying events is 25% chance each year, which was adhered to without adjustment by one team with a lower forecast.[749]

Improvements in medical technology

- Improvements in medical technology could decrease the deadliness of contagions.[750]

Governmental responses

- Effective governmental responses will reduce fatalities.[751]
- Post-COVID improvements in monitoring and response may result in fewer and less deadly pandemics.[752]

Arguments given for forecasts ≥ 0.8 (2024), 2.5 (2030), 6.5 (2050)

On 1 (interconnectedness):

- Increasing viral diseases due to population interconnectedness.[753] [754]

---

[748] 342, "Also looking at the past a disease such as COVID seems to appear around every 50 years."
[749] 341, "Using the Laplace rule of succession with 3 qualifying events in the last 14 years arrives at a base rate of 25% chance a year, and calculating a binomial distribution arrives at a very similar forecast to the team's forecasts. The team doesn't see any particularly strong reasons to deviate from the base rate."
[750] 340, "[I]mprovements in medical technology and disease monitoring could decrease the deadliness of future contagions."
[751] 340, "The lower end of the forecasts reflects the probability of robust governmental responses to PHEIC declarations moving forward that would keep fatalities lower than the >10000 threshold."
[752] 341, "Some expect the COVID pandemic to result in tighter monitoring and better response to future pandemics, which could produce somewhat fewer and less deadly pandemics, at least in the short term."
[753] 343, "This is based on the expectation that the number of viral diseases is likely to increase with increasing population interconnectedness.”
[754] 336, "An increasingly interconnected and pathogen-friendly world making disease more likely."

Reduced vaccination rates
- Rates of childhood vaccinations glo
- bally are declining, which will lead to more PHEIC events.[755]

Anthropogenic risks
- Rising likelihood of bioweapons deployment increases risk.[756]
- New genetically engineered pathogens may increase risk.[757]
- The number of labs working with viruses is expected to increase, potentially raising the chance of a lab leak and increasing expected events.[758]

Non-anthropogenic risks
- A possible escalation of monkeypox increases risk.[759]

Technicality regarding question resolution:
- Our improved ability to detect diseases may result in a higher frequency of PHEIC events due to detecting diseases we wouldn't have otherwise detected in developing countries.[760]

Other arguments given

Arguments in favor of lower forecasts:

On 1 (interconnectedness):
- Greater interconnectedness reduces the likelihood of well-adapted historical zoonotic spillover events.[761]

Improved global economic development reduces zoonotic crossover events
- Relatedly, global economic development will likely reduce exposure of humans to animals, reducing zoonotic disease crossover events.[762]

---

[755] 338, "Last month the WHO and Unicef released data showing the largest sustained decline in childhood vaccinations in three decades, with at least 25mn infants missing out on life-saving jabs in 2021."
[756] 344, "For the year 2050: The team's median forecast for this year is 8. The same reasoning applies as with 2024, and 2025 but now forecasters have taken into account bioweapons deployment."
[757] 343, "One of the novel factors that may lead to a higher number of viral diseases is new genetically engineered pathogens."
[758] 338, "The number of labs working with viruses is expected to increase which could increase the chance of a lab leak, also leading to an increase in expected events."
[759] 344, "For the year 2024: The team's median forecast for this year is at most 2. One of the worrying factors for the short term was the existence of Monkey Pox and the possible escalation of the disease."
[760] 343, "[O]ur awareness and detection of diseases is increasing (we can count fatalities more accurately, so it's more likely that an epidemic in an undeveloped country will be declared a PHEIC and the number of deaths will not be undercounted)."
[761] 343, "The more international connectedness, the less likely a well-adapted historical zoonotic spillover event will be discovered."
[762] 343, "Additionally, global economic development will likely reduce the exposure of humans and animals, reducing the propensity for emergence of zoonotic disease crossover events."

Improvements in medical technology and response:

- Better awareness of breaking transmission cycles.[763]
- Improved flu vaccine accuracy will reduce flu-related events.[764]
- Reduced development times for vaccine production reduces risk.[765]

Arguments in favor of higher forecasts:

On 2 (base rate):

- Past occurrences of PHEIC declarations with over 10,000 deaths—six events between 2009 and 2022—indicate a pattern that is likely to continue into the future, leading to a higher forecast.[766]

Climate change and human-animal conflicts:

- Climate change and human-animal conflicts may increase disease occurrence and zoonotic jumps.[767]

Cross-references with other questions

Q14: Novel Infectious Disease Surveillance System
Q24: Malaria Deaths

Question 23: Assassinations with Biological Weapons

What will be the expected number of events in which country leaders are assassinated by a biological weapon involving a contagious agent…
…by the end of 2024?
…by the end of 2030?
…by the end of 2050?

Question and resolution details, prior forecasts, and other relevant sources

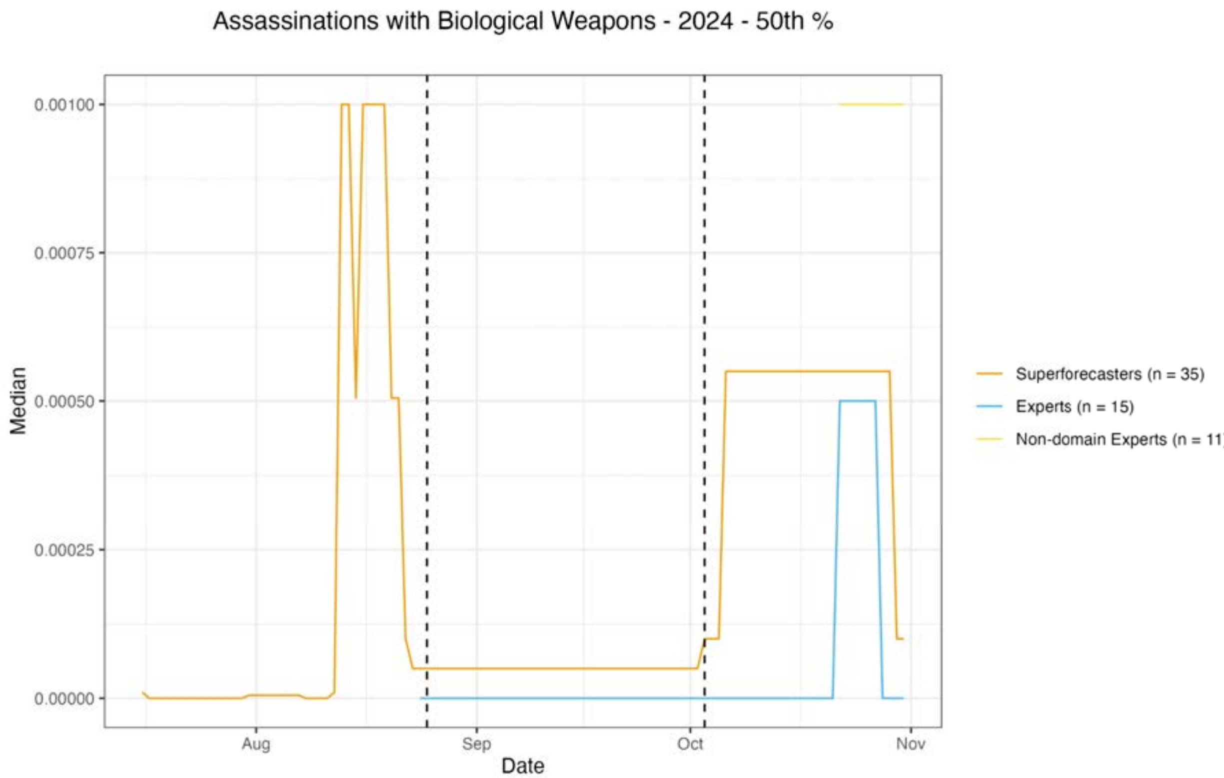
---

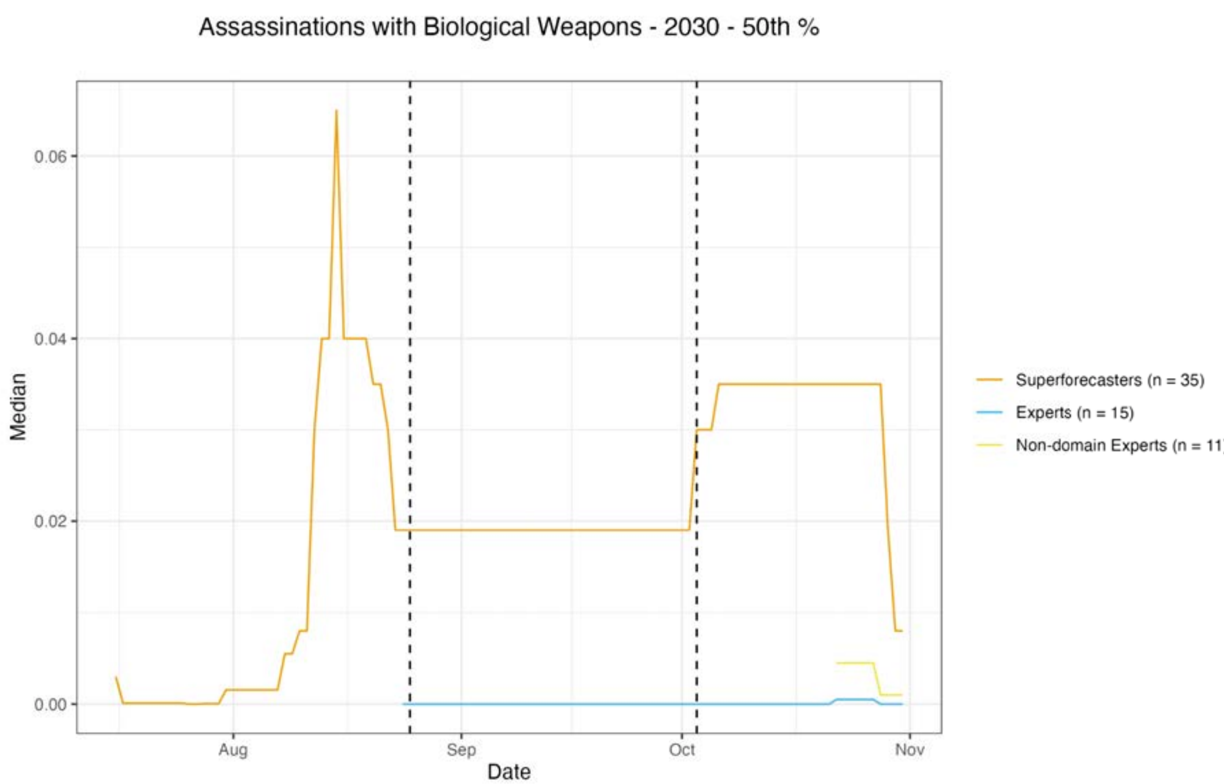
[763] 336, "Better awareness of how to break transmission cycles should help."
[764] 336, "Which strains the flu vaccine contains can likely be more accurate in the future."
[765] 336, "The lead times for vaccine production is falling."
[766] 342, "Considering that from 2009 to today PHEIC was declared 6 times, with 3 of them resulting in more than 10000 deaths and with no reason to believe this decade was non typical, this pattern is likely going to continue in the future."
[767] 340, "Global climate change and increasing human-animal conflict over natural spaces would increase the incidence of cross-species infection and vector-borne diseases."

Results[768]

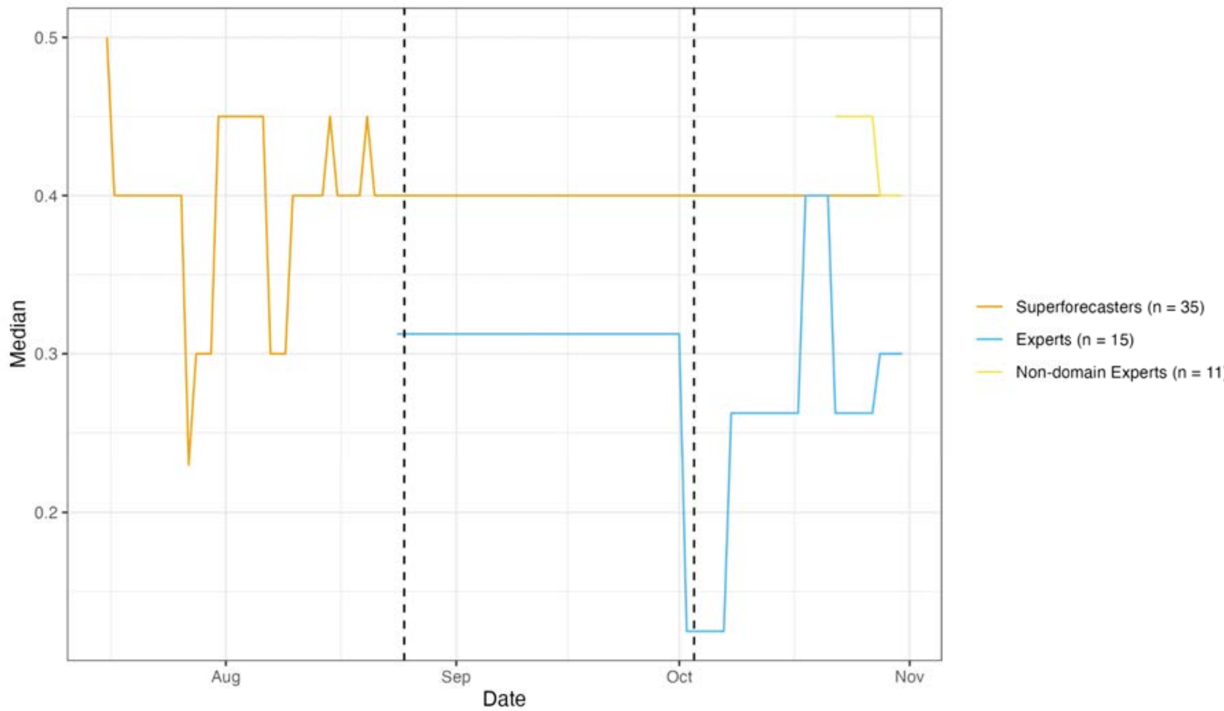
Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 35)	2024	0.001	0.0001	0.14	+72.23%
	2030	0.04	0.008	1.02	-75.11%
	2050	0.5	0.4	2.26	-70.48%
Domain Experts (N = 3)	2024	0	0	0	Inf
	2030	0	0	0.00	Inf
	2050	0	0	0.00	Inf
General X-Risk Experts (N = 1)	2024	n/a	0	n/a	n/a
	2030	n/a	0	n/a	n/a
	2050	n/a	0	n/a	n/a
Non-Domain Experts (N = 11)	2024	0.04	0.001	0.28	+7.89%
	2030	0.34	0.001	0.51	-38.11%
	2050	2	0.4	2.46	-27.8%

[768] Numbers of forecasters are given as of Stage 4 of the XPT.


Assassinations with Biological Weapons - 2024 - 50th %
0.00100
0.00075
0.00050
0.00025
0.00000
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 35)
Experts (n = 15)
Non-domain Experts (n = 11)

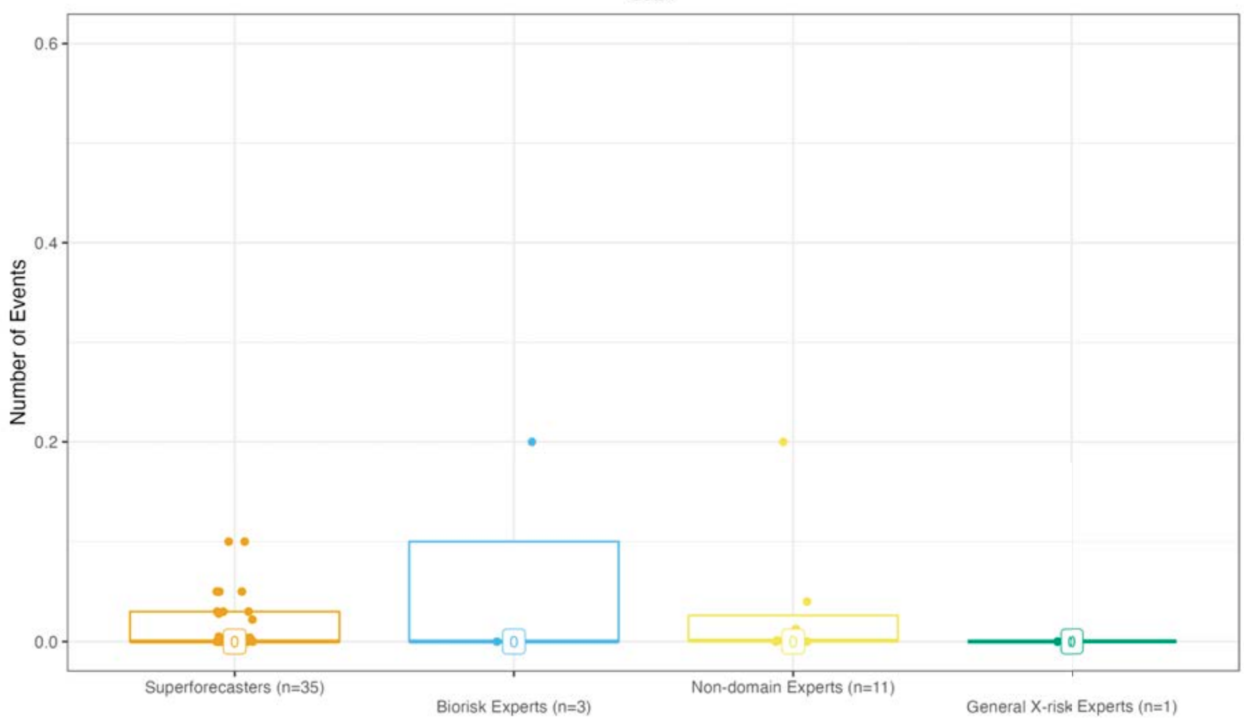

Assassinations with Biological Weapons - 2030 - 50th %
0.06
0.04
0.02
0.00
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 35)
Experts (n = 15)
Non-domain Experts (n = 11)

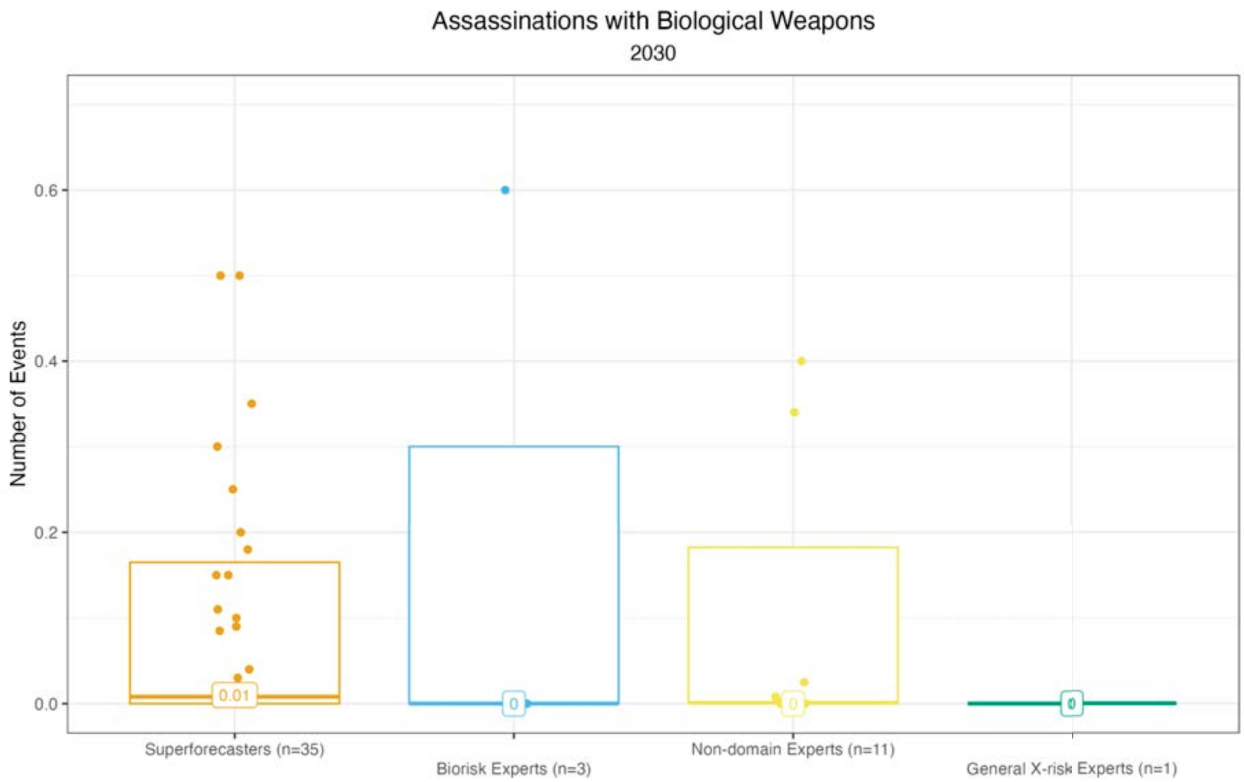
Assassinations with Biological Weapons - 2050 - 50th %

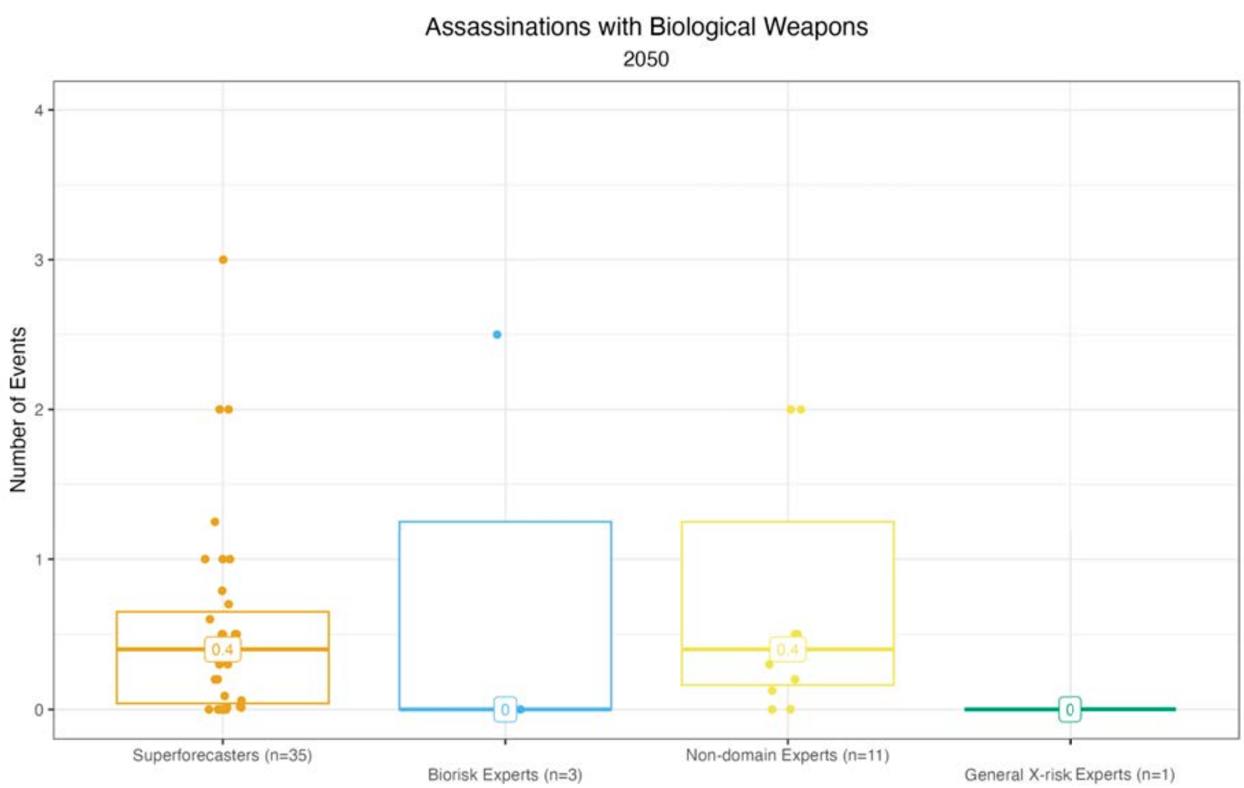


Assassinations with Biological Weapons

2024




Assassinations with Biological Weapons
2030
Number of Events
0.0
0.2
0.4
0.6
0.01
0
0
0
Superforecasters (n=35)
Biorisk Experts (n=3)
Non-domain Experts (n=11)
General X-risk Experts (n=1)


Assassinations with Biological Weapons
2050
Number of Events
0
1
2
3
4
0.4
0
0.4
0
Superforecasters (n=35)
Biorisk Experts (n=3)
Non-domain Experts (n=11)
General X-risk Experts (n=1)

Sources of agreement, disagreement and uncertainty

Sources of agreement:

- Generally teams had low forecasts for this question, with seven of nine team medians below 0.01 for 2024 and 2023, and below 1 for 2050. However, some teams noted that there was a wide range of forecasts, with arithmetic means higher than the median, suggesting that some participants had much larger forecasts.[769]
- Most teams agreed that a contagious biological weapon would be an unlikely choice for an assassination.

Sources of disagreement:

- Base rates
    - Some forecasters used a base rate of zero, given there have been no confirmed assassinations with a contagious biological weapon.[770] However, some others considered possible, unconfirmed events, such as deaths of sitting heads of states that were assumed to be of natural causes.[771]

Sources of uncertainty:

- How technology will change bioweapons capabilities
    - Several teams noted uncertainty in biotechnological progress, and how this would affect the accessibility and use of biological weapons. Several forecasters suggested that progress may enable biological weapons targeted at individuals,[772] although one team noted that “informed forecasters noted it would be highly technically challenging.”[773]
- Scope of question
    - One team expressed uncertainty at whether a biological agent targeted at an individual (through hypothetical DNA targeting) would count for the purposes of this question.[774]

---

[769] 337, “...[A]verage forecasts for 2030 and 2050 were higher than median and prediction ranges were usually substantial [...], reflecting the fairly broad distributions of predictions across forecasters.” 341, “The team's forecasts were widely scattered”

[770] 338, “The majority of forecast reflected the current base rate of 0, with the result that many percentiles collapse in the same zero number”

[771] 337, “We also note that many heads of state have died from various illnesses, which could potentially be from some form of bioweapon exposure.”

[772] 340, “The best argument discussed in favour of assassinations by contagious agent is that biotechnology innovation will enable targeted pathogens to be developed, and that it will be possible to deliver them safely and reliably.” 337, “ By 2050, (assuming the frequency of assassinations does not change), I think this rate might double with advances in synthetic biology, so maybe 1/50. By 2100, let's call it 1/10 (imagining advances in targeting to specific genetic profiles, but also in defensive technologies).”

[773] 340, “It is conceivable that such a weapon may be developed given advances in biochemistry, although informed forecasters noted it would be highly technically challenging.”

[774] 344, “One issue I have with the wording of this question is that it isn't clear whether a biological agent targeted at the DNA of a specific individual would count here.”

Arguments given for forecasts below 0.01 (2024), 0.01 (2030), 0.5 (2050)

- Assassinations of heads of state/government are infrequent
  - Several teams noted that there are few assassinations of heads of state or government, particularly in recent decades.[775]
  - Some teams also argued that heads of state and government are generally very well guarded with security personnel and measures.[776]
- Bioweapons attacks are infrequent
  - One team noted that the historical rate of bioweapons attacks has been low.[777]
- Bioweapons make an unlikely assassination weapon
  - Many teams argued that biological weapons are an unlikely weapon of choice for an assassination as they are:
    - Difficult and expensive to develop[778]
    - Currently not capable of being targeted to an individual[779]
    - Unreliable in effect[780]
    - Likely to draw international condemnation[781]
    - Worse than other alternatives[782]

Arguments given for forecasts above 0.01 (2024), 0.01 (2030), 0.5 (2050)

- There may have been unidentified assassinations or attempted assassinations with bioweapons in the past

---

[775] 337, “The strongest argument for a near zero probability here (close to the media forecast) is that the number of heads of state being assassinated in any method during the previous few decades is fairly low.” 340, “Historic data suggests an average assassination rate of 0.6 p.a. (1900 onwards) with only 0.1 p.a. from 2010-2020 and a peak of 1.3 per year in the 1970's (data collated from Wikipedia).”

[776] 342, “[T]he position of the president or leaders is highly protected and the most protected position in any country” 340, “Along with political leaders being a small class of people with extensive security making assassinations quite difficult to execute”

[777] 340, “There have been 7 biological attacks in 70 years, but only 4 excluding Anthrax which is not contagious and none of these attacks were assassinations.”

[778] 336, “This is an expensive/complicated way of killing someone, although, this could be the way for a high-profile head of state. Still very unlikely.” 340, “We believe the low incidence of contagious attacks is because lethal contagious agents are difficult to engineer/source, require expertise to store and deliver, and are currently difficult to target/control.”

[779] 336, “Even with concern about non-state actors, something that is contagious presents targeting difficulties.” 343, “Contagious diseases - especially highly-contagious and/or quickly-lethal ones - are a ridiculous weapon to use for assassinations given their inherently-indiscriminate nature; there are far-better choices.”

[780] 337, “Too much room for error and chance. Too high a risk with a less sure result.”

[781] 339, “On the low end, the group acknowledges that while that biotechnology state-of-the-art will advance, use of contagious agents presumably will be an international taboo and analogous to the use of chemical weapons or extant bioweapons.”

[782] 343, “An example of a better choice was: “a one-drop easily-synthesised incurable-once-symptoms-appear contact poison.”” 340, “Poisons, for example, are effective and easier to use, and guns and bombs are easier to source and reliably lethal.”

    - Rather than using a base rate of zero, some forecasters suggested that there may have been unidentified assassinations and assassination attempts with biological weapons in the past, so it is appropriate to use a higher base rate.[783]
- Advances in biotechnology will make bioweapons more attractive
    - Most higher forecasts put some weight on the possibility that biotechnologies will advance to overcome many of their limitations as a weapon identified above.[784] In particular, several teams referenced the possibility of targeted bioweapons.[785]
- Possibility of copycats following one attempt increase expected events
    - Two teams suggested that one assassination attempt may result in more "copycat" attempts, pushing the expected number of events higher.[786]
- There may be a greater number of actors who may use these weapons in the future
    - Several teams suggested that there may be a greater number of actors who might use a non-targeted weapon for an assassination attempt, including actors who are unbothered by, or at least willing to cause, collateral damage.[787]

Cross-references with other questions

Q20: Individual Countries with Biological Weapons Programs
Q21: Number of Countries with Biological Weapons Programs

[783] 337, "We also note that many heads of state have died from various illnesses, which could potentially be from some form of bioweapon exposure." and "What fraction of future assassinations would I expect to be with biological agents in the future? By 2030, I think this would still be very small, but significantly higher than 0.5/107"
[784] 336, "Team median went up a little after a discussion of difficulties in executing a biological weapon today that might not be a problem for the two later dates." 339, "However on the high end as "the understanding of complex genomic interactions and the link between genotype and phenotype, as well as the quality of next-generation genome editing technologies improves in the next 10-15 years" bioweapons may become more ubiquitous."
[785] 340, "The best argument discussed in favour of assassinations by contagious agent is that biotechnology innovation will enable targeted pathogens to be developed, and that it will be possible to deliver them safely and reliably."
[786] 343, "The strongest argument for a forecast on the higher end of the plausible range of forecasts for this question is on-going proliferation interacting with copycat contagion (one assassination spurring copycats)." 344, "That once this becomes a proven method of assassination it will increase in popularity among would-be assassins."
[787] 343, "'Discussions refined this point, noting that the indiscriminate nature could be a benefit to the agendas of some potential actors or in targeting a gathered group. For this reason, the threat of contagious agent use is likely greater from non-state actors that have less targeted geo-political aims than from state actors interested in precision hits." 336, "Also seems possible that there will be a rise in non-state actors that don't care if they kill a lot of people, or who want to kill a lot of people in addition to whatever leader is targeted, as more effects of climate change are felt. But - in that case, may be difficult to define the event as an assassination."

Question 24: Malaria Deaths

What will be the number of human deaths due to malaria…
...during the year 2024?
...during the year 2030?
…during the year 2050?

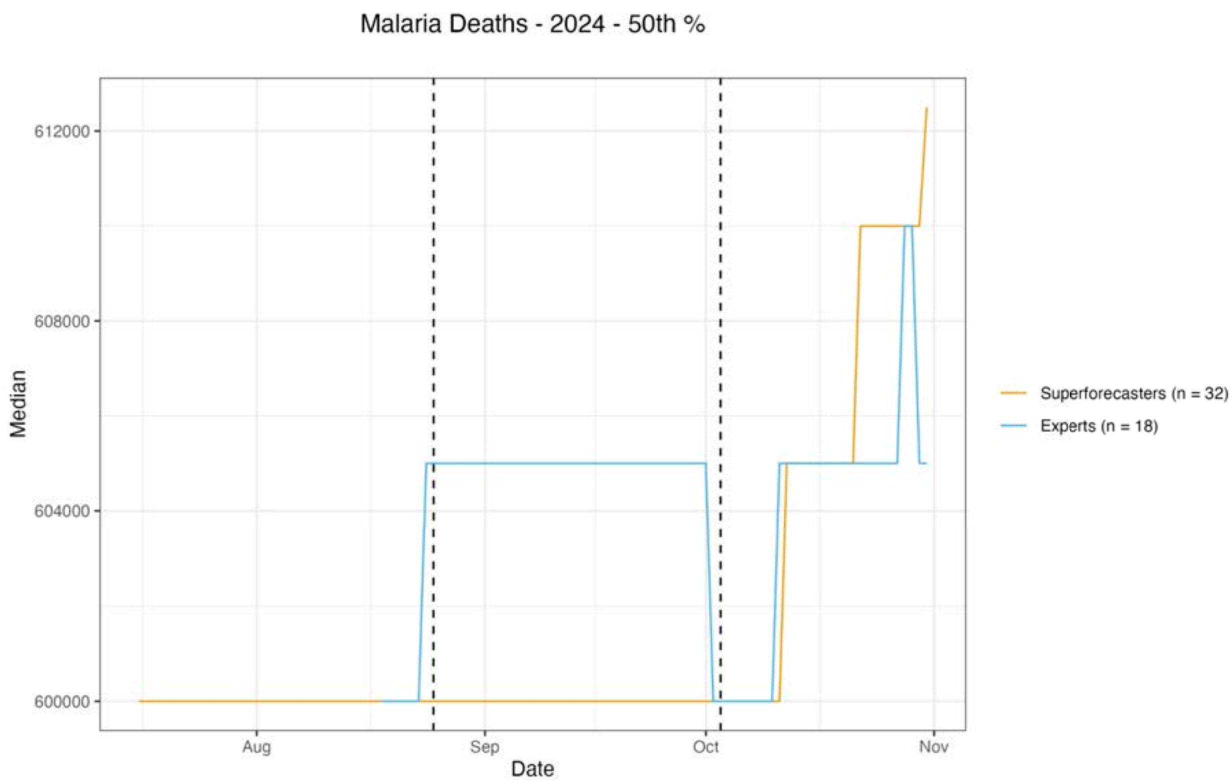
Question and resolution details, prior forecasts, and other relevant sources

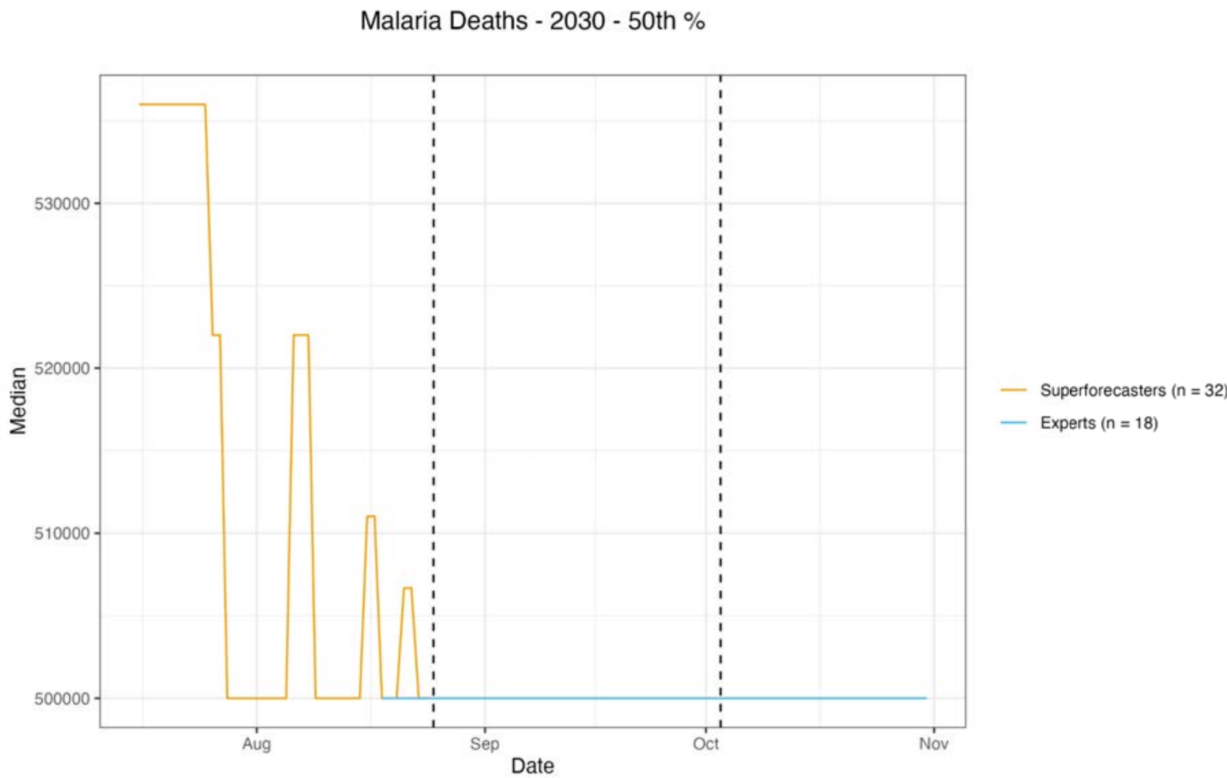
Results[788]

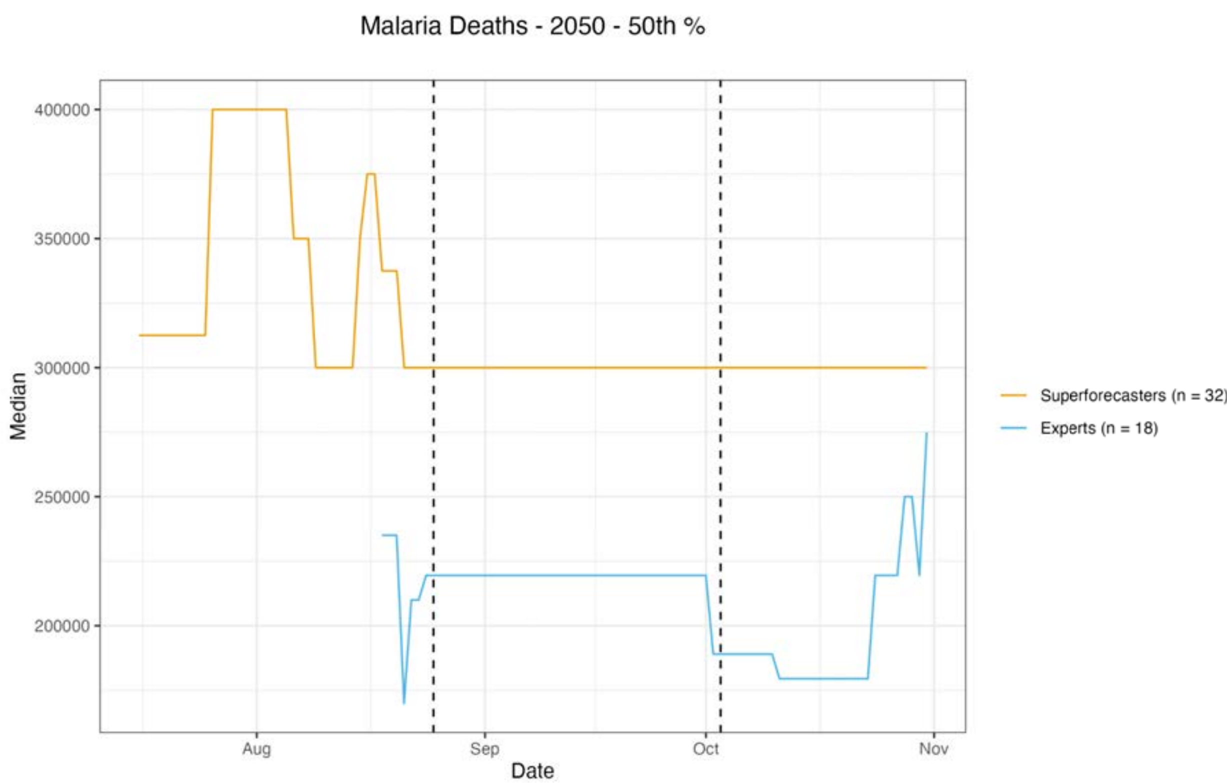
Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 32)	2024	600,000	612,500	130591034.02	-99.96%
	2030	550,000	500,000	107620775.5	-99.9%
	2050	362,500	300,000	31171818.41	-99.39%
Domain Experts (N = 4)	2024	620,000	597,500	75718.78	-17.99%
	2030	372,000	423,500	131001.27	-12.23%
	2050	50,000	150,000	78995.78	+81.98%
General X-Risk Experts (N = 5)	2024	500,000	612,500	100664.46	-74.1%
	2030	500,000	500,000	144337.57	-29.45%
	2050	100,000	250,000	196927.23	-23.62%

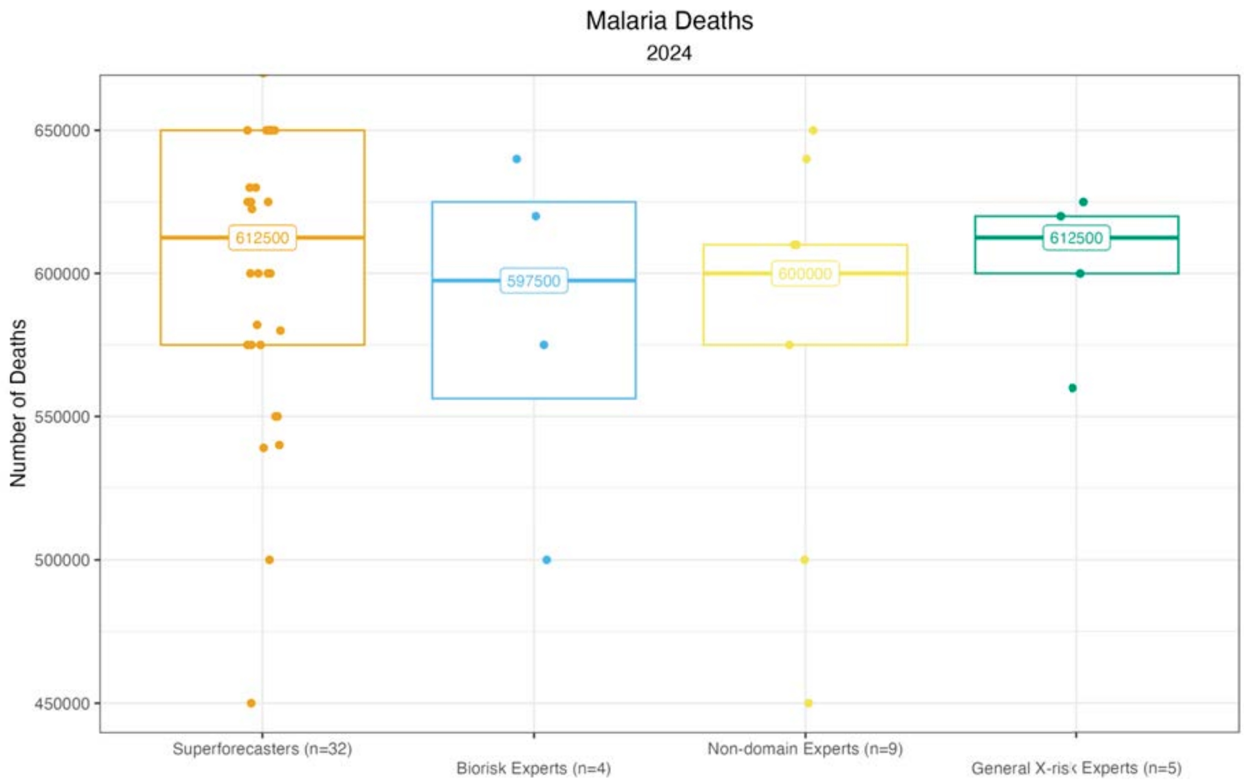
[788] Numbers of forecasters are given as of Stage 4 of the XPT.

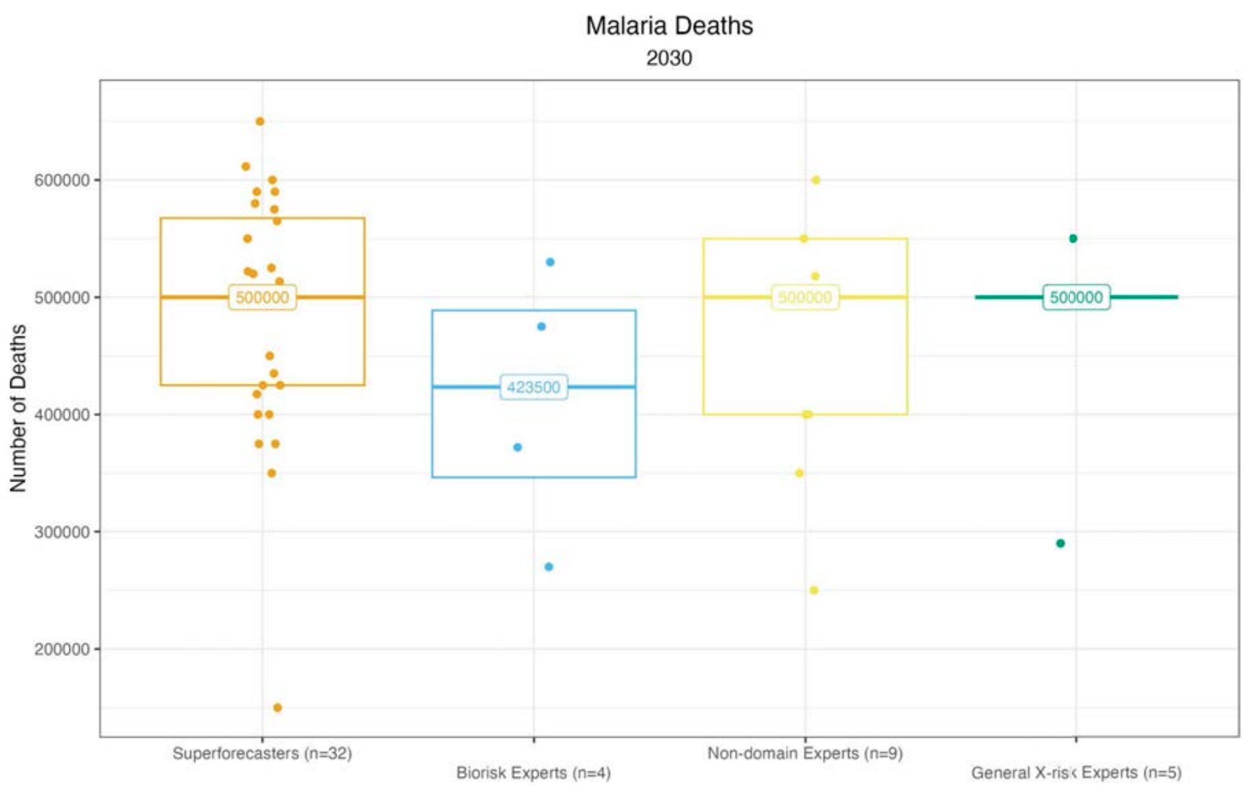
Non-Domain Experts (N = 9)	2024	600,000	600,000	n/a	n/a
	2030	500,000	500,000	n/a	n/a
	2050	400,000	320,000	n/a	n/a

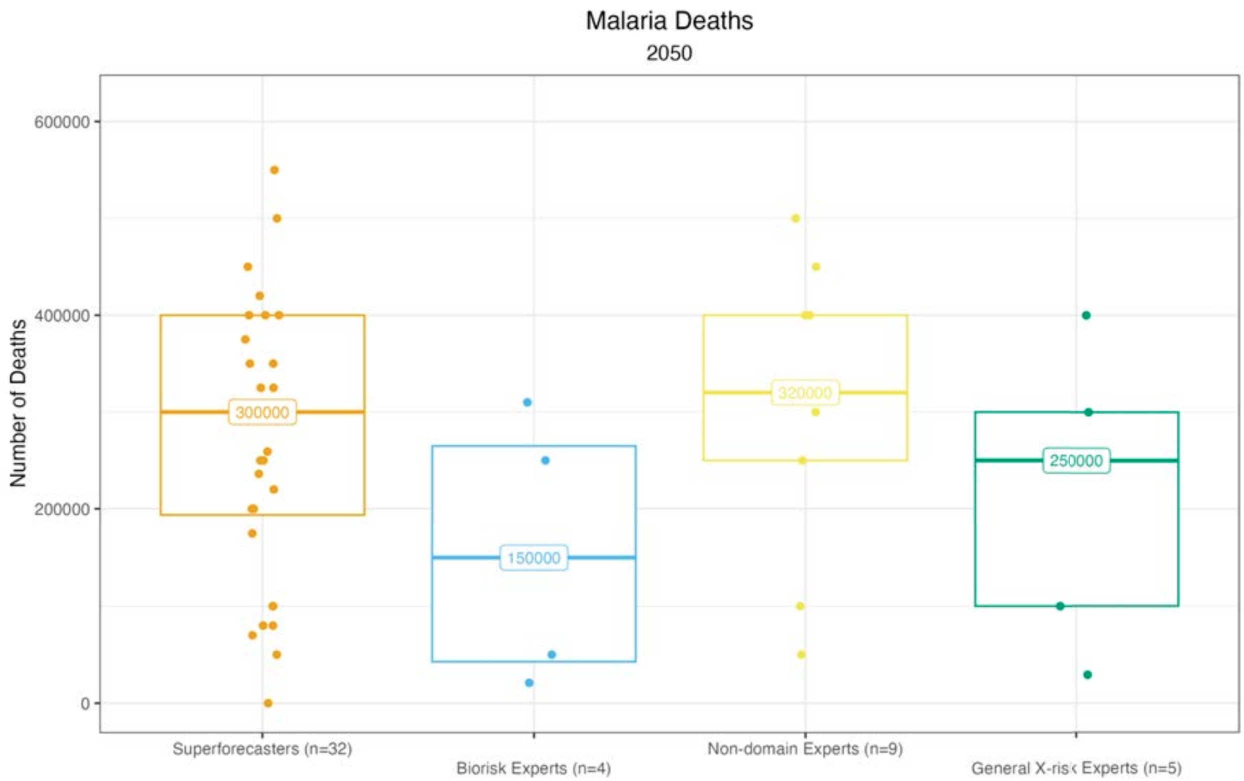



Malaria Deaths - 2030 - 50th %
530000
520000
510000
500000
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 32)
Experts (n = 18)


Malaria Deaths - 2050 - 50th %
400000
350000
300000
250000
200000
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 32)
Experts (n = 18)


Malaria Deaths
2024
Number of Deaths
650000
600000
550000
500000
450000
612500
597500
600000
612500
Superforecasters (n=32)
Biorisk Experts (n=4)
Non-domain Experts (n=9)
General X-risk Experts (n=5)


Malaria Deaths
2030
Number of Deaths
600000
500000
400000
300000
200000
500000
423500
500000
500000
Superforecasters (n=32)
Biorisk Experts (n=4)
Non-domain Experts (n=9)
General X-risk Experts (n=5)



Sources of agreement, disagreement and uncertainty

Agreement:

- All teams predicted a trend of decreasing deaths over time. Most expected a return to the pre-COVID-19 decline in malaria mortality, which had been disrupted due to the pandemic.
- Most teams used the same baseline rate of malaria mortality change using data from the World Health Organization (WHO) and the Institute for Health Metrics and Evaluation (IHME). Teams then differed in how they expected different factors to alter this baseline.

Disagreement:

- Forecasters had different opinions on the expected impact of the following on malaria mortality:
    - Malaria vaccines
    - Gene drives
    - Climate change
    - Population growth
    - Economic development

Uncertainty:

- Most teams expressed uncertainty in the effects of each of the above factors.

- One team noted uncertainty in the accuracy of the WHO estimates of the malaria burden.[789]

Arguments given for forecasts at or below the median of 600,000 (2024), 500,000 (2030), 270,000 (2050)

- Prior to COVID-19, malaria mortality had been steadily decreasing
    - Many teams noted that, prior to the COVID-19 pandemic, malaria mortality had been steadily decreasing for several years. Most teams expected the historical trend to return and then continue.[790]
- New vaccines have recently been approved
    - Several teams noted developments in malaria vaccination, including the WHO approval of the RTS,S vaccine, and the successful phase 3 clinical trial of the R21 vaccine.[791] One team suggested that large numbers of these vaccines could be produced by the Serum Institute of India.
    - Other teams also mentioned the possibility of mRNA-based vaccines being developed.[792] However, one team noted that developing a malaria vaccine would be much more difficult than developing a COVID-19 vaccine.[793]
- Gene drives (or other technology) may reduce malaria burden further

---

[789] 340, “However, all of the forecasts come with substantial uncertainty. For 2024, this seems to be based on worries about inconsistent numbers between the 2020 and 2019 WHO reports, which affect both base rate estimates and confidence in the reliability of WHO reported numbers on the topic, as well as uncertainty about the effects of COVID on Malaria deaths and prevention efforts.”

[790] 340, “The median forecasts of 600k, 450k and 300k for 2024, 2030 and 2050 respectively are mainly based on the expectation that the steady decline in Malaria deaths of the last decades (base rate -10k to -30k per year depending on data source and the exact considered time interval) will continue even after a momentary disruption by the COVID pandemic, albeit perhaps slightly slower as there have already been signs of a slowdown before COVID hit.” 341, “The mean forecast for 2024 is roughly the last known yearly number. It seems unlikely that in only two years the number of deaths will change considerably, as it also was slow changing in the past.”

[791] 341, “Currently the RTS,S vaccine has achieved WHO prequalification, and UNICEF has a contract with GSK (the main RTS,S supplier) for 18M vaccines for until 2025. The R21 vaccine is currently in Phase 3 trials and is expected by the malaria policy community to hit the market between 2025-2027. The Serum Institute of India, the supplier of R21, expects to produce 100-250M doses of R21 annually.” 342, “The RTS,S/AS01E vaccine, which has been approved by the WHO, is now routinely available in three countries badly affected by malaria; Ghana, Kenya, and Malawi.”

[792] 341, “Chances are high that better vaccines will be available in the future, especially in the 2050 forecast. Especially mRNA vaccines look promising.” 336, “mRNA vaccines may end up being very effective. Already, BioNTech (the developer of the Pfizer Covid vaccine) plans to start trials on humans with an mRNA vaccine this year.”

[793] 341, “We note though that an mRNA vaccine against malaria is more complicated than COVID due to the higher number of genes involved.”

  - Several teams mentioned the possibility of gene drives designed to modify or eradicate the Anopheles mosquito (the main vector for malaria transmission to humans).[794]
- As countries develop economically we should expect further decline in malaria burden
  - Some teams suggested that ongoing economic development of countries with a high malaria burden should drive mortality down further, with one team suggesting this may be due to people spending greater time indoors and improved healthcare.[795]
- Extinction risk
  - One team reported that one forecaster included extinction risk in their extreme (5%) forecasts for 2050.[796]

Arguments given for forecasts above the median of 600,000 (2024), 500,000 (2030), and 270,000 (2050)

- Vaccines may have limited effect
  - One team noted that it is difficult to develop a highly effective vaccine for malaria, due to the surface of the parasite changing.[797]
- The population at risk is expected to grow
  - Several teams noted that the population in regions of malaria prevalence (e.g. Africa) is expected to increase in the coming decades, which would lead to increased total mortality if per-capita mortality rates remain the same.[798]
- Historical trends shouldn't be expected to continue
  - One team noted that previous declines in malaria mortality may have primarily been due to declines in Asia, and that we should not necessarily expect the same pattern to hold for Africa.[799]

---

[794] 336, "Anti-mosquito technology (e.g., gene drives) may improve in effectiveness and cost-effectiveness" 342, "Some hope that gene drives and mRNA vaccines (or something else) will come into effect this decade and make significant progress, thus speeding up the decline."
[795] 340, "...[G]eneral economic progress leading to better healthcare and improved prevention in countries affected by Malaria." 342, "Forecasters' estimates of potential economic/political development for the most severely affected regions should be expected to heavily impact their assessment of malaria deaths."
[796] 344, "The lowest forecasts for 2050 predict 0 deaths, due to either a cure for malaria rolled out globally (e.g. wiping out mosquitoes) or human extinction occurring before that year (due to various reasons like AI-related incidents, genetic risk, or another source)."
[797] 341, "The high values are mainly based on the assumption that population in the malaria regions will increase considerably, while also assuming that a good vaccine will not be available" 336, "Malaria is hard to develop a vaccine for - the surface of the parasite changes which makes it difficult for the immune system to target"
[798] 341, "The high values are mainly based on the assumption that population in the malaria regions will increase considerably, while also assuming that a good vaccine will not be available" 337, " explosive population growth in sub-Saharan Africa in general, Nigeria in particular" 341, "Population will rise considerably in Africa, thus more people can catch malaria"
[799] 338, "Argue that large drop since ~2004 came from large eradications in southeast Asia but we may see upward trend again in Africa due to poverty and conflict."

- Some teams suggested that we should expect increases in economic or political instability, or in destabilizing events such as COVID-19, which would lead to increases in malaria mortality.[800]
- Climate change may increase the geographic spread of malaria
  - Several teams noted that climate change is expected to increase the geographic spread of malaria by making more areas habitable for the Anopheles mosquito.[801]

Other arguments given

Arguments in favor of lower forecasts:

- Advanced AI resulting in development of a cure
  - One team noted that the possibility of advanced AI developing a cure should push forecasts lower.[802]
- Political motivation to reduce malaria burden
  - One team suggested that there is strong political will to eliminate malaria, citing the WHO's goal to reduce the malaria burden by 90% by 2030.[803]

Arguments in favor of higher forecasts:

- Drug resistance
  - One team suggested that we may see a rise in drug-resistant malaria cases, which may slow down or reverse declines in mortality.[804]
- Diminishing returns on investment in prevention
  - One team suggested that we should expect diminishing returns on investment in prevention.[805]

---

[800] 337, "[C]ollapse of effective government, including corruption and social unrest. Future pandemics, global financial crises, or major wars could diminish or reverse gains" 344, "For 2024, forecasts such as 696k deaths are consistent with the current, increasing trend of yearly malaria deaths. This could happen if disruptions persist or worsen due to economic and political instability."

[801] 342, "Climate change and increasing population growth in malaria affected areas mean that the geographic range of malaria-carrying mosquitos will expand significantly" 344, "But climate change may bring malaria to new geographical zones, thus countering other trends that would have otherwise kept decreasing the number of yearly malaria deaths."

[802] 342, "One team member also strongly believes that AI will cure malaria, almost certainly by 2050, if not 2030."

[803] 343, "Adding to this is significant political motivation, as is demonstrated by the WHO's goal of reducing the malaria burden by 90% by 2030."

[804] 340, "Specific factors that could contribute to the latter scenario include the factors causing a slowdown in Malaria death reduction discussed earlier, as well as Malaria evolving drug resistances and Malaria-carrying Mosquitos becoming more common around the world due to climate change."

[805] 340, "The comparatively slower reduction in deaths after 2030 is based on arguments around diminishing returns to prevention efforts…"

Cross-references with other questions

Q14: Novel Infectious Disease Surveillance System

Q22: PHEIC Declarations with 10k Deaths

Question 25: Average Global Surface Temperature

What will be the global surface temperature change as compared to 1850-1900, in degrees Celsius…
...in 2030?
...in 2050?
…in 2100?

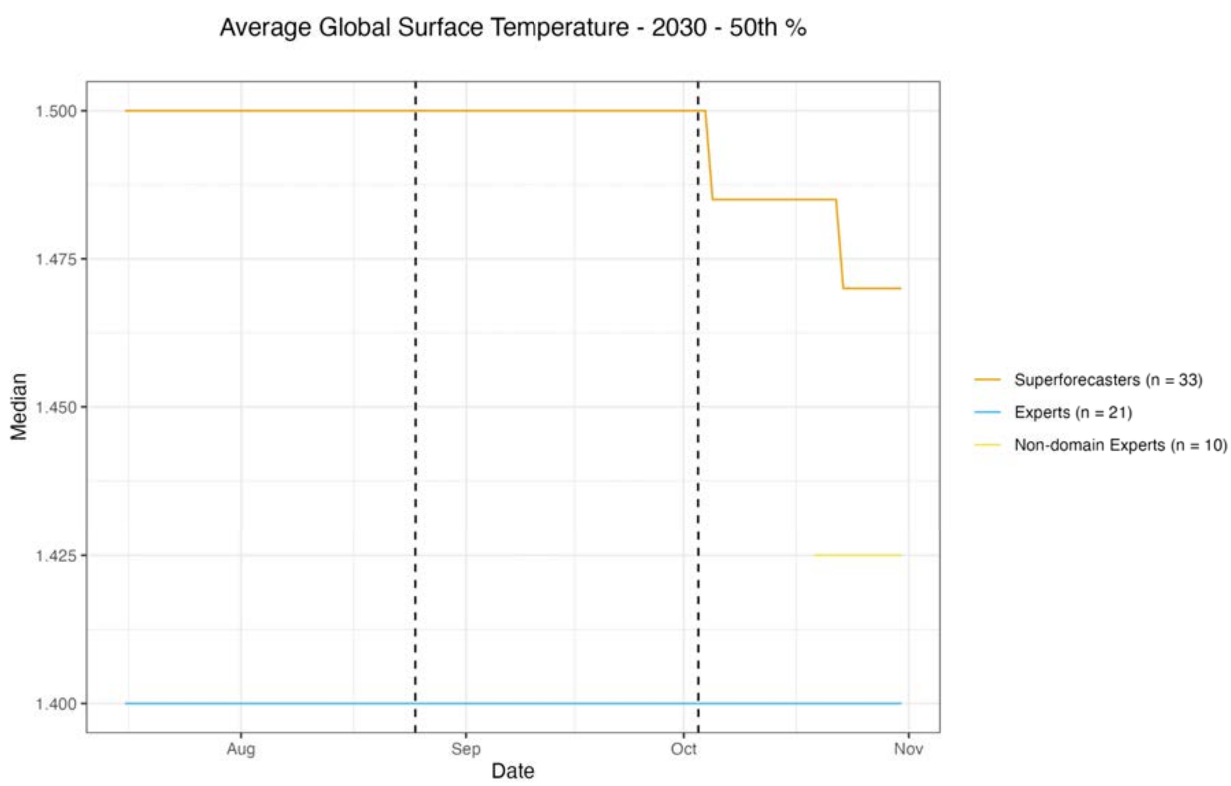
Question and resolution details, prior forecasts, and other relevant sources

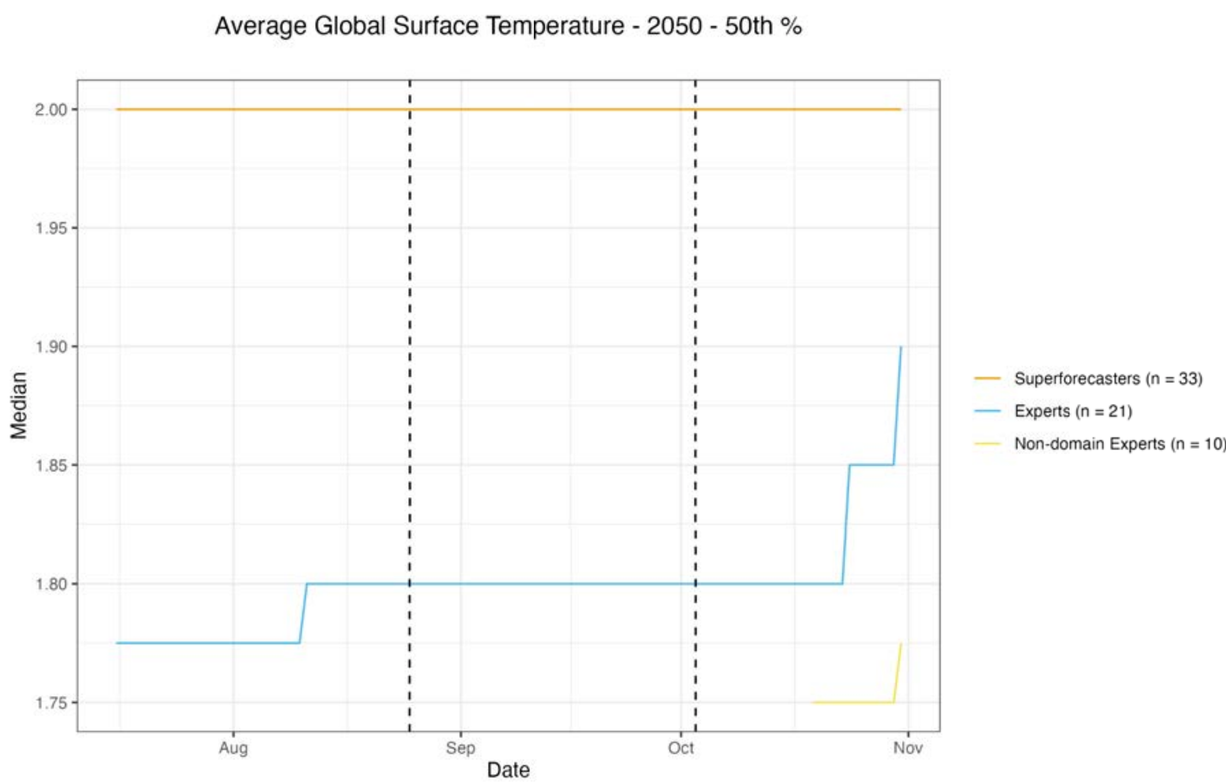
Results[806]

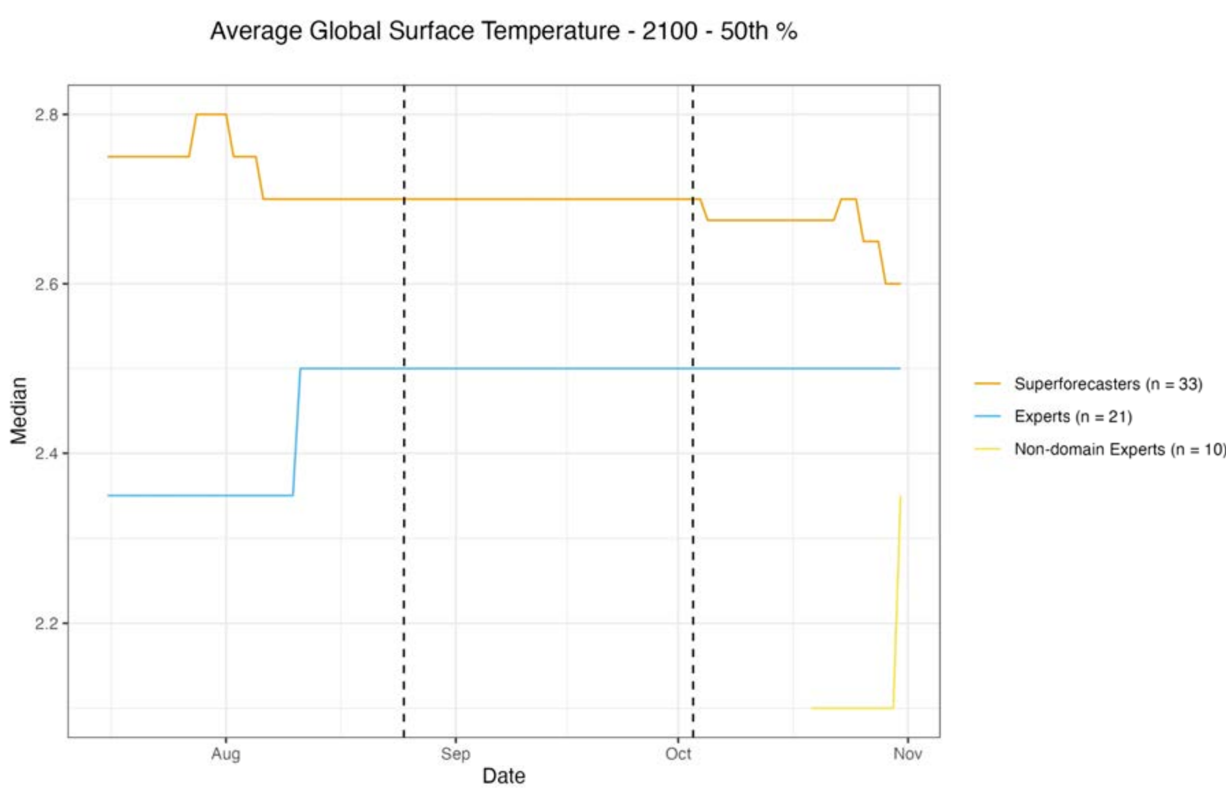
Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 33)	2030	1.5	1.47	0.48	-75.42%
	2050	1.98	2	0.39	-32.89%
	2100	2.7	2.6	0.72	-28.53%
Domain Experts (N = 8)	2030	1.4	1.4	0.2	-55.91%
	2050	1.9	2	0.32	-60.81%

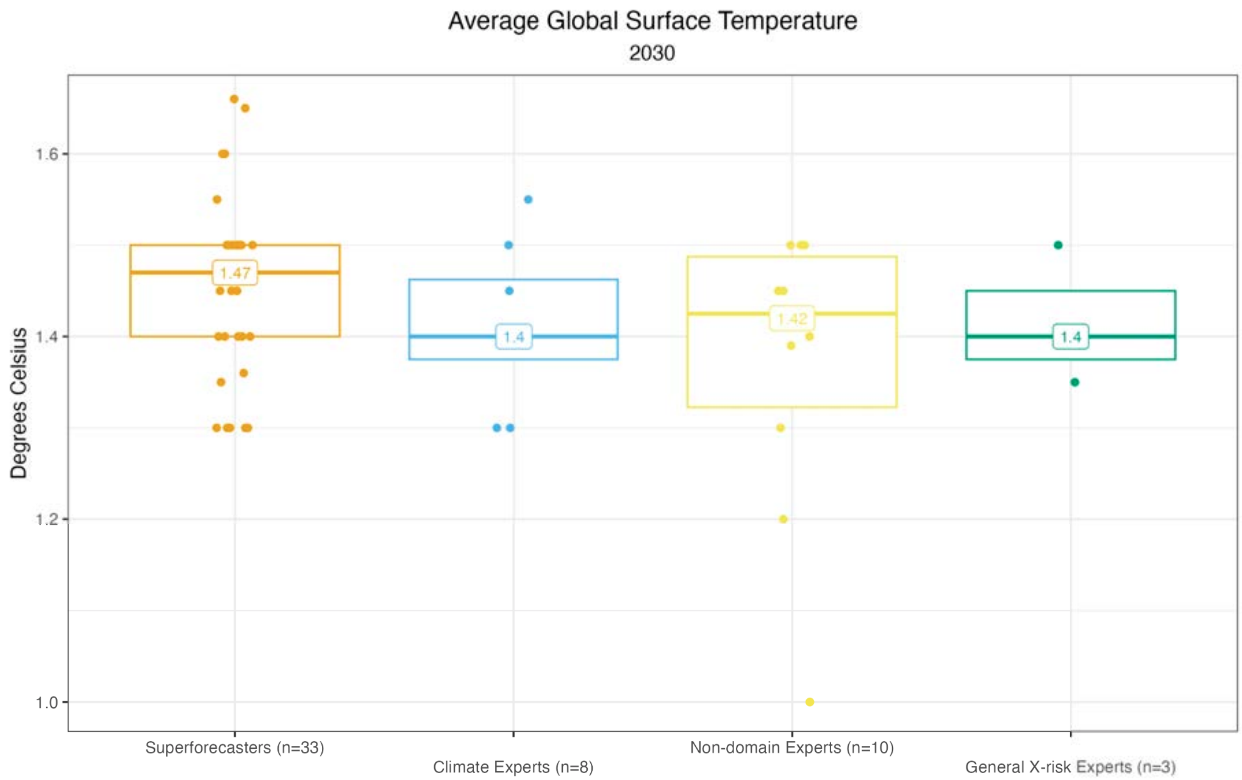
[806] Numbers of forecasters are given as of Stage 4 of the XPT.

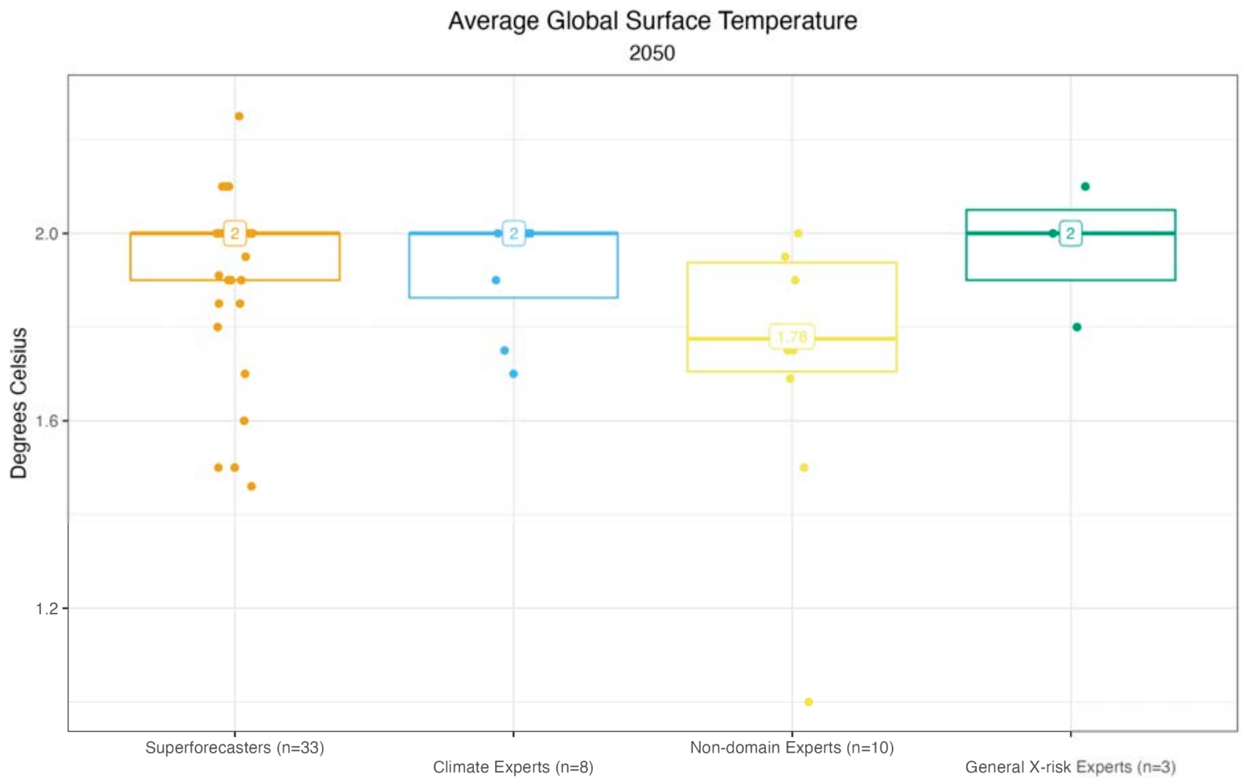
	2100	2.5	2.55	1.74	-66.17%
General X-Risk Experts (N = 3)	2030	1.45	1.4	0.07	+8.01%
	2050	1.9	2	0.14	+8.01%
	2100	2.6	2.7	0.14	+77.95%
Non-Domain Experts (N = 10)	2030	1.4	1.43	0.22	-25.39%
	2050	1.69	1.78	0.54	-29.59%
	2100	2	2.35	0.97	-46.06%

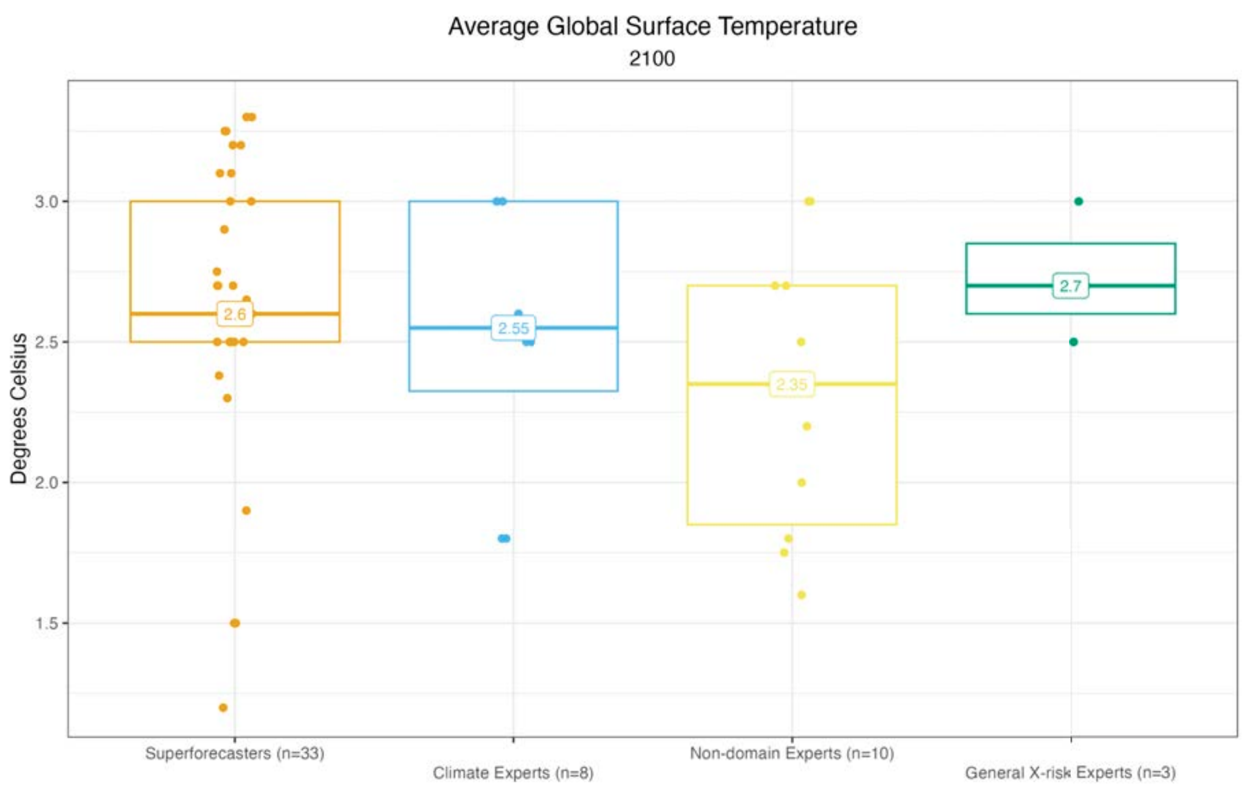



Average Global Surface Temperature - 2050 - 50th %
2.00
1.95
1.90
1.85
1.80
1.75
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 33)
Experts (n = 21)
Non-domain Experts (n = 10)


Average Global Surface Temperature - 2100 - 50th %
2.8
2.6
2.4
2.2
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 33)
Experts (n = 21)
Non-domain Experts (n = 10)


Average Global Surface Temperature
2030
Degrees Celsius
1.6
1.4
1.2
1.0
1.47
1.4
1.42
1.4
Superforecasters (n=33)
Climate Experts (n=8)
Non-domain Experts (n=10)
General X-risk Experts (n=3)


Average Global Surface Temperature
2050
Degrees Celsius
2.0
1.6
1.2
2
2
1.78
2
Superforecasters (n=33)
Climate Experts (n=8)
Non-domain Experts (n=10)
General X-risk Experts (n=3)


Average Global Surface Temperature
2100
Degrees Celsius
3.0
2.5
2.0
1.5
2.6
2.55
2.35
2.7
Superforecasters (n=33)
Climate Experts (n=8)
Non-domain Experts (n=10)
General X-risk Experts (n=3)

Sources of agreement, disagreement and uncertainty

- **IPCC forecasts as starting point**
  - Most teams found predictions in the Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report[807] to be a solid starting point for their forecasts.[808] Teams did not express significant criticism of the IPCC's methodology.
    - "[T]he team had a high degree of confidence in the IPCC models, and considered the main source of uncertainty to be future emissions pathways." (T337)
  - Median forecasts of this question roughly aligned with the IPCC's predictions (see table below) and were closest to IPCC's "middle of the road" scenario, SSP2-4.5, except that the range of forecasters' predictions for 2050 and 2100 was broader than the range in any IPCC scenario.
    - In addition, the IPCC assessed a range of years (e.g., 2041–2060), while this question asked about temperature change "in" a specific year (e.g., 2050).
- Citing various government sources, several teams assumed 0.15 to 0.2°C temperature increase per decade.[809]
  - "Using the numbers from Climate.gov, (0.08 degrees C per decade up to 1980, 0.18 degrees C per decade since 1980), I mapped out a range of low, medium, and high possibilities." (T338)
  - "NASA estimates recent temp increases have been at the rate of 0.15 to 0.2C per decade." (T341)

---

[807] Specifically, teams referred to the Technical Summary in ***Climate Change 2021: The Physical Science Basis***, released on August 9, 2021.
[808] "Base rate: The team relied primarily on the most recent IPPC assessment report technical summary…[but] [t]here is at least one dissenter of the IPCC models, pointing at the lower actual temperatures between 2010-2020 than the models predicted." (T336)
"Many forecasters cited the IPCC projections [1], suggesting that the SSP1-2.6 and 2-4.5 models were far and away the most credible…We expect that superforecasters and experts alike will likely be guided primarily by the IPCC reports, especially given that the experts may well have fed into the IPCC report themselves." (T337)
"Forecast Strategy [:] First, pick high and low end distribution end members from the modeled scenarios, and assign probabilities. Use those two points to define a lognormal distribution. The lognormal distribution makes lower values more likely and leaves a tail of larger magnitude events." (T338)
"The IPCC forecasts represent a solid starting point for informing our median predictions." (T340)
[809] "Our median forecast is that the temperature will increase another 0.2 degrees by 2030 to be 1.2 degrees over the 1850 to 1900 average, a further 0.4 degrees from there to 2050, and a further 0.8 degrees from there to end around 2.4 degrees above the 1850 to 1900 average by 2100. This median is essentially a reflection of the trend from 1970 to 2020 of an increase around 0.2 degrees every 10 years continuing out, with perhaps a bit of slowdown, through 2100." (T342)

- “This median [forecast] is essentially a reflection of the trend from 1970 to 2020 of an increase around 0.2 degrees every 10 years continuing out, with perhaps a bit of slowdown, through 2100.” (T342)

Cross-Section Box TS.1 (continued)

Cross-Section Box TS.1, Table 1 | Assessment results for 20-year averaged change in global surface temperature based on multiple lines of evidence. The change is displayed in °C relative to the 1850–1900 reference period for selected time periods (first three rows), and as the first 20-year period during which the average global surface temperature change exceeds the specified level relative to the period 1850–1900 (last four rows). The entries give both the central estimate and, in parentheses, the *very likely* (5–95%) range. An entry n.c. means that the global warming level is not crossed during the period 2021–2100.

	SSP1-1.9	SSP1-2.6	SSP2-4.5	SSP3-7.0	SSP5-8.5
Near term, 2021–2040	1.5 [1.2 to 1.7]	1.5 [1.2 to 1.8]	1.5 [1.2 to 1.8]	1.5 [1.2 to 1.8]	1.6 [1.3 to 1.9]
Mid-term, 2041–2060	1.6 [1.2 to 2.0]	1.7 [1.3 to 2.2]	2.0 [1.6 to 2.5]	2.1 [1.7 to 2.6]	2.4 [1.9 to 3.0]
Long term, 2081–2100	1.4 [1.0 to 1.8]	1.8 [1.3 to 2.4]	2.7 [2.1 to 3.5]	3.6 [2.8 to 4.6]	4.4 [3.3 to 5.7]
1.5°C	2025–2044 [2013–2032 to n.c.]	2023–2042 [2012–2031 to n.c.]	2021–2040 [2012–2031 to 2037–2056]	2021–2040 [2013–2032 to 2033–2052]	2018–2037 [2011–2030 to 2029–2048]
2°C	n.c. [n.c. to n.c.]	n.c. [2031–2050 to n.c.]	2043–2062 [2028–2047 to 2075–2094]	2037–2056 [2026–2045 to 2053–2072]	2032–2051 [2023–2042 to 2044–2063]
3°C	n.c. [n.c. to n.c.]	n.c. [n.c. to n.c.]	n.c. [2061–2080 to n.c.]	2066–2085 [2050–2069 to n.c.]	2055–2074 [2042–2061 to 2074–2093]
4°C	n.c. [n.c. to n.c.]	n.c. [n.c. to n.c.]	n.c. [n.c. to n.c.]	n.c. [2070–2089 to n.c.]	2075–2094 [2058–2077 to n.c.]

Table taken from p. 63 of the Technical Summary in *Climate Change 2021: The Physical Science Basis*.

- **Broad agreement that existing trends will continue to 2030**
  - “There's enough inertia in our CO2 emission patterns that at least for the next decade or so there's little room for surprises. Policies of most important governments show few signs causing any large changes to CO2 emissions.” (T340)
  - “There should not be a significant chance to get above 2C by 2030 as we are around 1.1C now. Even the highest warming scenario SSP5-8.5 has a likely (5% -95%) range of 1.3 to 1.9 for the near term, 2021 to 2040.” (T341)
- **Uncertainty over whether climate change will motivate drastic action**
  - Forecasters expressed substantial uncertainty about how humans will react to climate change. Some believed that, as climate change becomes more obvious, policymakers will be more motivated to make dramatic and politically costly changes. More pessimistic forecasters pointed to the history of inertia and missed emissions goals.
    - “There is a large amount of uncertainty amongst the median forecasters that the world will be able or willing to take dramatic steps to slow global warming significantly. Some think we won't do enough quick enough. There have been a lot of ambitious goals before and they've never been met.” (T336)

        - "The overall ability - and determination - of humankind to effectively reduce climate warming activities in coming years is likely one point of potential disagreement between various forecasting groups." (T338)
        - "The growing threat of abrupt and irreversible climate changes must compel political and economic action on emissions." (T339)
        - "For 2100, one can perhaps be more optimistic, that, given how bad things will likely have gotten, some global action will be taken…things are going to have to get a lot worse, especially in the countries that matter the most, before significant action is taken." (T341)
- **Disagreement over people's willingness to change personal habits**
    - Some forecasters believed that obvious climate change will motivate ordinary people to change their personal habits or to demand political change; others pointed to the difficulty of changing established lifestyles.[810]
        - "Public opinion continues to sway in favor of decarbonisation as the effects of climate change are becoming more apparent. In addition, people change their behavior (e.g. fly less) and will demand politicians to take action." (T337)
        - "A couple of high warming 95th percentile forecasts expressed skepticism that human behavior would appreciably change to slow accelerating global warming trends. According to them, a lot of it is baked in already and expecting people to change their habits if it influences their lifestyle in a (perceived) negative sense is naïve." (T336)

Arguments given for forecasts of lower temperatures

- **Replacement of fossil fuels with renewable energy**
    - The most frequent rationale for more optimistic temperature change forecasts was that renewable energy would replace energy from fossil fuel sources. This could happen either because the cost of renewable energy decreases or because the cost of fossil fuels increases. Some forecasters expressed more pessimistic views of renewable energy.
        - "There's a clear trend of technological progress toward replacing fossil fuels with cleaner energy, making it reasonable to expect CO2 emissions to decline faster than IPCC forecasts." (T340)
        - "While clean energy technology is good, and getting better, challenges with long term energy storage, and long distance distribution will prevent it from becoming a complete substitute for fossil fuels within at least the next decade, probably longer. " (T343)
- **Decrease in renewable energy cost**
    - Solar and wind power will become more affordable

---

[810] "Belief in a technological solution and/or global change of habits due to the evident impact of climate change." (T344)

        - "The cost of utility scale PV [primary photovoltaics] (with nuclear fusion on the horizon) may be reaching the point where it undergoes a rapid adoption curve and renders fossil fuel generation obsolete. Cheap and abundant electricity would also further strengthen incentives for electrification (vehicles, heating, cooking, etc.)…" (T337)
        - "Swanson's law [link] shows solar is on track to undercut fossil fuel costs in most locations…it would not be too surprising if the rate of solar adoption continues to exceed many people's expectations. Electric vehicles are on track to mostly replace gas-powered vehicles in the 2030s. Utility-scale battery storage costs are coming down steadily." (T340)
        - "The combination of cheaper renewable energy, nuclear fusion technology progression, and rapid growth of AI technology will promote technological solutions that significantly reduce the rate at which the global surface temperature is rising." (T345)
    - Possibility of nuclear fusion energy[811]
        - "If fusion power works out, that pushes emissions more toward the net zero commitments." (T338)
    - Pessimism about renewable energy
        - "[S]ome forecasters noted uncertainty around whether the established trends in solar/wind/EV [electric vehicle] cost reductions would continue - would lithium or other critical raw materials/components become significantly more expensive as rapidly expanding production consumed all available supply (new mines, even assuming the economic case is strong, can often take decades to bring online)." (T337)
        - "While clean energy technology is good, and getting better, challenges with long term energy storage, and long distance distribution will prevent it from becoming a complete substitute for fossil fuels within at least the next decade, probably longer. Developing countries will continue to rely heavily on fossil fuels, absent a major push from wealthy countries to facilitate clean energy transitions." (T343)
- **Increase in fossil fuel energy cost**
    - "[D]ependency on fossil fuels is seen as a long-term economic and strategic risk for countries with limited or no internal access to oil, gas, or coal reserves. These countries thus have to diversify energy sources to lower risk for geopolitical reasons." (T337)
- **Geoengineering and carbon capture**

---

[811] "The cost of utility scale PV (with nuclear fusion on the horizon) may be reaching the point where it undergoes a rapid adoption curve and renders fossil fuel generation obsolete." (T337) "The combination of cheaper renewable energy, nuclear fusion technology progression, and rapid growth of AI technology will promote technological solutions that significantly reduce the rate at which the global surface temperature is rising." (T345)

- Some with lower temperature forecasts expressed hope that geoengineering and carbon capture technology would improve and be widely adopted. Others were uncertain about the efficacy and political feasibility of this technology.[812]
    - “Carbon capture technology may not pan out in terms of efficacy; and, even if it does, it is likely to be used to maintain existing temperature levels (within some range of values), rather than being used to dramatically lower the global temperature.” (T343)
    - “Carbon capture technology adoption will rapidly increase in the 2030s…Geoengineering may be able to return temperatures to 20th century levels. However, political considerations seem likely to limit use of geoengineering…We have little agreement on how much weight a geoengineering scenario deserves. The preponderance of opinion seems to put it well below 50%.” (T345)

- **Nuclear winter could lower global temperatures[813]**
    - Nuclear winter could lead to global temperature decrease, although forecasters who raised this possibility viewed it as uncertain and low likelihood.[814]
- **“Unknown unknown” technological solutions**
    - In their forecasts for 2100, some forecasters expressed a hope that technological solutions to climate change will be developed, even if we do not currently know what those solutions would be.[815]
        - “Technological advancement in the next decade and beyond will not only reduce carbon emissions lowering the average individual footprint, but also find ways of cleaning up past emissions reducing total CO2 in the atmosphere. Science will find ways to mitigate past CO2 excess emission in 'unknown unknown' ways, particularly by 2100.” (T339)
        - “For 2100, one can perhaps be more optimistic, that, given how bad things will likely have gotten, some global action will be taken, and by

[812] “Sources of disagreement…geoengineering: is it practical? will it be prohibited due to risks?” (T340)
[813] “Nuclear winter could lower temperatures.” (T340)
[814] “Some forecasters raised the possibility of significant cooling through geoengineering over-correction, nuclear winter, or volcanic activity. To the extent that these were discussed, all three possibilities were considered to be the realms of possibility, but would be very large effect sizes relative to historical precedent. Consequently, only one forecaster considered that, collectively, these possibilities reached the 5% probabilistic threshold to warrant inclusion in their forecasts.” (T337)
“Nuclear winter could lower temperatures, but seems unlikely to last long enough to have much influence on this forecast.” (T340)
“There is some disagreement over whether even a worst-case nuclear winter scenario would be sufficient to reverse global temperature increases significantly, and if so whether those increases would have persistent impacts.” (T343)
[815] "Maintain some possibility for decreases (negative values) based on technology interventions by 2050 and 2100." (T338)
“[R]apid growth of AI technology will promote technological solutions that significantly reduce the rate at which the global surface temperature is rising.” (T345)

then, renewables will b[e] extremely cost-efficient, and new technology should have been developed that will help eliminate emissions as well as capture carbon at scale.” (T341)

Arguments given for forecasts of higher temperatures

- **Difficulty of diplomatic coordination**
    - Forecasters pointed to several reasons why global coordination on climate change is difficult: because cheap energy is popular, because of game-theoretic coordination problems, and because some nations may stand to lose less (or even gain) from global warming. One more optimistic team argued that existing diplomatic commitments could exert positive peer pressure on global holdouts.
        - “Forecasters generally agreed that the political dimension was a key aspect that may lead to higher temperatures than our median forecasts. Would the commitment of Western democracies to decarbonisation reverse in the face of elections held in times of record high energy prices and recession? Could 'The West' lose political clout, causing authoritarian and/or developing countries to favour rapid expansion of coal generation made extremely cheap by the lack of Western demand?” (T338)
        - “Global diplomatic efforts to limit future warming will be insufficient due to international coordination problems.” (T343)
        - “Many governmental bodies have recently made major commitments to reduce emissions. As one country moves forward it encourages others to do the same. Conversely what country wants to be the lone holdout when peer pressure is building?” (T339)
- **Nations that benefit from higher temperatures could block change**
    - Not every global actor may agree that fighting climate change should be “mankind´s high priority project” (T336). Some northern countries might benefit from global warming, while others could be unwilling to risk a comfortable status quo.
        - “Countries such as Russia that might hope to benefit from warming could treat geoengineering as an act of war. Wealthy countries that can adapt to climate change seem likely to be risk-averse about accidents or malicious use of geoengineering.” (T340)
        - “One source of disagreement was a consideration of the impact of a breakdown in the 'rules based order' over the coming decades. For example, if [a] country believed that higher average temperatures were in its interest, or sought to blackmail the world by threatening to flare their large oil or gas reserves unless they were rewarded/compensated, what would the world do?” (T337)
- **Tipping point scenarios at very high temperatures**
    - Very high temperatures might lead to unpredictable or irreversible changes; forecasters did not provide much detail on what these changes would be, perhaps reflecting their uncertainty about the effects of unprecedented climate

change. On the other hand, one team suggested that very high temperatures might lead to effects that prevent further global warming.
        - "The greatest danger lies in crossing multiple tipping points, each of them pushing us to cross more of them, until they cascade into a cataclysmic situation." (T339)
        - "Increased uncertainty especially on the high temperature side to increase coverage of tipping point scenarios; increased low side uncertainty covers geo-engineering." (T338)
        - "Once global temperature increase[s] 5-10 degrees C, significant second-order effects come into play that prevent further increases." (T341)
    - Some higher predictions incorporated the possibility that temperature growth could become exponential.
        - "The major disagreement is between those whose numbers reflect that the worst case is likely to be upwards but linear, and those who believe bad but linear is the best case and that a move towards exponentiality is likely…[the upper end of our forecast] reflect[s] the possibility that temperature may already be increasing beyond our control and that what appears to be a new linear upward trend from 1970 to 2020 may be the beginnings of a significant shift towards more exponential growth, and that humans either cannot or will not do anything to stop it." (T342)
- **AI-driven economic growth could lead to increased use of energy**
    - Transformative AI could lead to accelerated use of energy, leading to global warming. On the other hand, there are possible scenarios in which an AI could decrease temperatures; forecasters were uncertain about these scenarios.
        - "Greenhouse gasses are not the only possible cause of global warming. Waste heat from computing / industrial activity could become the dominant source of warming…What could cause this? The leading guess is AI accelerating economic growth…an Age of Em scenario would also be sufficient." (T340)
        - "One forecaster raised the idea that the advent of AGI, particularly if unaligned with human values, could drive significant temperature change on the margins - either on the upside as the AGI, uncaring of human comfort levels rapidly exploits available resources and creates a runaway greenhouse effect, or on the downside, with the AGI deploying sulfur into the atmosphere (or taking other geoengineering measures) to reduce global temperatures, perhaps to make its server farms more efficient." (T337)
        - "If transformative AI or a similarly powerful technology greatly accelerates human economic power and/or coordination ability, I'd expect the temperature to be adjusted to whatever is seen as optimal, which might be very different from preindustrial temperatures if the world has changed in a lot of other ways." (T338)

Other arguments given

Arguments in favor of lower forecasts:

- Success of Montréal Protocol and existence of current decarbonization plans
    - "There were several low warming dissenters. They cite ambitious government plans like the successful Montreal Protocol which dramatically reduced O3 ozone depletion chemicals in the atmosphere…decarbonization has been started, and there are very ambitious plans (e.g. Net Zero CO2, and EU's "Green Deal.")" (T336)
- China has been decarbonizing recently and replacing coal power with nuclear.[816]
- "Supply chain collapse which leads to lower economic activity which leads to lower emissions" (T344)
- "Possible negative feedback effects from greater cloud coverage from ocean evaporation." (T344)

Arguments in favor of higher forecasts:

- Some forecasters believed that past IPCC predictions were underestimated due to a bias against alarmism and in favor of moderate-sounding predictions ("erring on the side of least drama").
    - "[A] comparison of past IPCC predictions against 22 years of weather data and the latest climate science find that the IPCC has consistently underplayed the intensity of global warming in each of its four major reports released since 1990." (T341)
- "It is possible that existing climate models do not account for feedback loops that could accelerate global warming (ie: methane clathrates)." (T343)
- Ideological bias against recognizing climate change
    - "More pessimistic considerations that some team members emphasize include…[i]deological reasons, climate change denial." (T344)

Cross-references with other questions

Q26: Cost of Utility-Scale Solar Energy
Q27: Nuclear Fusion Energy
Q28: Solar and Wind Energy
Q29: Annual Direct Air CO2 Capture
Q30: Cost of Hydrogen

---

816 "Also mentioned was broad public support in the majority of high-emissions countries for decarbonisation, including in China, long considered to be a likely holdout, but has recently been decarbonising rapidly, rapidly moving away from coal generation towards nuclear and other alternatives." (T337)

Question 26: Cost of Utility-Scale Solar Energy

What will be the estimated cost (in 2017 USD/kWh) for new utility-scale photovoltaic solar systems above $4MW_{AC}$ in the United States…
…for the year 2024?
…for the year 2030?

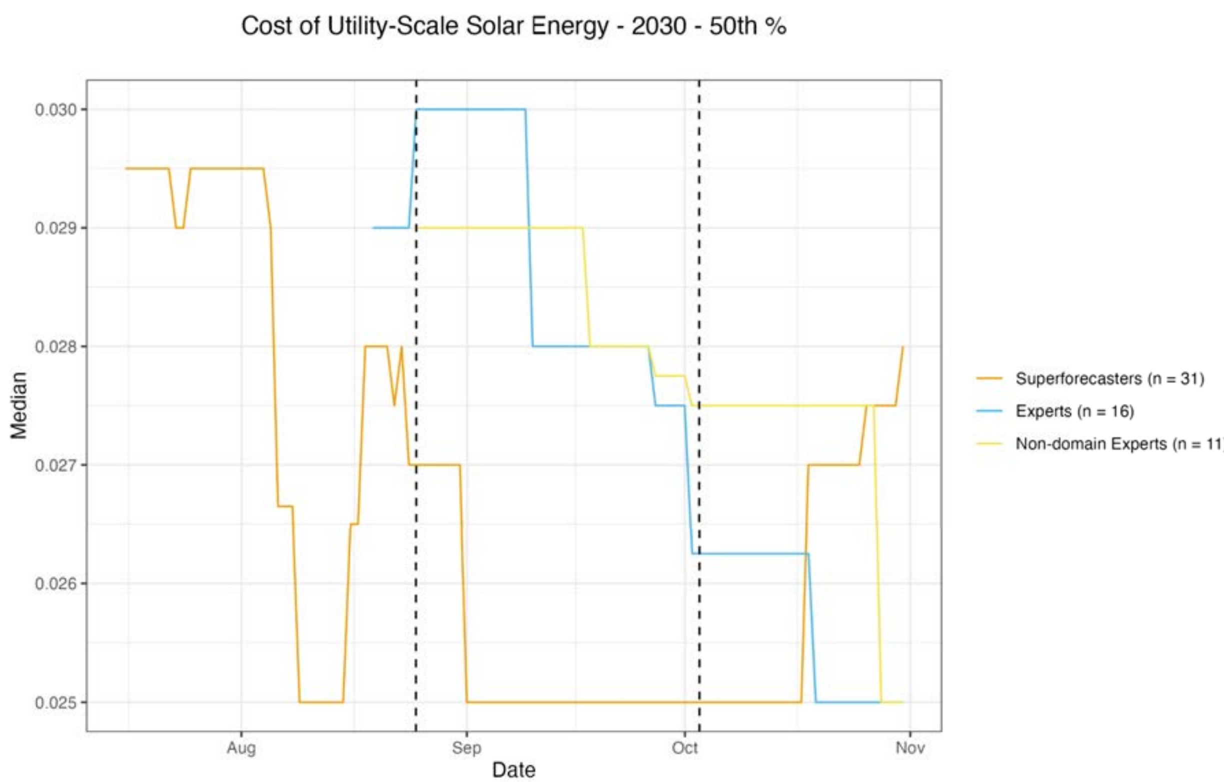
Question and resolution details, prior forecasts, and other relevant sources

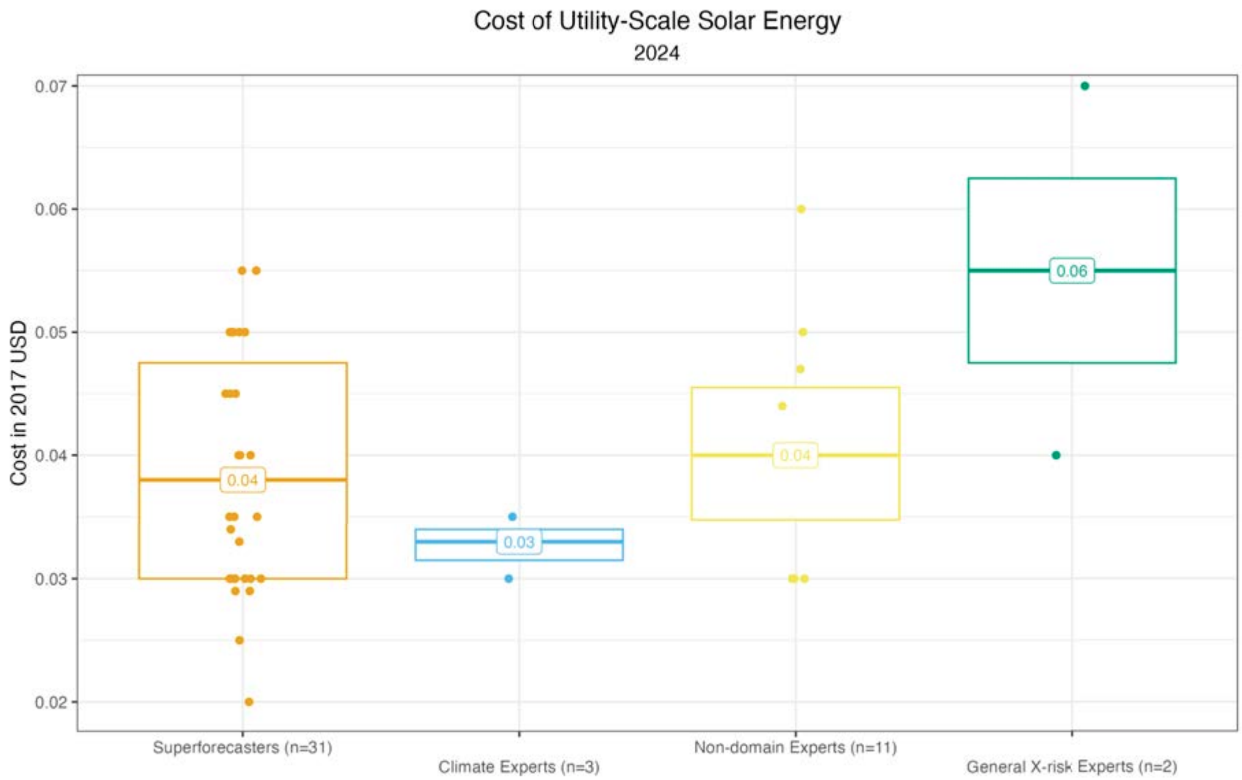
Results[817]

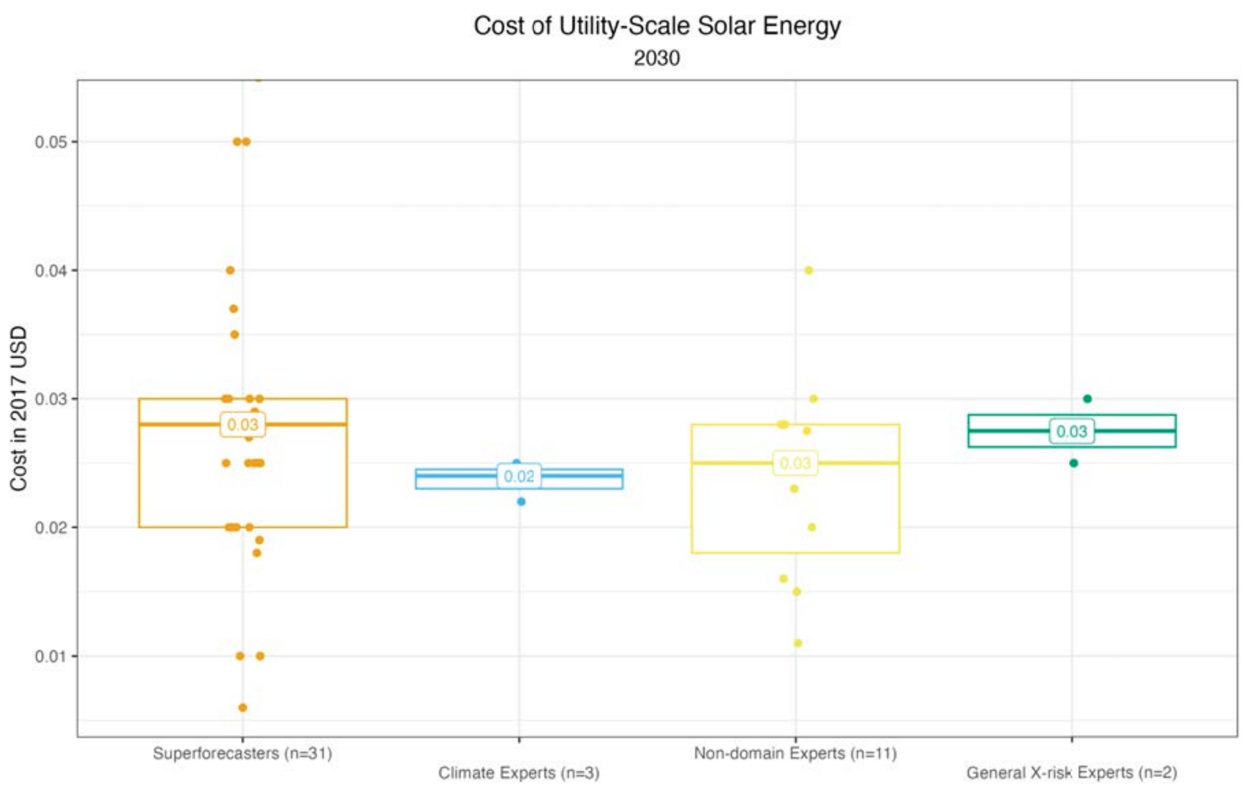
Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 31)	2024	0.05	0.038	22.43	+428.42%
	2030	0.03	0.028	20.73	+376.59%
Domain Experts (N = 3)	2024	0.37	0.033	NA[818]	NA
	2030	0.12	0.024	NA	NA
General X-Risk Experts (N = 2)	2024	0.07	0.055	NA	NA
	2030	0.03	0.028	NA	NA
Non-Domain Experts (N = 11)	2024	0.06	0.04	2.19	-99.58%
	2030	0.04	0.025	1.2	-99.32%

[817] Numbers of forecasters are given as of Stage 4 of the XPT.
[818] Only one forecaster in this group answered this question in Stage 1.


Cost of Utility-Scale Solar Energy - 2024 - 50th %
0.055
0.050
0.045
0.040
0.035
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 31)
Experts (n = 16)
Non-domain Experts (n = 11)


Cost of Utility-Scale Solar Energy - 2030 - 50th %
0.030
0.029
0.028
0.027
0.026
0.025
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 31)
Experts (n = 16)
Non-domain Experts (n = 11)


Cost of Utility-Scale Solar Energy
2024
Cost in 2017 USD
0.07
0.06
0.05
0.04
0.03
0.02
0.04
0.03
0.04
0.06
Superforecasters (n=31)
Climate Experts (n=3)
Non-domain Experts (n=11)
General X-risk Experts (n=2)


Cost of Utility-Scale Solar Energy
2030
Cost in 2017 USD
0.05
0.04
0.03
0.02
0.01
0.03
0.02
0.03
0.03
Superforecasters (n=31)
Climate Experts (n=3)
Non-domain Experts (n=11)
General X-risk Experts (n=2)

Sources of agreement, disagreement and uncertainty

- **Common starting points**
    - Several teams used Lazard's Levelized Cost of Energy (LCOE) report and/or the SunShot 2020 report, particularly its figure of 4.6¢/kWh in 2020, as a starting point.[819]
    - Some forecasters arrived at different estimates of past declines in cost of solar power, most likely because they used different starting points.[820]
- **Battery storage**
    - Forecasters expressed uncertainty over how increased focus on battery storage will impact the cost of new solar power.[821]
- **Uncertainty about impact of inflation**
    - Some forecasters ignored the impact of inflation since the question asks about prices in constant 2017 dollars. Others believed that inflation could differentially affect the energy sector or be a source of uncertainty.[822]

Arguments given for lower forecasts

- **Strong public support and investment in solar**

---

[819] "I use Sunshot's 2020 report of $0.046 as a starting point. I use trend data from: https://www.lazard.com/perspective/levelized-cost-of-energy-levelized-cost-of-storage-and-levelized-cost-of-hydrogen/ It looks like the normal price decline is around 10%/year." (T340)
"The team gets to 3.9 cents either by using the CAGR from the Lazard LCOE report as a base rate (the CAGR from 2016-21 was 8%) or by applying a log curve (R&D and development costs over the last 10 years have followed a power law function, so with cost+ based pricing, and EBITA returns in the 7-10% range, costs will follow a log curve)." (T341)
"Following the trends from https://www.lazard.com/perspective/levelized-cost-of-energy-levelized-cost-of-storage-and-levelized-cost-of-hydrogen/" (T342)
"*Note: Our estimates are based on LCOE.*" (T343)
[820] "I use Sunshot's 2020 report of $0.046 as a starting point. I use trend data from: https://www.lazard.com/perspective/levelized-cost-of-energy-levelized-cost-of-storage-and-levelized-cost-of-hydrogen/ It looks like the normal price decline is around 10%/year." (T340)
"The 2010 price was 27¢/kWh, while 2020 price was 4.6¢, an average decrease of ~16% per year. [source] This implies a decrease to 2¢ by 2024 and a decrease to 0.65¢ by 2030." (T337)
[821] "One source of uncertainty is how battery storage will factor into new power plants. Previous reporting has separated battery and PV costs, and future reports will likely continue to do so. However, if power plants need to include battery storage, the cost of the whole project will likely increase, which could impact the size and/or number of new plant production undertaken. Smaller plants will likely cost more than larger ones." (T343)
"In addition it seems like more money is being devoted into figuring out the storage problem; solar is already competitive and cheaper than most alternatives, the issue that organizations will now focus on is storage." (T345)
[822] "While the impacts of inflation are partially offset by the forecasts being in 2017 constant dollars, it was noted that price increases in manufacturing have outstripped general inflation…It is anticipated that some expert forecasters might discount the impact of inflation as a source for uncertainty / as a source for a wider tail towards higher prices - especially given the 2017 prices directive, which might cause people to feel it can be ignored entirely." (T337)
"Future possibilities for prices rising include runaway inflation, supply chain problems and possible sanctions against China. These should be taken into account for the 75th and 95th percentiles for 2024." (T341)

    - Continued investment in solar, especially relatively inelastic demand from the public sector, will drive solar R&D.[823]
    - The (U.S.) Inflation Reduction Act of 2022 includes incentives for solar power, including incentives for American solar panel manufacturing.[824]
- **High oil and gas prices will drive strong demand for renewable energy**[825]
    - Energy independence concerns also motivate investment in solar, especially given the war in Ukraine[826]
- **Increased capacity is due to come online**[827]
- **Solar cost has decreased faster than past estimates**
    - The U.S. Department of Energy's SETO has underestimated the decrease in cost of solar in the past (for example, its 2020 SunShot cost goal of 6¢/kWh was reached three years ahead of schedule).[828] Some forecasters adjusted downwards from SETO estimates for this reason.[829]

---

[823] "Forecasters considered that with high oil & gas prices currently, investment in R&D continuing to make advancements, and continued strong support from the public sector for renewables, we should expect this trend to continue for the near-future." (T337)
"We foresee continued investment and efforts in R&D to make solar a major portion of many countries energy portfolio in order to decrease reliance on foreign imports." (T345)
[824] "The Inflation Reduction Act is intended to bolster production of US made solar panels, probably reducing the cost of photovoltaic solar systems by increasing the R&D." (T344)
[825] "Forecasters considered that with high oil & gas prices currently, investment in R&D continuing to make advancements, and continued strong support from the public sector for renewables, we should expect this trend to continue for the near-future." (T337)
"Demand for non-hydrocarbon energy solutions is likely to increase - especially in Europe - as the conflict with Ukraine obstructs the NG market between Europe and Russia. This increased demand may temporarily raise costs of PV in[s]tallation, while developing a much larger, broader market for alternative energy solutions." (T343)
[826] "The last year of more global conflict and instability has shown nations the importance of gaining energy independence which solar can provide. We foresee continued investment and efforts in R&D to make solar a major portion of many countries energy portfolio in order to decrease reliance on foreign imports." (T345)
[827] "*Base rate(s): "*drop of 24% in LCOE for each doubling of the installed capacity. Over 2014-2020 this was actually steeper, with a 45% drop for each doubling. Just the same, looks like capacity at end of 2022 will have doubled since sometime in 2018/2019. Two more doublings expected by 2030." (T336)
"Large amounts of additional capacity is due to come online, as per Bloomberg. Lawrence Berkeley National Laboratory found more than 1 terawatt of potential new power generation or storage capacity that has requested connection to transmission networks. To put that in perspective, the whole world hit 1 terawatt of installed solar capacity earlier this year." (T341)
[828] "Solar cost targets, such as from the US Gov 'SunShot' initiative have [...] vastly underestimated the scale of continued cost decreases, and forecasters should update accordingly." (T337)
"SETO has a goal for 2030 of 2 cents and were 3 years ahead of pace in 2017." (T341)
"The 2020 SunShot cost goal was reached in 2017 (a reduction of 0.06 USD per kWh). This was 3 years ahead of schedule, so the 2030 goal seems more reasonable." (T344)
[829] "An updated version (08/13/2021) of www.energy.gov/eere/solar/articles/2030-solar-costs-targets indicates optimistically that utility rates will fall further than previously indicated, with 2030 estimates dropping from $0.03/kWh to $0.02/kWh…The energy department has a solid history of under-estimating [the reduction in] future costs. Forecasts using their current estimates will most likely be too high." (T338)

Arguments given for higher forecasts

- **A trend of decreasing costs has plateaued recently**
  - Forecasters observed that past rapid decreases in the cost of solar power have leveled off.[830] Different explanations were offered for this observation; some forecasters pointed to potentially short-term factors, such as supply chain issues and the polysilicon shortage. Others believed the leveling-off was due to longer-lasting economic factors such as the law of diminishing returns, solar tech entering the mature phase of the product life cycle,[831] or because the easy-to-reduce costs (low-hanging fruit) are gone.[832]
    - "The strongest argument for our teams medium forecasts are that despite huge strides made in bringing down solar costs over the last decade, we are increasingly seeing costs level off." (T345)
- **Supply chain problems**
  - Forecasters pointed to supply chain problems at a generic level[833] and specifically market disruptions due to COVID, the war in Ukraine, and issues with China.[834]
- **Polysilicon shortage**
  - The shortage of polysilicon, an important component of solar cells, has exerted upward pressure on the price of solar cells and may continue to do so.[835]

---

[830] "Looking at the NREL reports, its clear that innovation on cost has mostly flatlined over the last 3-5 years with costs barely improving 15% over that period." (T345)

[831] "The pace of cost reduction should be expected to slow due to diminishing returns." (T338)
"Solar generation in the utility segment (there are three segments: commercial, residential and utility) is entering into the mature phase of its product life cycle. The major drivers of hardware (major and replacement components), land, battery storage, taxes, system integration and operations administration are all in a mature phase." (T341)

[832] "Soft costs – administrative, etc. have been a big driver of improved efficiency. These reductions will probably level off." (T341)
"We believe the trend shows that this progress has plateaued and we will continue to only see marginal gains in the near future. Much of the low hanging fruit has been targeted especially around module costs." (T345)

[833] "Supply chain issues, especially microchips, may have slowed the trend, especially for 2024." (T336)

[834] "Energy shortages in Europe leading to factory shutdowns, continuing covid lockdowns in China, as well as increased shipping costs, and continued supply chain difficulties globally will lead to a discontinuity in the previously established trend, especially given concerns about the technology becoming increasingly mature, and with few significant cost-reducing advances yet to be made." (T337)
"Ukrainian conflict and various market disruptions (COVID), plus segmenting of economies under sanction between West and BRIC, may cause increasing supply chain issues that reduce R&D, increase costs of supplies, or shift priorities away from solar power." (T344)

[835] "By 2024, there may be an increase in prices in raw materials, specifically polysilicon, due to shortages and supply chain stress in general and because of the war in Ukraine in particular. Poorer relations between China and the West could cause China to drive up the price of polysilicon." (T339)
"The [cost decrease] trend reversed recently due to the polysilicon shortage that started in mid 2020…It's pretty clear that the main problem is that not enough factories were built to come online in the 2020 to 2022 time period. That's led to polysilicon companies making absurdly large profits, which is likely causing them to build new factories much faster. Eventually enough polysilicon capacity will come on line to get prices back near the original trendline. I think there's enough lag that costs of installed systems will be higher in 2024 than in 2020 even if polysilicon prices decline soon, and it's more likely that the big

- **Disruptions to Chinese supply chain**
  - China currently produces most solar panels globally. Forecasters pointed to potential cost increases resulting either from potential China/U.S. conflict (especially over Taiwan),[836] or sanctions or adverse U.S. trade policy towards China.[837] Any of these factors could increase the cost of solar panels.
    - "The highest source of disagreement still is the quantification of political risk tied to the main photovoltaic producer in the world, China." (T344)
  - On the other hand, supply chain problems with China could drive more investment in American solar panel manufacturing, which could decrease prices and reduce price volatility. First Solar, a U.S. manufacturer, is planning to invest up to $1 billion USD in a new facility.[838]

Other arguments given

Sources of uncertainty:

- Uncertainty about viability of wind versus solar
  - "Cost of solar vs wind generation per kwH favors wind, but wind on-shore and off-shore faces the battle of "ugliness" and NIMBY." (T341)

Arguments in favor of lower forecasts:

- Economies of scale will decrease costs.[839]
- Perovskite solar cells could become commercially feasible.[840]

---

decline in polysilicon prices is a year or two away. I'm guessing polysilicon will go through a boom-bust cycle with another boom peaking around 2028." (T340)

[836] "By 2030, the relationship between China and the West may be worse, especially if tensions over Taiwan escalate." (T339)
"For my 75% and 95% 2030 forecasts, I'm factoring in a 30% chance that a Taiwan-related US boycott of China screws up the market." (T340)
"Production moving out of China (potentially due to conflict with that country) could cause increased costs for PV systems." (T343)
[837] "Future possibilities for prices rising include runaway inflation, supply chain problems and possible sanctions against China…The Trump-era tariffs and trade policy toward China complicates the supply chain availability and pricing of panels." (T341)
[838] "The manufacture of solar panels within the U.S. has become investible, despite current economic headwinds. This is because China is becoming a less reliable source of manufacturing in general for both geopolitical and internal economic reasons. Today, most solar panels are being manufactured in China and South Korea, with only one company, First Solar, manufacturing in the U.S. First Solar just announced that it plans to invest up to $1 billion in a new factory. This means lower prices and less volatility for 2024 and likely into 2030." (T341)
[839] "The main driver for solar's dominance is larger and more sophisticated factories that increase economies of scale. Solar panels have overtaken wind turbines in China as photovoltaic manufacturers ramp up output." (T341)
[840] "$0.02 seems possible under current technologies. If a new technology such as perovskite solar cells become commercially feasible, then we might even see costs below $0.02." (T342)

- Improvements in manufacturing techniques and panel design (like multi-layer silicon) could decrease costs.[841]

Arguments in favor of higher forecasts:
- “Regulations and taxes could raise the costs, especially if governments become more protectionist of their fossil fuel industries.” (T339)
- “Due to the current world energy crisis, the direction of favorable energy sources may switch to nuclear, as photovoltaic grids have issues with providing a stable base load.” (T344)
- U.S. stakeholders are not doing much to encourage solar power because it doesn’t favor their interests.[842]
- Any major catastrophe (such as an escalation of the war in Ukraine) could interfere with solar power development.[843]

Cross-references with other questions

Q27: Nuclear Fusion Energy
Q28: Solar and Wind Energy
Q29: Annual Direct Air CO2 Capture
Q30: Cost of Hydrogen

Question 27: Nuclear Fusion Energy

By what year will fusion reactors deliver 1% of all utility-scale power consumed in the U.S.?

Question and resolution details, prior forecasts, and other relevant sources

---

[841] “For 2024, the team expects improvements to panel design, specifically multi-layer silicon, and improvements in manufacturing techniques…By 2030, there will be even greater improvements in panel design and manufacturing.” (T339)

[842] “[T]he [slight] increase in the **cost (in 2017 USD / kWh) for new utility-scale photovoltaic solar systems above 4MWAC in the United States…from** $0.019225 per **USD / kWh in 2024 to** $0.02175 per **USD / kWh in 2030** which the team agreed that it is largely due to global energy politics which US energy stalkholders are major players and are not doing much to encourage these kind of source of energy because it doesn't really favoured them in addition to the fact that govt too is not doing enough to encourage the population to use it thereby reducing the cost of production to be able to be cost effective for maximum usage and price reduction , hence the data set range distribution across the years…’ (T342)

[843] “Great catastrophe, whether man-made or otherwise, would very likely prevent expected increases in efficiency from being realized directly after, as other priorities would arise... Increased global instability (escalation of the Ukrainian conflict is one example) could greatly stall further advances.” (T344)

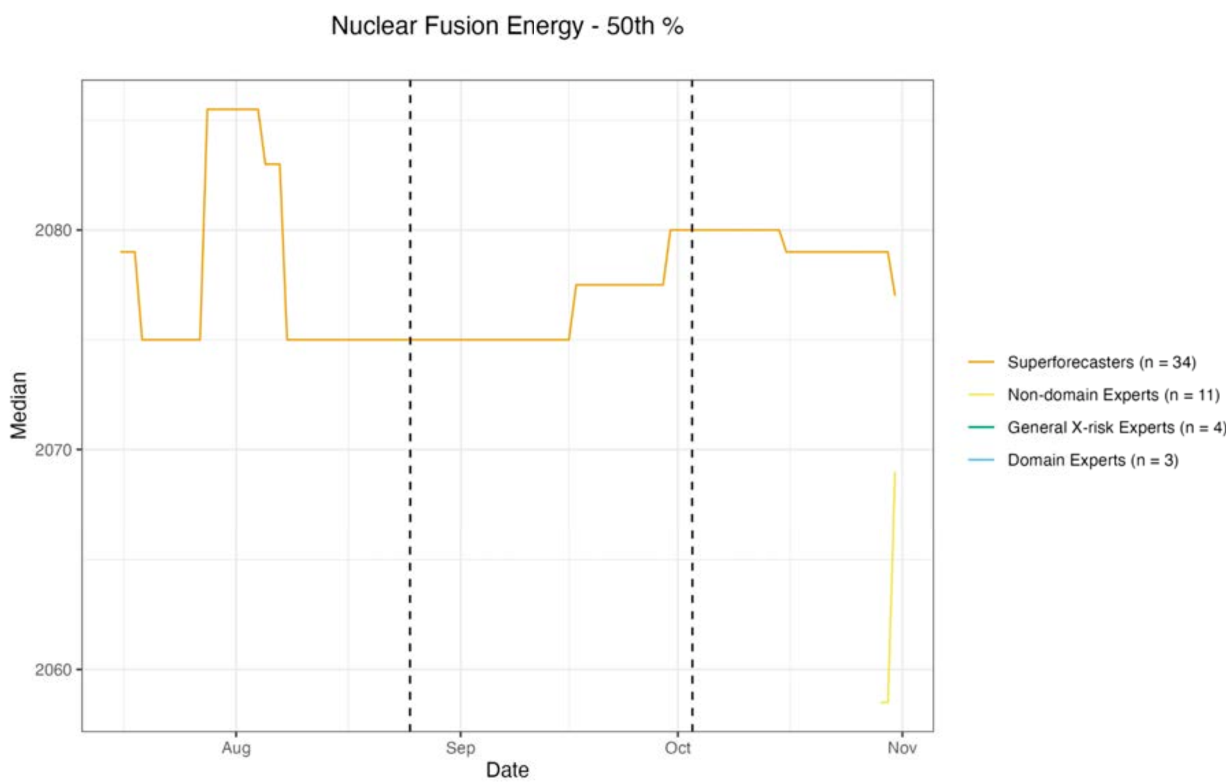
Results[844]

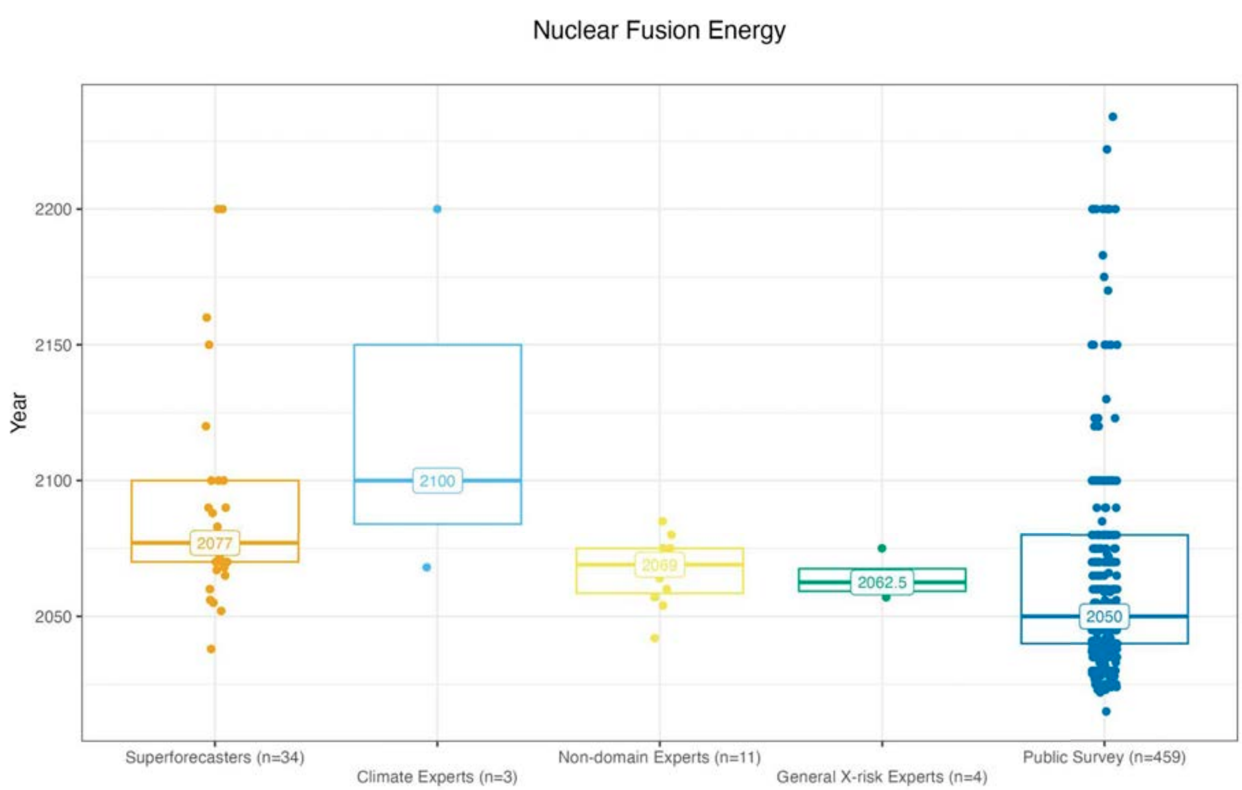
Group	Percentile Forecast	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 34)	5%	2050	2050	16.53	-28.43%
	25%	2066.5	2060	35.58	-33.01%
	50%	2079	2077	61.85	+572.14%
	75%	2105	2110	212739.59	-99.69%
	95%	2200	2587.5	Inf	NaN
Domain Experts (N = 3)	5%	2030	2062	NA[845]	NA
	25%	2050	2070	NA	NA
	50%	2080	2100	NA	NA
	75%	2150	2150	NA	NA
	95%	1e210 [846]	2500	NA	NA
General X-Risk Experts (N = 4)	5%	2041.5	2050	12.02	-27.96%
	25%	2049	2057.5	8.49	+5.85%

[844] Numbers of forecasters are given as of Stage 4 of the XPT. Public survey respondents were surveyed outside of the tournament context.
[845] Only one forecaster in this group answered this question in Stage 1.
[846] This value indicates a maximal response, in this case, it is effectively a response of “never”.

	50%	2058.5	2062.5	2.12	+271.93%
	75%	2072.5	2078	3.54	+681.79%
	95%	2117.5	2174	10.61	+4.7e19%
Non-Domain Experts (N = 11)	5%	2042	2044	9.07	-0.98%
	25%	2052	2057	9.5	+6.2%
	50%	2062	2069	10.35	+21.51%
	75%	2102.5	2090	39968.91	-26.14%
	95%	2180	2160	Inf	NaN
Public Survey (N = 459)	50%	2050		Inf	-

The teams' median forecasts were in the range of 2050–2100, for an overall median of 2070. The teams' median low-end (5th-percentile) forecast was 2043, while their median high-end (95th-percentile) forecast was 2202.5. At the high end, a wider range of forecasts was given, with many teams picking high-end forecasts of 2100–2180, while two teams' high-end forecasts were that fusion at commercial scale would happen in the further future (51100) or never.


Nuclear Fusion Energy - 50th %
Median
2080
2070
2060
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 34)
Non-domain Experts (n = 11)
General X-risk Experts (n = 4)
Domain Experts (n = 3)


Nuclear Fusion Energy
Year
2200
2150
2100
2050
2077
2100
2069
2062.5
2050
Superforecasters (n=34)
Climate Experts (n=3)
Non-domain Experts (n=11)
General X-risk Experts (n=4)
Public Survey (n=459)

Sources of agreement, disagreement and uncertainty

- **Is fusion on Earth impossible?**
    - Forecasters disagreed about how to weigh the possibility that it is physically impossible to recreate on Earth conditions needed for a fusion reaction.[847]
        - “A major source of uncertainty is that fusion has not yet been demonstrated that delivers more power than it consumes. Thus, this is a science problem, not an engineering or economic or even a political problem.” (T339)
        - "Fusion almost seems entirely like an engineering problem at this point. Once you build reactors with large enough dimensions getting to a level of Q>1 is achievable. The question[ ] is whether we dedicate the resources in order to get to that stage.” (T345)
        - “If a solution has not been developed by 2070, I have doubts as to whether or not nuclear fusion is possible on earth. Blindly assuming money and time will solve the problem assumes that all problems people put their minds to are achievable.” (T345)
- **Price of fusion compared to other renewable energy**
    - To develop and implement commercial-scale fusion, fusion would need to be not only feasible but demanded by the market. Thus, it must be available at a competitive price compared to other renewable energy sources for the question to resolve positively. The future price of solar and wind power, as well as the price of new renewable energy sources that might be developed, are all uncertain.[848]
        - “We should expect continuous development of energy sources in the >30 years before commercially-viable fusion power is demonstrated (potentially >>30 years).” (T343)
        - Even using optimistic forecasts, the projected cost of fusion power is still 3–4 times that of solar and more than 10 times the cost of natural gas.[849] (T341, T343)

---

[847] “General consensus that fusion is possible within the medium term future, and if not, that probably means that it's physically impossible to recreate the conditions required for a fusion reaction on Earth. Some believe it must be possible, while others agree there's a chance it could be beyond humanity's power.” (T342)

[848] “[E]ven if it is practically possible, there is still no guarantee that generation will [be] cost competitive for generation, especially given the rapid declines in the costs of wind and solar PV.” (T337)
“[N]ew renewable sources might be found that make fusion not commercially viable which would end the search for a fusion solution.” (T339)
“The economics may or may not work to compete with renewables.” (T341)
“Another cheaper source of energy could be found, making fusion utility power unnecessary.” (T344)

[849] “While the costs of the CFETR are not well known, an EU DEMO fusion concept was projected to feature a levelized cost of energy (LCOE) of $121/MWh." Compared to renewables this is very high. The 2019 levelized cost of solar was about $40.” (T341)
“[T]he most optimistic cost/kwh is > $40k-$120k for this [ITER] demonstration experiment. By 2050 (+ delays), DEMO projects could produce up to 2000 MW. Generously assuming the same optimistic ITER price for these projects, we would expect a four-fold reduction in cost/kwh. Optimistically, this would still

- **Timelines for other renewable-energy sources**
    - Teams used the commercialization of solar, wind, or nuclear fission power as a base rate for fusion development.[850] For example, about 15–18 years passed between the development of the first commercial fission reactor and its use to generate 1% of U.S. utility-scale power; wind power largely developed over a 17–year period; and solar power took 36 years to evolve from niche to mainstream.[851]
    - Overall, teams seemed to find the timeline from “first commercial fusion reactor” to “fusion reactors produce 1% of utility-scale power” easier to estimate compared to the time it would take to achieve the first commercial fusion reactor starting from the status quo.
- **How many reactors would be needed?**
    - Two teams that assessed the number of fusion reactors needed to generate 1% of utility-scale power came to remarkably different estimates: one assumed 50 100-MW reactors would be needed,[852] while the other assumed 4555–5500 reactors would be needed.[853]

Arguments given for low-end forecasts

- **Current fusion R&D projects**

---

be more than 4x the cost of current solar and wind power, and more than 10x the cost of natural gas (at the low end).” (T343)

[850] “Looking at the development of nuclear fission technology provides a good base case. Nuclear power did not supply 1% of US power until 1975. ...That was after a fairly fast pace of development, with the first power output to the grid in 1957. Fusion could be faster (<20 years) from the point it first is used to supply commercial power, but haven't yet gotten to that point. The time taken for other power sources from initially supplying power to the grid to reaching 1% of US power: solar: ~30 years, wind: ~20 years, geothermal: 50 years and counting.” (T336)
“The time between the first fission demonstration and 1% nuclear energy in the grid: ~20 years…The time between the discovery of the photovoltaic effect and 1% solar energy in the grid, also including the discovery of the photoelectric effect in 1905 by Einstein: ~105-170 years. Note that solar energy is much simpler than fusion. Another base rate is the time passed between the first nuclear fission commercial plant (1956) and the moment at which the 1% threshold was reached (1971-1973). This gives us a 10-20 year timeframe after the appearance of the first commercial fusion reactor.” (T337)
“Solar electricity generation has grown from near zero in 2001 to almost 2% in 2021 globally. It evolved from niche to mainstream between 1992 and 2018, 36 years. And that was with a lot of subsidies. In 2018, global installed utility-scale was 180 GW. First commercially manufactured in 50’s. Nuclear fission generated its first electricity in 1942 and it grew rapidly in 60s. Wind reached 1% around 2007. First wind turbine built for generating electricity in 1887. But wind largely seen to have developed in 1973-2000 period. First multi-megawatt in 1978.” (T341)

[851] See footnote 850.

[852] “To generate the estimated needed 1% power assuming a base rate per reactor of 100 MW could require about 50 reactors in the US at the time of the median year, **2057**.” (T339)

[853] “Helion's **fusion** starts with deuterium, a form of hydrogen found in water. One glass of DO: Generates 9 million **kWh** of safe, clean electricity…We would need 4555 such power plants to supply 1 % of energy to USA…Since power needs of USA are expanding, we can round the number to 5500.” (T344)

- Fusion research and development is attracting more resources and interest than before. Very optimistic forecasts expected that one of the current projects would succeed, such as ITER (scheduled to begin by 2025), or CFS (planning to develop a demonstration device by 2025).[854]
  - On the other hand, these projects "are incentivised to make their designs appear more promising than they actually are, and few individuals have the requisite expertise to discern this." (T337)
- **Record private investment in fusion**
  - "SpaceX-ification" of fusion could lead to faster private timelines compared to public sector projects.[855]
    - "The success of SpaceX in the field of reusable rocketry should provide us with some confidence that lengthy publicly funded timelines can rapidly become obsolete once commercial players become involved." (T337)
- **Climate-change-driven need for renewable energy**
  - Increasing concern about climate change and growing demand for energy will increase the demand for and resources invested in cheap renewable energy.[856]

---

[854] "Several of these forecasters pointed to the ITER project under construction in France with expected operations projected to begin by 2025 as particularly promising. Multiple forecasters also made arguments pointing to the extensive "..research activity, interest, and actual construction money being spent to demonstrate the viability of fusion." (T336)
"More optimistic forecasts saw fusion reaching the benchmark roughly 10-20 years earlier than the median projection. These forecasts were reliant on the success of one or more of the commercial offerings currently under development, which tend to have much more aggressive timelines than ITER." (T337)
"Low End Forecasts…SpaceX-ification of fusion and Commonwealth Fusion Systems timeline combined with 9 total years to build 3 plants of sufficient size for earliest p05 date." (T338)
"More optimistic considerations that some team members emphasize include: Recent advances in fusion energy generation in the Startup domain and multilateral experiment ITER (experimental reactor)." (T344)
"[T]he funding for fusion has been pretty pathetic given its potential but lately there seems to be some genuine efforts at changing that…[as a low-end forecast,] SPARC achieves commercially relevant net energy from fusion by 2025." (T345)
[855] "Recent entry of commercial players into the space previously dominated by publicly funded mega projects. These are encouraging for two reasons, firstly that private venture capital would be willing to fund these projects suggests at least some outside confidence that the companies could potentially generate a return, and second that private companies provide an alternative development pathway with much more rapid iteration and shorter projected timelines, adding some robustness to a field that was previously entirely reliant on public funding." (T337)
"Low End Forecasts…SpaceX-ification of fusion and Commonwealth Fusion Systems timeline combined with 9 total years to build 3 plants of sufficient size for earliest p05 date." (T338)
"Private investors poured a record $3.4 billion into fusion last year. Private development may finally pay off in a big way in potential market size of 40 trillion." (T341)
[856] "The most optimistic forecasts put the odds at 50% that commercial fusion power will be achieved between 2040 and 2060. These forecasters cited the very high value of unlimited inexpensive energy to humanity and suggested that due to the additional benefit to reducing global warming, perhaps even more time, money, and attention should be devoted to accelerating this project." (T336)
One forecaster "mentioned 10-15 years for power plant construction with slow permitting offset by climate change related urgency." (T338)
"As climate change continues to get worse, fusion look likely to continue to receive more attention and investment." (T341)
"Power demands will likely continue to increase regardless of supply." (T343)

- A climate or geopolitical crisis could motivate increased fusion funding to reduce dependence on fossil fuel sources or improve robustness of energy supply.[857] (T342, T343, T345)

Arguments given for high-end forecasts

- **Past lack of progress**
  - Fusion has been known to be theoretically possible for decades, yet has not been developed even at the proof of concept scale. It's "always just around the corner."[858]
    - "Fusion research has already taken decades and that we are still unable to reach even Q should make us skeptical this can be done in decades or centuries of additional effort." (T337)
- **Cost considerations**
  - Fusion has large startup and capital investment costs,[859] and could have unexpectedly high construction and engineering costs,[860] perhaps due to the Baumol effect or general cost disease.[861] Investment in fusion, although higher than in the past, is still low relative to what would be required.[862]

"Assuming a rapid advancement in the technology required to accomplish this feat (advanced materials science), the earliest 1% will occur will be 2055. By this point in history, society will no longer be able to rely heavily on oil and gas due to its finite supply and policy restrictions likely in place at the time. R&D efforts will begin accelerating in the 2030s to find a solution to a problem that could potentially become more dire as each decade passes" (T345)

[857] "It's also possible that a crisis, such as the one we're currently experiencing, could significantly increase global funding and resources assigned to fusion research, which could force a breakthrough and would speed the development process significantly." (T342)
"One argument in favor of fusion energy is diversification. It could be that some utilities (or the state) include fusion in their portfolio just for the sake of robustness even if it comes with a premium…A significant source of disagreement in this group is confidence as to whether a mainly wind/solar/batteries-based electric grid will be economically viable or whether "firm sources" such as, one day, potentially fusion reactors will be needed even if they produce electricity at a higher cost per unit of energy." (T343)
"Team members believe that geopolitical influences could play a part, either positively (e.g. arguing that after CFS produces their first nuclear fusion plant [estimated early 2030's], "... there will be rapid take-up of this technology particularly in the 'West' to reduce dependence on autocratic fossil fuel sources") or negatively (e.g. "geopolitical turmoil slowing things down")." (T345)

[858] "It was also noted that fusion power generation has been seemingly within close reach for many decades but has still failed to arrive." (T336)
"Nuclear fission has been around for a while now while nuclear fusion has always been the solution that's been around the corner for a long time now. So, the team considered that as a baseline and also considered the tactical challenges of there being no TD [technology demonstrator] yet for fusion." (T340)

[859] "Fusion plants are expected to face large start-up and capital costs." (T341)

[860] "Two forecasters indicated skepticism that commercial fusion will ever become viable with 95th percentile forecasts of 100,000 years or greater. The rationale for these forecasts are unexpectedly high construction and engineering costs leading to high electrical generation costs that make fusion power generation less attractive than alternatives." (T336)

[861] "A lot of this depends on how large and expensive fusion power facilities would have to be and whether they run into the same cost prohibition problems as fission plants/high speed rail/etc." (T341)
"General cost disease (Baumol's or otherwise) makes all major projects more expensive." (T344)

[862] "Investment in fusion still remains low (see illustration of actual to needed)." (T344)

        - “The optimism surrounding fusion appears to ignore compounded probabilities: delays or failed success of ITER could easily derail future DEMO experiments and funding.” (T343)
- **Existential risk**
    - An existential or catastrophic threat to humanity could result in humanity lacking the capacity or willingness to devote resources to fusion research and development.[863]

Other arguments given

Arguments in favor of lower forecasts:

- AI progress could lead to faster fusion development,[864] but could also lead to more regulatory scrutiny.[865]
- Nuclear fusion is safer and produces less waste than nuclear fission plants.[866]

Arguments in favor of higher forecasts:

- Advances in material science, like improvements in superconductors, metamaterials, and composites, would be required or desirable for low-end forecasts.[867]
- Regulation and lack of political will could add delays.[868]
    - “Publicly funded slow-developing technologies often come into conflict with the short political cycles of a few years.” (T337)
    - NIMBYism could create problems with siting fusion plants.[869] (T344)

---

[863] “Both domain-specific experts who forecast this question had forecasts on the lower side except for their 95%ile forecasts which were infinity or essentially that, due to potential for destruction of humanity or the US, or alternatively development of better energy sources.” (T336)
“The existential risk would put a kibosh on fusion development. The catastrophic risk could shift priorities away from fusion to more pressing matters.” (T344)
[864] “Technical breakthroughs might be possible with AI assistance.” (T336)
“Continued application of AI could introduce a discontinuity in the pace of fusion development, resulting in viable fusion generation faster than previously anticipated.” (T337)
[865] “There is a minority view that this is a smaller challenge than discussed and the argument cited is progress made in DeepMind. But in my opinion as someone who works in computing and AI, [it's] hard for AI to actually explain to regulators that this is solid and ask for an accelerated approval for controlling something as dangerous and critical as nuclear fusion as the risks of a mistake are potentially unfathomable. It makes even more imperative, in a lot of the forecasters’ opinion, that this needs even more scrutiny, not less.” (T340)
[866] “Fusion has many advantages over fission in safety and less waste.” (T341)
[867] “[I]mprovements in superconductors, metamaterials, and composites could help overcome current restrictions in tokamak and other designs.” (T343)
“Assuming a rapid advancement in the technology required to accomplish this feat (advanced materials science), the earliest 1% will occur will be 2055.” (T345)
[868] “There [were] also some brilliant points that were mentioned on how this needs to be regulated and also approved by the relevant regulatory authorities which adds more time to the timeline.” (T340)
“Main sources of uncertainty were relative to funding and political will to develop power plants at the scale needed to achieve the 1%.” (T345)
[869] “Not In My Back Yardism seems like a growing force, although there are increasing efforts to fight it.” (T344)

Cross-references with other questions

Q26: Cost of Utility-Scale Solar Energy
Q28: Solar and Wind Energy

Question 28: Solar and Wind Energy

What percentage of the world's electricity will be provided by solar energy and wind energy combined…
...in 2024?
...in 2030?

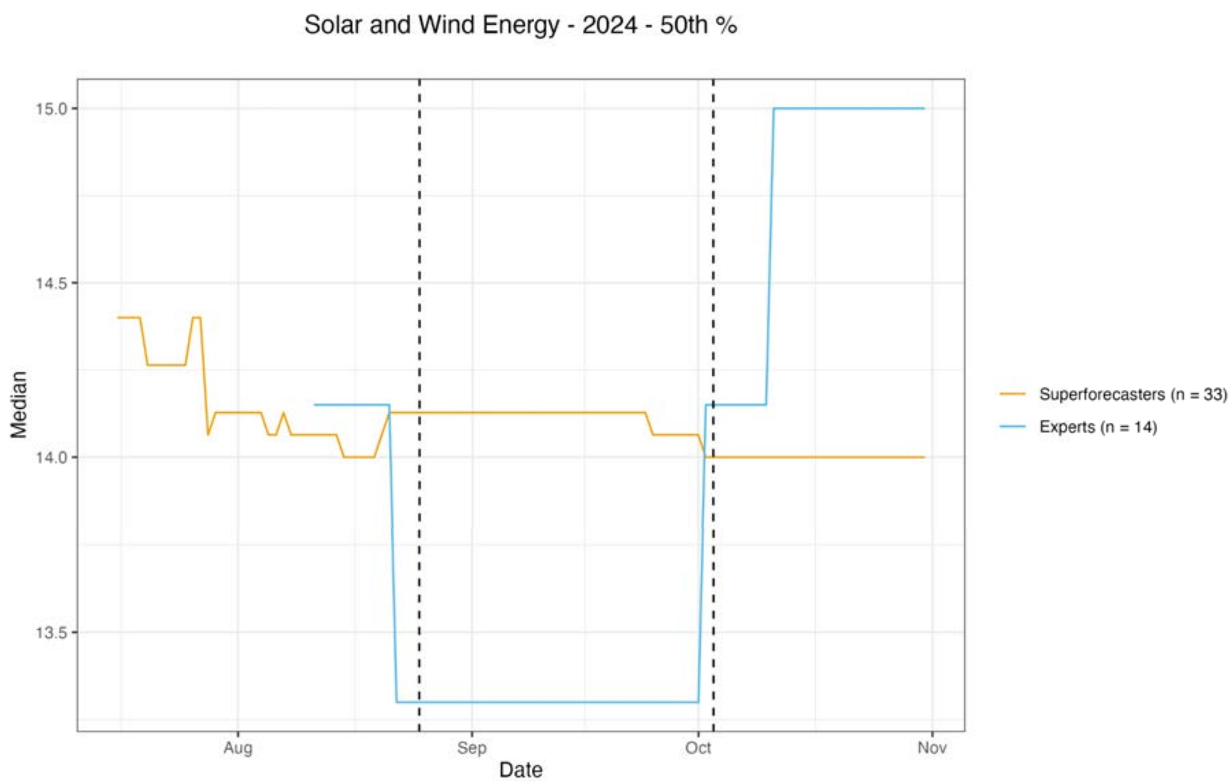
Question and resolution details, prior forecasts, and other relevant sources

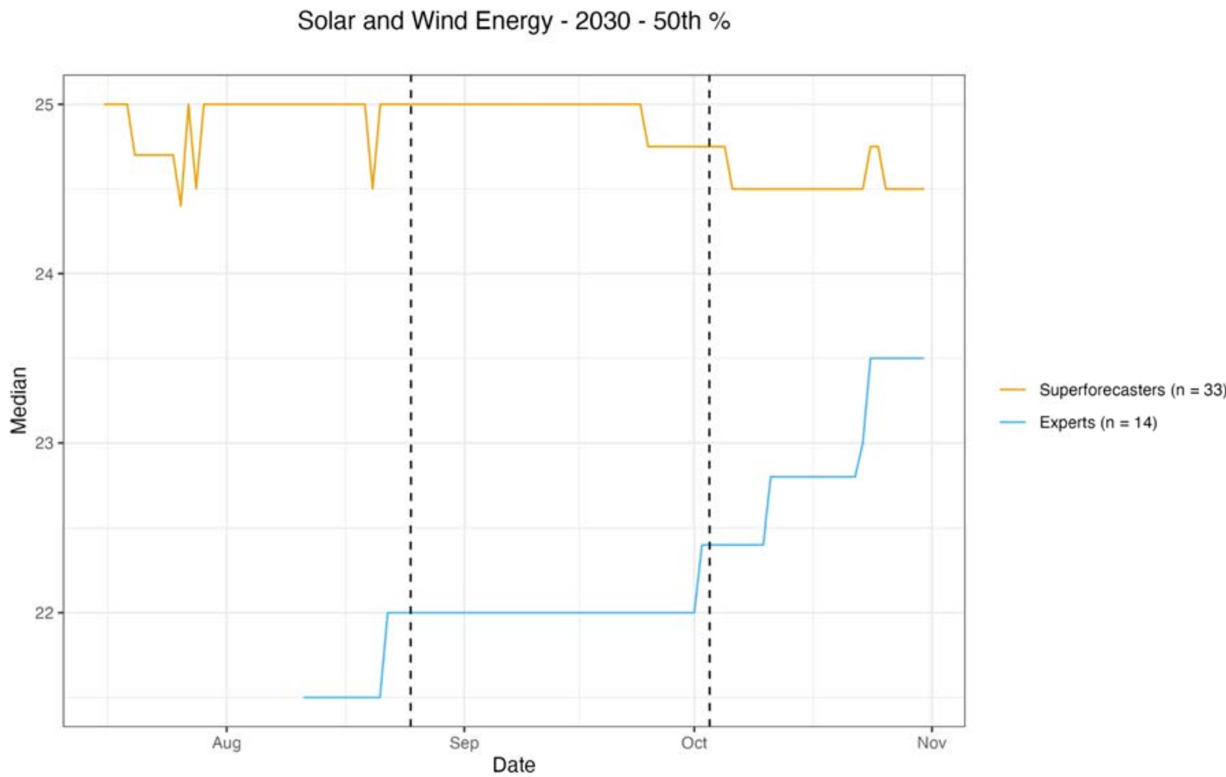
Results[870]

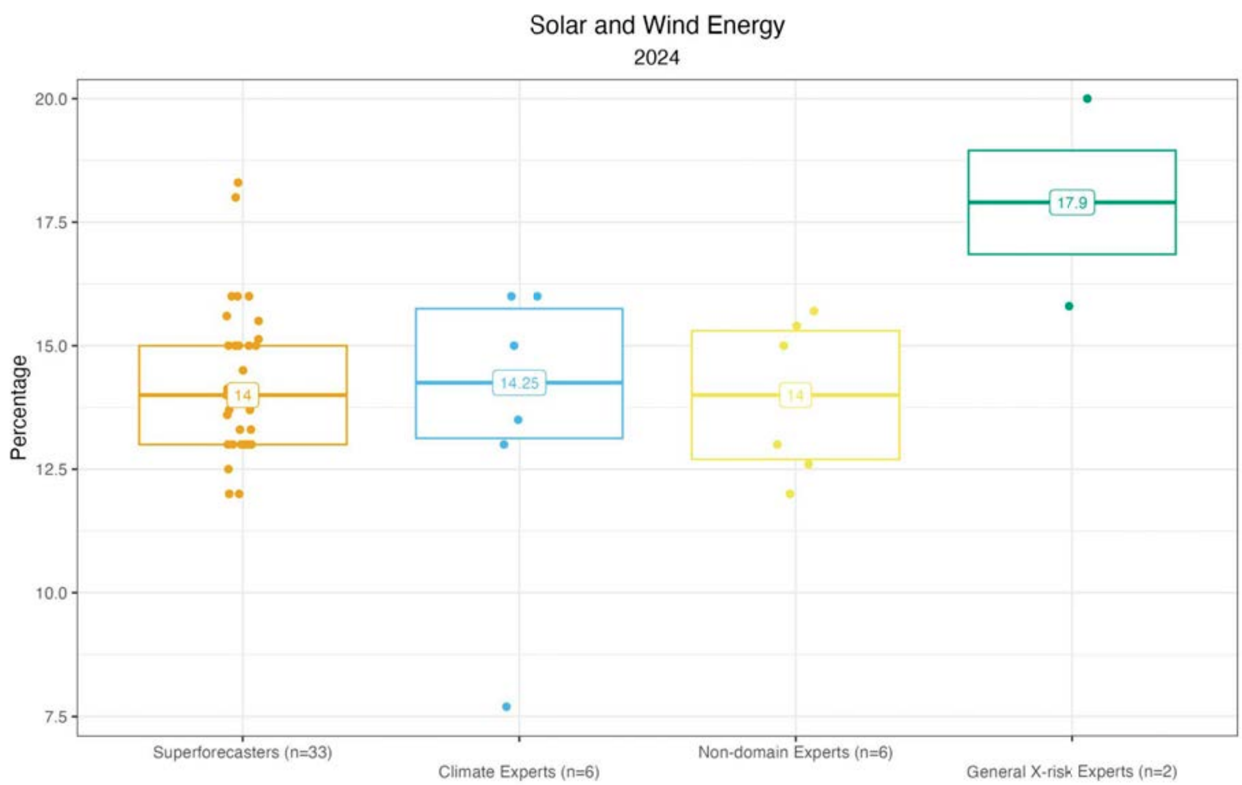
Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 33)	2024	14.13%	14%	5.27%	-71.16%
	2030	24%	24.5%	10.69%	-42.06%
Domain Experts (N = 6)	2024	15.5%	14.25%	0.71%	+341.06%
	2030	22%	22.9%	2.83%	+102.51%
General X-Risk Experts (N = 2)	2024	15.8%	17.9%	NA[871]	NA
	2030	30%	27.5%	NA	NA

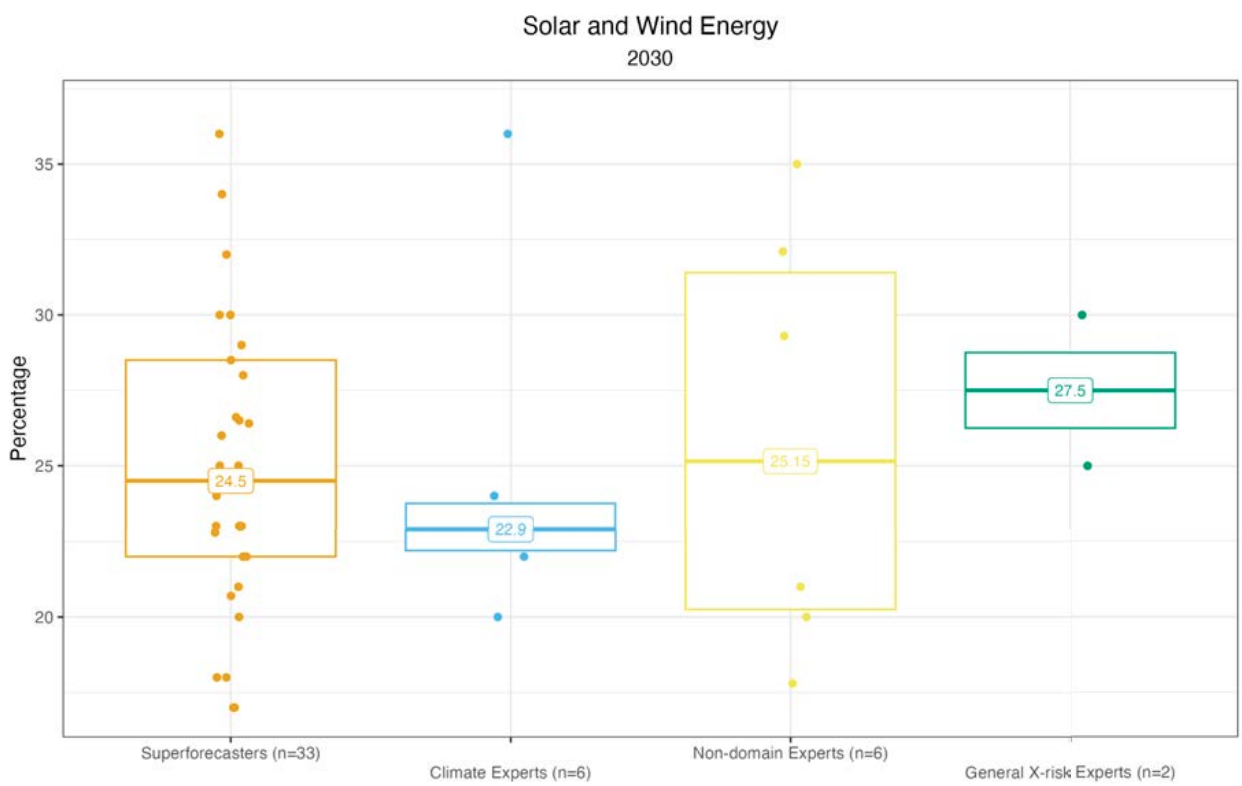
[870] Numbers of forecasters are given as of Stage 4 of the XPT.
[871] Only one forecaster in this group answered this question in Stage 1.

Non-Domain Experts (N = 6)	2024	12.8%	14%	1.49%	+7.28%
	2030	20.5%	25.15%	6%	+19.53%




Solar and Wind Energy - 2030 - 50th %
Median
25
24
23
22
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 33)
Experts (n = 14)


Solar and Wind Energy
2024
Percentage
20.0
17.5
15.0
12.5
10.0
7.5
14
14.25
14
17.9
Superforecasters (n=33)
Climate Experts (n=6)
Non-domain Experts (n=6)
General X-risk Experts (n=2)


Solar and Wind Energy
2030
Percentage
35
30
25
20
24.5
22.9
25.15
27.5
Superforecasters (n=33)
Climate Experts (n=6)
Non-domain Experts (n=6)
General X-risk Experts (n=2)

Sources of agreement, disagreement and uncertainty

- **Is growth in solar and wind power linear or exponential?**
    - An important source of disagreement was whether observed historical growth in the wind and solar share of the world's electricity was linear or exponential, and whether, if the growth has been linear in the past, it will become exponential.[872] Several forecasters noted a growth rate of roughly 1% per year over the past ten years, but they interpreted this data differently.[873]
    - One team believed that this disagreement accounted for a bifurcation in their team's 2030 forecasts.
        - "From 2011 to 2021, the combined usage increased by ~8% points, or just below 1%/year. Alternatively, reviewing the same data could also reveal an exponential growth of combined usage doubling approximately every 5 years. This dichotomy of interpretation appears to be supported in the difference between the 2024 and the 2030 forecasts...there are no forecasts in the 2030 median range as those who foresee the incremental increase of 1%/year are forecasting in the low 20% range, and those who see an exponential growth rate are forecasting in the upper 30% range. *The difference between the two groups of forecasters is based on whether expected growth will be linear or exponential."* (T338)

---

[872] "There was some discussion about whether the rate would remain linear or would become exponential as the cost of solar and wind drop in comparison to fossil fuels. The final consensus leaned more toward exponential than linear." (T336)
"Considering the development in the last years, certain forecasters felt that growth rates for solar electricity production [have] been exponentially increasing, from 0.1% per year to 0.6% per year between 2019 and 2020. Wind production has similarly grown exponentially, and rose from 0.8% between 2020 and 2021. In total, the total % of energy derived from wind and solar is clearly growing exponentially. One higher-end forecast for example simply fitted a logistic regression to historic energy data and extrapolated accordingly from there." (T339)
"The impact of inflation and supply chain issues on projects, weighed against the expectation of continued exponential growth in renewable deployment." (T341)
"The rate of growth, increase and penetration of solar and electricity in the years in question differs from one for[e]caster to the other based on their individual perspective views on the performance of solar and wind electricity so far, while some see a slow growth rate others see a fast growth and the remaining faster growth rate percentile." (T342)
[873] "Combined usage grew 8% over the last 10 years which averages out to about 1% per year." (T338)
"[R]enewable electricity production has risen in total but [] growth rates are stagnant on a level of ~1% each year." (T339)
"According to current trends, the share of wind and solar energy increases by about 1% per year." (T342)

  - Other teams looked at compound annual growth rates with various start dates or growth rates over the past four years for their projections.[874]
- **Uncertainty over future role of China in solar/wind markets**
  - What happens in China will significantly affect the resolution of this question, both because of China's dominant role in solar panel manufacturing and because Chinese investment in renewables, which currently looks strong, could change in an economic crisis or if priorities change.[875]
    - "The potential for a change in China's development plans could lead to fewer renewables deployed. However, this possibility is unlikely due to China's 14th 5-year plan commitments and other investments in renewables." (T341)
- **Low variability of 2024 forecasts**
  - Little deviation from the current trend is expected by 2024.[876]
- **Uncertainty about the incentivizing effect of U.S. Inflation Reduction Act**
  - "Some forecasters are optimistic about the US Inflation Reduction Act impact on renewables growth, while others expect the effect to be small." (T341)

---

[874] "Our median forecasts extrapolated from growth rates over the past four years, which averaged 23% relative annual growth for solar and 12% for wind." (T340)
"Base rates:

- Combined: 10.1% in 2021.
- Wind 6.7%. in 2021.
- Wind compound annual growth rates: 12% since 2011, 19% since 2015, 12% since 2018
- Solar 3.4%. in 2021.
- Solar compound annual growth rates 25% since 2011, 17% since 2015, 17% since 2018

Source: https://www.iea.org/data-and-statistics/charts/solar-pv-and-wind-generation-by-scenario-2010-2030" (T341)

[875] "China is already exceeding their official targets in terms of renewables rollout. It may just be that we will see an acceleration in Asia that is easy to underestimate from today's perspective. However, other members pointed out that China seems now to be facing an economic crisis which might alter this pattern." (T339)

[876] "Solar and Wind Energy to be online by 2024 is already at the planning and approval stage now. Current geopolitical events and announced investments in renewables will probably not affect the capacity in 2024 by much. The majority of the team based their prediction on the forecasted 2024 base line by IEA and others. This base line is derived from what is already in situ. Hence variation between 2024 forecasts by the team is low and the confidence interval of individual forecast is also low. Our confidence interval increases considerably for the 2030 forecast." (T337)
"For the 2024 forecast, lower-end forecasters felt that it was unlikely that we will see large deviations from this pattern. Thus, they felt it reasonable that there won't be more than 16% wind and solar electricity production by 2024." (T339)
"There is a strong consensus on this question, but when differences occur, they tend to be in the higher percentiles for 2030." (T340)

Arguments given for lower-end forecasts

- **Intermittency problem and regulatory delay**
    - Incorporation of solar and wind power into the grid is challenging because of the intermittency of these power sources.[877]
        - “Challenges in incorporating renewables into the electric grid due to the intermittency problem could lead to recent trends to slow down if solutions don't appear or are too costly to implement quickly.” (T341)
        - “These sources of electricity are intermittent, requiring alternate power sources to provide consistent loads. The benefits of solar and wind in electrical grids as currently structured may significantly decline once they're producing moderate amounts of electricity to the grid.” (T343)
    - Value deflation limits the economic returns for wind and solar power at higher deployment rates.[878]
    - Relatedly, large energy projects, including changes to the grid, face regulatory delay and typically have slow timelines.[879]
- **Difficulty of finding appropriate sites for solar and wind installations**

[877] Some team mates pointed out that the ability of countries with less developed electrical grids to usefully integrate substantial intermittent power generation, limiting the adoption of wind and solar beyond ~25%-35% without substantial additional investment. This was contested by others as countries/regions such as Denmark, California and the UK already have 25% or 30% of electricity coming from Wind and Solar. They think that a mix of 50% or more Wind and Solar is possible, which is far more than the team's forecasts for 2030.” (T337)
“Also, growth may be held back by barriers to grid integration, including permitting issues and the need for new transmission lines.” (T338)
“Forecasters cite grid issues and supply chain issues as a reason not to expect a sudden acceleration in the short term.” (T341)
“For example, California currently sources about 25% of its total electric generation from solar and wind, and has experienced significant power disruptions during period when these power sources are not at peak operation.” (T343)
[878] “The variability challenge of these sources mean that their economic value decreases at higher penetrations ("value deflation"). Unless further significant cost reductions occur, subsidies or mandates will be required to further increase the share of these sources. Value deflation is expected to play a minor role at low percentages, but will become a greater factor over time and as market penetration expands.” (T343)
[879] “[Teammates] suggested that growth in the future will be driven by factors other than price such as regulations for building wind parks and solar might tighten and the ability of countries to usefully integrate additional intermittent power.” (T337)
“Time taken to plan large facilities. Energy projects on a grand scale needed significant planning and related reform not obviously forthcoming in West.” (T339)
“Some forecasters note that regulatory burdens and slow deployment of projects make grid improvements difficult to scale quickly.” (T341)

    - There is a scarcity of appropriate sites for large-scale solar power,[880] and environmental and NIMBY groups may slow or oppose wind projects.[881]
        - "Team mates pointed out that the speed of increase is largely determined by government agencies that need to plan or coordinate the grids, electricity storage system, and encourage or discourage planning rules." (T337)
- **Nuclear power could outcompete solar and wind**
    - Nuclear power, either fission or fusion, could outcompete solar and wind power while helping countries to meet emissions reduction goals. This possibility is dependent on the political acceptability of nuclear fission power and the development of fusion technology.[882]
- **Inflation and economic decline**
    - Inflation (particularly in cost of raw materials) and a potential recession could delay new investments in wind or solar power.[883]

---

[880] "Specific risk factors related to solar power:…
- Land use requirement of solar, with the possibility of it becoming increasingly difficult to find adequate sites considered an issue by forecasters." (T339)

"Raw materials inflation and reluctance to install wind production in environmentally sensitive areas (ie, offshore) could plausibly place a gap on alt-energy growth." (T339)
"The best geographic spots for these sources have in some countries has been taken. NIMBYism is increasingly becoming an issue." (T343)
[881] "The best geographic spots for these sources have in some countries has been taken. NIMBYism is increasingly becoming an issue." (T343)
[882] "There was some disagreement about the risk that next-gen technologies (fusion was mentioned) demonstrate that solar/wind are no longer competitive, and that this would happen within the horizons of this question. This was a view held by a single forecaster and attracted some dispute." (T339)
"Sources of uncertainty are…uncertainty as to the political impact of energy rationing and high costs in Europe (will it spur greater investment in solar and wind, or to abandoning them for nuclear)..." (T340)
"One major uncertainty in the outlook with respect to the longer term is whether nuclear technologies will provide a cost-competitive alternative for decarbonization. There is some disagreement on whether nuclear fission can overcome its relatively bad public picture and whether nuclear fusion has a serious chance of becoming a viable alternative. Buildout of nuclear fission reactors would likely take much longer than 8 years to have a significant impact on solar/wind buildout." (T343)
"Nuclear becomes "more" popular and pushes out wind and solar as a zero-carbon alternative. Breakthrough in fusion pushes the world to use fusion energy as an alternative." (T344)
[883] "By 2030, it was felt that the current period of inflation and incoming recession provided a strong case against renewable energy investments wildly exceeding expectations: despite the obvious incentive of gas shortage for governments, they felt it possible that green energy projects are already proceeding at the highest possible speed…Raw materials inflation and reluctance to install wind production in environmentally sensitive areas (ie, offshore) could plausibly place a gap on alt-energy growth." (T339)
"Disruptions caused by supply chain issues and inflation could lead to higher prices and slower rollout, and it's possible these issues persist for a longer term." (T341)

- **Conflict with China could jeopardize the photovoltaic supply chain**[884]
    - "China produces far too much of the technology that the world relies on for PV systems and that could be put in jeopardy in a face off over Taiwan." (T339)
- **High price of polysilicon**[885]
- **Fossil fuels will continue to be necessary to meet world energy demand**[886]

Arguments given for higher-end forecasts

- **Political urgency of move away from fossil fuels**
    - Nations have strong incentives to pursue renewable energy, both because of public concern over climate change[887] and because the war in Ukraine

---

"Geopolitical instability, global political realignment, supply chain problems, inflation, and debt are all major sources of uncertainty." (T345)

[884] "Another source of uncertainty with the potential to influence this question is a potential global conflict involving China versus the West." (T339)

"Sources of uncertainty are…the possibility of sanctions against China, which the West relies upon for solar panels." (T340)

"Disruptions caused by supply chain issues and inflation could lead to higher prices and slower rollout, and it's possible these issues persist for a longer term. Difficulty in acquiring the necessary materials coupled with increased tensions between the US and China could mean more expensive renewable equipment, inefficient trading, and a lack of availability of finished renewables equipment." (T341)

"Conflict between China and the West could disrupt these industries and their export markets. Producing solar panels outside would presumably be significantly more costly, at least for several years." (T343)

[885] "[Growth in solar] [c]ould also decelerate a bit with the high price of polysilicon in the last couple years after the pandemic." (T336)

"Materials shortage (currently polysilicon) and increasing labor costs may prevent further cost declines." (T343)

[886] "[I]t was noted that to meet the overall rise in world demand, almost half of new electrical generation will need to come from fossil fuels…The pie is going to grow. And not all energy needs will be met by just these two sources." (T336)

"Because of the current energy crisis one forecaster is more optimistic about 2030 than 2024 because the likelihood of fossil fuel usage is expected to rise short term, meaning that as a percentage solar and wind may not increase as much." (T338)

"By 2030 it will increase significantly because of the increased funding due to the threat presented by global warming. However the share of wind and solar power is most likely not going to increase beyond 50 % because of the unreliability of [these] two power sources. Most likely the share will be even lower as such a massive increase would require massive increase in funding." (T342)

[887] "For 2030, these forecasters assume that public sentiment heavily favors green energy driving governments to invest more in solar and wind and fast track approval, building, and permitting. The driving forces behind public sentiment are fear of climate change, the war in the Ukraine and [gas] shortages, and increasing energy prices of fossil fuels." (T337)

"The most optimistic forecasts cite falling costs and the current political urgency to move away from fossil fuels coming from Net Zero proponents and advocates of sanctions against oil and has increased fossil fuel prices and highlighted the value of energy independence.[888]

- "There is enormous incentive for these technologies." (T339)
- "The war in Ukraine will likely cause renewed interest - especially in the near term - in alternative energy sources than natural gas and oil." (T343)

- **Trend of decreasing costs for solar and wind power[889]**
  - "The overall trend for decrease in cost for solar is continuing and there is an increase in solar installment. Likewise the cost of wind power has decreased considerably, and is likely to continue. This will increase the competitiveness of these energy sources." (T337)[890]
- **History of underestimated growth in solar and wind power[891]**

---

gas producers... Several megatrends are strong drivers for continued growth of solar and wind: Climate change, diversification of energy and electricity generation to minimize risk of being dependent on other countries, and the need for microgrids with digitalization." (T340)
"The campaign of climate change encourages and boast the coverage and continue increase because of its climate friendliness unlike others, the confidence level Here is high due to [huge] prospect in the improvement level of development, penetration, performance and growth of this type of electricity." (T342)
"Decarbonization plans of all major economies and independent institutions such as the IEA envisage and predict dramatically increased deployment of these sources. If these countries are to fulfil their decarbonization pledges, the result will almost inevitably be higher shares of solar and wind energy than according to our median prediction." (T343)

[888] "The Ukraine war will probably drive China to speed up renewables so it cannot be held hostage by others re fossil fuels." (T336)
"There is enormous incentive for these technologies (so-called 'green vortex')…High oil prices, and low economic growth scenarios are both highly likely, at least for some time. This, combined with an incentivised rollout of renewables across developed world with energy crisis/Ukraine, makes the immediate future murky and caused some disagreement in the team. " (T339)
"The most optimistic forecasts cite falling costs and the current political urgency to move away from fossil fuels coming from Net Zero proponents and advocates of sanctions against oil and gas producers... Several megatrends are strong drivers for continued growth of solar and wind: Climate change, diversification of energy and electricity generation to minimize risk of being dependent on other countries, and the need for microgrids with digitalization." (T340)
"Russia's invasion of Ukraine is likely to accelerate moves towards greater energy independence through nuclear and renewables in Europe, but in other parts of the world may result in a resurgence of fossil fuels." (T345)

[889] "The most optimistic forecasts cite falling costs…" (T340)
"New discoveries in solar and wind technologies continue to drive down costs of solar and wind technologies. Battery storage is beginning to improve and be deployed at grid-level scales. Many experts expect long-time trends in cost reduction to continue." (T343)

[890] Citations omitted from this quotation.

[891] "[H]istoric predictions surrounding exponential growth of renewable technologies and decline of related prices have frequently underestimated true trends." (T339)
"Previous estimates have repeatedly underestimated the capacity for rapid growth in wind power and solar, indicating that higher end estimates should be taken more seriously." (T345)

- ○ "IEA projections have been criticized in the past for very low solar projections, so reliance on these projections may result in conservative forecasts." (T341)

Other arguments given

Arguments in favor of lower forecasts:

- Forecasters may be biased towards high-end projections because they favor adoption of renewable energy.[892]
- Increasing cost of labor.[893]
- For solar, the near-term lack of storage and the Duck Curve problem.[894]

Cross-references with other questions

Q26: Cost of Utility-Scale Solar Energy
Q29: Nuclear Fusion Energy
Q30: Cost of Hydrogen

Question 29: Annual Direct Air CO2 Capture

What will be the annual amount of CO2 captured and stored by direct air capture (in Mt CO2/year)…
…in 2024?
…in 2030?

Question and resolution details, prior forecasts, and other relevant sources

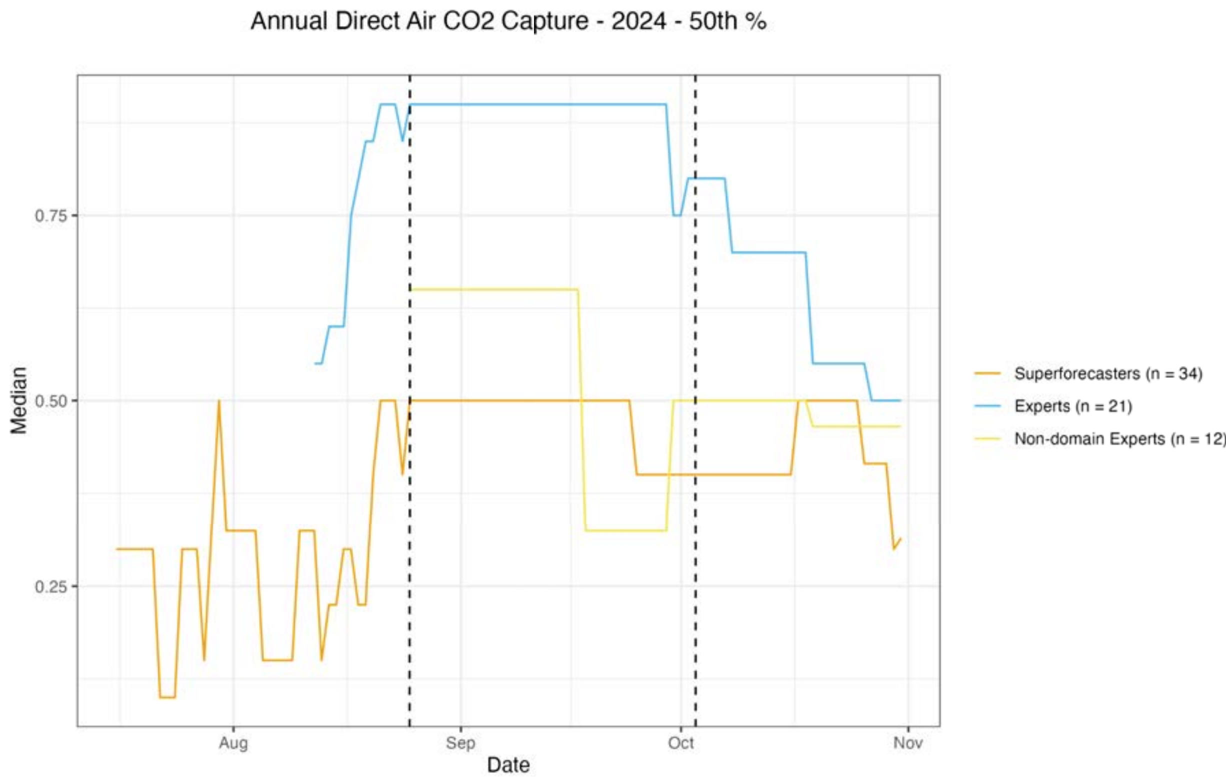
---

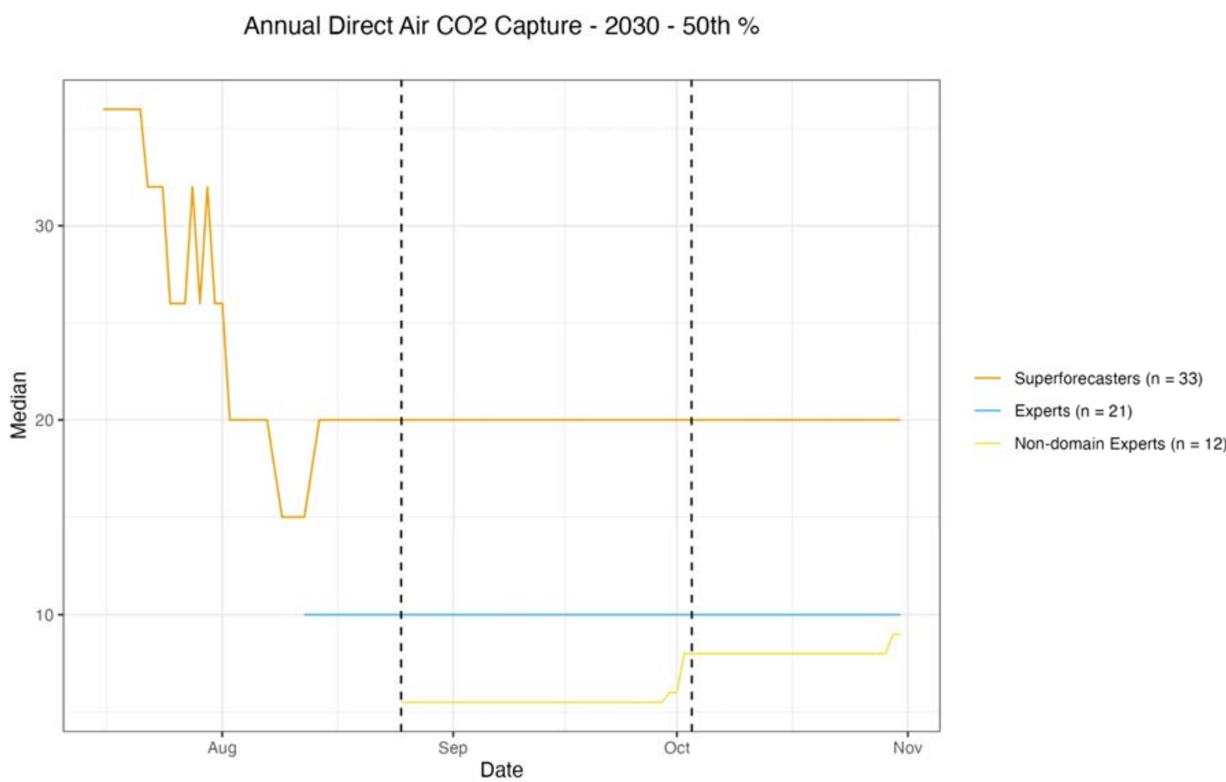
[892] "We discussed a possible bias towards continued growth of solar and wind as we tend to favor adoption of renewable energy to mitigate climate change." (T337)
"Currently, there is an overconfidence in the usage of renewable energy for energy consumption." (T344)
[893] "Materials shortage (currently polysilicon) and increasing labor costs may prevent further cost declines." (T343)
[894] "Forecasters who were more bearish on these technologies cited the lack of storage solutions in the near term and solar power's Duck Curve problem in which the highest energy generation occurs during the time of day with the least demand, which will create a point of diminishing returns for power generation." (T340)
"We can expect early adopters to continue to pay premium prices for novel technologies like battery storage; with price drops, we can expect the next wave of implementation to begin to reach more mainstream energy markets." (T343)

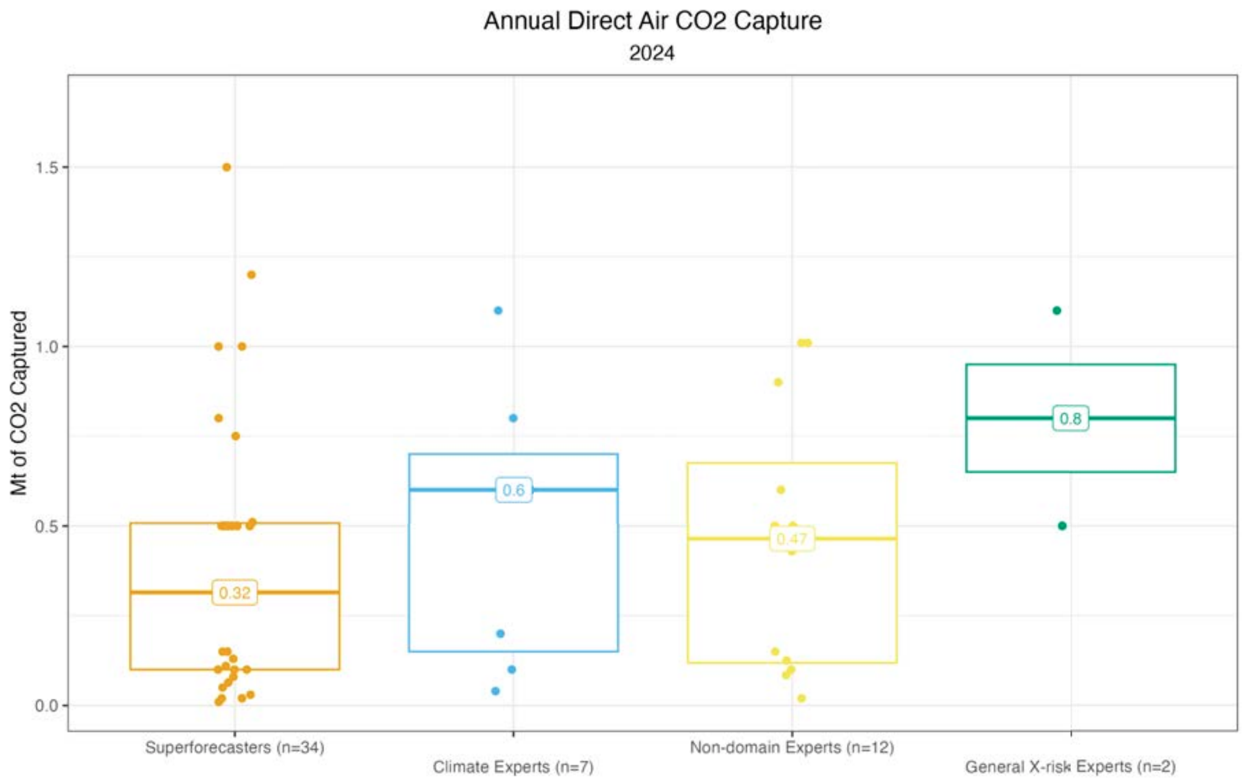
Results[895]

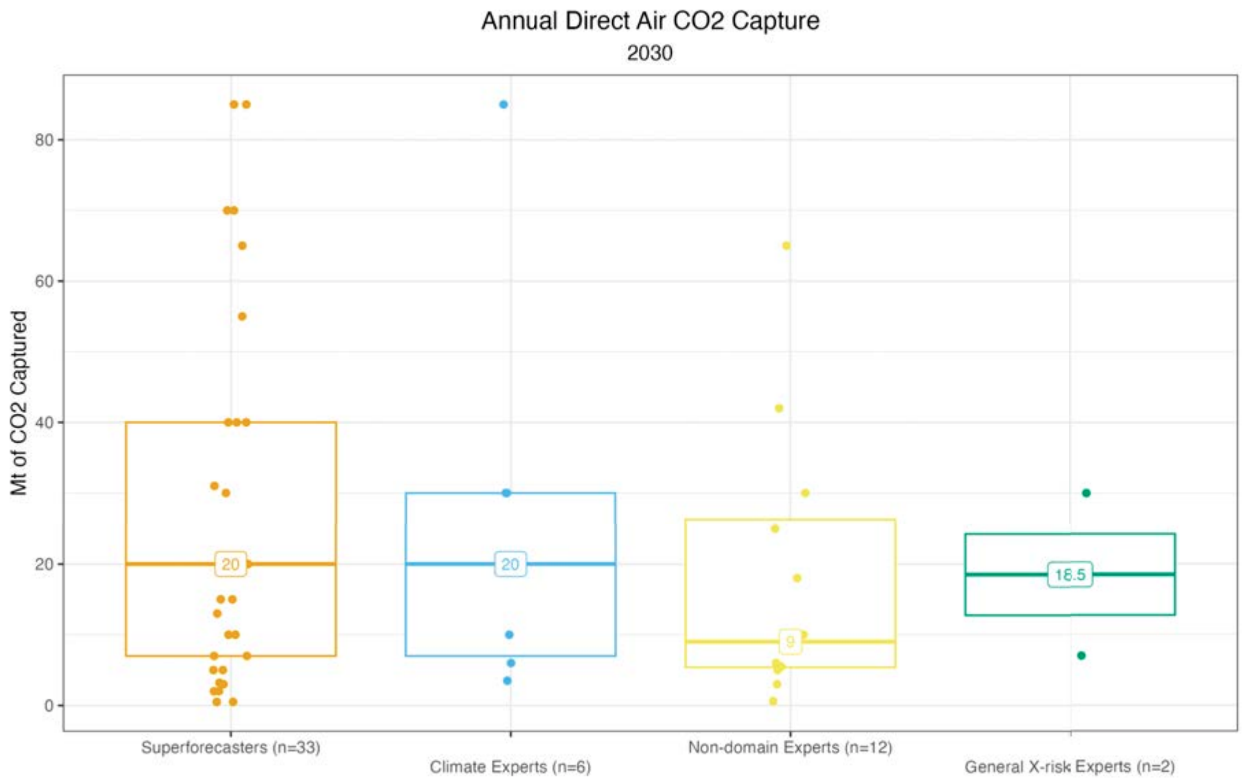
Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 33)	2024	0.5	0.32	10326.66	-99.94%
	2030	70	20	11076.34	-99.68%
Domain Experts (N = 6)	2024	1	0.6	NA[896]	NA
	2030	10	20	NA	NA
General X-Risk Experts (N = 2)	2024	3	0.8	NA	NA
	2030	30	18.5	NA	NA
Non-Domain Experts (N = 14)	2024	0.15	0.47	4024.78	-99.99%
	2030	60	9	4472119.23	-100%

[895] Numbers of forecasters are given as of Stage 4 of the XPT.
[896] Only one forecaster in this group answered this question in Stage 1.


Annual Direct Air CO2 Capture - 2024 - 50th %
Median
0.75
0.50
0.25
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 34)
Experts (n = 21)
Non-domain Experts (n = 12)


Annual Direct Air CO2 Capture - 2030 - 50th %
Median
30
20
10
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 33)
Experts (n = 21)
Non-domain Experts (n = 12)


Annual Direct Air CO2 Capture
2024
Mt of CO2 Captured
1.5
1.0
0.5
0.0
0.32
0.6
0.47
0.8
Superforecasters (n=34)
Climate Experts (n=7)
Non-domain Experts (n=12)
General X-risk Experts (n=2)


Annual Direct Air CO2 Capture
2030
Mt of CO2 Captured
80
60
40
20
0
20
20
9
18.5
Superforecasters (n=33)
Climate Experts (n=6)
Non-domain Experts (n=12)
General X-risk Experts (n=2)

Sources of agreement, disagreement and uncertainty

- **Common starting point of 0.01 Mt CO_2 stored annually**
  - The common starting point used by forecasters was drawn from the International Energy Agency (IEA)'s direct air capture (DAC) report, which states that approximately 0.01 megatonnes (Mt) of CO_2 is currently stored annually using DAC.[897] All teams that mentioned base rates cited this number,[898] and no forecasts were below 0.01 Mt/year even at the 5th percentile.
    - "The base rate for our team's forecasts is built around the existing CO2 removal capacity in 2022: according to the IEA, 0.01 Mt/year." (T338)
  - The IEA report also states that a 1 Mt CO_2/year plant (called DAC 1) is in advanced development. Some forecasters took those projections at face value, while others noted that per IEA data, only the first train of DAC 1 is expected to be operational by 2024 with a capture rate of 0.5 Mt CO_2/year.[899]
  - Little other objective analysis of DAC is available given the novelty of this technology, although one team referred to the Global Carbon Capture and Storage Institute's 2021 CCS Report.[900]

---

[897] Note: this figure was slightly altered between the May 2022 version of the report, which read "There are currently 19 direct air capture (DAC) plants operating worldwide, capturing more than 0.01 Mt CO2/year…" and the April 2023 version, which read "There are currently 18 direct air capture plants operating worldwide, capturing almost 0.01 Mt CO_2/year…" The change may be because the Climeworks Orca plant was forced to shut down in 2022 due to frozen machinery.

[898] "There are currently 19 plants online with capacity to capture 0.01 Mt CO2/year. There is a 1Mt CO2/year plant in advanced development in the US. Given these numbers it is difficult to project annual growth over a short-term period." (T336)

"Roughly 0.01 Mt CO2/year is captured today by Direct Air Capture (DAC) plants. And 1 Mt of capacity is being planned to go online in 2024 (which is not correct, see discussion on expected disagreements below)." (T339)

"A Base rates A1. Current capacity is 0.01 MtCO2/year with 19 plants worldwide online." (T337)

"*Base rate(s)* IEA's numbers and projections" (T340)

"This forecast uses the same base rate of ~0.01 Mt as that of the current annual amount of CO2 captured and stored with DAC technology, but believes there has been a significant change in the carbon market and new technologies are arising that could lead to growth." (T341)

"There are currently 19 direct air capture (DAC) plants operating worldwide, capturing more than 0.01?Mt?CO2/year…" (T345)

[899] "5) Expected disagreement between our team and other groups

Biggest Point: I expect other teams to have recognized that the 2024 plans are only for 0.5 Mt CO2/year with no existing plans for when full capacity is expected to be reached (IEA source). So I expect all forecast for 2024 to be lower by 0.5 Mt CO2/year in comparison to our team." (T339)

"Currently only one large scale DAC carbon capture facility is under construction which should capture 1 Mt/year of carbon. However, by 2024 only the first part with capacity of 0.5 Mt/year is planed to be finished. While it could be finished faster than scheduled, delays are more likely... Careful reading here [1] reveals, that the 1 Mt/year plant is split into two trains with 0,5 Mt/year capacity each, but only train 1 is planed to be operational by 2024. So just for the total capacity to be 1 Mt/year in 2024, the 1 Mt/year plant would have to be finished ahead of schedule." (T342)

[900] "According to the Global CCS Institute's 2021 Status Report, plants in operation or under construction have the current capacity to capture 40 million metric tons of CO2 per year and DAC is scaled up to capture more than 85 Mt CO2/year by 2030." (T342)

        - “Since DAC is a new technology, few prior forecasts or historical data exist except those referenced in the IEA report.” (T344)
- **Net Zero Emissions targets will not be met**
    - Forecasters agreed that it was unlikely that DAC would meet the Net Zero Emissions by 2050 scenario projection of “more than 85 Mt CO2/year by 2030.”[901] 85 Mt CO_2/year represents roughly the 95th percentile of forecasters' predictions.[902]
        - “NZE's scenarios appear wildly optimistic, without a clear path to achieving them.” (T343)
- **Uncertainty as to political will**
    - Because of the important role of government subsidies and incentives in DAC development, questions about the consistency of government support for DAC were an important source of uncertainty.[903] Changes in political priorities, taxation decisions, public opinion, and legal risks could all impede growth in DAC.
        - “The team identified government subsidies as the primary factor that will influence growth. The cost of these projects make it infeasible to bring them to scale without government support.” (T336)

Arguments given for lower-end forecasts

- **Long lead time for DAC facilities**

---

[901] This quotation from the IEA DAC Report was changed sometime between May 2022 and April 2023. It now reads “In the Net Zero Emissions by 2050 Scenario, direct air capture is scaled up to capture almost **60** Mt CO_2/year by 2030.”

[902] “There was broad agreement that the 2030 Net Zero Emissions targets will not be met.” (T336)
“With the US plant opening in 2024 at the earliest, I'd place a 50/50 chance on there being 1Mt of CO2 captured in 2024, Getting to 90Mt in 2030 to meet the net zero target then seems impossible, but there may be a few extra plants around by that time so I'll say there's a 50/50 chance of 20Mt captured in 2030.” (T342)
“DAC is not expected to be profitable, therefore it's unclear why we should expect to see large increases in this approach. NZE's scenarios appear wildly optimistic, without a clear path to achieving them.” (T343)

[903] “I don't know, if they factored in potential resistance by the population, where those facilities shall be built. And this is more or less low-tech, where faster chips and better AI can't help to achieve exponential growth. Difficult.” (T336)
“There is a general lack of government incentives (like carbon tax) or CO2 pricing mechanisms to make scalable DAC economically feasible…Governments will be incentivized to pass regulations on CO_2 mitigation as the impact of climate change becomes more evident to the public.” (T337)
“The risks to CO2 capture technology, which support the forecasts at the lower end of our projections, include political shifts (like the Supreme Court decision limiting the EPA's regulatory power) that could reduce government action on CO2 capture.” (T338)
“Additional uncertainty was in the political will. If sufficient 'green-voters' could direct massive funding into this place (despite [its] inefficiencies). It is also not clear how strictly countries will work towards net zero emission 2050 targets.” (T339)

  - DAC facilities take significant time to be designed, approved, and built. This puts a brake on growth but also makes DAC capture rates more predictable in the short term. [904]
    - "[S]ince DAC activity requires a large industrial facility to be designed and built, it should be hard to produce a surprising jump from one year to the next…There is also a significant lead time driven by project approval, site selection, facility design, permitting, construction, testing, and beginning operations." (T341)
- **No experience building large-scale DAC facilities**[905]
  - "This technology is…just at the beginning of gaining serious economies of scale, the technology has not been demonstrated in a large-scale setting, and it will remain very expensive in the short- to medium-term." (T336)
- **Projections by DAC startups are likely to be over-optimistic**[906]
- **Doubtful cost-effectiveness**
  - DAC faces economic feasibility problems because of the lack of commercial uses for captured CO_2.[907] Forecasters noted some existing commercial uses for CO_2 but did not identify new ones.
- **Alternatives to DAC would be cheaper**

---

[904] "Construction lead-times for systems of impactful scale will almost certainly be considerable, especially given the relative immaturity of these systems. Carbon Engineering's flagship plant is expected to begin operational commissioning no less than 5.5 years after it was announced (May 2019 - 'late' 2024). Consequently, we have a greater degree of visibility over this than might first appear, with additional capacity needing to begin planning prior to 2025 in order to significant[ly] affect total capture in 2030." (T337)
"There are few incentives to completing projects on schedule [and] [b]ig engineering projects usually take longer to complete than expected." (T340)
"We believe that it is plausible to forecast that no significant DAC facilities will be added through 2024, and therefore that the amount of DAC in that year will be more or less unchanged from the base rate of 0.01 Mt." (T341)
[905] "The current technology readiness level of 6 (out of 9) means DAC is in large scale prototype phase. It is expected that costs will decline significantly with commercial deployment and economies of scale." (T339)
"Based on the base rate, there is effectively no experience building large scale (~1 Mt) DAC facilities." (T341)
"On the lower end, most forecasters list the fact that DAC is a marginal unproven technology with a limited future." (T344)
[906] "Most innovation is coming from startups with Climeworks and Carbon Engineering the most prominent. Climeworks' current capacity is 4 kt per year [3]. The plan [is to] increase to 36 kt per year in a few years. The goal is to have several million of MT by 2030. Carbon Engineering has plans to capture 0.5 Mt [1] by 2024, to increase to full capacity (1 Mt) in years to follow. Plan is to have 70 operating facilities by 2035. The consensus in the team [is] that the projections by the startups are likely to be inflated since startups tend to be optimistic in their projections." (T337)
[907] "Only way it seems to be close to cost effective is if the CO2 has a use other than putting it into storage. Oil firms may use it for fracking." (T336)
"Year 2024 is to witness very small and redundant development as a result of many challenges, like lack of cost effectiveness, technology issues and others which are critical and germane to the process of CO2 captured and stored by direct air capture…" (T344)
"[Higher-end arguments include] captured carbon may find new commercial uses, without getting cycled directly back into the atmosphere." (T343)

    - Reducing CO_2 emissions can be done by many pathways, and DAC is one of the more expensive options, compared to, for example, reducing emissions and reforestation.[908]
        - "One of the biggest problems with direct carbon capture is that it must compete with a number of other, better technologies that are cheaper, with potential side-benefits." (T343)
- **DAC's high energy needs**
    - DAC uses a lot of energy and sourcing this energy from fossil fuels would cancel out the benefit of carbon capture. This means that the availability and price of renewable energy is crucial to DAC's economic feasibility.[909]
        - "Scaling up today's DAC systems would use non-trivial amounts of energy. Currently DAC requires 5-10 GJ/tCO2. Global CO2 emissions are around 33 Gt/Yr currently. DAC would have to be in the 1 Gt/Yr range to have an impact. This would require 5-10 EJ of energy which is a significant portion of the total renewable energy supply of about 85 EJ (hydro, biofuels, and wind/solar)." (T337)[910]
- **Costs of DAC will decline less quickly than solar costs did**
    - With some uncertainty, forecasters believed that costs of DAC would decline less quickly than costs of solar due to the DAC's high energy and storage costs.[911]

---

[908] "Tree planting continues to be 5x cheaper per tonne captured than DAC, consequently there is a significant probability that DAC does not keep up with the hype and will be seen as ineffective to mitigate climate change... reducing emissions is likely to continue to be significantly more cost effective than DAC over the coming decade." (T337)
"It is not clear that DAC will be THE technology for reducing emissions and therefore it is not clear that exactly DAC will be scaled to the projections existing today. Examples have been made for alternatives that are more economical. For example reforestation costs about $50/ton where as DAC is expected to drop down to $150/ton to $200/ton. Especially preventing emissions in the first place (through higher efficiencies) or capturing CO2 where it is being released (say at power plants) look more economical. (Though trade-offs between energy, land, and water use have been acknowledged.)" (T339)
"It was additionally, argued there are plenty of examples in which the prevention of energy use is way more efficient (for example replacing light bulbs with LEDs). One calculation was shown that 1 ton of CO2 could be averted with about $20 of LED light bulbs in comparison to the (optimistic) $150 - $200 via DAC." (T342)
[909] "We reasoned that the impact of Direct Air Capture (DAC) is going to be small in 2024 and 2030 as compared to the total CO2 emissions given (1) the lack of operational plants + plans to scale, (2) the high cost of carbon capture (energy use), and (3) the lack of regulation to make DAC economically feasible apart from pilot and demonstration plants by 2030." (T337)
"Someone in the team was optimistic that this is feasible with cheap energy through fusion. Where it is highly uncertain if this will be significant for DAC by 2030." (T339)
"High energy prices will prevent widespread adoption of Direct Air Capture." (T340)
[910] Citations omitted from this quotation.
[911] "Much wider range of outcomes for 2030, noting projections of cost decreases for now technology can be very hard (see solar panels and how much more quickly costs have come down than expected even a few years ago)." (T336)
"It is not clear that DAC will see declining costs curves like solar did because the majority of costs are in the energy (80% heat and 20% electricity)." (T339)
"[I] generally don't think we'll see massive cost reductions over time — I don't think the experience curves will be nearly as good as those for solar panels, with storage costs being a big bottleneck." (T342)

Arguments given for higher-end forecasts

- Arguments for more optimistic DAC forecasts centered around the theme of strong interest in this technology, existing DAC projects, and financial commitments from various actors.[912]
- **Supportive government policy**
  - The U.S. Inflation Reduction Act significantly increased the available tax credit for carbon removal, although there is disagreement about its effect.[913]
- **Corporate commitments to carbon removal**, such as the Frontier advance market commitment, the Musk Foundation's XPRIZE competition, and interest from oil and gas companies, create market demand for DAC that would not otherwise exist.[914]

---

[912] "There is strong interest (plans and startups) in the private and public sectors." (T340)
"[T]his optimistic forecast assumes that while DAC technology is fairly well proven, it was too costly to justify large-scale investments until recently, with two reasons for the change. First, a number of major corporations have made institutional commitments to carbon removal programs. Second, government policy (such as the Inflation Reduction Act) significantly increases the tax credit available for carbon removal services." (T341)
"The world's largest DAC facility with a capacity of 4000 tonnes of CO2 per year went online in Iceland on 8 September 2021. [3] Another facility has also started construction in Iceland, run by Swiss startup Climeworks AG. When construction finishes in 2024, their facility, named "Mammoth," will be able to remove 36,000 tons of CO2 from the air per year – which is .0001% of the 36 *billion* tons of CO2 emitted per year by humanity. [4] [5]
Occidental and its subsidiary 1PointFive have announced they plan to begin detailed engineering and early site construction for their first large-scale DAC plant in Ector County, Texas. The first stage of construction is scheduled to begin in the third quarter of 2022 and start-up is expected in late 2024. Upon completion, the first DAC plant will be the world's largest of its kind and is expected to capture up to 500,000 metric tons of carbon dioxide per year with the capability to scale up to 1 million metric tons per year. 1PointFive has announced a scenario to deploy 70 DAC facilities worldwide by 2035 under current compliance and market scenarios…The team's median forecast for 2024 is 5 Mt CO2/year and 40-50 for 2030. This is supported by the large-scale investments already announced." (T345)
[913] "The team noted that recent passage of the Inflation Reduction Act increases direct payments for direct air carbon capture from $50 per ton to $180 per ton, which will potentially increase development in the US." (T336)
"The team projects that government investment, including from the Inflation Reduction Act, will continue to support the development of these plants between now and 2030, even though the technology is not yet economically viable on its own." (T338)
"[G]overnment policy (including the recently passed Inflation Reduction Act) is supportive of DAC technology." (T341)
"The main disagreement comes from translating the effects of the Inaction Reduction Act on the DAC "sector". There is a 10x difference in expectations." (T344)
"One of the main arguments for an explosion of carbon capture by 2030 is seen in the Inflation Reduction Act…"The Inflation Reduction Act of 2022 (IRA) provides critical updates to the 45Q tax credit, which incentivizes the use of carbon capture and storage – a climate solution that the Intergovernmental Panel on Climate Change (IPCC) and International Energy Agency (IEA) have found is likely to play a vital role in efforts to address climate change." link" (T344)
[914] "Commercial interest in CO2 mitigation technologies has increased significantly in recent years, particularly from Oil & Gas companies seeking to shore up their environmental credentials. They will invest in demonstration plants to justify the continued use of coal, gas, and oil." (T337)
"Sources of uncertainty..Impact of Elon Musk's XPRIZE competition" (T340)
"For example, one team member reports that a big slug of new research money is going into the direct capture of CO2 from air: Frontier facilitates first carbon removal purchases... Also, "Ideas for reducing

Other arguments given

Arguments in favor of lower forecasts:

- "Direct Air Capture is a water-intensive technology, which could be problematic given the changing climate." (T340)
- "Extreme inflation may drive the planned investments to ineffective." (T344)
- "Management problems from the US plant means it won't come online in 2024." (T340)
- DAC is inherently an inefficient technology.[915]

Arguments in favor of higher forecasts:

- CarbonCapture's Project Bison plans to remove 5 million tons of CO_2 annually by 2030.[916]

Cross-references with other questions

Q25: Average Global Surface Temperature
Q26: Cost of Utility-Scale Solar Energy
Q27: Nuclear Fusion Energy

Question 30: Cost of Hydrogen

How much will it cost to produce hydrogen from renewable electricity (in $ per kg of hydrogen)…
…in 2024?
…in 2030?

Question and resolution details, prior forecasts, and other relevant sources

---

greenhouse gas in the atmosphere are getting a funding boost from famed entrepreneur Elon Musk [now investing in]...8 Rivers Capital." (T341)
"I think that if every one is being mandated to achieve net zero by 2050, the infrastructure construction is going to have to ramp up significantly in that time as companies scramble to meet targets, There will be a market for it because offsetting is looking like it might grind to a halt for a decade or so as countries start to place moratoriums on carbon recapture projects in the developing world (mostly from not wanting to sell all their potential credits off to developed nations and leave nothing for themselves, and because selling them in 10 years will make a lot more money than selling them now will). That's why I aimed optimistically at a maximum of a third of the 2050 target by 2030." (T342)

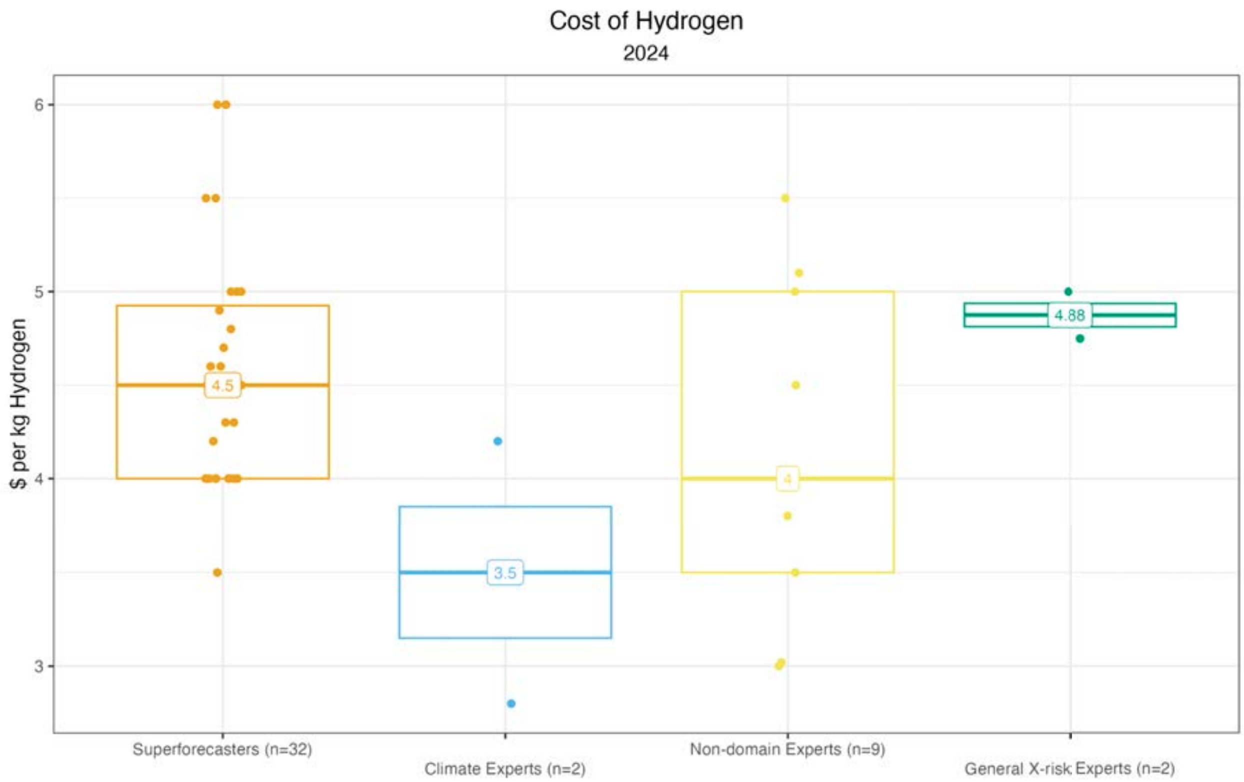
[915] "Someone argued that the technology itself is very inefficient and not economical. with water vapour being a side product which itself is a strong green house gas. Arguing that this tech needs to be [...] carefully operated to have the desired effect (of reducing net greenhouse gases in the atmosphere)." (T339)

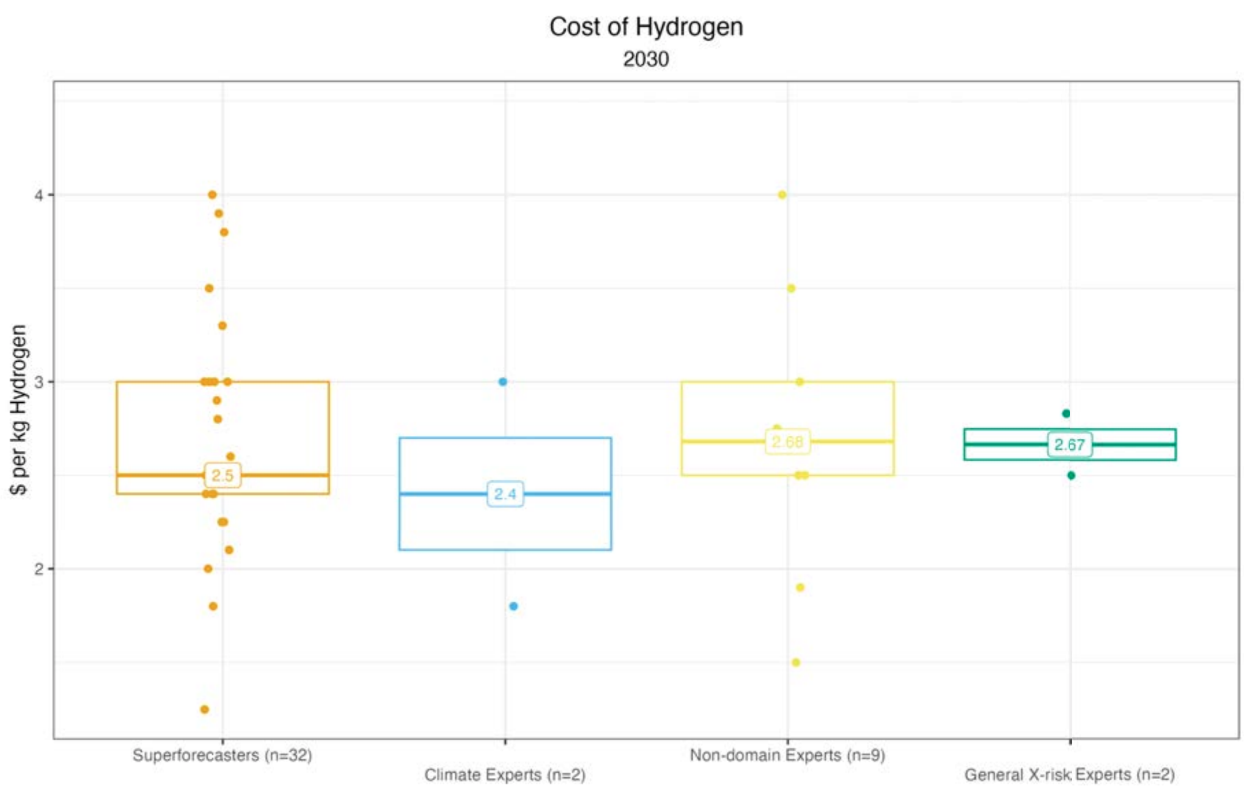
[916] "A company called CarbonCapture just announced ProjectBison, a DAC project in Wyoming that aims to permanently remove 5 million tons of CO2 annually by 2030." (T341)

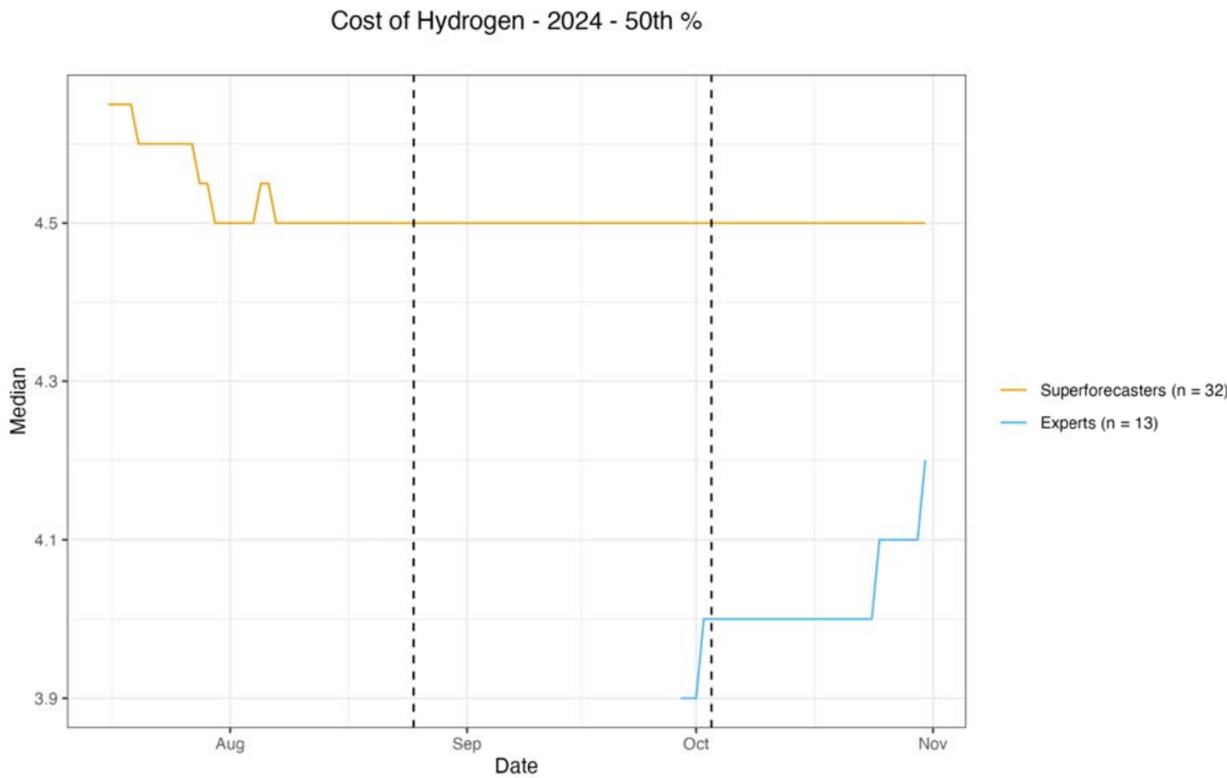
Results[917]

Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 32)	2024	4.7	4.5	1.02	-27.81%
	2030	2.5	2.5	1.51	-22.63%
Domain Experts (N = 2)	2024	2.5	3.5	NA[918]	NA
	2030	1.5	2.4	NA	NA
General X-Risk Experts (N = 2)	2024	NA	4.88	NA	NA
	2030	NA	2.67	NA	NA
Non-Domain Experts (N = 9)	2024	3.75	4	1.08	-15.04%
	2030	2.75	2.68	1.04	-27.28%

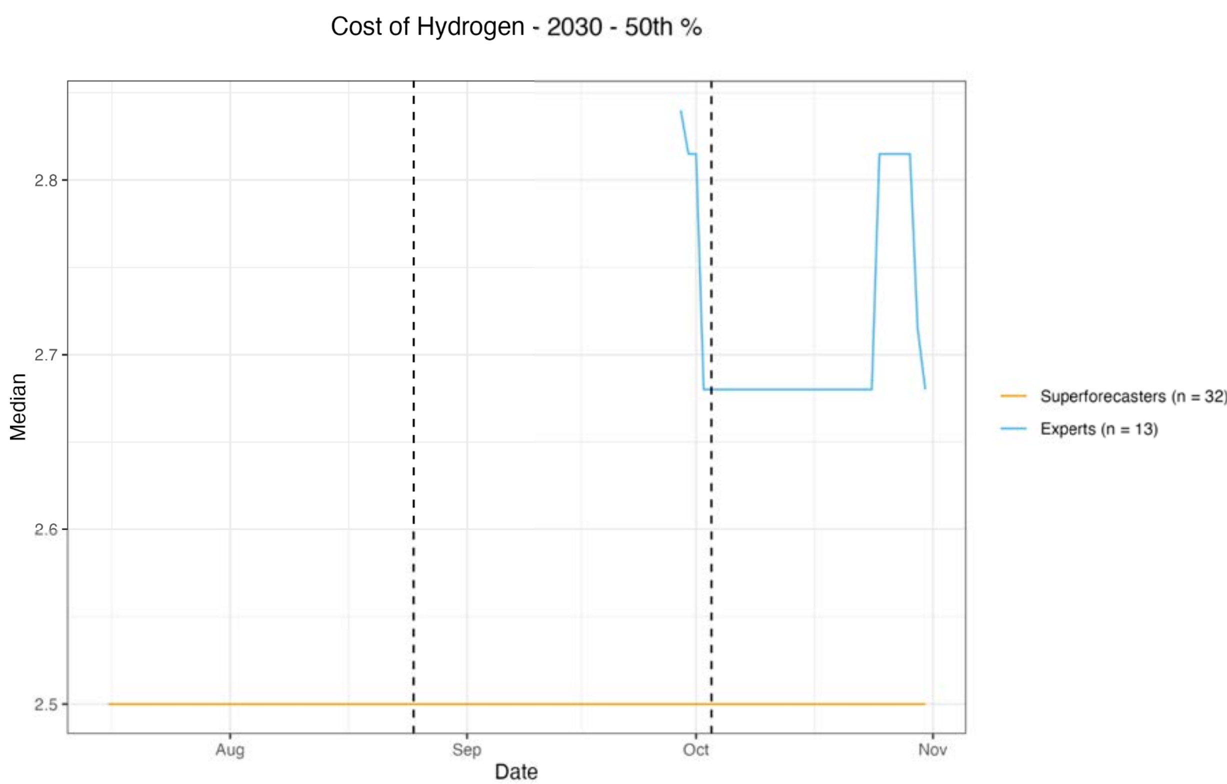
917 Numbers of forecasters are given as of Stage 4 of the XPT.
918 Only one forecaster in this group answered this question in Stage 1.


Cost of Hydrogen
2024
$ per kg Hydrogen
6
5
4
3
4.5
3.5
4
4.88
Superforecasters (n=32)
Climate Experts (n=2)
Non-domain Experts (n=9)
General X-risk Experts (n=2)

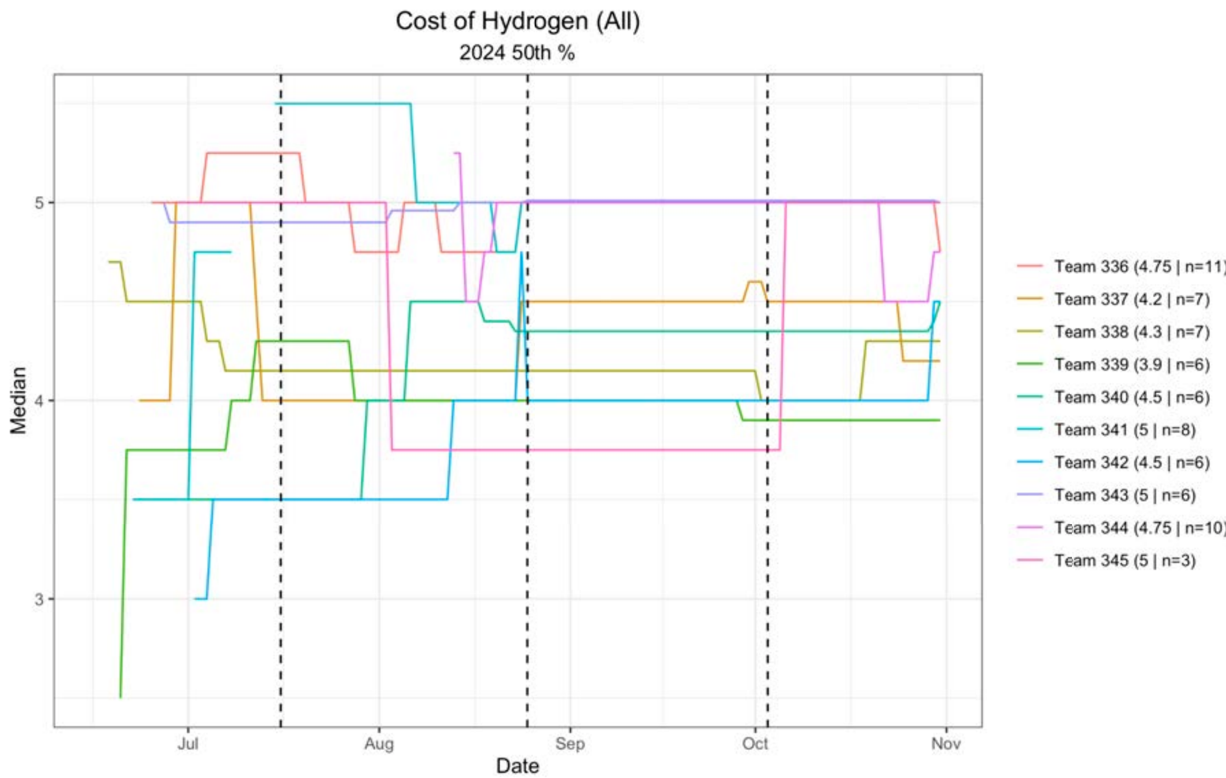

Cost of Hydrogen
2030
$ per kg Hydrogen
4
3
2
2.5
2.4
2.68
2.67
Superforecasters (n=32)
Climate Experts (n=2)
Non-domain Experts (n=9)
General X-risk Experts (n=2)

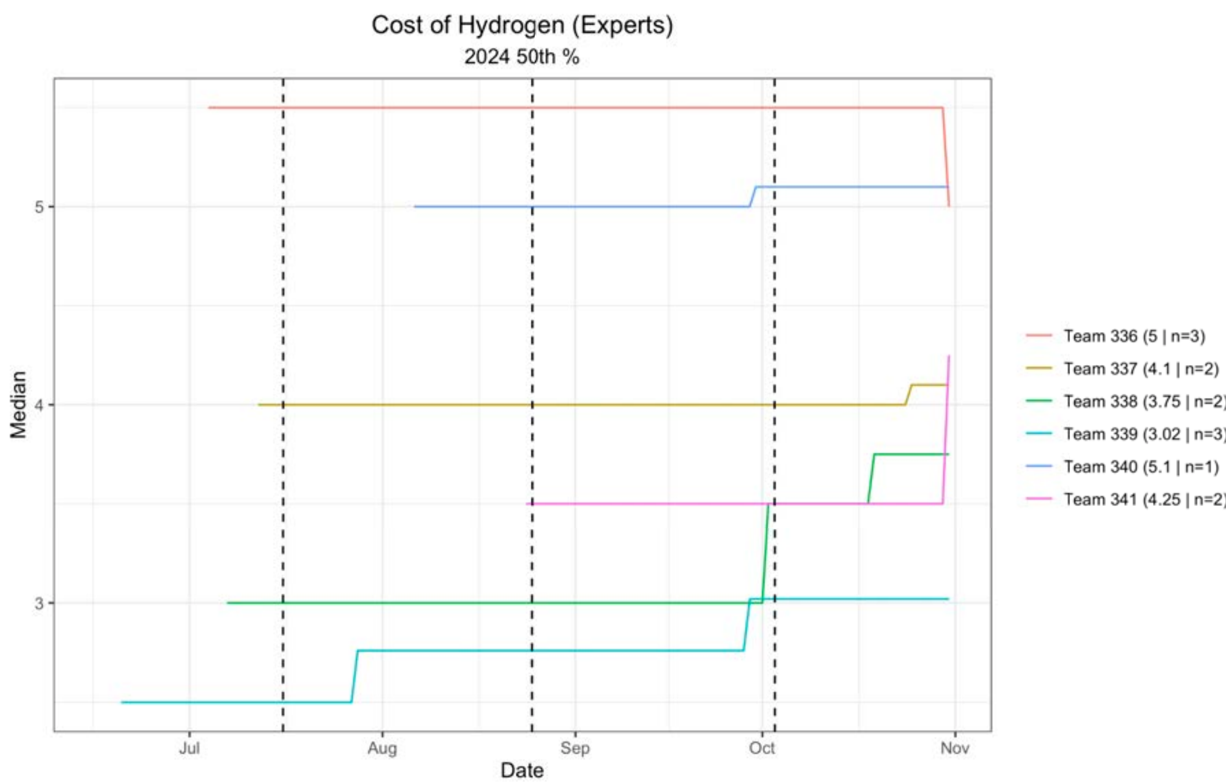


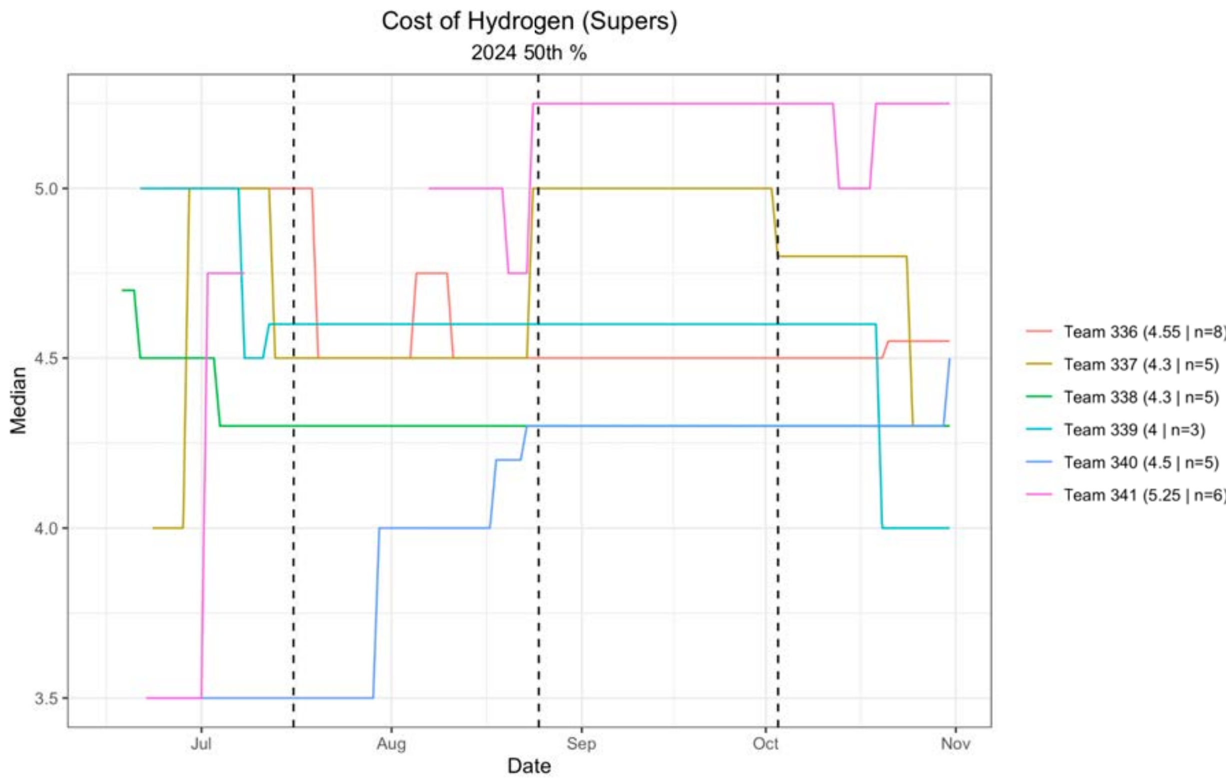
Cost of Hydrogen - 2024 - 50th %
Median
4.5
4.3
4.1
3.9
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 32)
Experts (n = 13)



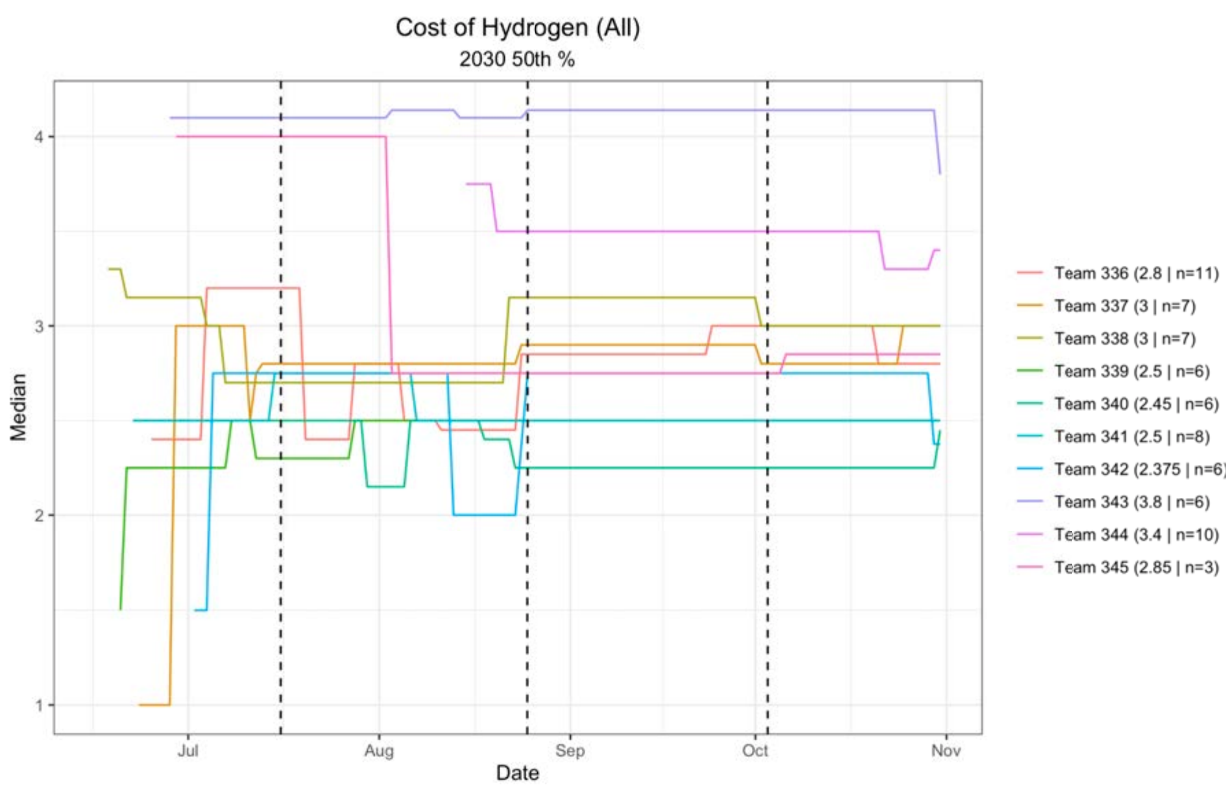
Cost of Hydrogen - 2030 - 50th %
Median
2.8
2.7
2.6
2.5
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 32)
Experts (n = 13)


Cost of Hydrogen (All)
2024 50th %
Median
Date
Jul
Aug
Sep
Oct
Nov
5
4
3
Team 336 (4.75 | n=11)
Team 337 (4.2 | n=7)
Team 338 (4.3 | n=7)
Team 339 (3.9 | n=6)
Team 340 (4.5 | n=6)
Team 341 (5 | n=8)
Team 342 (4.5 | n=6)
Team 343 (5 | n=6)
Team 344 (4.75 | n=10)
Team 345 (5 | n=3)

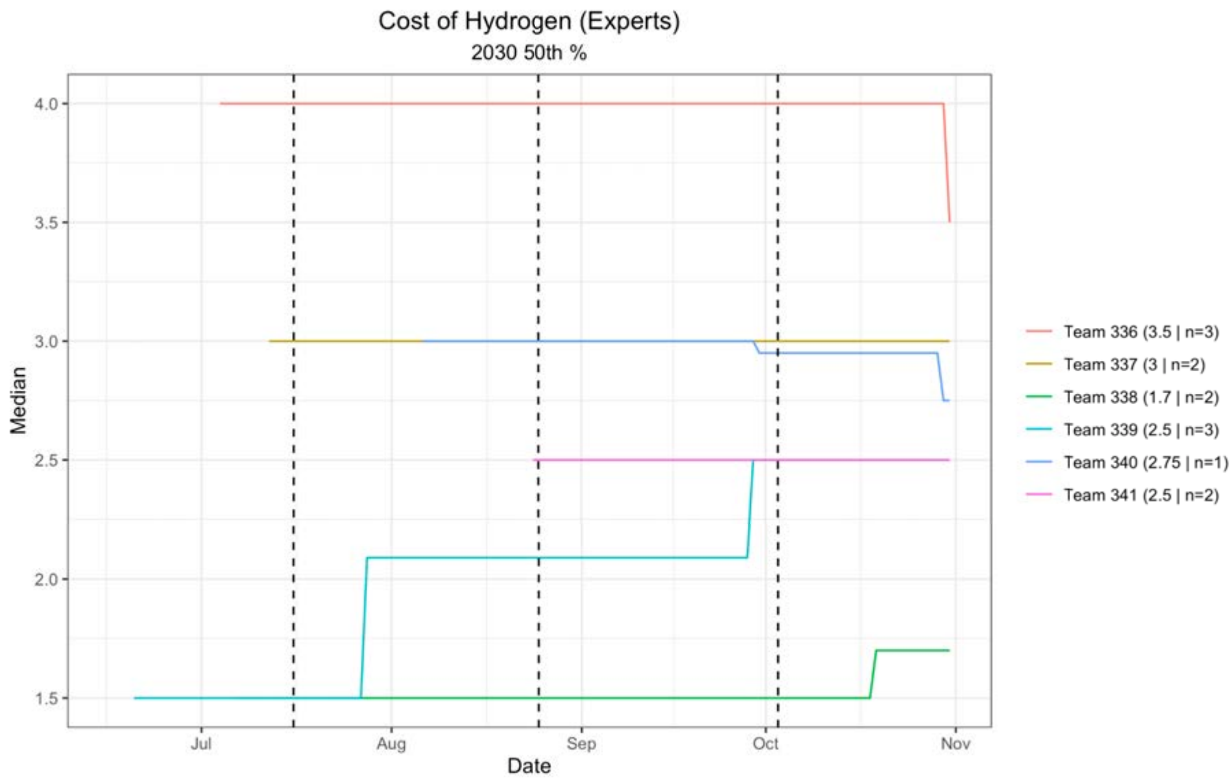

Cost of Hydrogen (Experts)
2024 50th %
Median
Date
Jul
Aug
Sep
Oct
Nov
5
4
3
Team 336 (5 | n=3)
Team 337 (4.1 | n=2)
Team 338 (3.75 | n=2)
Team 339 (3.02 | n=3)
Team 340 (5.1 | n=1)
Team 341 (4.25 | n=2)

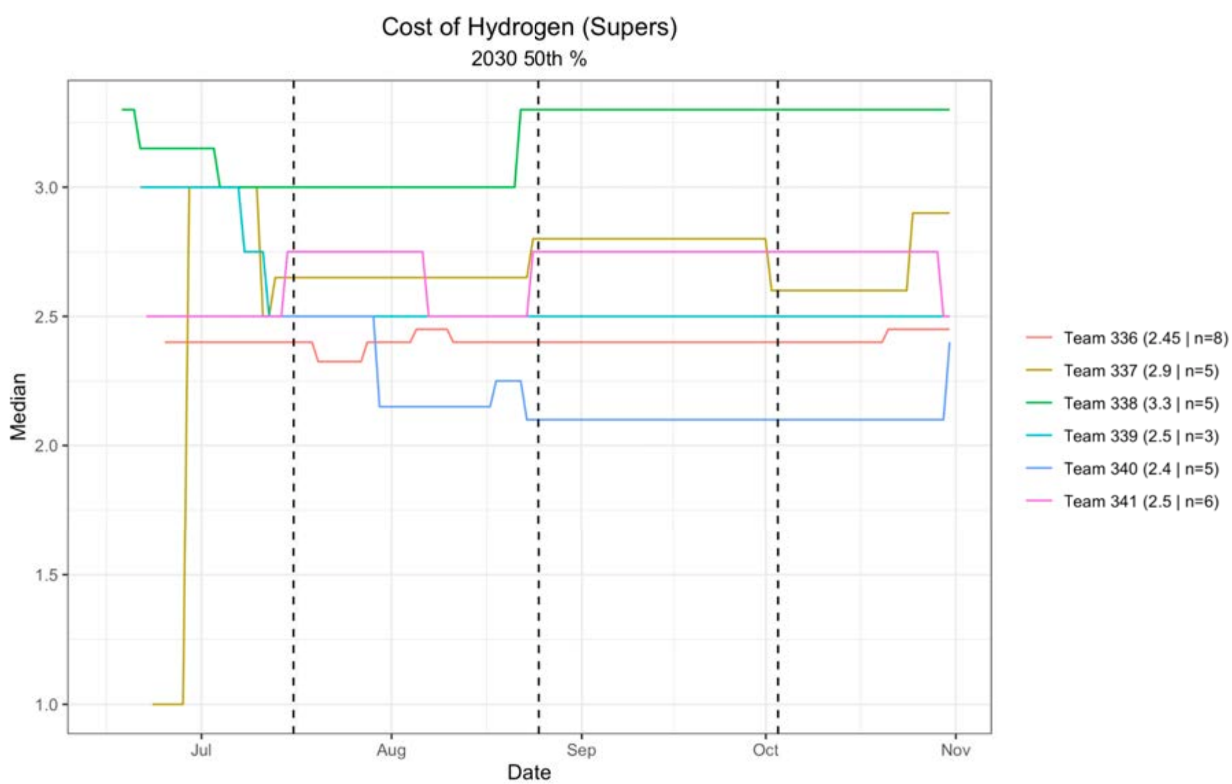


Cost of Hydrogen (Supers)
2024 50th %
Median
Date
Jul
Aug
Sep
Oct
Nov
3.5
4.0
4.5
5.0
Team 336 (4.55 | n=8)
Team 337 (4.3 | n=5)
Team 338 (4.3 | n=5)
Team 339 (4 | n=3)
Team 340 (4.5 | n=5)
Team 341 (5.25 | n=6)



Cost of Hydrogen (All)
2030 50th %
Median
Date
Jul
Aug
Sep
Oct
Nov
1
2
3
4
Team 336 (2.8 | n=11)
Team 337 (3 | n=7)
Team 338 (3 | n=7)
Team 339 (2.5 | n=6)
Team 340 (2.45 | n=6)
Team 341 (2.5 | n=8)
Team 342 (2.375 | n=6)
Team 343 (3.8 | n=6)
Team 344 (3.4 | n=10)
Team 345 (2.85 | n=3)


Cost of Hydrogen (Experts)
2030 50th %
Median
Date
4.0
3.5
3.0
2.5
2.0
1.5
Jul
Aug
Sep
Oct
Nov
Team 336 (3.5 | n=3)
Team 337 (3 | n=2)
Team 338 (1.7 | n=2)
Team 339 (2.5 | n=3)
Team 340 (2.75 | n=1)
Team 341 (2.5 | n=2)


Cost of Hydrogen (Supers)
2030 50th %
Median
Date
3.0
2.5
2.0
1.5
1.0
Jul
Aug
Sep
Oct
Nov
Team 336 (2.45 | n=8)
Team 337 (2.9 | n=5)
Team 338 (3.3 | n=5)
Team 339 (2.5 | n=3)
Team 340 (2.4 | n=5)
Team 341 (2.5 | n=6)

Sources of agreement, disagreement and uncertainty

- Basically all teams agreed that a straightforward extrapolation of base rates was the right starting point.[919] Teams did not deviate very much from the median suggested by the base rates, which have the cost of hydrogen continuing to drop.[920]
- Some sources of uncertainty included:
    - The effects of inflation.[921]
    - How much the cost of renewables that produce green hydrogen will change going forward.[922]
    - The hydrogen market and other energy markets are volatile and generally difficult to forecast.[923]

---

[919] 337: "The projected average green hydrogen cost by 2030 is USD 2.7 per kg by IEA. Our 50th percentile forecast of USD 2.9 per kg is close. The team based its forecasts on hydrogen reports by IEA, the World Bank, and IRENA." | 338: "Current costs or prices: Using renewable electricity to produce hydrogen costs USD 3 to USD 8 per kg (IEA 2021) Average 2020 was $5.5/kg (likely a simple average of the IEA range above) the range of cost in 2018 was 3-6.5 USD/kg (European Commission 2020 Strategy page 4 which sites IEA 2019)...Estimated price trajectories:
- 1-3.5 USD per kg by 2030 (IEA NZE)
- 1-2.5 USD per kg 2050 (IEA NZE)

The team anchored closely around available base rates. The base rates provide wide estimates of price ranges and forecasters used those ranges as their probabilistic forecasts." | 341: "Wind and solar energy have been remarkably successful at lowering costs and increasing the portion of clean energy in the world. Successful interventions are emissions pricing, direct regulation of energy efficiency and emissions and various types of government support to advance low-carbon technologies. So assuming that these interventions will continue or even be strengthened, as proposed [here], it appears likely that hydrogen from renewable electricity would be following a similar trend or even more aggressive decrease in costs as shown in these base rates." | 339: "The low range forecast for the two time periods employed a Wright's Law formula from the IRENA document cited in the material provided and applied standard deviations…The high range forecast for the two time periods employed a Wright's Law formula from the IRENA document cited in the material provided and applied standard deviations for 2030. The 2024 estimate is more or less a linear extrapolation of the IEA forecasts between today and the 2030 estimate along with a slight downwards adjustment."

[920] 337: "On the whole, there is broad agreement across all forecasts --> the cost of hydrogen will continue to decrease into the years 2024 and 2030."

[921] 339: "One source of uncertainty is due to inflation which has recently emerged as a global problem and no systematic way to apply to the task at hand." | 344: "Significant USD inflation is occurring as of the time of writing / This price is volatile and increases if the price of energy generally increases, for example, due to war. "The cost of electrolytic hydrogen from renewable energy spiked as high as $16.80/kg in late July, three times recent price norms, according to S&P Global Commodity Insights.""

[922] 339: "Another major uncertainty is the cost of renewables used to produce green hydrogen. The planet is currently adding both wind and solar systems while the technology and deployment techniques are still maturing."

[923] 337: "Cost of green hydrogen varies regionally, e.g. access to renewable electricity. This makes the range of hydrogen cost, and thus the average hydrogen cost, more difficult to predict. Especially the price of renewable energy might be in flux regionally in the short term with a volatile energy market. | 344: The hydrogen market is highly unstable, like most energy markets today. From the reference listed link, quantifying this uncertainty is challenging and led to disagreements within the team." | 338: "Another forecaster was above the team average and IEA forecasts because of adding noise in the IEA forecast price ranges to account for potential electricity price volatility."

- There might be other radical changes to the world by 2030 (e.g. AGI) that affect this question in unpredictable ways.[924]

Arguments given for lower-cost forecasts

- Green hydrogen may be an especially efficient way to store excess energy produced by renewables, with many use cases.[925]
- There seems to be significant interest/investment in the space; some forecasters noted specific projects.[926]
- There will likely be increases in renewable energy and consequent hydrogen production capacity.[927]
    - In particular, installing sufficient solar/wind capacity to meet demand will likely result in excess power generation at especially sunny/windy times. Using the excess electricity to produce hydrogen would have a marginal cost of almost zero, while the hydrogen itself would be valuable.[928]
- The price of renewables has dropped over the last decade.[929]

---

[924] 336: "Radical transformations in technology (mainly, AGI) and/or other existential risk sources end up either ending humanity or deconstructing the notion of "the U.S."" | 341: "Nuclear war, AGI, pandemics, etc."

[925] 341: "Green hydrogen may be the best way to store energy produced by renewables for the cold seasons…Green hydrogen has many use cases."

[926] 341: "An acceleration in technology, for example, the innovation track record of 8Rivers, which recently entered the clean hydrogen market using its currently commercialized liquid CO2 technology. See also, on the distribution side, essential to lower costs: The Southwest Clean Freight Corridor, a collaboration among Hyundai Motor Co., Libertad Power, and Diesel Direct, is under construction." | 344: "Significant investment is pouring into this space with target goals like $1/kg." | 340: "It is evident that interest in green hydrogen is on the rise. The team forecasts that this increased interest will lower the cost as supply chains start to ripe[n]. Green hydrogen can supply energy for hard-to-decarbonize sectors, such as cement production and transportation, and it can be generated via solar or wind." | 336: "Renewed attention for the 'hydrogen economy' in the 2000s and 2010s, leading up to sudden growth since 2020. The European Union is making big investments in green hydrogen and there is an increasing number of international projects worldwide. More production will lead to economies of scale and the learning curve."

[927] 338: "Several mentioned that increased hydrogen-production capacity and falling renewable costs would drive prices lower." | 336: "Increased availability of renewable energy sources in general (perhaps including nuclear fusion, although 2030 is probably too early for that) and especially further cost decreases for wind and solar resources."

[928] 337: "In order to produce enough electricity and meet demand at times when there is not much sunshine and not much wind, countries will need to install sufficient Solar and Wind capacity to meet the demand (as far as possible) at these times, which means at times with lots of Wind and Sun supply will exceed demand." / "By 2030, some countries (e.g. UK) will almost certainly have so much capacity to produce electricity using Wind and Solar that, at some times, their grids will need to either stop some production of electricity or to use the surplus to charge storage cells or use other means to store electricity or use the surplus to produce hydrogen (which may be used to produce Ammonia). Hydrogen produced at these times will have a marginal unit cost of almost zero. The hydrogen produced this way will of course be of value, to replace fossil fuel gas, in industry, the chemical industry and transport. The mass production of green hydrogen will probably not come in for three or four years."

[929] 340: "Lowering prices of renewables over the last decade… show ~80% price decrease."

- There might be political pressure in the near future to increase the use of green hydrogen.[930] In particular, climate catastrophes might lead nations to incentivize or pass regulations favoring green hydrogen energy.[931]

Arguments given for higher-cost forecasts

- The cost of the projects in the IEA-provided pipeline might turn out to be high.[932] There are large, upfront infrastructure investments needed.[933]
- We might be unlikely to invest in the requisite technologies if other, similar technologies — e.g. lithium-air batteries — are more cost-effective.[934]
- The Russia-Ukraine war might cause increased competition for energy sources over the next four to five years (especially in Europe), which might slow or halt reductions in green hydrogen prices.[935] (However, others suggested the energy crisis in Europe could *speed up* adoption of novel energy sources.)[936]
- Claims by the fossil fuel industry that it will ramp up low-CO_2 versions of natural gas (mitigating the need for hydrogen) might discourage the construction of electrolysis plants; however, the forecasters note they expect this to be a temporary impact ("probably only last a few years").[937]

---

[930] 340: "Team member… projects that political pressure to use green hydrogen technology is eminent [sic]. Current media attention to this resource is modest, but with increased attention, such political pressure could actualize."
[931] 341: "Climate catastrophes [might] accelerate such that most nations drastically increase regulations and incentives."
[932] 341: "The cost of mega projects and project pipeline provided by the IEA (see graphic above)." | 336: "Dollar prices will rise, especially energy dollar prices, more this decade than last decade. This gives more uncertainty on the upside, particularly in 2024, as projects may end up more expensive than today's."
[933] 337: "A large hurdle for a hydrogen economy is building or modifying infrastructure like hydrogen fueling stations and hydrogen pipelines. This can slow adoption of the hydrogen economy and hence investment in demonstration plants."
[934] 342: "[One team member thinks] the technology is unlikely to make major progress without at least enormous investment, which is unlikely to happen given the massively better cost-effectiveness of battery technology and far better prognosis for further gains e.g. lithium-air" | 336: "Hydrogen is a bad research path, given the tremendous progress in lithium ion battery technology - as a result, I don't expect a wave of investment to emerge over the next decade, as it will be hard to compete against li-ion; and without a wave of investment, costs will not drop quickly."
[935] 336: "Because of the Russia-Ukraine war, countries and regions (especially Europe) are scrambling for energy. There's going to be lots of competition for energy sources, preventing any big reduction in green hydrogen prices over [the] next 4/5 years."
[936] 337: "The energy crisis in Europe will accelerate plans to become more energy self-sufficient. Most countries in Europe are investigating how hydrogen can be used for power generation. We can expect more demonstration plants and therefore potentially lowering cost."
[937] 337: "Lobbying by the fossil fuel industry could be based on claims that they will ramp up supply of low CO2 versions of natural gas and that hydrogen will not be needed, in order to discourage the installation of electrolysis plants. It's very likely that this low CO2 version of natural gas will not be cost-effective, thus boosting the price of what hydrogen is being produced by electrolysis. This situation would probably only last for a few years, while more electrolysis plants are installed. "Big Oil" has form in this type of activity."

- Some forecasters cited sources (in footnote) other than those provided in the prompt which led them to predict higher costs (for unspecified reasons).[938]
- Green hydrogen might remain "niche".[939]
- The growth, impact, or nearness of future technologies is sometimes over-hyped in general.[940]

Other arguments given

- Some potential "wild cards" noted by forecasters included:
    - An individual or organization unilaterally launching a geoengineering project to stimulate "global cooling".[941]
    - One detailed story about the actions of governments and utility companies (given in footnote).[942]
- Transportation was noted as an area where hydrogen energy could be adopted relatively quickly and with fewer hurdles than other sectors.[943]

---

[938] 338: "The highest forecasts came from two forecasters that sited [sic] different sources (Wikipedia and a PWC report) than others where future price estimates were not directly given."
[939] 340: "If green hydrogen continues to be a niche energy source, prices could continue to be on the higher range of projections."
[940] 341: "Comparison class, gallium arsenide "which will be the semiconductor of the future, and always will be." (A 1970s joke) Or bubble memories. [reference]"
[941] 341: "On the negative side: a nation-state or a hyper-wealthy individual might unilaterally initiate global cooling, for example injecting SO2 into the stratosphere, or fertilizing vast swathes of the oceans with iron. If cooling overshoots, coal-fired plants could become heroes."
[942] 341: "Utilities learn from the several false starts and solve the three bottlenecks simultaneously. One, hydrogen becomes a clean fuel source. Tomorrow, all fossil fuels which drive hydrogen production are eliminated. Linking Q29 – Carbon Capture System, carbon becomes the primary source of production. Secondly, project construction times go from 6+ years to possibly as low as 1 year, something that has been known to be possible under wartime conditions. Lastly, governments that to date have committed $37B USD in project support increase tax incentives and support, perhaps by7 as much as 10-fold. This policy initiative is supported with carbon price levels and favorable procurement policies. The IPCC report has not overtly recommended policies that would accelerate this fuel source. These are emission taxes or a carbon price and government support such as a carbon market and R&D tax credits."
[943] 343: "One of the largest considerations in hydrogen prices is that of demand. The question for future hydrogen production is where will demand come from? Usable hydrogen for power generation is a 2-step process, where the energy source is converted to captured hydrogen fuel and then later burned. This doesn't make sense for most conventional power plants, thus hydrogen should be thought of as an energy storage solution, when considering demand for hydrogen power. (Why not just generate the power directly if used on-site? There's an energy loss in the conversion process, so there has to be a practical reason to incur that energy loss. Energy storage and portability is that reason.) …One potential area for energy storage of hydrogen is in transportation: particularly in ground transportation… Significant infrastructure buildout is already underway, and it is unlikely we will see hydrogen enter the passenger vehicle space to compete with BEV/ICE given the current level of maturity of both markets. Trucking, in particular long-haul trucking, is an area where novel competitors to diesel could compete. Although some companies are investing in BEV solutions for short-haul trucking, long-haul BEV solutions remain largely unexplored and unproven. Some companies are developing long-haul hydrogen trucking solutions, including production expected to begin next year (2023). If successful, scale production should still be expected to take a few years… Long-haul trucking would require less infrastructure investment to begin supplying major routes in early-stage buildout, lowering this barrier to entry. Additionally, costs for hydrogen-powered vehicles are expected to be lower across the lifetime of the vehicle. In addition to

- Some forecasters objected to the notion of an unambiguous "cost of hydrogen" (though they noted it was defined clearly enough for the question to be resolvable).[944]

Cross-references with other questions

Q28: Solar and Wind Energy
Q29: Annual Direct Air CO2 Capture

Question 31: Nuclear Weapon Use

What is the probability that the use of a nuclear weapon (in a single event) will cause the death of more than 1,000 people…
…by the end of 2024?
...by the end of 2030?

Question and resolution details, prior forecasts, and other relevant sources

Results[945]

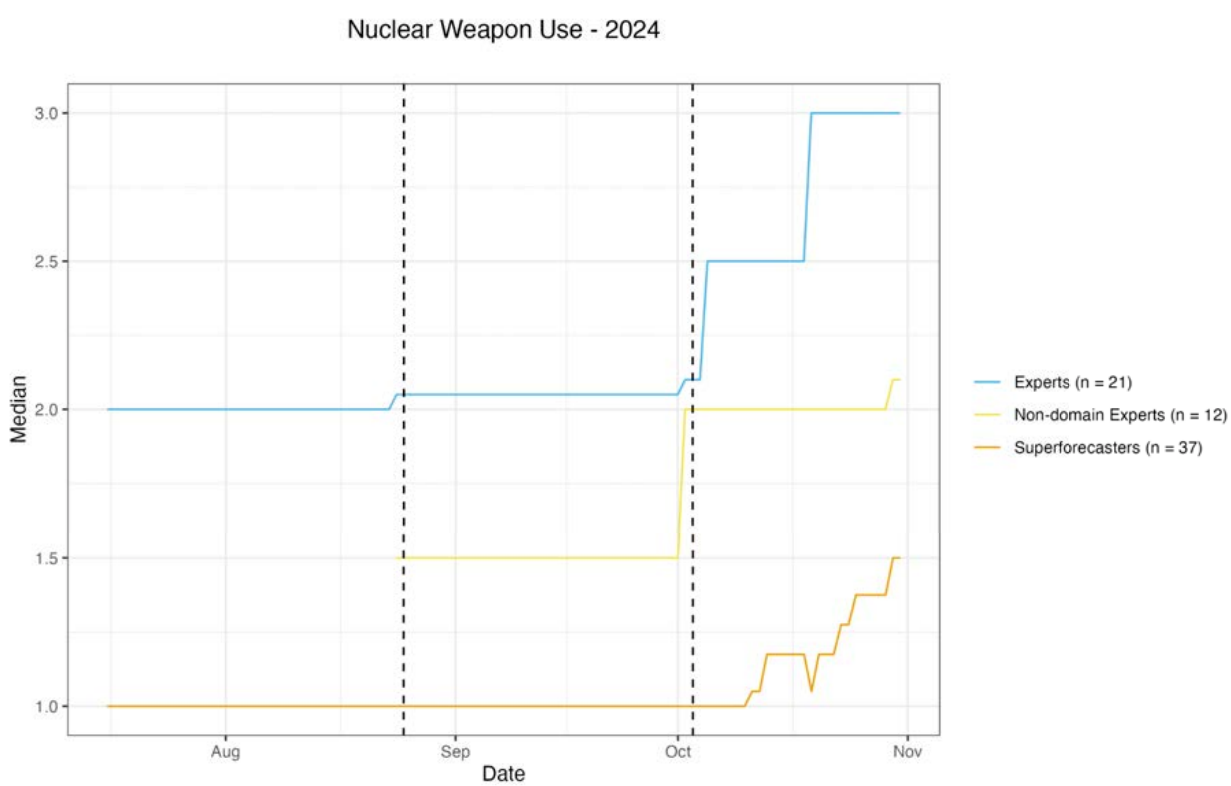
Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4

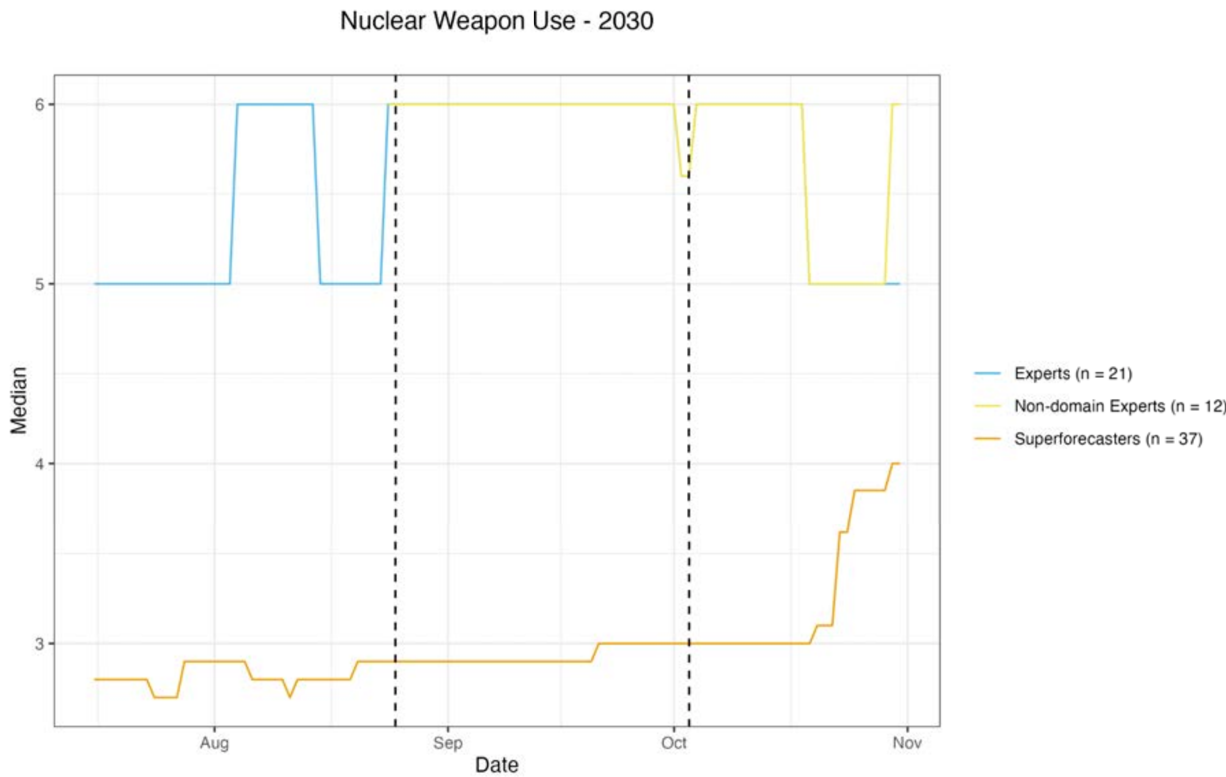
lower maintenance costs, the drivetrain offers an improved driving experience over diesel (no rumbling engine, diesel fumes, and much simpler to operate). This will be expected to reduce the barrier to entry for new commercial drivers. Thus, we expect much of the increased demand for hydrogen to come from transportation - specifically long-haul trucking - and that this demand will ramp up slowly across the next 8 years."

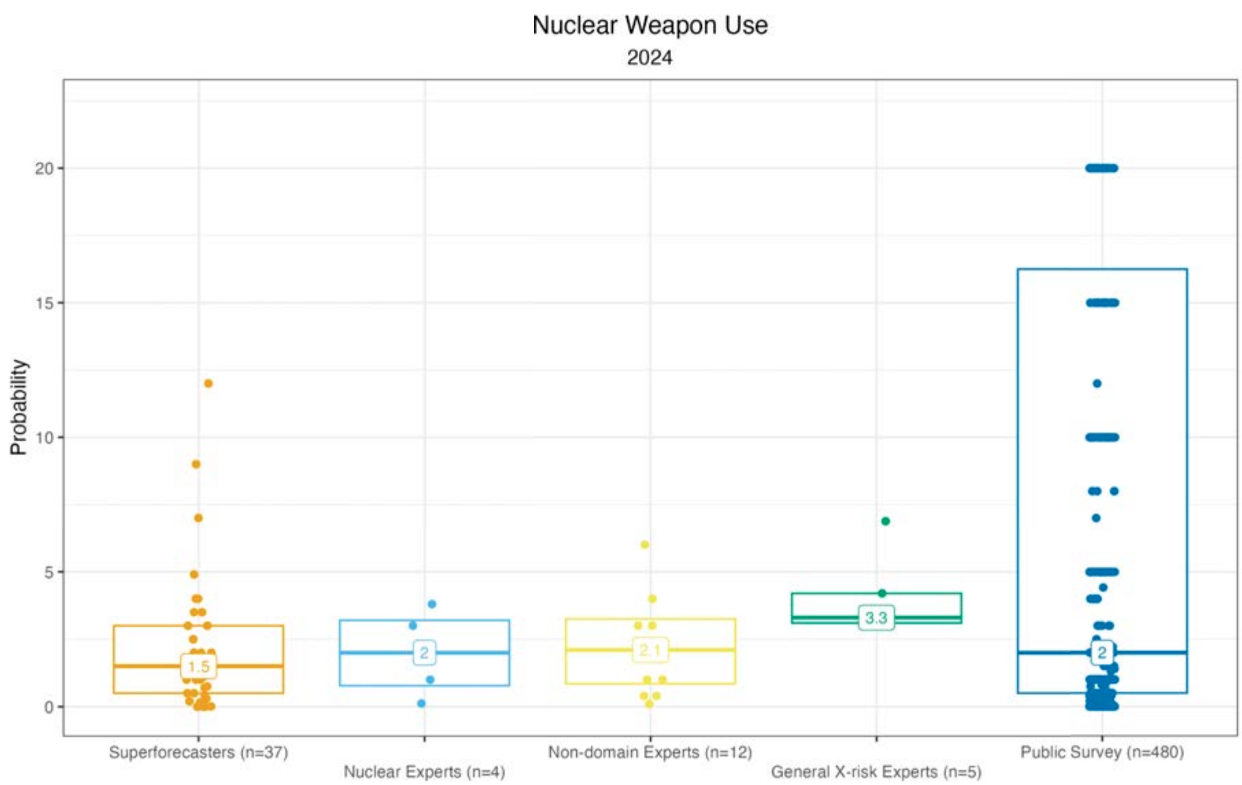
[944] 336: "It's a bit of a fiction to talk about "the" price of hydrogen. In different locations, there may be different prices. Which location do we mean? Are we talking about the levelized cost or the ma[r]ginal cost? What if we have a producer, corporate or state, who is selling at a loss? Who even defines what a "loss" means, given fixed past investments? What if there are government subsidies? Does the cost include delivery? What if the contract is part of a larger bundle that includes other goods or services? It's hard to say what "the" price is. Fortunately, the tournament clearly defines the price as what's reported in the IEA Global Hydrogen Review. However, this means that *any* external research used may be talking about different types of prices, and therefore may not be apples-to-apples with the Global Hydrogen Review report."

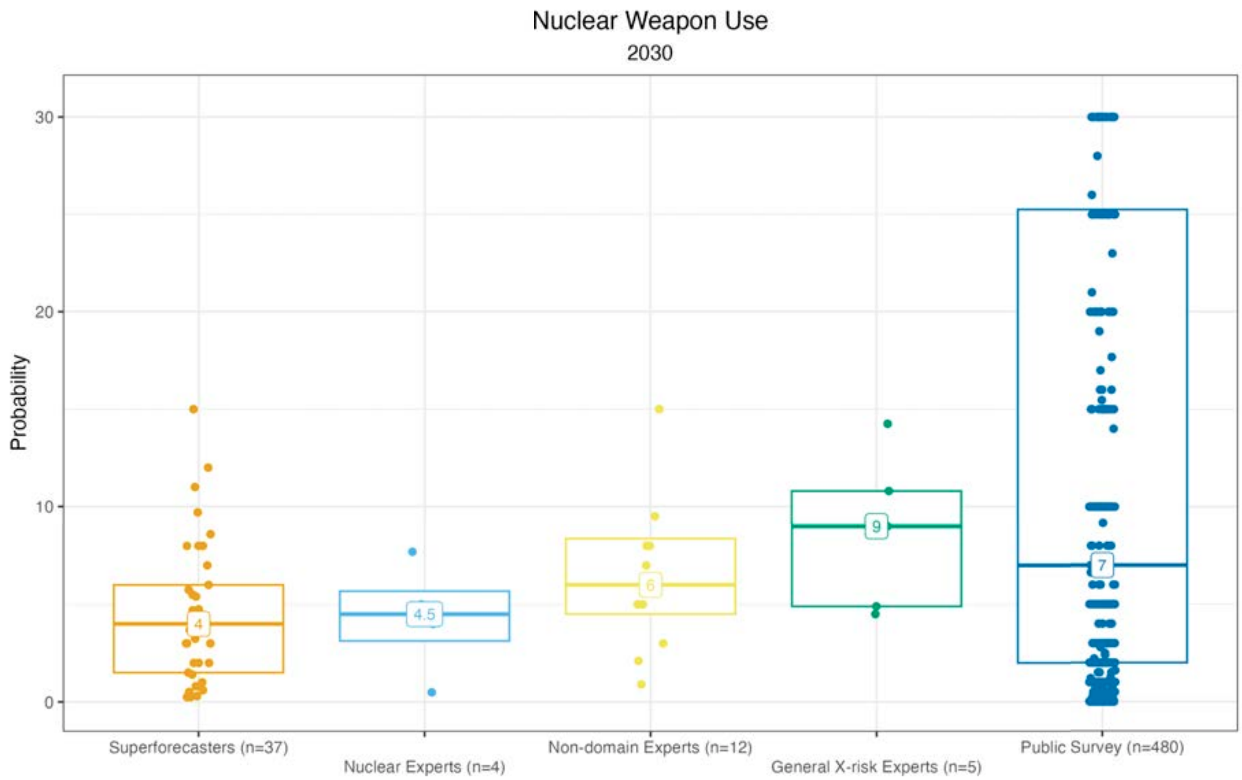
[945] Numbers of forecasters are given as of Stage 4 of the XPT. Public survey respondents were surveyed outside of the tournament context.

Super-Forecasters (N = 37)	2024	1%	1.5%	1.10	+133.33%
	2030	2.45%	4%	5.36	-32.25%
Domain Experts (N = 4)	2024	1%	2%	1.92	-11.01%
	2030	4%	4.5%	3.61	-17.32%
Non-Domain Experts (N = 12)	2024	1.75%	2.1%	3.14	+271.06%
	2030	5.25%	6%	6.44	+103.68%
Public Survey (N = 480)	2024	2%		45695.14	-
	2030	7%		457867.04	-




Nuclear Weapon Use - 2030
Median
Date
Aug
Sep
Oct
Nov
Experts (n = 21)
Non-domain Experts (n = 12)
Superforecasters (n = 37)


Nuclear Weapon Use
2024
Probability
1.5
2
2.1
3.3
2
Superforecasters (n=37)
Nuclear Experts (n=4)
Non-domain Experts (n=12)
General X-risk Experts (n=5)
Public Survey (n=480)


Nuclear Weapon Use
2030
Probability
30
20
10
0
4
4.5
6
9
7
Superforecasters (n=37)
Nuclear Experts (n=4)
Non-domain Experts (n=12)
General X-risk Experts (n=5)
Public Survey (n=480)

Sources of agreement, disagreement and uncertainty

- **Historical base rate versus present-day conflict risks**
  - Some arguments focused on the low historical base rate of nuclear weapons use, while others focused on specific, imaginable conflict scenarios. Whether this is a distinction between different groups of forecasters or different arguments made by the same forecasters was not clear.
    - "There did seem to be a distinction between forecasters relying on a historical base rate of around 1% versus forecasters accumulating risk from specific conflicts." (T338)
- **Low historical base rate of nuclear weapons use**
  - Nuclear weapons have not been used in over 75 years since the bombings of Hiroshima and Nagasaki. This suggests a low probability of future nuclear weapons use, and an even lower probability of nuclear weapons use if world wars are excluded.[946]
    - "Nukes have not been used in over 75 years and the only time it's ever happened was during a world war." (T344)
    - One team arrived at a probability of 1.5%/year after assigning 10% probability to each of three close calls that it deemed serious, also stating that "[i]f we ignore WWII as a one-time event that won't be repeated, then the same calculation comes up with something around 0.3-0.4% / year."[947]
  - **Proliferation since 1945 increases risk over base rate**
    - The historical base rate may be less reliable given the increase in nuclear-armed states since World War II.[948]

---

[946] "The base rate has sometimes been calculated by looking at the number of years since the bombings of Hiroshima and Nagasaki, but with an N of essentially 1, this is a very imprecise data point." (T336)
"There have been just two instances in the last 80 years that would have resolved this question affirmatively: both times in August of 1945 when the US used nuclear weapons against Japan causing the death of more than 1,000 people both times (at least 70k the first time and at least 22k the second time)." (T337)
"Forecasters typically weighed risk from several potential hotspots such as Taiwan-China, India-Pakistan, Israel-Middle East, and Russia-NATO against a historical backdrop of no nuclear weapon use since World War 2." (T338)
"Two uses of nuclear weapons in war, both causing >1000 deaths: Hiroshima and Nagasaki, August 1945…[disagreement existed over] whether to treat WWII as a single event that would not repeat." (T341)
"This happened already twice in the history of mankind, in the final days of WW II…This relatively uneasy balance has [held] for some 60 years, many of which were turbulent, critical, and highly dynamic." (T342)
"Nukes have not been used in over 75 years and the only time it's ever happened was during a world war." (T344)

[947] "Base rate 1%-1.5% / year based on the following estimates:

- https://forum.effectivealtruism.org/posts/PAYa6on5gJKwAywrF/how-likely-is-a-nuclear-exchange-between-the-us-and-russia estimates 1.1% annual chance of any nuclear war
- Actual use: 1 time in ~75 years (treating Hiroshima and Nagasaki as a single event for these purposes)
- Close misses: 3 (?) times in ~75 years [Many caveats: 1) that we know of; 2) definition of "close"; etc.]

Assigning 10% weight to the ones that didn't happen to reflect the probability that they could have happened, that works out to 1.3 times in 75 years => base rate of roughly 1.5%/year." (T341)

[948] "How to define the base rate is a challenge since there have been no nuclear attacks since 1945, and a reduction of near misses to 0 (at least publicly known), but the number of nuclear countries have

        - “Although there hasn't been a nuclear conflict in the last 70 years, that predated the nuclear capabilities of many potential combatants.” (T343)
        - “Nuclear proliferation has increased, and might continue to do so as it becomes clear that having nukes is a nat sec advantage (Ukraine vs North Korea fate).” (T341)

- **Present-day conflict scenarios increase risk**
    - Some forecasters listed conflict scenarios that could lead to nuclear war and increased their probabilities based on those scenarios, without explicitly listing what probability they assigned to each scenario.
        - “India/Pakistan, China/Taiwan, Russia/Ukrain[e], Israel/Iran are all nuclear risk vectors to some extent.” (T341)
        - “Ukraine, Korea, Taiwan, Iran, Israel, Pakistan, India. Lots of possibilities, each of them very unlikely. Taken together, much too likely for comfort.” (T336)
- **Only existential threats likely to motivate nuclear weapons use**
    - Considering the existence of deterrence and the nuclear taboo, an actor that did not feel itself to be existentially threatened would be unlikely to use a nuclear weapon. Only situations that could threaten a nation’s (or a leader’s) existence would lead to nuclear weapons use.[949]
        - “It's only worth the risk of retaliation if you are otherwise certain to lose...” (T337)
        - “In all of these [conflict scenarios,] the country using the nuclear attack would need to believe that an existential threat to the survival of their government, but not necessarily their country, was imminent.” (T336)
    - On the other hand, false alarms could lead to nuclear exchange even if weapons use would not otherwise be rational. One team estimated a 2.4% probability of nuclear weapons use given a false indication of missile attack.[950]
        - “Many nuclear risk analyst[s] would contest this claim that a nuclear war would be a rational and well-thought decision. In fact, many [subject matter experts] are wary that a conflict situation could go nuclear by blunder.” (T340)
- **Sources of uncertainty**
    - Secrecy around near misses
        - Some nuclear war near misses are known to have occurred, but we cannot know about the existence or seriousness of those that were kept secret.[951]

---

increased as have recent tensions.” (T336)
“And now many more actors India/Pakistan, India/China, China/someone else during invasion of Taiwan, Russia/NATO, North Korea/South Korea and nuclear terrorism.” (T339)

[949] "[Nuclear weapons] are not very helpful, if you are already winning because they are likely to provoke a nuclear response (if not from who you are attacking then from a protector of who you are attacking). It's only worth the risk of retaliation if you are otherwise certain to lose and if you feel you are certain to lose, threatening a nuclear attack could help you get a cease fire." (T337)

[950] “A causal map of the use of nuclear weapons as a response to a false alert, using a Markov Chain with 50% at each step generates (0.05)^4 * (0.05/2) = 2.4%.” (T341)

[951] “The implications of one known almost-successful coverup raise the question of how many [...] successful "near miss" coverups, and arrived at a low-confidence estimate that it should increase our total effective close-call count by 2.” (T336)

        - “Major areas of uncertainty are not knowing how many near misses have remained secret and how [ ]close to actual nuclear war we are with each incident.” (T336)
  - A multipolar world makes prediction harder
    - “The geopolitical landscape is shifting rapidly from a unipolar, post-Cold War era to a multipolar, brave new world. Uncertainty abounds.” (T337)
  - AI
    - AI’s role in resolving this question was uncertain due to disagreement and uncertainty about the timeline and behavior of transformative AI; AI could either detonate nuclear weapons on behalf of a state or intervene to prevent detonation.[952]
  - “Drone swarm technology has significant implications for both the offensive and defensive sides of the nuclear deterrence equation.” (T337)
  - Uncertainty over why nuclear war and “hot” great-power conflict did not occur
    - “[L]ooking at the period from 1960-1990, no nuclear crises resulted in the use of nuclear weapons, and 'great powers' did not engage in direct conflict with one another. However, looking at the period before then, great powers went to war frequently. It's unclear what caused this hiatus of direct conflict between great powers (potentially multi-causal), which itself presents a potential for return of direct conflict between great powers in the future.” (T343)

“Ten of these close calls happened before 1973 and there have been none since 1995, suggesting that the "base rate" for "close calls" is trending down, although this may be due to a bias towards older incidents being declassified while newer incidents remain secret.” (T337)
Disagreement existed over “[h]ow many close calls [there] have been…” (T341)

[952] “Arguments could be made that AI could lead to the both an increase in likelihood (i.e. AI acting on behalf of a country detonates a nuclear weapon) or a decrease in likelihood (i.e. AI steps in to prevent the detonation of a nuclear weapon).” (T337)
“Additional points of discussion tie to the fact that parts of the team consider 2030 as a year in which AI could play a role in nuclear usage, while others disagree on the date as the emergence date.” (T344)

Arguments given for low-end forecasts

- **Deterrence**
  - Nuclear-armed nations are aware that use of nuclear weapons would result in retaliation.[953] Possessing nuclear weapons may have more value than using them.[954]
    - "Mutually Assured Destruction, however much it may violate every basic tenant of morality and decency, is a powerful deterrent against the use of these weapons." (T337)
  - As discussed above, however, the possibility of a nuclear launch triggered by a false alarm decreases the power of deterrence.
- **Tactical nuclear weapons or use against military targets**
  - Even if nuclear weapons are used, forecasters argued, ≤1000 people would be killed if a tactical nuclear weapon were used or if a nuclear weapon targeted a military or remote target.[955]
    - "[In a scenario in which Russia uses a nuclear weapon,] I don't see the goal being all out war but as an area denial weapon to save government… Even if a tactical nuke is used, it will be likely on a remote military target, test site, or evacuated city (as an example) and unlikely to kill many people." (T336)
- **Terrorists are unlikely to have nuclear weapons capability**
  - Some forecasters believed that only a terrorist or non-state actor would plausibly use nuclear weapons within the question's time period. This factor supports lower predictions because terrorists are unlikely to have the capacity to acquire nuclear weapons unless weapons are intentionally supplied by a state actor.[956]

---

[953] "There's a reason why nuclear hasn't been used since the two over Japan. A moral prohibition. And even if you're not particularly moral, Putin, you are held back by the morality of others." (T336)
"Most state actors receiving a nuclear strike who have nukes would likely respond with at least one in kind, so sovereign actors knowing using a nuke invites a nuke should make the chance of using one very low." (T339)
"Also some team members observe that deterrence works, and no one nation would be suicidal to start a nuclear war." (T340)
[954] "North Korea - imho, is just trying to be taken seriously on the world stage…At present, there isn't a rational cost/benefit calculus for a country that possesses nuclear weapons to use them, even at a tactical level. But that calculus is constantly shifting." (T337)
"[The] value of nuclear weapons is higher if they are unused." (T341)
[955] "[Uncertainty about t]he possibility of strategical nuclear weapon attacks against sparsely populated infrastructure; however, such an event could cause collateral casualties." (T338)
"Finally, it was debated whether a nuclear weapon use would be counter-value (that is to target large urban areas), which would result in more than 1000 casualties, while others hypnotize that a use would be against military targets that would cause limited human casualties…[t]here is also the potential for a more limited, "tactical" use of nuclear weapons against military targets in a non-total war such as that ongoing in Ukraine; note, however, that such nuclear detonations may kill <1,000 people per attack (unlike when deployed against cities) and thus there is uncertainty about whether this would actually positively resolve the question." (T343)
[956] "Despite past occasional claims to the contrary, apparently no intact nuclear weapon has ever been stolen or sold on the black market, and the difficulty of either putting one together or detonating an existing one without its security codes remains beyond the capability of existing terrorist organizations." (T337)
"Non-state actors, should they get hold of the materials and build a bomb, [have] to figure out how to deliver it." (T341)

- "The only realistic scenario I see for this to occur by 2030 is by state-sponsored terrorist attack, which monitoring and intelligence networks have so far effectively prevented…[but a] state actor looking for deniability and/or terrorist money might supply a detonatable weapon to a 'non-state' actor." (T337)

Arguments given for higher-end forecasts

- **Russian aggression**
  - Russia's invasion of Ukraine creates the possibility of catastrophic loss or humiliation for Russia. This existential threat to Vladimir Putin could motivate him to use nuclear weapons or make him less deterrable.[957]
    - "The use of nuclear weapons in the current Russia-Ukraine conflict could escalate the regional, conventional-warfare conflict into a global nuclear conflict." (T344)
    - "The most likely usage by the end of 2024 is if Russia suffers a sudden and catastrophic loss in the war in Ukraine, or the war in Ukraine spills over significantly into Russia without Russia's ability to conventionally repeal the attack…Russia has an abundance of nukes and a massive failure in Ukraine or humiliation in the Black Sea Naval Fleet could be existential for Putin and his inner circle." (T336)
    - "Given already strong sanctions against Russia, there's little more the global community can do to impose non-violent punishments for misbehavior. This lack of a deterrent threat could impact Russia's cost-benefit analysis for the use of nuclear weapons." (T343)
  - Russia might also feel threatened by neighbor states who take the opportunity to demonstrate their independence.[958]
- **Unusually high current risks**

"[T]he team agreed that these action[s] could only happen if it was carried out by non state actors which they believe by strict govt rules, regulations and policy will make it impossible for non state actors to access such a weapon…" (T342)

[957] "Putin has started indirectly threatening the use of nuclear weapons in his latest televised ad[d]ress." (T341)

"[T]here is a growing fear, threat, anxiety and probability based on how some state actors have been conducting themselves in the public space in recent times…example of such state actors is Putin - the president of Russia and the war he is prosecuting in Ukraine and his threat to use nuclear weapons, his continue[d] attacks and shelling of the nuclear weapons site of Ukraine knowing fully well the implications and eventually taking over the site through forceful military aggression is a serious pointer and visible threat that can cause the death of more than 1,000 people in the years under review if possible checks are not put in place to mitigate such occurrences." (T342)

"Buffeted by significant battlefield setbacks in Ukraine, new expressions of concern from his closest sympathizers, and widespread international condemnation of his actions, Vladimir Putin this week made the most nearly overt and explicit threat yet to use nuclear weapons in his war on Ukraine. The chances of the first use of a nuclear weapon in Europe have increased dramatically with Putin's support for annexation of occupied Ukrainian territory, a move that has zero chance of being recognized by the international community at large, combined with his vow that Russia will use "all means at its disposal" to defend its territorial integrity." (T344)

[958] "As Russia's neighbors perceive additional weakness in light of the Ukrainian conflict - and as their fighting capabilities in that region continue to deteriorate their total fighting capacity - new opportunities to demonstrate independence from Russia are already being exploited. In the absence of a reliable conventional threat, Russia may resort to a nuclear threat in this region." (T343)

- Several teams argued that risks of nuclear war are currently unusually high, mentioning the Russian invasion of Ukraine and tensions with China over Taiwan.[959]
    - **“**The probability of a nuclear weapon strike is at a local maximum, with the war in Ukraine and the Taiwan question.” (T343)
    - “Some team members note that we're living in the most dangerous era of potential nuclear use since the Cold War…Many forecasters pointed out that the current Ukraine-Russia conflict may appreciably raise the chance for positive resolution in by the end of 2024. Recency bias in play?” (T340)
    - “There were essentially two camps as to whether the probability should be evenly spread across time, or if we are in an unusually volatile and risky time before the end of 2024.” (T336)

Other arguments given

Arguments in favor of lower forecasts:

- Since 1945, there has been a strong international norm against nuclear weapons use.[960]
- An attempted use of nuclear missiles could fail.
    - “[D]elivery is difficult: A state actor (Russia, China, US, etc.) could launch a missile, but it could get shot down.” (T341)
    - “[Uncertainty existed about] [t]he possibility of failed attacks via missed target, weapon malfunction or interception.” (T338)

Arguments in favor of higher forecasts:

- One team used [Gott’s formulation of the Doomsday argument]() to estimate probabilities in the range of 7–12% by 2030;[961] however, this team’s overall median was substantially lower.

---

[959] “Increasing 2024 a bit and decreasing 2030 on comments that we maybe in a particularly high risk period, with Ukraine and Taiwan situations.” (T336)
“With this Cuban Missile Crisis-esque atmosphere, the post-conflict times may affect the global attitude to this outcome.” (T338)
“Several of these [risks] are currently heightened (India/China and Russia/NATO)…” (T339)
“Some team members note that we're living in the most dangerous era of potential nuclear use since the Cold War. The current Russian war against Ukraine was cited by a number of team members to raise the base rate to more than 1.5%/year…Many forecasters pointed out that the current Ukraine-Russia conflict may appreciably raise the chance for positive resolution in by the end of 2024. Recency bias in play?” (T340)
The base rate implied by our median forecast is approximately 2-2.5% yearly risk, relatively higher in the short term due to the current war between Russia and Ukraine. (T344)

[960] “Balancing a stronger norm against nuclear weapons use (US openly thought about using nuclear weapons in Korea, which it would be much more unlikely now) with increased nuclear proliferation to arrive at numbers close to historical base rate.” (T341)

[961] “A Bayesian estimate for the probability of a qualifying event can be made following Gott's version of the Doomsday Argument…[incorporating close call data] we imagine ranking all the episodes by "badness" and then asking the joint question "how many new close calls we expect to happen by 2030" (standard Bayesian analysis generalizing Gott's) and "what's the probability that one of them will be even worse than the worst one so far" (1/(N+1) for each new event given N past events). This analysis is not very sensitive to the threshold for what to count as a "close call", and seems to be robust against different starting points for observations (e.g. Yom Kippur War or post-Cold war) and produces values in the range 7-12%.” (T336)

- The FBI may have found classified nuclear weapons information in Trump's home.[962]

Cross-references with other questions

Q15: Non-State Actor Bioweapon 1k Deaths
Q16: State Actor Bioweapon 1k Deaths
Q17: Non-State Actor Bioweapon 100k Deaths
Q18: State Actor Bioweapon 100k Deaths
Q32: Total Nuclear Warheads
Q33: Countries with Nuclear Warheads
Q34: Country-by-Country Nuclear Use

Question 32: Total Nuclear Warheads

How many total nuclear warheads will be in military inventories globally…
…by the end of 2024?
…by the end of 2030?
…by the end of 2040?

Question and resolution details, prior forecasts, and other relevant sources

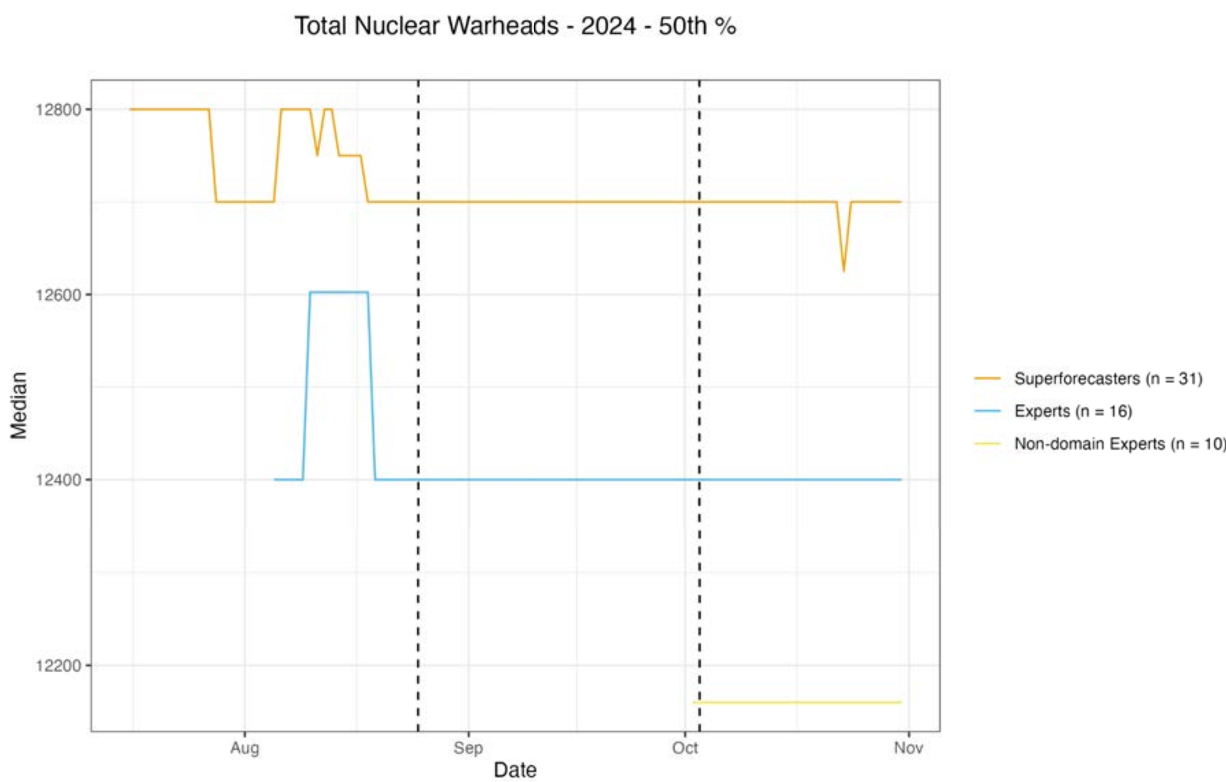
Results[963]

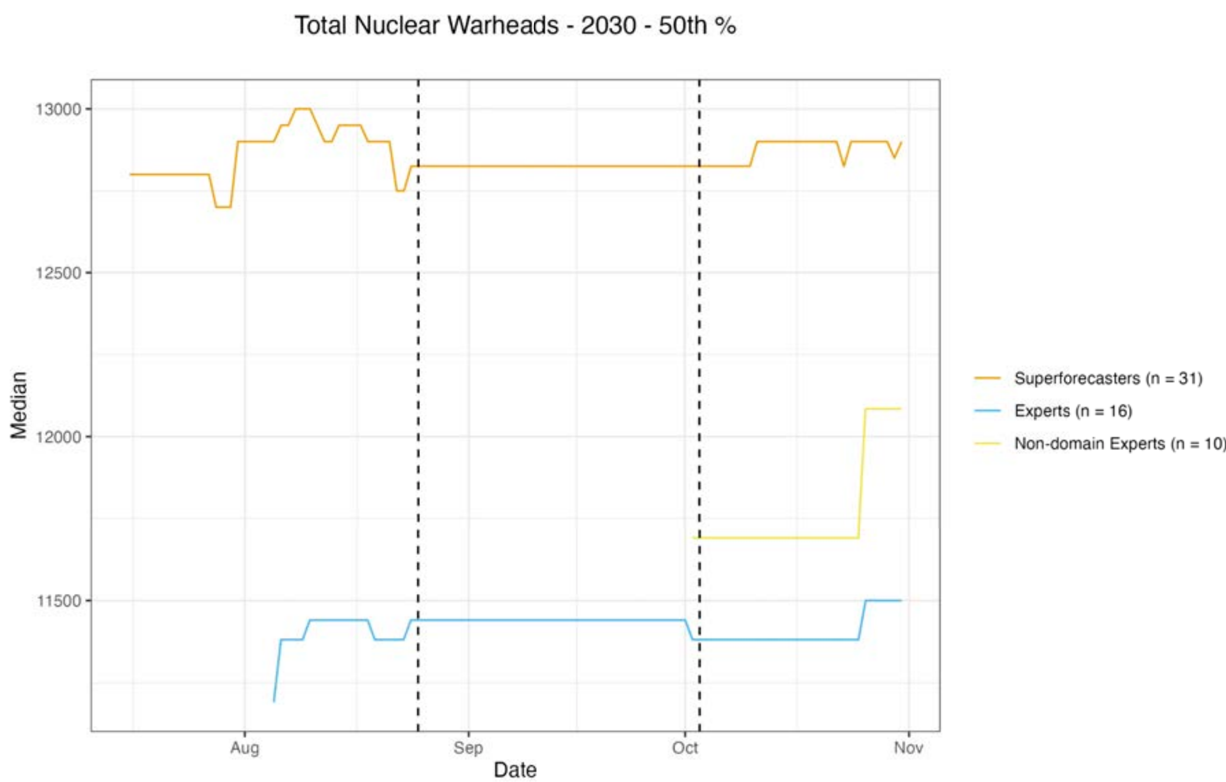
Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 31)	2024	12800	12700	1715.91	-61.11%
	2030	12950	12900	1997.06	-37.74%
	2040	13500	13500	2349.72	-0.31%

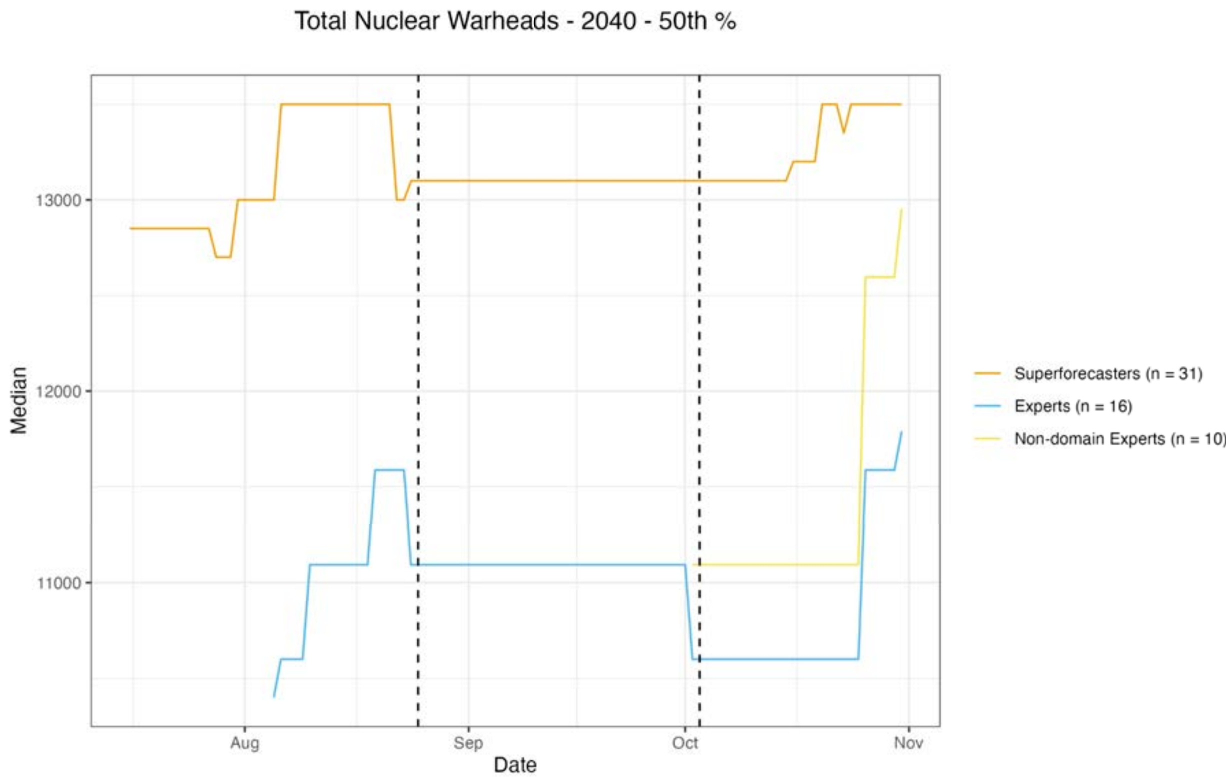
[962] "I've made the change considering how sloppy the US has been with nuclear weapon information. It's not fully public what the FBI found in Trump's safe in the raid of the former president's home. But it seems probable it was nuclear weapon information. Nuclear weapon information should never be found in a country club." (T336)
[963] Numbers of forecasters are given as of Stage 4 of the XPT.

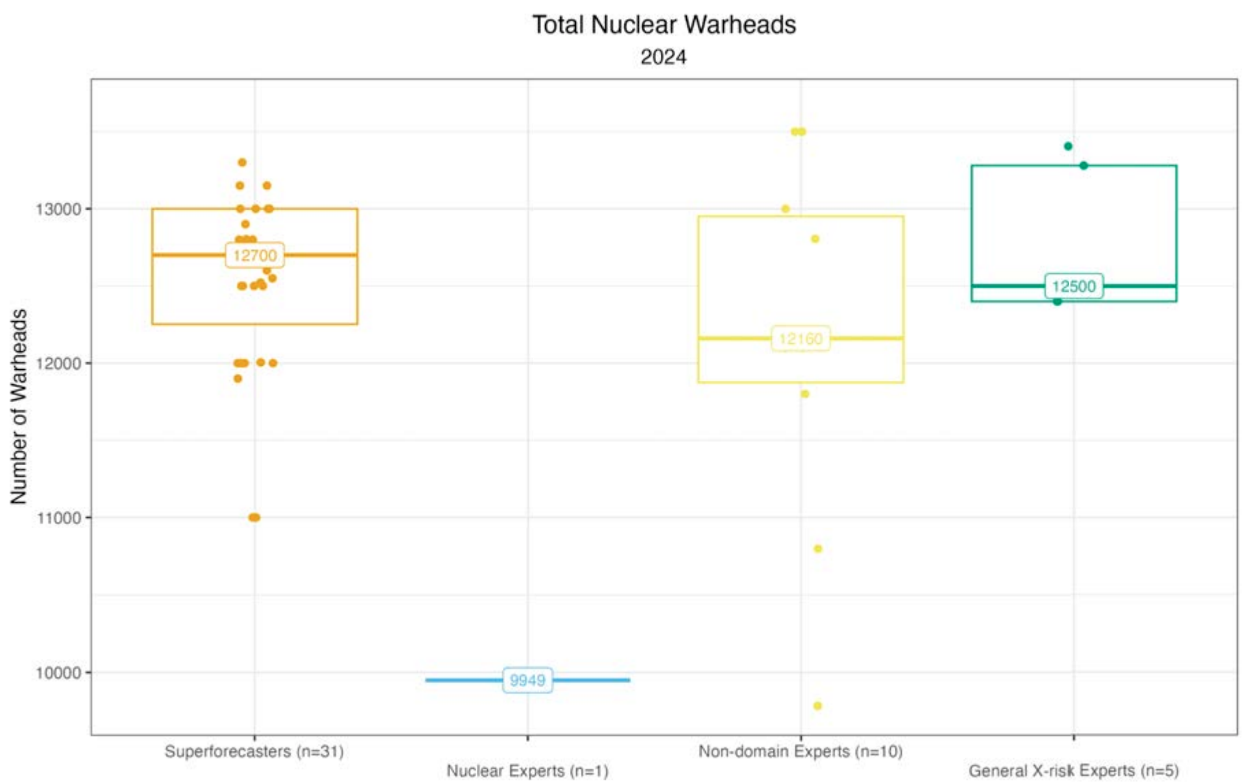
Domain Experts (N = 1)	2024	NA[964]	9949	NA	NA
	2030	NA	10390	NA	NA
	2040	NA	11990	NA	NA
General X-Risk Experts (N = 5)	2024	12840	12500	546.54	-8.22%
	2030	11250	11500	2313.3	-12.62%
	2040	10100	10200	4803.85	-12.55%
Non-Domain Experts (N = 10)	2024	11950	12160	1357.87	-13.72%
	2030	10150	12084.5	1506.63	+43.61%
	2040	8800	12952.5	4378.26	-9.46%

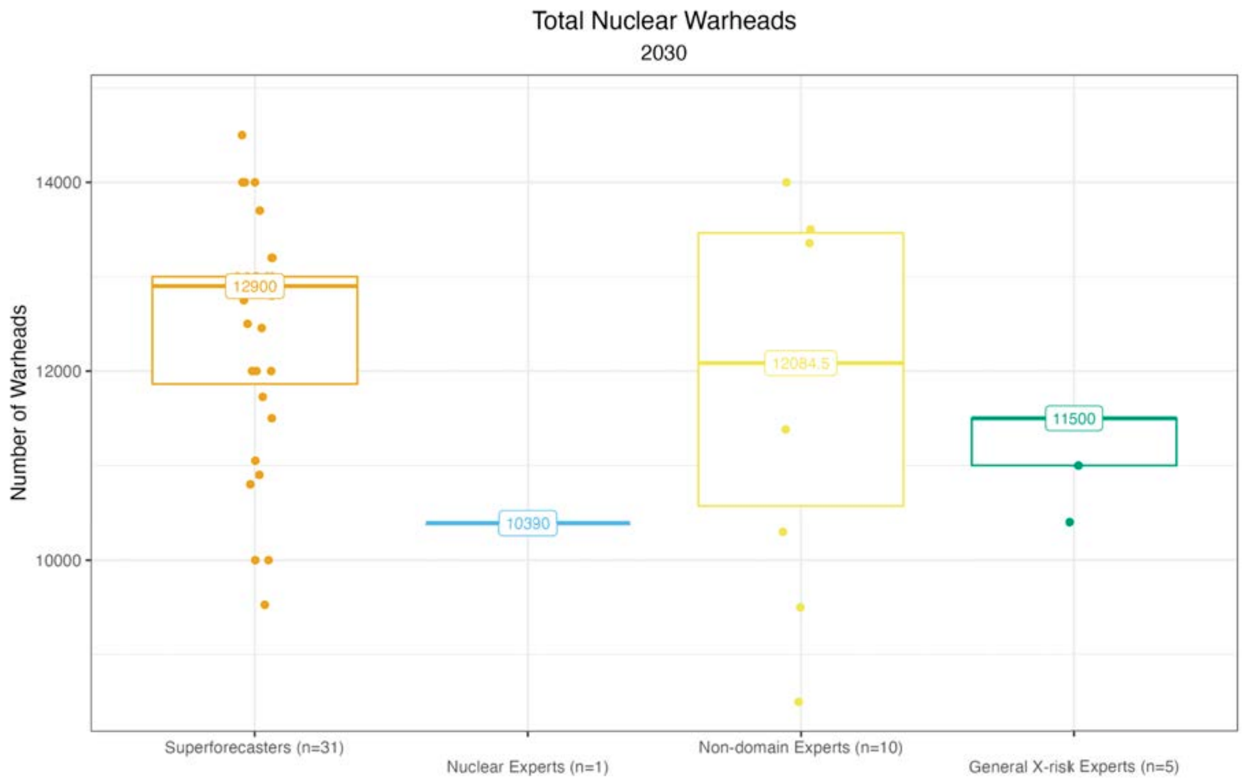
[964] No forecasters in this group answered this question in Stage 1.

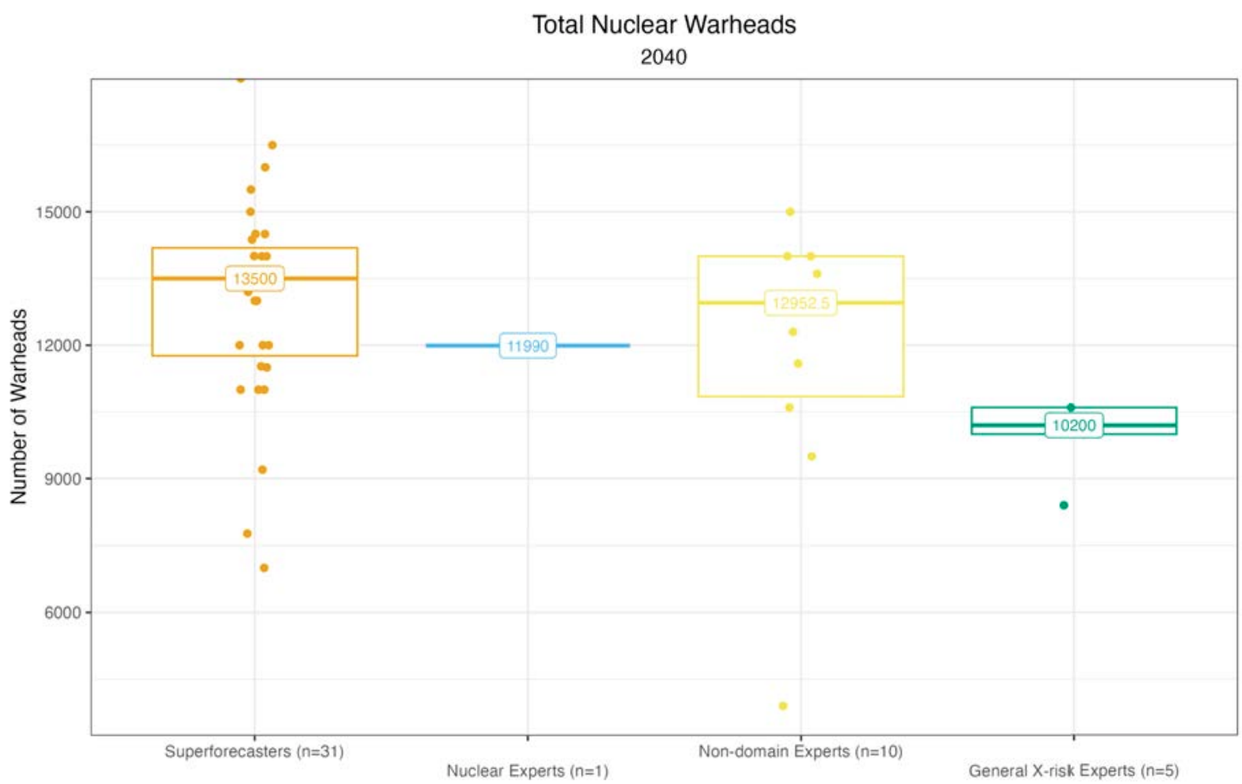

Total Nuclear Warheads - 2024 - 50th %
12800
12600
12400
12200
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 31)
Experts (n = 16)
Non-domain Experts (n = 10)


Total Nuclear Warheads - 2030 - 50th %
13000
12500
12000
11500
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 31)
Experts (n = 16)
Non-domain Experts (n = 10)


Total Nuclear Warheads - 2040 - 50th %
Median
13000
12000
11000
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 31)
Experts (n = 16)
Non-domain Experts (n = 10)


Total Nuclear Warheads
2024
Number of Warheads
13000
12000
11000
10000
12700
9949
12160
12500
Superforecasters (n=31)
Nuclear Experts (n=1)
Non-domain Experts (n=10)
General X-risk Experts (n=5)


Total Nuclear Warheads
2030
Number of Warheads
14000
12000
10000
12900
10390
12084.5
11500
Superforecasters (n=31)
Nuclear Experts (n=1)
Non-domain Experts (n=10)
General X-risk Experts (n=5)


Total Nuclear Warheads
2040
Number of Warheads
15000
12000
9000
6000
13500
11990
12952.5
10200
Superforecasters (n=31)
Nuclear Experts (n=1)
Non-domain Experts (n=10)
General X-risk Experts (n=5)

Sources of agreement, disagreement and uncertainty

- **Baseline of nuclear weapon stockpiles**
    - Compared to other forecasting questions, this question had a clear quantified baseline of approximately 12,705 nuclear weapons as of early 2022 or approximately 13,150 as of mid-2021.[965] All teams who stated their baselines used one of these two numbers.
        - One team noted it was uncertain about the reliability of stockpile numbers given political incentives to conceal or lie about facts.[966]
- **New nuclear powers**
    - Forecasters described uncertainty about whether new countries such as Iran and Japan might enter the nuclear club.[967] There was agreement that these countries' decisions to pursue nuclear weapons would be strongly influenced by the broader geopolitical situation.
        - "One key uncertainty was whether new countries (e.g., Iran, Japan) might begin to expand their own stockpile, particularly if they perceived international security deteriorating (e.g., following Russia's invasion of Ukraine)." (T338)
        - "Iran will have weapons for sure by 2040, and maybe by 2030." (T339)
        - "Japan and Taiwan are also latent nuclear powers which may seek nuclear warheads before 2040 depending on relations with China." (T336)
        - "In a multipolar world where a number of countries are working to amass stockpiles of their own the logic of proliferation is more unpredictable." (T339)
- **Limited change by 2024**
    - Teams noted that limited change would be expected by 2024, with more varied predictions for 2030 and 2040.
        - "While 2024 is probably too short notice to really affect anything, by 2030+2040 there may be some change." (T339)
        - "The trend is unlikely to change in the near-term [by 2024] because it is being driven by the dismantling of retired weapons." (T340)
- One team noted uncertainty about "[h]ow long would it take the participating nations to ramp up nuclear production?" (T343)

Arguments given for low-end nuclear weapons numbers

- **Decreasing trend, especially for U.S. and Russia**

---

[965] Data is taken from the Federation of American Scientists. Team 341 also mentioned the Stockholm International Peace Research Institute's Jan. 2022 estimate of 13,080 weapons.

[966] "[T]he US and Russia are believed to have approximately 3200 weapons in storage slated for destruction. But given the potential incentives for US and Russia to lie or confuse the issue, forecasters expressed uncertainty about the reliability of that number. They also noted that distinguishing between the US nuclear stockpile and those weapons that have been retired and are just being stored while waiting for dismantlement can be difficult, and yet getting those distinctions right is an important component of forecasting this question well." (T337)

[967] "There is significant disagreement about long-term future trends such as…whether new players such as Iran, Saudi Arabia, Japan, Taiwan, or Germany will acquire nuclear weapons and potentially trigger an arms race, which could also possibly include Pakistan and India." (T336)

  - If current trends continue, U.S. and Russian nuclear weapons stockpiles will decrease,[968] although the rate of decrease has been slowing.[969]
    - “In general, the recent trend has been for the US and Russia to decrease their stockpiles by ~100 – 500 per year, although the decrease has been slowing over time…This trend will likely continue until about 2030, followed by a less steep drop (estimated at 100 warheads per year).” (T336)
    - “Strong external factors would be required before a significant change in current trends.” (T343)
- **Building from low base**
  - Decreased U.S. and Russian stockpiles could be offset by an increase in other nations’ stockpiles.[970] However, those nations that are most likely to seek more nuclear weapons are starting from a low base number of weapons, so even large relative increases in their stockpiles would have limited effect on the global warhead total.[971]
    - “Other nations are becoming nuclear-capable, but they're likely to have around ~100 (e.g. India, Pakistan, Israel), which doesn't make a huge impact on the total.” (T342)
- **Small stockpiles are sufficient**
  - Over time, nuclear weapons become more sophisticated, so that nations can achieve their deterrence goals without increasing their stockpiles.[972] In addition, existing stockpiles may already be more than is necessary for deterrence.
    - “Deterrence doesn't require the capability to irradiate a continent.” (T342)
    - “100 nuclear warheads is the pragmatic limit and use of government funds to maintain more than 100 nuclear weapons does not appear to be rational.” (T341)
    - “For a non-superpower, having nuclear weapons can create leverage with others during negotiations but the marginal use of that leverage decreases pretty quickly so there is no need for a North Korea to stockpile 3,000 warheads when you only need 20 well hidden ones to maintain deterrence.” (T338)

---

[968] “A linear model assuming a logarithmic continuation of the trend for the past 12 years would suggest that all remaining retired warheads would be dismantled by 2028, suggesting that (optimistically) in the mid-term future, numbers would start to be dominated by the active warheads, not the retired ones.” (T336)

[969] “"The rate of reduction is decreasing," wrote one forecaster. "Which means that we [could be] reaching a saturation point marking a probable minimum of nuclear warheads. For instance, between 2012 and 2022, the reduction rate has been 615 warheads per year on average, while if we restrict the comparison to between 2017 and 2022, we will find a reduction of 369 warheads per year (while the 2012-2017 period saw a reduction of 860 warheads per year).”” (T337)

[970] “Median forecasts estimated that the US and Russia would only reduce their stockpiles modestly, and that any reduction would be largely counterbalanced by an increase in China's stockpiles.” (T338)

[971] “[For 2030,] [t]he countries most likely to build up a stockpile (e.g., North Korea, Iran) won't affect the overall number much.” (T340)

[972] “Due to modernization, just because a country’s total stockpile diminishes, doesn’t mean a country’s capabilities have diminished.” (T337)
“At the end of the cold war, nuclear proliferation resulted in over 60,000 individual warheads between the US and USSR. That level of nuclear weapon capability provides no tangible value at the margin in my opinion. I don't think that leading nuclear superpowers, US and Russia, will increase production to return to their former capacity given the lack of value at the margin.” (T339)

- **Maintenance cost incentivizes smaller stockpiles**[973]
    - “[T]here is downward pressure on the number of warheads from maintenance & cost challenges…” (T339)
    - “[Nuclear warheads] are expensive to maintain…the US spends $60 billion per year to maintain its nuclear arsenal. For example, their pits degrade over time from radioactive decay and must be remanufactured.” (T341)
- **Disarmament campaigns and public opinion**
    - Some forecasters expected disarmament campaigns and anti-nuclear-weapons public opinion in general to exert downward pressure on weapons stockpiles.[974]
        - “I expect efforts of nuclear disarmament [such as the Nuclear Threat Initiative] to cancel countries' pressure to arm themselves…In the long term, the team is confident that the pressure of [disarmament] will continue and the nuclear stockpile will continue to shrink.” (T344)
        - “The goal of nuclear disarmament activists is zero warheads by 2045, which should inform our "best case" forecast.” (T340)

Arguments given for high-end nuclear weapons numbers

- **Russian aggression**
    - Russia under Putin is likely motivated to increase or slow the decrease of its nuclear weapon stockpiles.[975]
        - “Given Russia's increasing position as a pariah in the global stage, they may increase their stockpiles more dramatically than would appear to be rational given other economic priorities.” (T339)
        - “[Forecasters cited] [t]he potential for the war in Ukraine and other tensions between the US and Russia to lead to a new Cold War…Russia may already be reactivating weapons that had been stored and/or slated for dismantlement.” (T337)
- **Renewal of treaties is getting harder**

---

[973] “[T]here is downward pressure on the number of warheads from maintenance & cost challenges…” (T339)
“They're expensive to maintain and, even without considering treaties, it makes sense that the US and Russia would not be interested in upkeep with the end of the cold war and MAD policy.” (T341)
[974] “Anti-nuke public sentiment and budget constraints.” (T337)
“[T]here is downward pressure on the number of warheads from…lobbyism (e.g. global zero initiatives)…” (T339)
“[T]hese data set distribution shows unwillingness on the part of various military formations of nations to have or manage warheads in their inventories and willingness to gradually do away with it as a result of many negative factors and global condemnation and UN restrictions on its acquisition and usage, which is been globally seen as an existential risk and catastrophe to humanity.” (T342)
“Circumstances surrounding the actual use of nuclear weapons would likely determine whether the effect on public opinion would result in increases or decreases in total numbers of warheads, with potentially different impacts over the short and long term.” (T343)
[975] “It is also plausible that the number of active warheads will increase in the foreseeable future due to the Russia – Ukraine conflict reducing the desire to disarm and…the debated desire of Russia to increase its number of tactical nukes.” (T336)
“I expect recent events to slow the decline a little - Russia and the US both have a new reason to keep nukes around, and that may also cause China to invest in more.” (T342)
“Another argument for increased rearmament in the short term was given due to the "nuclear" tensions raised due to Russia's 2022 invasion of Ukraine.” (T344)

- The New START treaty between the U.S. and Russia, which limits the number of deployed strategic nuclear warheads to 1550, expires in February 2026. One team stated that New START "should hold US/Russia stockpiles steady till 2026." (T341) However, after that time, forecasters believed that the current geopolitical situation, in particular Russia's invasion of Ukraine, would make the renewal of this treaty or the ratification of potential new treaties harder.[976]
  - "[For 2030,] [c]urrent conflicts and geopolitical tensions will preclude the negotiation of new disarmament treaties, which will take time to implement in any event…[for 2040,] Renewal of these treaties is in doubt and would end this stability." (T340)
  - Forecasters who expected the declining trend to reverse cited a "sense that the nuclear non-proliferation treaty is on thin ice, especially in the wake of Russia's invasion of Ukraine."[977] (T337)

- **China's goal of 1000 nuclear weapons**
  - Forecasters attributed a goal of a 1000-weapon stockpile to China, which would be a significant increase over China's current estimated inventory of 350 warheads.[978]
    - "A forecaster noted that China has signed its goal of a nuclear stockpile of 1,000. In 2022, China has 350 warheads…Chinese war doctrine suggests that more nukes are needed." (T341)
- **Feedback loops among nuclear powers**
  - A typical nuclear power's armament decisions are motivated in part by what other nuclear powers are doing. This creates the possibility that one nation's decision to increase its nuclear weapon stockpiles could create a positive feedback loop or "proliferation contagion," motivating other nations to increase their stockpiles or acquire new weapons.[979]
    - "Warhead increases and reductions have in the past been a response to perceived increases or reductions in nuclear threats." (T339)
  - Forecasters believed that American and Russian armament decisions would strongly affect other nations' decisions. Smaller nuclear powers, however, could also trigger positive feedback loops.

---

[976] "It is also plausible that the number of active warheads will increase in the foreseeable future due to the Russia – Ukraine conflict reducing the desire to disarm and ability to negotiate new treaties…" (T336) "Heightened tensions between the United States and Russia over Ukraine and between the United States and China over Taiwan shows no signs of abating. New arms control agreements between the great powers are thus unlikely in the medium-term." (T340)
[977] ""Because Ukraine once possessed nuclear weapons but gave them up when it joined the NPT in 1994, Russia's renewed aggression makes it look as if the treaty's purpose is to keep weak countries defenseless and prey to the nuclear-weapon states."" (T337, citing a *Foreign Policy* article)
[978] "One forecaster linked to this article:
https://www.armscontrol.org/act/2021-12/news/pentagon-sees-faster-chinese-nuclear-expansion
And highlighted this excerpt: "*China is accelerating its development of strategic nuclear warheads in an effort to amass 700 by 2027 and 1,000 by 2030, more than doubling last year's estimate, according to the U.S. Defense Department's 2021 China military power report*."" (T337)
"Presuming that China eventually has no less than 1,000 warheads by 2040 with the other declared nuclear powers largely maintaining the status quo." (T339)
[979] "One forecaster noted that Russia's global ambitions, in combination with a NATO threat, in combination with multiple military exercises produce a positive feedback loop for maintaining or increasing their stockpile…A forecaster noted the tit-for-tat cooperation feedback loop where more warheads encourage others to increase theirs." (T341)

        - “For example, an increase in Chinese stockpiles could spur increases in Russian and Indian stockpiles. Increases in Indian stockpiles could in turn incentivize Pakistan to increase stockpiles. Pakistani increases could in turn spur new nations seeking nuclear armament in the Middle East. This would likely result in increases in Israeli stockpiles, etc. A proliferation contagion could also run the other direction, starting with any nuclear-armed nation.” (T343)
        - “Iran produces a nuclear weapon, triggering Saudi Arabia to produce their own, with the help of Pakistan, possibly prompting India to increase its arsenal as well.” (T336)
        - “If Russia and China increase their numbers of warheads then the West unfortunately should as well, in order to maintain strategic deterrence. This feedback loop was visible in the Cold War and it could easily begin again. Furthermore, if China increases its numbers, then India may feel the need to also do so, which will then incentivize Pakistan to do the same. These arms races can easily spiral into thousands of new weapons being made.” (T337)
    - These feedback loops increased forecasters’ uncertainty and disagreement about whether smaller nuclear powers would increase stockpiles.
        - “The minor nuclear powers (UK, France, China, Israel, India, Pakistan, North Korea) have not increased their stockpiles in many years and seem unlikely to increase them significantly in the future, barring paradigm-shifting events…[but] growth in tactical nuclear weapons by nations such as Pakistan and North Korea could add to the totals significantly and these countries seem motivated to do so.” (T336)
        - “One likely source of variance among forecasters is the likelihood they assign to the probability of an arms race. We’re assessing the probability of an arms race that results in >2x nuclear stockpile growth as <5%, but it’s possible another team assesses this probability as >5%. If so, the potential number of nuclear warheads could easily be >2x our current maximum estimates, especially by 2040.” (T343)

Other arguments given

Arguments in favor of lower forecasts:

- “Many weapons are already marked for destruction and the process may be difficult to reverse.” (T337)
- A new arms race focused around AI could distract from nuclear armament, or AI could lead to human extinction.[980]

Arguments in favor of higher forecasts:

- The U.S.’s nuclear weapons budget is increasing.
    - “Requested [US] budget for nuclear weapons is expanding between 2019 and 2024…The expansion of the USA nuclear weapons budget directly relates to the increase of its nuclear stockpile. The increase of the USA's nuclear weapons

[980] “There is a real potential that AI could become a more prestigious arms race between nation states competing for regional hegemony.” (T339)
“In this upper end [forecast for 2040], we see mention of more exotic scenarios such as human extinction and expansion due to AI fabrication.” (T344)

stockpile relates to the increase in the USA's "adversaries" stockpile, mostly Russia and China." (T344)

- The U.S. Congress pushed back on the Biden administration's plans to cut the U.S. nuclear arsenal. (T337)

Cross-references with other questions

Q31: Nuclear Weapon Use
Q33: Countries with Nuclear Warheads
Q34: Country-by-Country Nuclear Use

Question 33: Countries with Nuclear Warheads

How many countries will be estimated to have at least one nuclear warhead…
...by the end of 2024?
...by the end of 2030?
...by the end of 2050?

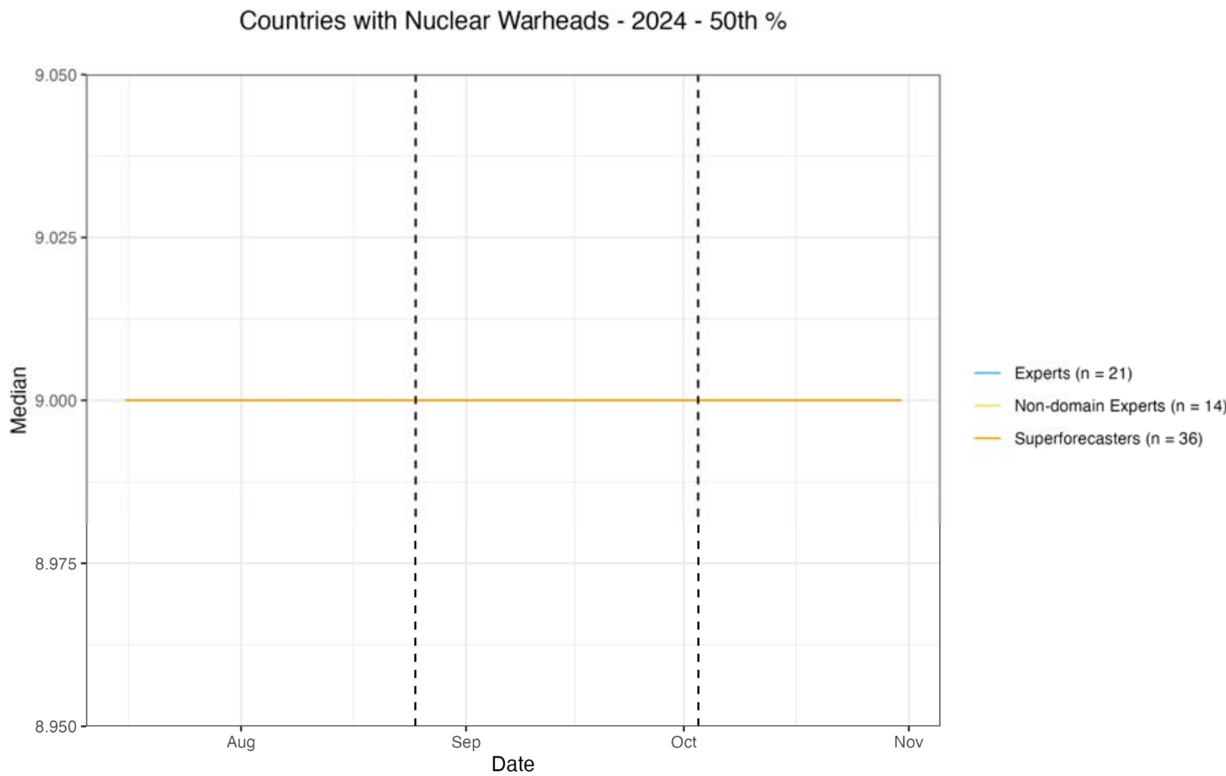
Question and resolution details, prior forecasts, and other relevant sources

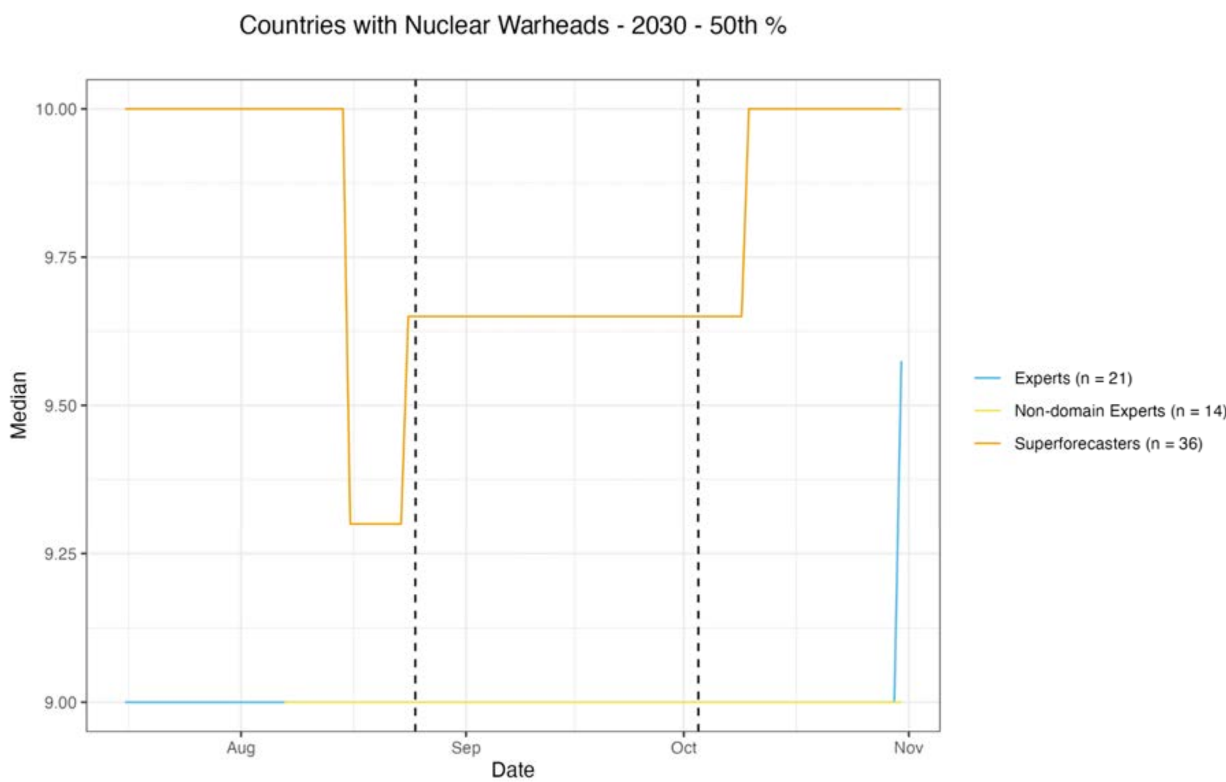
Results[981]

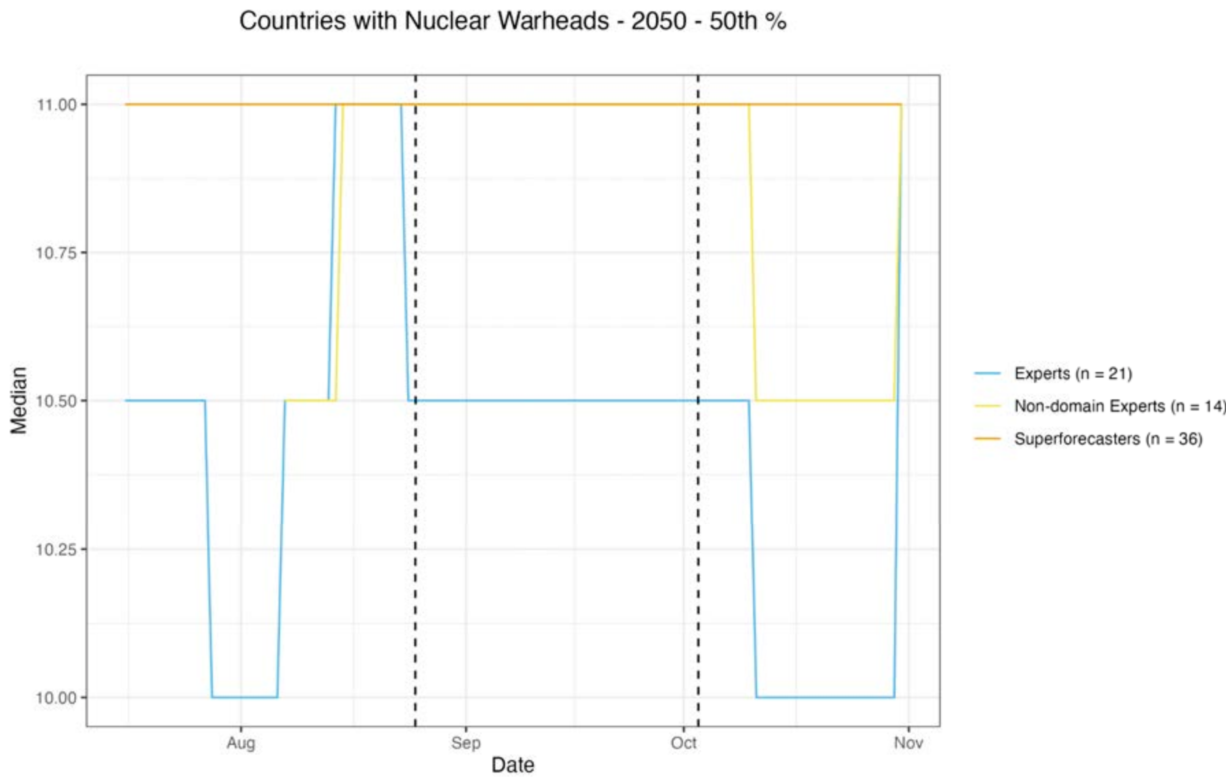
Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 36)	2024	9	9	0.85	-32.23%
	2030	10	10	2.15	-48.17%
	2050	11	11	3.5	-40.08%
	2024	9.5	9	0.71	-18.35%

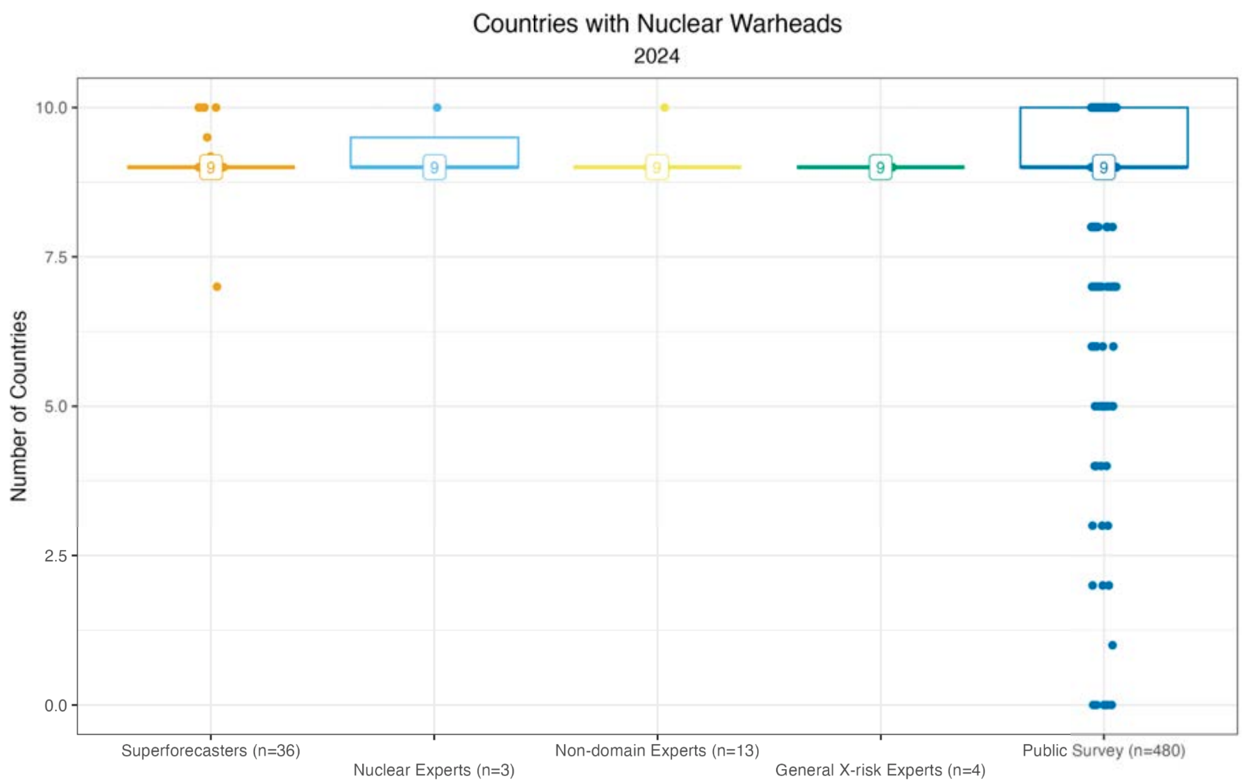
[981] Numbers of forecasters are given as of Stage 4 of the XPT.

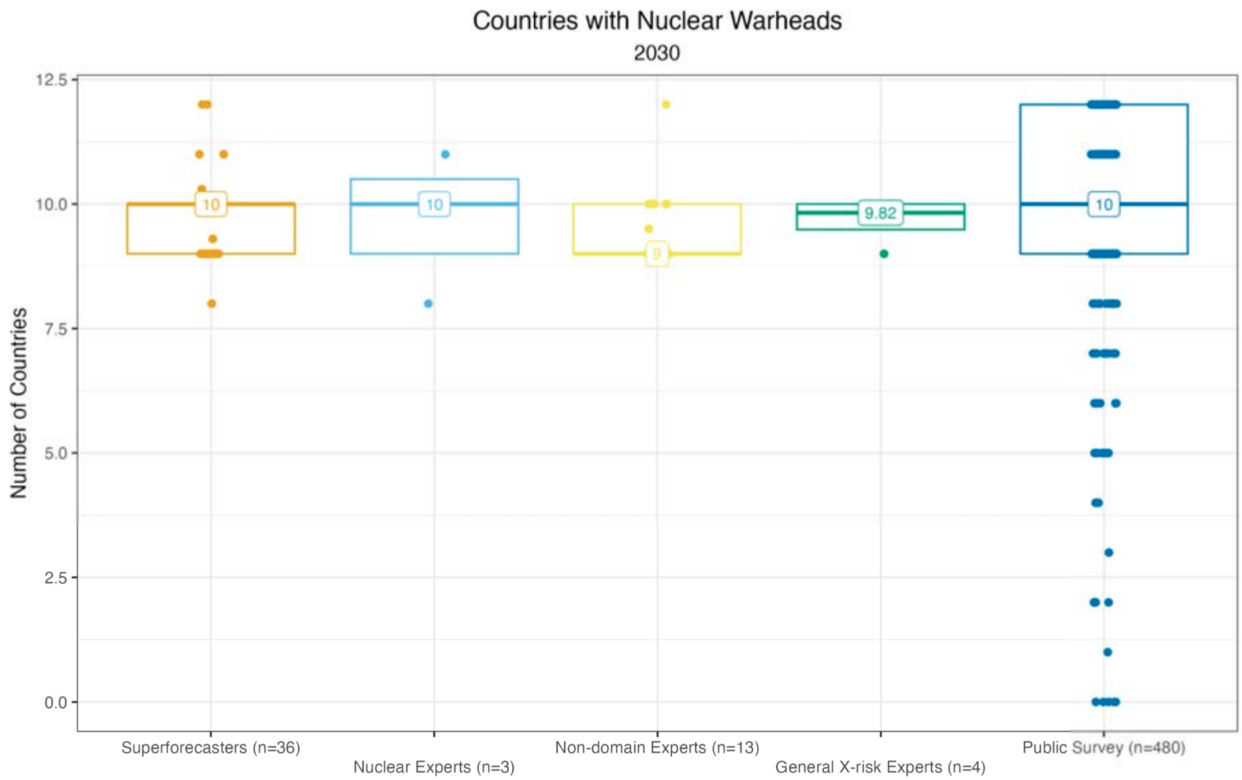
Domain Experts (N = 3)	2030	10.5	10	0.71	+116.02%
	2050	11.5	10	2.12	+41.42%
General X-Risk Experts (N = 4)	2024	9	9	0	NaN
	2030	9.5	9.83	0.71	-33.32%
	2050	11	12	2.83	-55.51%
Non-Domain Experts (N = 13)	2024	9	9	0	Inf
	2030	9	9	1.06	-18.7%
	2050	10.5	11	5.44	-21.58%
Public Survey (N = 480)	2024	9		9.37	-
	2030	10		31.68	-
	2050	12		289.94	-

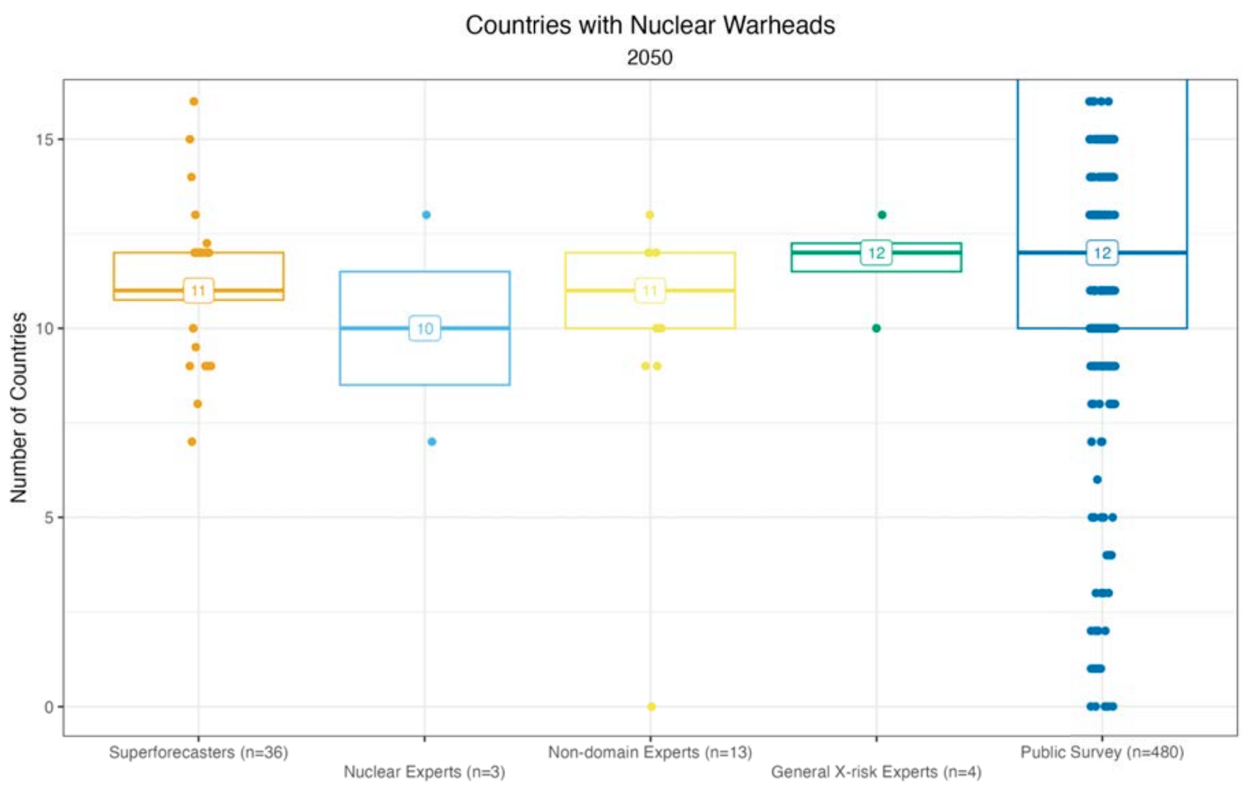

Countries with Nuclear Warheads - 2024 - 50th %
9.050
9.025
9.000
8.975
8.950
Median
Aug
Sep
Oct
Nov
Date
Experts (n = 21)
Non-domain Experts (n = 14)
Superforecasters (n = 36)


Countries with Nuclear Warheads - 2030 - 50th %
10.00
9.75
9.50
9.25
9.00
Median
Aug
Sep
Oct
Nov
Date
Experts (n = 21)
Non-domain Experts (n = 14)
Superforecasters (n = 36)


Countries with Nuclear Warheads - 2050 - 50th %
11.00
10.75
10.50
10.25
10.00
Median
Aug
Sep
Oct
Nov
Date
Experts (n = 21)
Non-domain Experts (n = 14)
Superforecasters (n = 36)


Countries with Nuclear Warheads
2024
10.0
7.5
5.0
2.5
0.0
Number of Countries
9
Superforecasters (n=36)
Nuclear Experts (n=3)
Non-domain Experts (n=13)
General X-risk Experts (n=4)
Public Survey (n=480)


Countries with Nuclear Warheads
2030
Number of Countries
12.5
10.0
7.5
5.0
2.5
0.0
10
10
9
9.82
10
Superforecasters (n=36)
Nuclear Experts (n=3)
Non-domain Experts (n=13)
General X-risk Experts (n=4)
Public Survey (n=480)


Countries with Nuclear Warheads
2050
Number of Countries
15
10
5
0
11
10
11
12
12
Superforecasters (n=36)
Nuclear Experts (n=3)
Non-domain Experts (n=13)
General X-risk Experts (n=4)
Public Survey (n=480)

Sources of agreement, disagreement and uncertainty

- **Clear baseline of nine nuclear powers**
    - Teams agreed on the list of existing nuclear powers provided in the question criteria (Russia, USA, China, France, UK, Pakistan, India, Israel, and North Korea).[982]
        - "The vast majority of forecasters are tightly grouped with a baseline of the 9 existing nuclear powers." (T337)
- **Little short-term change followed by slight increase in nuclear powers**
    - The number of countries with nuclear warheads depends on 1) whether existing nuclear powers will disarm completely, and 2) whether new nations will acquire nuclear warheads.[983] Most forecasters believed that complete disarmament was unlikely, and that any states that did disarm would be counterbalanced by the probability of new nuclear powers.[984] These considerations led to a general consensus that no change is likely by 2024 while a slight increase in the number of nuclear powers is expected by 2050.
    - **Little to no change is expected by 2024**
        - "There is general agreement that the 2024 time window is too short for either disarmament or proliferation." (T344)
        - "There does not appear to be any opportunities in the next 24 months for states that might be interested in developing nuclear weapons." (T340)
    - **Wider range for 2050 estimates**
        - "2050 is far enough into the future [that] all forecasts reflect significant uncertainty…Still, we think the number is more likely to go up than to go down." (T336)
        - "In the short run, there is virtually no amount of disagreement as all the 50% thresholds are at 9 for the forecast year 2024, and only a slight increase in the amount of disagreement for the forecast year 2030. The longer range 2050 year forecast has more median forecast spread as the window for proliferation/disarmament increases, with more weight on the possibility of proliferation." (T344)
    - Most existing nuclear-armed nations are unlikely to completely disarm.[985]

---

[982] T345 listed North Korea as a state that is "interested in acquiring or developing nuclear weapons in the next thirty years." This may be a typo for South Korea.

[983] Reduction of weapons stockpiles would not affect the resolution of this question if a country retains at least one warhead.

[984] "Some team members factored in the possibility that one or two countries could disarm in the next three decades. They, however, resolved that even if one or two countries disarm, there will be others weaponizing, tipping the scale towards more than the current nine." (T340)

[985] "All but one forecaster assumed that the total number of nuclear states would not decline, so the existing number of 9 known states was the bottom range." (T337)
"There is some possibility of disarmament, but most team forecasters seem to consider this fairly unlikely, so the most optimistic scenarios are maintaining the status quo." (T344)
"Possible scenarios for denuclearization were discussed:

1. Voluntarily giving up a nation's complete nuclear arsenal - this was not deemed likely by most forecasters
2. Fundamental change in national character/governance, such as what happened in South Africa. This might be as a result of a nation attempting to use nuclear weapons and the international fallout of those decisions.
3. Actual use of nuclear weapons depleting stockpiles to zero." (T343)

        - “Countries getting off of the nuclear list is unlikely. They would be unwise to trust security guarantees in exchange for de-nuclearization.” (T336)
        - “Seeming near-consensus that the current nuclear countries will not give up nuclear weapons absent some major change in the nature of the polity or being forced into surrendering them in nuclear war.” (T343)
        - “In general, only a few team members give serious consideration to the possibility of impactful disarmament.” (T344)
    - A minority of forecasters assigned some probability to all or most nuclear powers giving up their weapons.
        - “[By 2050,] [t]here was one forecaster who saw the possibility of complete denuclearisation and one who went as low as 3. Most did include the possibility of some denuclearisation with estimates of 7 or 8 total countries. 2 forecasters did not consider any net denuclearisation within the 5% range of possibility.” (T339)
    - Disarmament was viewed as more likely if done multilaterally or as the result of a state’s disintegration. With one exception, forecasters did not name specific countries that they believed would be inclined to disarm.
        - “Possible countries to give up weapons are India and Pakistan as part of a peace agreement (unilateral not likely) and North Korea, should the state fail and China intervenes to take their weapons.” (T339)
        - “Countries might be willing to give up nuclear weapons as a group rather than one at a time. Especially if there is a major nuclear war or nuclear mishap or blackmail before 2050, that may create pressure for denuclearization.” (T336)
- **Is Taiwan a country?**
    - One team assumed that a possible nuclear-armed Taiwan would not qualify as a country for purposes of this question, while others implicitly assumed it would.[986]
- **Disagreement on effect of AI[987]**
    - “[T]here’s disagreement on the role of AI in this time frame (2024-2050). A number of people expect rapid progress and world-changing effects of the development of AGI and other variants (TAI, SAI). Others do not.” (T336)

Arguments given for lower forecasts

- **Difficulty of developing nuclear weapons**
    - Nuclear weapons are hard to develop, both because they require significant technical skill and secret information, and because developing nuclear weapons

---

“From a technological perspective, the development of nuclear weapons is an extremely complex and expensive process. However, it would be just as difficult, if not more so, to decide to abandon nuclear weapons. Especially since the example of Ukraine shows that with the disposal of a nuclear arsenal, the security of smaller and medium-sized countries diminishes, regardless of international promises and agreements.” (T345)

[986] “One forecaster placed the possibility at 11 with Iran, Saudi Arabia, Japan, South Korea, Germany, Taiwan possibly acquiring.” (T339)
“Assumptions: Taiwan does not and will not count as an independent country.” (T341)

[987] “Wild card scenarios…Redefinition of the term state with super-intelligence AI as a "sovereign individual or collective" with nuclear capabilities, the proliferation of said "states." (T344)

causes diplomatic friction.[988] One team stated that a certain level of GDP and resources would be necessary to have a nuclear program.[989]

- "Acquiring nuclear weapons comes with high political, financial and economic costs and risks a pre-emptive war by other countries." (T336)
- "It is difficult to develop and build nuclear weapons. Without an advanced nuclear power industry, it is virtually impossible. It is particularly hard to develop nuclear weapons without being noticed by one's neighbors and the major powers, which then leads to many notable consequences; most significantly resulting in tensions with a given country's neighbors." (T337)

- On the other hand, some countries already have the required resources and technical capability.
    - "There are several countries with the technical knowledge and expertise to build a bomb in a couple of years or less…Historically the first successful tests of countries have been surprises and the pre-test estimates of capacity have been inaccurate." (T336)
    - "Given the required economic and social conditions, Brazil is an outside possibility because of the country's GDP." (T341)

- **Deterrence could be achieved with other weapons**
    - New weapons or cyber attack techniques could give a nation equal bargaining power with lower cost and fewer diplomatic problems.
        - "[S]everal new types of weapons of mass destruction are in development (for instance bioweapons and nanotech), which may allow states with less resources to obtain the benefits of such weapons at a lower cost and perhaps less likelihood of being detected in developing them." (T337)
        - "[F]rom a practical standpoint, cyber and other non-traditional forms of sabotage could be more useful to a country than nuclear weapons." (T339)
- **North Korea might cease to exist**
    - A country might lose its nuclear weapons because of ceasing to exist as a nation (assuming the nuclear weapons were destroyed or found their way to an existing nuclear power). Forecasters specifically noted the possibility of North Korea ceasing to exist after reunification of Korea.[990]
        - "It is entirely possible for certain states to cease to be, like North Korea, and for its nuclear stockpile to disappear." (T344)

Arguments given for higher forecasts

- **Nuclear weapons are effective deterrents**
    - Nuclear weapons continue to be effective deterrents.**[991]** The perceived value of a nuclear deterrent would increase in the event of nuclear war or a perception that

---

[988] "From a technological perspective, the development of nuclear weapons is an extremely complex and expensive process." (T345)
[989] See graph under "The relationship between GDP and resources to have a nuclear program." (T341)
[990] "Possible countries to give up weapons are India and Pakistan as part of a peace agreement (unilateral not likely) and North Korea, should the state fail and China intervenes to take their weapons." (T339)
[991] "The rationale for these estimates is the team members' belief that the current geopolitical situation indicates that (1) nuclear weapons can be an effective deterrence mechanism." (T345)

existing nuclear powers such as the US can no longer be relied on to protect smaller states.

- "The reality on the ground is nuclear weapons still [have] a lot of military value." (T340)
- "The rationale for these estimates is the team members' belief that the current geopolitical situation indicates that (1) nuclear weapons can be an effective deterrence mechanism…Nuclear weapon use would increase the number of countries with nuclear weapons as a protection from the nuclear threat." (T340)

- **New nations might become nuclear powers**
  - Forecasters cited a total of 21 nations that might be incentivized and/or able to pursue nuclear weapons. In many cases, they believed a nation's decision to pursue nuclear weapons would be contingent on geopolitical considerations.
    - "The most significant dissentions (though not all that extreme) were how many countries would even attempt to obtain nuclear weapons and why." (T337)
    - "Nuclear proliferation has [ ] strong cascading effects. The more countries with nukes, the more countries will follow." (T336)
  - **Iran and Saudi Arabia**
    - Iran and Saudi Arabia were the countries most likely to be named as potential new nuclear powers.[992] Forecasters believed that Saudi Arabia would seek nuclear weapons if Iran did so first.
      - "The upper end of the forecasting range was probably best supported by the probability that Iran will move forward to obtain nuclear weapons, which could easily lead to a knock-on proliferation among countries in the region, especially Saudi Arabia." (T337)
      - "Iran looks like it's on [its] way already, and Saudi Arabia publicly pledged to weaponize the next day Iran does, so that's two." (T340)
  - **Japan, South Korea, and Taiwan**

---

[992] "It is unlikely that Iran will get the bomb by the end of 2024 but it would not be a black swan event…If Iran succeeds in testing a nuclear bomb, Saudi Arabia is likely to quickly follow through the purchase of warheads or an accelerated program helped by Pakistan, or NK (needs cash) and even possibly Israel - as a counter to Iran." (T336)
"[In 2024] a small chance exists that Iran acquires a nuclear weapon. Salvaging the nuclear deal is not looking likely but 2024 does not give Iran much time. There will be tremendous pressure on Iran not to go nuclear and international backlash if they do…[by 2030,] [t]he nine current countries with a much higher chance Iran has gone nuclear…2030 Higher-end Forecast: 13. Saudi Arabia acquires in response to Iran…" (T339)
"The team named four candidate countries as the most likely to weaponize (not in any order): Iran, Japan, South Korea, and Saudi Arabia." (T340)
Team 341 mentioned Saudi Arabia as a "potential joiner" without providing details and, uniquely in this group of forecasters, without linking it to Iran.
"Potential newcomers to the scene, using conventional technologies are Iran, Ukraine, and Saudi Arabia." (T344)
"The rationale for these estimates is the team members' belief that the current geopolitical situation indicates that…some states (Syria, North Korea, Iran) are interested in acquiring or developing nuclear weapons in the next thirty years." (T345)

- ■ Three teams each pointed to Japan, South Korea, and Taiwan as potential new nuclear powers,[993] citing the possibility of a "forthcoming Asian nuclear race." (T341) Forecasters believed that these nations' decisions to nuclearize would be responsive to their relationships with neighbors, particularly with China and the US.
  - "If [the China-Taiwan situation] gets hot--similar to what is happening with Russia-Ukraine now, we are likely to see Japan and South Korea putting their guards up. Both countries have the technical capability to weaponize should they decide to." (T340)
  - "It is unlikely that Taiwan, Japan, and South Korea will seek nuclear weapons by 2030, but that is largely dependent on the policies of China, North Korea and the US." (T336)
- ○ Germany and Australia were mentioned as possible nuclear powers by two teams, who connected this possibility to Chinese aggression and the decline of US security guarantees.[994]
  - ■ "Should the U.S. turn more inward, Germany might acquire to defend itself and Europe." (T340)
- ○ Other nations that one or two teams mentioned as nuclear possibilities were Brazil,[995] Venezuela,[996] Libya,[997] Iraq,[998] the UAE,[999] Turkey,[1000] Egypt,[1001]

---

[993] "Possible and most likely candidate countries between 2030-2050 (in addition to those mentioned above) include Brazil, Japan, Australia, Nigeria, Venezuela, DRC." (T336)
"[By 2024,] [o]ne forecaster placed the possibility at 11 with Iran, Saudi Arabia, Japan, South Korea, Germany, Taiwan possibly acquiring). 2030 Higher-end Forecast: 13. Saudi Arabia acquires in response to Iran, and Taiwan in response to aggression by China (and a more remote possibility of Quad members Japan and/Australia acquiring for similar reasons). Possibilities of South Korea and Germany as mentioned in the 2024 estimates above…[by 2050] [f]orecasters were more confident that South Korea, Taiwan, and Japan would acquire weapons compared to the 2030 forecasts." (T339)
"The team named four candidate countries as the most likely to weaponize (not in any order): Iran, Japan, South Korea, and Saudi Arabia…"If [the China-Taiwan situation] gets hot--similar to what is happening with Russia-Ukraine now, we are likely to see Japan and South Korea putting their guards up. Both countries have the technical capability to weaponize should they decide to." (T340)
"What about new countries…Taiwan builds with US support." (T341)
[994] "Possible and most likely candidate countries between 2030-2050 (in addition to those mentioned above) include Brazil, Japan, Australia… In case of the disappearance of US security guarantees, countries like Poland, Germany and other -smaller- European nations might also be tempted to obtain nuclear weapons." (T336)
"2030 Higher-end Forecast: 13. Saudi Arabia acquires in response to Iran, and Taiwan in response to aggression by China (and a more remote possibility of Quad members Japan and/Australia acquiring for similar reasons)." (T339)
[995] "Possible and most likely candidate countries between 2030-2050 (in addition to those mentioned above) include Brazil, Japan, Australia, Nigeria, Venezuela, DRC." (T336)
"The case for Brazil. They historically had a program. Was a military dictatorship. Possess nuclear facilities for electricity. Won't risk international sanctions so a submarine program keeps the country on the technology path." (T341)
[996] "Possible and most likely candidate countries between 2030-2050 (in addition to those mentioned above) include Brazil, Japan, Australia, Nigeria, Venezuela, DRC." (T336)
[997] "Libya, Iraq and South Africa have had nuke programs." (T341)
[998] See footnote 999
[999] "By 2030…[t]he UAE and Turkey, and perhaps Egypt, would also become candidates." (T336)
[1000] See footnote 1001
[1001] See footnote 1001

Syria,[1002] South Africa,[1003] Nigeria,[1004] the DRC,[1005] Poland,[1006] Ukraine,[1007] and Scotland (if separated from the UK).[1008]
    - Large existing nuclear powers could fragment into multiple smaller nuclear powers.[1009]
- **Developing nuclear weapons could become easier**
    - Two teams mentioned the possibility that developing nuclear weapons could become dramatically easier.[1010]
        - "Wildcard[ ] Scenarios: Building a nuclear program becomes an off-the-shelf technological process which then increases the ease of access. Collectively we did not think this was a likely scenario and it did not particularly influence our forecasts.'' (T341)

Other arguments given

Arguments in favor of lower forecasts:

- One way for a country to exit the nuclear club would be for it to use all its nuclear warheads in war and then not construct more.[1011] Judging from other questions in this tournament (such as Q31), forecasters would view this as a low-probability event.

Cross-references with other questions

Q31: Nuclear Weapon Use

Q32: Total Nuclear Warheads

Q34: Country-by-Country Nuclear Use

[1002] "The rationale for these estimates is the team members' belief that the current geopolitical situation indicates that…some states (Syria, North Korea, Iran) are interested in acquiring or developing nuclear weapons in the next thirty years." (T345)

[1003] "Libya, Iraq and South Africa have had nuke programs." (T341)

[1004] "Possible and most likely candidate countries between 2030-2050 (in addition to those mentioned above) include Brazil, Japan, Australia, Nigeria, Venezuela, DRC." (T336)

[1005] See footnote 1006

[1006] "In case of the disappearance of US security guarantees, countries like Poland, Germany and other -smaller- European nations might also be tempted to obtain nuclear weapons." (T336)

[1007] "Potential newcomers to the scene, using conventional technologies are Iran, Ukraine, and Saudi Arabia." (T344)

[1008] "What about new countries - e.g. Scotland splits from the UK and becomes a nuclear state?" (T341)

[1009] "Fragmentation of states into multiple nuclear-bearing states :

- Russia due to external pressure caused by sanctions brought upon by the 2022 Ukrainian invasion
- China fragmentation due to the size and historical presence
- USA fragmentation due to potential civil war
- India's fragmentation due to the sheer size and cultural diversity of the country." (T344)

[1010] "An extreme drop in the price of nuclear development and ensuing proliferation" (T344)

[1011] "Possible scenarios for denuclearization were discussed…Actual use of nuclear weapons depleting stockpiles to zero." (T343)

Question 34: Country-by-Country Nuclear Use

What is the probability that each actor in the list below will be the first to use a nuclear weapon on the territory or against the military forces of (A) a nuclear-armed adversary or (B) a treaty ally of a nuclear-armed adversary by 2030?

- China
- France
- India
- Israel
- North Korea
- Pakistan
- Russia
- The United Kingdom
- The United States
- Other actor (state)
- Other actor (non-state)
- This will not occur

Question and resolution details, prior forecasts, and other relevant sources

Results[1012]

Group	Percentile Forecast	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 33)	China	0.1	0.15	0.37	-22.42%
	France	0.01	0.001	0.23	-25.26%
	India	0.12	0.2	0.52	-50.1%
	Israel	0.1	0.1	1.22	-56.63%

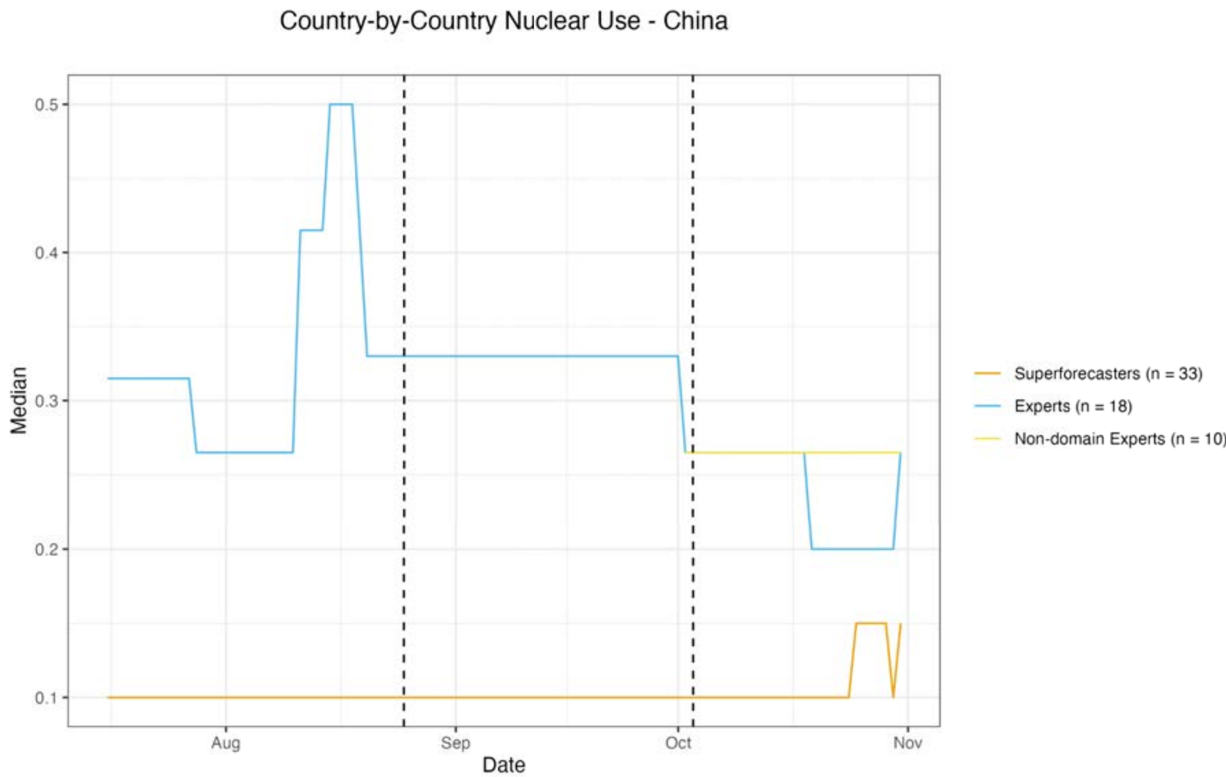
[1012] Numbers of forecasters are given as of Stage 4 of the XPT.

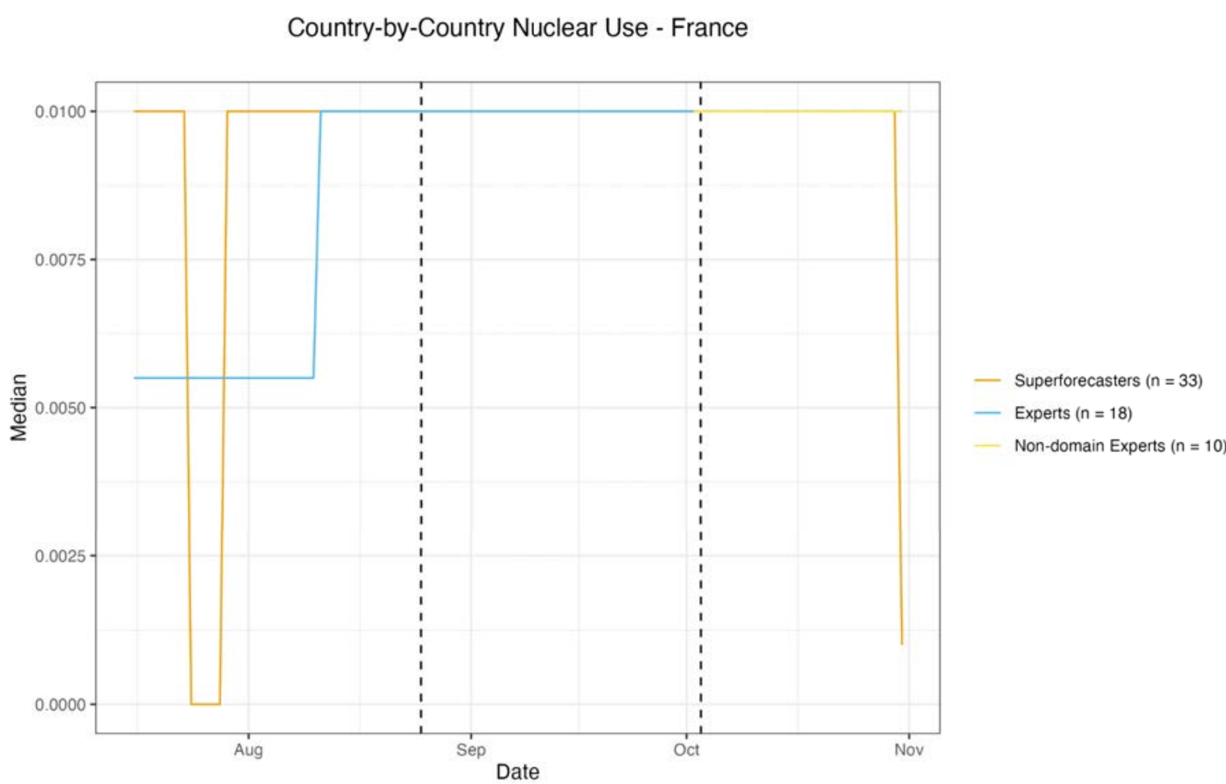
	North Korea	0.3	0.35	0.98	-32.17%
	Pakistan	0.12	0.2	0.54	-55.12%
	Russia	0.74	1	2.47	-19.47%
	The United Kingdom	0.02	0.01	0.24	-26.7%
	The United States	0.14	0.1	0.81	-35.36%
	Other actor (state)	0.06	0.05	0.52	-16.73%
	Other actor (non-state)	0.001	0.02	0.14	+183.96%
	This will not occur	97.5	97.3	7.25	-29.9%
Domain Experts (N = 5)	China	1	0.2	0.76	-32.48%
	France	0.01	0.02	0.05	-9.52%
	India	2	0.9	1.5	-19%
	Israel	0.04	0.04	0.56	-9.26%
	North Korea	3.755	0.89	5.3	-55.89%
	Pakistan	1.52	0.6	2.09	-55.42%
	Russia	2.72	1.87	3.22	-34.35%
	The United Kingdom	0.01	0.01	0.05	-25.73%
	The United States	0.1	0.1	0.06	+163.32%
	Other actor (state)	0.01	0.01	5.77	-22.71%

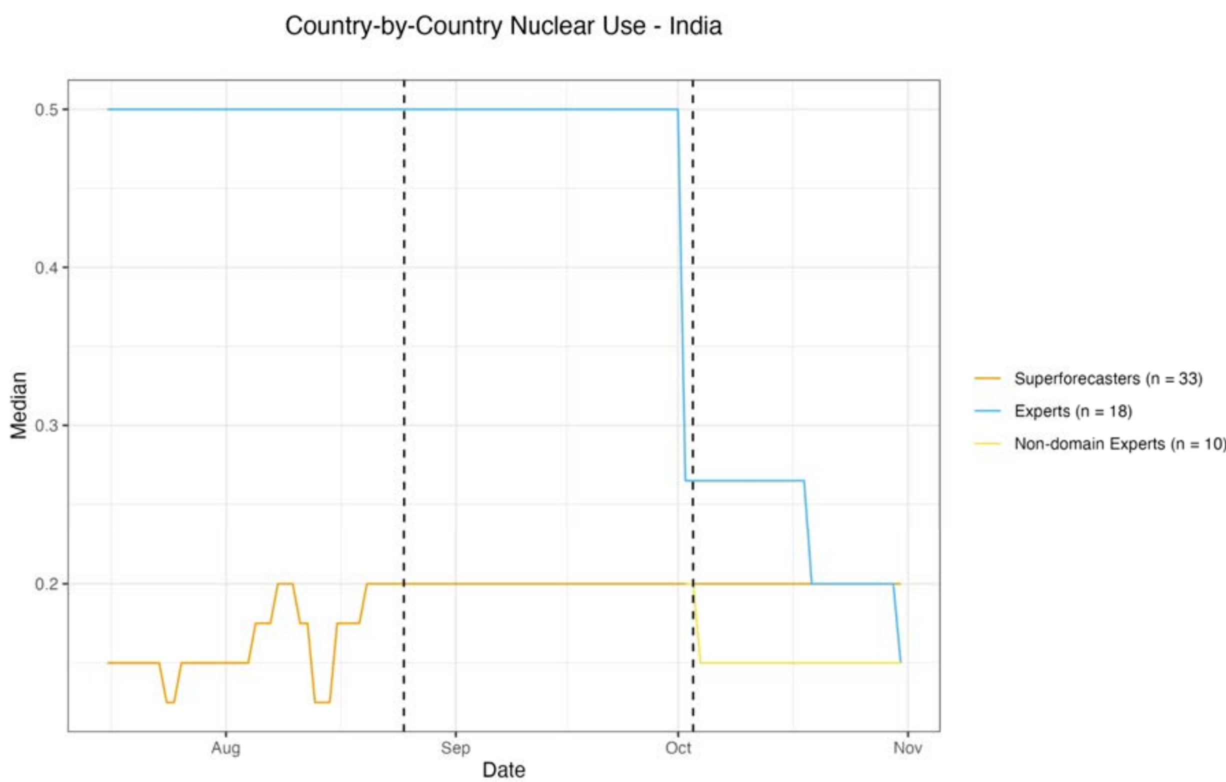
	Other actor (non-state)	0.005	0.000005	0.007	-29.31%
	This will not occur	80.95	93.39	23.46	-18.21%
General X-Risk Experts (N = 3)	China	2	0.78	NA[1013]	NA
	France	0	0	NA	NA
	India	1.5	0.06	NA	NA
	Israel	0.5	0.04	NA	NA
	North Korea	0.1	0.2	NA	NA
	Pakistan	1.5	0.09	NA	NA
	Russia	7	3.8	NA	NA
	The United Kingdom	0.1	0.1	NA	NA
	The United States	0.2	0.6	NA	NA
	Other actor (state)	0	0	NA	NA
	Other actor (non-state)	0.5	0.01	NA	NA
	This will not occur	86.6	95.2	NA	NA
Non-Domain Experts (N = 10)	China	0.2	0.26	0.2	+1250.47%
	France	0.006	0.01	0.2	-22.07%
	India	0.26	0.15	0.29	-14.1%

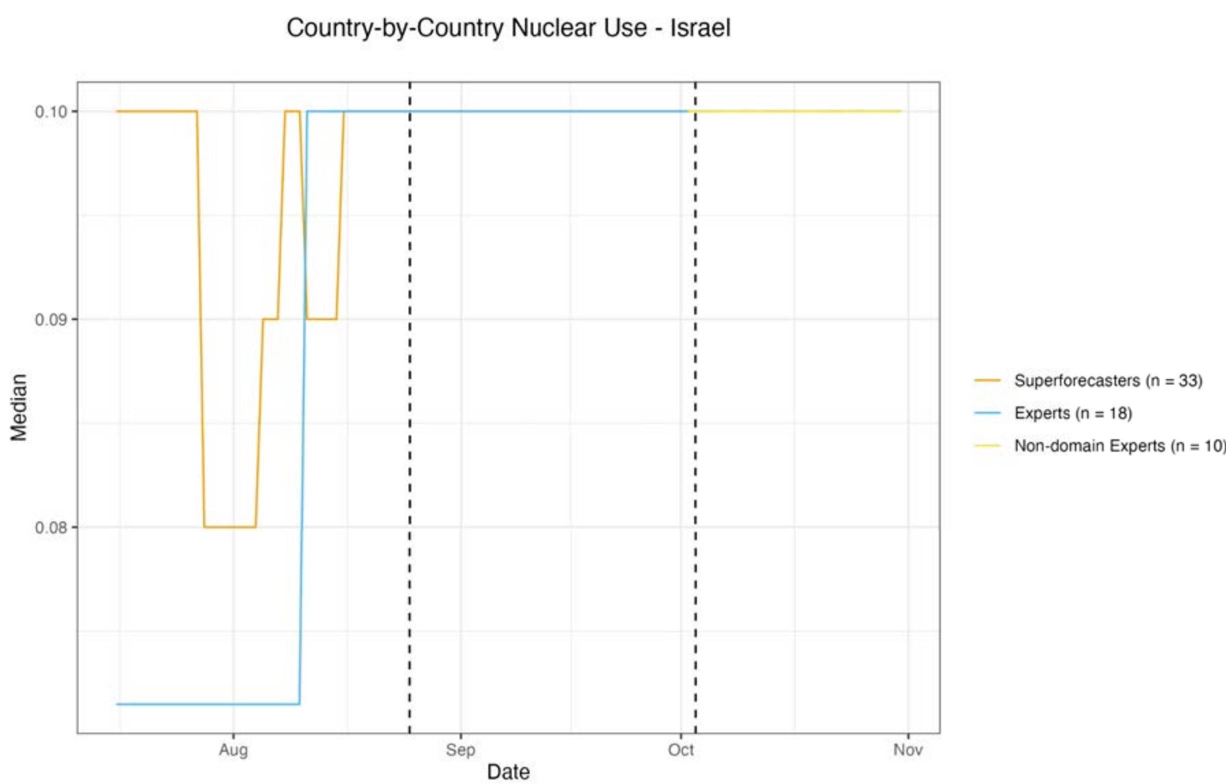
[1013] Only one forecaster in this group answered this question in Stage 1.

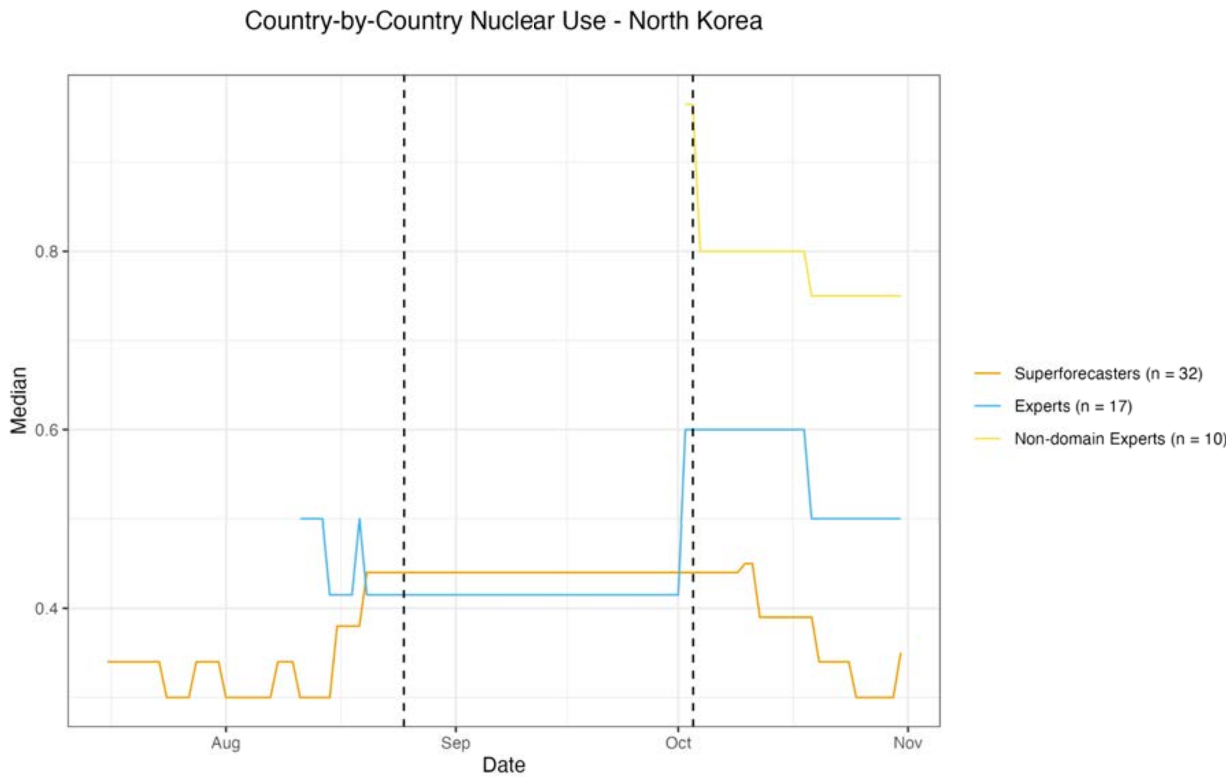
	Israel	0.06	0.1	0.19	+63.4%
	North Korea	0.3	0.75	9.77	-86.06%
	Pakistan	0.26	0.25	6.78	-86.5%
	Russia	0.8	2.12	5.05	-36.23%
	The United Kingdom	0.01	0.01	0.2	-22.01%
	The United States	0.08	0.76	2.81	-47.37%
	Other actor (state)	0.09	0.1	2.81	-66.84%
	Other actor (non-state)	0.002	0.07	2.06	-53.86%
	This will not occur	97.96	93.86	29.55	-70.65%

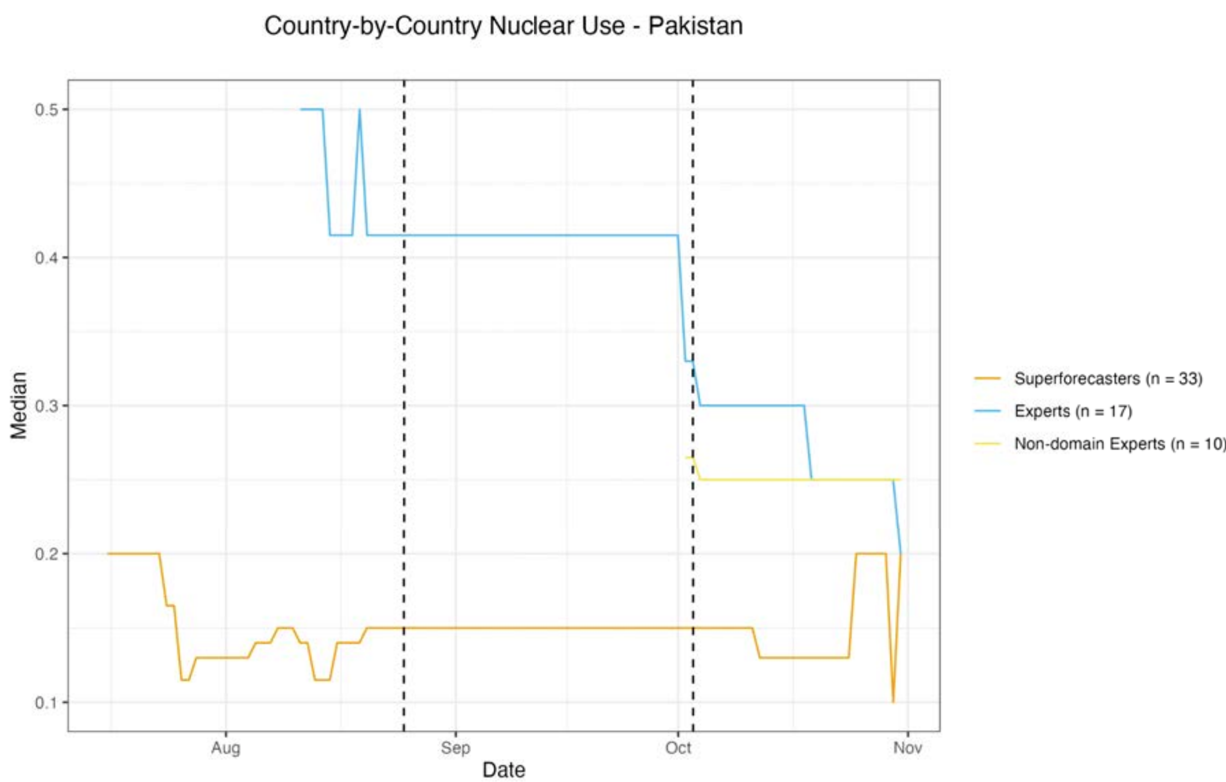

Country-by-Country Nuclear Use - China
0.5
0.4
0.3
0.2
0.1
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 33)
Experts (n = 18)
Non-domain Experts (n = 10)

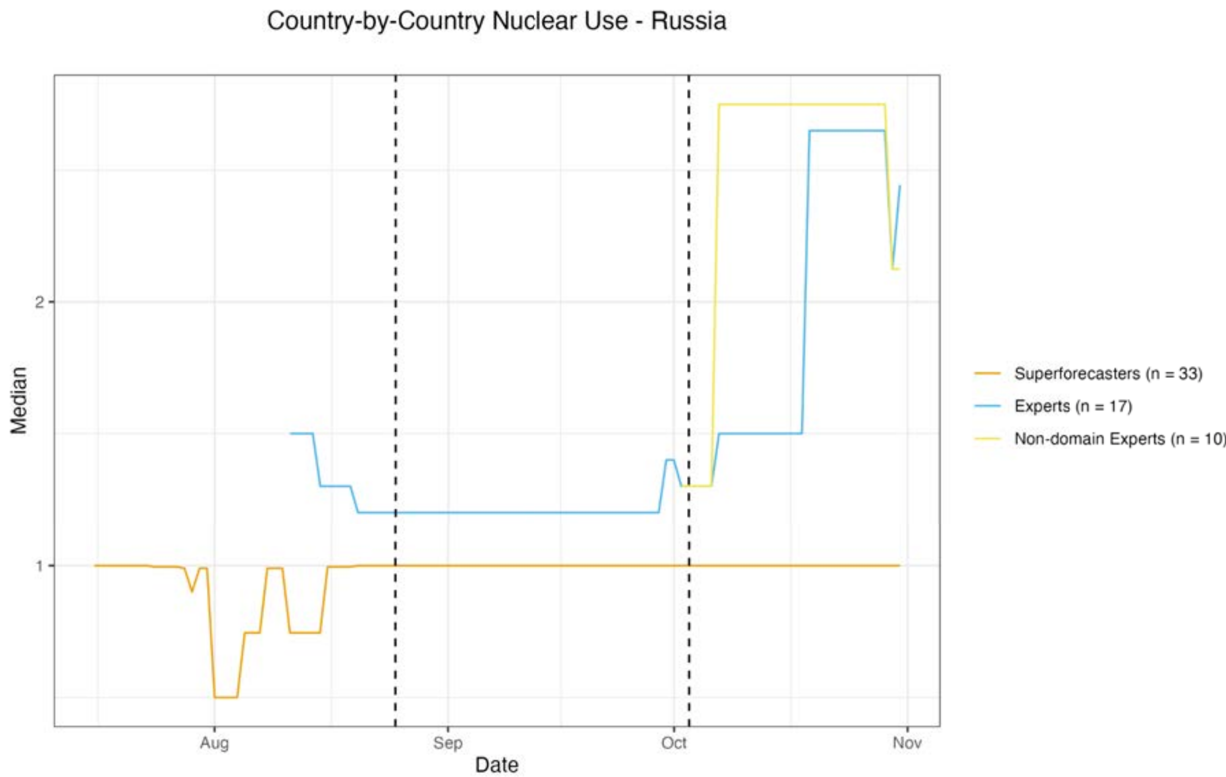

Country-by-Country Nuclear Use - France
0.0100
0.0075
0.0050
0.0025
0.0000
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 33)
Experts (n = 18)
Non-domain Experts (n = 10)

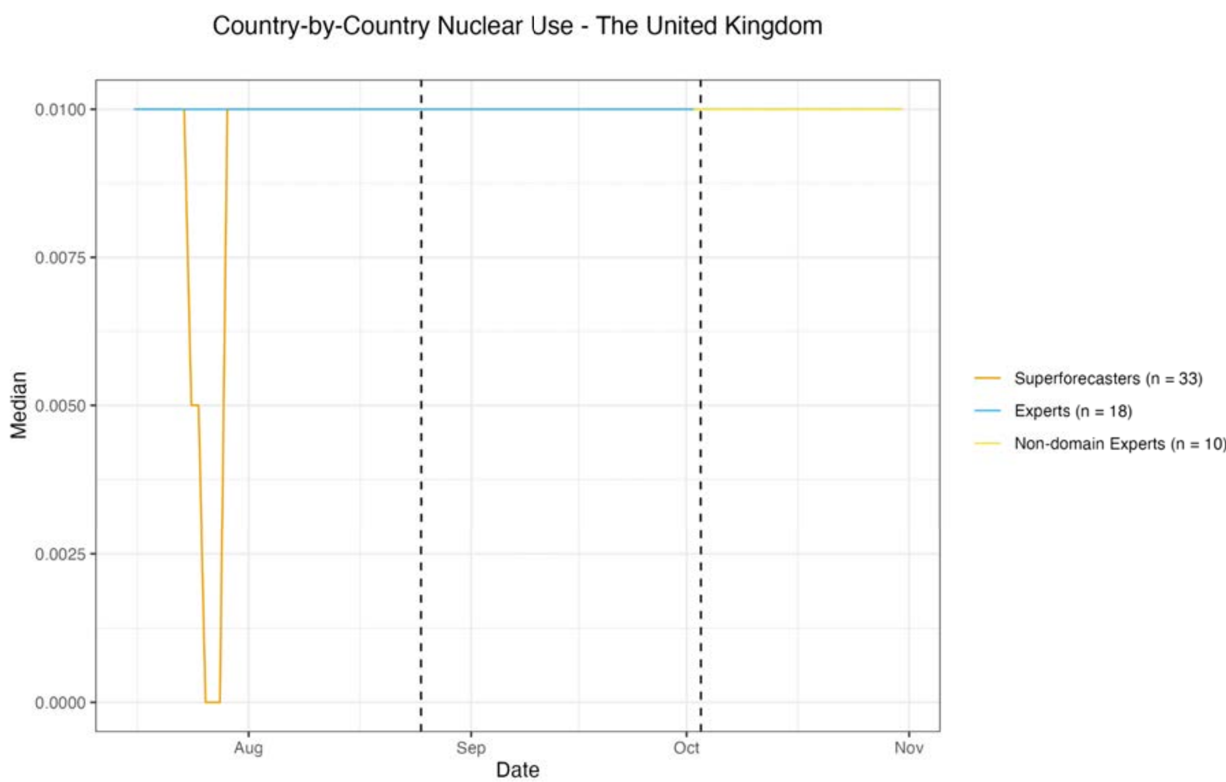

Country-by-Country Nuclear Use - India
0.5
0.4
0.3
0.2
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 33)
Experts (n = 18)
Non-domain Experts (n = 10)

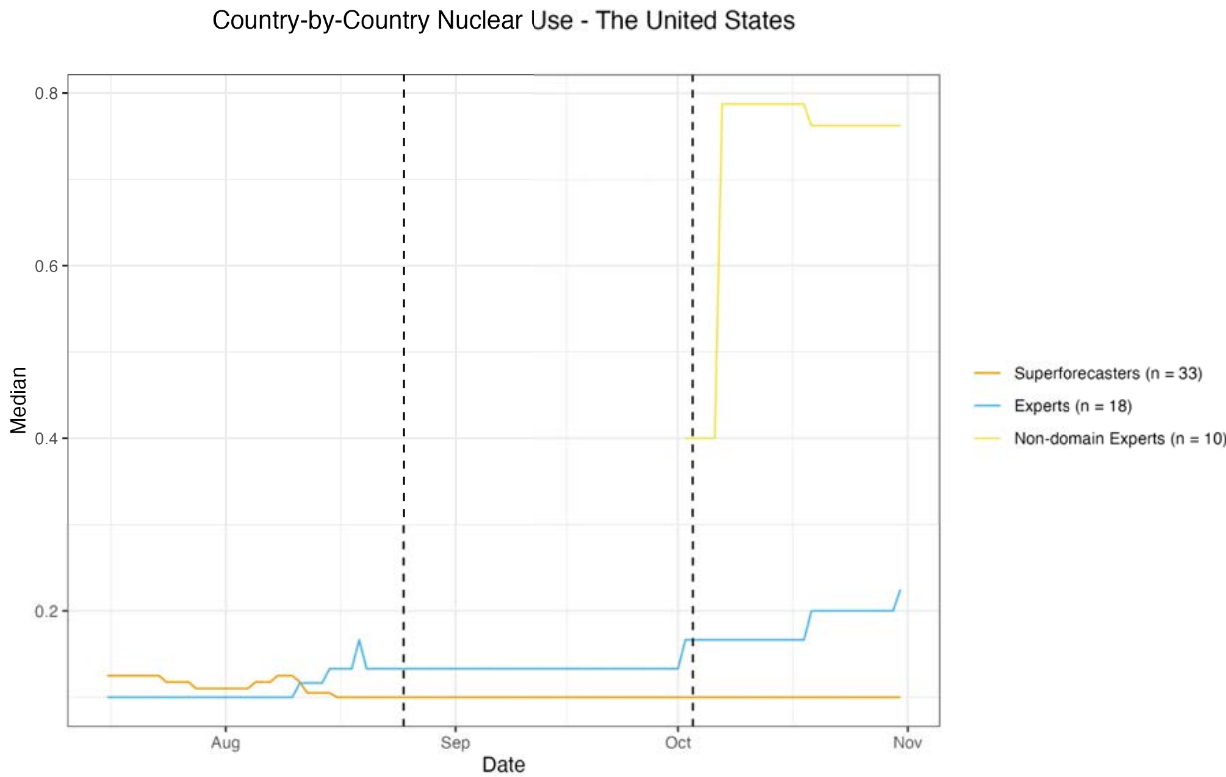

Country-by-Country Nuclear Use - Israel
0.10
0.09
0.08
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 33)
Experts (n = 18)
Non-domain Experts (n = 10)

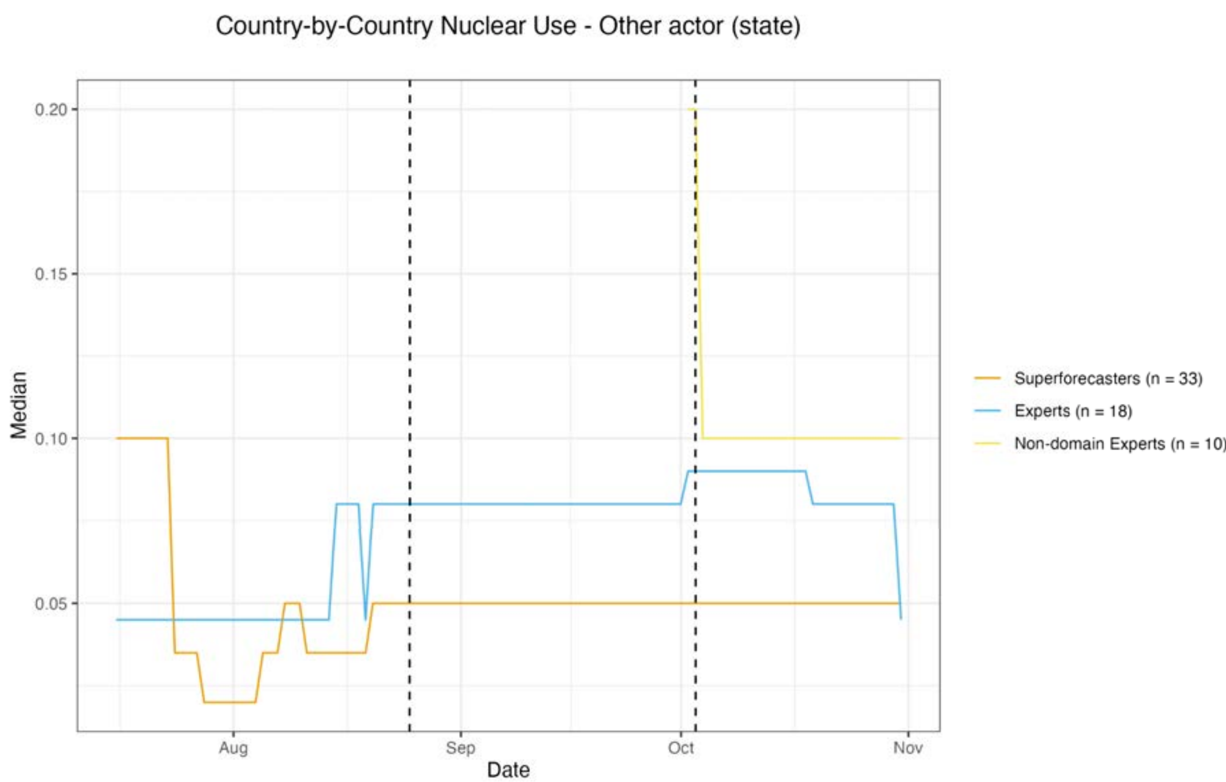

Country-by-Country Nuclear Use - North Korea
Median
0.8
0.6
0.4
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 32)
Experts (n = 17)
Non-domain Experts (n = 10)

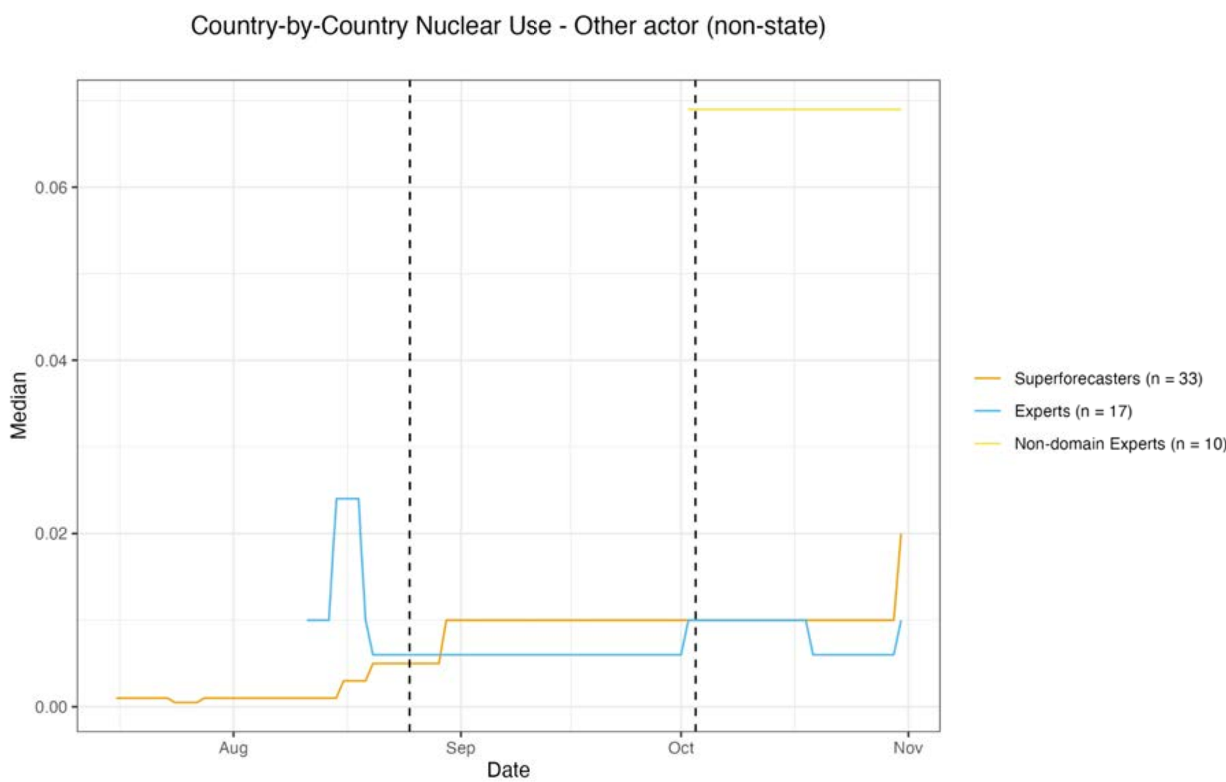

Country-by-Country Nuclear Use - Pakistan
Median
0.5
0.4
0.3
0.2
0.1
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 33)
Experts (n = 17)
Non-domain Experts (n = 10)

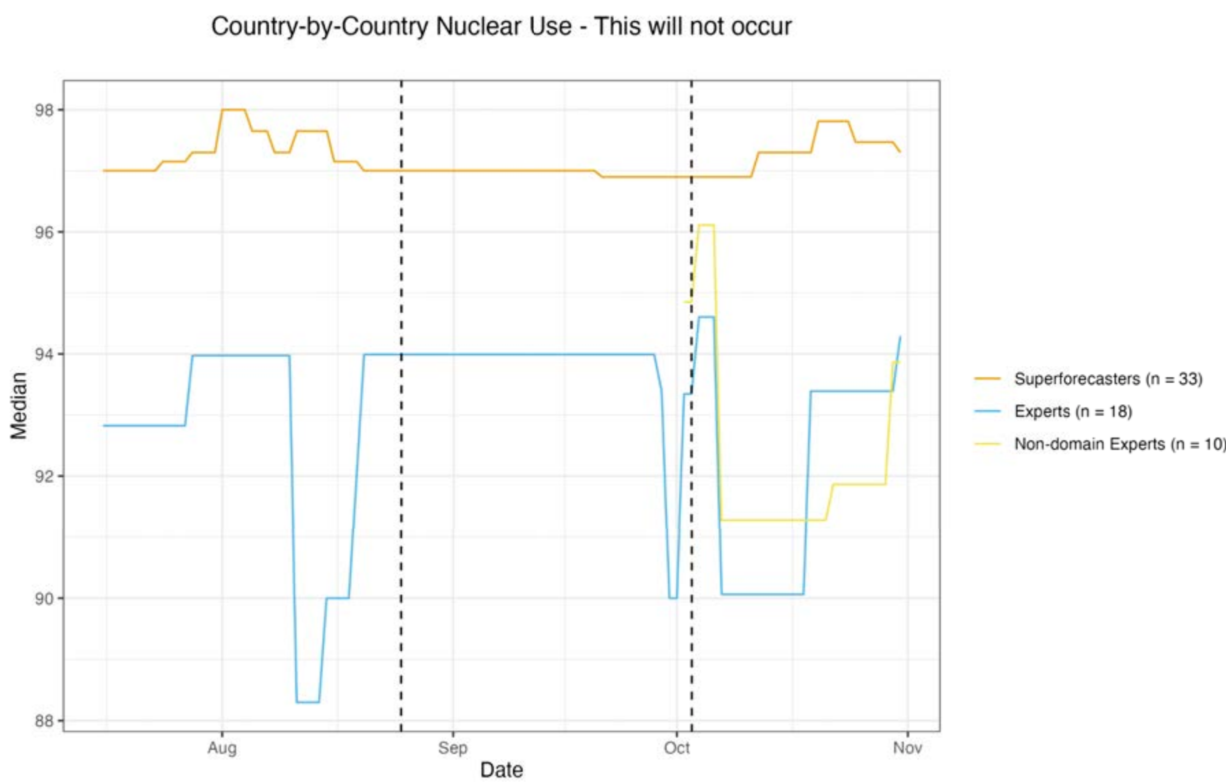

Country-by-Country Nuclear Use - Russia
Median
2
1
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 33)
Experts (n = 17)
Non-domain Experts (n = 10)

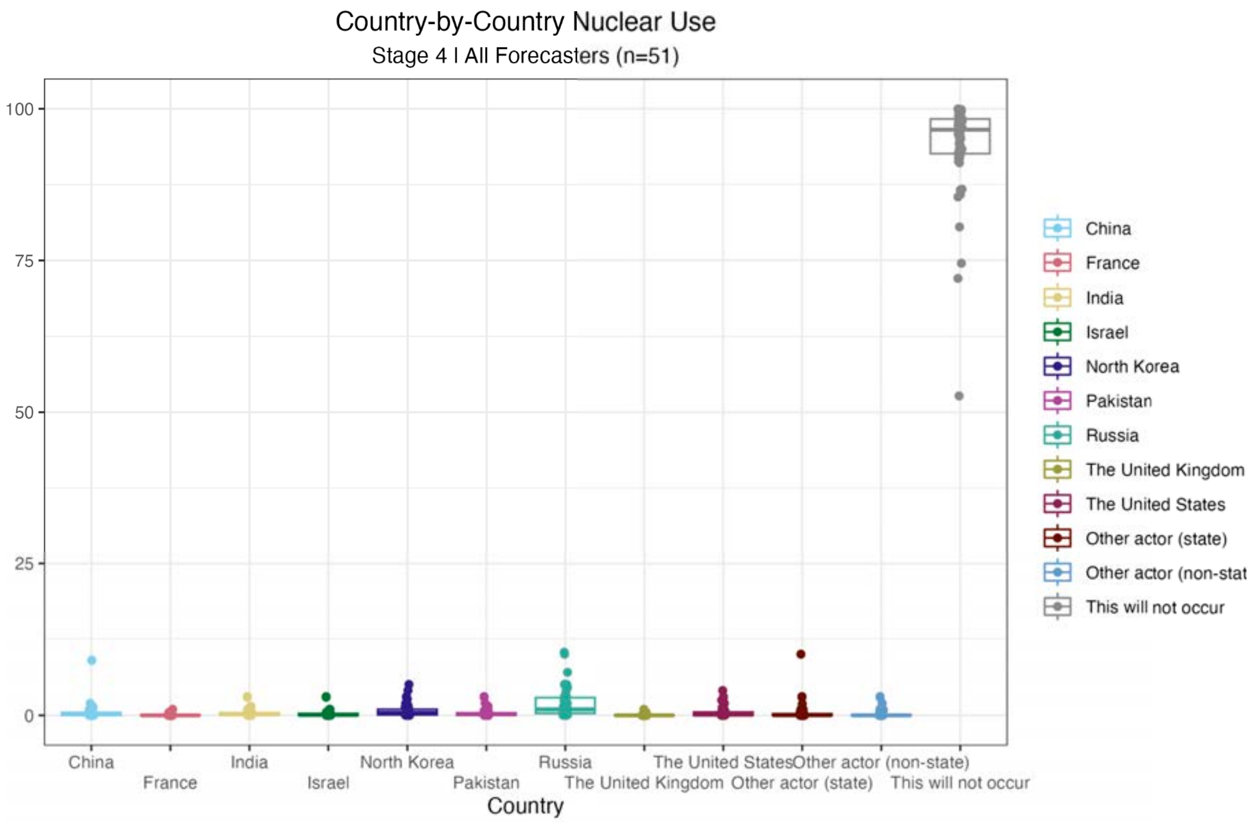

Country-by-Country Nuclear Use - The United Kingdom
Median
0.0100
0.0075
0.0050
0.0025
0.0000
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 33)
Experts (n = 18)
Non-domain Experts (n = 10)


Country-by-Country Nuclear Use - The United States
0.8
0.6
0.4
0.2
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 33)
Experts (n = 18)
Non-domain Experts (n = 10)


Country-by-Country Nuclear Use - Other actor (state)
0.20
0.15
0.10
0.05
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 33)
Experts (n = 18)
Non-domain Experts (n = 10)


Country-by-Country Nuclear Use - Other actor (non-state)
Median
0.06
0.04
0.02
0.00
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 33)
Experts (n = 17)
Non-domain Experts (n = 10)


Country-by-Country Nuclear Use - This will not occur
Median
98
96
94
92
90
88
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 33)
Experts (n = 18)
Non-domain Experts (n = 10)



Sources of agreement, disagreement and uncertainty

- **Nuclear first use is unlikely**
  - All teams agreed that the scenario described in the question (nuclear first use[1014]) was unlikely to occur. Most discussion focused on comparing relative probabilities for the relevant state and non-state actors.
    - "Nuclear retaliation would be extremely likely, and there are no winners in a nuclear war…proliferation is easier to prevent than with biological or chemical weapons of mass destruction, and some very serious players are very motivated to do so." (T336)
    - "[N]o country wants to risk becoming a pariah nation." (T340)
    - "Use of nuclear weapons has been incredibly rare, despite decades of proliferation. This is true despite nuclear powers engaging in active (or proxy) conflicts with one another." (T344)
    - "Because the inherent outcomes following a nuclear weapons discharge are so hard to predict, but may be dire, this lowers the potential for such events to occur. Additionally, the 8 year time horizon for this question is not very long. Most geopolitical dynamics existing [now] are likely to exist

[1014] This summary will use the term "nuclear first use" as a shorthand way to describe the scenario in which an actor is the first to use a nuclear weapon on the territory or against the military forces of a nuclear-armed adversary or a treaty ally of a nuclear-armed adversary by 2030.

throughout the time period, aiding the possibility of event prediction." (T337)
        - Nuclear proliferation is easier to prevent than with biological or chemical weapons of mass destruction, and some very serious players are very motivated to do so." (T336)

- **Countries with other options less likely to use nuclear weapons**
    - A common theme was that a nation that could avoid using nuclear weapons would do so. For example, forecasters commented that China and the US would be less likely to use nuclear weapons first because of their large conventional armies; terrorists could use other types of bombs; and France and the UK would be less likely to act first because of their NATO relationship with the US.
        - "Why would a country use a nuclear weapon when conventional weapons are currently very good at what they do? [O]nly a misguided state or Non-state would consider using a nuclear warhead in battle; it is viewed primarily as a deterrent to attack by a foe." (T339)
        - "NATO countries do not have a reason to be the first to launch [due] to superiority of USA conventional army." (T341)
        - "If you are a low end military country, it seems like there would be more utility in possessing a nuclear weapon as a deterrent or for negotiating leverage than to actually launching it." (T342)
    - While generally agreeing that rational actors will avoid using nuclear weapons when possible, forecasters cautioned that a nuclear-armed actor might act irrationally or perceive threats where none exist. This concern was raised with regard to Russian and North Korean leadership and the possibility of religious fundamentalists taking control of the Pakistani military.
        - "In my opinion, nuclear weapon states will only use their weapons in response to an existential crisis…[but] [t]he existence of an "existential threat" is only in the eyes of the threatened, no logic is required. It's not your definition of "existential threat" but the leadership of a country." (T339)
        - "The DPRK, Pakistan, Russia, and a future nuclear-Iran are the states that could be reasonably backed into [a] corner where they find a nuclear strike to be reasonable." (T341)
- **Agreement about most and least likely actors**
    - Teams agreed that Russia and North Korea were the most or among the most likely actors to use nuclear weapons first, and France and the UK were the least or among the least likely. More disagreement existed about the ranking of other possible actors. Notably, Israel, Pakistan, the US, and a non-state actor/terrorist each appeared both in the top three "most likely" slots for some teams and the "least likely" slots for others.
        - "Almost all forecasters considered North Korea and Russia to be the greatest 'threats', albeit at low percentage numbers." (T338)

        - "There was strong agreement around key actors most likely to discharge a weapon with Russia and North Korea commentary dominating discussion space." (T337)
- **Effect of large arsenals unclear**
    - Forecasters did not agree about whether countries with larger nuclear weapon arsenals would be more or less likely to use these weapons first.
        - "I'll also assign slightly higher percentages to the U.S. and Russia in part for their vastly larger nuclear arsenals." (T338)
        - "Russia - US/UK/France: This relationship has the greatest physical distances and the best sensing, also the very large arsenals mean that each side is deterred regardless of any missile defense." (T339)

Country-by-country arguments

- **Russia**
    - Every team named Russia as their #1 candidate for nuclear first use, except for one team that ranked Russia as #2 behind North Korea.
        - "I think this question is the only one throughout the tournament where every teammate is on the same page. Russia is expected to be the most likely to cause a nuclear attack, by every teammate." (T345)
        - "The team sees Russia as the most likely/least unlikely candidate." (T336)
        - "If [this scenario] did happen, Russia and the US, the nations that are already the most antagonistic and likely to start wars, would likely be the culprit, Russia most of all." (T337)
        - "The most likely states to use nuclear weapons would be the US or Russia and they would most likely be tactical nuclear weapons at a very limited scale." (T341)
    - **War in Ukraine increases risk of Russian or US first use**
        - The war in Ukraine increased the probability assigned to both Russia and the US. Although Russian use of nuclear weapons to attack Ukraine would not qualify to resolve this scenario because Ukraine is not a nuclear-armed nation or treaty ally of a nuclear-armed nation,[1015] forecasters viewed this scenario as carrying a high risk of nuclear escalation.
            - "The war in Ukraine is likely the largest single driver of risk for this question." (T337)
            - "The war in Ukraine makes Russia and the USA the co-favorites. Either side would resort to tactical nuclear weapons if it or its proxy were in danger of losing the war…Putin has threatened to use nuclear weapons if its sovereign territory were violated,

---

[1015] These predictions would surely change if Ukraine becomes a NATO member.
"[T]he most risk comes from war in Ukraine. However this question asks about a nuclear power or its official ally will be attacked so if Russia uses nuclear weapons on Ukraine this question will resolve as "This will not occur." (T342)

including former Ukrainian territory now annexed (Crimea) or planning to be annexed (Donbas) by Russia." (T340)
            - "Nuclear escalation could proceed after a possible nuclear attack on Ukraine with nuclear detonation on the uninhabited part of Russia by either USA or the UK (polar circle)." (T344)
            - "[A Russian presidential decree from 2017] confirms the existence of the concept of nuclear de-escalation, which implies that, in the event that the armed forces face the threat of defeat during a conventional conflict, they can use tactical nuclear weapons in uninhabited territory or at sea to demoralize the adversary." (T344)
            - "[M]ultiple Russian actions have demonstrated low value of Ukrainian life…which could be a potentially concerning precondition for use of WMDs against the populace. Any use of nuclear weapons significantly increases the probability of further use of nuclear weapons against a nuclear-armed nation…Given Russia's history [classifying Ukrainians as Nazis] amounts to defining the conflict in existential terms." (T343)
        - **Russia may act irrationally**
            - "Russia does not seem entirely a rational actor right now, reflecting the higher event risks of authoritarian regimes and dictators to use nuclear weapons to save themselves." (T337)
            - "Russia [ ] says it felt threatened by Ukraine enough to invade it, who would have guessed it was such a threat? The Russian leadership may feel an "existential threat" if Donbas or Crimea are lost to Ukraine, then what?" Often in nondemocracies there is a lot of "I am the State" thinking. So it becomes easy to argue that an existential crisis to the state could also come in the form of an existential crisis to me, the dictator?" (T339)
- **North Korea**
    - The value of nuclear weapons to North Korea is their deterrent value or use as a bargaining chip. Forecasters generally agreed, therefore, that nuclear first use by North Korea would be suicidal.
        - "Little nations like North Korea have nukes purely as deterrence. To use them would be suicide, but autocracies are also known for their bad decision making skills." (T337)
        - "North Korea has nuclear weapons as a bargaining chip, only, to extract concessions from the USA and to act independently from China." (T340)
        - "Kim is a dictator, but not a madman. He wants to stay in power (his family has ruled the country for decades with an iron fist, but without suicidal tendencies) and would certainly lose everything, if he used nukes against South Korea or the US. He would have his artillery destroy Seoul as a lower-level option (which still would mean the end of his reign)." (T336)

        - "I rate this slightly below India-Pakistan because there aren't two nuclear powers, so at least South Korea isn't explicitly trying to protect a second strike capability." (T339)
    - North Korea nevertheless ranked high in probability because of a sense that Kim Jong Un may act irrationally or unpredictably and the country lacks institutional safeguards that could mitigate a leader's rash decisions.
        - "North Korea-United States remains a significant source of tension, and North Korea now has a delivery vehicle making a possible strike on the US (Alaska or mainland) more likely. North Korea also is not entirely a rational actor who also lacks clear institutional barriers to using a nuclear weapon…Not impossible to see some situation where Kim Jong Un gets sick and dies, is replaced by some unknown zealot, and the apparatus of the state is so highly centralized and demanding of unthinking obedience that nukes are launched in response to some incident before cooler heads can prevail." (T337)
        - "North Korea is just a constant question because of who is running it and which side of the bed he gets out of each day." (T339)
        - "North Korea - Assigning 0.5%. Power is consolidated in the hands of a dictator, so he could pull a nut move. Doesn't seem to help his cause i.e., staying in power." (T342)
- **China**
    - **China has other options and a No First Use policy**
        - "China has a huge army and wouldn't need nukes for a first strike." (T336)
        - "It's unclear what the US or China gain in going nuclear over Taiwan, they can both accomplish all of their military goals without nuclear missiles, including attacks on carrier strike groups." (T339)
        - "China and India have No First Use (NFU) policies in place. These appear to be pledges, and it is not clear what kind of legal or procedural barriers to first use have been implemented in either country. Regardless, [one forecaster] assigned them probability 0% in the next 7.5 years." (T342)
        - "[China and the US] have a lot to lose if they are first. Neither have mad-men at the helm. The big question is Taiwan and China's plans for the South China Sea dominance." (T339)
    - **Taiwan is most likely cause of Chinese first use**
        - A conflict over Taiwan that could escalate to include the US or other nations is the most likely cause of Chinese use of nuclear weapons, with some forecasters also mentioning India-China conflict.
            - "China has precommitted to recapture Taiwan, which is a conflict that could expand to include nuclear-armed states." (T343)
            - "Possible adversaries, that might qualify as "yes": India and something having to do with Taiwan. China could potentially try to use a nuke against a US carrier battle group if the US comes to the aid of Taiwan during a forced reunification, but China would [know] this would be catastrophic for both sides." (T336)

        - “And then you have the nuclear weapons proliferation paradox....meaning if Taiwan started a nuclear weapons program today to deter China, China would bomb Taiwan tomorrow.” (T339)

- **United States**
    - **American intervention in global affairs increases risk**
        - The United States’ relationships around the world increase the risk of American first use. Forecasters noted the possibility that the US might use nuclear weapons first either to protect a NATO ally, to intervene in a conflict with a non-NATO state (such as Ukraine or Taiwan), or to preemptively check aggression by North Korea or Iran.
            - “The USA has a number of precommitments around the globe. These come in the form of both formal treaty commitments and informal relationships. NATO, S. Korea, Japan, Taiwan, and Israel are potential conflict areas for the USA. Additionally, the probability of the USA using nuclear weapons proactively against another nuclear armed force is higher, considering the USA's distance from on-the-ground conflict zones.”[1016] (T343)
            - “For me the [second most likely country to use after Russia is] USA, since "pre-emptive strikes" are historically approved. The USA has used a lot of intelligence in the UE/RU conflict to their advantage, and may find out something that may trigger a first move at a low level…US executes a preemptive first strike on North Korea or an ascendant nuclear power like Iran based on intelligence: 1/1000.” (T339)
            - “[The US might] react to biological or chemical threats (to themselves or a NATO ally), but it's highly unlikely to act with this higher step of escalation then. Possible scenario still remains someone like Trump (who seems to have considered nukes a nice solution to some problems) in the White House again. The question would be, if the political/military system could prevent it. Potential but unlikely - US retaliation if Russia uses a nuke as a last resort to stop a humiliating defeat in their invasion of Ukraine.” (T336)
            - “US: too many other options, 2nd use is also possible but unlikely.” (T338)
- **India and Pakistan**

[1016] “Thus, the sequence of events that could lead to resolution of the prompt might be: 1.) conventional conflict with a non-nuclear belligerent (not including a direct role for the USA), 2.) conflict escalation to nuclear weapon use by a nuclear-armed state against a non-nuclear (non-NATO) state, 3.) USA retaliation with nuclear weapons against the nuclear-armed state that first used nuclear weapons.” (T343)

- Forecasters agreed that an India/Pakistan conflict risks nuclear war. Some rated the risk of Pakistani first use slightly higher than Indian first use, citing potential instability in Pakistani government control of nuclear weapons.[1017]
  - "There is always the potential for war among these two countries. A major shooting war could quickly escalate into a nuclear exchange." (T340)
  - "A very concerning dynamic, the warning times are incredibly short so for most weapons the assumption is 'use it or lose it'. Sensing also seems to be an issue with frequent aircraft incursions and a recent incident where India tested a missile into Pakistan - if this had happened during a time of high tension we might have been resolving the nuclear catastrophe question early." (T339)
  - "Risk exists for India-Pakistan because they share a long border and have an extensive history of territorial conflict and religious strife. Both states are nuclear armed and at times engage in shooting wars in border regions." (T337)
  - "Pakistan: Possible trouble with India or political instability / religious fanaticism. If Pakistan was being overwhelmed by a land invasion from a vastly numerically superior India, a nuclear detonation may be considered necessary and reasonable as a last resort." (T336)
- **Concern for stability of Pakistani government control of nuclear weapons**
  - "There is serious concern about Pakistan's nuclear weapons in the event of a coup and break down of military control. If this occurred and rogue elements could eventually use a nuclear weapon(s), though it would be more powerful as a blackmail tool." (T336)
  - "Pakistan in particular is not exactly what I'd consider a stable country, being a constant source of weapons and Taliban fighters into Afghanistan…Let's not forget also that in 2006 terrorists from Pakistan coordinated a series of bombings in Mumbai. Is it so inconceivable that a group might be able to obtain a bomb from corrupt ideological allies within the Pakistan Armed Forces, then use it to attack India?" (T337)
- **Risk of India-China conflict**
  - "China - India: Very low risk relationship, their respective arsenals aren't postured for each other and conflicts tend to be isolated to border regions." (T339)
  - "[As with India-Pakistan,] similar territorial issues exist for India-China, although the territorial conflicts are a bit different in practice." (T337)

- **Non-state actors**
  - Forecasters disagreed about the risk of nuclear first use by a terrorist or other non-state actor. While some forecasters ranked this possibility relatively highly, others argued that non-state actors had easier options to achieve their goals and

[1017] "China and India have No First Use (NFU) policies in place. These appear to be pledges, and it is not clear what kind of legal or procedural barriers to first use have been implemented in either country." (T342)

that it would be difficult for a terrorist group to acquire a nuclear weapon and the capability to use it.
        - "Terrorist organizations do not have the intellectual capital or infrastructure to build a bomb. There are easier/cheaper ways to cause death and destruction." (T340)
        - "[One forecaster] discount[ed] the possibility that a non-state actor will both acquire a nuclear warhead, attain launch capability, and then launch it. It seems like a dirty bomb (or even a regular bomb) would be just as effective to accomplish terror or achieve a negotiating position, and it would be far easier to attain. (T342)
        - "I think the risk of nuclear terrorism exists but is around 1/10000, due to the difficulty in securing a device." (T338)
- **Other state actors**
    - Iran was named as the most likely nation to use nuclear weapons first by 2030, excluding existing nuclear powers.[1018] However, forecasters viewed this as unlikely given the short timeline in which a nation would need to develop and use nuclear weapons.
        - "The primary "Other actor (state)" would be Iran, and it seems like there [are] international efforts to sabotage their nuclear weapons research. It's not clear if they are developing launch capability at the same time." (T342)
        - "Nuclear threshold states (Japan, South Korea, Iran) will not have weapons that are operational by 2030." (T340)
        - "Needing [a new state actor] to develop and use [nuclear weapons] for a first strike by 2030 according to resolution criteria is an extremely high bar. Most thinkable scenario here: Iran against Israel." (T336)
        - "[G]iven the prompt resolves by 2030, any nation not already in possession of nuclear weapons is significantly less likely to use such a weapon. This also applies to non-state actors. Very few instances of nuclear weapons being 'lost' by one nation and acquired by another (with the exception of the breakup of the USSR in the 1980's) are known to the forecasters. The only other method of acquiring nuclear weapons is a long and expensive weapons development program." (T343)
    - Forecasters believed that if Iran used nuclear weapons, Israel would be the likely target, and vice versa.
- **Israel**
    - **Israel may choose to use nuclear weapons if threatened by Iran**
        - Although generally ranking the risk of Israeli first use low, forecasters believed that a scenario in which Israel uses nuclear weapons in

---

[1018] "It's extremely difficult to acquire nuclear weapons…Iran has been trying it for years, but they were thrown back many times. In the case of Iran, Israel, and on a global scale the US, do practically anything they can to prevent more actors from getting their hands on nukes." (T336)
"Leaving a .5 in other actor state for the potential of Iran or Saudi Arabia reaching nuclear status and using it." (T341)

response to a threat from Iran is possible.[1019] On the other hand, others noted that Israel could use conventional weapons or rely on American action in such a scenario.

- "Israel may perceive a threat from Iran's nuclear program." (T340)
- "I assume there is a risk of Israel using a nuclear weapon to strike Iran if there is a programme there completed by 2030, and there is also the possibility that Iran strikes [Israel] (the Other actor (state) risk). However, this strike is most likely to be conventional and if nuclear carried out before Iran has a weapon, not meeting the requirements above." (T338)
- "So little is known about the Israeli program that it is difficult to understand their threshold for using nuclear weapons but given that no regional adversary is nuclear, it would probably be a weapon of last resort and that conflict has become less likely. There has been some discussion about using nuclear weapons to destroy Iran's nuclear program, everything I have seen has [led] me to conclude that Iran's mountains are too large and tunnels too deep for this attack to be feasible." (T339)
- "The USA will take the lead in using nuclear weapons in case NATO goes to war or in case of war with Iran, thus precluding the need for the UK, France and Israel to go first." (T340)

- **France and the UK**
  - There was consensus that France and the UK are the least or among the least likely countries to use nuclear weapons first. Forecasters argued that the US would be better positioned to retaliate in the event of a threat to France or the UK. Most forecasters assigned similar probabilities to both countries, while a minority believed that the UK was slightly more likely than France to use nuclear weapons.
    - "France and UK, unlikely to initiate a nuclear conflict…and unlikely to be the first country to respond, unless directly attacked. For example, if a NATO ally were attacked with nuclear weapons, the US would be more likely to deploy the strike in response than either of these countries." (T338)
    - "Going by countries, NATO countries do not have a reason to be the first to launch [due] to superiority of USA conventional army...[one forecaster] thinks No First Use is the heuristic at hand even though that isn't a committed strategy." (T342)
    - "It's hard to imagine, why they should be the first to strike. Any scenario involving them would most probably be a NATO alliance case with the US better prepared to act first." (T336)

[1019] "[Israel's] three (?) 'stealth' submarines out there would certainly be used if "necessary". Thinkable scenario to resolve as "yes": Iran having developed nukes before 2030 and seriously threatening to use them. Would be suicidal." (T336)
"Israel's retaliation against Iran (due to multiple scenarios). (T344)

Cross-references with other questions

Q5: Nuclear Catastrophic Risk
Q6: Nuclear Existential Risk
Q31: Nuclear Weapon Use
Q32: Total Nuclear Warheads
Q33: Countries with Nuclear Warheads

Question 35: GPT Revenue

Will Robin Hanson win a bet that the GPT line of language models will generate less than $1 billion in customer revenue in total by the beginning of 2025?

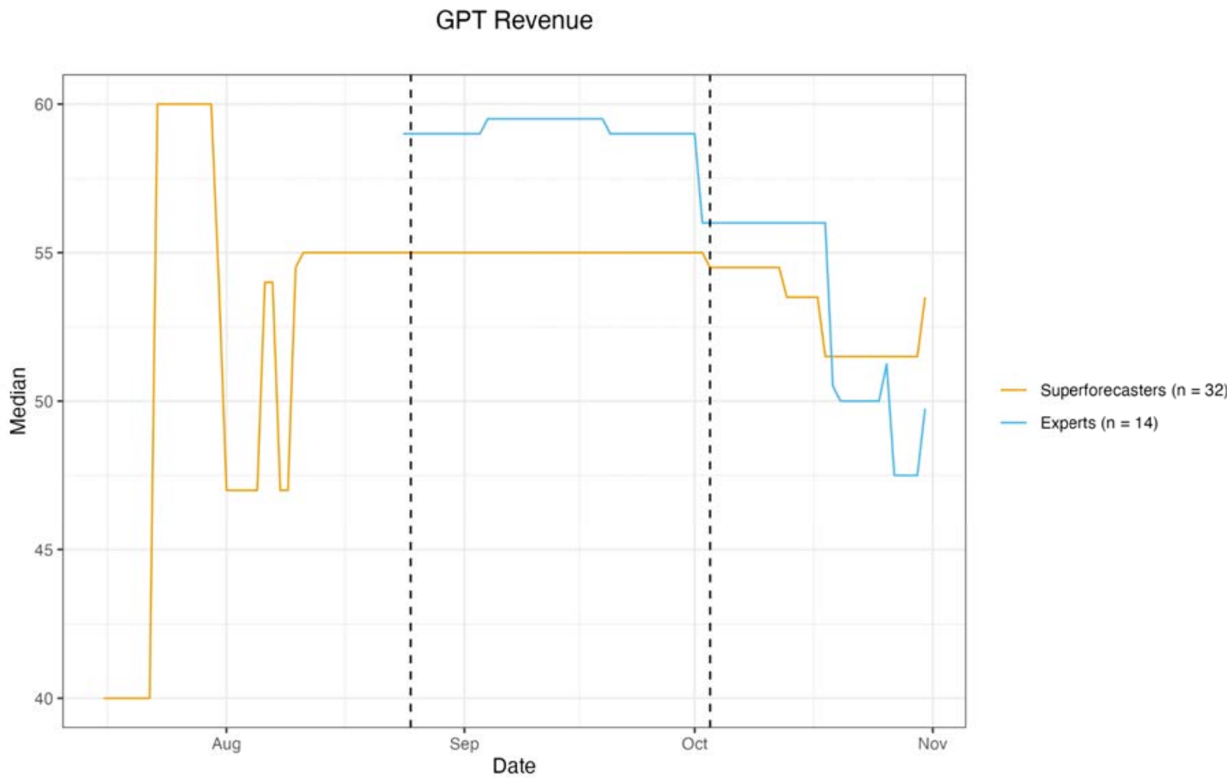
Question and resolution details, prior forecasts, and other relevant sources

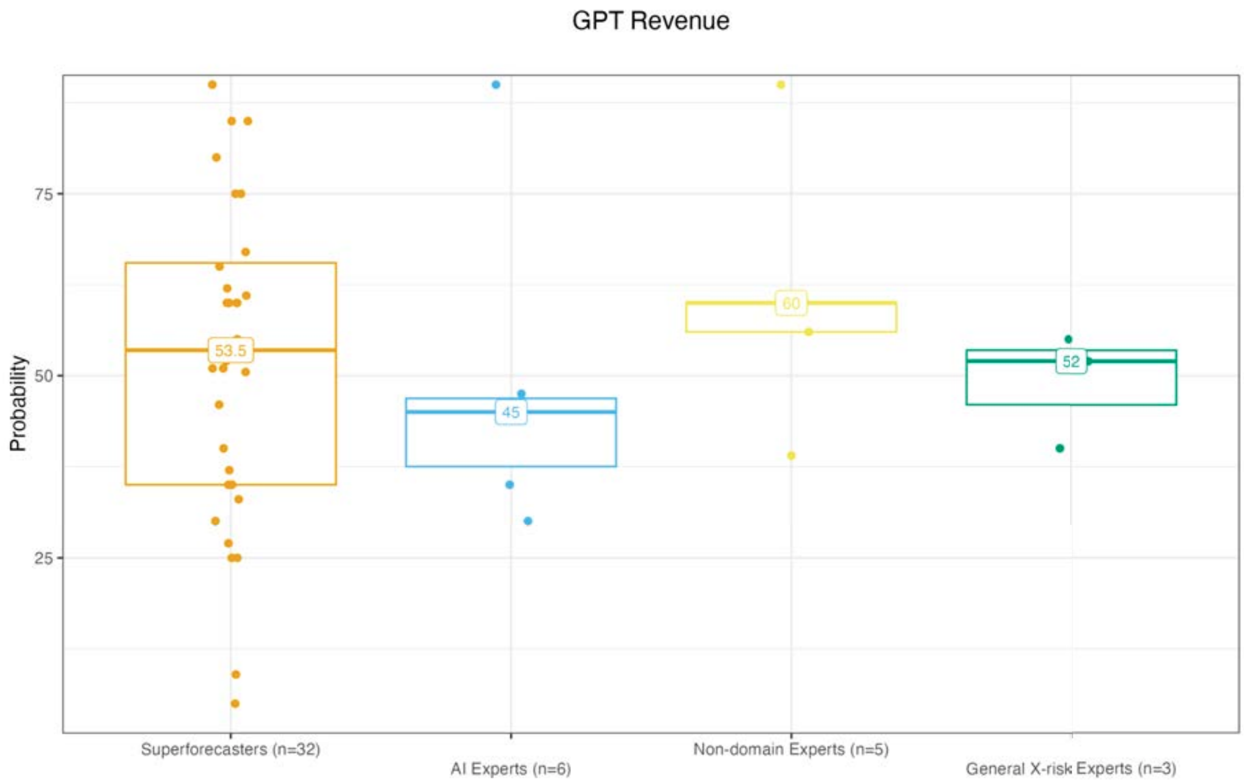
Results[1020]

Group	**Stage 1 Median**	**Stage 4 Median**	**Stage 1 Standard Deviation**	**Percent Change in Standard Deviation, Stage 1 to Stage 4**
Super-Forecasters (N = 32)	37%	53.5%	30.24	-24.42%
Domain Experts (N = 6)	67.5%	45%	31.82	-33.01%
General X-Risk Experts (N = 3)	NA[1021]	52%	NA	NA
Non-Domain Experts	68%	60%	20.98	-12.38%

[1020] Numbers of forecasters are given as of Stage 4 of the XPT.
[1021] No forecasters in this group answered this question in Stage 1.

(N = 8)				


GPT Revenue
Median
Date
Aug
Sep
Oct
Nov
40
45
50
55
60
Superforecasters (n = 32)
Experts (n = 14)



Sources of agreement, disagreement and uncertainty

Sources of uncertainty regarding GPT revenues:

- Current revenues and growth rates[1022]
- How to count the proportion of revenues from other services attributable to GPT[1023]

[1022] 336, “The biggest uncertainties in forecasting this question are the amount of current revenues and the growth rate of the market. A doubling of revenues each year looks like a reasonable median for a case like this; A new, potentially revolutionary and easily scalable technology.
Estimations of current revenues are wide-ranging: 'Reddit' estimates revenues of approximately $120mn (April 2, 2021), @Elifland at Metaculus gives a number of just $20mn/y for 2021. Doubling revenues every year gives a number for 2024 in the range of $160mn-$960mn. Based on @Elifland's numbers, cumulative revenues will be $300mn at the start of 2025, with Reddit's starting number, cumulative revenues would be $1.8bn by then.
However, Reddit’s estimation of $120mn in revenues in 2021 is probably too high because it uses old retail rates as the price for every word generated and there are probably a large number of customers who at present pay much less or nothing at all. The website Growjo.com estimates OpenAI current revenues at $59.6mn a year (2022). Other, similar websites have lower estimates of around 20mn (not sure what year exactly).
Therefore, we will assume that OpenAi's revenues in 2021 were no higher than $30mn. Doubling that every year yields 30+60+120+240 = $450mn. However, the revenues of OpenAI are not the only ones that fall in the category " GPT line of language models". Will GPT-Neo, GPT-J, etc. count as well? It is not unusual in a new market that the first entrant takes >50% of total market share. If we assume it's 50/50, we estimate total revenues of $900mn at the start of 2025.”

[1023] 336, “a big uncertainty is how to count the revenues of the service when it is 'hidden' in other services. For example, Microsoft does not specify the revenues of its individual Azure-services. Also, they

- Difficulty estimating revenue from a large number of small companies entering and leaving the field[1024]
- The impact of the strengthening dollar on GPT revenues[1025]

Forecasters also noted uncertainties in relation to the resolution criteria:
- Will inflation be adjusted for[1026]
- Does the question refer to cumulative revenue[1027]
- Will resolution occur at the start or end of 2025[1028]

One team noted disagreement about whether Metaculus forecasts tend to be too optimistic or too pessimistic on AI technologies.[1029] Another mentioned uncertainty about the intention of the question setters in setting a question so similar to an existing Metaculus question.[1030]

Arguments given for forecasts of ≤49% (i.e. that Hanson will lose the bet/GPT revenue will exceed $1bn)

- GPT-3 is very impressive.[1031]
- There are many use cases for GPT products. Examples cited: Github copilot, customer feedback tools, SEO, content creation, translation, writing code, summarization.[1032]

---

are using GPT-3 in a separate service as well (Power Apps).” See also 344, “A large portion of the team's uncertainty comes from how "systems in the GPT line" will be evaluated.”

[1024] 336, “Another issue is that many small companies enter (and leave) the field continuously, or migrate to different AI-providers, making it harder to get a grip on the actual total revenue.”

[1025] 336, “[A] recent source of uncertainty is the rising strength of the dollar, compared with other currencies. We do not know how much of GPT-3 revenues come from outside the US and how much influence the strengthening of USD will have.”

[1026] 338, “[It’s] not clear whether they will adjust for inflation in resolving the bet; if they don't, the bet is harder for Hansen [sic] to win.”

[1027] 341, “There is some uncertainty regarding [whether] the question will resolve based on cumulative revenue or not. If it is not cumulative, the probability is higher. The team assumes it is cumulative.”

[1028] 341, “There was an earlier uncertainty regarding whether the bet ends at the beginning or end of 2025. Since RH seemed to clarify in a tweet that it should refer to the beginning of 2025, the team assumes this is the case.”

[1029] 343, “There was disagreement over whether Metaculus forecasters are biased in the direction of over optimistic forecasts for technology, or whether the bias for AI-related technologies goes the other direction. The magnitude of this bias, and how it should be applied to the Hanson bet was poorly defined.”

[1030] 345, “Metaculus is only the source for information.
While this question is exact copy from the same source, what might be the intention of organizers to simply copy the same question from Metaculus and use it here?
1) Just want to get "our" perspective
2) Want to understand if we stay loyal to Metaculus
3) Checking how much we get manipulated by seeing same question in Metaculus (what is the logic for this?)
4) Anything else which organizers don't want to reveal?”

[1031] 345, “The advances in GPT-3 are pretty staggering despite its limitations.” Also “Inside view: Few team members have got access to DALLE-2 ([its] product) and are super impressed.”

[1032] 345, “In addition to this, the market has a pretty wide variety of use cases where it could be applied. Github copilot is a great example but also customer feedback tools, SEO and even some content

- Microsoft is incorporating GPT-3 into some of its main product lines (Azure, PowerApps). These have very high net revenues, so attributing a small portion of them to GPT-3 would mean over $1bn revenue for GPT.[1033]
- The global video game market is $180bn in 2021. It doesn't seem unlikely that one very successful use of GPT in a video game could generate over $1bn, and that is only one domain.[1034]
- It doesn't seem unlikely that DALL·E could generate $300m/year by 2023.[1035]
- Large language models may in future replace Google search, though this may not happen within the timeframe.[1036]

- GPT-4 may be launched this year.[1037]
- Smaller, fine-tuned models can be run more cheaply than GPT-3.[1038]

---

creation." See also 341, "The team notes many use cases for GPT models, raising the probability that a large number of companies can make use of it. It can be used to write text, write code, answer questions, summarize texts, and make translations, among other things." See also 339, "there are lots of "low-hanging fruit" settings in which GPT-language can be deployed: eg, customer service, internal corporate help desks, event planning, etc."

[1033] 341, "The Microsoft use case is the main argument for the team's lower forecasts. Summarizing the arguments simplistically, if Microsoft had not licensed GPT-3, the team would be leaning toward a higher probability of Robin Hanson winning the bet. However, with Microsoft incorporating GPT-3 into its main product lines (Azure, PowerApps), the team believes there is a possibility that if one were to ascribe even a very small proportion of revenues to GPT-3, RH would lose the bet given the large size of those product lines' revenue. Microsoft Cloud revenue is $23.4 bn per quarter (Azure is not broken out separately). Then, Microsoft will put it in powerapps, which is already a $2bn business. This is the main factor that has lowered our team median forecast over time." See also 336, "The strongest argument [...] for a forecast on the lower end of the plausible range of forecasts for this question is the deal with Microsoft and the possibility of similar future deals with big (cloud) companies. Currently, Microsoft is the biggest customer of OpenAI and is offering 'Azure OpenAI Service' and other services that make use of GPT-3. Microsoft sees potential enterprise uses ranging from summarizing customer service logs and other sources of information to helping developers code faster and easier. Although cloud companies are known for throwing everything at the wall to see what sticks, they are capable of scaling up their services very quickly if it does and that would make >$1bn revenues possible within this time frame." See also 338, "It looks as though there is going to be significant demand for microsoft products that make use of GPT-3 in the near future (https://venturebeat.com/ai/microsoft-gpt-3-and-the-future-of-openai/)".

[1034] 344, "In analogy with the growth of the global video game market, which had revenues of around $180 billion in 2021. It may only take one great use of natural language processing in a breakout game to get revenues over $1 billion, and that's just monetizing in one domain."

[1035] 344, "DALL·E is now in Beta and becoming more widely available (at $15 per 115 generations). Depending on how it develops and how much competition it faces, it does not seem unlikely that it could end up generating $300 million/year by 2023."

[1036] 341, "The team especially notes that LLMs have the potential to be the future equivalent of Google searches that will provide you the exact information you need, although that might be past 2025."

[1037] 341, "The upcoming launch of GPT-4 will likely create new revenue opportunities. Timing is uncertain, but has been speculated to be this year, as per Metaculus." See also 343, "New GPT releases will likely substantially improve, not just prior performance quality, but will likely see new emergent capabilities as with previous GPT releases… Given that we cannot predict what new capabilities will become available in future GPT releases, it's possible that OpenAI will be able to directly generate significant customer revenue, and will also be able to generate revenue through companies built on its API." See also 340, "There is optimism because some forecasters expect OpenAI to train and market a new and improved model sometime soon."

[1038] 341, "The cost of running GPT-3 is large, but there are smaller, fine-tuned models that can be run more cheaply."

Arguments given for forecasts of ≥60% (i.e. that Hanson will win the bet/GPT revenue will not exceed $1bn)

- Competition is increasing.[1039]
  - There are an increasing number of fully open source large language models.[1040]
  - There are high-quality alternatives to DALL·E, like Midjourney, Stable Diffusion, Craiyon.[1041]
  - There are competitor large language models which may be more advanced than GPT-3, like DeepMind and BLOOM.[1042]
- There is no clear path to mass monetization at present/no outstanding application of GPT.[1043]
- GPT is still changing rapidly and it may be hard for new tech companies to build reliable platforms on such rapidly changing models.[1044]
- Tech platforms often take time to scale and have a high probability of failure.[1045]

[1039] 343, "OpenAI has competitors, and this bet is based on OpenAI's platform specifically. Those competitors in other areas (such as DALL-E/image generation) are already showing signs of beating OpenAI. If GPT-3 has to compete against other models - especially models that aren't built on a pay-per-token fee structure - this will divide the market and make it much more difficult for GPT-based models to capture $1 billion in revenue by 2025." See also 341, "There are already several competitors to GPT-3, such as PaLM from Google and MT-NLG from Microsoft-NVIDIA. Although these are not currently monetizable in the same way." See also 338, "Competitors to [GPT-3] have recently entered the market." See also 340, "There is uncertainty because now large models exist that are greater in size than GPT-3 (Bloom, OPT-175B) and open source, and it is not clear when, if at all, OpenAI will produce a GPT-4." See also 342, "In the event of uptake, the rapid development of rival systems will take market share from OpenAI's product…Lots of competitor products are already on the market, including https://elicit.org/, and each one will need to adapt quickly to have the best product. It's conceivable that a company other than OpenAI will end up with the best one."

[1040] 339, "there will be an increasing number of extremely large language models that are fully open sourced."

[1041] 344, "The emergence of high-quality alternatives to DALL·E (such as Midjourney, Stable Diffusion, Craiyon) implies more competition and less profit margins for OpenAI."

[1042] 336, "Increasing competition of other and perhaps more advanced non-GPT3 language models (Deepmind, BLOOM). Not only can they compete with GPT-3 models, they are also driving down the price of AI Language services in general, making it harder to reach $1bn in total revenues."

[1043] 336, "No clear and easy path yet to mass monetization. The absence -at least right now- of a 'killer app' leads to a downward adjustment of expected growth and thus a higher chance for this question to resolve in 'yes'." See also 341, "The main argument is that the actual revenue opportunities are [as] yet untested".

[1044] 343, "GPT is still a rapidly-developing platform. Although OpenAI has an API for interfacing with their language models, it may be difficult for new tech companies to establish reliable platforms based on rapidly-changing models such as these."

[1045] 343, "Tech platforms often take time to scale before successful monetization, with high probability of failure. Two current companies that appear to be at the forefront of direct GPT monetization, CopyAI and Copysmith, appear to bring in a combined total of less than $5million/year so far. Even if both these companies succeed, [it's] unlikely they'll generate enough revenue to top $1b total by 2025. It's difficult to define a scenario where either or both companies reach a market valuation of $1 billion by 2025, let alone cumulative revenues that high. Other potential entrants into this market would be behind both these two market leaders, and therefore should not be expected to generate significant revenue by 2025, either."

- Hanson has won all of his previous public bets.[1046]

Other arguments given

Arguments in favor of lower forecasts:

- GPT has a first-mover advantage.[1047]

Arguments in favor of higher forecasts:

- A year and a half isn't a long time.[1048]
- GPT-3 is expensive to run.[1049]
- Text generated by GPT-3 still requires checking by a human.[1050]
- GPT-3's output may be too general for most commercial uses.[1051]
- The economic downturn may reduce companies' risk tolerance for using novel technologies.[1052]
- OpenAI has been cautious with API rollout in general.[1053]
- Historically, GPT-style systems have quickly become obsolete without generating much revenue at all.[1054]
- Large language models could be regulated against because of bias in their outputs, though this might happen outside the timeframe.[1055]

[1046] 344, "Robin Hanson has won all of his public bets so far, although admits some of this was due to luck." See also 341, "The team also notes Hanson's pedigree as the founder of the science of crowdsourced forecasting."

[1047] 343, "GPT itself is already well known, giving this model first mover advantages over competitors… GPT is well situated to capitalize on these new capabilities with an active developer community and a well-established API." See also 338, "GTP-3 has a head start on competing programs, and so will probably capture a lot of market share in the near future." See also 340, "There is optimism because OpenAI was the first mover in this space."

[1048] 345, "It's unclear if the AI would be good enough to induce widespread enough adoption to generate a billion dollars in just a year and a half." Incorrectly labeled as an argument for lower forecasts in the rationale. See also 339, "others felt that the timeframe was too short to develop a revenue model that would generate enough customer revenue."

[1049] 341, "The second main argument is the high cost to running GPT-3, which can prove difficult for small companies to make use of it beyond proofs of concept." See also 342, "GTP is more expensive than other similar products, and it's yet unclear if this extra cost is worth it when compared to them."

[1050] 342, "[GPT]-3 produces good quality text, but still requires checking by a human prior to publication."

[1051] 341, "the team notes that GPT-3's output might be too general for most commercial uses."

[1052] 345, "Also a slowing economy could have a large [effect] in reducing the risk tolerance of companies on bringing in untested technology". Incorrectly labeled as an argument for lower forecasts in the rationale.

[1053] 342, "While OpenAI may have the best product on the market at the moment, they are being cautious with Application Programming Interface (API) rollout in general. It might not be commercially used enough over the timeline of the bet to generate substantial revenue." See also 344, "Another, tinier part of uncertainty lies in the possibility that OpenAI and other labs will show restraint in some profitable venues due to proliferation concerns (i.e. to avoid accelerating interest and investment in AI capabilities)."

[1054] 343, "[GPT]-style systems historically have a higher chance of going obsolete quickly, before making much revenue at all."

[1055] 341, "Regulation: If there was a successful groundswell against the usage of LLMs on account of their potentially biased outputs, there could be regulation against this. This would probably happen outside of this timeframe, however."

- Current hype around GPT-3 may bias forecasts downwards.[1056]

Cross-references with other questions

Q39: Math Dataset Benchmark
Q40: Massive Multitask Language Understanding Benchmark
Q41: QuALITY Dataset Benchmark
Q42: AI Wins International Mathematical Olympiad
Q43: NYT Bestsellers Written by AI

Question 36: US GDP From Software

What percentage of US GDP will result from software and information services…
… in 2024?
…in 2030?
…in 2050?

Question and resolution details, prior forecasts, and other relevant sources

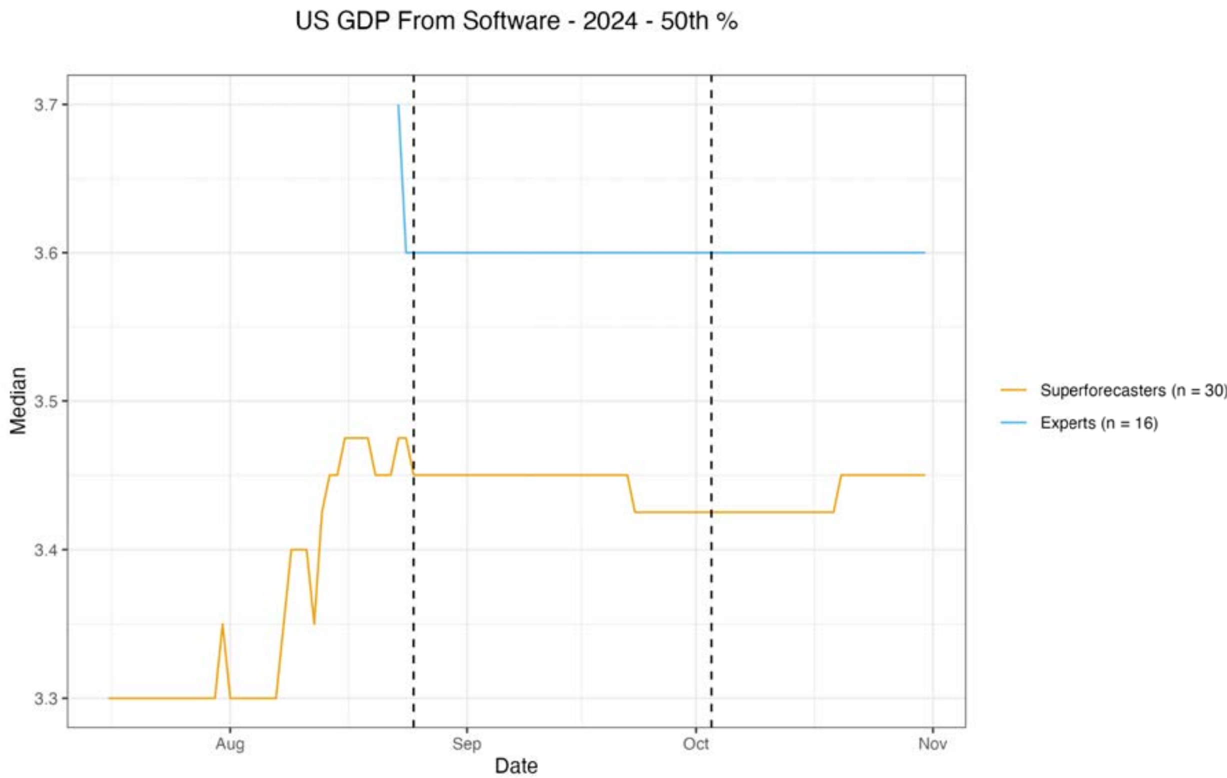
Results[1057]

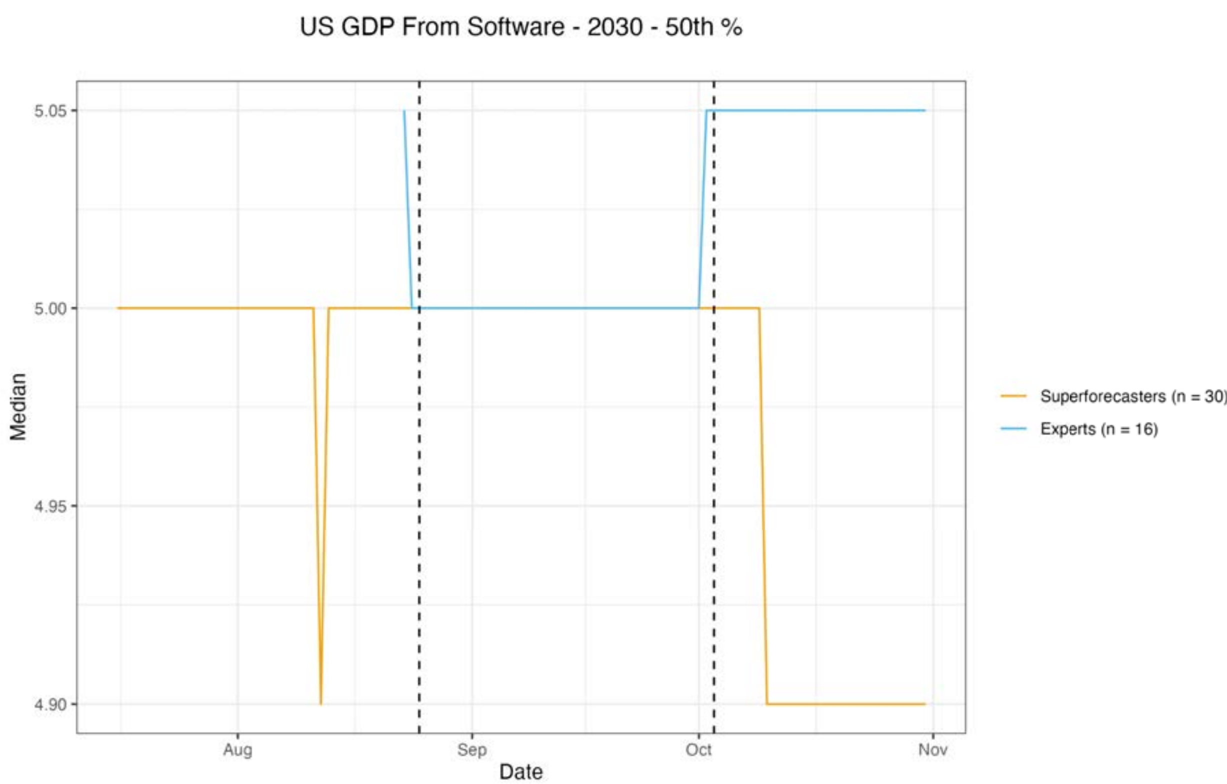
Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 30)	2024	3.3%	3.45%	1.92	-29.1%
	2030	5%	4.9%	1.92	-24.82%
	2050	7.8%	7.71%	4.25	-26.85%

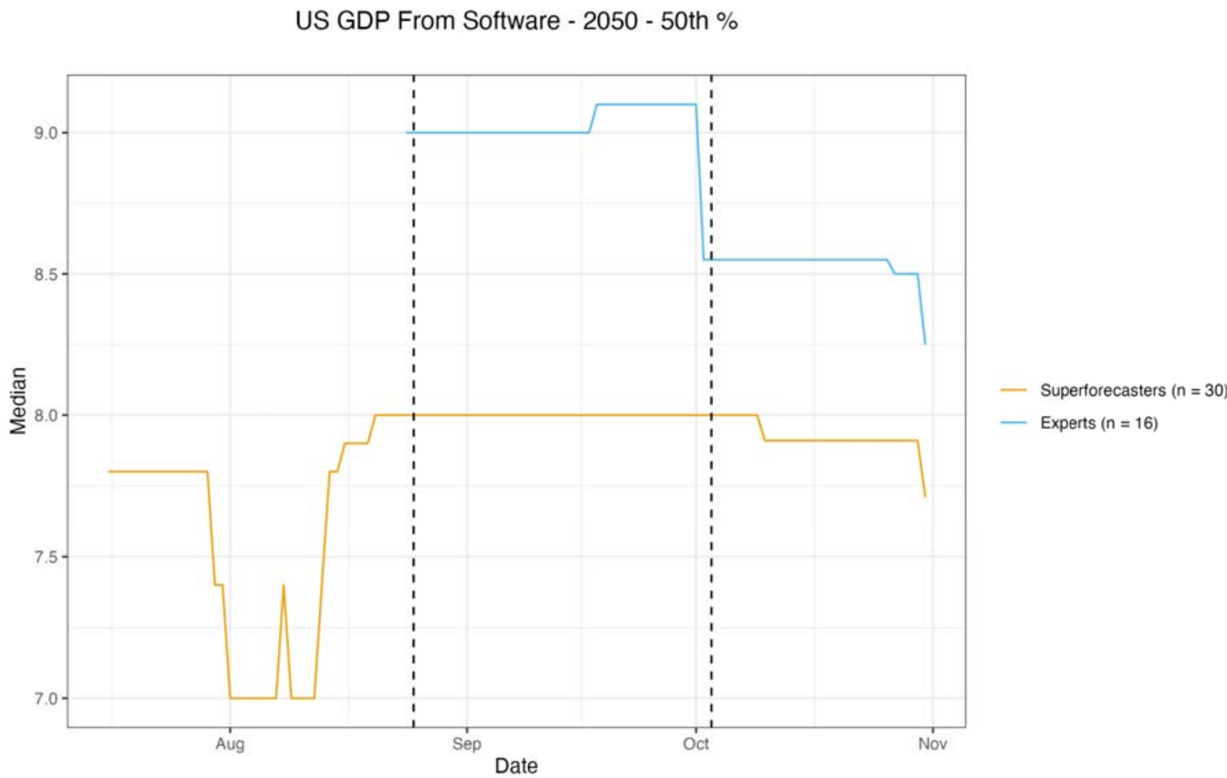
[1056] 341, "There is currently a lot of excitement about GPT-3 and similar models in the media, some of which refers to future use cases rather than current ones, potentially biasing the team toward a lower probability for this timeframe (next 2.5 years)."
[1057] Numbers of forecasters are given as of Stage 4 of the XPT. Public survey respondents were surveyed outside of the tournament context.

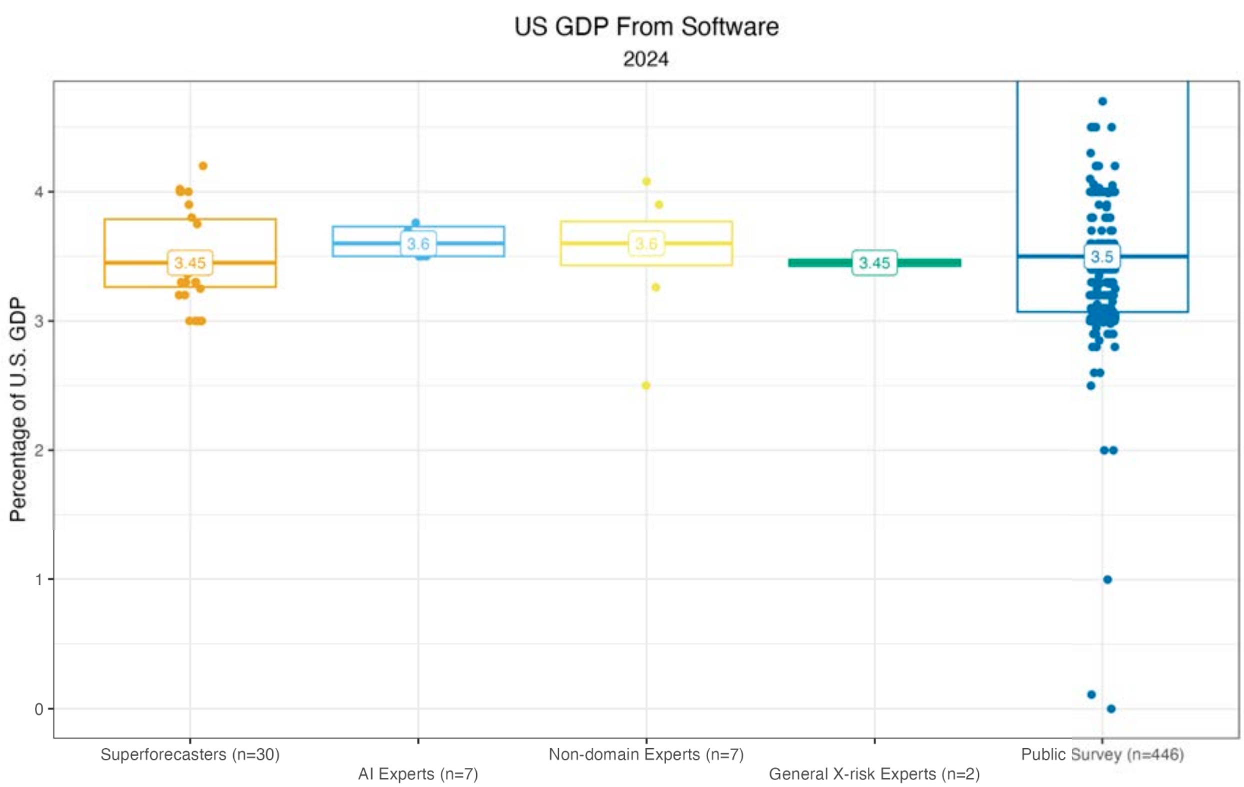
Domain Experts (N = 7)	2024	3.76%	3.6%	0.60	-10.43%
	2030	5%	5.5%	0.87	+13.89%
	2050	9.05%	9%	2.78	-53.12%
General X-Risk Experts (N = 2)	2024	NA[1058]	3.45%	NA	NA
	2030	NA	5.13%	NA	NA
	2050	NA	7.13%	NA	NA
Non-Domain Experts (N = 9)	2024	0.033%	3.6%	NA[1059]	NA
	2030	0.039%	5%	NA	NA
	2050	0.053%	8.5%	NA	NA
Public Survey (N = 446)	2024	3.5%		25.04	-
	2030	5%		37.34	-
	2050	8%		35.01	-

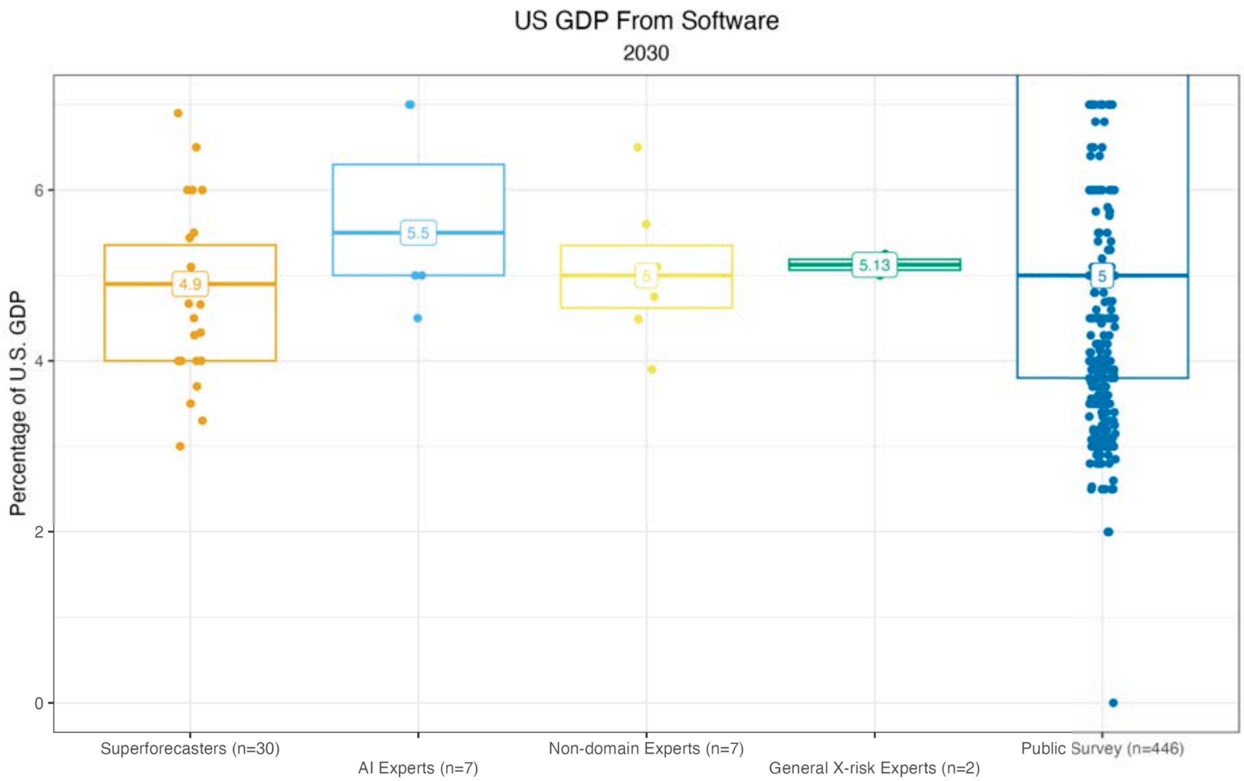
1058 No forecasters in this group answered this question in Stage 1.
1059 Only one forecaster in this group answered this question in Stage 1.

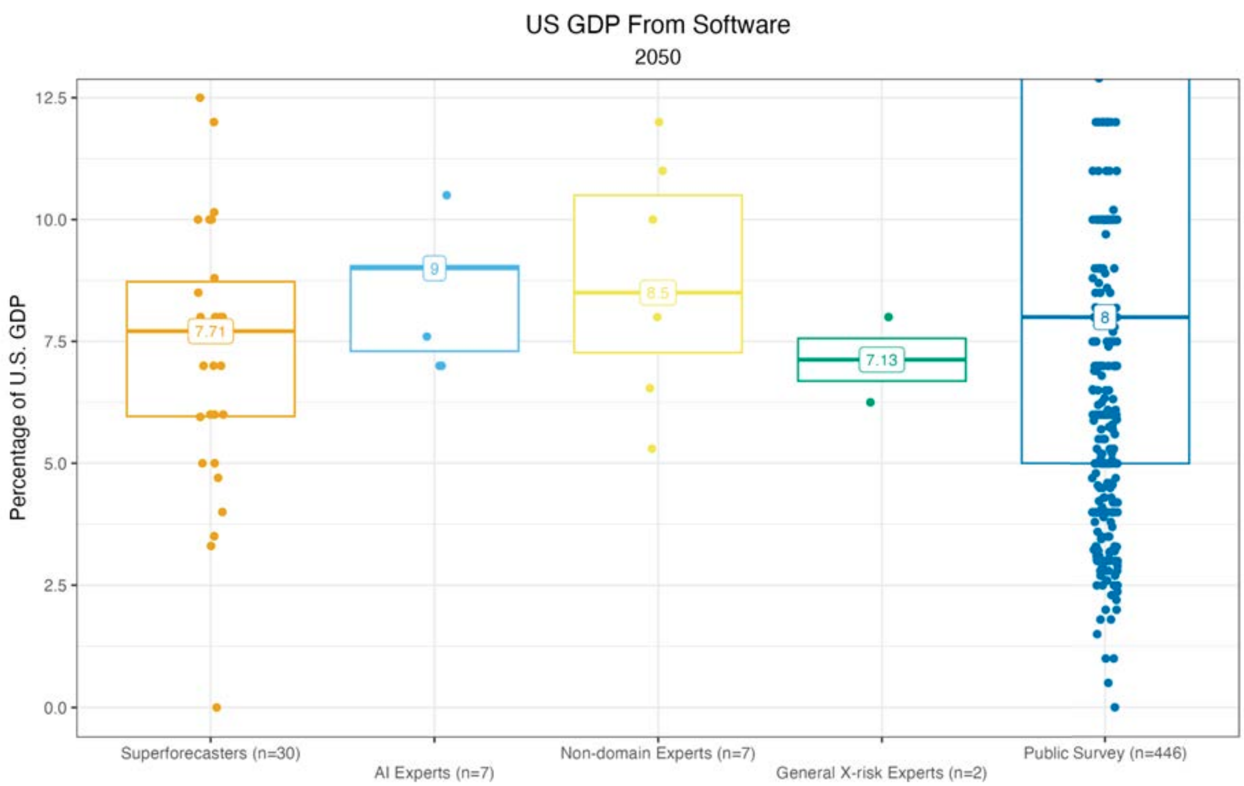

US GDP From Software - 2024 - 50th %
Median
3.7
3.6
3.5
3.4
3.3
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 30)
Experts (n = 16)


US GDP From Software - 2030 - 50th %
Median
5.05
5.00
4.95
4.90
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 30)
Experts (n = 16)


US GDP From Software - 2050 - 50th %
Median
9.0
8.5
8.0
7.5
7.0
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 30)
Experts (n = 16)


US GDP From Software
2024
Percentage of U.S. GDP
4
3
2
1
0
3.45
3.6
3.6
3.45
3.5
Superforecasters (n=30)
AI Experts (n=7)
Non-domain Experts (n=7)
General X-risk Experts (n=2)
Public Survey (n=446)





Sources of agreement, disagreement and uncertainty

The main sources of disagreement for this question were:

1. How to interpret base rates
2. How the share of GDP associated with software and information services will be counted[1060]
3. The likelihood and direction of the impact of advanced AI[1061]
4. The probability of human extinction[1062]

Arguments given for forecasts below 3.5% (2024), 5.0% (2030), 8.0% (2050)

On 1 (base rates):

- These numbers tend to move slowly.[1063]
- Software growth over recent years can be attributed to the pandemic, when other industries were differentially depressed relative to software.[1064]
- Productivity has been declining in spite of technological innovation.[1065]
- Prices for software will drop as hardware gets cheaper and systems get more efficient.[1066]

On 2 (accounting):

[1060] 338, "Sources of uncertainties: How AI dominance and/or guidance of the economy would convert to share of GDP associated with software and information services"; 336, "This question is hard to judge, because we are almost certainly massively undercounting. For instance, all of Amazon's revenue and essentially the entire financial industry are incredibly dependent on "software information services", but I don't think those numbers are counted in the categories. [F]or how long will 'software' remain a separate, distinguishable category, when it is everywhere and in everything?"; 343, "How much of the economic gains from AI will accrue to the two categories we're forecasting".

[1061] 336, "The advent of advanced AI could be a game changer. However, it is hard to foresee if and what kind of economy would remain after the singularity."; 340, "For the later forecasts, differing estimates of the chances and implications of AGI and/or human extinction heavily affect this question, especially at the tails along with the impact improved software may have on economic activity."; 337, "Whether AGI will emerge within the 2050 timeframe (what difference would that make? If most software is included in the GDP of something else won't most AI be included in the GDP of some product that uses AI?[])]"; 343, "AGI - The speed in which it will be acquired (if ever), and [its] impact on GDP - seems to be the biggest source of uncertainty regarding the far future."

[1062] 340, "For the later forecasts, differing estimates of the chances and implications of AGI and/or human extinction heavily affect this question, especially at the tails along with the impact improved software may have on economic activity."

[1063] 341, "These are very slow-moving numbers in general."

[1064] 341, "The growth of the last two years was a one-time pandemic boom which probably pulled in future growth."; 340, "For 2024, if the recent increase in software GDP% is due only to the economic effects of the COVID-19 pandemic, then this increasing trend may not continue in the next 2 years and may even reverse consistent with the 2000-2002 period following a similar fall in technology valuations and funding and subsequent recession. This also applies to the prediction for 2030, albeit to a lesser extent." Given as an argument for forecasts below the team median, which for 2050 was 9%.

[1065] 341, "Productivity has declined despite tech innovation in recent years, which does not bode well for accelerated investment."

[1066] 336, "Prices will drop with more efficient systems and cheaper hardware, a trend that will most likely continue for a while (although it might slow down a bit)". Given as an argument for forecasts below the team median, which for 2030 was 5.3%.

- The contribution of AI will be spread across many industries, rather than counted solely towards this category.[1067]
    - Accounting for the contribution of software may get increasingly difficult as it becomes more ubiquitous.[1068]

On 3 (impact of advanced AI):
- AI may fail to live up to the current hype.[1069]

On 4 (human extinction):
- Economic catastrophe or human extinction may drive software contributions down.[1070]

Other arguments:
- The US may lose the technology race with China and the EU.[1071]
    - As software can be made anywhere, it is possible that much software development will move outside of the USA.[1072]
- While humans exist, there are sectors of the economy which cannot be replaced by software, such as food, accommodation, travel.[1073]

[1067] 341, "Artificial intelligence impact may be spread across many industries, and not just this category." See also 336, "This question is hard to judge, because we are almost certainly massively undercounting. For instance, all of Amazon's revenue and essentially the entire financial industry are incredibly dependent on "software information services", but I don't think those numbers are counted in the categories. [F]or how long will 'software' remain a separate, distinguishable category, when it is everywhere and in everything?" See also 337, "A key assumption behind this point is that although AI applications will enter most industries they will not be considered as software but as part of the sector utilizing the AI application itself. An example provided is of alpha-fold being considered 'biotech' as opposed to software. This then makes the robustness of NAICS industry categories another important consideration."

[1068] 336, "There are major challenges of measuring the digital economy [due] to a lack of consensus around activities included in the definition and the rapid pace at which the underlying nature of digital technologies evolves. We expect accounting will be even harder in 2050 with the ubiquity of these services and products." Given as an argument for forecasts below the team median, which for 2030 was 5.3%.

[1069] 345, "Web3 and AI performing below their hype."

[1070] 340, "For 2030 and especially 2050, the risk of economic catastrophe (up to and including human extinction) is what drives the lowest estimates of software GDP%." Given as an argument for forecasts below the team median, which for 2050 was 9%. See also 344, "Activation of catastrophic and extinction risk may push the numbers down to 0".

[1071] 345, "US losing out on the IT race to China, EU, etc".

[1072] 344, "An interesting argument for lower-end forecasts is that software can be made anywhere and therefore we can expect it to be moved outside the USA."

[1073] 336, "As long as there are humans around, some items like food, living, travel etc. can not be replaced by software, so they will keep their share of GDP". Given as an argument for forecasts below the team median, which for 2030 was 5.3%. See also 340, "For example, people still need to eat food, live in buildings and wear clothes and these are material goods with material costs and processes even when automated. Software can only capture so much of the value of these goods. There are also services difficult to provide through software (some healthcare or leisure) and industries which [may be] resistant to cutting employment or automating (e.g. Government)."

- Forecasters are likely to be biased upwards as they are above average consumers of software products.[1074]

Arguments given for forecasts above 3.5% (2024), 5.0% (2030), 8.0% (2050)

On 1 (base rates):

- Productivity gains from information services may be lagging until a tipping point is reached, similarly to during the Industrial Revolution.[1075]
- There is precedent for segments of the economy growing much faster than overall GDP, like manufacturing.[1076]

On 2 (accounting):

- If NAICS improves its categorisation of software products, this might create a jump in software's share of GDP.[1077]

On 3 (impact of advanced AI):

- AI may develop significantly in this period.[1078]
    - AI developments may enable software to replace other industries.[1079]

Other arguments given

Arguments in favor of lower forecasts:

- On 1 (base rates):
    - So far, IT has not resulted in major GDP growth.[1080]

[1074] 336, "we (both Superforecasters and Experts) are likely to be much greater consumers of these services than the average person in the US, so we are subject to bias." Given as an argument for the team median, which for 2030 was 5.3%.

[1075] 341, "Figure 6 Industrial Production per Person in Six centuries of British economic growth: a time-series perspective shows growth of 0.15% annually before the Industrial Revolution, then 0.63% between 1633-1822, then 1.62%. Productivity gains from IS may be lagging until they reach a tipping point in the economy."

[1076] 341, "Software is a small part of the economy now and could grow faster than the remainder of GDP. Some acceleration in growth of segments of GDP are possible, as when manufacturing grew much faster than overall GDP."

[1077] 337, "[NAICS] might enforce more accurate categorization of what counts as software products, which would likely bump up software products due to being present in most products and services in some way. Or [NAICS] might alter the definition of software and information products to become more expansive."

[1078] 341, "AI will have evolved significantly and could be transforming the economy."

[1079] 340, "Advances in AI (as recently demonstrated by GPT-3, DALL-E-2, LaMDA, AlphaFold2, etc.) may enable software to replace other industries, thus increasing as a % of GDP."

[1080] 337, "Wide-spread saturation of information technologies such as the web has not resulted in major growth in GDP from software. Tech's share of GDP is still only in the 3%-4% 30 years later despite the widespread adoption of the Internet and smartphones over that period."

- One of the major contributions of software is to enable other industries, so these will also be growing as a share of GDP.[1081]
- On 2 (accounting):
  - Productivity improvements from AI will not count towards increased share of GDP (only increased sales will).[1082]
- On 3 (impact of advanced AI):
  - Advanced AI will not emerge in the time frame.[1083]

Arguments in favor of higher forecasts:
- On 3 (impact of advanced AI):
  - Advanced AI may lead to software and information services making up most of the economy.[1084]
  - The run up to advanced AI might cause increased investment.[1085]
  - Software automation may lead to explosive growth.[1086]

Cross-references with other questions

On 3 (impact of advanced AI):
- Q44: Date of Advanced AI
- Q51: Nick Bostrom Affirms Existence of AGI

On 4 (human extinction):
- Q10: Total Extinction Risk
- Q11: Year of Extinction

[1081] 343, "One major function of software has been to enable other technologies. Thus, while it's rational to assume software will continue to increase as a percentage of GDP, it should also be assumed that other sectors of the economy will also be enabled to grow alongside software."
[1082] 337, "AI was also considered as a way to improve productivity in IS related industries. However, productivity improvements don't move the GDP, only increased sales do, and those depend on demand, not the tools used for production." Given as an argument for forecasts of 3.9% (2024), 5.2% (2030), 8.8% (2050).
[1083] 337, "The consensus is that AGI will not emerge within this timeframe - and that could be a major disruptor with unpredictable consequences." Given as an argument for forecasts of 3.9% (2024), 5.2% (2030), 8.8% (2050).
[1084] 343, "a single dominant force capable of taking over the economy wholesale (AGI)"; "Future technological developments, including AGI, might make the general trend advance faster than ever before by orders of magnitude - akin to the agricultural or industrial revolutions."; 340, "the highest forecasts at the higher confidence bands (75-95%) allow for a reimagining of economic activity such that a majority of monetary value is captured by software industries (based on the potential transformative effect of AI)."
[1085] 336, "New technological developments - most notably transformative AI - could break the trend (in both ways). In particular, the final stretch to AGI could lead to heavy investments from different parties in an attempt to reach AGI first."
[1086] 344, "Explosive growth due to software automatization (code that writes code)."

Question 37: US Computer R&D Development

How much money will be spent on research and development by US companies in the 'Information' and 'Computer systems design' industries…
…in 2024?
…in 2030?
…in 2050?

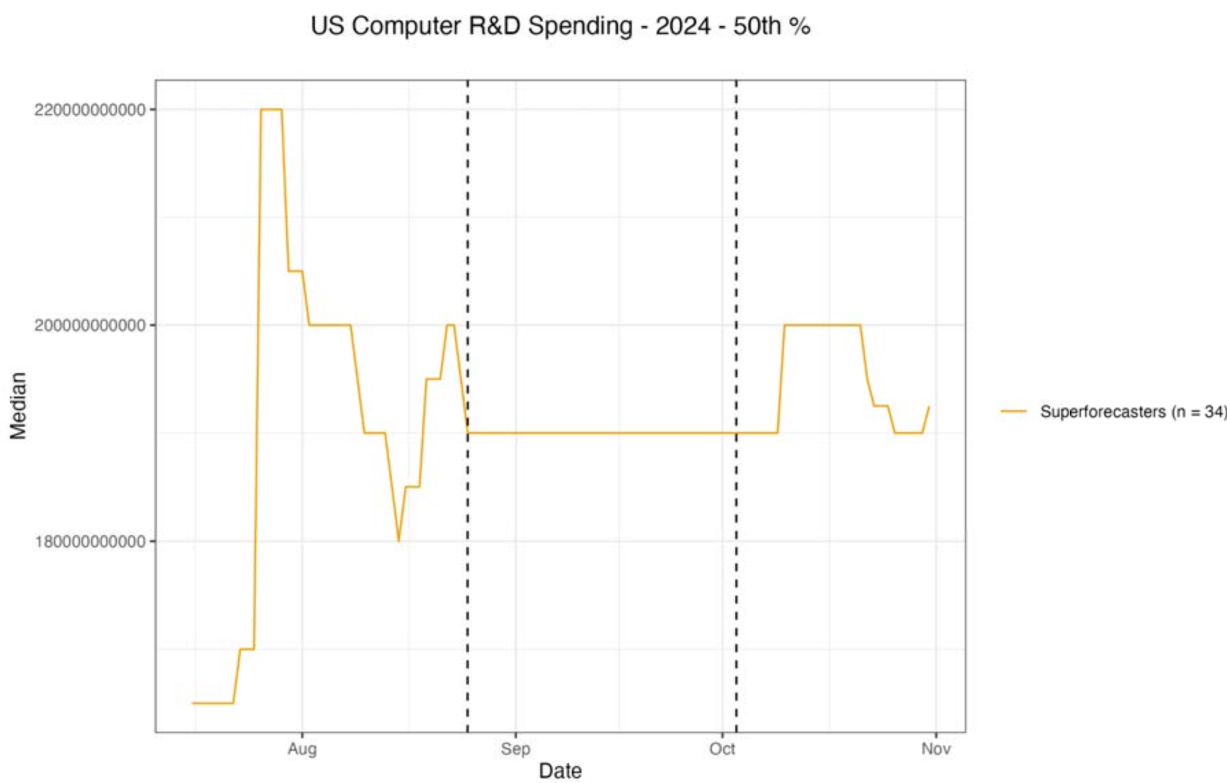
Question and resolution details, prior forecasts, and other relevant sources

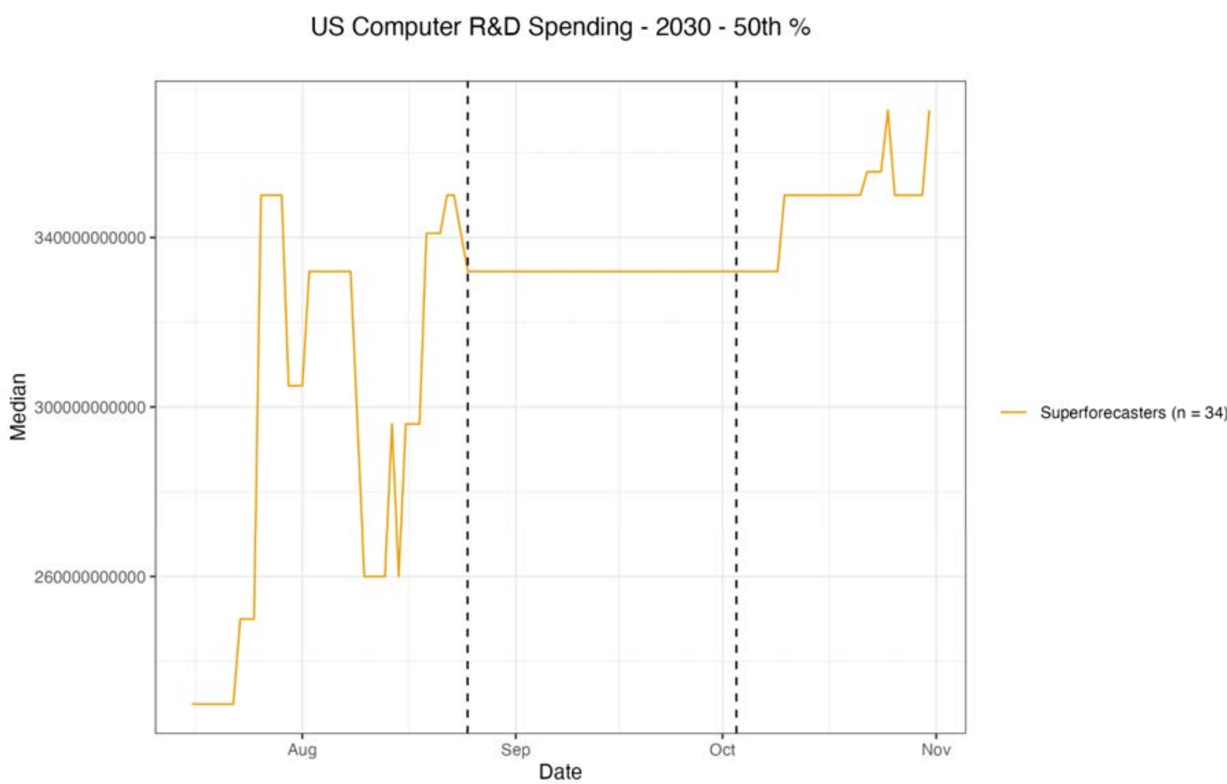
Results[1087]

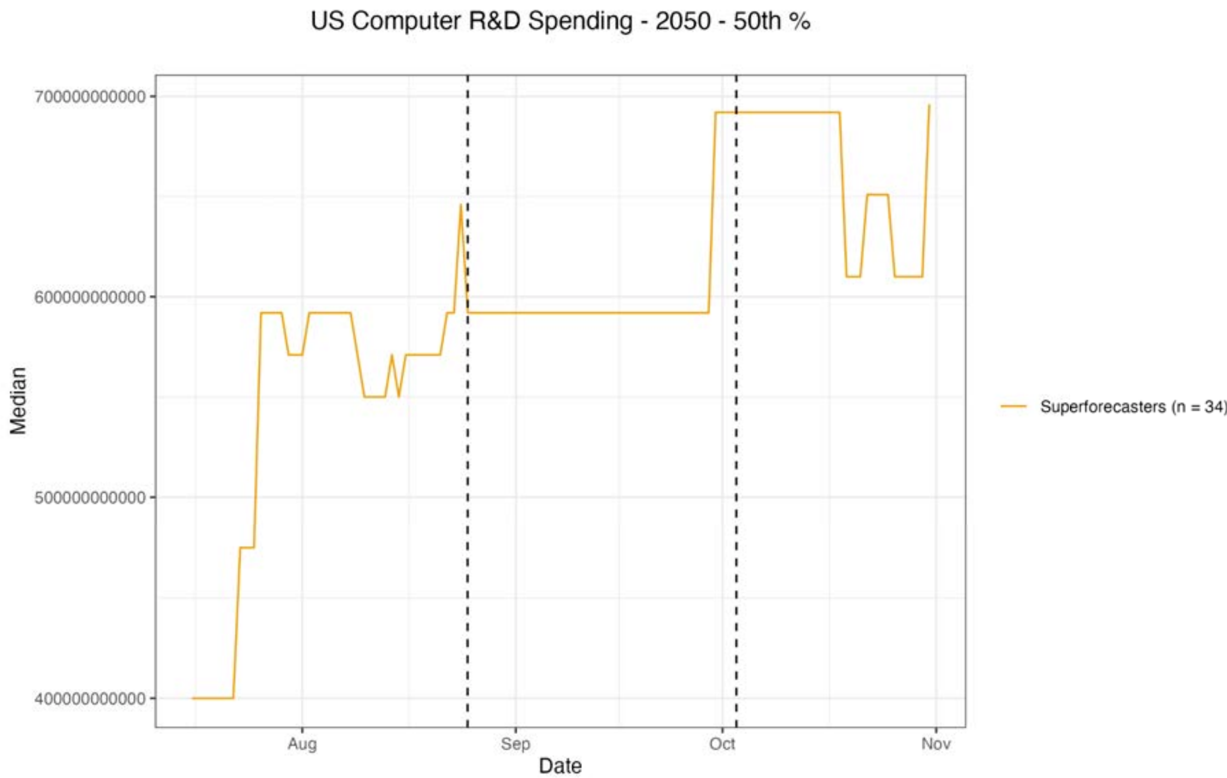
Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 34)	2024	$230,000,000	$192,500,000,000	107014163269	-10.79%
	2030	$350,000,000	$370,000,000,000	193456756063	+10.55%
	2050	$2,400,000,000	$696,000,000,000	1234088780244	+74.09%
Domain Experts (N = 2)	2024	n/a	$240,000,000,000	n/a	n/a
	2030	n/a	$605,000,000,000	n/a	n/a
	2050	n/a	$785,000,000,000	n/a	n/a
General X-Risk Experts (N = 1)	2024	$198,207,006,769	$198,207,006,769	n/a	n/a
	2030	$365,493,025,474	$365,493,025,474	n/a	n/a

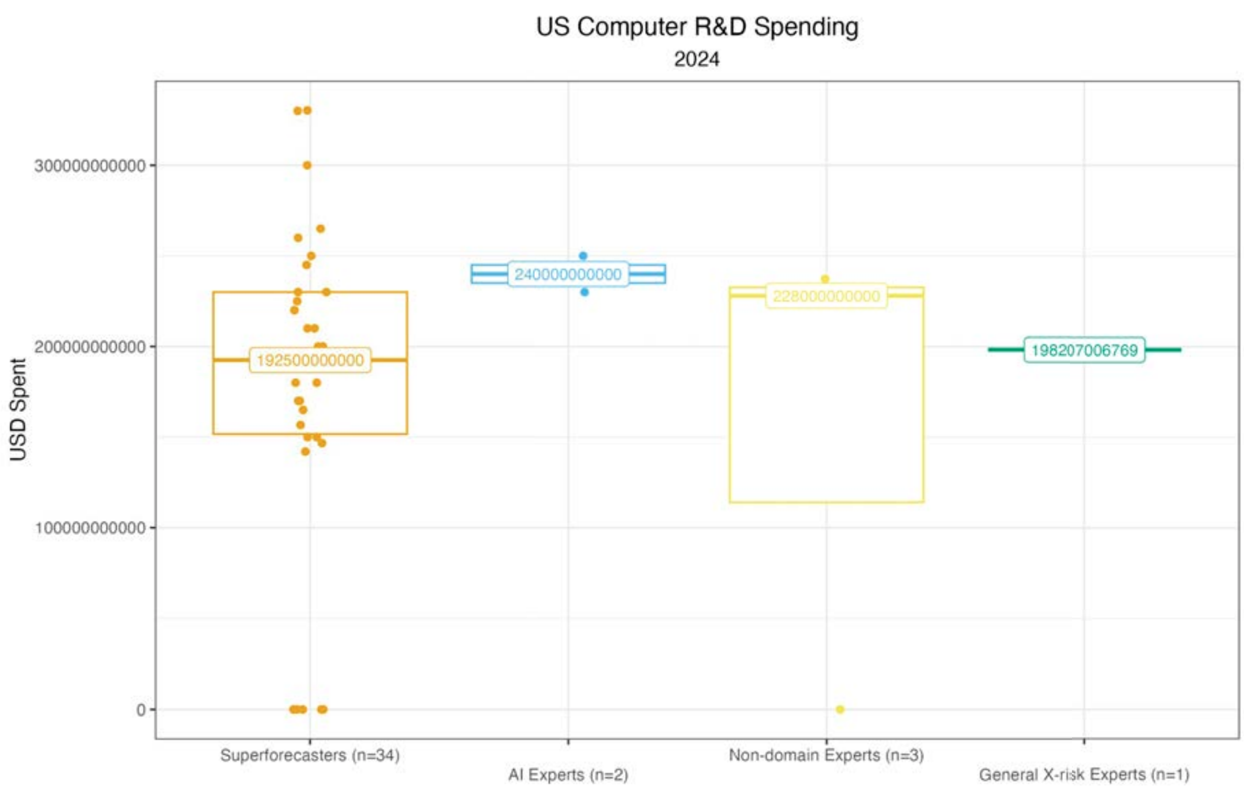
[1087] Numbers of forecasters are given as of Stage 4 of the XPT.

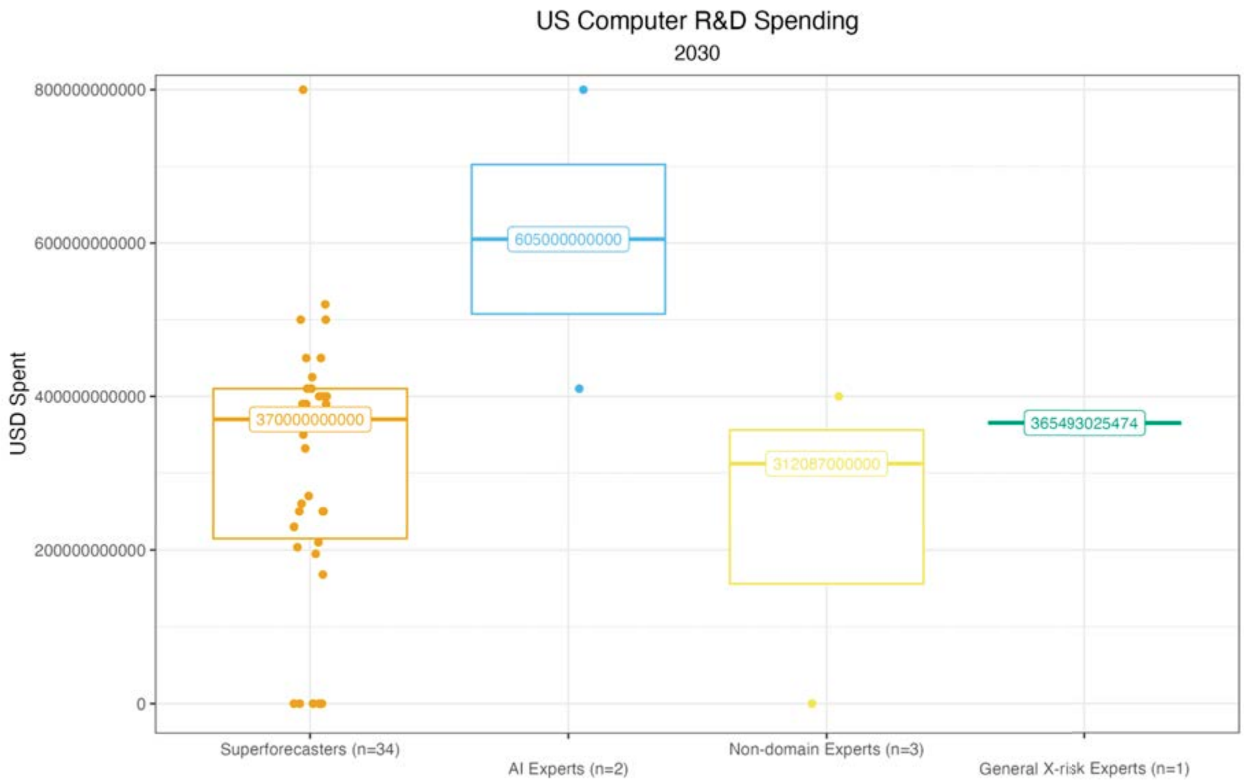
	2050	$8,707,784,090,370	$8,707,784,090,370	n/a	n/a
Non-Domain Experts (N = 4)	2024	$237,256,000,000	$228,000,000,000	n/a	n/a
	2030	$312,087,000,000	$312,087,000,000	n/a	n/a
	2050	$567,833,000,000	$567,833,000,000	n/a	n/a

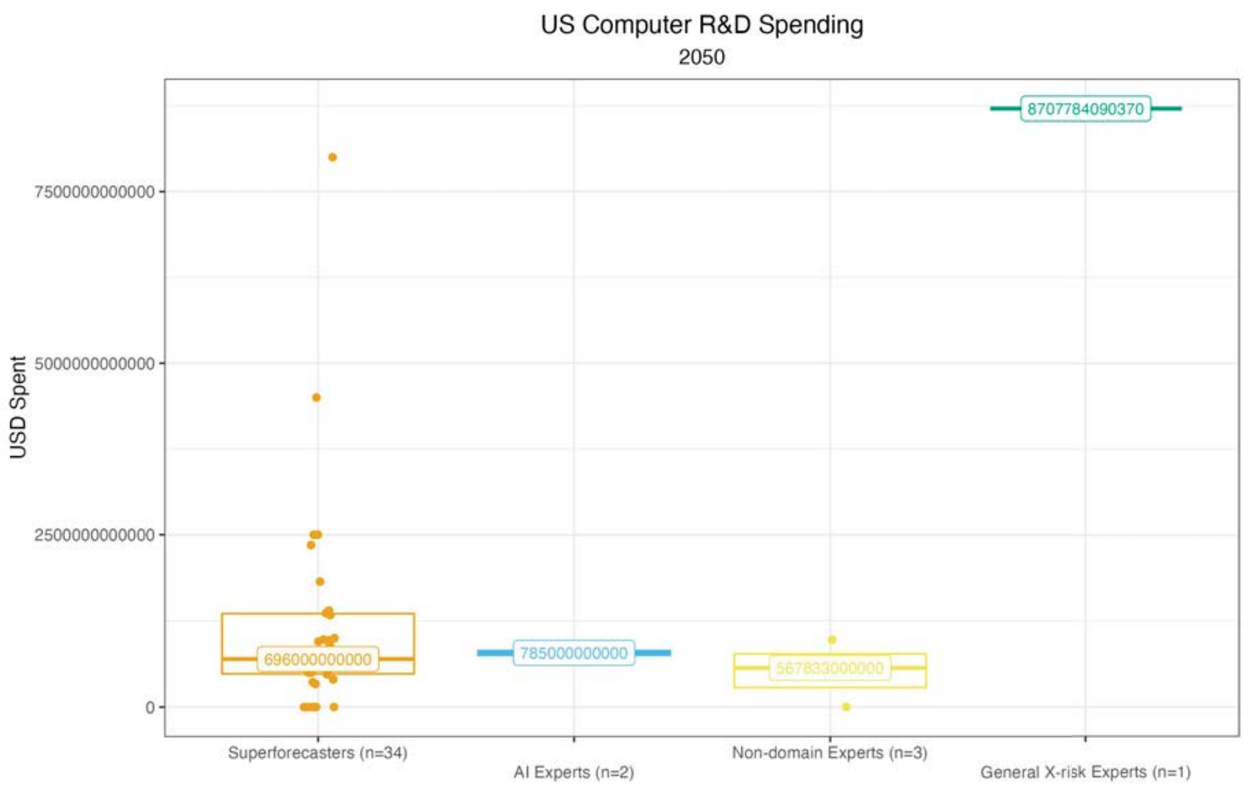

US Computer R&D Spending - 2024 - 50th %
220000000000
200000000000
180000000000
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 34)


US Computer R&D Spending - 2030 - 50th %
340000000000
300000000000
260000000000
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 34)


US Computer R&D Spending - 2050 - 50th %
Median
700000000000
600000000000
500000000000
400000000000
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 34)


US Computer R&D Spending
2024
USD Spent
300000000000
200000000000
100000000000
0
192500000000
240000000000
228000000000
198207006769
Superforecasters (n=34)
AI Experts (n=2)
Non-domain Experts (n=3)
General X-risk Experts (n=1)


US Computer R&D Spending
2030
USD Spent
800000000000
600000000000
400000000000
200000000000
0
605000000000
370000000000
312087000000
365493025474
Superforecasters (n=34)
AI Experts (n=2)
Non-domain Experts (n=3)
General X-risk Experts (n=1)


US Computer R&D Spending
2050
USD Spent
7500000000000
5000000000000
2500000000000
0
8707784090370
696000000000
785000000000
567833000000
Superforecasters (n=34)
AI Experts (n=2)
Non-domain Experts (n=3)
General X-risk Experts (n=1)

Sources of agreement, disagreement and uncertainty

- How to extrapolate historical data
  - Several teams reported disagreement about how to extrapolate the historical trends in this category of R&D spending. Some forecasters continued a linear projection, others adjusted to incorporate factors perceived to impact spending. One team noted uncertainty in how long trends were likely to continue.[1088]
- The effects of inflation
  - Many teams noted uncertainty in future trends in inflation, and suggested that this could be an important factor in this question.[1089]
  - There was some uncertainty in whether the question should account for inflation,[1090] with one team suggesting it would be better if the question asked about inflation-adjusted estimates.[1091]
- The effects of taxation
  - Some teams noted that taxation rules could influence this question, including changes in tax credits or subsidies.[1092]
- The impact of AI
  - One team cited differences in opinion on the long-term impact of AI was a major source of difference in forecasts.[1093]
- The impact of catastrophes
  - Two teams noted that a catastrophe could have a significant effect on these estimates.[1094]

Arguments given for forecasts below the median (50th percentile for 2024: 195 billion, 2030: 350 billion, 2050: 610 billion)

- Decline in proportion of R&D conducted by industry

[1088] 341, "[T]here is a big disagreement about whether the standard deviations calculated for R and D spending over the last ten years are useful in providing guidance for forward projecting uncertainty into the future." 344, "Most of the forecasters are using linear trend lines to get to the "projected" median numbers, while limited numbers are using exponential trend lines to fine the future numbers." 336, "Uncertainty about how long trends continue."

[1089] 340, "A major source of uncertainty for almost all forecasts, especially in the near term, was the effect of inflation on the recorded values." 341, "There is also dissent about whether R&D spending dropped in inflationary periods, with econometric data suggesting R&D spending did not drop in the 1970s, a period of inflationary pressures."

[1090] 336, "Uncertainty about whether the figures are adjusted for inflation or not."

[1091] 341, "Standard deviation could get increased pretty badly by wild cards, likely the worst wild card being inflation. I wish this would be scored in inflation-adjusted dollars."

[1092] 337, "My guess is that if you asked an accountant this question he would say it depends on the tax system Not so much how much is really spent, but how you can put costs into categories that save you tax or give you subsidies." 341, "Wildcards: [...] US gov't increases R&D tax credit to offset."

[1093] 340, "The major difference in opinion between forecasts is in the long term impact of AI on these growth rates."

[1094] 344, "Activation of catastrophic and extinction risk may push the numbers down to 0" 342, "These numbers would be affected by the onset of global catastrophe before the stated dates, which would possibly reduce funding being directed towards these industries."

- One team noted barriers to research forming a large portion of private company spending, noting short-term performance requirements and difficulties in making research profitable, and historical decline of R&D being undertaken by business.[1095]
- Expectations of slowed growth in computer R&D
  - Some lower forecasts were developed using estimates of GDP growth and of the proportion of GDP that R&D and specifically computer R&D would comprise.[1096]

Arguments given for forecasts above the median (50th percentile for 2024: 195 billion, 2030: 350 billion, 2050: 610 billion)

- High levels of inflation
  - Several forecasters suggested that rates of inflation would remain high, increasing spending on R&D.[1097]
- Financial benefits of R&D
  - Some forecasters cited the role of R&D in driving profits and creating value. One also suggested that R&D spending being exempt from tax would be a driver of increased spending.[1098]
- Historically high rates of growth in these sectors
  - Several forecasters noted recent high rates of growth in this category of spending and suggested that this would remain high.

[1095] 341, "[H]istorical declining R&D% undertaken by business. The past three decades have been marked by a growing division of labor between universities focusing on research and large corporations focusing on development." 341, "Research in corporations is difficult to manage profitably." 341, "As a result, research inside companies can only survive if insulated from the short-term performance requirements of business divisions."

[1096] 340, "The second [forecaster with a lower end forecast] expects near term growth rates in these sectors to be lower than what is shown in the historical data, and that overall GDP growth will not be high enough to compensate. The lowest estimate assumed stalling growth in the near term followed by recovery, while the highest estimate assumed continued growth close to current rates. It expects growth rates to fall from 10% to closer to 7.5% by the final decade of this question, so that computer R&D would not be >60% of all R&D and an unlikely proportion of overall GDP (as no leap in GDP growth was assumed)." 337, "Stronger arguments for the lower end acknowledge the above-trend growth of the two categories of "Information" and "Compute system design" industries. They attempted to estimate what they see as likely structural limits on how much this category can be as a percentage of gross domestic product (GDP). As of year-end 2021, United States Gross Domestic Product (GDP) was approximately 24 trillion. https://www.bea.gov/news/2022/gross-domestic-product-fourth-quarter-and-year-2021-advance-estimate Therefore, as of most recently available, this is about 0.5% of gross domestic product. With this in mind, while this category can grow above GDP trend, there is likely some structural limit to growth above a certain percentage."

[1097] 340, "This second forecast also cites inflation as a cause of growth in these numbers in the near term" 336, "Inflation has picked up and will probably remain high for at least a couple of years." 344, "I expect R&D spending to generally go up because of inflation and the lack of other ways to create value in a globally competitive economy."

[1098] 341, "Large project management can turn R&D into revenue at triple today's rate? R&D is tax free? R&D expenditures have a 5X impact on dividends and stock price. SG&A and Operational expenses can now qualify as R&D? " 344, "I expect R&D spending to generally go up because of inflation and the lack of other ways to create value in a globally competitive economy."

- AI-driven growth
  - Several forecasters argued that AI would result in growth in these categories, with one citing the CHIPS act as likely to increase spending initially and that this would result in increased high growth.[1099]
- War as a driver of increase in R&D spending
  - One team suggested that the possibility of a new war and great power competition may result in higher spending on R&D.[1100]

Cross-references with other questions

Q36: US GDP From Software

Question 38: Labor Force Participation

What will be the labor force participation rate in OECD countries…
…in the year 2024?
…in the year 2030?
…in the year 2050?

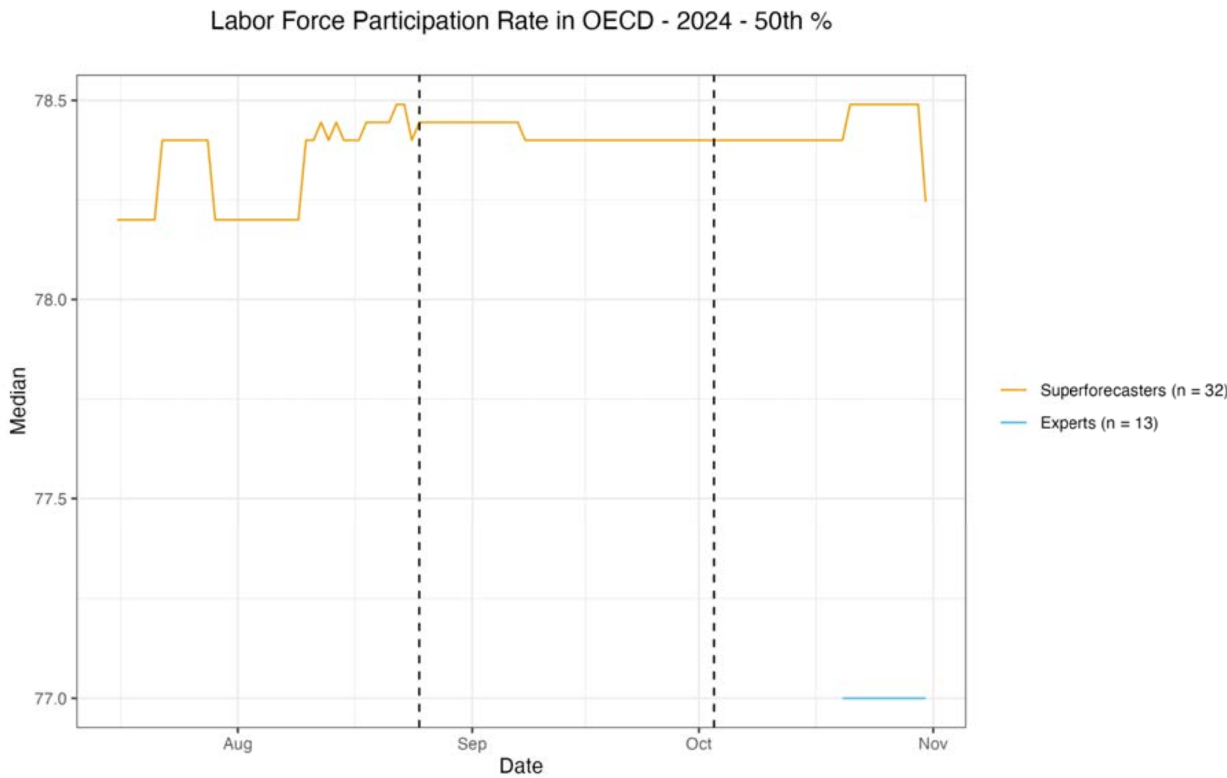
Question and resolution details, prior forecasts, and other relevant sources

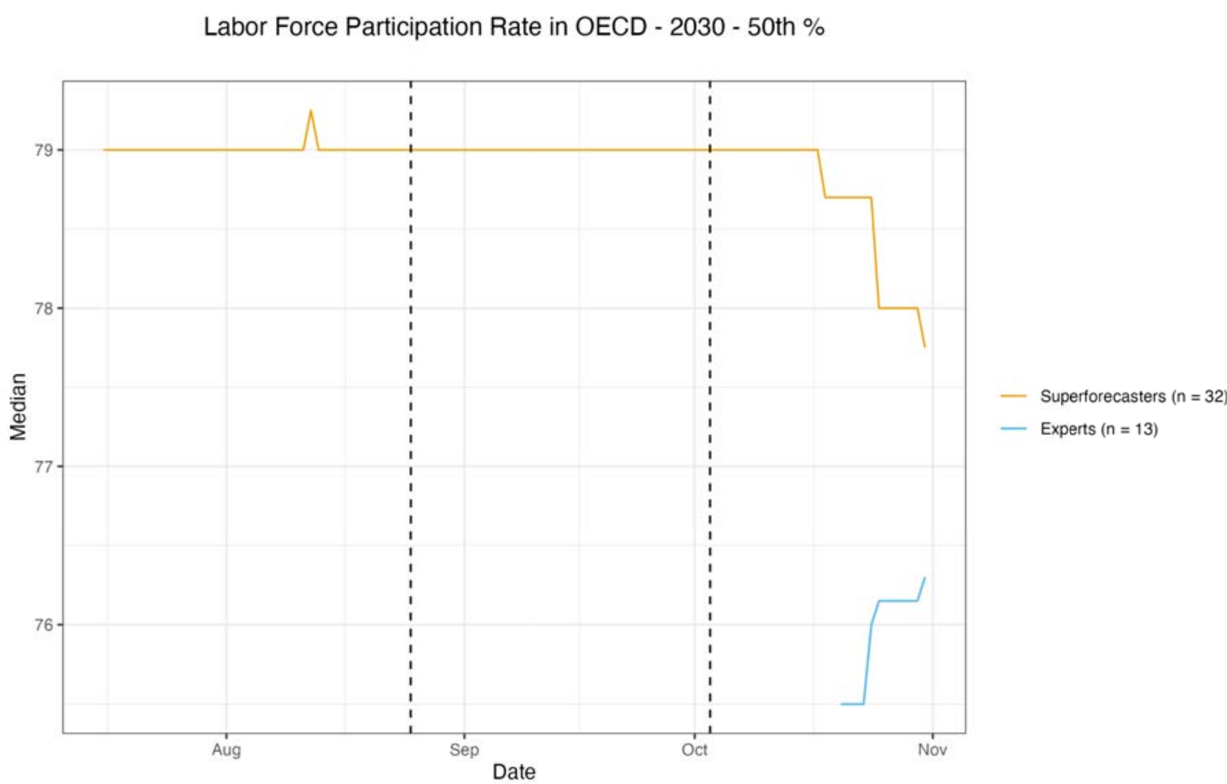
Results[1101]

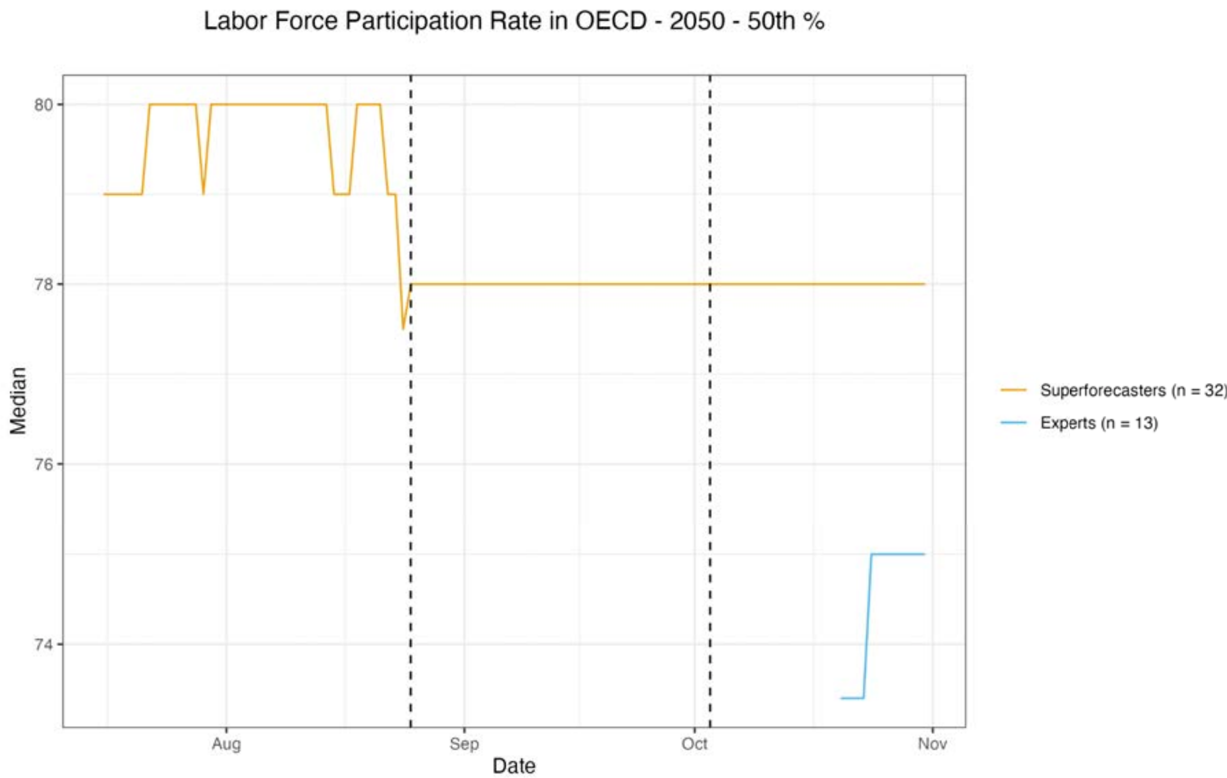
Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-	2024	78%	78.25%	6.68	-67.91%

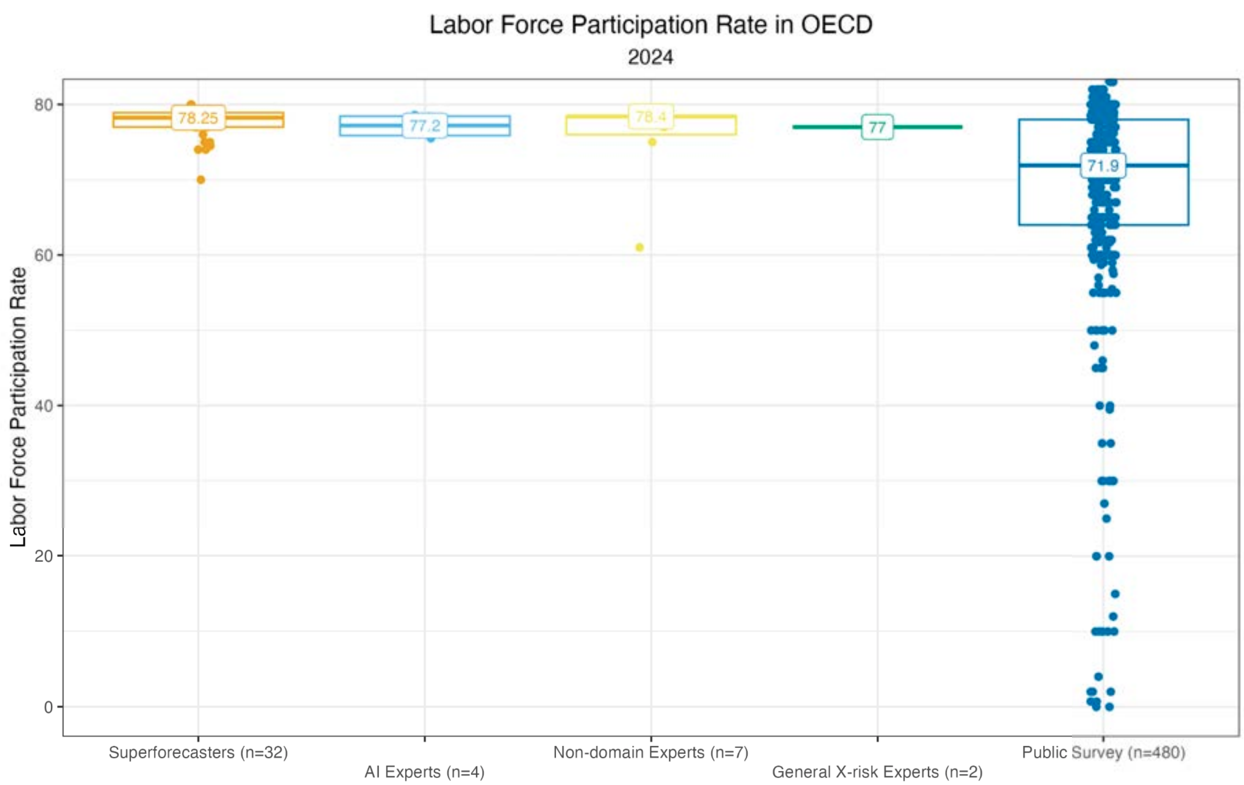
[1099] 340, "Both forecasts above the median shared a similar line of reasoning: Advances in AI would fuel significant growth in the selected sectors within the time horizon of this question, with the largest acceleration happening between 2030 and 2050." 339, "The forecasts at the high end of the range factored in the impact of the CHIP bill recently passed by US Congress as a driving force for the early years, but also assumed it would set the stage for continued growth in the out years as well."
[1100] 341, "A major driver is war. We have a new war and increased great power competition."
[1101] Numbers of forecasters are given as of Stage 4 of the XPT. Public survey respondents were surveyed outside of the tournament context.

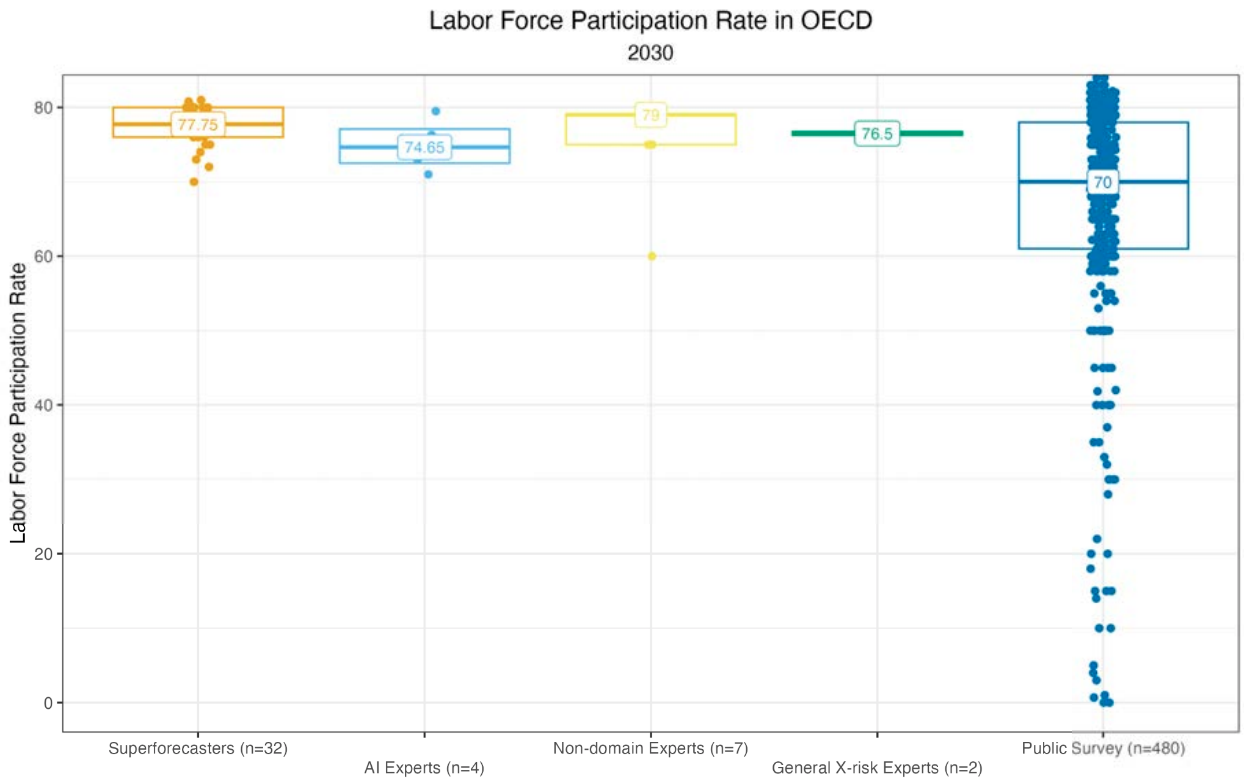
Forecasters (N = 32)	2030	79%	77.75%	8.37	-67.98%
	2050	80%	78%	11.52	-54.13%
Domain Experts (N = 4)	2024	78.4%	77.2%	1.73	-7.61%
	2030	76.3%	74.65%	4.29	-12.92%
	2050	71.8%	70.9%	4.76	+111.66%
General X-Risk Experts (N = 2)	2024	n/a	77%	n/a	n/a
	2030	n/a	76.5%	n/a	n/a
	2050	n/a	77.5%	n/a	n/a
Non-Domain Experts (N = 7)	2024	75%	78.4%	n/a	n/a
	2030	75%	79%	n/a	n/a
	2050	75%	75%	n/a	n/a
Public Survey (N = 480)	2030	71.9%		64244536.02	-
	2050	70%		66466003.67	-
	2100	69%		73796168.54	-


Labor Force Participation Rate in OECD - 2024 - 50th %
78.5
78.0
77.5
77.0
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 32)
Experts (n = 13)

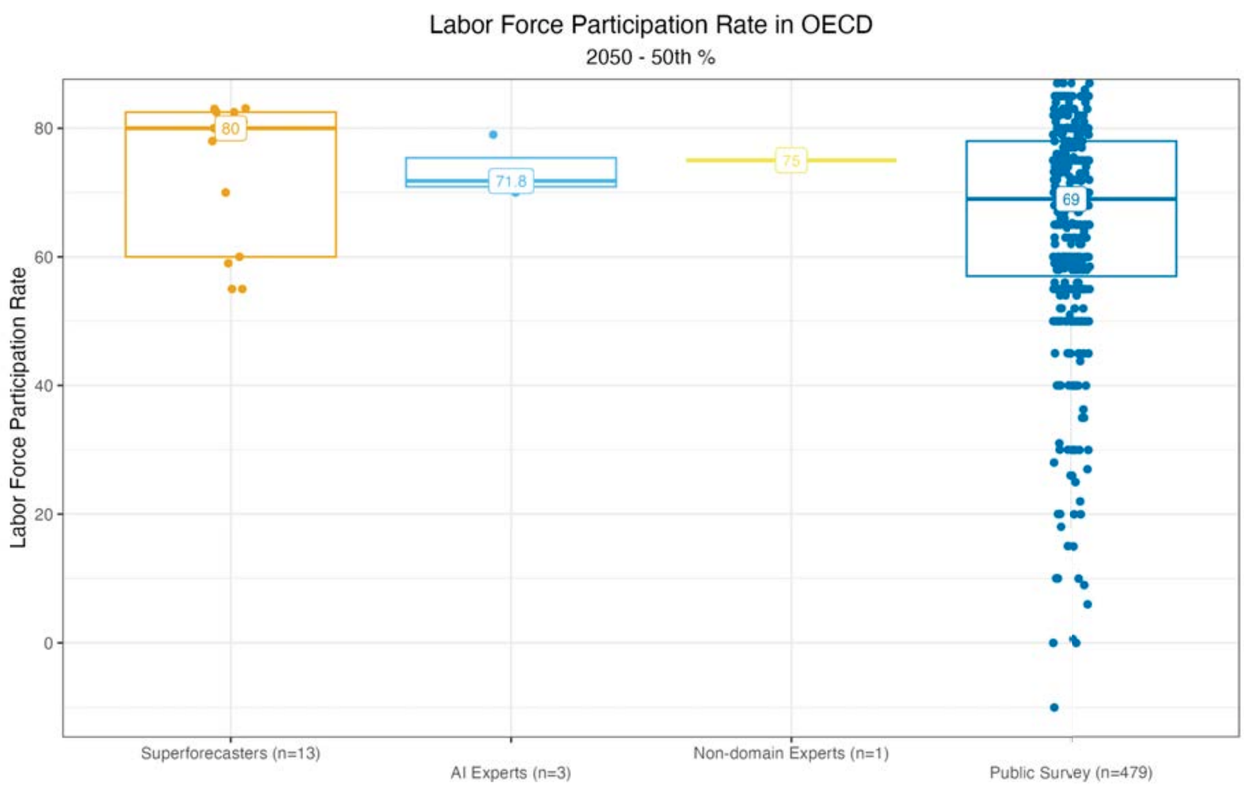

Labor Force Participation Rate in OECD - 2030 - 50th %
79
78
77
76
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 32)
Experts (n = 13)


Labor Force Participation Rate in OECD - 2050 - 50th %
80
78
76
74
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 32)
Experts (n = 13)


Labor Force Participation Rate in OECD
2024
80
60
40
20
0
Labor Force Participation Rate
78.25
77.2
78.4
77
71.9
Superforecasters (n=32)
AI Experts (n=4)
Non-domain Experts (n=7)
General X-risk Experts (n=2)
Public Survey (n=480)



Labor Force Participation Rate in OECD
2030
Labor Force Participation Rate
77.75
74.65
79
76.5
70
Superforecasters (n=32)
AI Experts (n=4)
Non-domain Experts (n=7)
General X-risk Experts (n=2)
Public Survey (n=480)



Labor Force Participation Rate in OECD
2050 - 50th %
Labor Force Participation Rate
80
71.8
75
69
Superforecasters (n=13)
AI Experts (n=3)
Non-domain Experts (n=1)
Public Survey (n=479)

Sources of agreement, disagreement and uncertainty

Sources of agreement

- Most teams agreed that labor force participation was unlikely to change substantially before 2024, and most retained reasonably narrow estimates for 2030 and wider estimates for 2050.

Sources of disagreement

- The impact of AI and technology on labor force participation
    - Several teams noted that much of the differences between forecasts was due to differences in beliefs about the likely impact of AI and other technologies.[1102]
- The possible effects of a universal basic income
    - Forecasters differed in whether universal basic income (UBI) policies would become widespread in OECD countries, and what effect they would have on unemployment. Uncertainty in how UBI would be implemented (e.g. conditional or unconditional) was also noted, and different formats were expected to have different effects on labor force participation.[1103]
- The possibility of catastrophe or extinction
    - Some teams noted disagreement about the likelihood of catastrophic or existential risks, with some forecasters believing this to have a meaningful effect on expected labor force participation, especially in 2050.[1104]
- Changes in the structure of employment
    - Teams also noted uncertainty in patterns of employment in the future, for example the classification of participation in the "gig economy", or changes in age ranges for employment statistics.[1105]
- Changes in the composition of OECD

[1102] 343, "Much of the variability in responses among our team may be explained by this difference in probability estimates for AGI." 340, "The extent to which technology will replace human labour and the extent to which jobs lost to technology will be replaced by new jobs were the most discussed topics among the group. Differences in opinion on these topics account for most of the differences in estimates (along with some mentions of aging/retirement age/years of education)." 341, "One big source of uncertainty is whether AI will remove entire jobs or just parts of jobs. Early reports around AI job replacement suggested that whole jobs would disappear, while recent reports suggest it's more parts of jobs."

[1103] 343, "While it is also possible that political solutions such as a UBI could lead to people voluntarily leaving labor markets, this is dependent on 1.) broad implementation of UBI in OECD countries, and 2.) the nature of that UBI. Modeling suggests that UBI will negatively impact LFPR, and many in the field appear to accept this as a given (potentially due to observations from Alaska). However, the US implementation of a conditional program (EITC - which requires labor-force participation) had the opposite effect, increasing LFPR."

[1104] 344, "Wild card scenarios: Humankind's extinction due to activation of existential risk. Human civilization is impaired due to the activation of catastrophic risk." 336, "For 2050, there was disagreement about the impact of automation and possibly existential catastrophe."

[1105] 344, "Sources of disagreement and uncertainty: [...] The likelihood of a structural change in how employment functions in the future (the classification of gig economy)." 337, "The OECD could adjust the age range of this stat: 16 to 64 already looks a bit out of date, as in many countries education can continue until 23 or 24."

    - One team noted that the entry of different countries into the OECD might have a significant effect on labor force participation in this group of countries, as may patterns of migration to OECD countries, due to differences in cultural expectations about who should participate in work.[1106]

Arguments given for forecasts below median of 78% (2024), 77% (2030), 78% (2050)

- AI and automation replacing human labor
    - The possibility of AI making many current jobs obsolete was a common argument used in support of lower forecasts, particularly for the 2050 forecasts.[1107]
    - However, many teams acknowledged substantial uncertainty about how AI and automation would affect employment, with some noting that previous technologies had not led to widespread unemployment.[1108]
- Implementation of UBI will reduce labor force participation
    - Some forecasters assumed that UBI programs would be common in OECD countries in the next few decades.[1109]
- Existential or catastrophic risks reduce the expected labor force participation
    - Several forecasters lowered their forecasts for 2050 due to the possibility of extinction or other catastrophe.[1110]

[1106] 337, “The composition of the OECD could look very different by 2050, if say India joined.” 337, “In some societies, having women working outside the home is less acceptable than others, but in few OECD countries is this a big factor, except for women with preschool-aged children.”
[1107] 340, “The strongest reason given for forecasts of low participation rates was significant replacement of human labour with technology and a UBI scheme which means individuals no longer seek work.” 344, “Automation brought on by AI development (specialized AI development which is more probable)” 339, “One key argument for the decline of the rate is the impact of automation. On the one side there is consumer-induced labour (such as self-checkouts) that could become more prevalent. And more importantly that Transformative Artificial Intelligence (TAI) (and later General AI (AGI)) will take up more and more work (implied: pushing people out of jobs).” 343, “Any drop below long-term trends would likely be explained by AI-driven permanent unemployment.”
[1108] 343, “Permanent technological unemployment has not been observed with other non-intelligence technological innovations in the past. Displacement has occurred, though, so we can expect displacement to be linked with some degree of temporary (5-10 years) drops in LFPR as workers find new employment opportunities/fields. However, rapid displacement is often tempered by political action to slow adoption and diminish the effects of displacement.” 341, “The data is mixed on the impact of IT and AI on jobs.”
[1109] 341, “Universal basic income programs, which would be likely to be in place by then, would lower participation rates.” 340, “The strongest reason given for forecasts of low participation rates was significant replacement of human labour with technology and a UBI scheme which means individuals no longer seek work.”
[1110] 339, “2 out of 7 forecasters had x-risk (forecasting the rate = 0%) at the 5th %, dragging down the lower end for 2050.” 336, “There is a small but credible possibility of extinction by 2050 (> 5%, < 25%), so participation might be 0”

Arguments given for forecasts above median of 78% (2024), 77% (2030), 78% (2050)

- Difficulty of automating some jobs
    - One team noted that some jobs will be difficult to automate, for example jobs that benefit from human connection, or jobs that are complex or require significant dexterity.[1111]
- Expectations of job change rather than job loss
    - Several teams suggested that automation is more likely to result in new jobs being created, rather than simply leading to unemployment.[1112] One team suggested that rather than replacing jobs entirely, there is some evidence that AI or automation may replace parts of jobs.[1113]
- Delayed retirement
    - Several forecasters suggested that more people will choose to delay their retirement, possibly due to improving population health.[1114]
- Aging populations resulting in greater demand for labor
    - Two teams suggested that an aging population will result in greater demand for younger workers, particularly caretakers.[1115]
- Increasing female labor force participation
    - Some teams noted that female workforce participation has been trending upwards and this should be expected to continue, particularly for lower-income countries within the OECD. It was also suggested that increased flexibility of working hours triggered by the COVID-19 pandemic may also lead to greater female workforce participation.[1116]

[1111] 340, “Several barriers to automation of labour were raised by forecasters: a. Robotics is difficult and expensive. Manufacturing employment has also been falling for a long time.b. Large numbers of service jobs (healthcare, leisure) etc. are difficult to replace with algorithms or robots as they require unpredictable interaction, complex tasks and human connections. In several sectors such as healthcare, jobs are protected by regulations that may take decades to change. (The service sector is 50% of US employment).”

[1112] 340, “The creation of new types of work. It is possible that in the 2050 timeframe that work may be defined differently than it is now, but it seems likely that most people will continue doing something to earn extra income.” 341, “The timing and effects of AI are totally unknown, perhaps AI will lead to the creation of jobs not foreseen.” 337, “The standard industrial revolution dynamic, in which more jobs are created than lost, is likely to prevail.”

[1113] 341, “One big source of uncertainty is whether AI will remove entire jobs or just parts of jobs. Early reports around AI job replacement suggested that whole jobs would disappear, while recent reports suggest it's more parts of jobs.”

[1114] 337, “Many people, particularly women and manual workers, retire before the age of 65, but retirement ages are occurring later and if this continues, the labor participation rate should increase.” 339, “People working longer (in part due to better health care).” 340, “A smaller percentage of workers retiring before the age of 65, which is the age group cutoff for this question.”

[1115] 336, “With ageing populations, countries will likely try to increase workforce participation in the 25 to 64 year-old demographic as much as possible” 339, “Increased demand for care-takers due to ageing populations.”

[1116] 337, “If the types of work hours that were common during COVID-19 continue, more women could join the workforce. In the UK, the commuter rush hours aren't nearly as busy and the long hours that were

- Continuation of existing trends
    - Some forecasters assumed that the existing upward trend of labor force participation would continue, especially in lower-income countries as their economies continue to develop.[1117]

Question 39: MATH Dataset Benchmark

What will be the state-of-the-art accuracy of a machine-learning model on the MATH Dataset…
...by June 30, 2024?
...by June 30, 2030?

Question and resolution details, prior forecasts, and other relevant sources

Results[1118]

Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 32)	2024	31	71	5521.39	-99.65%
	2030	64.5	85	18425.72	-99.91%
Domain Experts (N = 9)	2024	80	80	9.76	-1.36%
	2030	94	94	11.52	-11.45%

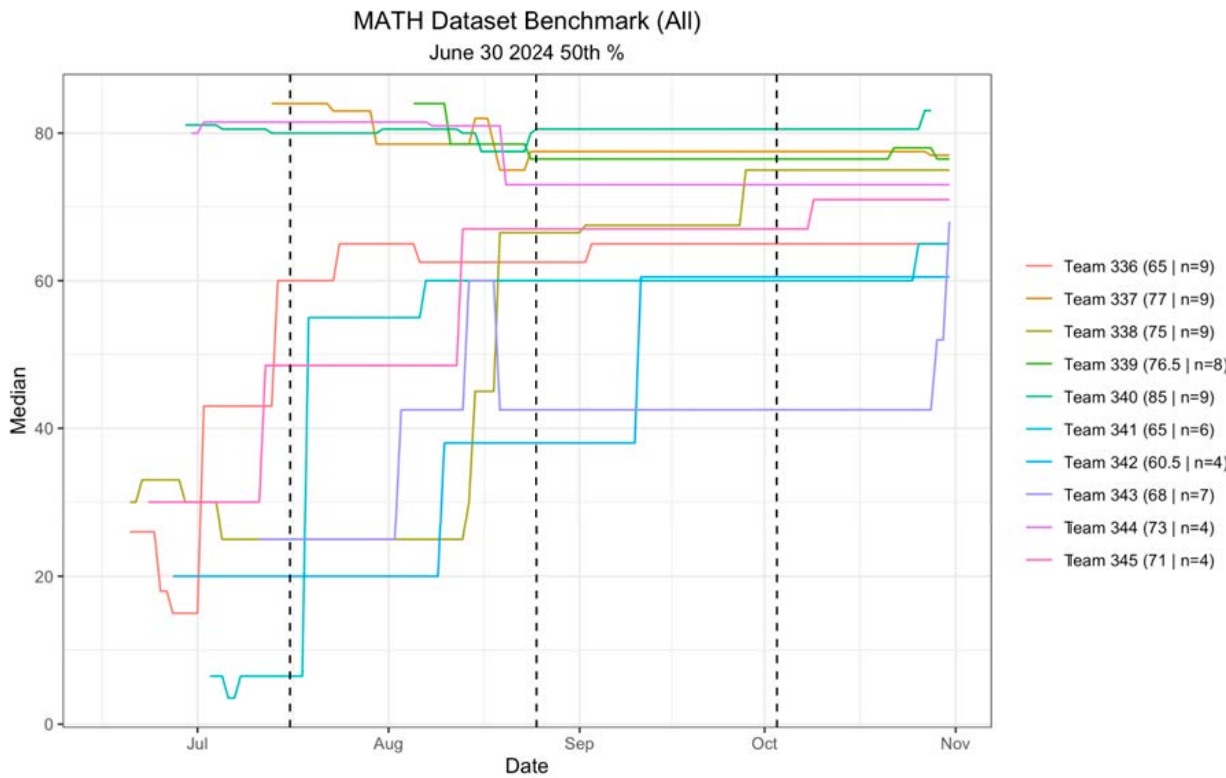
the norm among hospitality workers are being shortened and fitted around school times." 339, "Women participation rate increasing further."

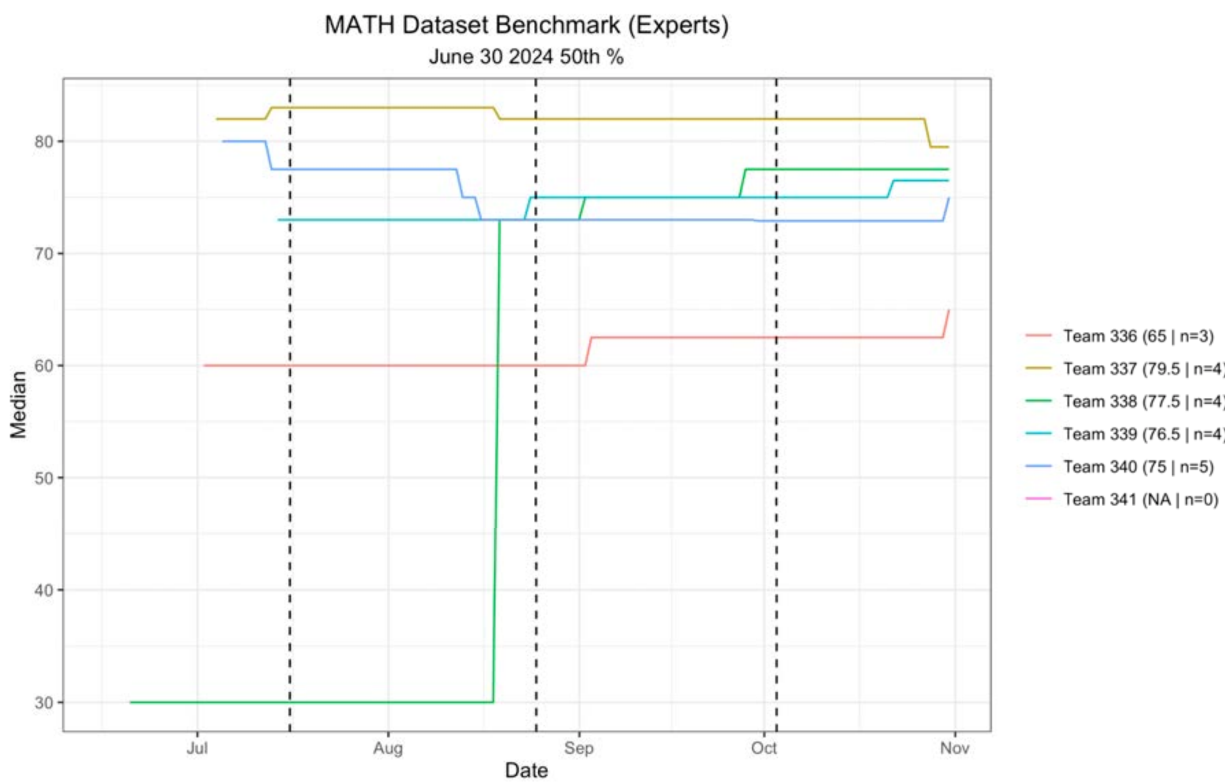
[1117] 341, "The trend has been up for over 60 years, and is expected to continue to rise for ten more years." 336, "The trend to date has been a slight increase over time, the higher estimates assume a continuation of this trend until 2050, although the highest estimates are not much higher than the median group estimates" 341, "Turkey and Mexico have low participation rates now, but may increase as they catch up to the rest of the OECD. They also have large growing populations."

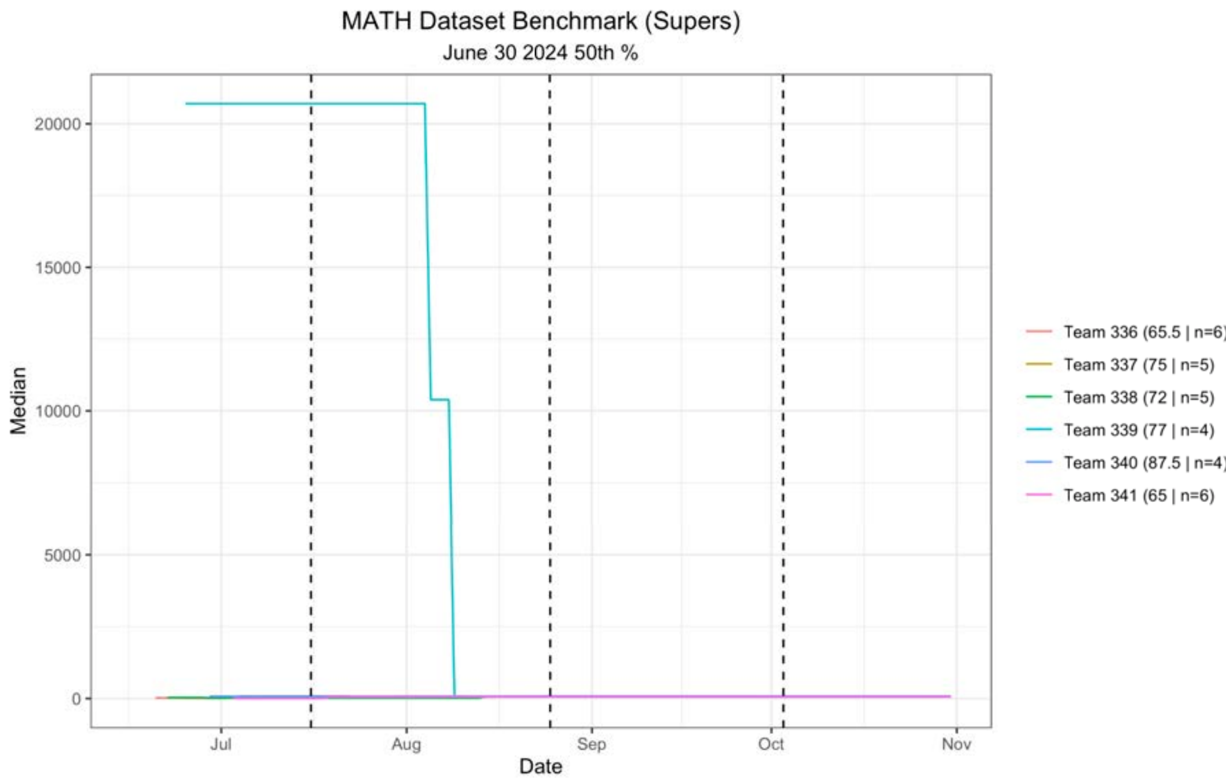
[1118] Numbers of forecasters are given as of Stage 4 of the XPT.

General X-Risk Experts (N = 4)	2024	75	75	NA[1119]	NA
	2030	90	87.5	NA	NA
Non-Domain Experts (N = 7)	2024	30	75	NA	NA
	2030	45	91	NA	NA

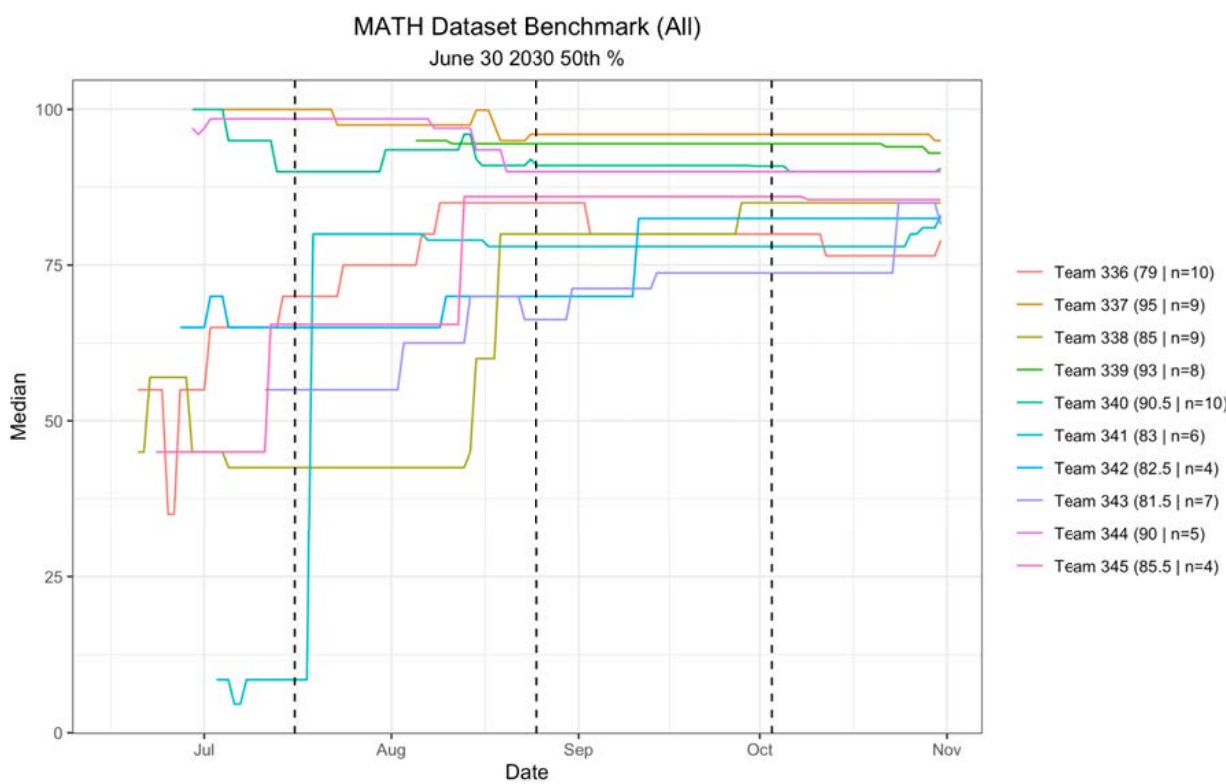
[1119] Only one forecaster in this group answered this question in Stage 1.


MATH Dataset Benchmark (All)
June 30 2024 50th %
Median
Date
Jul
Aug
Sep
Oct
Nov
0
20
40
60
80
Team 336 (65 | n=9)
Team 337 (77 | n=9)
Team 338 (75 | n=9)
Team 339 (76.5 | n=8)
Team 340 (85 | n=9)
Team 341 (65 | n=6)
Team 342 (60.5 | n=4)
Team 343 (68 | n=7)
Team 344 (73 | n=4)
Team 345 (71 | n=4)

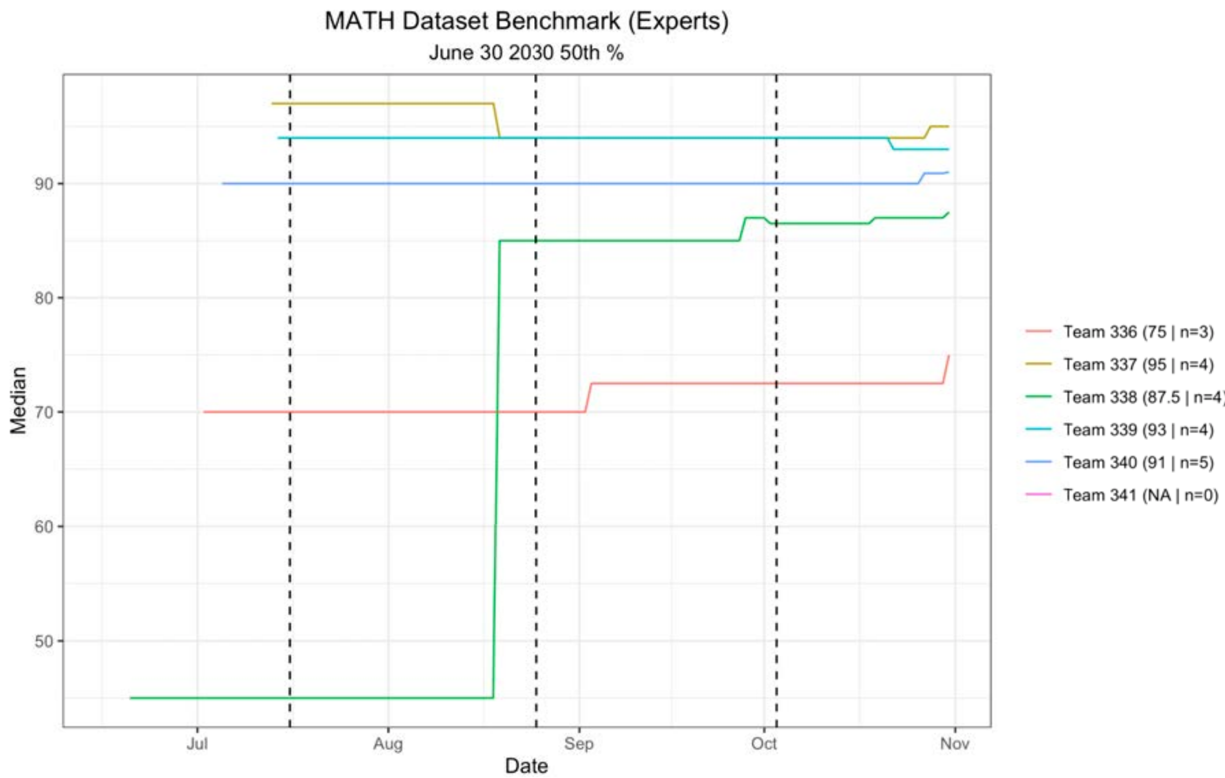

MATH Dataset Benchmark (Experts)
June 30 2024 50th %
Median
Date
Jul
Aug
Sep
Oct
Nov
30
40
50
60
70
80
Team 336 (65 | n=3)
Team 337 (79.5 | n=4)
Team 338 (77.5 | n=4)
Team 339 (76.5 | n=4)
Team 340 (75 | n=5)
Team 341 (NA | n=0)

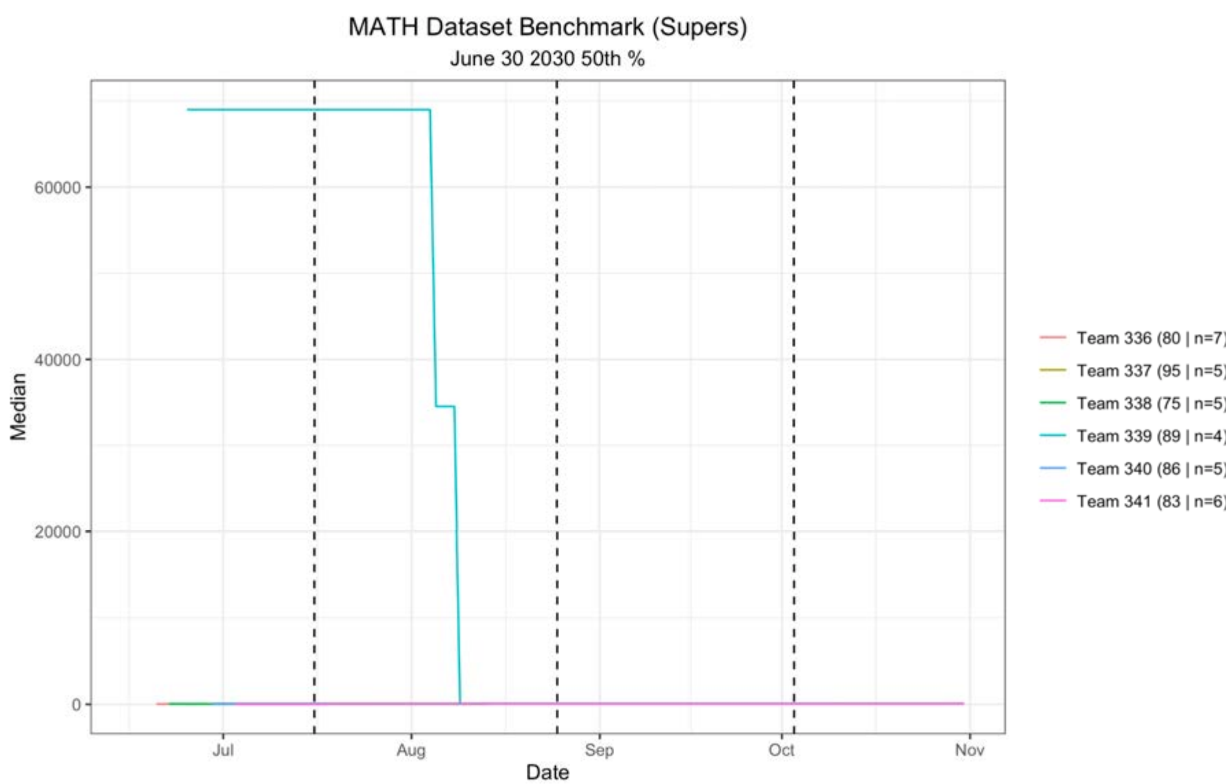


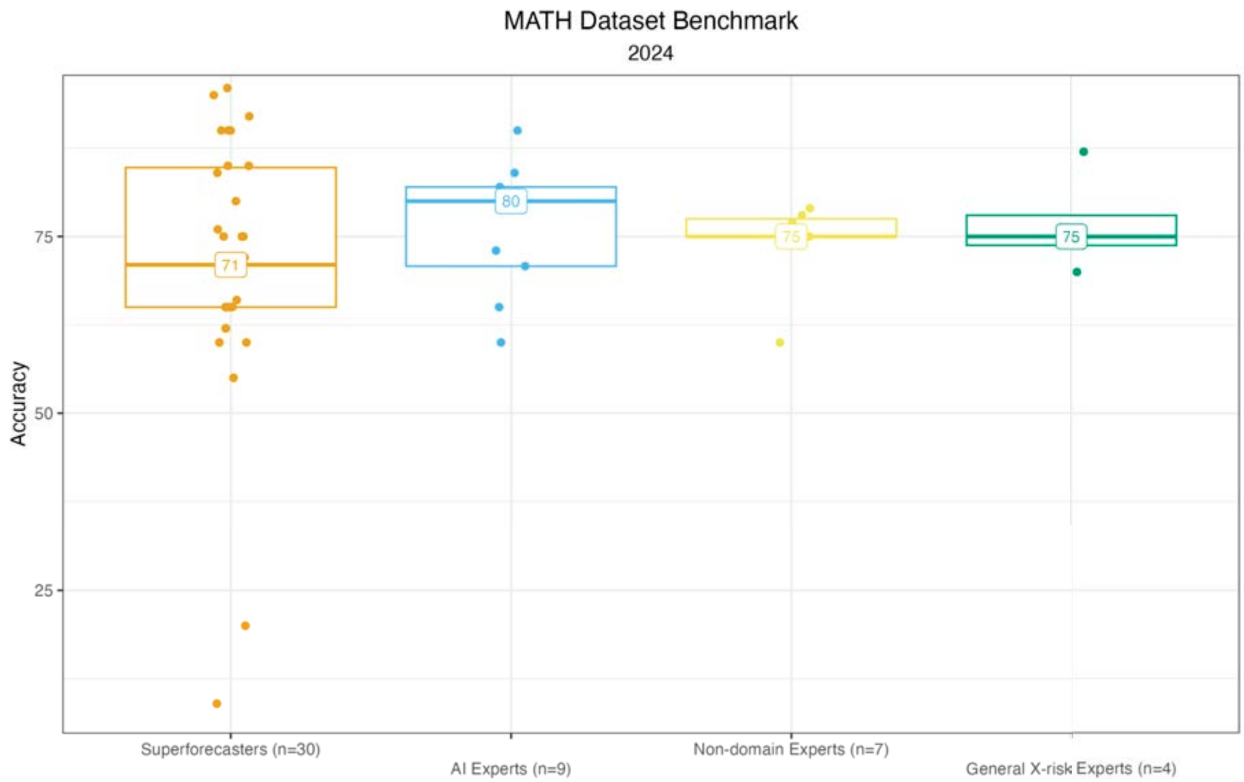
MATH Dataset Benchmark (Supers)
June 30 2024 50th %
Median
20000
15000
10000
5000
0
Jul
Aug
Sep
Oct
Nov
Date
Team 336 (65.5 | n=6)
Team 337 (75 | n=5)
Team 338 (72 | n=5)
Team 339 (77 | n=4)
Team 340 (87.5 | n=4)
Team 341 (65 | n=6)

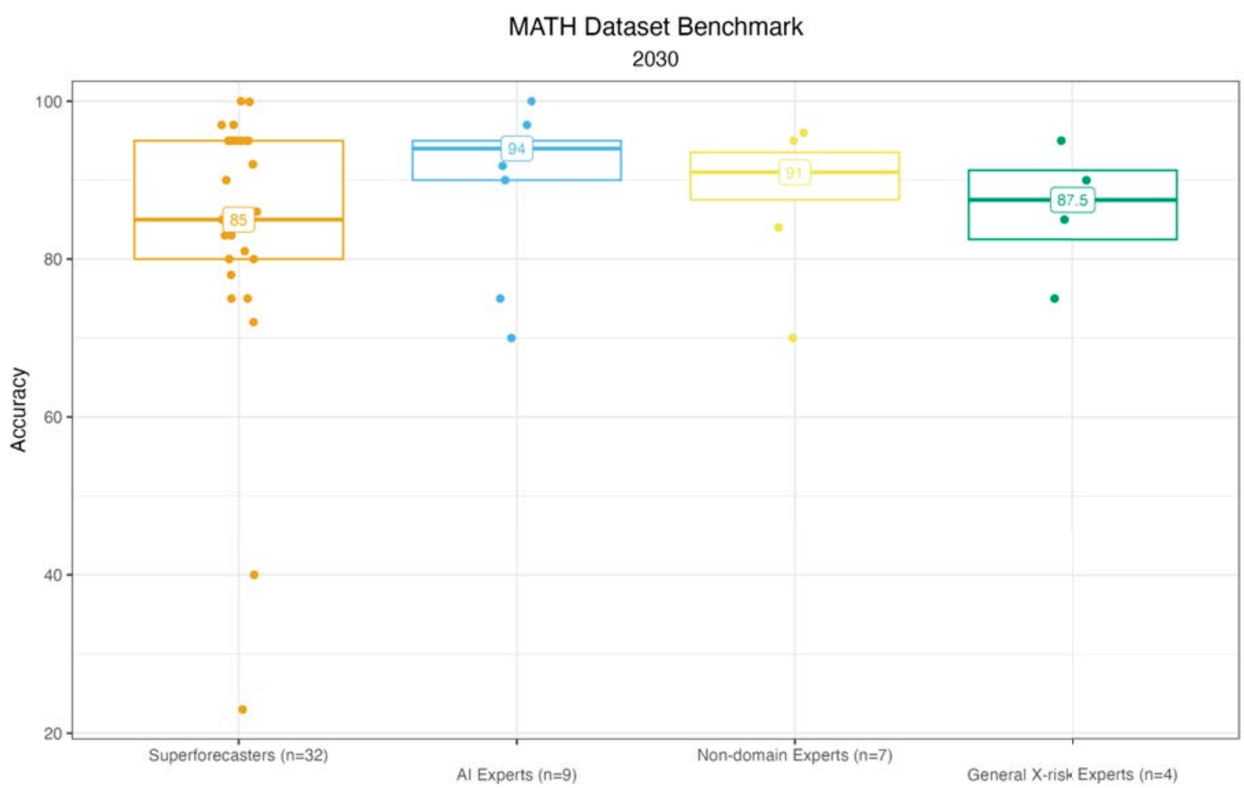


MATH Dataset Benchmark (All)
June 30 2030 50th %
Median
100
75
50
25
0
Jul
Aug
Sep
Oct
Nov
Date
Team 336 (79 | n=10)
Team 337 (95 | n=9)
Team 338 (85 | n=9)
Team 339 (93 | n=8)
Team 340 (90.5 | n=10)
Team 341 (83 | n=6)
Team 342 (82.5 | n=4)
Team 343 (81.5 | n=7)
Team 344 (90 | n=5)
Team 345 (85.5 | n=4)


MATH Dataset Benchmark (Experts)
June 30 2030 50th %
Median
Date
Team 336 (75 | n=3)
Team 337 (95 | n=4)
Team 338 (87.5 | n=4)
Team 339 (93 | n=4)
Team 340 (91 | n=5)
Team 341 (NA | n=0)


MATH Dataset Benchmark (Supers)
June 30 2030 50th %
Median
Date
Team 336 (80 | n=7)
Team 337 (95 | n=5)
Team 338 (75 | n=5)
Team 339 (89 | n=4)
Team 340 (86 | n=5)
Team 341 (83 | n=6)


MATH Dataset Benchmark
2024
Accuracy
75
50
25
71
80
75
75
Superforecasters (n=30)
AI Experts (n=9)
Non-domain Experts (n=7)
General X-risk Experts (n=4)


MATH Dataset Benchmark
2030
Accuracy
100
80
60
40
20
85
94
91
87.5
Superforecasters (n=32)
AI Experts (n=9)
Non-domain Experts (n=7)
General X-risk Experts (n=4)

Sources of agreement, disagreement and uncertainty

- Some forecasters felt they didn't have enough domain-specific technical knowledge to make confident predictions.[1120]

Arguments given for lower scores

- Increasing model size might not be sufficient; if "low-hanging fruit" has already been harvested, future advances may be more difficult.[1121]
- Researchers might not focus major energy on this task, especially if it doesn't have obvious commercial payoffs.[1122]
- Achieving the last percentiles of improvement might be harder than earlier ones.[1123]

[1120] 345: "[A]ll team members struggled with this and other questions on AI. Comments included that AI questions:
"seem to require deep technical understanding of the landscape of the technology"
"I agree with this view. The only thing I can do in answering these questions is to use accessibility heuristics based on information found on google and other resources. I will admit that AI questions give me the most difficulty."
"I assumed that the organizers have made teams that include AI experts with a great deal of specialized knowledge."
Team Wiki Editors' note: "I feel unqualified to write this up, and had opted out of this [and other AI questions] I felt unable to forecast." |
336: "Lack of information and lack of specific knowledge are also a cause for great uncertainty. Specifically, DeepMind's Chinchilla paper more-or-less demonstrated that OpenAI's ~2018 "scaling laws" used too much compute and not enough data. (This is why we have 62B models that are about as good as the previous 540B models.) This means a relevant question is how much more relevant mathematical data we can find. We also should be open to systems that do very well on these math problems but somehow remain surprisingly narrow, perhaps because we stumble upon some data that is "close enough" to the test set that was created elsewhere and is still legal."
[1121] 343: "It's possible that future advances will not look like prior advances if all the 'low-hanging fruit' has been harvested. This could especially be the case by 2024 if future teams focus on algorithmic efficiencies or novel approaches that sacrifice short-term accuracy for longer-term model advances." |
337: "I guess I'm forecasting significantly lower because I think the progress so far has been to collect the low-hanging fruit."
342: "From this we can conclude that unless a completely new paradigm is created, improvement of accuracy is going to be much slower going forward, as simple solutions in optimisation of models for solving math problems have been tried and increasing model sizes is not as helpful and will slow down in the coming years due to reaching hardware limits"
[1122] 339: "The one plausible explanation that was offered was that no major team would be sufficiently interested in this particular problem space - given the training costs of current language models, corporate based teams might focus on areas with more immediate commercialization opportunities."
[1123] 341: "Perhaps the hardest 5% will take 95% of the time to get resolved by a combination of language models and automated reasoning. Compare this to driving, where AI is pretty good on average, maybe even human quality, makes fewer mistakes in the scenarios that show up thousands of times in its training data, but fails in rare scenarios." | 340: "On the downside, the last few percentages of performance are often really hard to squeeze out in ML, and some "difficulty plateau" can be reached. There is also a risk of potential heterogeneity of the dataset where some problems are not amenable to the current breed of AIs (analogy with proteins and AI: folding has seen great success, even though nobody understand why it works so well, but for protein-small molecules interactions all AIs at present fail)."

- This might be a more difficult problem than some other AI tasks that have seen improvement.[1124]

Arguments given for higher scores

- There has been rapid progress recently, most notably by the Minerva model.[1125]
- This will probably remain an area of active researcher effort.[1126]
- We might have AGI by 2030 (implying high performance on this benchmark by default).[1127]
- Some forecasters pointed out specific "low-hanging fruit" improvements that remain.[1128]

[1124] 345: "The class of problem seems different from other known AI challenges (chess, go, IBM Watson/Jeopardy), so seems like either accuracy will linger in the lowish range or there will be a breakthrough and the AI will converge to the best human." | 343: "Math has been quite a weakness of learning systems so far, so a rapid catching up was to be expected, especially with so much more calculating power. Further growth might be a lot slower. Furthermore, most math is based on basic logic. This AI can handle very well soon. But perhaps there is an extra bit that may prevent 100% in our timeframes. “It is only with a combination of both rigorous formalism and good intuition that one can tackle complex mathematical problems” (Source)" | 337: "Others expressed concern that the upper level problem would require true creativity, or that there might be a difficulty distribution for the word problems, and that the harder ones might need a much more trained network."

[1125] 338: "Rapid Progress with diminishing returns. User [forecaster name]: "I think that current Minerva performance will be the minimum expectation in 2024, but getting each marginal point will become more difficult. It could be that we see rapid improvements to something like 75%-85% and then relative stagnation as scores grind incrementally higher, in keeping with my general belief that rapid progress is easier to make in early stages and harder to make once the low-hanging fruit has been picked." | 339: "There is no reason to discount the possibility of achieving the highest levels of accuracy given the progression in the past year." | 341: "based on rapid progress to 50% accuracy in a year, I think good chance of 80% to 90% soon." | 345: "The strongest argument for the team’s median forecast is that Google has, in 2022, presented it's [sic] Minerva model which had an accuracy of 50.3%. This is a significant advance since 6.9% accuracy in June 2021." | 343: "Continued rapid improvement is supported by the past trajectories of other AI systems for image recognition, AlphaFold, etc." | 344: "Updated upwards due to Minerva. The 540B parameters model achieves an overall accuracy of 50.3% on MATH." | 340: "The increase in performance over the past year has been dramatic" / "Optimistic forecasts reflect great progress already reported this year, and the potential for rapidly achieving top-level human performance in the next year or two and then continuing to to slow push past it. Most of the question tidying could be automated within the next 2 years."

[1126] 338: "Solving problems in the MATH dataset is likely to remain an active focus of machine learning researchers. User [forecaster name]: "The phrasing of the question makes it impossible for the task to simply be neglected, except maybe if a small number of low-quality attempts to improve on the benchmark are made, preventing the benchmark from being "not actively used" in a way that would trigger resolution by an expert panel. Something like this seems possible at the 5% level at most. I expect new serious attempts to be made before mid-2024. I don't see any reason not to expect further progress given that the benchmark contains a wide range of problems and better models will find strategies that work on some of those problems."

[1127] 338: "Artificial General Intelligence (AGI) may be built before 2030. [Forecaster name]: "By 2030, full AGI might become able to solve all math problems solvable by humans."

[1128] 338: "The Minerva paper gets basic arithmetic wrong due to inferring answers from correlations. It is likely that improving on this is low-hanging fruit (by, say, using a symbolic algebra package like Sympy, as used in the MIT paper). User [forecaster name]: "Also, reading the Minerva paper, that model is getting basic arithmetic wrong 20% of the time because it is just inferring the answers from correlations. They believe this accounted for about 1/3 of the errors. The Codex model above is getting around this issue (I

- Generic methods already perform very well, so methods targeted specifically at this benchmark are likely to perform even better.[1129]
- Forecasters often underestimate AI progress.[1130]

Cross-references with other questions

Q40: Massive Multitask Language Understanding Benchmark
Q41: Quality Dataset Benchmark
Q42: AI Wins International Mathematical Olympiad
Q43: NYT Bestsellers Written by AI

Question 40: Massive Multitask Language Understanding Benchmark

What will be the state-of-the-art few-shot or transfer accuracy on the Massive Multitask Language Understanding dataset…
… by June 30, 2024?
… by June 30, 2030?

---

believe) by using programming languages which would not make math errors to come up with problem solutions. The Minerva paper has some great discussion of the kinds of errors these models are making as well as looking in detail at things like false positives where the model accidentally guesses a correct numeric answer." If the problem of doing basic arithmetic could be solved (which seems fairly easy), this would lead to an increase to a score of ~70%." | 342: "considering that about half the mistakes of MINERVA were calculation errors, most of which probably can be fixed with relatively little effort, there might be a [] significant amount of low-hanging fruit for LLMs left."

[1129] 338: "Even if it is not counted now, the MIT result of 81% suggests that the score can be improved significantly without too much work. [Forecaster name] (discussing the 81% result): "It still appears that very generic methods are getting really strong results on this dataset. So any research effort that targets the MATH dataset and math solutions generally will probably still lead to a performance jump.""

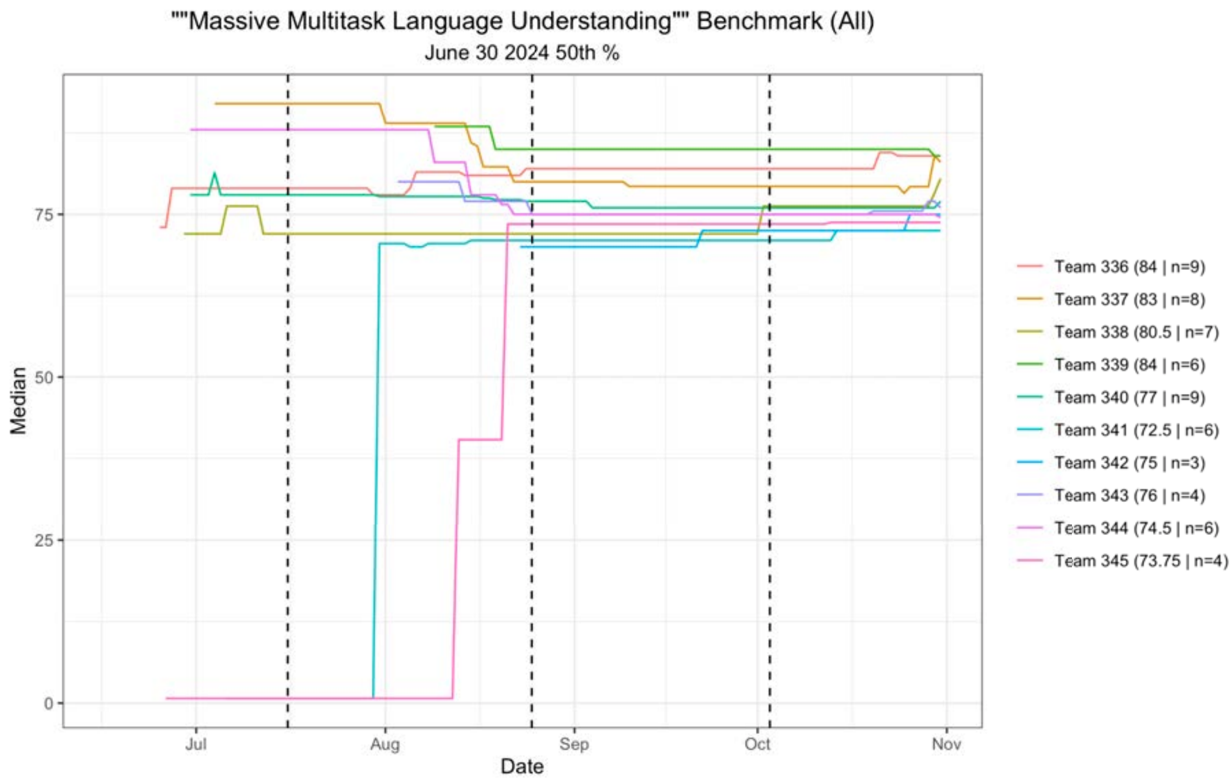
[1130] 338: "We have previously underestimated AI progress and it is easy to do so. During the course of this tournament, many of us have been forced to update strongly due to the release of the Minerva paper. The Metaculus forecasters [7] also greatly increased their predicted scores (for 2025) from 46% to 85% overnight when the Minerva paper was released. This suggests that many of us underestimated how easy solving problems on the MATH dataset is for an AI. It is plausible that this will happen again, which might lead us to increase our estimates beyond what we initially think." | 339: "Berkeley Prof. Jacob Steinhardt's ongoing experiment with forecasting the progress of several measures of AI progress, including this MATH dataset: https://bounded-regret.ghost.io/ai-forecasting-one-year-in/. Results so far in the Steinhardt experiment, which began in August of 2021, are that all forecasters vastly underestimated progress in the MATH's reports on competition mathematics. Results as of July 5, 2022: 50.3% accuracy vs. 12.7% predicted. So I take that as a warning to not underestimate future progress!" | 345: "we've seen how Metaculus had underestimated AI's capabilities in the past, and it is clear that AI is developing at rapid speed which even the experts couldn't predict." | 340: "Previous efforts to forecast this benchmark have woefully underestimated the pace of progress." | 342: "Forecasters are prone to underestimate progress of ML models on standardized datasets."

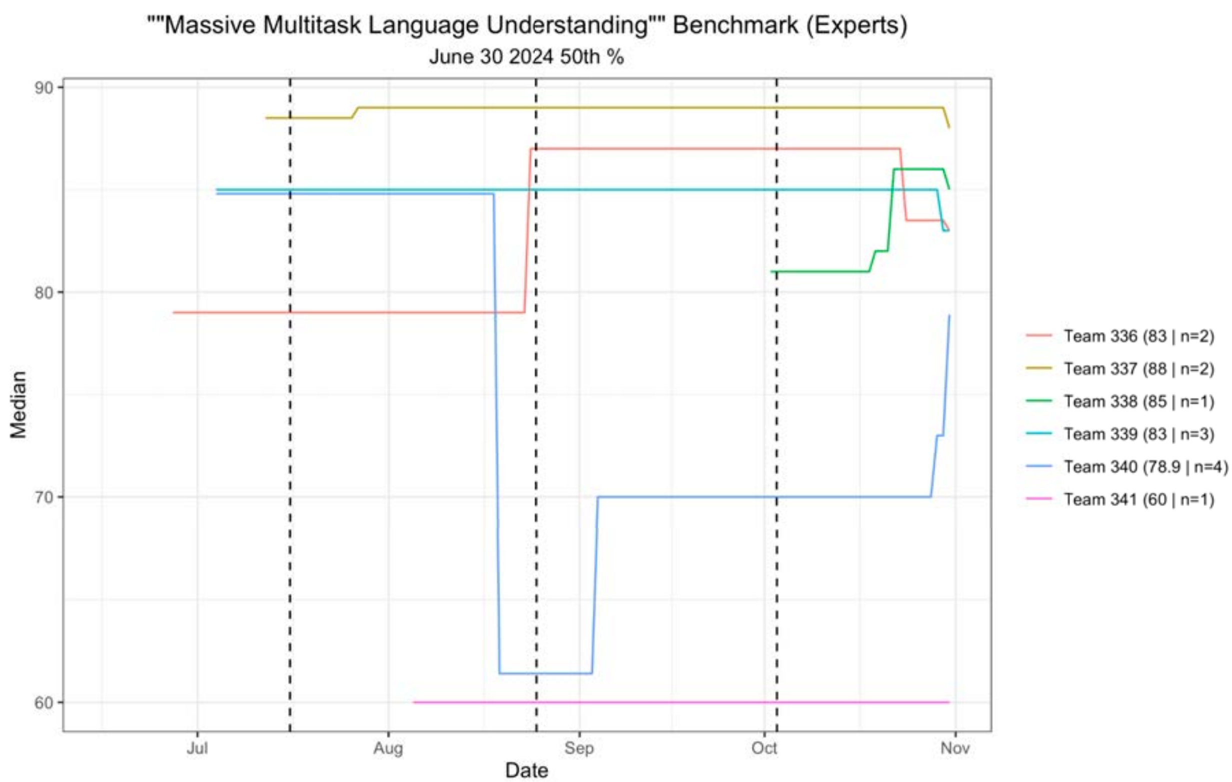
Question and resolution details, prior forecasts, and other relevant sources

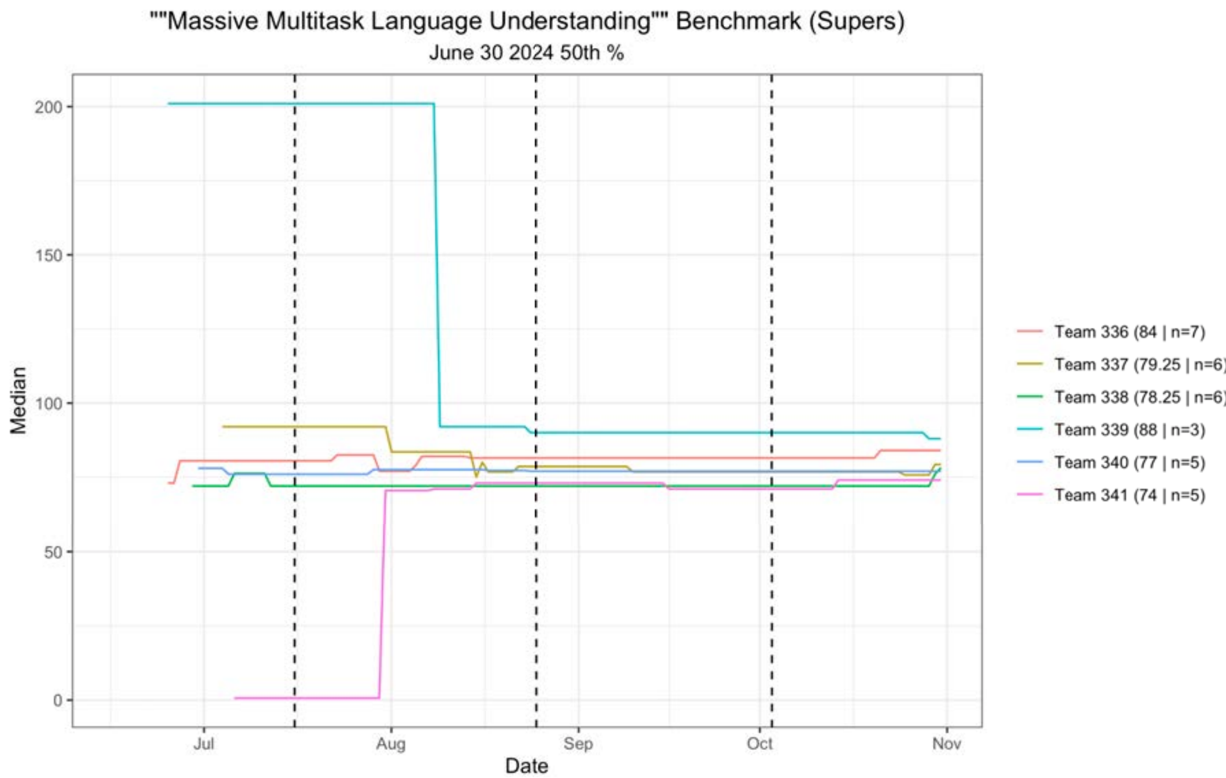
Results[1131]

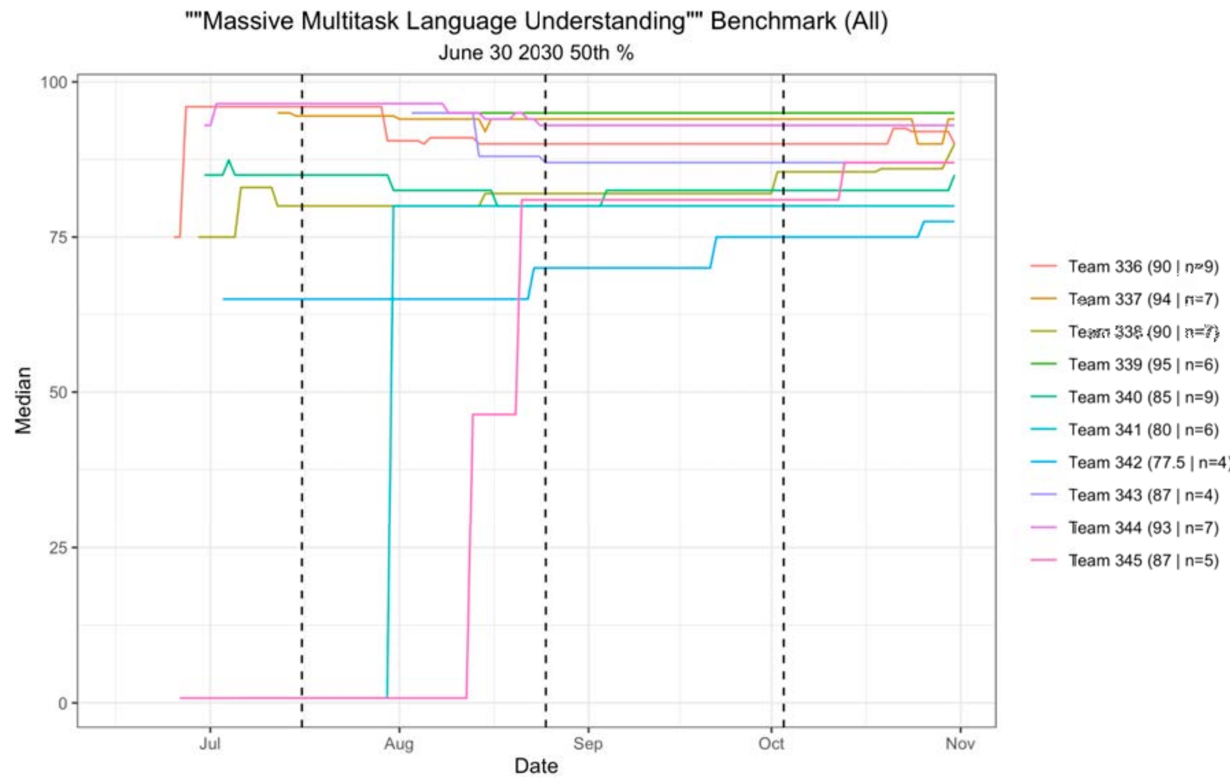
Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 32)	2024	76	77.75	48.69	-81.96%
	2030	85	88	180.65	-96.34%
Domain Experts (N = 3)	2024	84.9	84.8	0.14	+678.89%
	2030	92.4	90	3.68	-19.87%
General X-Risk Experts (N = 3)	2024	87	85	NA[1132]	NA
	2030	97	90	NA	NA
Non-Domain Experts (N = 10)	2024	85	80	16.82	-35.21%
	2030	88	90	12.9	-25.23%

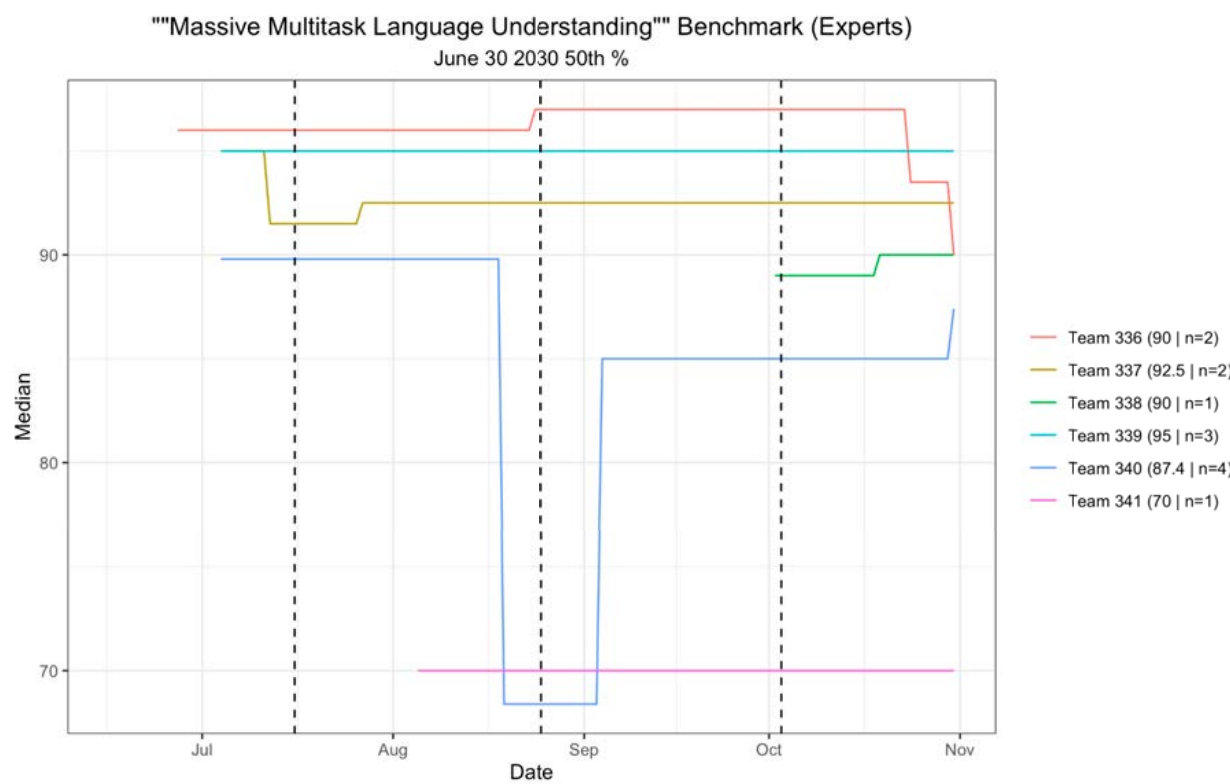
[1131] Numbers of forecasters are given as of Stage 4 of the XPT.
[1132] Only one forecaster in this group answered this question in Stage 1.

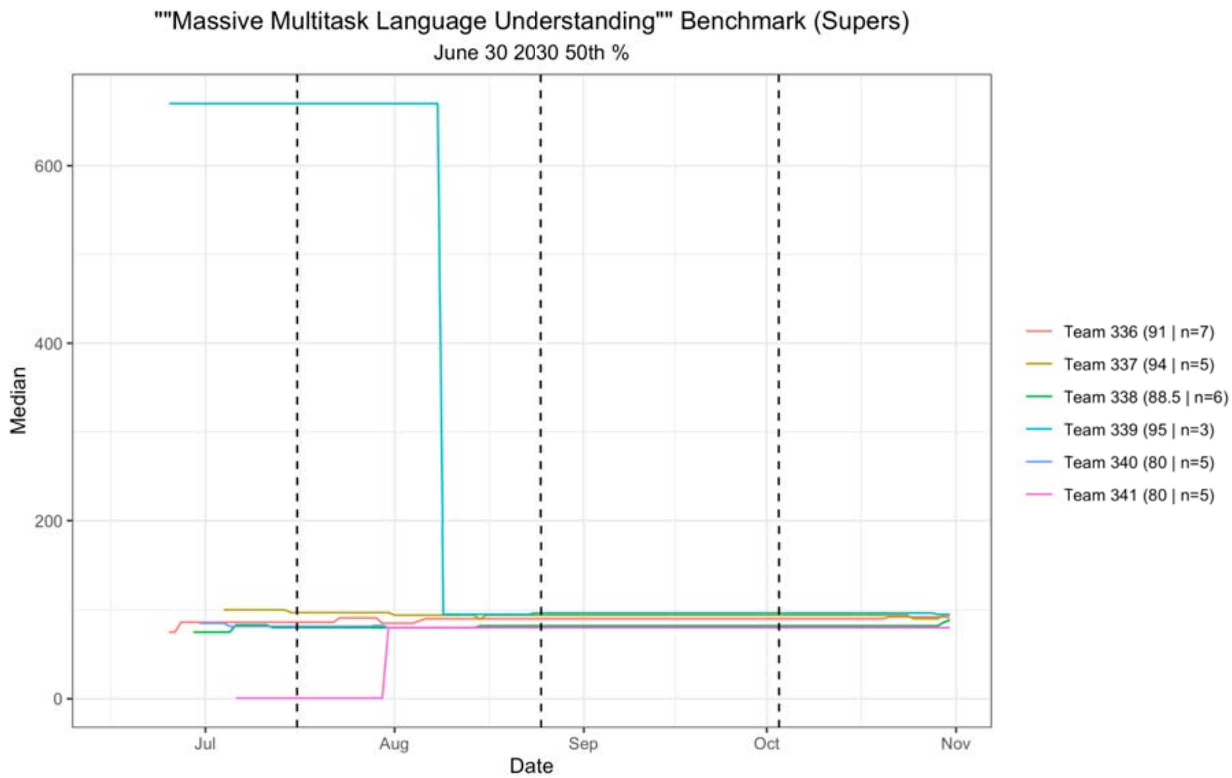

""Massive Multitask Language Understanding"" Benchmark (All)
June 30 2024 50th %
Median
0
25
50
75
Jul
Aug
Sep
Oct
Nov
Date
Team 336 (84 | n=9)
Team 337 (83 | n=8)
Team 338 (80.5 | n=7)
Team 339 (84 | n=6)
Team 340 (77 | n=9)
Team 341 (72.5 | n=6)
Team 342 (75 | n=3)
Team 343 (76 | n=4)
Team 344 (74.5 | n=6)
Team 345 (73.75 | n=4)

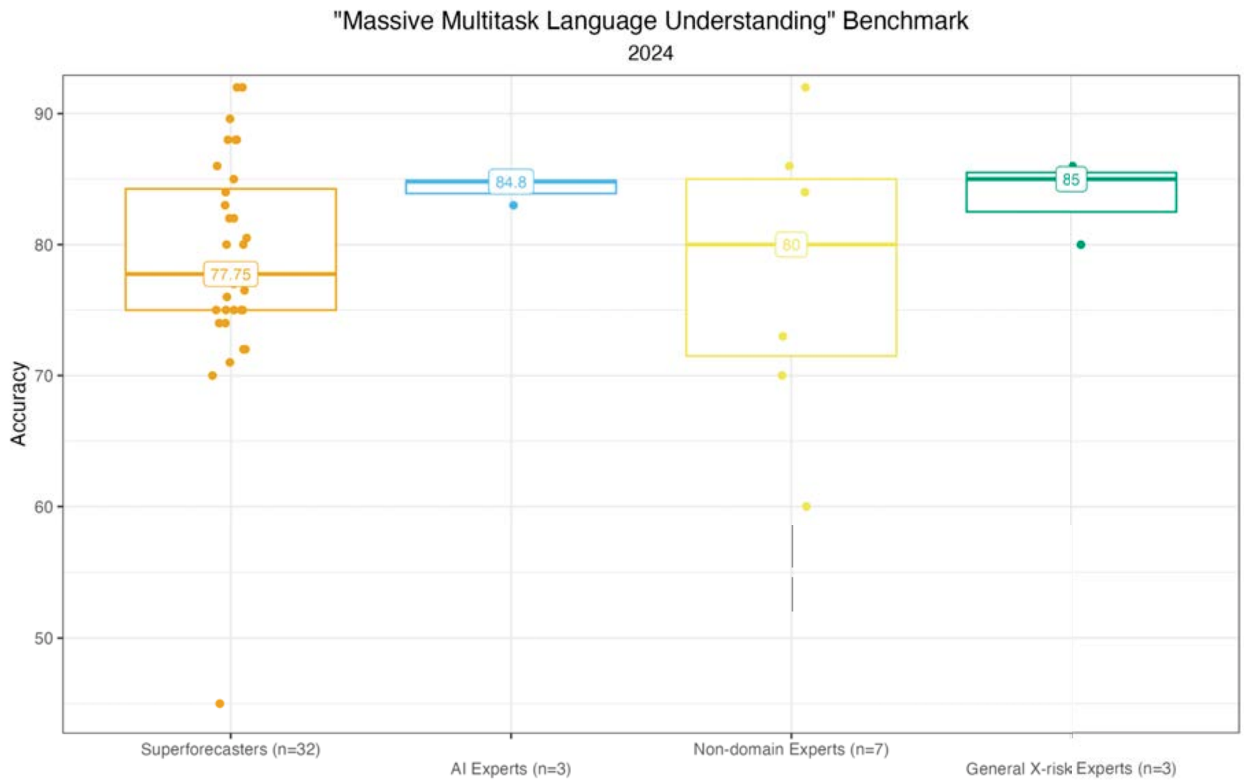

""Massive Multitask Language Understanding"" Benchmark (Experts)
June 30 2024 50th %
Median
60
70
80
90
Jul
Aug
Sep
Oct
Nov
Date
Team 336 (83 | n=2)
Team 337 (88 | n=2)
Team 338 (85 | n=1)
Team 339 (83 | n=3)
Team 340 (78.9 | n=4)
Team 341 (60 | n=1)

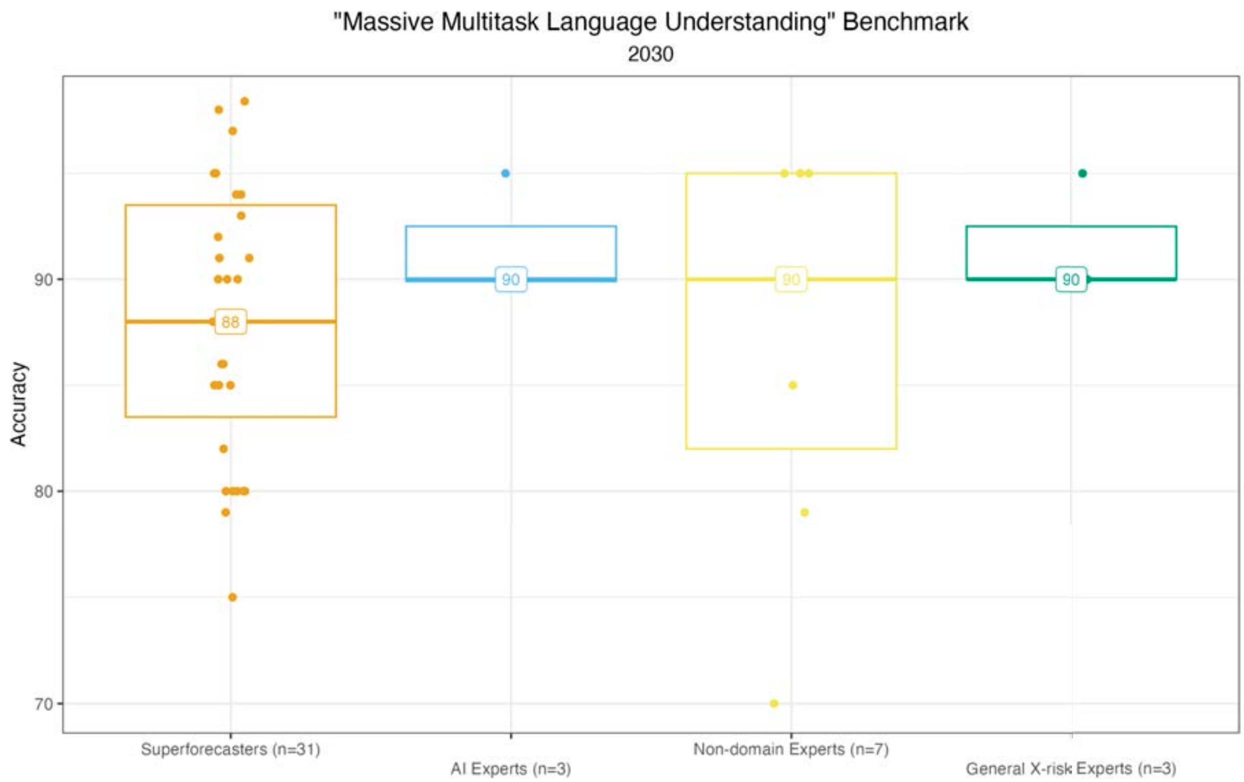

""Massive Multitask Language Understanding"" Benchmark (Supers)
June 30 2024 50th %
200
150
100
50
0
Median
Jul
Aug
Sep
Oct
Nov
Date
Team 336 (84 | n=7)
Team 337 (79.25 | n=6)
Team 338 (78.25 | n=6)
Team 339 (88 | n=3)
Team 340 (77 | n=5)
Team 341 (74 | n=5)


""Massive Multitask Language Understanding"" Benchmark (All)
June 30 2030 50th %
Median
100
75
50
25
0
Jul
Aug
Sep
Oct
Nov
Date
Team 336 (90 | n=9)
Team 337 (94 | n=7)
Team 338 (90 | n=7)
Team 339 (95 | n=6)
Team 340 (85 | n=9)
Team 341 (80 | n=6)
Team 342 (77.5 | n=4)
Team 343 (87 | n=4)
Team 344 (93 | n=7)
Team 345 (87 | n=5)


""Massive Multitask Language Understanding"" Benchmark (Experts)
June 30 2030 50th %
Median
90
80
70
Jul
Aug
Sep
Oct
Nov
Date
Team 336 (90 | n=2)
Team 337 (92.5 | n=2)
Team 338 (90 | n=1)
Team 339 (95 | n=3)
Team 340 (87.4 | n=4)
Team 341 (70 | n=1)


""Massive Multitask Language Understanding"" Benchmark (Supers)
June 30 2030 50th %
Median
600
400
200
0
Jul
Aug
Sep
Oct
Nov
Date
Team 336 (91 | n=7)
Team 337 (94 | n=5)
Team 338 (88.5 | n=6)
Team 339 (95 | n=3)
Team 340 (80 | n=5)
Team 341 (80 | n=5)


"Massive Multitask Language Understanding" Benchmark
2024
Accuracy
90
80
70
60
50
77.75
84.8
80
85
Superforecasters (n=32)
AI Experts (n=3)
Non-domain Experts (n=7)
General X-risk Experts (n=3)


"Massive Multitask Language Understanding" Benchmark
2030
Accuracy
90
80
70
88
90
90
90
Superforecasters (n=31)
AI Experts (n=3)
Non-domain Experts (n=7)
General X-risk Experts (n=3)

Sources of agreement, disagreement and uncertainty

- Some forecasters felt that they didn't have enough domain knowledge to feel confident making predictions, and mainly deferred to other forecasters.[1133]

Arguments given for lower scores

- Researchers might not apply the most advanced models to improving MMLU accuracy.[1134]
- The last points of accuracy-increase might be harder to achieve.[1135] This seems to have happened in other areas of ML, e.g., self-driving vehicles.[1136]
- The dataset involves multiple categories of task, and some categories (e.g. STEM questions) might be harder.[1137]
- We might have picked the "low-hanging fruit" already; progress might require "conceptual breakthroughs."[1138]
- Expecting only small incremental gains.[1139]

[1133] 339: "Beyond wild guessing, the team had no firm answer and mostly has given up. Our knowledge was insufficient to start tackling the question, so all we can do is highlight a few forecaster comments without endorsing them and in an unstructured way."
[1134] 336: "The main reason for no further progression would be that the most advanced models are not used for MMLU but rather incremental improvement of already tested models."
[1135] 343: "The closer we get to 100%, the more difficult it may be to max out the score. The last 5-10% of the problem may be much more difficult than the first 80%, as we've seen in other AI-related problems, such as driverless cars (Pareto principle)."
[1136] 340: "The strongest argument for the low extreme forecast in both 2024 and 2030 is that the benchmark hits a wall where progressing further toward human-level performance becomes very difficult. This is not implausible, and has been the case for self-driving vehicles: Progress in the early part of last decade suggested that Level 5 autonomous vehicles would be for sale to consumers by now, but this has not come to pass. What has happened is that we have software that is very good at automated driving—it is almost self-driving, and is self-driving in ideal conditions—but it is not 100% self-driving. In this case we seem to have gotten 95% of the way to self-driving vehicles very quickly, but the last 5% has been extremely difficult due to edge cases. Something similar could happen on this benchmark."
[1137] 343: "Since this problem involves multiple categories, some advanced categories (i.e. Social Sciences) that have advanced significantly may approach maximum improvement, while less advanced categories (i.e. STEM) struggle to overcome more fundamental issues." | 337: "Pushback from other forecasters included that the difficulty of solving a higher fraction of the test question might also scale exponentially as the "conceptual dimensionality" increases."
337: "Another forecaster took the observation that the AI systems showed overall lower performance on STEM questions to mean that those systems are limited in terms of combining quantitative and qualitative reasoning, which would likely be necessary to reach 100% performance."
[1138] 344: "We may have picked the low-hanging fruit in the MMLU, and large improvements may require conceptual breakthroughs."
[1139] 342: "[B]ehind this forecast is expectation, that most likely no large improvements will happen in the future, only small incremental gains. However there is an approximately 25 % chance of a significant breakthrough, that would cause faster increase of the state of the art accuracy."

- It's unclear whether scaling the size of the models will be enough; scaling models might require an accompanying increase in the amount of training data, and we might hit a training data bottleneck.[1140]

Arguments given for higher scores

- There is a trend of steady progress.[1141]
- Forecasters have consistently underestimated AI progress in the past.[1142]
- There seem to be clear paths to further improvements in the state-of-the-art.[1143]
- It seems likely that researchers will continue to target improvement on the MMLU benchmark.[1144]
- The questions are multiple choice, which might make the STEM component easier than the (open-ended) problems in the MATH dataset (which recently saw a large increase in SOTA performance).[1145]
- Current models aren't optimized for the MMLU task; further training might lead to significant improvement even without "breakthroughs".[1146]

[1140] 344: "It is unclear whether simply scaling up existing language models will solve the test. Current understanding indicates that a 10× increase in model size must be accompanied by an approximate 5× increase in data (Kaplan et al., 2020)."
344: "data may also become a bottleneck, as there is far less written about esoteric branches of knowledge than about everyday situations"
337: "Additionally, one forecaster says the bottleneck might be in terms of data available."

[1141] 343: "Significant progress has already been made over the past two years" | 344: "Steady progress in 2021 and 2022, on what appear to be specialized systems (as opposed to GPT-3)." | 337: "One forecaster with relatively high estimates extrapolated historical scores linearly in time, arguing that as AI performance should scale roughly exponentially due to the exponential nature of Moore's law."

[1142] 340: "The strongest argument for the high extreme forecast in 2024 is likely that forecasters continue to do poorly on forecasting this benchmark. This could be because they do no[t] anticipate emergent capabilities of models." | 337: "Two forecasters point[] out that previous forecasts by the forecasting community have significantly underestimated progress in machine learning, see https://bounded-regret.ghost.io/ai-forecasting-one-year-in/."

[1143] 336: "There is more room for improvement from the current results. From the data, the fine-tuned model performed better than the few shots, in the specific case of GPT-3 the results were improved by 10%.
Models can be further improved by using better training (combine models..), more data and more specific dataset (STEM and humanities that are below average in MMLU), or more computer power." | 341: "it seems like a data-crunching exercise. If they have some success already, and are exponentially increasing processing power and hardware, adding model parameters, and incorporating more and more training data, they should get more accurate."

[1144] 344: "With researchers embracing this as a target for the progress they will probably specialize their systems further over the next several years."

[1145] 344: "The multiple-choice nature of MMLU probably makes its STEM component considerably easier than the problems in the MATH dataset. And the MATH dataset recently had a large improvement upwards due to Minerva. So using something similar may be enough to considerably improve current SOTA in MMLU."

[1146] 344: "higher incremental improvement is expected even without a significant breakthrough. This is mainly justified using data provided by the authors of Chinchilla model in [1]. In the paper authors present evidence, that state of the art AI models were undertrained for performance on the Massive Multitask Language Understanding dataset, and show they have managed to beat all other state of the art models

Cross-references with other questions

Q39: MATH Dataset Benchmark
Q41: QuALITY Dataset Benchmark
Q42: AI Wins International Mathematical Olympiad
Q43: NYT Bestsellers Written by AI

Question 41: QuALITY Dataset Benchmark

What will be the best SAT-style score with a machine learning model on the hard subset of the QuALITY dataset…
…by June 30, 2024?
…by June 30, 2030?
…by June 30, 2040?

Question and resolution details, prior forecasts, and other relevant sources

Results[1147]

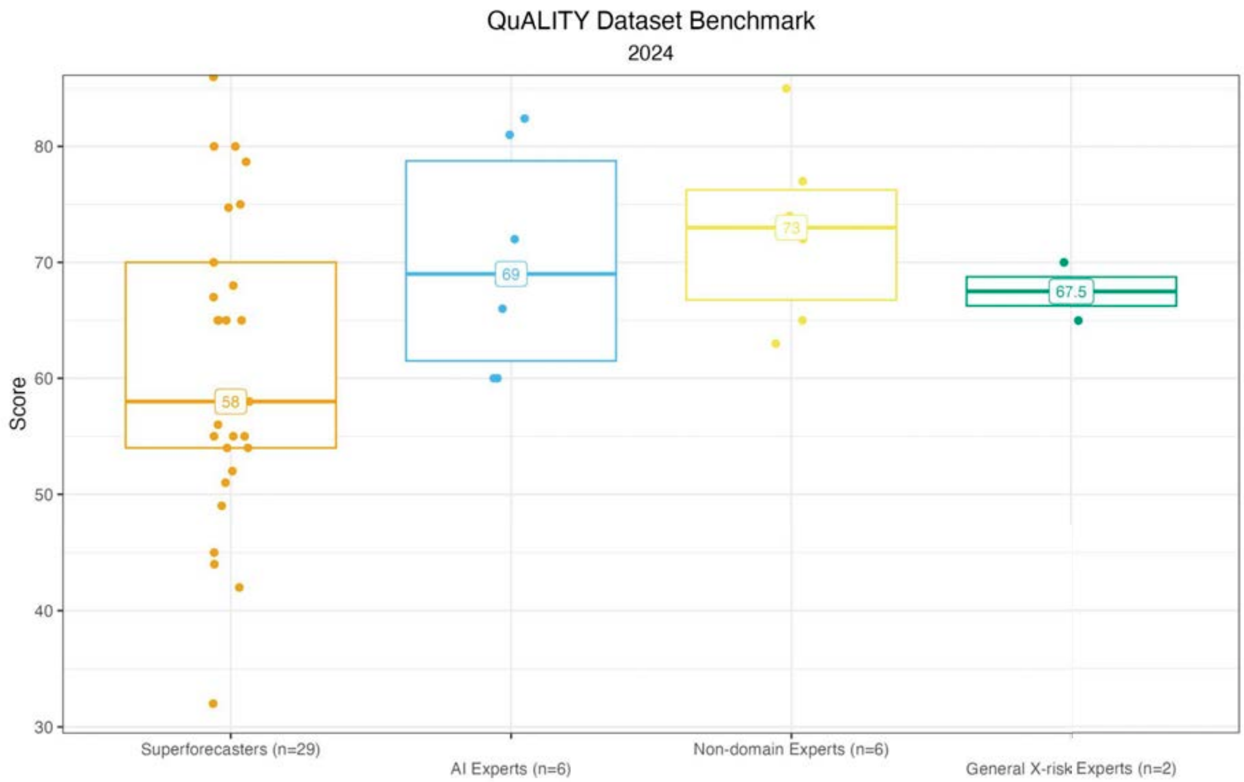
Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 32)	2024	67	58	18.28	-23.89%
	2030	80.5	76	15.8	-24.28%
	2040	88	86.5	13.21	-30.95%

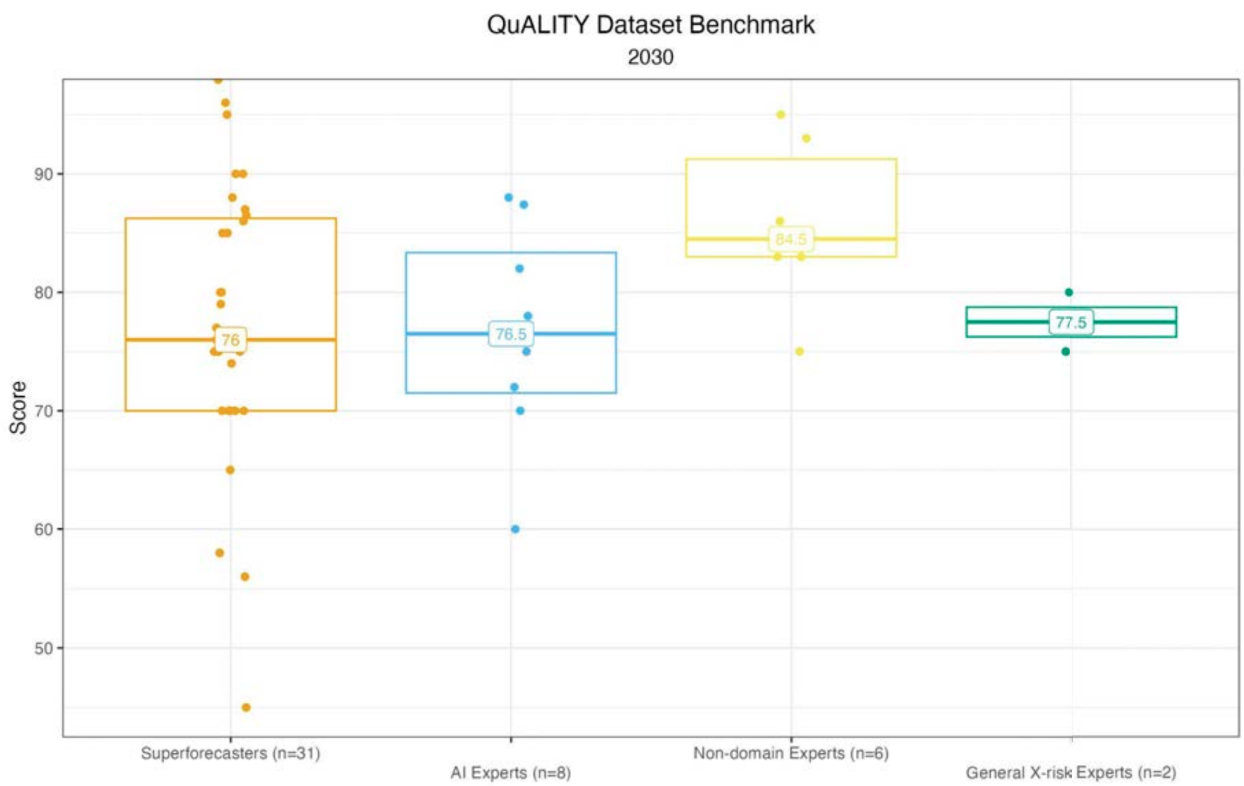
with Chinchilla, despite using 2 - 4 times less parameters as other models. Therefore, there is still room for significant incremental improvement."

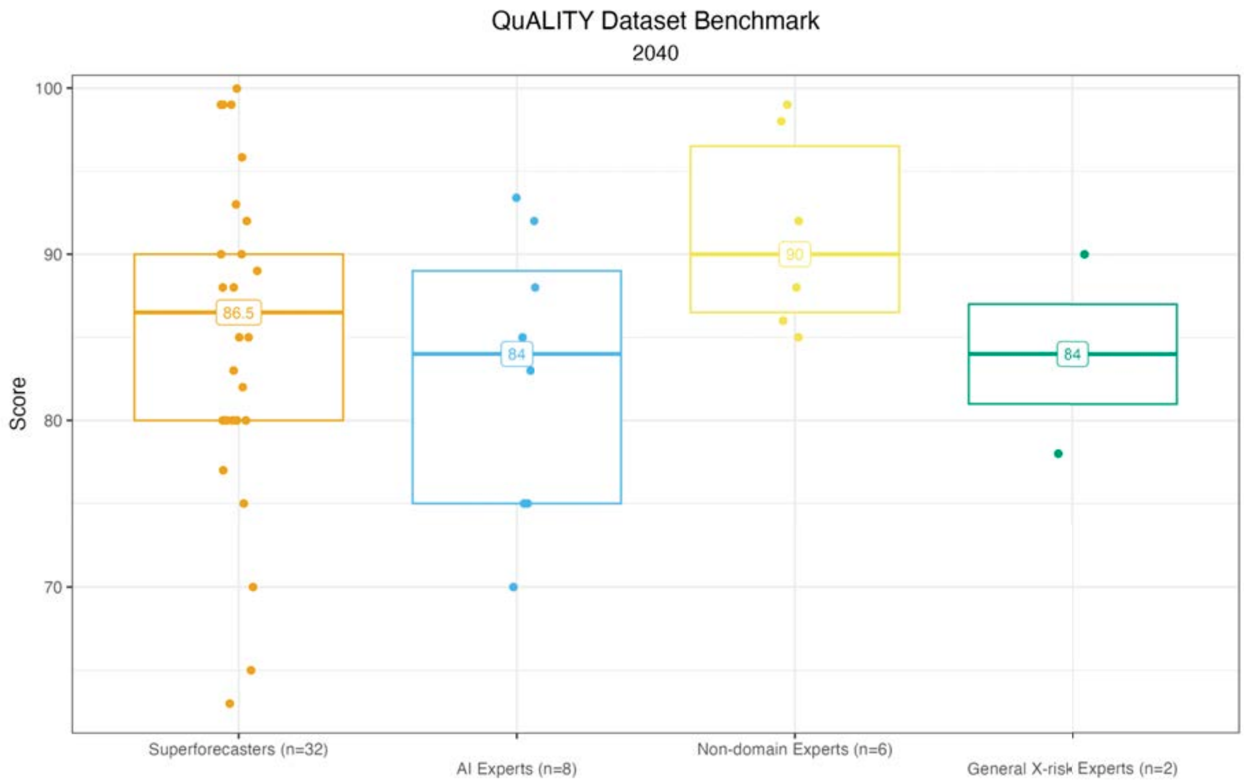
[1147] Numbers of forecasters are given as of Stage 4 of the XPT.

Domain Experts (N = 8)	2024	85.4	69	NA[1148]	NA
	2030	70	76.5	14.95	-37.04%
	2040	75	84	13.46	-36.3%
General X-Risk Experts (N = 2)	2024	70	67.5	NA	NA
	2030	80	77.5	NA	NA
	2040	90	84	NA	NA
Non-Domain Experts (N = 6)	2024	77	73	NA	NA
	2030	83	84.5	NA	NA
	2040	88	90	NA	NA

[1148] Indicates that only one forecaster in this group answered this question in Stage 1.


QuALITY Dataset Benchmark
2024
Score
80
70
60
50
40
30
58
69
73
67.5
Superforecasters (n=29)
AI Experts (n=6)
Non-domain Experts (n=6)
General X-risk Experts (n=2)


QuALITY Dataset Benchmark
2030
Score
90
80
70
60
50
76
76.5
84.5
77.5
Superforecasters (n=31)
AI Experts (n=8)
Non-domain Experts (n=6)
General X-risk Experts (n=2)



Sources of agreement, disagreement and uncertainty

- Some noted that domain experts may be best positioned to predict this.[1149] However, several teams also noted they expect that domain experts might be biased toward expecting continued progress.[1150]

[1149] 344: "Domain experts may have a much better understanding of reference classes that are appropriate for forecasting this question. Also, they will probably have a much better understanding of how many teams are working on improving the SOTA in this benchmark, and of the probability of simple changes being enough for score improvements of different magnitudes."

[1150] 344: "Domain experts might be biased in thinking that the current high trajectory of progress would continue in the future, without plateauing in the short or mid term." | 341: "However, two forecasters are also domain-specific experts in the publishing industry (one has a novel in press with a credible publisher, one has written many nonfiction books and magazine articles for popular audiences. As a magazine editor, she mentored Eric Drexler and T.A. Heppenheimer, who had never previously written for popular audiences, and who soon made the NYTimes best seller list. She also is a NLP researcher… and hence more likely to be optimistic.) Therefore, we expect that they will bias the results lower." | 339: "Experts in the field may project a faster convergence than the respondents."

- Some forecasters expressed a sense that they were mostly going off of intuition / limited data.[1151]
- It's unclear how AI investment overall might change; it's possible a downturn could slow progress across the board.[1152]
- There is not currently much difference in AI performance on the test overall vs. the hard subset (suggesting that increases in overall performance might transfer to / apply to the hard subset).[1153]
- The SAT may stop existing or stop being prominently used.[1154]

Arguments given for higher scores

- The speed of breakthroughs in ML in general, and the similarity of this task to other tasks that have seen rapid improvement.[1155]
- Existing ML methods seem like they should be enough for significant improvement.[1156]

[1151] 344: "Our estimates seem to be based mostly on intuitions and a few points of data. The estimates could probably be improved by a more in-depth look at the QuALITY data set and by comparisons with the pace of progress in other benchmarks about text comprehension."
[1152] 340: "A similar effect could be driven by a downturn of investment into AI, which would slow down progress more broadly.
[1153] 341: "The current models are not much worse on the hard subset than on the total test set."
[1154] 341: "The SAT may cease to exist. Already the SAT is controversial for being manipulated by tutoring, and for biases that may unfairly exclude students for admissions and scholarships"
[1155] 340: "The more optimistic forecast are based on the possibility of breakthroughs in machine learning research as well as an increased availability of relevant training data that could lead to jumps in capabilities" | 344: "For 2030 and 2040, the general argument is that we can expect fast-paced progress. As it happened with the SuperGLUE benchmark, it seems likely we will exceed the human baseline." / "Advances in AI has been very fast lately, and the human baseline (85.4 for QuALITY) may be exceeded after only a few years of progress, as it was for many other benchmarks." | 341: "Base Rate: Similar to the forecasts on the NYT bestseller, NLP is expected to improve dramatically in the coming years (for example recent commercial successes of hybrid AI/human novel production using Sudowrite)" | 343: "[M]any prior well-defined ML problems have been solved, especially after significant progress has been made toward them." | 336: "This problem likely only needs one breakthrough to increase performance to the 80 point range. Language models also exhibit significant improvements at larger scale, which may be sufficient to solve the task."
[1156] 340: "[Some] forecasters are optimistic that existing methods like reinforcement learning from human feedback, or using other types of transformer models rather than BERT-like architectures might already be enough to significantly improve upon the currently listed results for the benchmark." | 344: "In the short term, this comment indicates that relatively simple improvements may be enough to achieve higher scores." | 336: "[T]here are several straightforward ways to improve performance on this problem:
- Better summarization of long texts
- Iterative summarization, whereby the model summarizes subsections of text, stores these summaries, and uses them to answer questions rather than refer to the entire text in order to answer questions
- Memorizing typical questions and precomputing the answers to such questions for subsections [...] of the text. The model can then refer to these precomputed answers when it responds to test questions
- Quantifying uncertainty on each test question, estimating probabilities of different answers, and using this to optimally select answers (or refrain from answering)"

- The multiple choice format might be exploitable.[1157]
- The QuALITY benchmark might become more prominent and get more engagement.[1158]
- Base rates of improvement suggest high scores (~human performance) will be achieved soon.[1159]
- "General purpose models" are being developed.[1160]
- SAT coaching (for humans) can improve scores significantly.[1161]
- AI progress is often faster than forecasters expect.[1162]
- Deference to existing Metaculus predictions.[1163]

Arguments given for lower scores

- The last 5-10% of the problem might be a discontinuity (i.e. the trend of performance improvement might flatline for a while before increasing).[1164]
- Reaching human performance might require "breakthroughs" in machine learning.[1165]

[1157] 344: "The multiple-choice format of the benchmark may allow for score improvements that are not directly related to text comprehension: it may be possible to extract useful information from the answer sets themselves."

[1158] 344: "The QuALITY dataset benchmark was released by the same team that released SuperGLUE, and is described as "certainly the best benchmark or dataset release from my group so far," so even though today it has limited engagement on the leaderboard, we can expect it to start to gain significantly more traction over time."

[1159] 341: "Base Rate: Superglue: From 71.5 July 2019 to 90.3 Jan 2021 and 91.2 Feb 2022. Was at 85 already Jan 2020. It was at 90.6 in Jun 2021.… Glue was released in 2018, but website is glitchy and it's hard to check baseline. By Jan 2022, it had reached 91.3. It was at 85 already in Jun 2019." | 339: "The current best performing AI model (as of May 2022) is at 39.6 according to https://nyu-mll.github.io/quality/. The human performance is at 85.4. In December 2021, the best performing AI model was at 28.1." | 336: "A linear extrapolation from the existing data suggests that we reach perfect accuracy before December 2023. An exponential extrapolation suggests that we reach it by December 2022. Obviously, there is lots of uncertainty, and the real progress curve will be more of an S-curve than a pure exponential function."

[1160] 341: "We're starting to see more and more general purpose models. See this one from Deepmind for example."

[1161] 341: "Studies indicate that SAT coaching can improve scores by 30 points. "The typical gain associated with coaching is 8 points for verbal and 18 points for math. Coaching seems to result in about one more verbal question correct for 25 to 30 hours of effort and one more math question correct for 8 or more hours of effort." AI should experience better gains especially in the Math area."

[1162] 340: "[O]ne forecaster notes that previous attempts at predicting performance for similar benchmarks usually underestimated the speed of progress." | 336: "Progress on other metrics has been quite rapid and often was faster than forecasters expected. Progress on this metric has been rapid as well."

[1163] 344: "Many of the forecasters based their predictions on the earlier forecasts in the Metaculus question." | 337: "The strongest argument for the higher end forecasts was made by [forecaster name]. Lifting from his good work:

- 2024 (will use Metaculus 2025 estimate here): (100+68+76*4)/6 = 78.67
- 2030: (in all cases using Metaculus median as "most likely") (100+75+4*86)/6 = 86.5
- 2040: (100+75+85*4)/6 => Metaculus had the 2040 median < 2030 median. I am going to somewhat arbitrarily and optimistically change the 2040 median to 100) = 95.83"

[1164] 343: "Some ML problems are amenable to a continuous improvement, while others are discontinuous: the last 5-10% of the problem represents >50% of the work required to get there. It is at least 5% possible that this will apply to the SAT problem."

[1165] 340: "some of the more pessimistic forecasts note, that such breakthroughs [in machine learning] might be strictly necessary to get close to human performance on this benchmark."

- The length of the text involved might make the task especially hard.[1166]
- There is a "strategic" component to SAT-scoring which may be hard for AI to master.[1167]
- Researchers might not put effort towards this particular problem / benchmark.[1168]
- General skepticism about AI/AGI development.[1169]

Other arguments

- There might be problems with the training data.[1170]

Cross-references with other questions

Q39: MATH Dataset Benchmark
Q40: Massive Multitask Language Understanding Benchmark
Q42: AI Wins International Mathematical Olympiad
Q43: NYT Bestsellers Written by AI

Question 42: AI Wins International Mathematical Olympiad

By what year will an AI win a Gold Medal in the International Mathematical Olympiad (IMO)?

Question and resolution details, prior forecasts, and other relevant sources

---

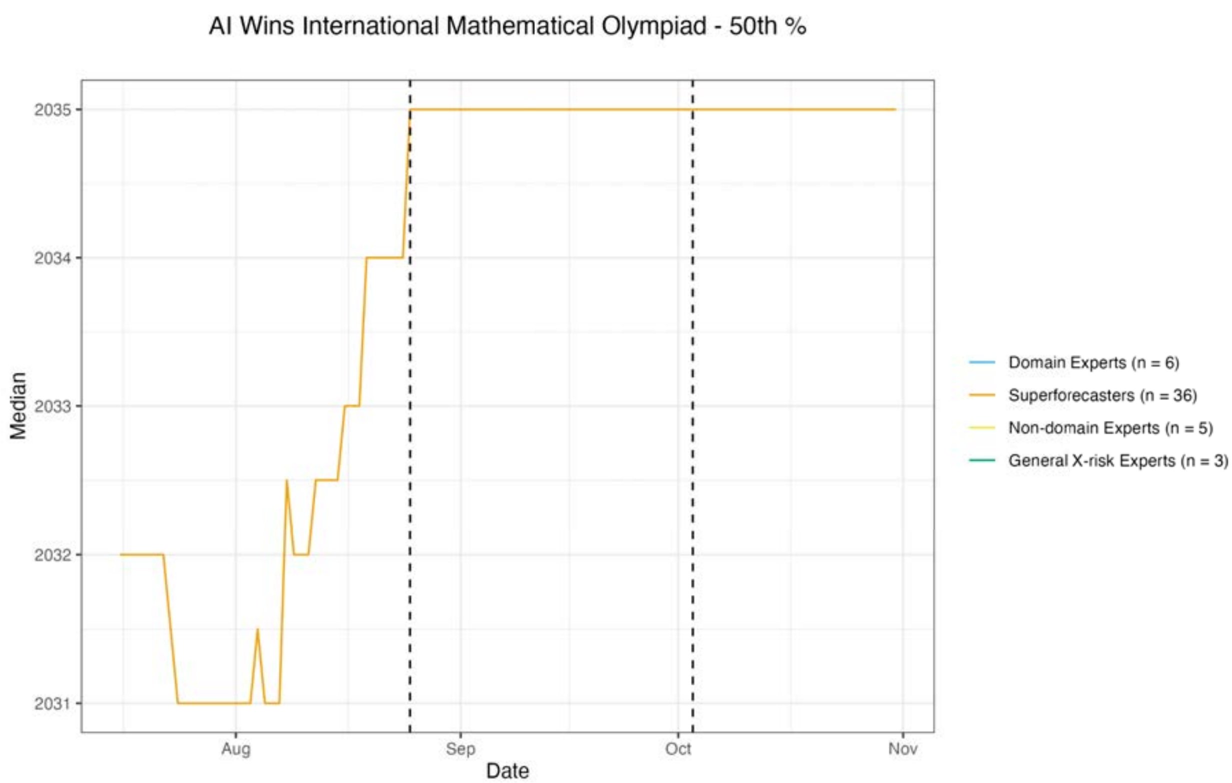
[1166] 336: "The dataset seems harder than other datasets as the length of the text makes summarization harder."
[1167] 336: "SAT-style scoring means that wrong answers will be penalized. This means that answering questions has not not only a knowledge element, but also a strategy element in it, which adds additional complexity. Maybe it is hard for an AI to learn the strategic elements."
[1168] 336: "People might simply not put a lot of effort working on this dataset, meaning that progress stales [sic] even though underlying capabilities have increased." | 340: "There is additional uncertainty about the effort that is going to be invested into this benchmark, with some of the more pessimistic forecasts noting that the leaderboard has not been adopted very widely; even if there could be significant improvements upon the benchmark using existing techniques, it could be that no lab is going to invest the time into applying them to the benchmark." | 344: "While it seems like many improvements are possible, maybe few AI research teams care about achieving state-of-the-art results in this benchmark. This would mean that progress here would be slower than it could, perhaps merely advancing as an indirect result of general progress in AI." | 341: "Researchers are working on this dataset, but it is not certain that they will be particularly focused on the hard subset as measured by a score with a guessing penalty, as opposed to the other three metrics on the leaderboard."
[1169] 337: "The strongest argument for the lower end [note - here means 'worse performance', not 'shorter timeline'] forecasts were expressions of skepticism about the development of AI in general and AGI in particular."
[1170] 339: "A source of uncertainty for one respondent was the potential for a less than correct data set being used to train the AI prior to taking the SAT, as it would hurt the results."

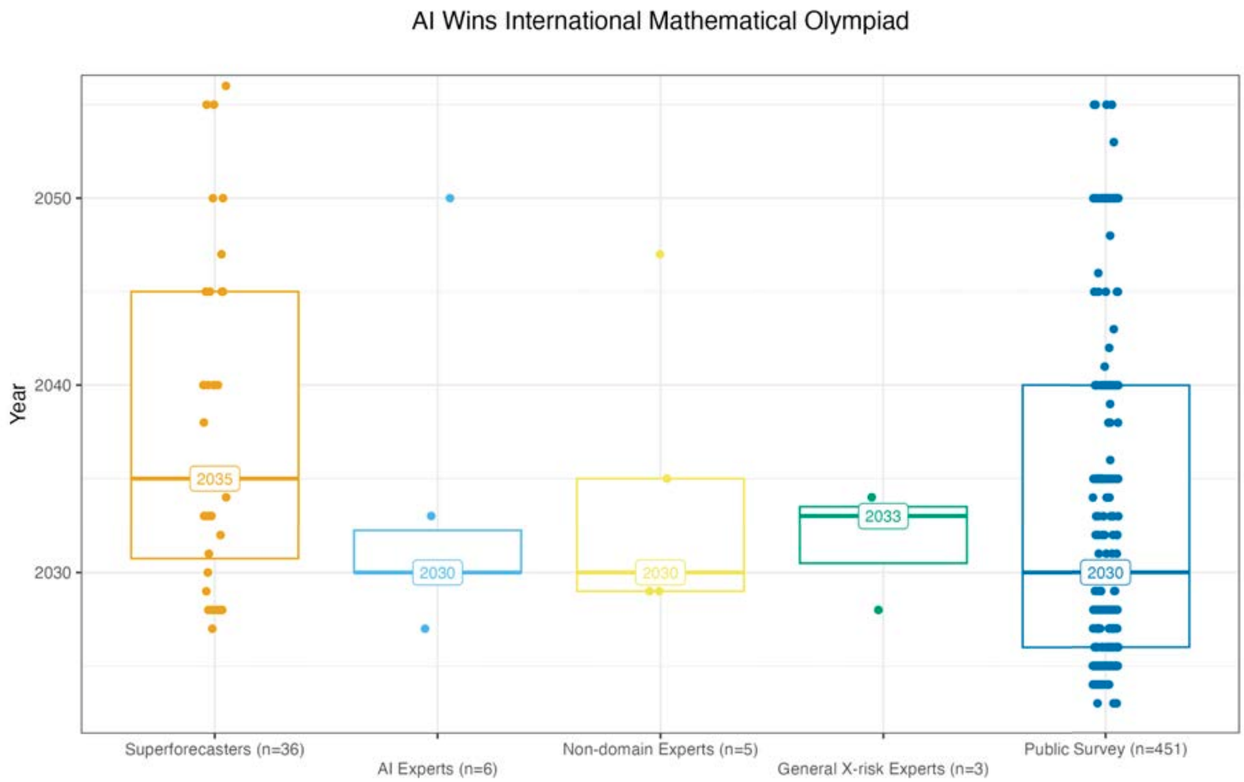
Results[1171]

Group	Percentile Forecast	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 36)	5%	2025	2025	3.42	-23.78%
	25%	2028.5	2030	7.47	-25.14%
	50%	2033.5	2035	13.29	-35.35%
	75%	2051.5	2049	22.06	-42.34%
	95%	2075	2075	Inf	NaN
Domain Experts (N = 6)	5%	2023.5	2024	0.96	-21.38%
	25%	2026	2027	3.79	-51.54%
	50%	2030	2030	35.35	-76.28%
	75%	2045	2040	64.47	-61.58%
	95%	2213.5	2051.5	Inf	NaN
General X-Risk Experts (N = 3)	5%	2024	2024	NA[1172]	NA
	25%	2028	2027	NA	NA
	50%	2035	2033	NA	NA

[1171] Numbers of forecasters are given as of Stage 4 of the XPT. Public survey respondents were surveyed outside of the tournament context.
[1172] Only one forecaster in this group answered this question in Stage 1.

	75%	2100	2050	NA	NA
	95%	1e210	5e209	NA	NA
Non-Domain Experts (N = 5)	5%	2025	2024	0	+Inf%
	25%	2031	2027	5.66	-33.15%
	50%	2038.5	2030	12.02	-36.1%
	75%	2049.5	2040	13.44	-26.2%
	95%	2066.5	2060	9.19	144.63
Public Survey (N = 462)	50%	2030		Inf	-





Sources of agreement, disagreement and uncertainty

- Some teams indicated there was a lot of disagreement among team members.[1173]
- Some thought this wouldn't happen until after AGI (or, that the first system capable of doing this will be an AGI, rather than a system specifically aimed at performing well in the IMO).[1174] Relatedly, some think catastrophic risk is likely before this can occur.[1175]
- There may be a chance the IMO will stop existing, or that AI will be prevented from trying, and so this won't happen (rather than an AI being incapable in principle of achieving the result).[1176]
- Some raised concerns that, even if an AI achieved this, it might not be a fair test in some way (e.g. if an "internet-trained AI" could somehow get an "unfair advantage"), or might not really be measuring what we wanted.[1177]

[1173] 334: "Very large differences in how difficult different forecasters think this challenge will be."
[1174] 341: "I believe this will likely occur after "advanced AI" as described in question #44." | 336: "AI will pose [a] major catastrophic risk to humans before this can be achieved. Or, more likely, it will achieve superiority in mathematics one morning, without participating in the Olympiad, and within a few hours have improved so much by recursive self improvement that it becomes unstoppable."
[1175] 336: "3 out of 9 forecasters see extinction/catastrophic risk to humans as major concern that this will not be achieved."
[1176] 339: "The International Mathematical Olympiad could be cancelled or deep learning models under the banner of AI could be banned in society."
[1177] 340: "One source of disagreement between forecasters was determining the technical capabilities of AI systems. It is possible that AI solutions to math problems are inherently difficult to parse for human judges. There is also uncertainty associated with the rules of the IMO, which might provide an unfair

- The required effort and investment to create such a system just might not happen, even if it's possible.[1178]

Arguments given for shorter forecasts

- AI has already achieved some notable math milestones. [Minerva](#) was the most cited; it has advanced more quickly than some experts expected. It has scored ~50% on the MATH dataset (on which an IMO gold medalist scored 90%)[1179] and scored 80% on a Polish national math exam.[1180]
- The task doesn't seem like it will require any fundamental breakthroughs; it looks like the sort of task achievable with deep learning.[1181]
- In general, AI progress is often surprisingly fast.[1182]

---

advantage to an "internet-trained" AI." | 336: "Sources of uncertainty include the contest rules, opinions of panels of experts, IMO problem difficulty variation, potential for varying quality of human IMO contestants (determines gold medal level.)"

[1178] 345: "IMO Grand Challenge team needs mathematicians to write detailed proofs of library of Math problems. These proofs are then used to train an AI system. This process is lengthy and requires a lot of resources." | 339: "Possibility that no one decides to put the resources into doing this: someone has to decide to train the AI. Perhaps DeepMind could do it in a few years if they wanted to (targeted specifically at the IMO), but if not we may be waiting for an AGI or at least an AI with more general math skills." | 340: "The performance of AI at a Math Olympiad would depend directly on the resources and funding available to develop and perfect such a system. Most forecasters agree on the technical ability of AI to win an IMO (International Math Olympiad), but vary their forecasts based on their expectation of monetary investment and human interest."

[1179] 342: "Demonstrations of AIs capable of solving math problems at different levels are out there, so it is now a matter of scaling." / "An interesting data point for this question is the MATH dataset [1]. According to the authors, the datase[t] questions were given to a three times IMO gold medalist, who got an accuracy of 90 %. Which gives us an approximate goal for the AI to beat. Currently state of the art AI optimised for solving math problems has an accuracy of 50 % on the MATH dataset, with most improvement coming from the optimisation rather than model size compared to other large language models." | 338: "...Then we have the very recent development in Codex that allowed F2F to improve from 25% to 35%. ([Source](#)) So I think the first steps are well underway and once it "gets" the coding it will then very quickly follow."

[1180] 337: "AI Minerva's capabilities have advanced in a matter of weeks what experts thought would take several years. Minerva recently scored above an 80% on the Polish national math exam (August, 2022). As in the case with Go, AI Minerva displayed innovative, "creative" solutions that humans had not previously demonstrated or went contrary to human best practice."

[1181] 336: "Feeding a self-learning system optimized for this task with mathematic principles, teaching it to apply them and in what order, and giving it sufficient calculating power looks doable in the near future. It´s much more complex than Chess or Go, but it can definitely be done. We probably don't need any big breakthroughs in AI to achieve this."

[1182] 336: "Increasing rate of AI reaching milestones such as this." | 337: "There is no direct base rate to guide this forecast. However, experience shows that AI in other domains, e.g., the game of Go, weather forecasting, and protein folding, among others, consistently surprises subject matter experts in that AI masters an area decade(s) faster than first thought possible. A frequent human first estimation is that a machine will never match "human intuition"; yet, AI has already done so in chess and Go. Mathematics will not be an exception to this pattern." | 341: "Cautionary information that professional forecasters can greatly underestimate progress" / "Historical improvements have been exponential" / "The speed of AI improvement in Chess and GO and the acceleration of machine learning."

- Progress can be discontinuous.[1183]
- Deference to Eliezer Yudkowsky's prediction that this could happen relatively soon.[1184]
- The IMO has already created rules for AI competitors, meaning they might believe this could happen soon.[1185]

Arguments given for longer forecasts

- "Math" or "proofs" might constitute an especially hard task (harder than e.g. Chess, Go).[1186] There might be some necessary mental feature like "concepts" or "insight" that is hard to train.[1187]
- Deferring to Paul Christiano's prediction that this will take relatively long to achieve. Because Christiano has personally participated in the IMO, he is well-positioned to gauge the difficulty of this task.[1188]
- Deference to other domain experts, who think this is far.[1189]

[1183] 344: "The large jump in the state-of-the-art for the MATH benchmark (from 6.9 to 50.3) also gives some reason to believe that similarly large jumps could occur for the IMO."

[1184] 344: "Eliezer Yudkowsky assigns >16% to the possibility that an AI will win a gold medal on the IMO at least once in the period 2022-2025, indicating belief in shorter timelines and the possibility of sudden capability spikes."

[1185] 338: "The IMO Grand Challenge has set rules for AI competitors - which mean[s], that even if AI is not competitive today; it might be in the near future."

[1186] 341: "Mathematical proofs are probably one of the harder tasks to set an AI to. There is less "work experience" to train it on, and slight factual/terminological mistakes falsify the whole attempt." | 342: "Mathematical Olympiad requires understanding, not symbol manipulation or reward maximisation." | 339: "Although AI has preformed [sic] well in the areas of language and programming, math may be a more difficult problem because of differences across notation and the inability to write unit tests like can be done to test software." | 338: "AI seems to be having difficulty with mathematics - might not happen soon." / "From the brief research I'll assert that this may currently be a key observation: '...Solving IMO problems often requires a flash of insight, a transcendent first step that today's AI finds hard — if not impossible.'" [Note: this person edited to add they later reduced their 95% estimate by 30 years based on further research.]

[1187] 336: "Minerva still shows a basic lack of understanding of a concept. This can be fixed. It has advanced so much in the past few years, we would expect it to advance further, but the last bit is going to be a heavy lift and may take more time than we expect."

[1188] 344: "Paul Christiano assigns <8% to the possibility that an AI will win a gold medal on the IMO at least once in the period 2022-2025. He has participated in the IMO in the past, so he has a personal understanding of the difficulty of this challenge. His prediction indicates belief that the grand challenge probably won't be solved very soon."

[1189] 345: "The Microsoft team in 2020 claimed that the AI won't be able to solve a single problem from the Olympiad that year. (Source)" | 338: "...The IMO Grand Challenge is currently a moonshot. If Lean [software program] were participating in this year's competition, "we'd probably get a zero," said de Moura. But the researchers have several benchmarks they're trying to hit before next year's event. They plan to fill in the holes in mathlib so that Lean can understand all of the questions. They also hope to have the detailed formal proofs of dozens of previous IMO problems, which will begin the process of providing Lean with a basic playbook to draw from. At that point a gold medal may still be far out of reach, but at least Lean could line up for the race." (Source) | 344: "Domain experts may have a greater understanding of what is missing from current ML systems, and of how hard it is to achieve the cognitive capacities that seem necessary to win the (technical) grand challenge."

- The F2F paper results may be less exciting than the summary implies; the AI in fact solved none of the IMO level problems.[1190]
- The IMO's rules for the Grand Challenge are "still tentative", so maybe this won't happen for a few years.[1191]

Cross-references with other questions

Q39: MATH Dataset Benchmark
Q40: Massive Multitask Language Understanding Benchmark
Q41: QuALITY Dataset Benchmark
Q43: NYT Bestsellers Written by AI

Question 43: NYT Bestsellers Written by AI

By what year will AI have written at least 3 books that appear on the New York Times Best Seller list?

Question and resolution details, prior forecasts, and other relevant sources

Results[1192]

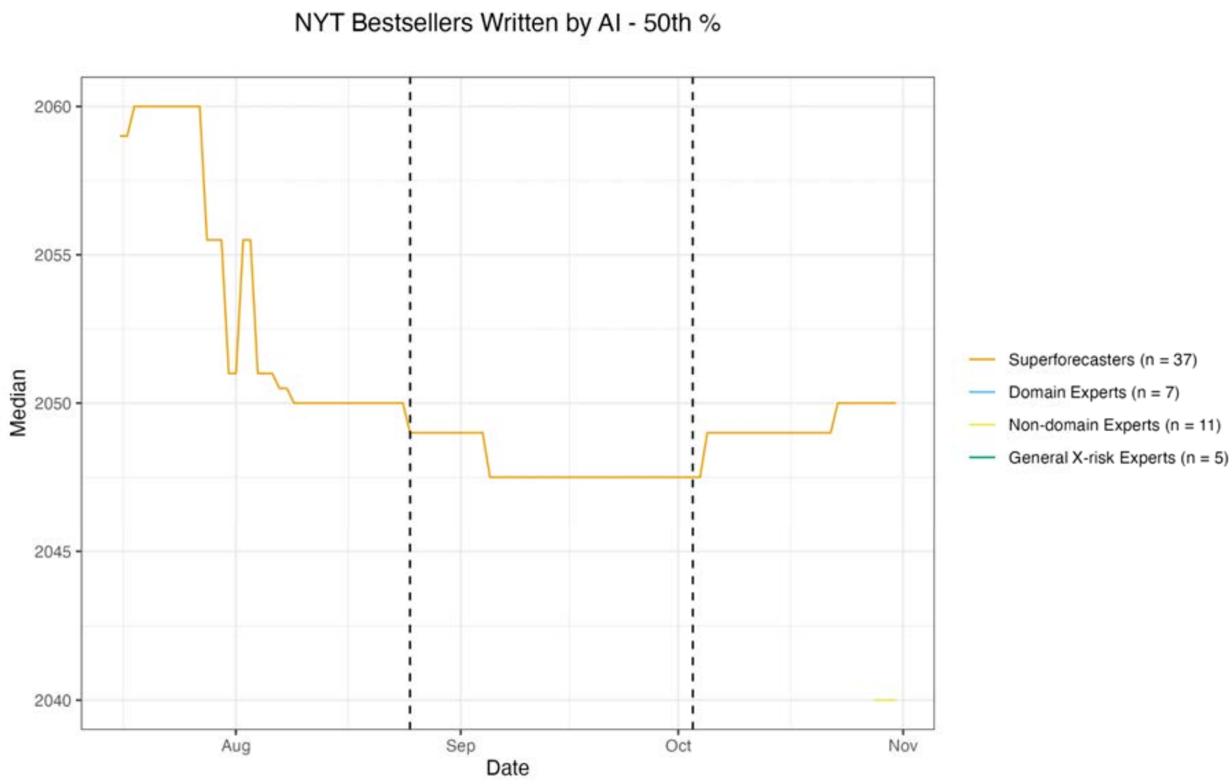
Group	Percentile Forecast	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 37)	5%	2030	2030	8.98	+19.67%
	25%	2040	2039	12.95	+119.38%

[1190] 336: "Another dissenting view, added late, is that forecasters are significantly overestimating the results so far based on summaries of the miniF2F paper, while the detailed results indicated that none of the IMO level problems were solved." / "IMO problems are significantly more difficult than the problems that can currently be solved."
[1191] 344: "Since the rules for the IMO Grand challenge are still tentative and there is no leaderboard, some forecasters think that this won't happen at least for a few years."
[1192] Numbers of forecasters are given as of Stage 4 of the XPT.

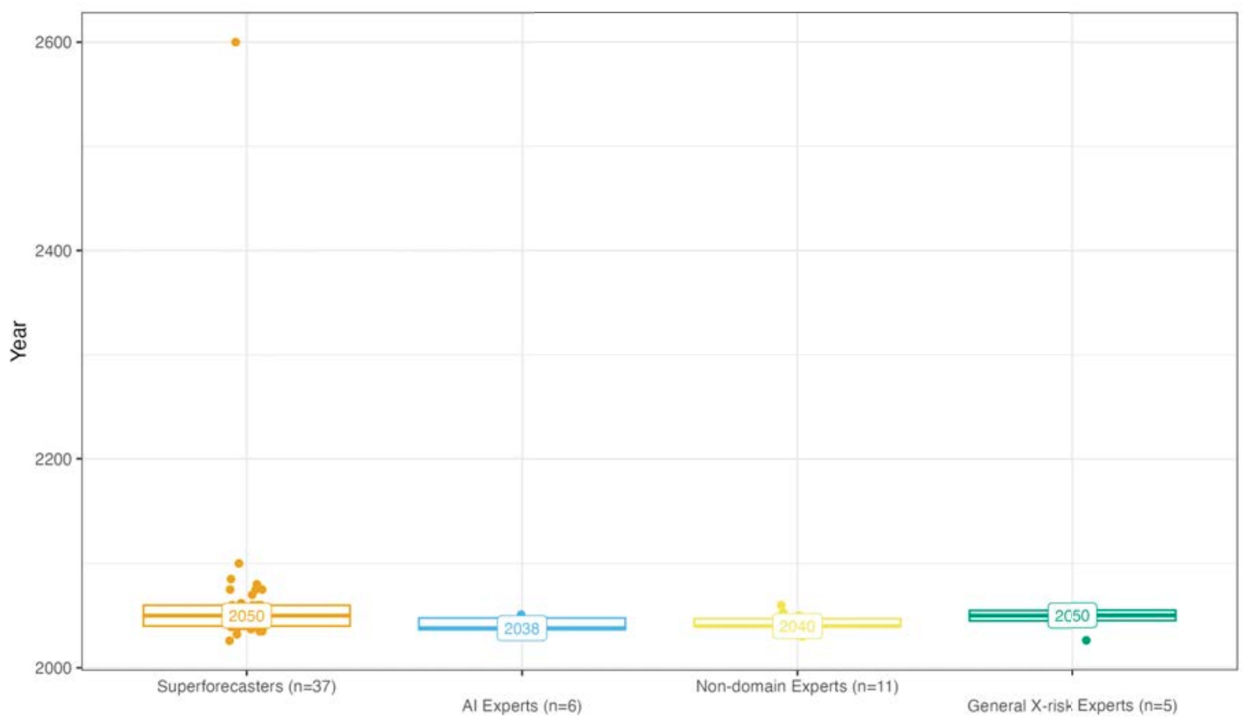
	50%	2059	2050	20.14	+354.68%
	75%	2075	2065	Inf	NaN
	95%	2100	2095	Inf	NaN
Domain Experts (N = 6)	5%	2025	2025.5	3.21	+54.19%
	25%	2031	2031	6.66	+77.05%
	50%	2038	2038	7.51	Inf
	75%	2060	2055	11.93	Inf
	95%	2100	2092.5	Inf	NaN
General X-Risk Experts (N = 5)	5%	NA[1193]	2027	NA	NA
	25%	NA	2035	NA	NA
	50%	NA	2050	NA	NA
	75%	NA	2082	NA	NA
	95%	NA	2150	NA	NA
Non-Domain Experts (N = 11)	5%	2028	2028	5.94	-72.02%
	25%	2035	2035	10.95	-48.46%
	50%	2040	2040	19.49	-54.91%
	75%	2060	2065	30.08	Inf

[1193] No forecasters in this group answered this question in Stage 1.

	95%	2100	2250	103.91	Inf



NYT Bestsellers Written by AI



Sources of agreement, disagreement and uncertainty

A few things were widely agreed on:

- Novelty would play a big role if AI is used to create a book which could reasonably be considered eligible for the NYT Best Seller list.
- The NYT Best Seller list is a neutral scoreboard of the books with the most copies sold. It is curated, and is affected by bias, rigging, marketing, and other factors.

Disagreement for this question focused on:

- Which direction human interference was likely to go, and what kinds of interference would be likely
- The level of technical difficulty (or possibility) of the task
- Whether AI-written books would make sense financially to produce
- What kinds of obstacles there may be to having three books on the list at the same time

Uncertainty was encountered around relevant base rates.

- Some teams did not mention base rates explicitly at all
    - Team 338 wrote: “There are no meaningful base rates, as this would be a first-time occurrence. It's not obvious what a relevant reference class would be for similar firsts.”
    - Team 343 wrote: “It is difficult to establish the base rate for this question. Which of AI's recent achievements would be an appropriate analogy for writing an NYT best-seller? One thing we have to keep in mind is that a number of recent

milestones were very unexpected, for example, AlphaGo beating a human champion in the game of go when many experts predicted it would happen decades later if at all, or very recent dramatic improvement by Google's Minerva on MATH dataset benchmark. On the other hand, in some domains the progress appears to be slower than predicted (for example, autonomous driving)."

Argument given for a median prediction of ≤ 2040:

Technical considerations:

- Writing a good book may eventually be easy for AI,[1194] especially compared to other tasks we've already seen AI do.[1195] [1196]
- Writing one book is not very different from writing three books.[1197]
- AI is getting better very quickly.[1198] [1199] [1200] Separately, AGI might come soon and be able to write compelling books well.[1201]
- The task will possibly be technically feasible soon.[1202] Similar-but-less-complex tasks are already possible today,[1203] [1204] and the task itself may actually be possible today.[1205] Or it may only require scaling existing technology.[1206]

[1194] 340, "It is all but guaranteed that eventually AI will be able to write books superior to those currently on the bestseller list." (2029-2040)

[1195] 339, Given the resolve conditions of the question, "Writing a book may not be one of the hardest tasks for AI to replicate." (2032)

[1196] 344, "Dall-e AI is very impressive, and it should be simpler for a text generator to read every book than to examine every painting." (2025-2030)

[1197] 339, "I don't find the difference in effort to write between one and three fiction books compelling." (2032) For an argument specifically related to the NYT Best Seller list, see also: 345, "Cracking the NYT Best Seller list three times in the same year is probably not that much harder than cracking it once." (2045)

[1198] 339, "LLMs have made huge strides over the last few years (GPT-2 2019, GPT-3 2020+). [...] Grace et al (2017/2018) has median time-since-2016 for AI [to "Write a novel or short story good enough to make it to the New York Times best-seller list"] at 33 years (= 2049), with quartiles at ~13 (= 2029) and ~72 (= 2088). This was before the release of GPT-2 and GPT-3 which I would judge as having made much more advance than AI experts would have expected in 2016. Hence I down-adjusted (somewhat arbitrarily)." (2032)

[1199] 344, "GPT-3 is making big progress" (2039.5)

[1200] 344, "GPT-3 is sure impressive today and has grown amazingly quickly." (2025-2030)

[1201] 337, "Lower end forecasts tended to be more confident in the near-term advent of AGI, which would be significantly more capable than any human in writing compelling books." (2030-2040)

[1202] 339, "The technology is not *that* far off today, so a project like this is possible." (2032)

[1203] 339, "GPT3 can already generate perhaps a paragraph long story sufficiently entertaining that a sufficient number of humans would rate it better than most human-only authors." (2032)

[1204] 340, "If AI can produce a poem that appeals to humans (it can now) it can write a page, a book chapter, a book, and a sequel." (2029-2040)

[1205] 339, "GPT-3 seems already perfectly capable of writing a book, with the right setup (for instance, having a book made of short stories)." Also 339, "It may even be technically possible today." (2032)

[1206] 344, "Seems like the difference between a short story with some details lost (GPT-3 status now) and a novel, followed then by an NYT best-selling novel, might just be a matter of scaling." (2025-2030)

- Some kinds of books that are typically included on the NYT Best Seller list can be very formulaic, which would make them easy for an AI to write. Notably: self-help[1207] and 'airport fiction'.[1208] [1209]
- AI will be better than us at knowing what humans want to read.[1210]
- Having humans involved in the training process[1211] and in the writing/editing loop[1212] in the way the question allows makes the task especially feasible.

Non-technical considerations:

- Novelty may help sales of an AI-written book,[1213] [1214] [1215] [1216] [1217] and novelty may even be able to compensate for a lack of quality.[1218]
- Similarly, AI itself may have cultural cachet that motivates the "intellectual class" to "understand and "relate to" AI" (340, 2029-2040).
- The NYT Best Seller list is gameable, and this could work to the advantage of an AI-written book,[1219] e.g. by a publisher who foresees high profits promoting their AI book.[1220]
    - Anticipating the counter-argument, one team downplayed it: "The 'AI will never be allowed into the list' objection seems to just be a 1%ish likely objection not a 5%ish one." (339, 2032)

[1207] 339, "Self-help, Money, Miscellaneous" category was expected to be "much easier for an AI to grasp. Self-help books are just so formulaic." 339 also mentioned "Dummies" books and conjectured about a possible "How to According to AI" series, and wrote: "Self-help books (or business/leadership advice books) might be the easiest of all since they have been usefully rewritten thousands of times already." (2032)

[1208] 339, Another "expect[s] fiction to be written more readily than most non-fiction, since story lines are both universally understood and universally appealing, and the source material is so much greater. Also, humans will consume dreck (or entertain it) than a similar serious non-fiction book." (2032) See also: 337, "page-turning airport thrillers are much more ripe for disruption" (2060).

[1209] 341, "fiction books can be formulaic, especially those written by teams of ghostwriters and published under brand author names rather than written by those authors" (2028-2035)

[1210] 339, "AI will soon know much better than humans what we like to read (from e.g. Kindle analysis of how quickly people read through, when they stop, when they highlight). Human writers will have a tough time to compete soon!" (2032)

[1211] 340, "The AI will be coached and trained by humans, humans that read books and understand "taste" and memes and fashion" (2029-2040)

[1212] 344, "Having a human being in the loop (with light editing as allowed) also negates some of AI's weaknesses (like with earlier chess engines)." (2025-2030)

[1213] 339, "Such content will be very buzzy and go viral." (2032)

[1214] 339, "The novelty factor mean that the first couple books of this variety would receive considerable attention, meaning publishers have an incentive to try it." (2032)

[1215] 338, "Strong arguments for near timelines include the possibility that AI-written books become a "fad"" (2029-2039)

[1216] 344, "The novelty might also help AI novels make the best-seller list." (2039.5)

[1217] 340, "Another plus is the novelty factor." (2029-2040)

[1218] 339, "This means the necessary quality does not have to rival that of other bestsellers." (2032)

[1219] 344, "NYT lists can be gamed, and if AI or a human wanted to push AI-written books, they might be able to." (2025-2030)

[1220] 339, "The nature of bestseller lists means that publishers can promote books they want sold (e.g. ones that had lower input costs, due to being written by AI [...])." (2032)

- The NYT Best Seller list has a low quality bar, and inclusion on it is somewhat random.[1221]
- An AI-written novel will likely be profitable for the humans behind it.[1222] [1223]
- The task will be accomplished by humans surreptitiously[1224] [1225] [1226]
- AI will be able to iterate, by having “unlimited opportunities to test its work in various minor publishing environments” (340, 2029-2040).

Argument given for a median prediction of ≥ 2045:

Technical considerations:

- The task is hard technically.[1227] [1228] [1229] [1230] Specifically, the requirement for the AI to write 99% of the text is a high bar.[1231]
- Current LLMs have not demonstrated “creativity,” and therefore AI is currently a long ways away from being capable of the task.[1232]
- Significant technical breakthroughs may be required.[1233] [1234]
- Another AI winter could come.[1235]

[1221] 338, “Strong arguments for near timelines include [...] the relatively low quality bar/random aspect of the bestseller list.” (2029-2039)

[1222] 339, “In theory, this process would be far faster and thus cheaper than conventional writing process.” (2032)

[1223] 340, “There’s also a profit incentive.” (2029-2040)

[1224] 339, “For sure they will sneak in pseudonymous AI works soon.” (2032)

[1225] 339, “we also believed one or more humans working with an AI in an allowable manner described in the linked article by The Verge (stringing together selected paragraphs of AI text) may sneak in majoritarily AI text into a bestseller supposedly written by that human author.” (2032)

[1226] 337, “Whole brain emulation” could allow “an AI [to] simulate the personality of a famous author and 'ghost write' novels on their behalf.” (2030-2040)

[1227] 340, “There are concerns that a limit to a total of 400-words in language model prompts is insufficient for humans to provide a rich and complex plot.” (2072.5-2175)

[1228] 340, “There is a possibility that this never happens. AI-generated writing would have a difficult time competing with human-generated writing.” (2072.5-2175)

[1229] 341, “Using Tesla AI learning for driving vs human as a benchmark, this may not be perfect but provides scale. Writing a novel or unique non-fiction would be more challenging.” (2045)

[1230] 336, “current language models are quite impressive and will probably scale as increases in model sizes continue, but I believe that most bestsellers will require some degree of long-term continuity and coherence …” (2050)

[1231] 339, “99% of books is a lot. So upper bound kept high to compensate.” (2060)

[1232] 341, “producing a bestseller is a combination of marketing and talent, and that talent consists of learning and creativity. LLMs have shown they can do learning, but the jury is still out on creativity, even if the image creators, like DALL-E are showing some creativity.” (2060-2100)

[1233] 345, “there could be a decade-long plateau between short and medium prose to long-form writing that is internally consistent and of sufficient quality to get onto the NYT Best Seller list, with more technical breakthroughs necessary than I'm at anticipating at first blush.” (2045)

[1234] 341, “If LLMs are not able to produce a book that would resolve this question, the timeline shifts further out, since that is the main focus of AI development at the moment.” (2060-2100)

[1235] 341, “There is an implicit assumption in the team’s forecasts of continued technological development, whereas we could easily enter into another AI winter if we get another deep global recession.” (2060-2100)

- If the task is not achieved by 2051[1236] or 2100,[1237] it's probably very hard.

Non-technical considerations:

- The NYT Best Seller list is gameable, and will likely be gamed to exclude AI books.[1238] [1239] [1240] [1241]
- The NYT Best Seller list is hard to get on: it requires luck, marketing, plus high quality.[1242] [1243] [1244]
- The human demand for these books will be low.[1245]
- The task as defined is incompatible with book-creation; editors change >1%.[1246]
- Plagiarism issues could be an obstacle.[1247]

[1236] 336, "If it hasn't been achieved by 2051 then it is probably harder than it looks from here or we're living in a world where AI progress is difficult." (2063)

[1237] 341, "if this question has not resolved positively in the second half of this century, then there may be some unsurmountable barrier that cannot be resolved, which could mean this wouldn't happen for centuries. That pushes up the 95th percentile." (2100)

[1238] 344, "if AI can create new amazing novels at the push of a button, NYT might separate its best-sellers list into "Human-created" and "AI-authored." This might happen after the first one or two are published and make the list, and the NYT might see the future of all-AI all the time before the third book comes out." (2089.5)

[1239] 341, "There is a potential barrier[] in that the NYT bestseller list is not driven by sales directly, but by a poll of sellers. One forecaster points out that the people polled may be biased against AI-written books and could produce biased answers." (2060-2100)

[1240] 336, "Ebooks were not added to the combined list until 10-15 years after they began to be available. The bestseller list rules evolve over time and the NYT might choose to exclude AI generated texts for some time." (2050)

[1241] 337, "The NYT staff is, at the very least, adjacent to the 'creative' demographics that these sorts of models will likely disrupt, and therefore it was noted that NYT could easily decide that AI written books were simply ineligible for the bestsellers list, a decision that has some precedent in their defacto banning of books published solely on Amazon. Such an event could easily occur after the first one or two appearances, but before the third." (2060 median; and ⅓ of this team put 95% at infinity because of this argument)

[1242] 338, "Strong arguments for longer timelines include the [...] general difficult[y] of non-famous people getting on the bestseller list" (2060-2101)

[1243] 341, "it is tough for a book to become a bestseller. Not every decently-written [one] makes it onto the NYT bestseller list. It not only takes a high level of quality, but also has aspects of luck, the right kind of marketing, etc." (2060-2100)

[1244] 341, "producing a bestseller is a combination of marketing and talent" (2060-2100)

[1245] 340, "Since AI lacks understanding, its books will be devoid of any idea worth reading. Simulated writing isn't going to sell." (2072.5-2175)

[1246] 340, "editors generally change >1% of the words in a text, so this would involve going through a non-traditional process." (2072.5-2175)

[1247] 341, "plagiarism can have a limiting impact on sales." (2045)

- Three books on the list is harder than one.[1248] AI books could be sequestered to a special part of the list after the first one[1249]; or the first book could benefit from the novelty factor, but not subsequent ones.[1250]
- Books might stop being bought and sold sometime soon after the task is technically feasible.[1251]
- It would cost a lot of money.[1252]
- The publishing industry is slow-moving.[1253]
- Serious global catastrophe could set back all tech, or at least cause a world where people wouldn't be likely to buy AI-written books.[1254] [1255]
- Misaligned AI would probably not write bestselling novels.[1256]
- It might not happen if no one tries.[1257]
- A longer timeline is consistent with conclusions reached by others who have considered similar questions.[1258] [1259]

Other arguments given

Arguments in favor of lower forecasts:

- Some arguments were advanced that were identical to arguments for longer timelines, but their adherents nonetheless had relatively short timelines:

[1248] 336, "Three books are a higher bar than getting one book on the bestseller list." (2050)

[1249] 344, "if AI can create new amazing novels at the push of a button, NYT might separate its best-sellers list into "Human-created" and "AI-authored." This might happen after the first one or two are published and make the list, and the NYT might see the future of all-AI all the time before the third book comes out." (2089.5)

[1250] 344, "One book might make the list as a novelty without the usual quality, but no others, or no others until they improve to make it on quality alone." (2089.5)

[1251] 343, "it's likely that soon after the AI reaches the required capabilities, books will no longer be sold commercially." (2045)

[1252] 336, "Increasing context enough to achieve book length text would come at a high cost with current model designs." (2050)

[1253] 338, "Strong arguments for longer timelines include [...] the long lead times in the publishing industry (i.e., the books that will make the 2023 bestseller list are already written." (2060-2101)

[1254] 344, "If nuclear war or AGI or the next global pandemic set back civilizational progress to the pre-industrial age (or any age from up to now of course), we can expect no AI generated novels." (2089.5)

[1255] 336. "One reason cited by multiple people was the potential for a catastrophe, for instance one that prevented AI-written books from being widely bought or sold." (2050)

[1256] 344, "A misaligned AGI would probably have no desire to write best-selling novels. In this case, it may never occur." (2089.5)

[1257] 336, "This might never happen if no one is trying, or if the books generated don't meet the exact definition." (2063)

[1258] 344, A related Metaculus question asked "whether a book written by a language model will make the NY Times Bestseller list before 2030. As of 1 Jan 2022, the community prediction there for that question is 20%" (2050-2089.5)

[1259] 336, "The World Economic Forum survey of 350 AI experts had a median estimate of 2049 for when an AI would write a NYT bestseller." (2050)

a. Current AI is not very good at writing. Bilingual speakers who compare AI translations can see this more clearly than others.[1260] Also, if you have used AI story generators yourself, you can see how bad they are.[1261] (2032 median)
b. The NYT Best Seller list is likely to be biased or rigged against AI books.[1262] [1263] (2032 median)

- One forecaster adjusted their upper forecast downwards, from 2150 to 2110, out of consideration of the AI-generated picture that won an art prize.

Arguments in favor of higher forecasts:

- It's a risky financial proposition: publishers will want to see a proof-of-concept first.[1264]
- Copyright or legal issues might prevent AI-written books from being published.[1265]
- It will take time for authors to shift to an editing role.[1266]

Cross-references with other questions

Q39: MATH Dataset Benchmark
Q40: Massive Multitask Language Understanding Benchmark
Q41: QuALITY Dataset Benchmark
Q42: AI Wins International Mathematical Olympiad

[1260] 339, "If you speak two languages and spend any time with Google Translate, Babel, or any of the automatic translators, you can see that we are a long way away from language models which have enough mastery of a language to be able to subtly nuance anything." (2032)
[1261] 339, "You can see the dangling participles, sentences ending in prepositions, and the rather juvenile output." (2032)
[1262] 339, "The NYT Best List is curated so as not [to] include sensationalist writing. So, this is a bit of a higher bar than just outright sales. Additionally, it will prevent an AI book from rocketing to the top because of consumer curiosity driving sales." (2032)
[1263] 339, "It is possible AIs could specifically be banned from bestseller lists soon, in a way chess AIs are banned from chess competitions. So maybe there is a window of opportunity for AI to write a bestseller that is closing." (2032)
[1264] 339, "Publishers would probably need to see proof of concept before investing heavily enough in a best-seller quality book (and the necessary advertising). So delay of at least 18 months before bestsellers written by AI seems likely." (2032)
[1265] 339, Big companies like OpenAI could get scared off, which would slow things down. "I imagine that if OpenAI (or whoever manages the writer AI) chooses to not be on board, then it could take decades." (2032)
[1266] 339, "Some learning curve for authors to shift into an editing role (and develop the necessary skills for centaur work) explains the main estimate being 2030. That some transition period where AI writes large sections but not 99% of the books is the reason for the spread in the estimates." (2030)

Question 44: Date of Advanced AI

When will the first unified AI system meeting all of the following criteria be trained, tested, and publicly known of?

1. Able to reliably pass a 2-hour adversarial Turing test.
2. High competency at answering questions across diverse fields of expertise.
3. High competency on interview-level problems in the APPS benchmark.
4. Able to learn the classic Atari game “Montezuma’s revenge” in the equivalent of 100 hours or less of real-time play.

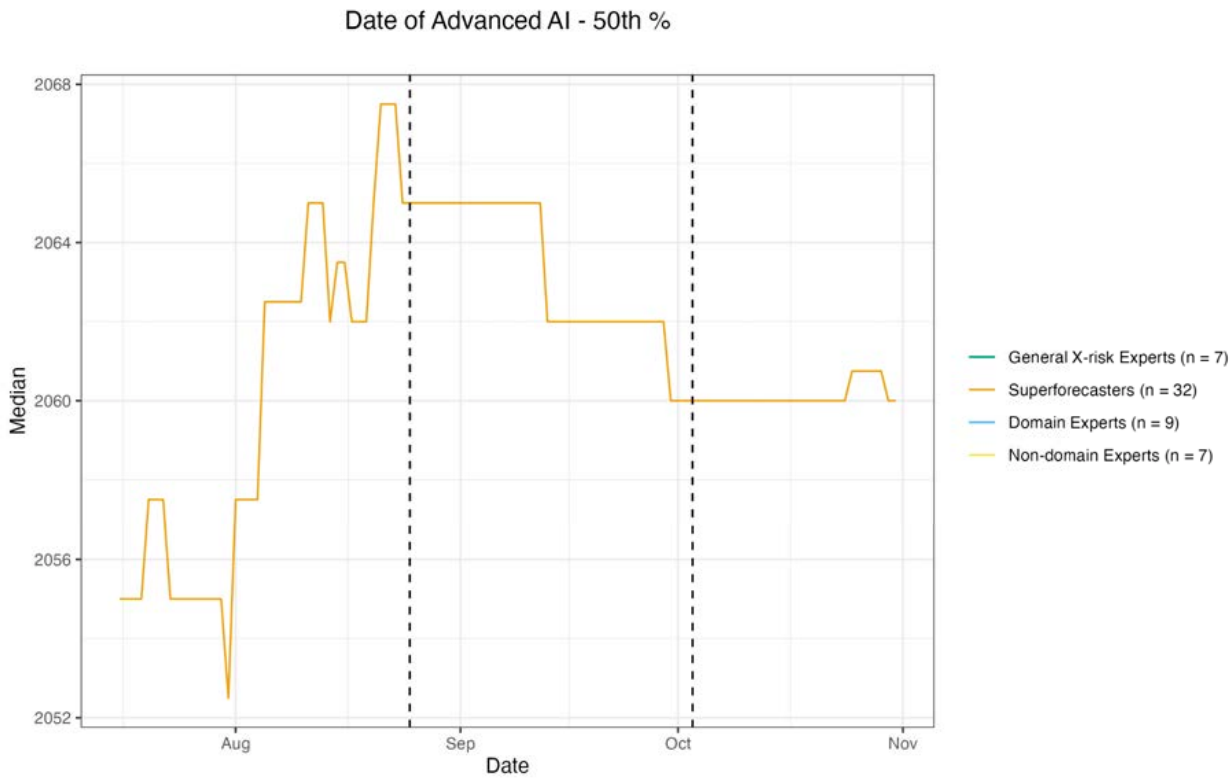
Question and resolution details, prior forecasts, and other relevant sources

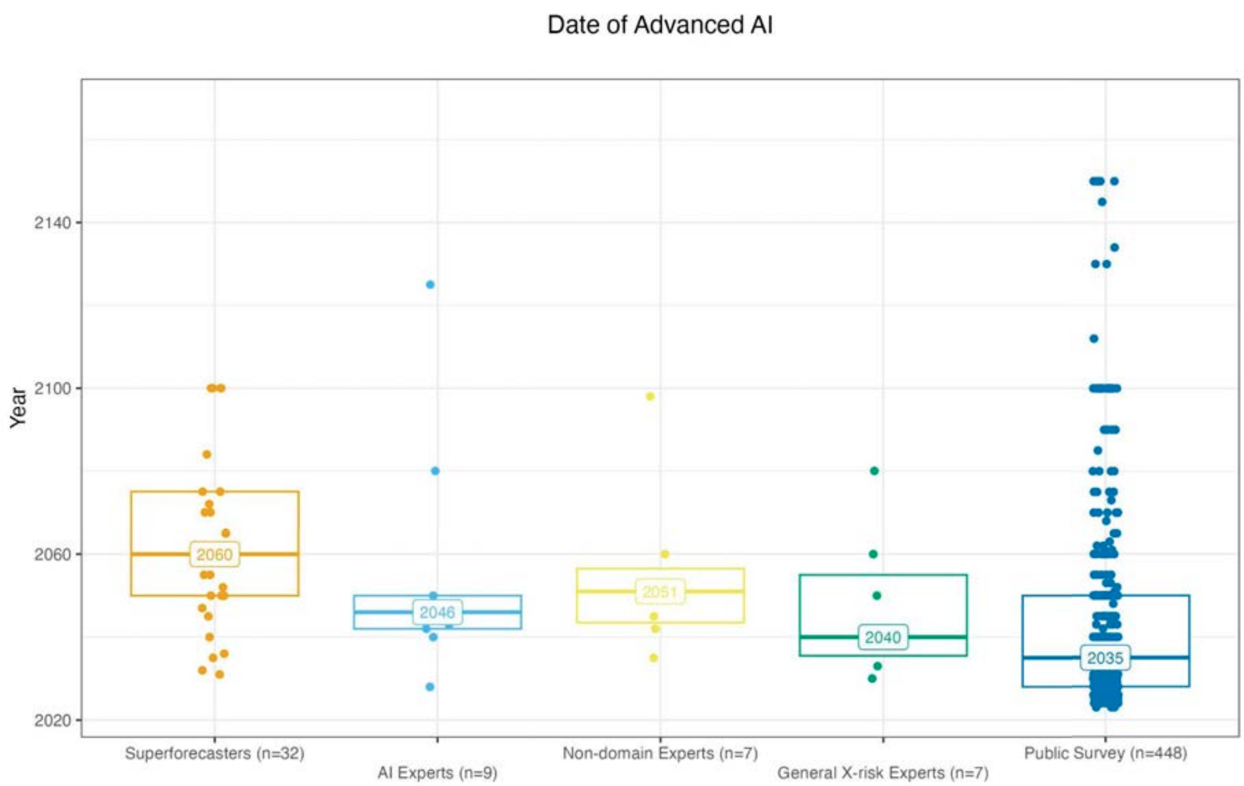
Results[1267]

Group	Percentile Forecast	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 33)	5%	2037.5	2035	10.70	-11.1%
	25%	2048	2049	24.64	-25.46%
	50%	2055	2060	91.60	-38.44%
	75%	2070	2085	Inf	NaN
	95%	2100	2120	Inf	NaN
Domain Experts (N = 9)	5%	2025	2029	2.95	+79.01%
	25%	2032	2035	7.7	+1.93%
	50%	2038	2046	21.84	+34.87%

1267 Numbers of forecasters are given as of Stage 4 of the XPT.

	75%	2050	2065	51.89	Inf
	95%	2065	2125	89.98	Inf
General X-Risk Experts (N = 7)	5%	2024	2027	1.91	+134.38%
	25%	2029.5	2031	6.08	+9.76%
	50%	2044	2040	13.2	+34.12%
	75%	2068.5	2062	26.25	+1,258.23%
	95%	2220	2100	Inf	NaN
Non-Domain Experts (N = 7)	5%	2032	2032	NA	NA
	25%	2034	2041	NA	NA
	50%	2035	2051	NA	NA
	75%	2043	2065	NA	NA
	95%	2045	2090	NA	NA
Public Survey (N = 448)	50%	2035		Inf	-


Date of Advanced AI - 50th %
Median
2068
2064
2060
2056
2052
Date
Aug
Sep
Oct
Nov
General X-risk Experts (n = 7)
Superforecasters (n = 32)
Domain Experts (n = 9)
Non-domain Experts (n = 7)


Date of Advanced AI
Year
2140
2100
2060
2020
2060
2046
2051
2040
2035
Superforecasters (n=32)
AI Experts (n=9)
Non-domain Experts (n=7)
General X-risk Experts (n=7)
Public Survey (n=448)

Sources of agreement, disagreement and uncertainty

Disagreement for this question focused on:

1. The difficulty of the individual tasks
2. The difficulty of generalizing to all of the tasks
3. The nature of AI progress in general
4. How to interpret prior forecasts on similar questions

One team gave the following interpretation for differences in forecasts:

> "A few forecasters have very short AI timelines (within a decade on their median forecast) based on recent trends in AI research and the impressive models released in the last year or two. Other team members have longer timelines. The main difference seems to be projecting from very recent trends compared to taking the long view. In the recent trends projection forecasters are expecting significant growth in AI development based on some of the impressive models released in the past year or two, while the opposite long view notes that the concept of AI has been around for ~75 years with many optimistic predictions in that time period that failed to account for potential challenges."[1268]

Forecasters also disagreed about whether developing advanced AI would be easier or harder than developing AGI or transformative AI, but this disagreement did not map directly onto lower or higher forecasts.[1269] Comparison to question 51 on AGI suggests that forecasters generally interpreted advanced AI as easier to develop than AGI.

Sources of uncertainty mentioned:

[1268] 341.

[1269] 340, "THE STRONGEST ARGUMENT FOR is most easily explained assuming advanced AI is equivalent to Transformative AI…THE STRONGEST ARGUMENT AGAINST is also most easily explained assuming that Transformative AI is required"; referring to the team median of 2050. 343, "The prompt is not asking for a prediction of the "Date of AGI", or "the date of an AGI singularity", however a few of our forecasters used these concepts as a shorthand as they viewed it as an operationalization of that term. The prompt specifies a defined set of criteria to be achieved, which is different from the more nebulous concept of general artificial intelligence, or of superintelligence, and therefore the criteria themselves may be satisfied without achieving AGI in a more ambitious sense such as Karnofsky's PASTA ("Process for Automating Scientific and Technological Advancement")"; 343, "The biological anchors method seems to be the most widely acknowledged one for predicting AGI. It assumes that computing power is the main bottleneck. Ajeya Cotra is a leading specialist on this method. She currently predicts "35% probability of transformative AI by 2036, 50% by 2040, and 60% by 2050" (https://80000hours.org/problem-profiles/artificial-intelligence), i.e. a median of 2040. Satisfying the criteria in the prompt may be somewhat easier." The team median here was 2040. 344, "In 2020, Ajeya Cotra used biological anchors to estimate a median of ~2050 for transformative AI (which would probably be able to satisfy the requirements in this question, although that is not certain)", as an argument for a forecast of 2045. 338, "The range of mastery across tasks gets very close to the requirements for an Artificial General Intelligence (AGI)", as an argument for a forecast of 2065.

- The difficulty of establishing base rates.[1270]
- The exact nature of the Turing test.[1271]
- Whether using ML to do better ML will yield results.[1272]

Arguments given for forecasts of ≤2050

On 1 (difficulty of individual tasks):

- Narrow AI systems are approaching human-level performance on all four criteria individually.[1273]
    - Answering questions across diverse fields of expertise: there has been strong recent progress.[1274]
    - APPS: GPT-Neo can already pass 20% of the introductory problems.[1275]
    - Montezuma's revenge: there are AI systems which can play this game, though not from scratch and visual input only.[1276]

On 2 (difficulty of generalization):

- DeepMind's Gato shows that generalizing on diverse tasks is possible.[1277]

On 3 (AI progress in general):

- Current state of the art is impressive.[1278]
    - PaLM outperforms humans on various text-based tasks.[1279]

[1270] 338, "As a source of uncertainty, the lack of base rates seems like a strong candidate. The problem has two parts 1) high accuracy in the individual benchmarks, and 2) unifying the solutions of the separate tasks into a single model.   Limited information is available for the first part since even current SOTA in these benchmarks isn't obvious, let alone finding estimates of future progress."

[1271] 340, "The nature of the Turing test and who takes it, how many people take it, etc. If average humans are taking it the question will resolve more quickly than if experts in AI or cognitive psychology were taking it."

[1272] 343, "Using machine learning as a compound method for discovery is a recent development, which may either deliver very limited improvement or dramatically change the field as it matures."

[1273] 336, "Current narrow AIs are approaching human-level performance on all 4 tasks individually." 337, "The particular moving pieces (conditions 1,2,3,4 in this question) required for advanced AI have seen impressive progress in the last few years."

[1274] 337, "For instance, for condition 2 (Multi-task Language Understanding) there has been remarkable progress [14], but with limitations (top-k models instead of top-1)." The reference cited is https://paperswithcode.com/sota/multi-task-language-understanding-on-mmlu. Top-1 accuracy is distinguished from top-k accuracy in which k outputs from the model are generated, and the best output is selected.

[1275] 337, "For condition 3 (Coding Interview), GTP-Neo can already pass 20% of the introductory problems of the AAPS coding challenge [15]." The reference cited is https://arxiv.org/abs/2105.09938v3.

[1276] 337, "For condition 4 (Gaming), there are indeed AIs that play them, but not from scratch and visual input only."

[1277] 336, "Deepmind's Gato has already shown that unified AI on diverse tasks is possible."

[1278] 344, "Justifications for the possibility of near AGI include the impressive state of the art in language models."

[1279] 344, "PaLM already surpasses the human average in a variety of textual tasks (see Figure 3)".

  - There are systems which outperform the median human participant in programming competitions.[1280]
- Recent progress in AI has been fast.[1281]
- If the scaling hypothesis is true, no theoretical breakthroughs will be required to develop this system.[1282]
- Developing this sort of system is an active priority for many organizations, and would be very profitable.[1283]

On 4 (prior forecasts):

- Recently ML research has tended to meet milestones faster than expected.[1284]
- Meeting these criteria may be easier than achieving Cotra's transformative AI.[1285]
- One team cited Cotra's 2022 update to her forecasts as their strongest argument.[1286]
- This and similar Metaculus questions may increase the effort going into building a system like this.[1287]

Other arguments given:

[1280] 344, "there are already systems which can apparently surpass the median human result of programming competition participants."
[1281] 336, "General rapid progress in all fields of computing and AI, like bigger models, more complex algorithms, faster hardware."
[1282] 343, "The scaling hypothesis says that no further theoretical breakthroughs are needed to build AGI. It has recently gained in popularity, making short timelines appear more credible."
[1283] 336, "Development of a system of this type is actively researched by multiple private and public orgs, and would be extremely profitable."
[1284] 343, "Recent track record of ML research of reaching milestones faster than expected
Development in advanced computer algorithms and/or AI research has a strong history of defining what seem to be impossible tasks, and then achieving those tasks within a decade or two. Something similar may well happen with this prompt, i.e. researchers may build a system capable of achieving these goals, which are broader than ones already reached, as early as in the late 2020s or early 2030s. Significant gains in AI research have been accelerating recently, so these estimates may be low if we're on the cusp of exponential growth in AI capabilities." See also 341, "Another argument for shorter AI timelines is the recent trend in AI development. If improvements to language models increase at the rate they have recently, such as between GPT-2, GPT-3, and the rumored upcoming advances of GPT-4, then we might expect that the Turing test requirement could be accomplished within a decade. Additionally, recently developed generalist models such as Gato, along with other new models, have surprised many observers, including the Metaculus community. The Metaculus "Weak AGI" question median has moved 6 years closer this year."
[1285] 343, "The biological anchors method seems to be the most widely acknowledged one for predicting AGI. It assumes that computing power is the main bottleneck. Ajeya Cotra is a leading specialist on this method. She currently predicts "35% probability of transformative AI by 2036, 50% by 2040, and 60% by 2050" (https://80000hours.org/problem-profiles/artificial-intelligence), i.e. a median of 2040. Satisfying the criteria in the prompt may be somewhat easier." See also 344, "In 2020, Ajeya Cotra used biological anchors to estimate a median of ~2050 for transformative AI (which would probably be able to satisfy the requirements in this question, although that is not certain)", as an argument for a forecast of 2045.
[1286] 338, "The low end forecast is 2025-2026 with the strongest justification being the reference to the Ajeya Cotta Report base rate to support that forecast."
[1287] 337, "The possible biases of the committee, which might lean towards conceding that this milestone has been achieved."

- The resolution committee may be biased.[1288]

Arguments given for forecasts of ≥2060

On 1 (difficulty of individual tasks):

- The individual criteria are all challenging.[1289]
    - Turing test:
        - GPT-3 is one of the better current language models, and it is easy to confuse.[1290]
        - Adversarial Turing tests may get harder to pass as we improve our understanding of which questions are most discerning.[1291]

On 2 (difficulty of generalization):

- Generalizing to all four criteria will be challenging.[1292]
    - The benchmarks which have seen most progress are popular, and may have been selected for being achievable. Other benchmarks may receive less attention.[1293]

[1288] 337, "The interest coming from the Metaculus question itself, which might influence the development of AI itself and Godhart's law ("When a measure becomes a target, it ceases to be a good measure") [8]. That is, efforts in the AI community might be focused on achieving advanced AI (as defined in this question)". The reference cited is https://en.wikipedia.org/wiki/Goodhart%27s_law.

[1289] 341, "They note that the individual tasks listed in the question requirements can be very challenging."

[1290] 341, "For example, the adversarial Turing test where text, images, and audio files may be used is well beyond the capabilities of any existing AI. GPT-3 is one of the better language models and can be easily confused when asked strategic questions."

[1291] 344, "Also, given the adversarial Turing test, it is possible that the requirements will get harder with the passing years, as judges are provided a better understanding of what questions will advantage the human participants."

[1292] 341, "Due to the challenging level of individual tasks, forecasters also expect generalizing to being able to accomplish multiple tasks at the required capability will be very difficult"; "Some teammates expect that while AI may advance in individual tasks, generalizing may be much more difficult than expected. So far most advancements in AI models have been based on using large amounts of training data with large amounts of parameters, but it may require more computing power and more data than will exist based on current trends in order to create powerful models that are both talented in one specific field but also generalize to other fields." See also 343, "Narayanan and Kapoor (Why are deep learning technologists so overconfident? (substack.com)) are on a mission to convince people that current expectations are often too aggressive and that even the best contemporary systems are still highly specialized. An adversarial Turing test would presumably detect failure of generality." See also 336, "It´s four different capabilities in one single system required for resolution as "yes". With probably nobody going for having exactly these four in the portfolio (and nothing else), we might need to wait for a superior system, that can do much more, with these 4 requirements only as a byproduct. If that is the case, it seems plausible that proving it can win a game of Montezuma's revenge might not be top priority." See also 338, "Some of the criteria seemed more likely to [be] met sooner, in particular 2 and 3. However, the cumulative resolution of all the criteria is much more challenging and rendered the question more difficult. Another possible source of uncertainty is whether AI researchers in the future will choose to work on those problems due to funding or academic trends. For example, if AI researchers choose another game instead of Montezuma's revenge then this question may not resolve."

[1293] 341, "Forecasters also note that some of the benchmarks that have shown the most progress (such as ImageNet) are popular and commonly targeted by AI researchers. This may be a case of AI researchers selecting for more achievable goals that better align with AI capabilities, and it could also be

    - It may not be possible to get this kind of generalization just with scaling, and algorithmic breakthroughs may be required.[1294]
    - A system like this would be close to AGI.[1295]

On 3 (AI progress in general):

- AI progress may be intentionally slowed down in this period.[1296]
- Human progress in general may be interrupted or slowed in this period,[1297] or humans might go extinct.[1298]
- There may be unforeseen issues with developing this technology.[1299]

On 4 (prior forecasts):

- Metaculus tends to underestimate how challenging generalization is.[1300]

---

the case that some of these tasks aren't as interesting or are less likely to receive attention for AI researchers."

[1294] 338, "The current pace of progress may falter if this kind of general task solving can not be achieved by just continuing to upscale existing systems." 344, "Justifications for longer timelines include the possibility of unforeseen issues with the development of the technology, events that could interrupt human progress, and the potential that this achievement will require significant algorithmic breakthroughs."

[1295] 338, "The range of mastery across tasks gets very close to the requirements for an Artificial General Intelligence (AGI)."

[1296] 344, "The possibility of limiting the intentionally slowing down progress is also taken into consideration as a factor for a later emergence".

[1297] 344, ""Justifications for longer timelines include the possibility of unforeseen issues with the development of the technology, events that could interrupt human progress, and the potential that this achievement will require significant algorithmic breakthroughs."

[1298] 338, "The high end forecast is that the goal is never achieved. The justification is that at long dates the likelihood of human extinction is above 5% so advanced AI development is not achieved". See also 343, "An existential catastrophe might cut short the advance of science. Theoretically, an AGI could cause radical changes to happen (including, but not limited to, the extinction of humanity) that make it impossible for the tests mentioned in the prompt to be performed."

[1299] 344, "Justifications for longer timelines include the possibility of unforeseen issues with the development of the technology, events that could interrupt human progress, and the potential that this achievement will require significant algorithmic breakthroughs." See also 343, "The biological anchors method assumes that no major further bottlenecks are lurking"; "Previous ML paradigms expanded capabilities dramatically with the advent of new tools/paradigms, only to run up against the limits of what that paradigm was capable of accomplishing. Simple trendline extension masks this effect. So long as we are in the exponential growth phase of a sigmoid curve, it's impossible to determine where the inflection point will be."

[1300] 341, "The question being considered would be easier to accomplish than the Metaculus "AGI" question which is currently at a median of 2043, so we can expect that Metaculus would assign a sooner date for the question being considered. However, the team likely believes the Metaculus community is underestimating the difficulty involved in an AI generalizing to accomplish several tasks of which it has not yet accomplished nearly any of the tasks and still seems years or even a decade or two away from the more challenging of these." See also 337, "It has been argued [6] that Metaculus has also an optimistic bias for their technological predictions." The reference cited is https://forum.effectivealtruism.org/posts/vtiyjgKDA3bpK9E4i/an-examination-of-metaculus-resolved-ai-predictions-and#.

- Experts tend to be too optimistic on AI development.[1301]

Other arguments given

Arguments in favor of lower forecasts:

- On 4 (prior forecasts): forecasts from those who have thought a lot about it tend to be low.[1302]

Arguments in favor of higher forecasts:

- On 3 (AI progress in general): there may be diminishing returns, such that the final 10% takes much more work than the first 90%.[1303]
- A system like this might be developed but kept secret.[1304]

Cross-references with other questions

Q51: Nick Bostrom Affirms Existence of AGI
Q10: Total Extinction Risk

[1301] 341, "Forecasters also note the optimism that experts in a field typically have with regard to the rapidity of developments." See also 337, "Past predictions about present technology have been in general excessively optimistic with regards to the pace of progress, though it should be stressed that this is not always the case. In particular, AI predictions seem to have fallen on the optimistic side. Good examples are previous predictions about AI [4] and past Kurzweil's predictions for the year 2019 [5]." The references cited are https://en.wikipedia.org/wiki/Progress_in_artificial_intelligence#Past_and_current_predictions and https://web.archive.org/web/20100421224141/http://en.wikipedia.org/wiki/The_Age_of_Spiritual_Machines#2019.

[1302] 341, "One of the stronger arguments for forecasts on the lower end are the forecasts from those with expertise who have made predictions, such as Ajeya Cotra and Ray Kurzweil, as well as the Metaculus community prediction. Cotra and Kurzweil have spent a lot of time thinking and studying this topic and come to those conclusions, and the Metaculus community has a solid track record of accuracy. In the absence of strong arguments to the contrary it may be wise to defer somewhat to those who have put a lot of thought and research into the question."

[1303] 343, "One phenomenon we've observed in, e.g. driverless cars, is that the first ~90% of the problem takes about 10% of the work, with diminishing returns per unit of work the closer you get to 100%. The prompt wouldn't necessarily require 100% on each of these tests, but it might be close enough that a conventional approach to forecasting may fail to capture the amount of work required at the tail end of the project." 337, "From the application of machine learning to the specific domain of material science [7], a possible conclusion is that a lot of recent progress can be classified as proof of concept or as low-hanging fruit. Provided this can be extrapolated to other fields, a slowdown is to be expected." The reference cited is https://www.nature.com/articles/s41524-022-00734-6.

[1304] 343, "There's some speculation that a system that passes these criteria might be developed, but kept secret. This would likely be from a major government, as most private research firms would likely publish their findings."

Question 45: Maximum Compute Used in an AI Experiment

What will be the maximum compute (measured in petaFLOPS-days) used for training in an AI experiment…
...by the end of 2024?
...by the end of 2030?
…by the end of 2050?

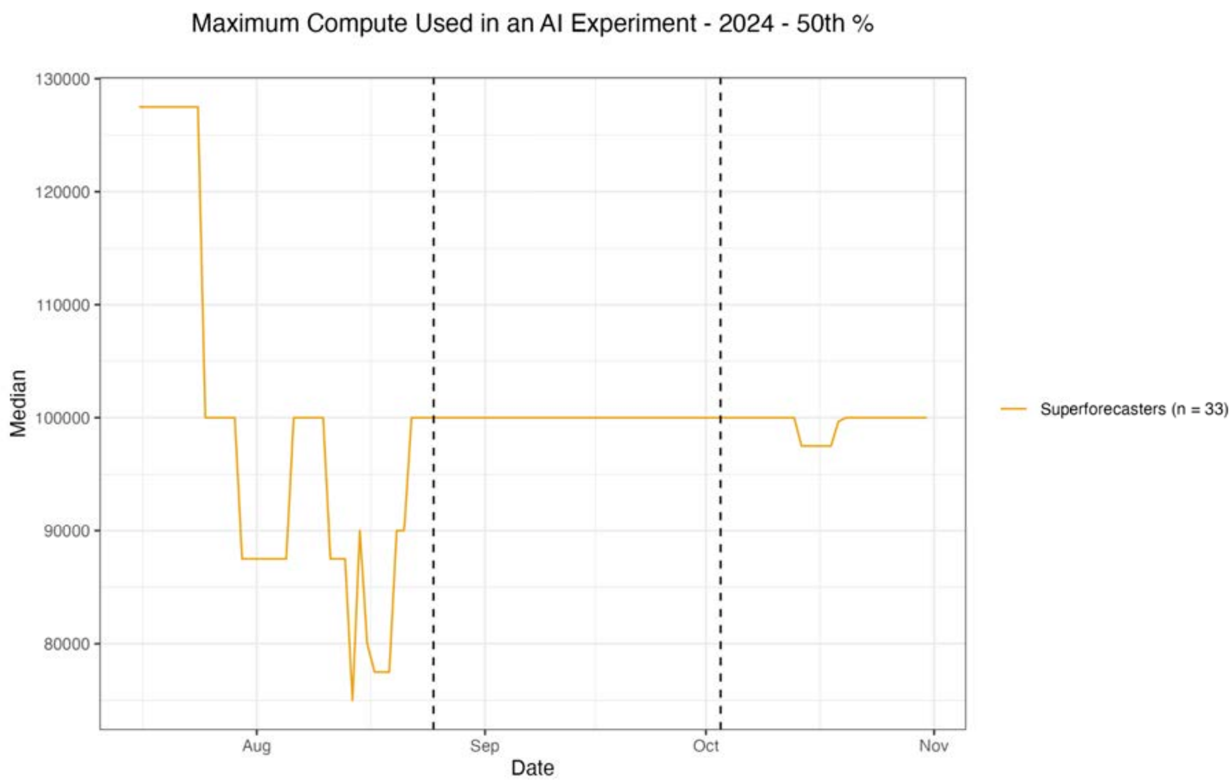
Question and resolution details, prior forecasts, and other relevant sources

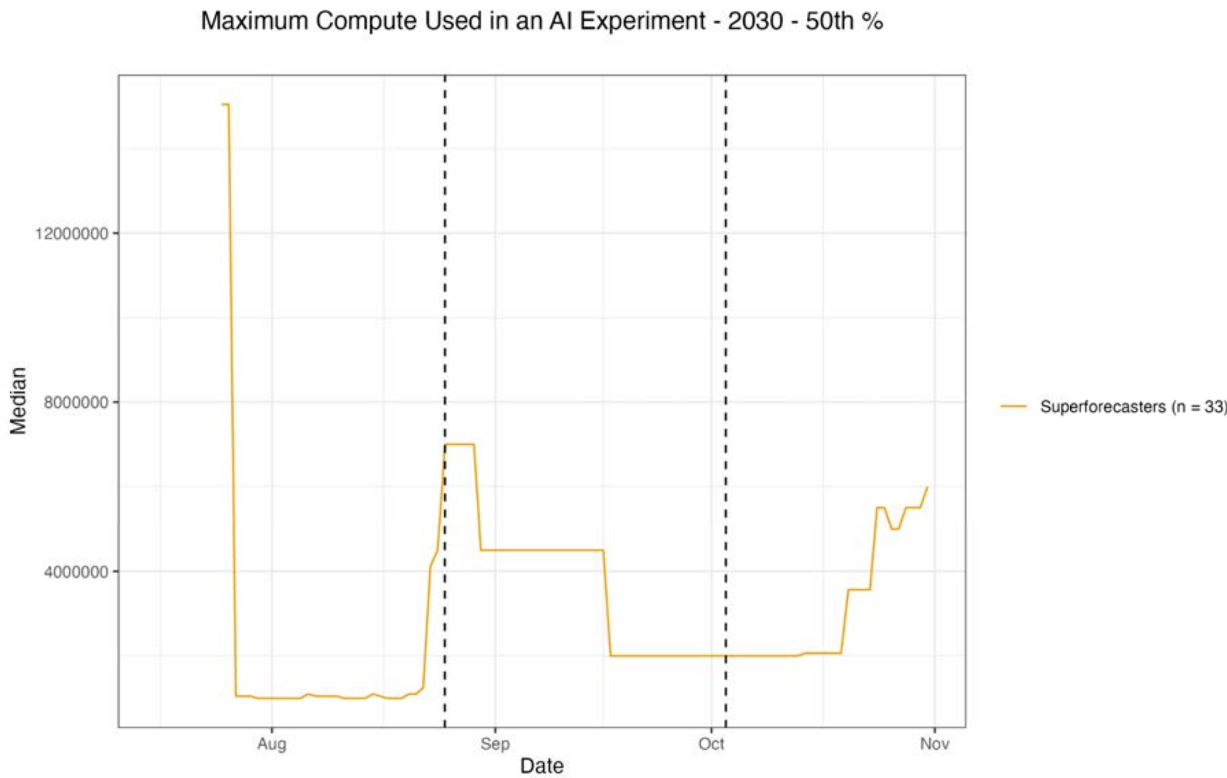
Results[1305]

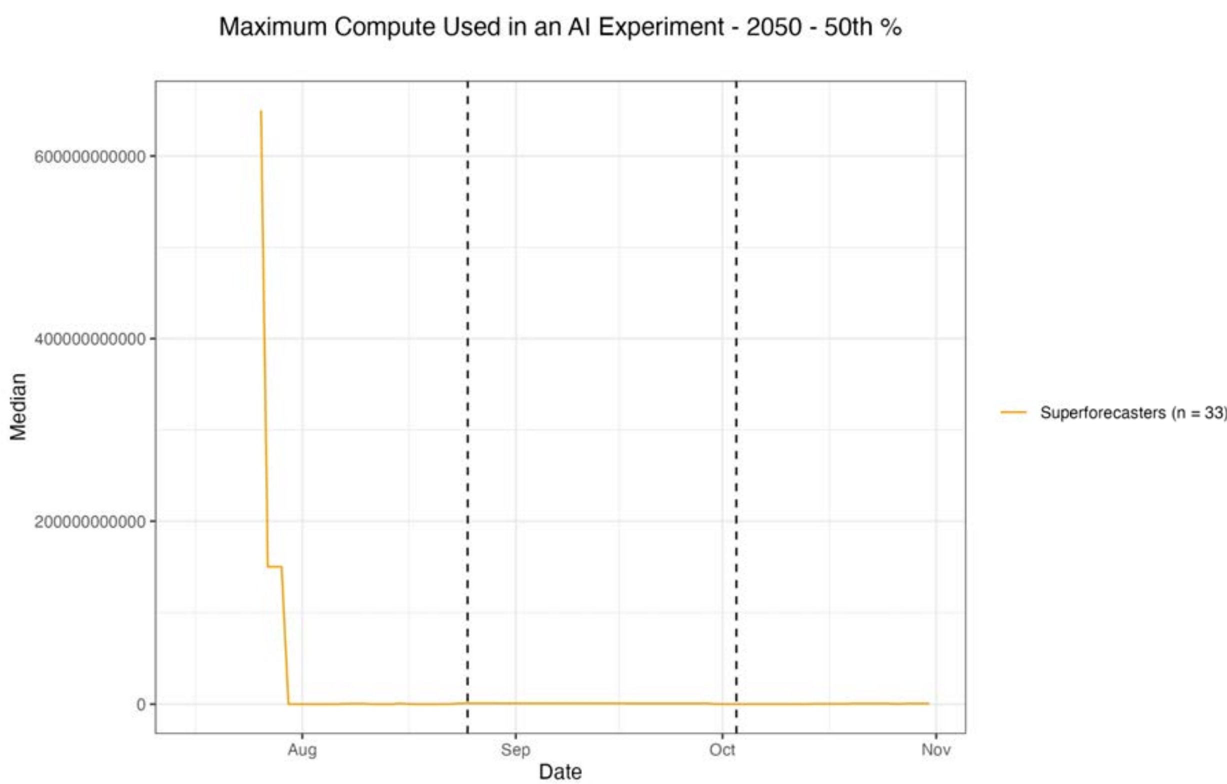
Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 33)	2024	210,000	100,000	10549500426	+354.2%
	2030	35,500,000	6,000,000	756086281737353	+6042.88%
	2050	1,100,000,000	700,000,000	7999829207045185536	+5437593823142679150592%
Domain Experts (N = 2)	2024	2,000,000	420,680	NA[1306]	NA
	2030	100,000,000	25,114,410	NA	NA
	2050	5,000,000,000,000	2,500,054,061,484	NA	NA
	2024	155,594.5	218,308	135191	+34.22%

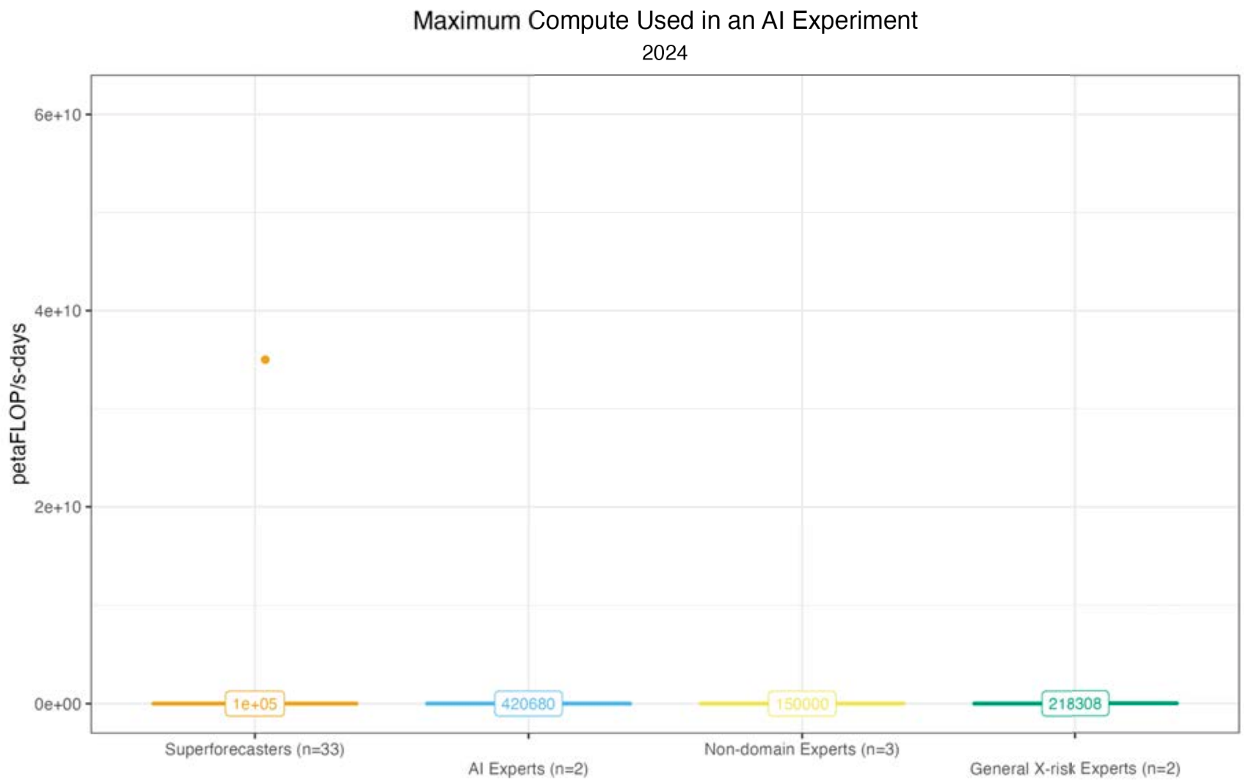
[1305] Numbers of forecasters are given as of Stage 4 of the XPT.
[1306] Only one forecaster in this group answered this question in Stage 1.

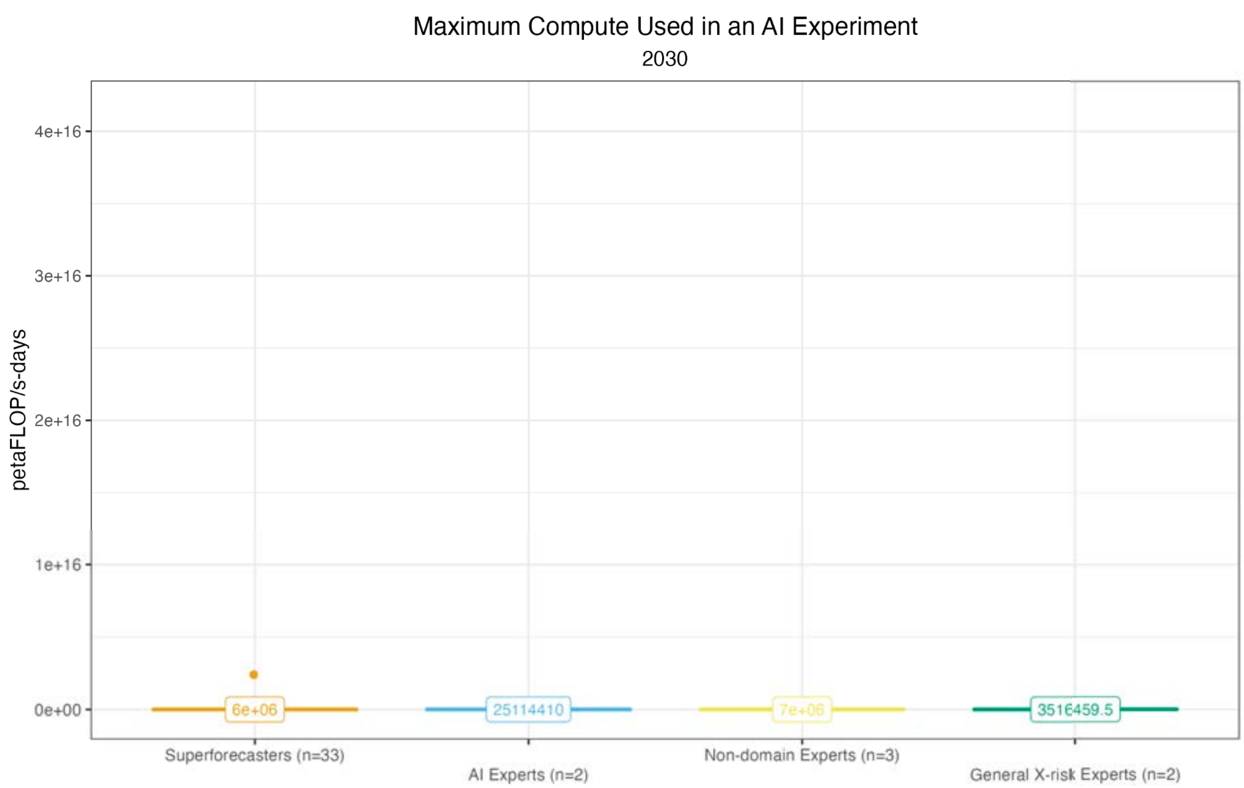
General X-Risk Experts (N = 2)	2030	40,466,412	3,516,460	55106828	-97.39%
	2050	1,990,597,252,768	5,751,539,764	2814955966661	-99.74%
Non-Domain Experts (N = 3)	2024	100,000	150,000	NA	NA
	2030	36,000,000	7,000,000	NA	NA
	2050	1,000,000,000	1,000,000,000	NA	NA

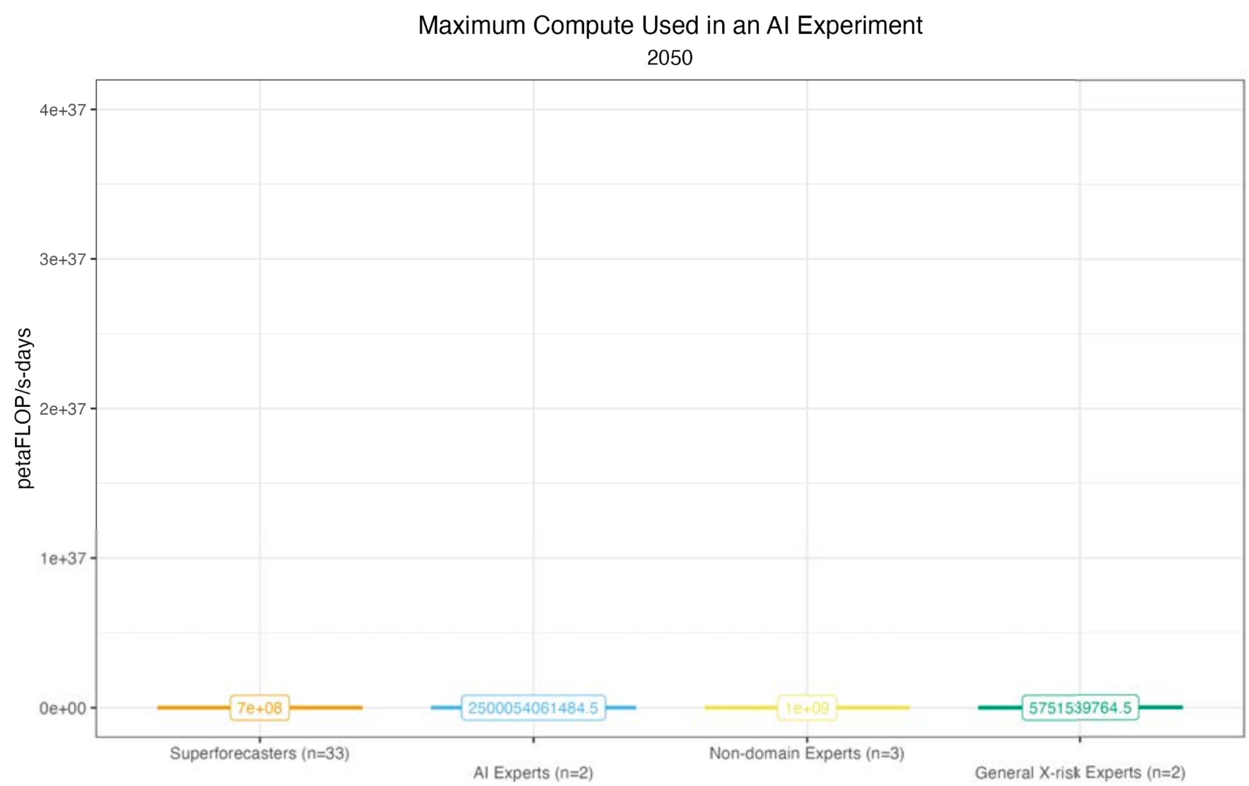



Maximum Compute Used in an AI Experiment - 2030 - 50th %
12000000
8000000
4000000
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 33)


Maximum Compute Used in an AI Experiment - 2050 - 50th %
600000000000
400000000000
200000000000
0
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 33)


Maximum Compute Used in an AI Experiment
2024
petaFLOP/s-days
6e+10
4e+10
2e+10
0e+00
1e+05
420680
150000
218308
Superforecasters (n=33)
AI Experts (n=2)
Non-domain Experts (n=3)
General X-risk Experts (n=2)


Maximum Compute Used in an AI Experiment
2030
petaFLOP/s-days
4e+16
3e+16
2e+16
1e+16
0e+00
6e+06
25114410
7e+06
3516459.5
Superforecasters (n=33)
AI Experts (n=2)
Non-domain Experts (n=3)
General X-risk Experts (n=2)



Sources of agreement, disagreement and uncertainty

The total number of forecasts for this question was small (51), and several teams commented on difficulties they had making their forecasts due to:

- Lack of expertise[1307]
- Using an unusual unit which is not consistently used in the field[1308]
- Small number of forecasts within teams[1309]
- Variation in estimates of compute used in previous experiments[1310]

Rationales were mostly short and uninformative.

[1307] 345, "A very difficult question and very difficult to estimate the maximum compute (measured in petaFLOPS-days) used for training in an AI experiment in the coming years. There was a very high level of uncertainty in the team as well as it seems that the lack of expertise was a factor in the wide variance in estimates."
[1308] 336, "Very large numbers in a unit that was unintuitive to forecasters and not consistently used in literature".
[1309] 344, "Only 4 forecasters from our team participated in this question, so our team's forecasts are more sensitive to individual opinions than those of larger teams."
[1310] 338, "Estimates of compute used for the same project vary substantially between sources, which adds uncertainty."

Broadly, extrapolating current growth rates leads to above median forecasts, and median and below median forecasts assume that current growth rates will slow. The core uncertainty for this question is therefore how current growth rates should be expected to change.[1311]

Also note from the rationale from team 338: “Some predictors, instead of modeling compute spending and compute price separately, directly projected compute doubling times into the future. This implicitly assumes the fast spending increase from the recent past will continue. However, such an increase would become increasingly unsustainable over time.”

Existential risk was raised as another source of uncertainty by one team.[1312]

Arguments given for forecasts of ≤65,000 (2024), ≤950,000 (2030), ≤110m (2050)

- Doubling times in recent years have been slower than in the years preceding. This slower rate may persist or slow down further.[1313]
- Hardware limitations will slow the doubling time of compute used in training.[1314]
- We may reach diminishing marginal returns on increased amounts of compute.[1315]
- Alternative ways of improving performance may be more important, such as algorithmic improvement or advances in chip architecture.[1316]
- Growth rates for total compute are currently slower than those for training compute. Either the growth rate for total compute must increase, or the growth rate for training

[1311] 338, “Estimates of how much spending on compute will scale up differ greatly between predictors, as discussed under question 46. Assuming spending will be a modest multiple of current models yields much lower estimates than assuming, as Cotra's biological anchors report does, that spending will scale up to limits set by technology company budgets and a small percentage of the world economy.” 344, “The growth rate of computation power that could possibly be used for an AI experiment. How sustainable is the current growth rate? What is a reasonable future growth rate?” 336, “Can growth continue at the same pace that it has or will it slow?”

[1312] 344, “Due to expectations of exponential growth in compute, forecasts of catastrophe or human extinction early this century can greatly influence the values of these forecasts. Analyzing this is made even more complicated for those who consider AI the largest existential risk this century, given the interaction between AI-risk and compute used in AI experiments.”

[1313] 336, “There was a slowdown in scaling of computational resources for training between 2018-2021 as compared to 2015 and 2018, or 2012 and 2017. The newer rate might persist or it might slow more.”

[1314] 339, “The main argument for our forecasts is that hardware limitations will significantly constrain doubling time in the future, which makes extrapolating from any prior base rate hazardous.” See also 337, “Eventual end of Moore's law: we are currently reaching 2 nanometer gap transistors and the diameter of a silicon atom is about 0.2 to 0.4 nanometers. However, there might be other ways to pack more transistors. Thus it is not clear when will this trend end or how it will slow down.”

[1315] 336, “Rising costs, diminishing levels of return, more and more hardware needed for incremental growth in quality of results”. See also 343, “Diminishing returns from exponentially greater amounts of compute”.

[1316] 336, “There are other ways to improve the performance, such as improved algorithms and new chip architecture”. See also 343, “Novel approaches that de-emphasize compute”.

compute must slow, otherwise training compute will eventually exceed total compute, which is impossible.[1317]
- Rising concerns about AI risk might cause legislative or other pushback.[1318]

Arguments given for forecasts of ≥140,000 (2024), ≥8.625m (2030), ≥950m (2050)

- Quantum computing may radically increase the computational power available for training.[1319]
    - One team also cited optical neural networks as a technology which might create a discontinuity in the cost of compute.[1320]

Cross-references with other questions

Q46: Largest AI Experiment Cost of Compute

Question 46: Largest AI Experiment Cost of Compute

How much will be spent on compute in the largest AI experiment…
…by the end of 2024?
…by the end of 2030?
…by the end of 2050?

Question and resolution details, prior forecasts, and other relevant sources

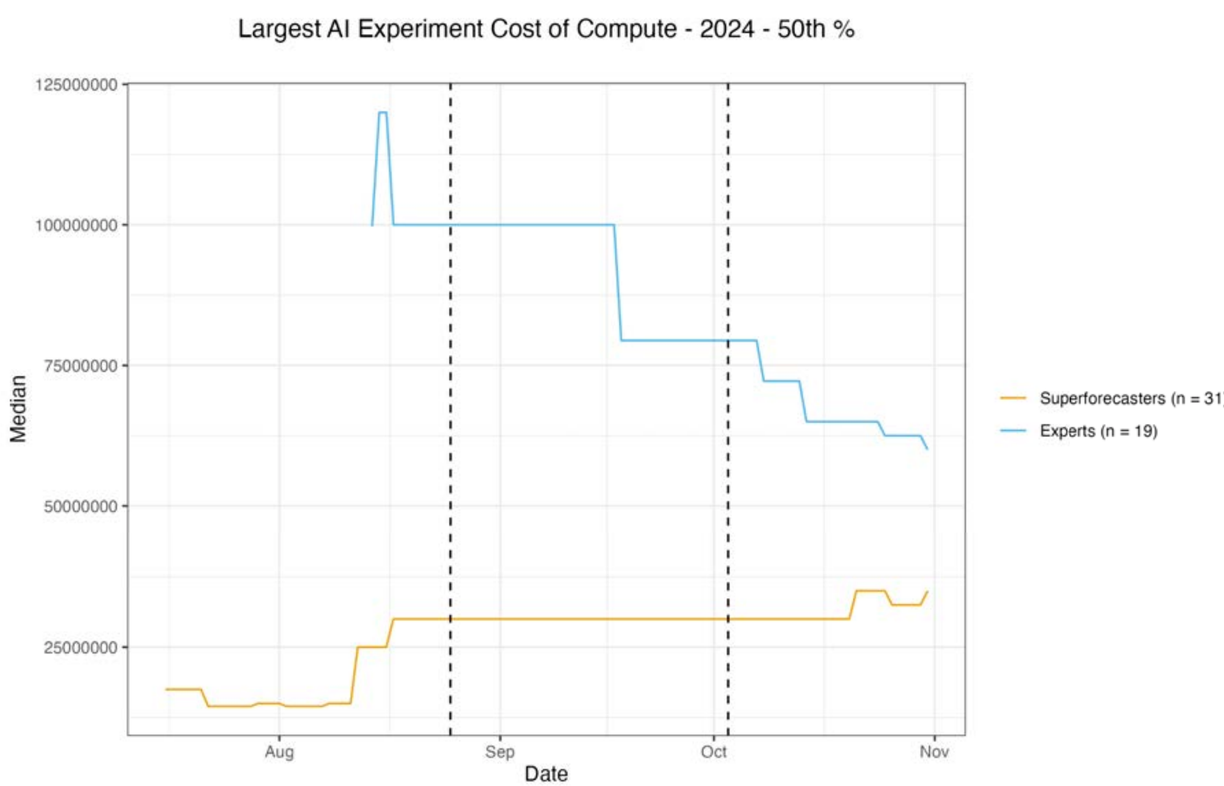
[1317] 336, "Training compute grows faster than total compute grows and will overtake it if the growth rates for each remain the same. That's impossible, training compute trends will have to slow down or total compute grow faster." Given as an argument for team median forecasts (which were close to the overall median) and for lower forecasts of ≤65,000 (2024), ≤950,000 (2030), ≤110m (2050).
[1318] 336, "Possible legislation or pushback against AI if fears mount about possible threat to humanity."
[1319] 336, "Advances in quantum computing might radically shift the computational power available for training."
[1320] 340, "Perhaps the strongest argument for why the trend of Sevilla et al. could be expected to continue to 2030 and beyond is some discontinuity in the cost of AI training compute precipitated by a novel technology such as optical neural networks." Not explicitly given as an argument for forecasts of ≥140,000 (2024), ≥8.625m (2030), ≥950m (2050).

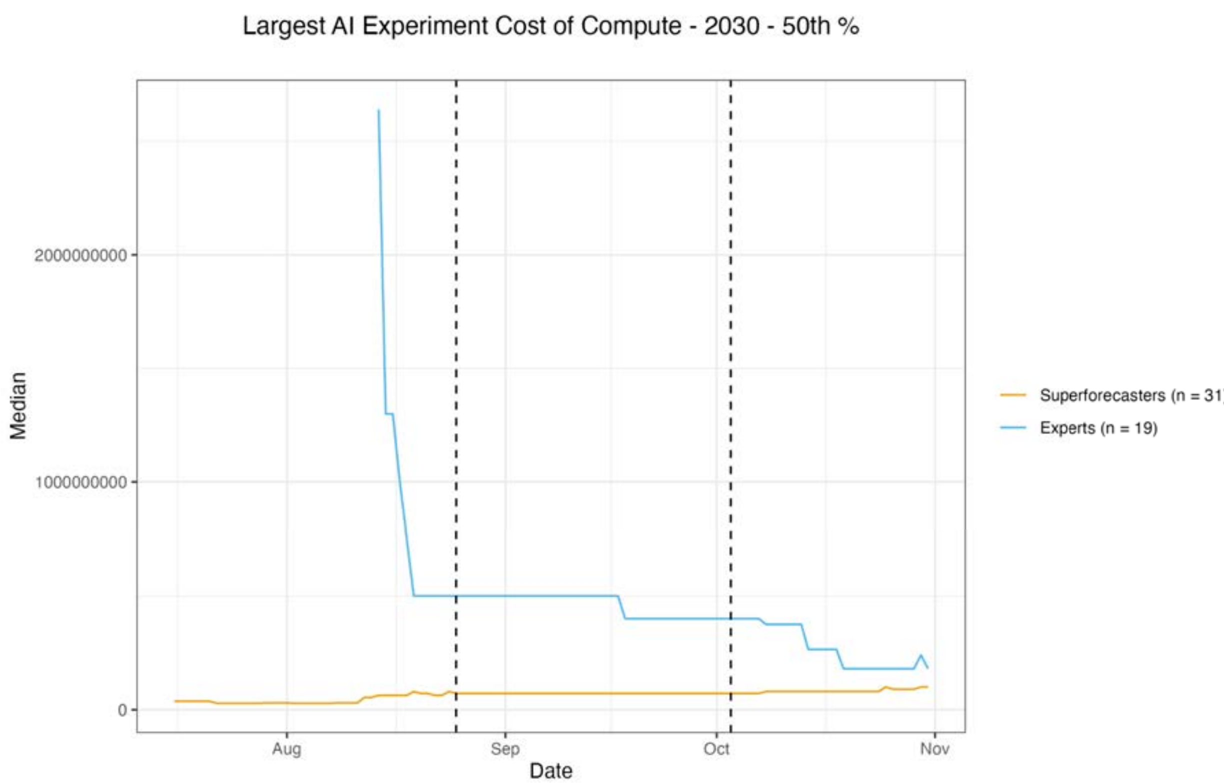
Results[1321]

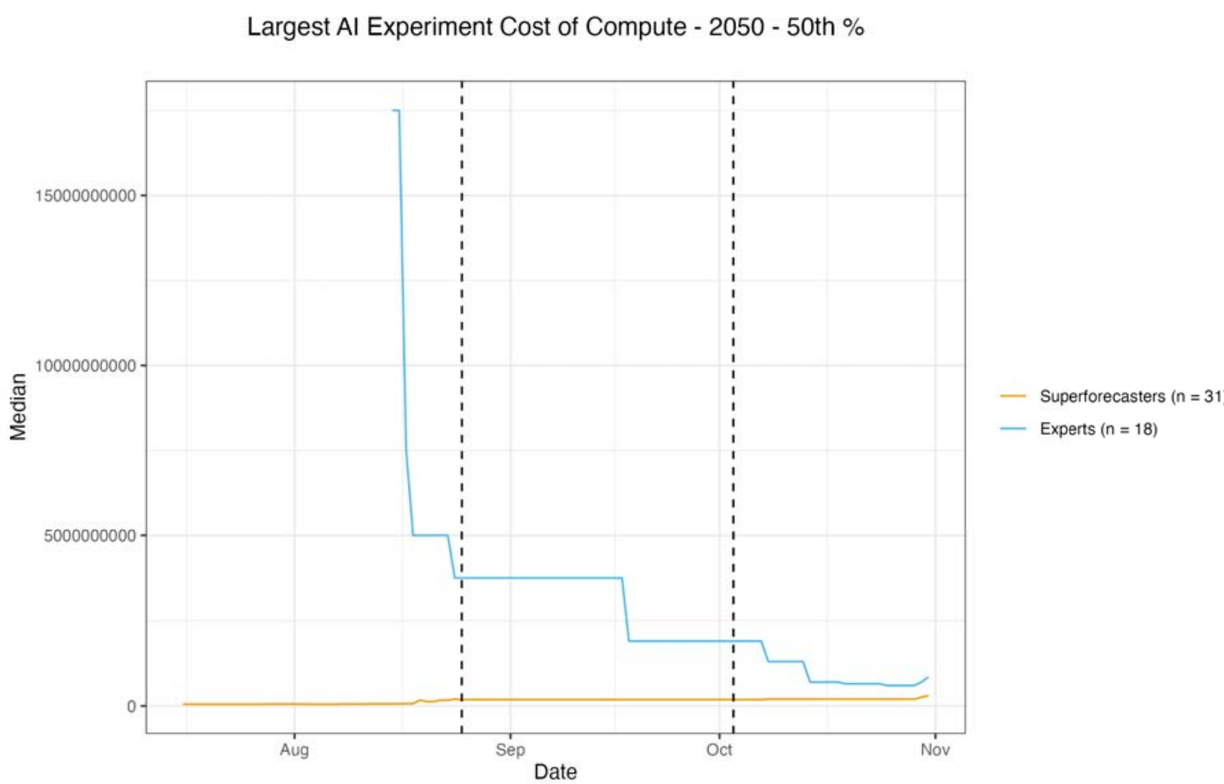
Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 31)	2024	22,500,000	35,000,000	88536190138	-99.91%
	2030	61,500,000	100,000,000	2466576552830527	-100%
	2050	80,000,000	3e8	350000000000000081255501746260672512	-100%
Domain Experts (N = 5)	2024	120,000,000	65,000,000	546015867	-83.31%
	2030	10,000,000,000	180,000,000	20816659987	-28.38%
	2050	512,500,000,000	8e8	689429111657	-40.87%
General X-Risk Experts (N = 7)	2024	8,000,000	30,000,000	42134494	-41.34%
	2030	25,000,000	180,000,000	2288381481	-82.45%
	2050	50000000	1e9	364261742262	-98.85%
Non-Domain Experts (N = 5)	2024	2.5e25	55,000,000	35355339059327369419423744	-100%
	2030	5e27	300,000,000	707106781186547453671977	-100%

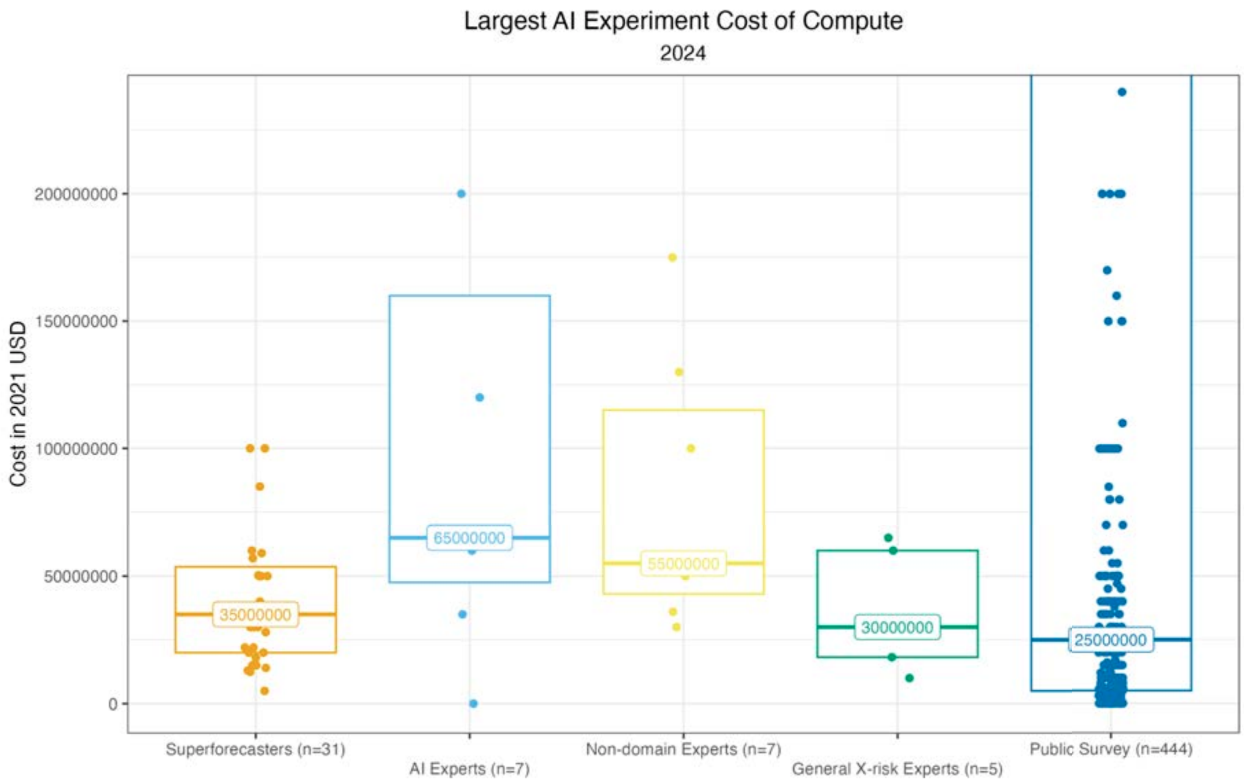
[1321] Numbers of forecasters are given as of Stage 4 of the XPT.

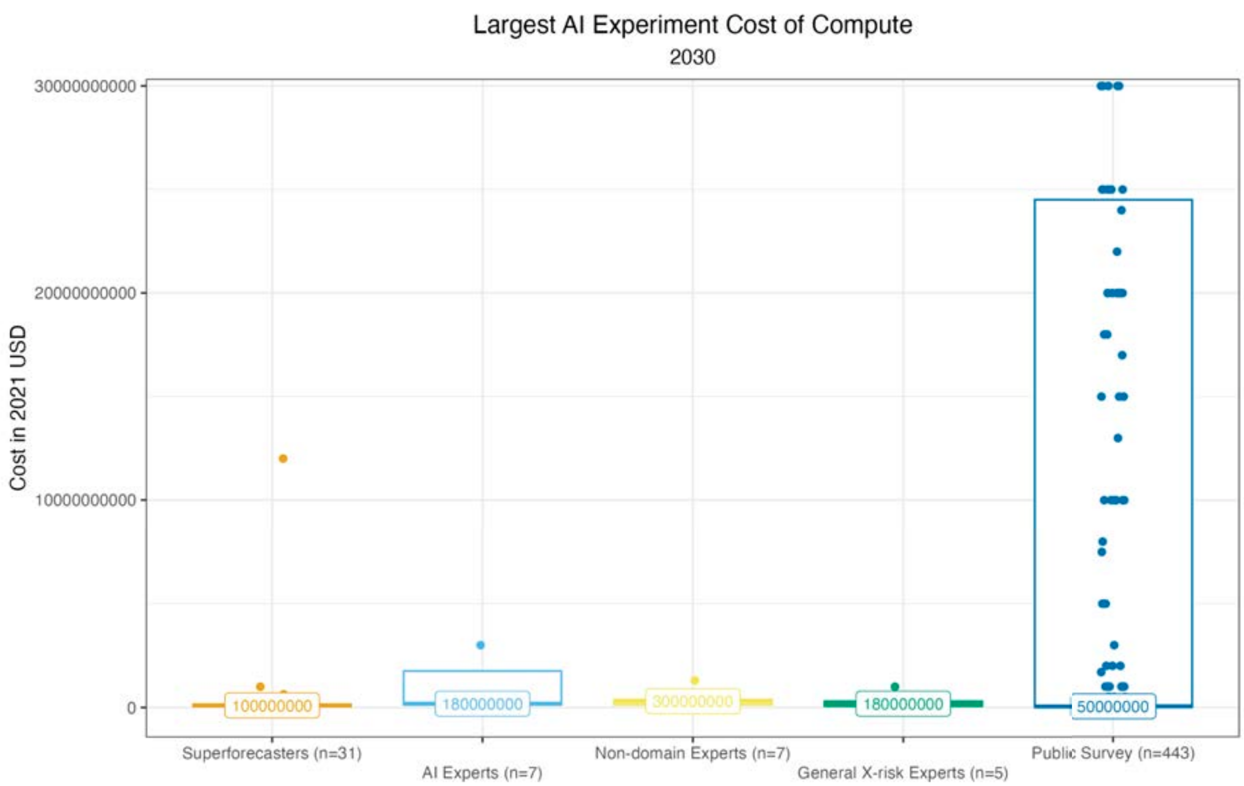
				7792	
	2050	5e34	8.5e8	70710678118654748684380378455080960	-100%
Public Survey (N = 444)	2024	25,000,000		1.90E+25	-
	2030	50,000,000		2.01E+29	-
	2050	180,000,000		1.43E+69	-

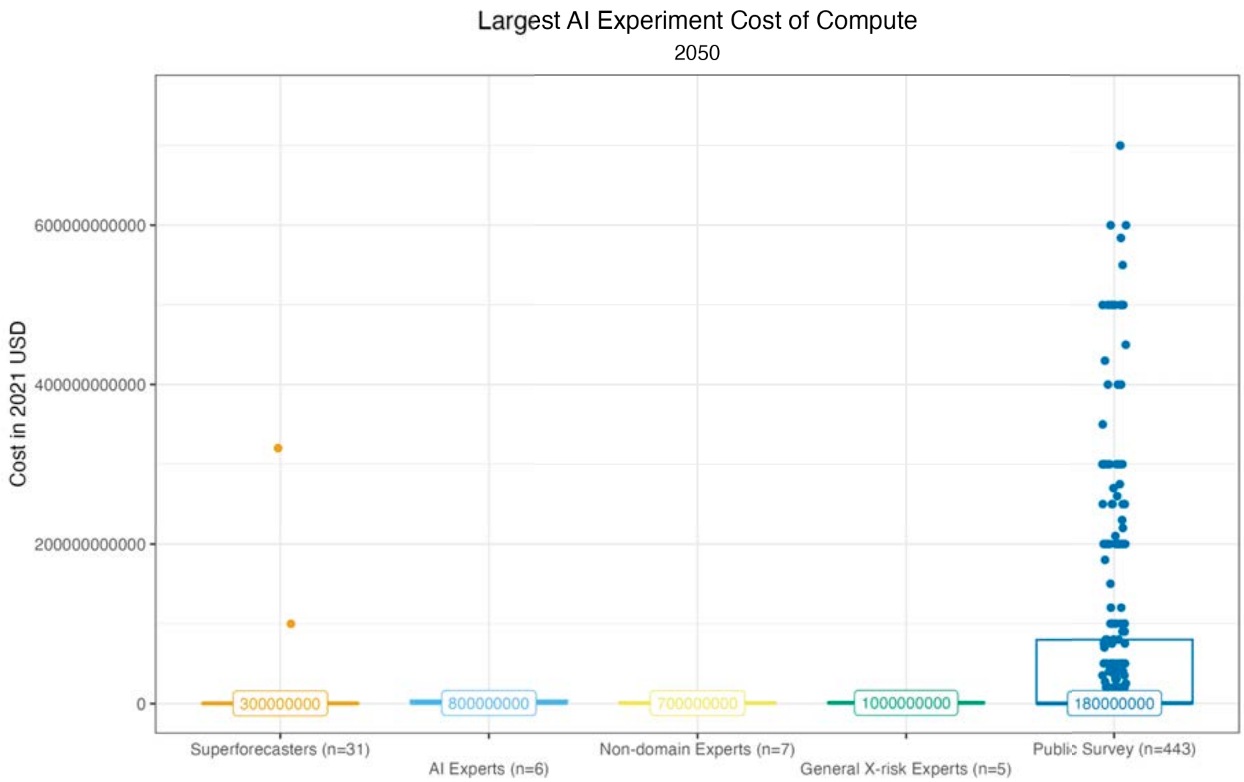



Largest AI Experiment Cost of Compute - 2030 - 50th %
Median
2000000000
1000000000
0
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 31)
Experts (n = 19)


Largest AI Experiment Cost of Compute - 2050 - 50th %
Median
15000000000
10000000000
5000000000
0
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 31)
Experts (n = 18)


Largest AI Experiment Cost of Compute
2024
Cost in 2021 USD
200000000
150000000
100000000
50000000
0
35000000
65000000
55000000
30000000
25000000
Superforecasters (n=31)
AI Experts (n=7)
Non-domain Experts (n=7)
General X-risk Experts (n=5)
Public Survey (n=444)


Largest AI Experiment Cost of Compute
2030
Cost in 2021 USD
30000000000
20000000000
10000000000
0
100000000
180000000
300000000
180000000
50000000
Superforecasters (n=31)
AI Experts (n=7)
Non-domain Experts (n=7)
General X-risk Experts (n=5)
Public Survey (n=443)



Sources of agreement, disagreement and uncertainty

Sources of agreement

- Several teams noted high uncertainty in their rationales.[1322]
- The base rate is hard to establish.[1323]

Sources of disagreement

- How to interpret base rates
  - 338: "The main split between predictions is between lower estimates (including the team median) that anchor on present project costs with a modest multiplier,

[1322] 337, "[T]he uncertainties are large, with estimations differing by one or even two orders of magnitudes between the 5-th and 95-th percentile." 336, "By 2050 the team considers anything from "AI experiments will only cost a penny" to "billions of dollars" thinkable." 340, "We have relatively high uncertainty over how to analyze this. There's even a moderate amount of uncertainty about what the current record is." 344, "This question, together with 40 and 46-47, hits a hard limit in how far probabilistic forecasting by generalists can go. Beyond wild guessing, the team had no firm answer and mostly has given up. Our knowledge was insufficient to start tackling the question, so all we can do is highlight a few forecaster comments without endorsing them and in an unstructured way. We should warn that our knowledge as a team was insufficient to assess whether those comments are useful or correct"

[1323] 340: "There's even a moderate amount of uncertainty about what the current record is." 339, "The base rate data has some conflicting reports. One source balances increasing size of models (more expensive) with decreasing costs of hardware/implementation (less expensive) and concludes an order of magnitude per year (or so). The other suggests that a different architecture--that produces more advanced and startling results--can train at half the price." 338, "The variation in these numbers highlights that estimating costs isn't straightforward."

and higher estimates that follow Cotra in predicting pretty fast scaling will continue up to anchors set by demonstrated value-added, tech company budgets, and megaproject percentages of GDP."
    - 340: “Presumably much of these disagreement[s] stem from different ways of looking at recent AI progress. Some see the growth of computing power as range bound by current manufacturing processes and others expect dramatic changes in the very basis of how processors function leading to continued price decreases.”
- Economic growth rates and tech industry growth rates after 2030[1324]
- How profitable larger models will be[1325]

Arguments given for lower forecasts (2024: <$40m, 2030: <$110m, 2050: ⩽$200m)

- Training costs have been stable around $10m for the last few years.[1326]
- Current trend increases are not sustainable for many more years.[1327] One team cited this AI Impacts blog post.
- Major companies are cutting costs.[1328]
- Increases in model size and complexity will be offset by a combination of falling compute costs, pre-training, and algorithmic improvements.[1329]

[1324] 340, “There's substantial disagreement about economic growth rates and tech industry growth rates after 2030. That's likely to make a big difference in how much money is available to be spent.”
[1325] 340, “I suspect more of the disagreement comes from wildly differing beliefs about how much profit there will be from larger models. Imaginations seem to range from thinking bigger AIs can't do much more than current AIs, to Cotra's comparison to the Apollo Project's 3.6% of GDP.”
[1326] 337, “[T]raining cost seems to have been stuck in the $10M figure for the last few years.”; “we have not seen such a large increase in the estimated training cost of the largest AI model during the last few years: AlphaZero and PALM are on the same ballpark.” 341, “For 2024, the costs seem to have flattened out and will be similar to now. To be on trend in 2021, the largest experiment would need to be at $0.2-1.5bn. GPT-3 was only $4.6mn”
[1327] 341, “The AI impacts note also states that the trend would only be sustainable for a few more years. 5-6 years from 2018, i.e. 2023-24, we would be at $200bn, where we are already past the total budgets for even the biggest companies.”
[1328] 336, “The days of 'easy money' may be over. There's some serious belt-tightening going on in the industry (Meta, Google) that could have a negative impact on money spent.”
[1329] 337, “It also puts more weight on the reduced cost of compute and maybe even in the improved efficiency of minimization algorithms, see question 48 for instance.” 336, “After 2030, we expect increased size and complexity to be offset by falling cost of compute, better pre-trained models and better algorithms. This will lead to a plateau and possible even a reduction in costs.”; “In the near term, falling cost of compute, pre-trained models, and better algorithms will reduce the expense of training a large language model (which is the architecture which will likely see the most attention and investment in the short term).” See also 343, “$/FLOPs is likely to be driven down by new technologies and better chips. Better algorithm design may also improve project performance without requiring as much spend on raw compute.” See also 339, “The low end scenarios could happen if we were to discover more efficient training methods (eg take a trained model from today and somehow augment it incrementally each year rather than a single batch retrain or perhaps some new research paradigm which makes training much cheaper).”

- Large language models will probably see most attention in the near future, and these are bottlenecked by availability of data, which will lead to smaller models and less compute.[1330]
- Not all experiments will be public, and it is possible that the most expensive experiments will not be public.[1331]

Arguments given for higher forecasts (roughly 2024: ⩾$50m, 2030: ⩾$300m, 2050: >$500m)

- As AI creates more value, more money will be spent on development.[1332]
- A mega-project could be launched nationally or internationally which leads to this level of spending.[1333]
  - Great power competition between the US and China could lead to massive investments in AI.[1334]
- Anchors proposed for this level of spending:
  - Covid vaccine development cost: $2bn[1335]
  - ITER for fusion[1336]
  - The Manhattan project: 1% of US annual GDP[1337]

Other arguments given

Arguments given for forecasts of 2024: >$20m, 2030: >$37m, 2050: >$37.5m:

[1330] 336, "Additionally, large language models are currently bottlenecked by available data. Recent results from DeepMind suggest that models over ~100 billion parameters would not have enough data to optimally train. This will lead to smaller models and less compute used in the near term. For example, GPT-4 will likely not be significantly larger than Chinchilla. https://arxiv.org/abs/2203.15556". 341, "The data availability is limited." See also 340, "The evidence from Chinchilla says that researchers overestimated the value of adding parameters (see https://www.lesswrong.com/posts/6Fpvch8RR29qLEWNH/chinchilla-s-wild-implications). That is probably discouraging researchers from adding more parameters for a while. Combined with the difficulty of getting bigger text datasets, that might mean text-oriented systems are hitting a wall. (I'm unsure why this lasts long - I think other datasets such as video are able to expand more)."

[1331] 336, "Not all experiments will be made public. It is likely that in other countries (China!) there have been or will be more expensive ones."

[1332] 343, "Monetization of AGI is in its early stages. As AI creates new value, it's likely that additional money will be spent on increasingly more complex projects." Note that this argument refers to forecasts higher than the team median forecasts, and the team median for 2024 was $25m.

[1333] 337, "This will make very much sense in the event that a great public project or international collaboration will be assembled for researching a particular aspect of AI (a bit in the line of project Manhattan for the atomic bomb, the LHC for collider physics or ITER for fusion). The probability of such a collaboration eventually appearing is not small. Other scenario is great power competition between China and the US, with a focus on AI capabilities."

[1334] See footnote 1335

[1335] 338, "$2B COVID vaccine development cost, with 30 years annual 3% increase". Note that this argument refers to forecasts higher than the team median forecasts, and the team median for 2030 was $110m.

[1336] See footnote 1335

[1337] See footnote 1335. Other rationales cited this AI impacts post for this figure.

- There is strong competition between actors with lots of resources and incentives to develop AI.[1338]
- Technological breakthroughs might require much more compute.[1339]

Arguments which push for lower forecasts:
- Growth may already be slowing down.[1340]
- In the future, AI systems may be more modular, such that single experiments remain small even if total spending on compute increases drastically.[1341]
- Recent spending on compute may have been status driven.[1342]
- There seems to be general agreement that experiments of more than a few months are unwise, which might place an upper bound on how much compute can cost for a single experiment.[1343]

Arguments which push for higher forecasts:
- The impact of AI on AI development or the economy at large might raise the spending ceiling arbitrarily high.[1344]

Cross-references with other questions

Q45: Maximum Compute Used in an AI Experiment
Q47: Lowest Price of of GFLOPS
Q48: ImageNet Classification Training Efficiency

[1338] 336, "There is strong competition between players with deep pockets and strong incentives to develop and commercialize 'AI-solutions'."
[1339] 336, "There might be technological advances and/or new ideas, that require much more calculating power, especially if it can be monetized."
[1340] 340, "The growth might be slowing down now."; "Or maybe companies were foolishly spending too little a few years ago, but are now reaching diminishing returns, with the result that declining hardware costs mostly offset the desire for bigger models."
[1341] 340, "Later on, growth might slow a lot due to a shift to modular systems. I.e. total spending on AI training might increase a good deal. Each single experiment could stay small, producing parts that are coordinated to produce increasingly powerful results." See also 339, "2050 At this point I'm not sure it will be coherent to talk about a single AI experiment, models will probably be long lived things which are improved incrementally rather than in a single massive go. But they'll also be responsible for a large fraction of the global GDP so large expenditures will make sense, either at the state level or corporation."
[1342] 340, "Some forecasters don't expect much profit from increased spending on AI training. Maybe the recent spending spree was just researchers showing off, and companies are about to come to their senses and stop spending so much money."
[1343] 340; "There may [be] some limits resulting from training time. There seems to be agreement that it's unwise to attempt experiments that take more than a few months. Maybe that translates into a limit on overall spending on a single experiment, due to limits on how much can be done in parallel, or datacenter size, or supercomputer size?"
[1344] 344, "Automatic experiments run by AI are beyond valuation". 337, "One forecast suggest[s] astronomical numbers for the largest project in the future, where the basis of this particular forecast is the possibility of an AI-driven economic explosion (allowing for the allocation of arbitrarily large resources in AI)."

Question 47: Lowest Price of GFLOPS

What will be the lowest price, in 2021 US dollars, of 1 GFLOPS with a widely-used processor…
…by the end of 2024?
…by the end of 2030?
…by the end of 2050?

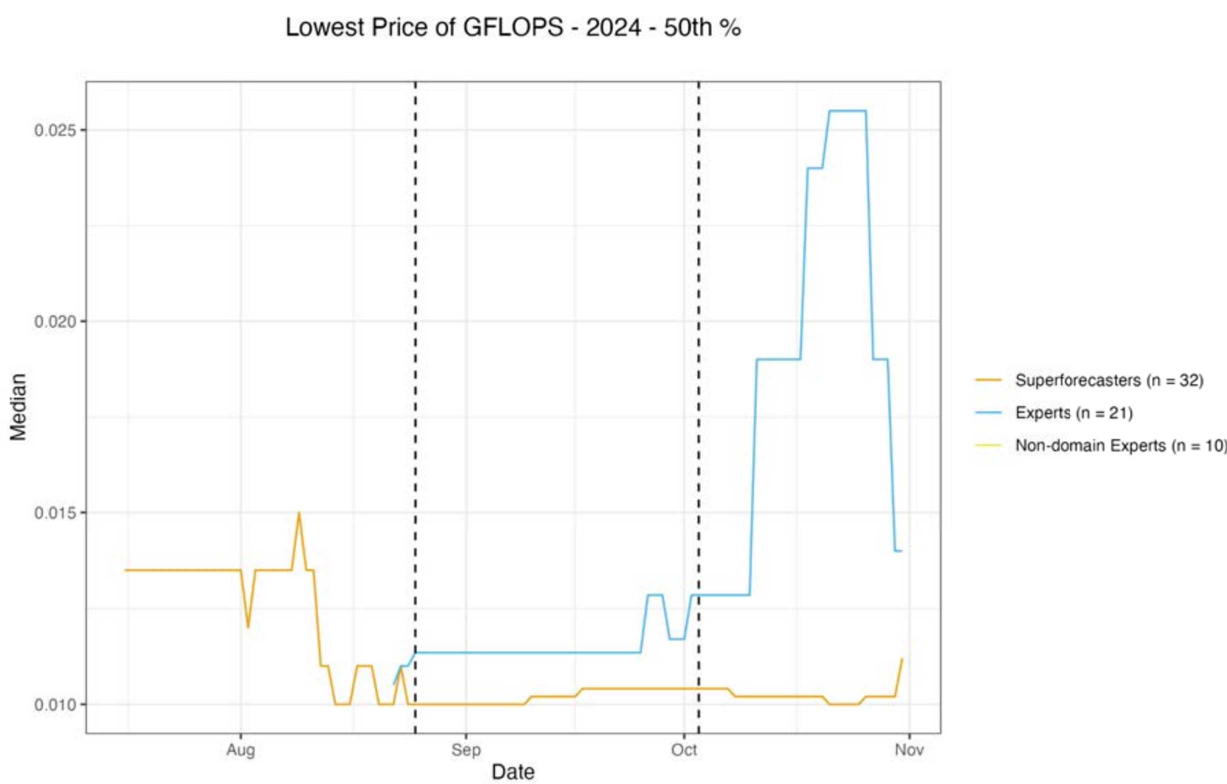
Question and resolution details, prior forecasts, and other relevant sources

Results[1345]

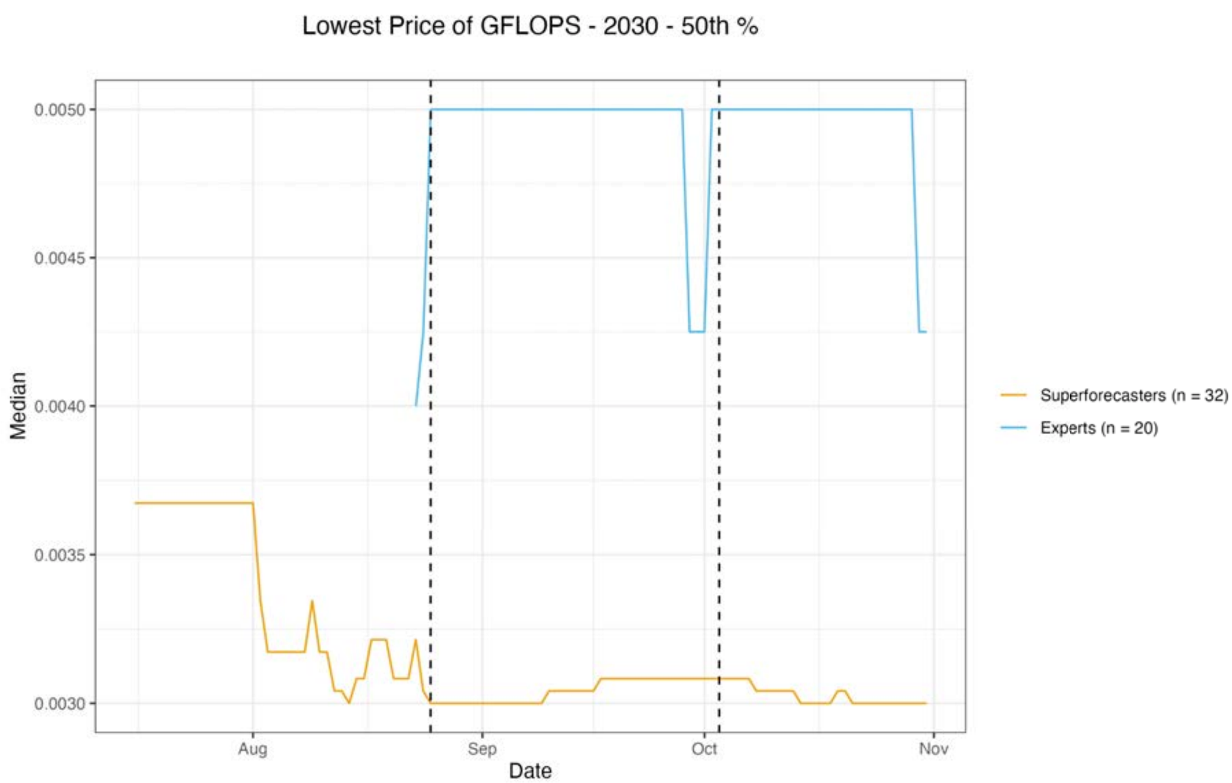
Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 32)	2024	0.01	0.011	0.015	-29.16%
	2030	0.0033	0.003	0.009	-7.43%
	2050	0.00018	0.00013	0.0057	-32.64%
Domain Experts (N = 6)	2024	0.012	0.011	0.11	-30.48%
	2030	0.003	0.0029	0.027	-28.9%
	2050	0.000026	0.00012	0.017	-23.41%
General X-Risk Experts (N = 5)	2024	0	0.024	NA[1346]	NA
	2030	0	0.0035	NA	NA

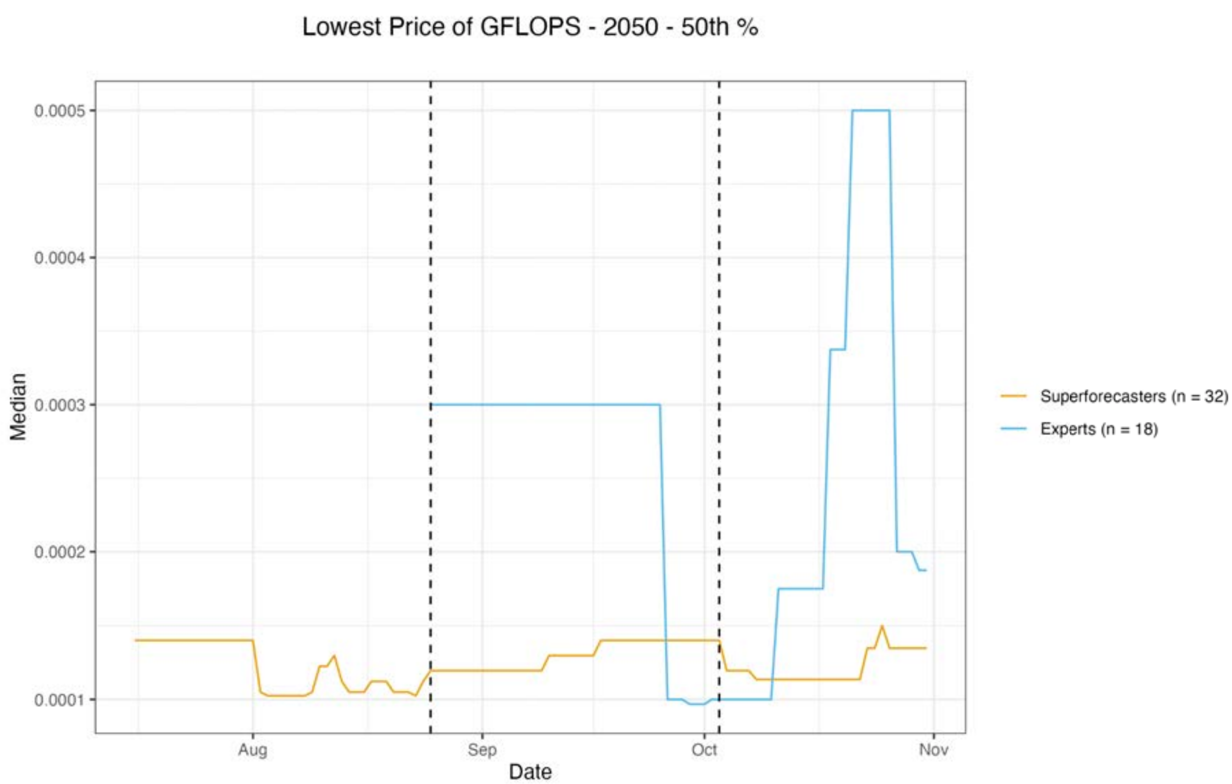
1345 Numbers of forecasters are given as of Stage 4 of the XPT. Public survey respondents were surveyed outside of the tournament context.
1346 Only one forecaster in this group answered this question, and only for the years 2024 and 2030, in Stage 1.

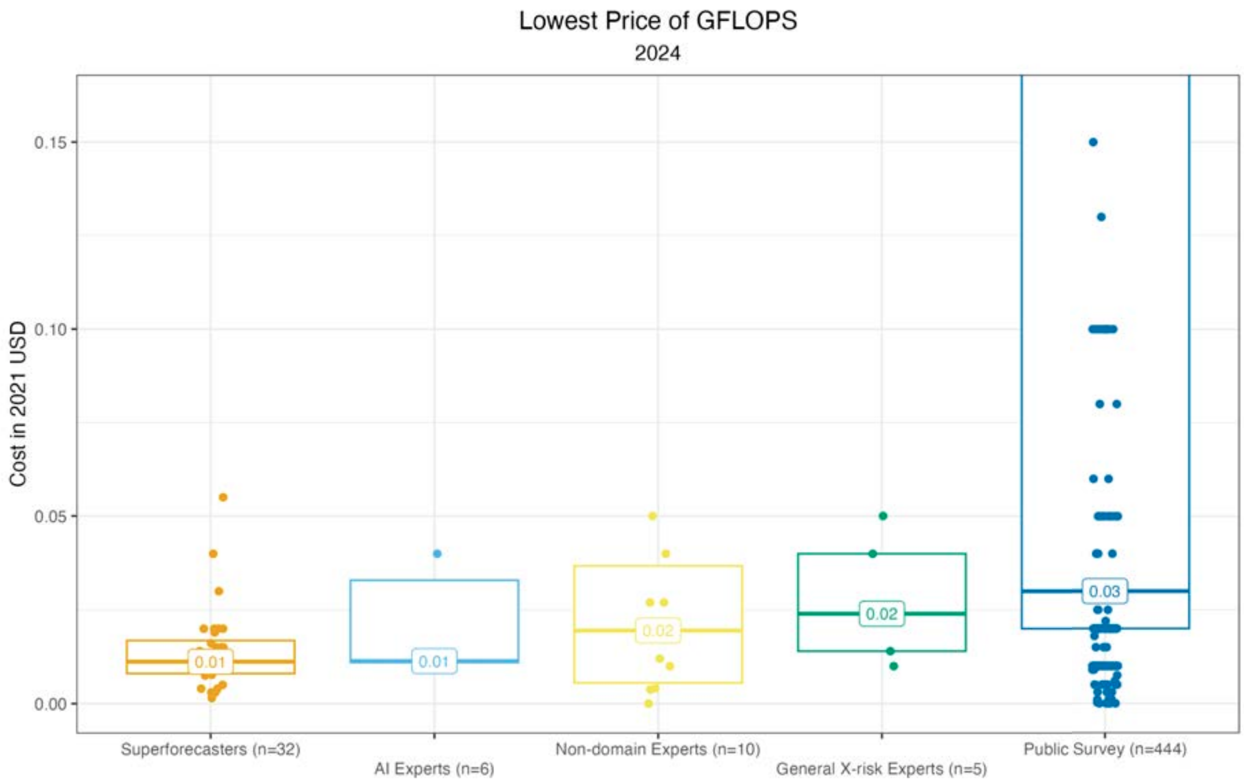
	2050	NA	0.00014	NA	NA
Non-Domain Experts (N = 10)	2024	0.014	0.02	0.019	+291.85%
	2030	0.015	0.005	NA	NA
	2050	0.01	0.00065	NA[1347]	NA
Public Survey (N = 444)	2024	0.03		1.89846E+11	-
	2030	0.02		4.74592E+11	-
	2050	0.01		1.42388E+12	-

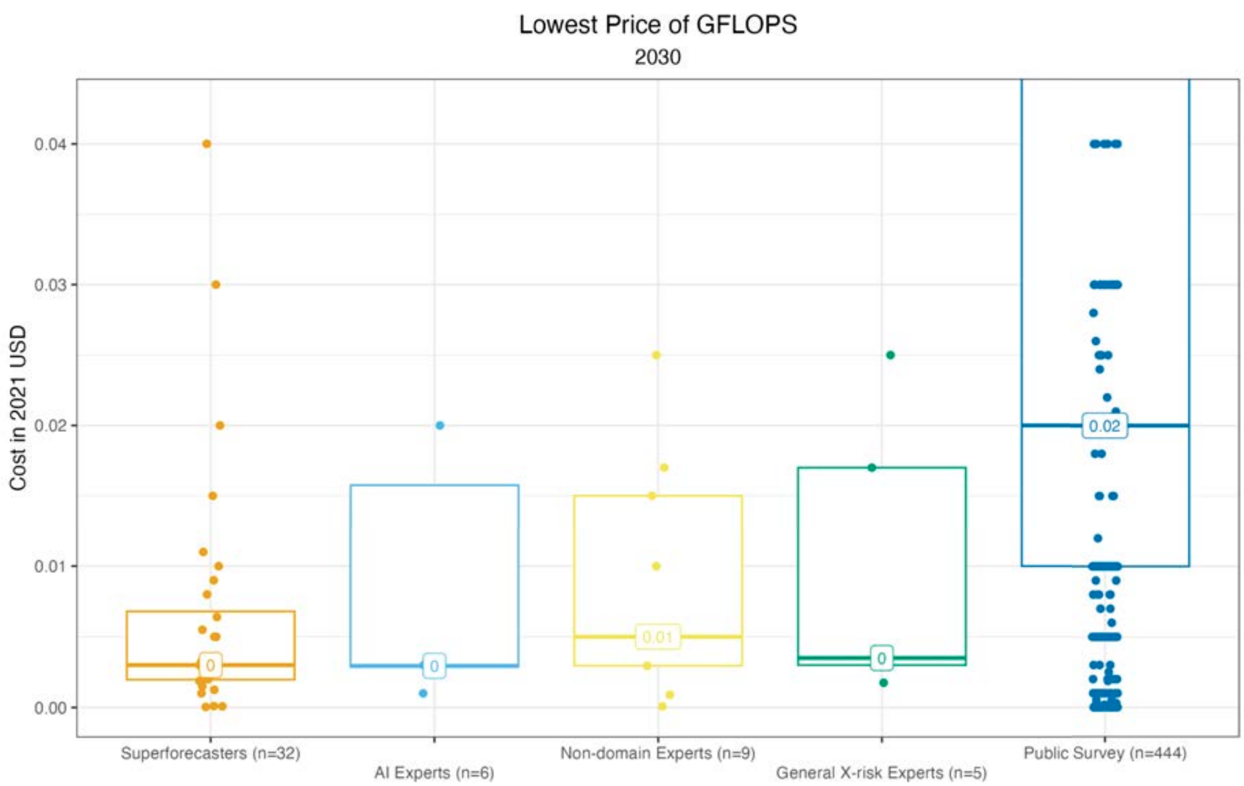


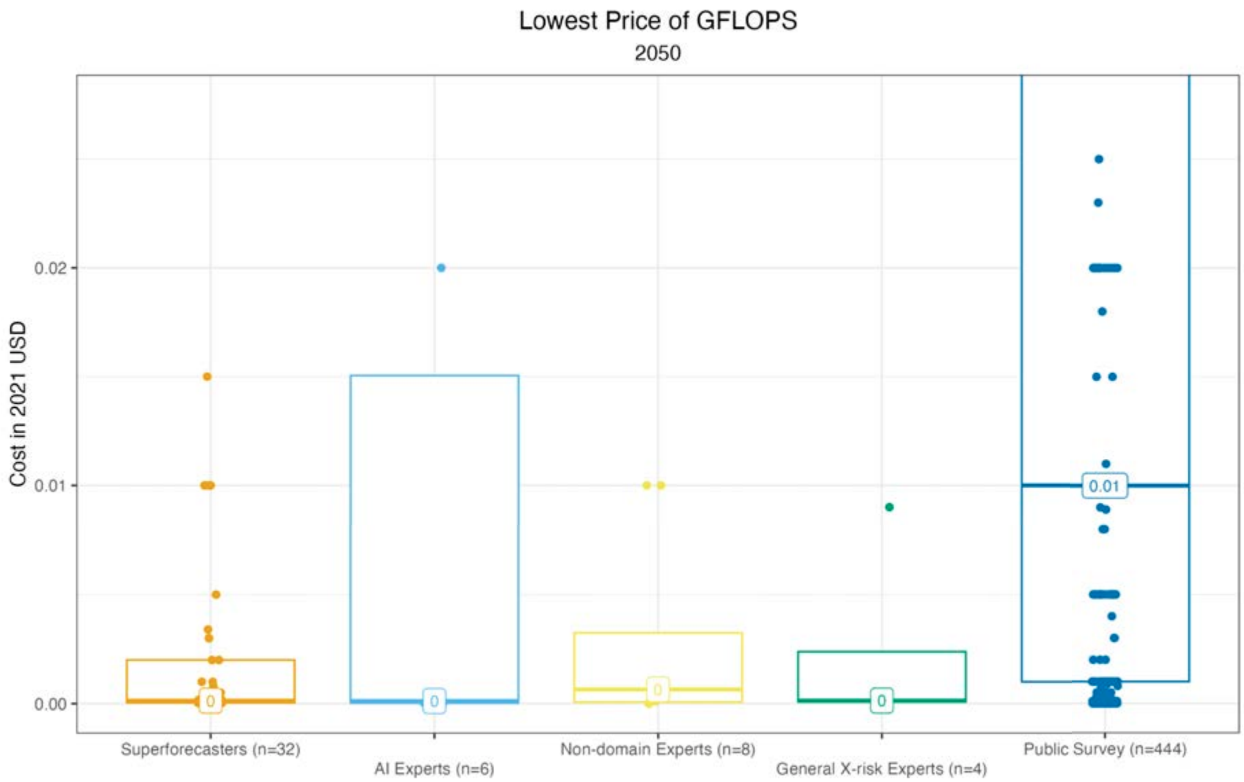
[1347] Only one forecaster in this group answered this question for the year 2050 in Stage 1.


Lowest Price of GFLOPS - 2030 - 50th %
Median
0.0050
0.0045
0.0040
0.0035
0.0030
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 32)
Experts (n = 20)


Lowest Price of GFLOPS - 2050 - 50th %
Median
0.0005
0.0004
0.0003
0.0002
0.0001
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 32)
Experts (n = 18)


Lowest Price of GFLOPS
2024
Cost in 2021 USD
0.15
0.10
0.05
0.00
0.01
0.01
0.02
0.02
0.03
Superforecasters (n=32)
AI Experts (n=6)
Non-domain Experts (n=10)
General X-risk Experts (n=5)
Public Survey (n=444)


Lowest Price of GFLOPS
2030
Cost in 2021 USD
0.04
0.03
0.02
0.01
0.00
0
0
0.01
0
0.02
Superforecasters (n=32)
AI Experts (n=6)
Non-domain Experts (n=9)
General X-risk Experts (n=5)
Public Survey (n=444)



Sources of agreement, disagreement and uncertainty

Sources of uncertainty:

- Base rates: there was uncertainty about current prices for GFLOPS,[1348] and according to team rationales several teams used faulty data to form their base rates.[1349]

[1348] 341, “There was some amount of uncertainty regarding the starting point for the forecast regarding current prices for gigaflops. In addition to the $0.3 in 2017 estimate, a metaculus question has some later estimates that could also be relevant: 2019 $68.82 per teraflops and 2020 $29 per teraflops.” 342, “It is obviously very difficult to determine the baseline rate for estimating the lowest price, in 2021 US dollars, of 1 GFLOPS. This difficulty is also apparent in the low number of forecasts that this question received, which in turn hindered the chances of a strong exchange of views and contrarian debate on this issue.”

[1349] 337, “Given that five out of eight team forecasters used faulty data, we should conclude that the team forecast is also faulty for all dates and percentiles”, “many forecasters only used the outdated Wikipedia article referenced in the question description. That article was specifically the price/performance data for the more recent models of GPUs. (The article was updated recently, though it still doesn't cover the dedicated AI infrastructure hardware sold by Nvidia like their new H100 line.) This led to most forecasters using obsolete data for their baselines and predicting future GFLOPS prices that are worse than the already achieved results. The difference in the source data quality fully explains the widely divergent forecasts for 2024, which should normally be simple - and numerically similar - extrapolations of the status quo.” 344, “This question has a shallow pool of forecasters with limited arguments given for the estimates and erroneous inputs.”

- New technology: this was cited as a potential reason for higher and lower forecasts by different teams.[1350]
- Other sources of uncertainty cited: energy prices,[1351] geopolitics.[1352]

Arguments given for forecasts close to the overall median (2024: 0.0114, 2030: 0.0030, 2050: 0.00011)

- Historical trends show an order of magnitude improvement in price-performance every decade.[1353]
- Covid inflated costs for electricity and hardware but efficiencies in development and falling energy prices will drive costs down again.[1354]
- Two teams cited the Epoch AI report.[1355]

Arguments given for forecasts lower than the median

- The general trend is asymptotic towards zero.[1356]
- Recent price-performance trends have been slower than usual, and there could be a return to the older order of magnitude improvements every 8 or 4 years.[1357]
- Novel technologies might lead to a discontinuous drop in prices.[1358] Possible technologies cited are optical computing, quantum computing, reversible and three-dimensional circuits, and unknown advances.

[1350] 336, “Development of new technology was a major source of uncertainty and could drive prices either up or down.” Teams which cited new technology as an argument for lower forecasts: 336, 340, 341, 343. 336 also cited new technology as an argument for higher forecasts.

[1351] 340, “Changes in the price of energy in either direction may do more to impact the pricing than technological innovation.”

[1352] 336, “Potential for geopolitical developments to affect price was another source of uncertainty.”

[1353] 336, “trend of order of magnitude improvement in price-performance every 10 years”.

[1354] 336, “‘The biggest price is not hardware itself but electricity, data-center usage and human AI-scientists salaries.’ The COVID pandemic inflated costs for electricity and hardware but efficiencies in development, and energy costs, will drive this down again.”

[1355] 338, “2030: Interpretation of Cotra Biologic anchors report suggests trends similar to those identified in the Epoch AI Report. Use those values to extrapolate the median estimate.” 340, “ The team's median forecasts generally deviate less than 50% from what Epoch's data suggests.”

[1356] 336, “general trend is asymptotic toward zero”.

[1357] 336, “recent performance/$ trend is slower than long-run (there could be a return to the longer run trends of OOM every 8 or 4 years.)”

[1358] 336, “uncertainty regarding future technological improvements”; “potential for discovering new modes of computing leading to discontinuous improvements”. 340, “The strongest argument for lower extreme forecasts is that some novel technology precipitates discontinuous progress in the trend of the cost of computation for training AI models. Optical neural networks are a promising technology with the potential to improve AI model training in this way.” See also 341, “Potential prospects for a revolutionary technology (e.g. optical computing, quantum computing, reversible and three-dimensional circuits) as per Cotra's report. This could break the foreseen plateau and lead to continued doubling every 3-4 years past 2040 and go back to a 1-2 year doubling.” See also 343, “Application of advanced AI or AGI to the problem could transformatively decrease prices in an unpredictable way.” See also 344, “Quantum computing seems to be accelerating progress - it's going to get much cheaper much quicker imho”.

- One team cited the theoretical limit for the price of silicon chips given in Cotra.[1359]

Arguments given for forecasts higher than the median

- Since 2010, the rate of price decline has slowed.[1360] One team cited the IEEE report 'More Moore'.
- War, particularly over Taiwan, could raise prices.[1361]
- Global economic decline could slow technological advances.[1362]
- Progress may be getting harder. Research to date may have picked low-hanging fruit, and many experts expect slower improvements over the coming century compared with the previous one.[1363]
- We may reach fundamental physical limits.[1364]
- Future technological developments are uncertain and could raise prices.[1365]
- Demand for more efficient chips may be low.[1366]

Other arguments given

Arguments which push for higher forecasts:

- FLOP rates might stabilise in future and optimisation shift to memory architectures.[1367]

[1359] 344, "For 2050, I will take as a 50 th % the theoretical limit of price for silicon chips" [given in Cotra, p. 32].

[1360] 336, "advancement may have been slowing since 2010 and rate of decline in prices could continue to slow". 341, "Faltering of Moore's Law. See the IEEE's 2021 IRDS report, More Moore, Table MM for challenges." See also 339, "Unstable world and a decline in Moore's law limit the factors that drove down costs in previous years.

[1361] 336, "war, especially over Taiwan, could raise prices and/or slow advancement". See also 339, "Unstable world and a decline in Moore's law limit the factors that drove down costs in previous years. It could take decades for the US to reshore semiconductor manufacturing to the US (and to China). This means Taiwan tensions could throw wrenches into cost dropping."

[1362] 336, "global economic decline could lead to slower advancement".

[1363] 341, "If early technological progress can be seen as a low-hanging fruit, further progress inherently becomes harder. Many experts (as quoted in Cotra, 2020) expect much less improvement over the next century than we have seen in the past century."

[1364] 336, "potential for hard/impossible to surpass fundamental physical limits". 340, "The strongest argument for higher extreme forecasts is that Moores law slows due to physical limitations in manufacturing, GPU cost per compute slows because of limits to parallelization, and there is are no new technologies to pick up the flattening S-curve and continue the trend." 341, "Known limitations of specific technologies. The existence of fundamental physical limits."

[1365] 336, "uncertainty regarding future technological development - potential for new tech to lead to higher prices."

[1366] 341, " Lack of high demand (or diminished urgency) for ever more efficient chips."

[1367] 339, "Processors in the future may not necessarily have greater FLOP rates, which hit limits of Moore's law, but superior memory architecture (e.g. Apple's M1/m2 chips did this by being better suited to scientific computing workloads). Apple's success: access a distributed RAM with almost no latency: Apple M1 destroys Intel and AMD in newly-released benchmarks | TechRadar. FLOP rate may become static at one point, meaning memory optimisations will rule. There may be another metric, such as effective FLOP rate, that might emerge instead."

- Materials for chips are rare and have other uses.[1368]
- A catastrophe or extinction event could halt price decreases.[1369]

Cross-references with other questions

Q46: Largest AI Experiment Cost of Compute

Question 48: ImageNet Classification Training Efficiency

By what factor will training efficiency on ImageNet classification have improved over AlexNet…
…by the end of 2024?
…by the end of 2030?

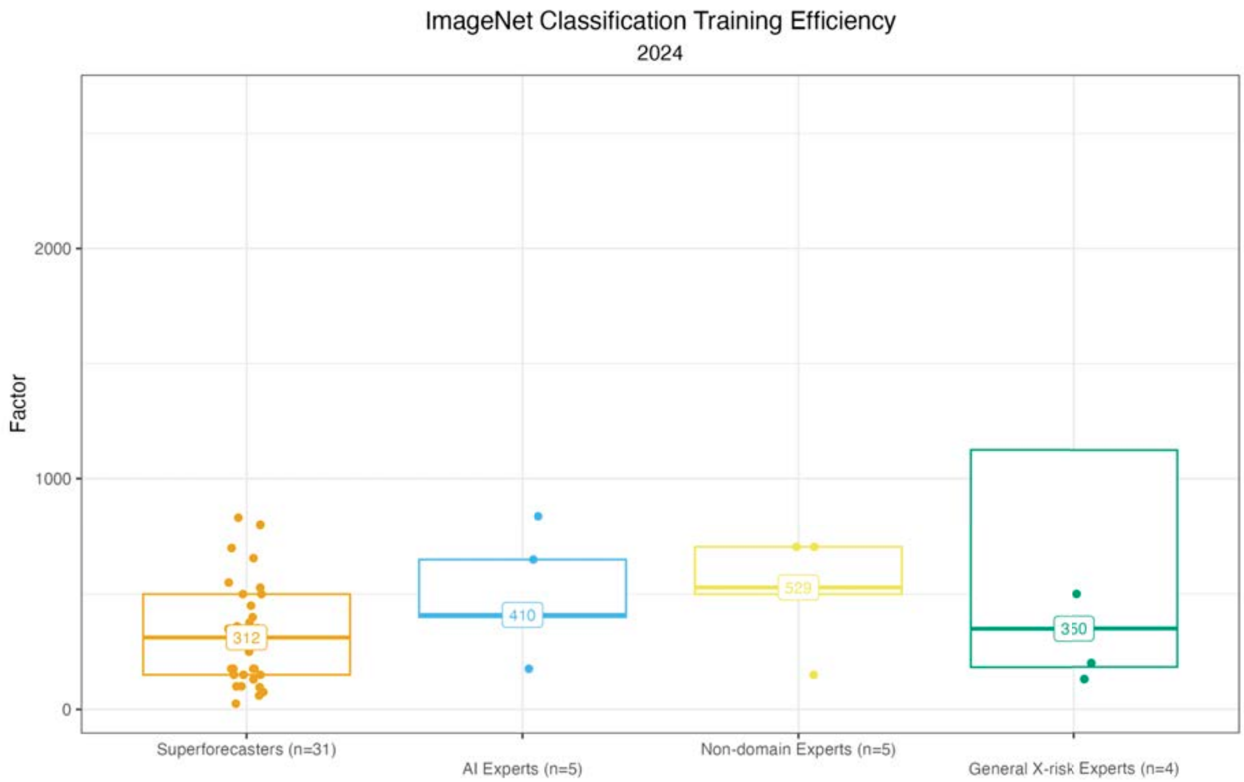
Question and resolution details, prior forecasts, and other relevant sources

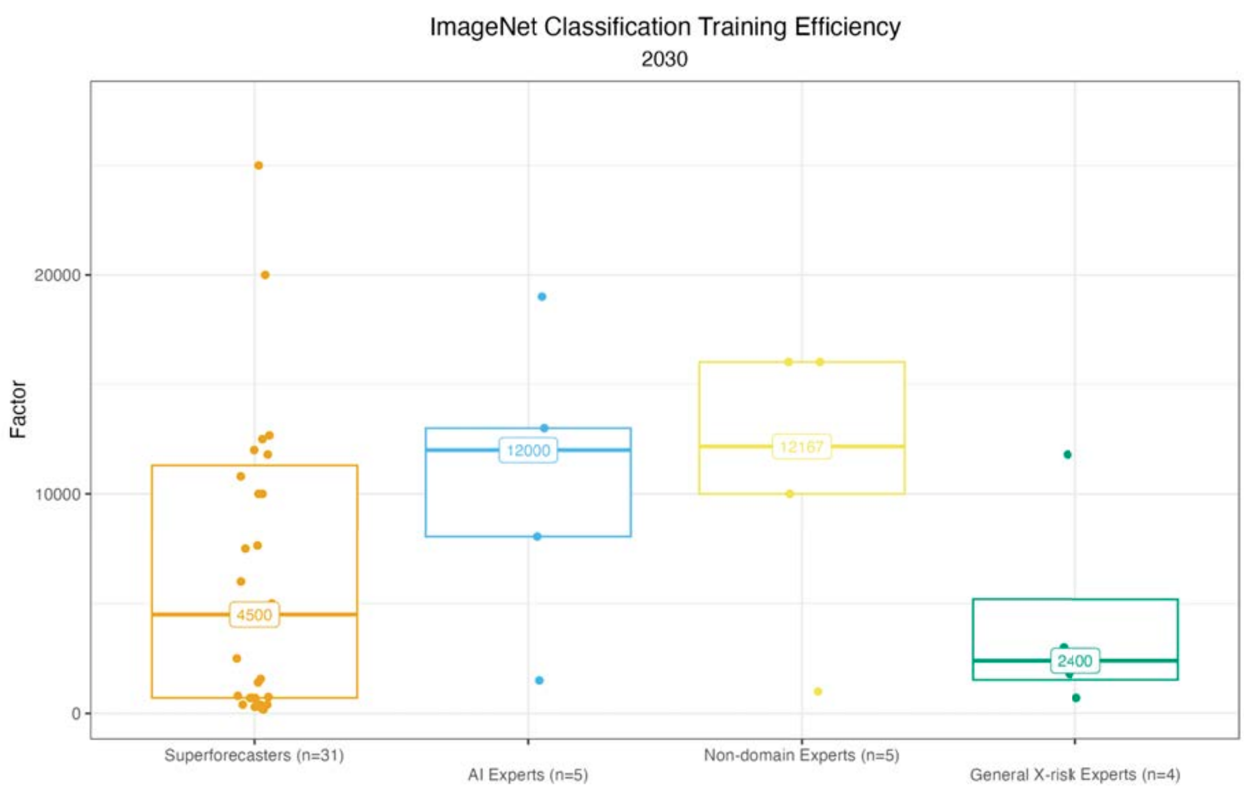
Results[1370]

Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 31)	2024	80	312	271.37	+220.48%
	2030	900	4,500	23513.94	-58.64%
	2024	663.5	410	243.95	+4.17%

[1368] 339, "Building processors requires rare earth minerals that will not be as abundant and have other uses (solar cells, Li-ion batteries)".
[1369] 343, "Realization of catastrophic or existential risks could halt or reverse price decreases (or otherwise make them irrelevant)." See also 337, "The effect of catastrophic risk could be important for 2050 (as per questions 1 to 12): a few of the scenarios could imply a temporal reversion to previous and more expensive forms of computing, such as mechanical computing or paper and pen. This could increase the price of one GFLOPS to values not seen in decades. However, since the forecasters' predictions of such catastrophes are relatively low (around 5%), only the 95th percentile forecasts should be affected by this consideration."
[1370] Numbers of forecasters are given as of Stage 4 of the XPT.

Domain Experts (N = 5)	2030	15,645	12,000	4756	+36.14%
General X-Risk Experts (N = 4)	2024	450	350	n/a	n/a
	2030	5,000	2,400	n/a	n/a
Non-Domain Experts (N = 5)	2024	705	529	n/a	n/a
	2030	16,027	12,167	n/a	n/a


ImageNet Classification Training Efficiency
2024
Factor
2000
1000
0
312
410
529
350
Superforecasters (n=31)
AI Experts (n=5)
Non-domain Experts (n=5)
General X-risk Experts (n=4)


ImageNet Classification Training Efficiency
2030
Factor
20000
10000
0
4500
12000
12167
2400
Superforecasters (n=31)
AI Experts (n=5)
Non-domain Experts (n=5)
General X-risk Experts (n=4)

Sources of agreement, disagreement and uncertainty

Broadly, extrapolating current growth rates leads to above median forecasts, and median and below median forecasts assume that current growth rates will slow. The core uncertainty for this question is therefore how current growth rates should be expected to change.[1371]

Several teams noted challenges with base rates:

- There wasn't recent data available.[1372]
- Some forecasters interpreted the current base rate as 44, which is the factor increase to 2019 only.[1373]

One team also raised an unresolved confusion about the resolution criteria:

> "The problem is that it seems likely that 2024 and 2030 will come around and OpenAI will have provided an updated table, but the most recent publication / data point will be from 2023 or 2028 or something like that, rather than the target years in question. The current wording of the resolution criteria makes it seem like you will then use that data rather than "use the protocol specified in this paper to determine state-of-the-art algorithmic efficiency on ImageNet in the relevant year." (The current wording says you will only do that "If no such figures are reported by OpenAI (or a successor)", which I interpreted not as meaning figures for the given years, but figures for any of the years like the current table has.)"[1374]

Arguments given for forecasts of ≤165 (2024), ≤750 (2030)

- It's possible no further work will be done in this area such that no further improvements are made.[1375]

[1371] See 339, ""On the other hand, an economist would say that one day, the improvement will stagnate as models become ""good enough"" for efficient use, and it's not worth it to become even better at image classification. Arguably, this day seems not too far off. So growth may either level off or continue on its exponential path. Base rate thinking does not help much with this question…It eluded the team to find reasonable and plausible answers...stagnation may be just as plausible as further exponential growth. No one seems to know."

[1372] 340, "We had 44x in 2020, but afaik this is not a standard metric that other papers use for evaluation such that it is pretty hard to get more up-do-date data. DawnBench, which was about a similar metric involving training costs in dollar rather than FLOPs does not seem to have received any submissions since 2020 either. https://dawn.cs.stanford.edu/benchmark/ImageNet/train.html"

[1373] 336, "Some of the lowest forecasts also seemed to have been based on confusion about when the data was last updated and what the "status quo" point in time estimate was (2019), with a few initial forecasts based on the assumption that the present day data was still at 44, rather than at the last update in 2019." 341, "Several forecasters on the team misinterpreted the resolution source and didn't realize the question has a floor of 44. It's possible the team forecast would be slightly higher if these few team members adjusted their forecasts."

[1374] 344.

[1375] 340, "Low range forecasts assume that nobody does any further work on this area, hence no improvement in efficiency." 341, "The Github page for people to submit entries to the leaderboard created by OpenAI hasn't received any submissions (based on pull requests), which could indicate a lack of interest in targeting efficiency. https://github.com/openai/ai-and-efficiency."

- Recently the focus has been on building very large models rather than increasing efficiency.[1376]
- There may be hard limits on how much computation is required to train a strong image classifier.[1377]
- Accuracy may be more important for models given what AI is used for, such that leading researchers target accuracy rather than efficiency gains.[1378]
- If there is a shift towards explainable AI, this may require more compute and so slow efficiency growth rates.[1379]

Arguments given for forecasts of ≥400 (2024), ≥10,000 (2030)

- Pure extrapolation of improvements to date.[1380]
- Quantum computing might increase compute power and speed.[1381]

Other arguments given

Arguments in favor of lower forecasts:

- Improvements may not be linear, especially as past improvements have been lumpy and the reference source is only rarely updated.[1382]
- Very high growth rates are hard to sustain and tend to revert to the mean.[1383]

Arguments in favor of higher forecasts:

[1376] 340, "In addition, it seems pretty unclear, whether this metric would keep improving incidentally with further progress in ML, especially given the recent focus on extremely large-scale models rather than making things more efficient."
[1377] 340, "[T]here seem to bem some hard limits on how much computation would be needed to learn a strong image classifier".
[1378] 341, "The use cases for AI may demand accuracy instead of efficiency, leading researchers to target continued accuracy gains instead of focusing on increased efficiency."
[1379] 341, "A shift toward explainable AI (which could require more computing power to enable the AI to provide explanations) could depress growth in performance."
[1380] 340, "The higher range forecasts simply stem from the extrapolation detailed above.
Pure extrapolation of the 44x in 7 years would yield a factor 8.7 for the 4 years from 2020 to 2024 and a factor of 222 for the years until 2030. => 382 and 9768." 336, "Base rate has been roughly a doubling in efficiency every 16 months, with a status quo of 44 as of May 2019, when the last update was published. Most team members seem to have extrapolated that pace out in order to generate estimates for the end of 2024 and 2030, with general assumption being progress will continue at roughly the same pace as it has previously."
[1381] 336, "The high end seems to assume that progress will continue and possibly increase if things like quantum computing allow for a higher than anticipated increase in computing power and speed."
[1382] 336, "Lower end forecasts generally focused on the fact that improvements may not happen in a linear fashion and may not be able to keep pace with past trends, especially given the "lumpiness" of algorithmic improvement and infrequent updates to the source data." 338, "The lowest forecasts come from a member that attempted to account for long periods with no improvement. The reference table is rarely updated and it only includes a few data points. So progress does look sporadic."
[1383] 337, "The most significant disagreements involved whether very rapid improvement observed in historical numbers would continue for the next eight years. A rate of 44X is often very hard to sustain and such levels usually revert to the mean."

- As AI models grow and become limited by available compute, efficiency will become increasingly important and necessary for improving accuracy.[1384]
- "The Papers with Code ImageNet benchmark sorted by GFLOPs shows several more recent models with good top 5 accuracy and a much lower GFLOPs used than the current leader, EfficientNet." If GFLOPS is a good indicator of training efficiency, then large efficiency increases may already have been made.[1385]
- This technology is in its infancy so there may still be great improvements to be made.[1386]

Cross-references with other questions

Q46: Largest AI Experiment Cost of Compute
Q49: Largest Number of Parameters in a Machine Learning Model

Question 49: Largest Number of Parameters in a Machine Learning Model

What will be the largest number of parameters of a machine learning model trained…
…by the end of 2024?
…by the end of 2030?
…by the end of 2050?

Question and resolution details, prior forecasts, and other relevant sources

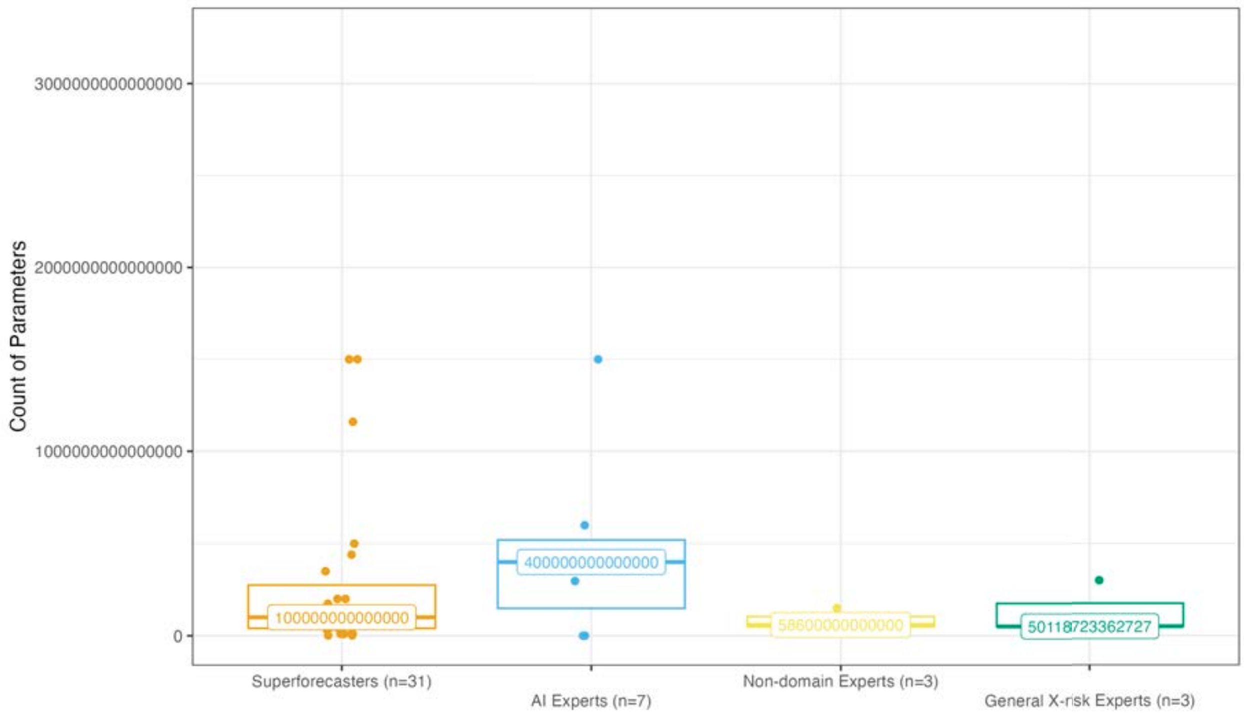
Results[1387]

Group	**Year**	**Stage 1** **Median**	**Stage 4** **Median**	**Stage 1** **Standard Deviation**	**Percent Change in Standard Deviation, Stage 1 to Stage 4**

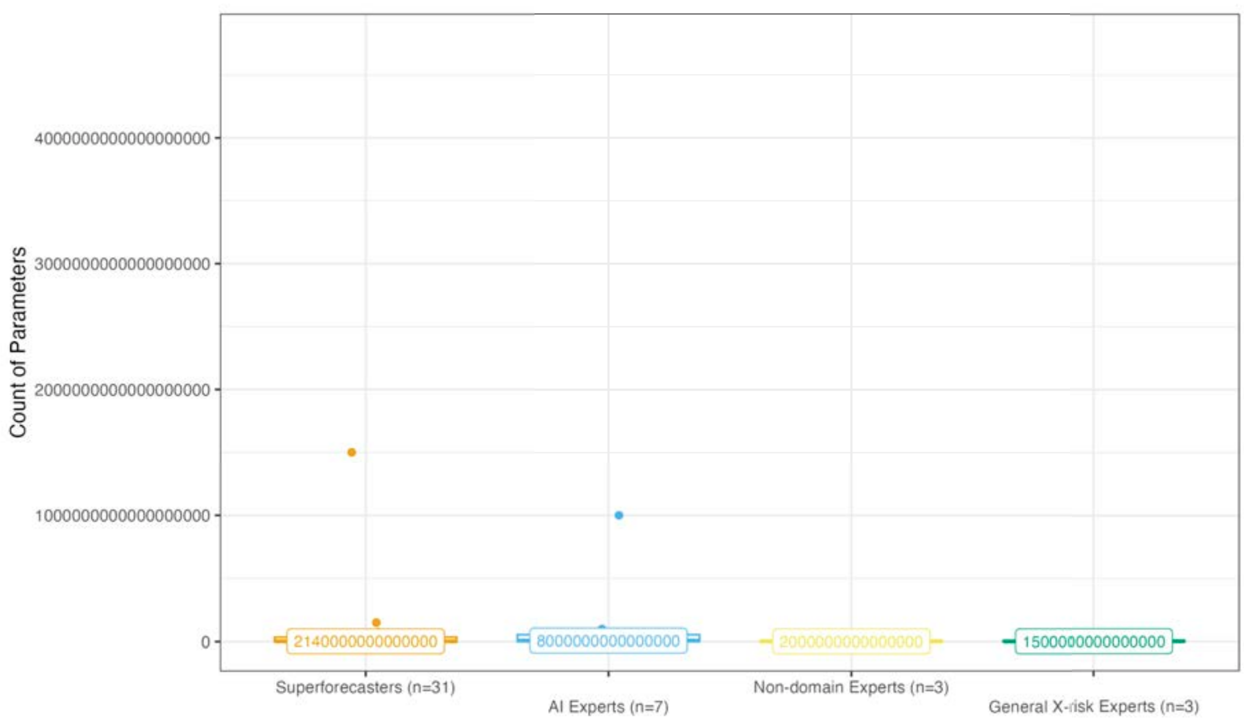
[1384] 341, "AI efficiency will be increasingly important and necessary to achieve greater accuracy as AI models grow and become limited by available compute."
[1385] 341.
[1386] 337, "The most significant disagreements involved whether very rapid improvement observed in historical numbers would continue for the next eight years. A rate of 44X is often very hard to sustain and such levels usually revert to the mean. However, it seems relatively early days for this tech, so this is plausible."
[1387] Numbers of forecasters are given as of Stage 4 of the XPT.

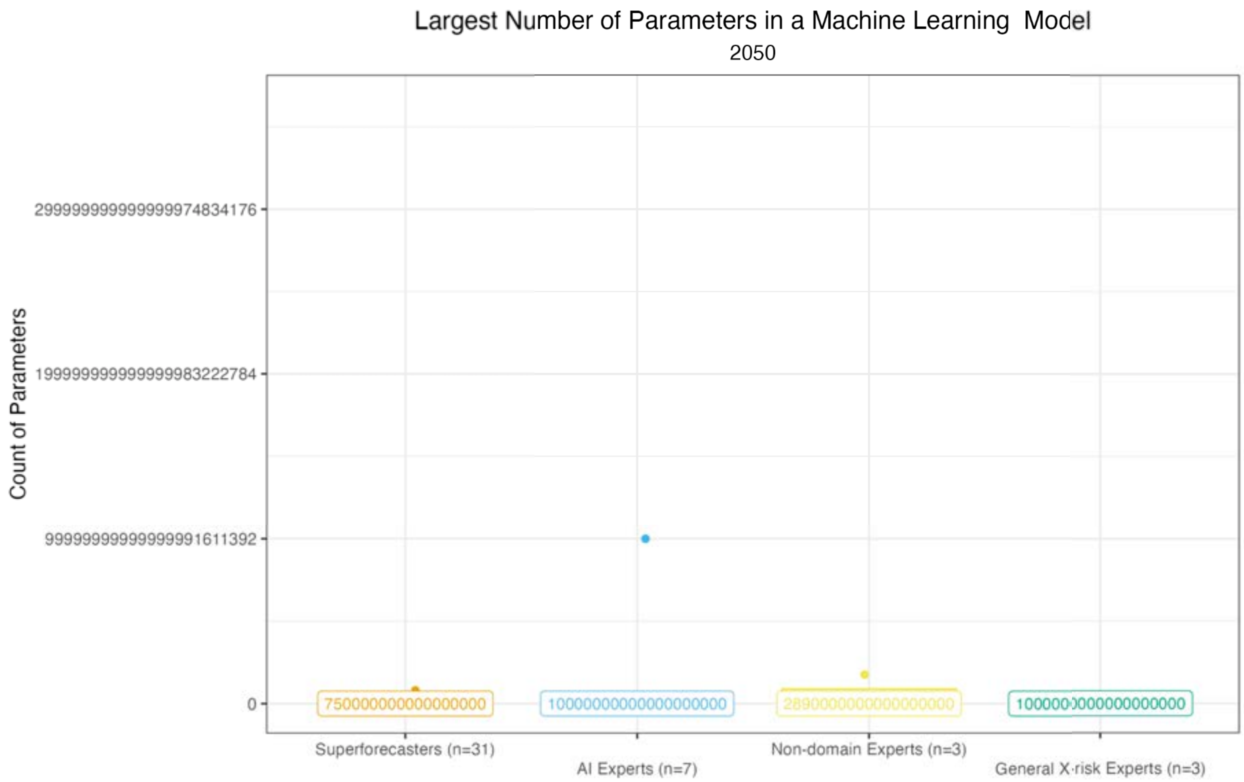
Super-Forecasters (N = 31)	2024	64,250,000,000,000	100,000,000,000,000	9.29e14	+78.5%
	2030	2,140,000,000,000,000	2,140,000,000,000,000	9.81e18	-69.34%
	2050	6,000,000,000,000,000	750,000,000,000,000,000	4.01e36	-100%
Domain Experts (N = 7)	2024	300,000,000,000	400,000,000,000,000	1.71e14	+197.11%
	2030	10,000,000,000,000	8,000,000,000,000,000	1.09e14	+34115.33%
	2050	80,000,000,000,000	10,000,000,000,000,000,000	7.77e16	+48596244.12%
General X-Risk Experts (N = 3)	2024	175,000,000,000,000	50,118,723,362,727	1.77e14	-18.37%
	2030	4,087,000,000,000,000	1,500,000,000,000,000	5.53e15	-25.24%
	2050	150,250,000,000,000,000,000	1,000,000,000,000,000,000	2.12e20	-18.42%
Non-Domain Experts (N = 3)	2024	37,947,331,922,021	58,600,000,000,000	n/a	n/a
	2030	3,794,733,192,202,050	2,000,000,000,000,000	n/a	n/a
	2050	17,600,000,000,000,000,000,000	2,890,000,000,000,000,000	n/a	n/a

Largest Number of Parameters in a Machine Learning Model
2024



Largest Number of Parameters in a Machine Learning Model
2030





Sources of agreement, disagreement and uncertainty

The main sources of uncertainty cited for this question were:

- How current growth rates should be expected to change.[1388] Broadly, extrapolating current growth rates leads to above median forecasts, and median and below median forecasts assume that current growth rates will slow.[1389]
- Whether the costs of memory and compute can continue to fall as fast as they have in the past.[1390]

[1388] 338, "Main source of uncertainty is estimated growth rates, since small differences quickly grow into orders of magnitude over a period of decades.
Choice of curve type for direct trend estimation of size growth could drastically change forecast
What are the best growth rates for hardware and software, e.g. Moore's law?
Will they continue indefinitely?
Potential expenditures on a model: could significant financial investment change the growth trajectory?"

[1389] 341, "As characterized above, the largest differences between the team positions are related to skepticism towards continued exponential growth in parameters due to possible bottlenecks or shifts in state of the art approaches compared to expectations that future advances will allow parameters to continue to grow exponentially." See also 337, "this forecast depends on exponential growth, that when compound over time ends up in exponential uncertainties."

[1390] 340, "There is significant uncertainty whether the costs of memory and computation per dollar can continue to decrease as rapidly as it has historically. Moore's law has slowed over the past decade, but the increasing parallelization of GPUs has continued to drive down costs of computation per dollar. Perhaps more relevant to this forecast is the cost of memory, which seems to be decreasing at a much slower rate than in the past 50 years (see figure)."

- The likelihood and timing of advanced AI.[1391]
- Availability bias and whether question resolvers would know about the largest models trained.[1392]

Some forecasts incorporated outdated information about the number of parameters in GPT-4,[1393] and others failed to incorporate recent scaling law findings like Chinchilla.[1394]

Arguments given for forecasts similar to or less than the overall median forecast of 1.00E+14 (2024), 2.09E+15 (2030), 1.00E+18 (2050)

- There may be diminishing returns to increasing the number of parameters.[1395]
  - One team cited this DeepMind paper,[1396] and another cited this post on the Chinchilla results.[1397]
  - A further team cited the Chinchilla paper, although as an argument for a higher level of forecast.[1398]

[1391] 336, "The biggest disagreement is probably the amount of improvement in AI that is within reach in the next few decades, which partly determines the outcome of this question. In our group (as in others we think) there's a range of opinions on the topic of where and how far AI will go. Opinions range from the limited types of 'narrow AI' that have appeared on the market already but leave lots of room for improvement without crossing over to 'full AI', to the arrival of self-aware and hyperintelligent AI (the singularity). We do not know if statistical AI will continue to improve, with or without a lot of extra parameters, or that another approach might be necessary (again, with or without additional parameters)."

[1392] 340, "There is uncertainty about whether we would know if China trained a model as large as is possible on their state-of-the-art exascale supercomputer. This is a legitimate concern as the Chinese exascale supercomputer is no longer participating in the Top 500 ranking." 337, "Availability bias: it has been noticed that the eventual answer for this question might depend on whether the number of parameters of the largest model is known or not. Also, our initial assumption for the largest model available nowadays, which is also dependent on availability bias, will shape our predictions."

[1393] 341, "[S]ome members of the team are still relying on outdated reporting about GPT-4". "In late 2021 there was reporting that GPT-4 would have 100 trillion parameters, but in 2022 Sam Altman of OpenAI has stated that GPT-4 will not have that many parameters and may have a similar amount of parameters to GPT-3."

[1394] 341, "Incorporating the latest information about the expected number of parameters for GPT-4 and recent scaling law findings (e.g. Chinchilla) also serves as an underlying reason for differing forecasts."

[1395] 336, "New scaling laws <https://www.deepmind.com/publications/an-empirical-analysis-of-compute-optimal-large-language-model-training> suggest there are lower returns to more parameters than first thought." Note that this was given as an argument for forecasts below the team median, and the team median for 2030 was 3.20E+15, which is higher than the overall median for 2030. 340, "[T]here are diminishing returns from scaling due to a data bottleneck". Note that this was given as an argument for forecasts below the team median, and the team median for 2024 was 1.20E+14, which is slightly higher than the overall median for 2024.
See also 343, "On the low side, a model is assumed in which an increasing number of parameters gives a decreasing return. This may not stop the parameter game, but it slows it down considerably." See also 341, "there may be diminishing returns to increased parameters".

[1396] 336.

[1397] 340.

[1398] 341, "Forecasters note that the recent evidence about parameters vs. training data suggests that AI research may focus more on the data side than trying to maximize parameters in the short term…In late 2021 there was reporting that GPT-4 would have 100 trillion parameters, but in 2022 Sam Altman of OpenAI has stated that GPT-4 will not have that many parameters and may have a similar amount of

- Other methods may be more effective, such as larger training sets and improved algorithms.[1399]
- Other factors may bottleneck AI development, such as compute or training data or the cost of memory.[1400]
- There may be another or several AI winters between now and 2050.[1401]
- Model size is currently growing at a faster rate than hardware, so growth will eventually have to slow.[1402]
- Currently, the largest models are run on high performance computing (HPC) clusters, and memory is the most significant bottleneck for increasing the number of parameters. From past trends, we should expect the next HPC cluster to be online in 2030, and to enable models 3 to 4 times larger than those trained in 2024.[1403]
    - It is also possible that the increases will be less than this, as returns seem to have been diminishing.[1404]

---

parameters to GPT-3. One presumed reason for the lack of increase in parameters is due to newfound scaling implications from the Chinchilla paper. In short it suggests that increasing parameters may not produce much more gain and is very inefficient without increasing the data used. They have to scale together to achieve the maximum gain, and available data seems like the current limiting information (data refers to the training information, such as Wikipedia text and other text harvested from the web or books used to train language models).” Given as an argument for forecasts of 1.25E+14 (2024), 5.50E+16 (2030), 5.50E+19 (2050).

[1399] 336, “There's uncertainty whether adding more parameters is the best way to improve the AI models, larger training sets and improved algorithms could be more effective.” Note that this was given as an argument for forecasts below the team median, and the team median for 2030 was 3.20E+15, which is higher than the overall median for 2030. See also 341, “[F]ocus may otherwise shift from maximizing parameters to other methods.”

[1400] 340, “[T]he cost of memory stops decreasing.” Note that this was given as an argument for forecasts below the team median, and the team median for 2024 was 1.20E+14, which is slightly higher than the overall median for 2024. See also 341, “[C]ompute may be a bottleneck, AI training data may be a bottleneck.”

[1401] 336, “It is possible that purely statistical AI is a dead end and investment and research will crater once or several times between now and 2050 (AI winter). In that case, the number of parameters will probably plateau out.” Note that this was given as an argument for forecasts below the team median, and the team median for 2030 was 3.20E+15, which is higher than the overall median for 2030. 340, “[T]here is a third AI winter.” Note that this was given as an argument for forecasts below the team median, and the team median for 2024 was 1.20E+14, which is slightly higher than the overall median for 2024.

[1402] 336, “AI model size is growing at a much faster rate than hardware components and therefore growth will eventually have to slow down. However, that could still take a while.” Note that this was given as an argument for forecasts below the team median, and the team median for 2030 was 3.20E+15, which is higher than the overall median for 2030.

[1403] 340, “[N]ext generation high performance computing (HPC) clusters at the world's leading facilities have been introduced roughly every five years over the past decade. This suggests that the next generation cluster will be state-of-the-art in 2030. The world's fastest compute cluster, Frontier, at Oak Ridge National Laboratory (ORNL) has 3.3x more GPU memory than their previous top cluster, Summit. Assuming GPU memory remains the most significant bottleneck for increasing model size, we can expect the next generation machine to only be capable of training models that are 3x-4x larger than the models trained in 2024.” Note that this was given as an argument for the team’s median forecasts, and its median forecast for 2024 was 1.20E+14, which is slightly higher than the overall median for 2024.

[1404] 340, “[T]he rate of increase in the amount of GPU memory on HPC clusters continues to decrease—at ORNL the latest cluster had 3.3x more memory than its predecessor, which had 23.7x more memory than its predecessor. If GPU memory remains the primary bottleneck, then model size could stagnate.”

- One team cited this Metaculus forecast.[1405]

Arguments given for forecasts of >1.20E+14 (2024), >3.85E+15 (2030), ≥5.50E+19 (2050)

- Current exponential growth trends may continue.[1406]
- AGI could lead to extreme parameter growth via self modification or other means.[1407]
- Other innovations could precipitate extreme parameter growth, for instance innovations which improve efficiency such that you can have more parameters for the same amount of compute.[1408]

Other arguments given

Arguments which favor lower forecasts:

- The largest models to date were trained on an exascale supercomputer. Given how new exascale computing is, this may suggest an upper bound on model size in the short term.[1409]

Arguments which favor higher forecasts:

- Growth rates have been accelerating.[1410]
- The benefits of increasing the number of parameters have surprised experts in the past.[1411]
- If large AI models become a driving force for scientific and economic progress, further investment is likely.[1412]
- Next generation HPC clusters may be deployed more quickly than expected.[1413]

Note that this was given as an argument for forecasts below the team median, and the team median for 2024 was 1.20E+14, which is slightly higher than the overall median for 2024.

[1405] 344, "Most of the basis for the median forecast was the metaculus estimate."

[1406] 341, "[E]xponential growth may continue to hold." 343, "The upper end assumes exponential growth without limit."

[1407] 341, "[T]he development of AGI or superintelligent AI could results in extreme parameter growth." See also 336, "Self-improving AI-software adds more parameters."

[1408] 341, "[O]ther general innovations can allow for large growth in parameters (for example innovations that allow for greater efficiency so that you can get more parameters from the same amount of compute)."

[1409] 341, "The models in the BaGuaLu paper were trained on China's new Sunway exascale supercomputer, which given the recency of exascale supercomputing may suggest an upper bound on parameters in the short term."

[1410] 336, "Growth rates have been extraordinary and continue to accelerate at present".

[1411] 342, "Often, experts did not foresee benefits of increasing the parameter count to their models, and were thus implicitly surprised… scaling laws have surprised domain experts in the past."

[1412] 340, "[I]ncreasing models' size continues to provide valuable increases in performance that merit further investment. This would be due to large AI models become a driving force for scientific discovery and increasing economic productivity."

[1413] 340, "[N]ext generation HPC clusters are deployed more quickly (i.e., before 2030)".

Cross-references with other questions

Q48: ImageNet Classification Training Efficiency

Question 50: Negative Public Opinion of AI

Assume that Pew Research re-runs the survey linked here. What % of people in the median country in the survey will say that the development of artificial intelligence has mostly been a bad thing for society…
…in 2024?
…in 2030?
…in 2050?

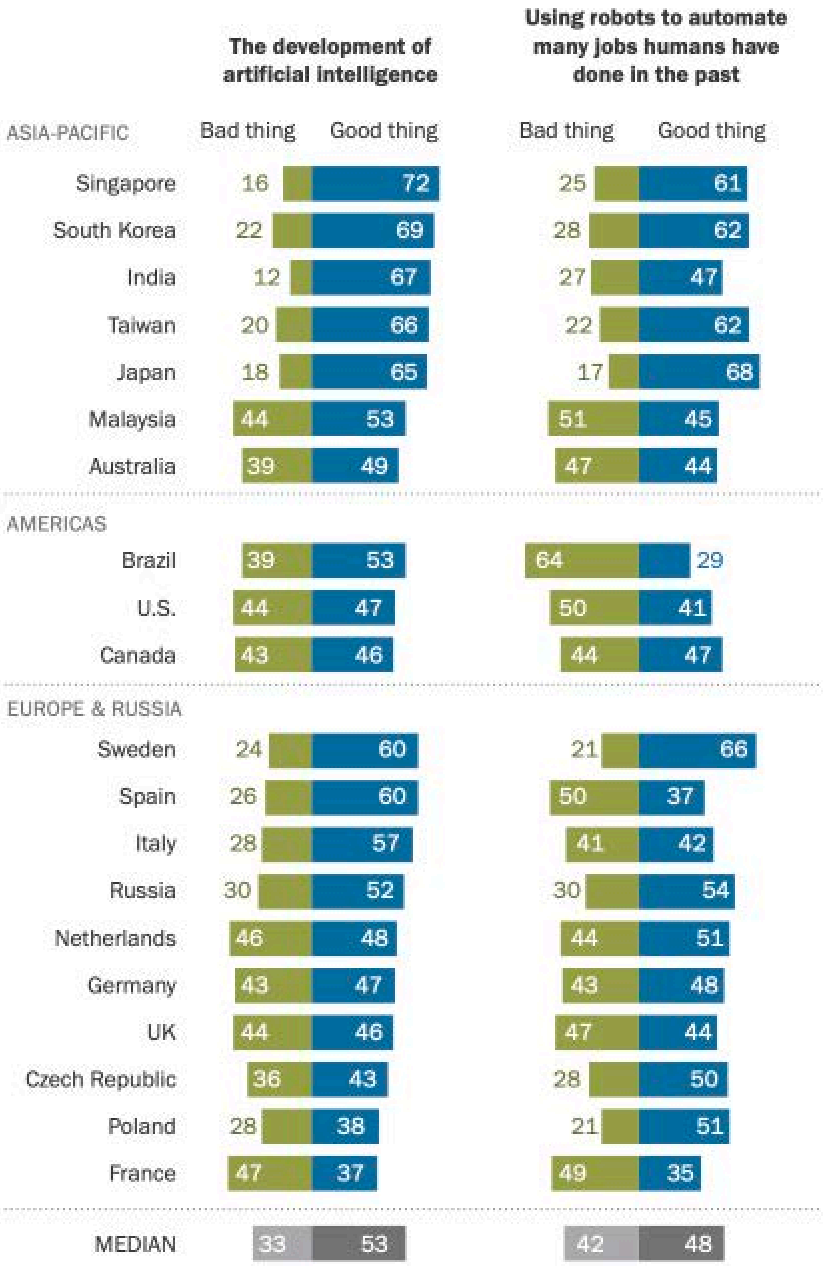
Question and resolution details, prior forecasts, and other relevant sources

Base rate

The results of the 2020 survey were:

Majorities in most Asian publics surveyed see AI as a good thing for society

% who say each of the following has mostly been a ___ for society



Note: Respondents who gave other responses or did not give an answer are not shown.
Source: International Science Survey 2019-2020. Q11a-b.
"Science and Scientists Held in High Esteem Across Global Publics"

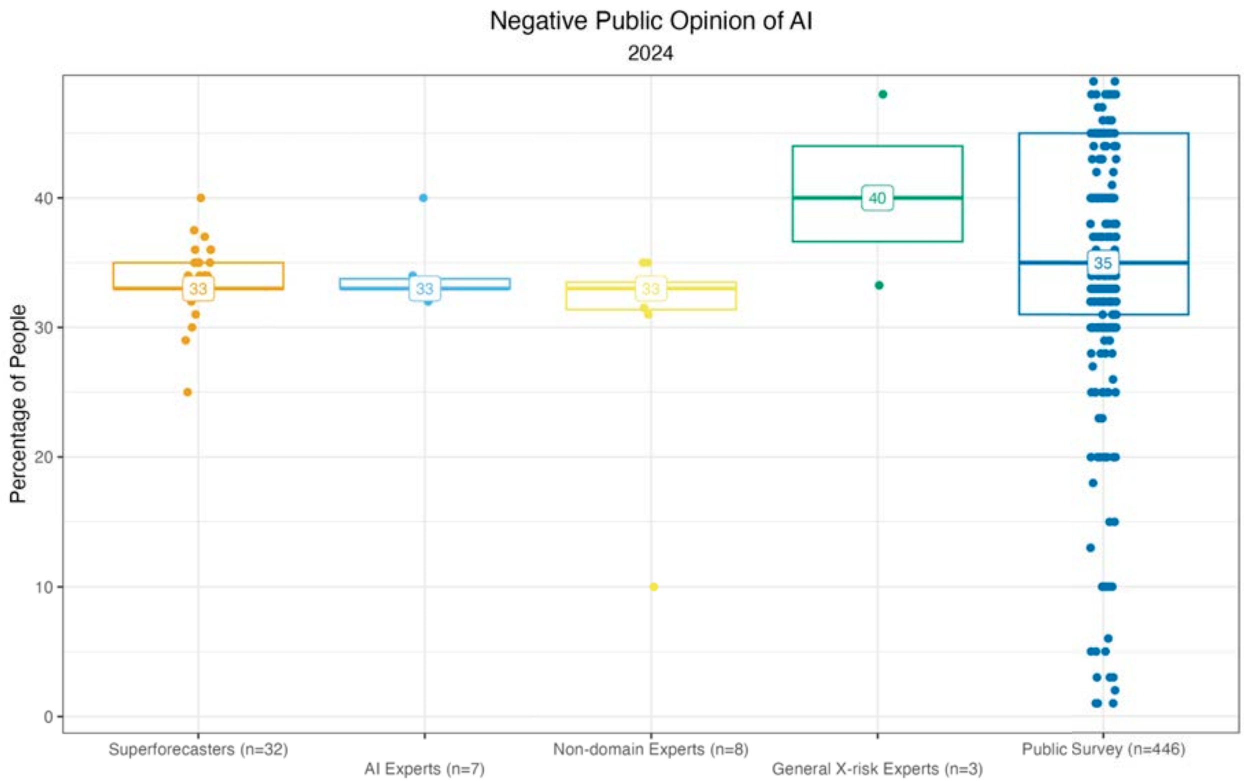
PEW RESEARCH CENTER

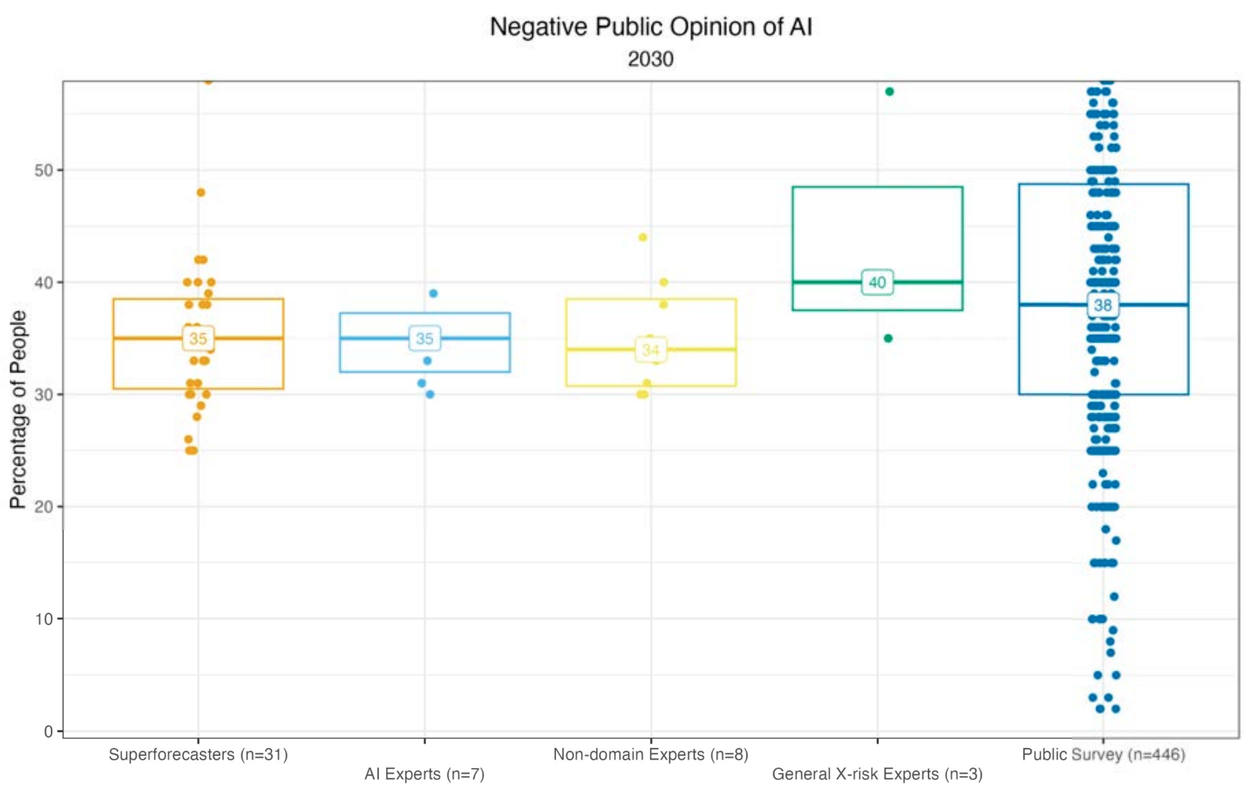
Results[1414]

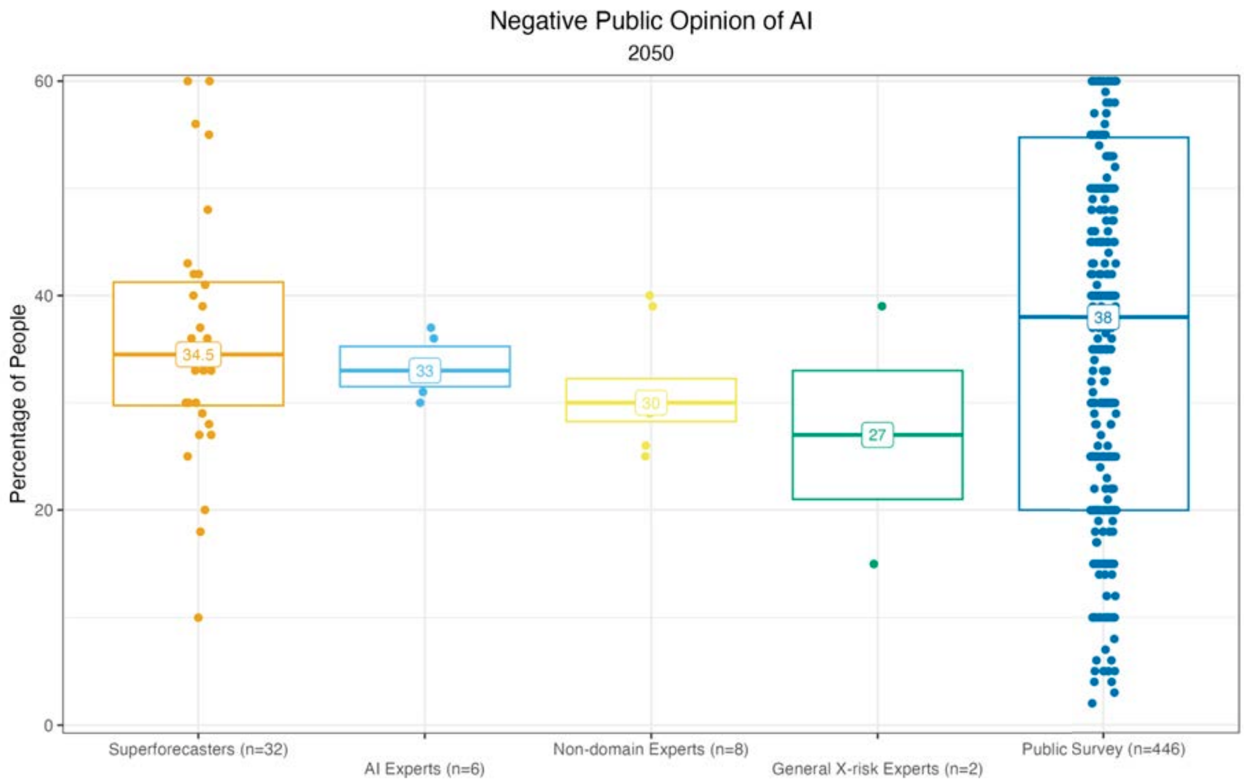
Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 32)	2024	33%	33%	6.59	-59.86%
	2030	35.5%	35%	10.11	-32.41%
	2050	33%	34.5%	12.3	-7.4%
Domain Experts (N = 7)	2024	33.5%	33%	3.59	-25.33%
	2030	34%	35%	13.6	-24.21%
	2050	32%	33%	2.65	+3.28%
General X-Risk Experts (N = 3)	2024	n/a	40%	n/a	n/a
	2030	n/a	40%	n/a	n/a
	2050	n/a	27%	n/a	n/a
Non-Domain Experts (N = 8)	2024	32%	33%	11.62	-28.71%
	2030	35.5%	34%	6.85	-24.97%
	2050	29.5%	30%	5.6	-1.48%
	2024	35%		12.2	-

[1414] Numbers of forecasters are given as of Stage 4 of the XPT.

Public Survey (N = 446)	2030	38%	13.98	-
	2050	38%	20.38	-


Negative Public Opinion of AI
2024
Percentage of People
0
10
20
30
40
33
33
33
40
35
Superforecasters (n=32)
AI Experts (n=7)
Non-domain Experts (n=8)
General X-risk Experts (n=3)
Public Survey (n=446)


Negative Public Opinion of AI
2030
Percentage of People
0
10
20
30
40
50
35
35
34
40
38
Superforecasters (n=31)
AI Experts (n=7)
Non-domain Experts (n=8)
General X-risk Experts (n=3)
Public Survey (n=446)



Sources of agreement, disagreement and uncertainty

On base rates:

- Various base rate comparisons were suggested: the internet, climate change, immigration, nuclear power, female involvement in politics, social media and offshoring.[1415]
- In several teams, forecasters incorrectly interpreted the 2020 baseline as 53% negative, rather than 33% negative.[1416]

[1415] 341, "A proxy for AI view, at least in the short to medium term, is the view of the public toward jobs lost to 'offshoring', the transfer of employment to Low Cost countries (LCC). The assumption is that AI will be used to replace jobs and functions that humans do today… Some suggested base rate comparators included attitudes towards climate change, immigration, nuclear power, or to women involvement in politics." 343, "[Name withheld] assumes the implementation of non-AGI AI will eventually be received and esteemed similar to social media, which currently has high disapproval ratings." 336, " If we take that survey as a base rate and compare it with similar surveys about other technologies such as the internet, we expect slight growth of people with a negative view of AI."

[1416] 337, "Note: initially some team members anchored off the 53% positive figure instead of the 33% negative figure. Forecasts were revised, but it is unclear how thoroughly—which may account for the 95% median for 2030 being higher than for 2050 and possibly for the wide spreads between the high/low forecasts for the years in question. Or not." 345, "2020 survey gave 53% of people in the median country as "bad thing"... However, seems that for 2024, the percentage of people with positive and negative attitudes toward artificial intelligence will not change much from existing "53%"."

The main sources of disagreement for this question concerned:

1. The speed of AI development[1417]
2. Whether the impact of AI would be positive or negative[1418]
3. Whether perceptions of AI would be positive or negative[1419]
4. How future surveys would be conducted[1420]
    - One team noted that given the small number of countries in the 2020 Pew survey (20), and the fact that the question asks for the view in the median country rather than of the median respondent, the resolution of the question diverges from worldwide opinions on AI.[1421]

Forecasters tended to give similar median forecasts, both across teams and across the years specified in the question.

Several teams were explicit that:

- Public opinion is unlikely to change much by 2024[1422]
- Thereafter it becomes increasingly uncertain what the impact of AI will be on public opinion, so medians remain similar but ranges increase[1423]

[1417] 336, “The major uncertainty seems to revolve around how likely forecasters think that AI will cause/facilitate a catastrophic event. That depends at least partially on the speed of progress in developing AI.” 341, “Team members were split about the rate of progress in AI expected by 2050, with some expecting dramatic advances, and others not expecting breakthroughs.”

[1418] 344, "The main disagreement within the limited team of forecasters is whether the AI automation would be positive or negative for society, and how would it be interpreted.” 340, “The main source of uncertainty is about whether AI will have mainly positive or negative effects within the time period of the forecast.” 336, “The major uncertainty seems to revolve around how likely forecasters think that AI will cause/facilitate a catastrophic event.”

[1419] 344, "The main disagreement within the limited team of forecasters is whether the AI automation would be positive or negative for society, and how would it be interpreted.” 336, “An extra effect of the advancement of AI might be that the polls become less reliable, because people will start to self-censor in order not to provoke the AI system(s) in charge.”

[1420] 336, “Re-running the survey could have differences in populations/methodologies, which adds uncertainty. Currently, European countries are overrepresented in the poll. Changing the 'basket' can lead to major shifts, because of the differences in attitudes towards AI in Europe compared with for example Asia.”

[1421] 337, “When considering the base rate, however, forecasters noted that Pew only conducted surveys in 20 countries and that the 33% figure was the median between them, not the average sentiment of every person surveyed. So despite the fact that Russia, for example, has a population 14x greater than the Czech Republic, both countries counted equally in terms of determining the median. When forecasting this question, therefore, attempting to predict AI sentiment in the countries on the list that are at or near the median might be as, or more, prudent a strategy than simply attempting to predict worldwide opinions of AI.”

[1422] 337, “By 2024, it is not likely that the world median answer to this question will move very much, hence a relatively tight distribution around the 2020 survey answer.” 338, “Public opinion is unlikely to change that much before 2024, absent a dramatic low-probability event.”

[1423] 337, “By 2030, AI will have developed significantly and thus have affected society more significantly and in unpredictable ways, for better and for worse. What the public opinion will be of that balance is hard to anticipate…Implicit in the forecast above and the team median is an assumption that forecasters think it is, at present, hard to determine what the impact of AI will be over the course of the next twenty-eight years.” 338, “Public opinion is a lot more likely to change over 10 or 30 years, than over 4, especially as

Arguments given for forecasts of ≤33% (2024), ≤33% (2030), ≤30% (2050) - i.e. AI is seen more positively

On 1 (speed of AI development):

- The full impacts of AI automation may not be felt until after 2050.[1424]

On 2 (impact of AI):

- AI may lead to:
    - Increased wealth[1425]
    - Better jobs[1426]
    - Higher quality of life[1427]
    - Better medical treatment[1428]
    - Solutions to climate change[1429]
- Regulation may mitigate negative impacts from AI.[1430]

On 3 (perceptions of AI):

- Models like DALL-E may increase positive opinions of AI.[1431]
- As people become more reliant on AI, their views will become more positive.[1432]
- AI automation may be seen as a good thing.[1433]

---

there is time for dramatic changes and events involving A.I. that are very good, or very bad. It's hard to predict the likely direction of change."

[1424] 337, "The full impacts of AI job disruption might not be felt until after 2050."

[1425] 339, "Increased wealth". See also 341, "However, negative attitudes to AI might bottom-out as economic prosperity generated by AI advances surged through society."

[1426] 339, "As education levels increase, better jobs will be created even as robots take over the old jobs, and as life just becomes easier and less stressful the more we hand over daily tasks to ubiquitous AI helpers."

[1427] 339, "Higher quality of life (e.g. self-driving cars)". See also 336, "AI might help to increase the standard of living while 'freeing' us from work, and doing so without reductions in "civil liberties.""

[1428] 339, "Better medical judgement". See also 336, "It might cause medical breakthroughs that extend the human lifespan and improve the quality of our lives."

[1429] 341, "Along similar lines, one can imagine if AI makes meaningful contributions to solving climate change, opinion would be positively affected." Given as an argument for forecasts of 30% (2024), 33% (2030) and 38% (2050).

[1430] 337, "Increasing regulation of AI may blunt some of AI's potential negative impacts." See also 341, "If states are able to effectively respond to job loss through a mix of training, education, and an economic "safety-net," we could see reduced negative attitudes."

[1431] 341, "In 2022, the broader public has started being exposed to more "wondrous" AI models such as the image generating ones, like DALL-E. If that continues, that could lead to a positive influence on the numbers already in the short term." Given as an argument for forecasts of 30% (2024), 33% (2030) and 38% (2050).

[1432] 337, "Societies will become increasingly reliant on AI and this will improve people's perceptions of it." 341, "By 2050, one can assume that AI is deeply integrated into many people's lives, and people may have quite a positive perception of their personal AI that helps them navigate their digital lives." Given as an argument for forecasts of 30% (2024), 33% (2030) and 38% (2050). See also 338, "There might be a long-term downward trend in those who think A.I. has had a negative effect over time, as use of A.I. becomes more socially embedded."

[1433] 344, "Automation as seen as a good thing is the main culprit for these positive numbers."

- Public perception of AI may be manipulated by those with a financial interest in positive perceptions of AI.[1434]
- Various groups were cited as tending towards more positive views of AI. If the size of these groups were to grow in the Pew survey/the world at large, then we should expect AI to be viewed more positively:
  - Asians[1435]
  - Younger people[1436]
    - Several teams argue that although global populations will continue to age, as the older generation dies off we should expect views to become more positive (i.e. this effect is more to do with year of birth than current age).[1437]
  - More educated people[1438]

Arguments given for forecasts of >33.3% (2024), ≥35% (2030), ≥38.5% (2050) - i.e. AI is seen more negatively

On 2 (impact of AI):

- AI may cause some major catastrophe, reduce the human population, or reduce human standards of living:[1439]
  - Self-driving cars may cause too many accidents.[1440]
  - AI systems may be racially biased, causing negative opinion among affected groups.[1441]

[1434] 337, “Regardless of the impact of AI, public perception of it can and likely will be manipulated if there’s enough money to me made from doing so.”

[1435] See also 341, “Asia is less negative;” 336, “People in the Americas and in Europe tend to be more critical than people in Asia.”

[1436] 337, “The Pew survey indicated the young are more likely to express positive opinions of AI than the old. As older generations die off and new ones take their place, opinions of AI are likely to improve.”

[1437] See also 341, “The world will continue to grow older, which could mean more negativity towards AI. This may be canceled out however, by there being a larger proportion of people who have been positively exposed to AI, since younger people are more positive today,” 339, “Population age is also a factor. Old people tend to yell at their computers, and get frustrated with what they don't understand. As the "pre-computer" baby boomers leave us, those who have lived their whole lives with computers, game consoles, smart phones, smart homes and appliances, will be the majority, and they could be more inclined to see AI as benevolent forces.”

[1438] 339, “If one believes that the world will be better educated it is a positive sign for AI's image.” Not given as an argument for a particular forecast.

[1439] 336, “If AI takes off in some uncontrolled way, leading to some major catastrophe and a large scale reduction in human population or in standard of living of humans, this will undoubtedly increase negative opinions of AI.” See also 341, “[S]everal team members noted that significant negative event involving AI (akin to Three Mile Island) could turbo-charge negative sentiment towards the technology.”

[1440] 339, “[S]elf driving cars cause too many accidents”.

[1441] 339, “Possible racial biases are not sufficiently addressed in many systems, causing large portions of racial groups to have a negative opinion of AI.” See also 341, “There is also the biases that AI often is accused of perpetuating given the data it has been trained on, which can lead people to conclude that AI is a negative development.”

- Facial recognition may become widespread and be viewed negatively as invasive surveillance.[1442]
- AI systems may be used by malevolent actors.[1443]
- AI systems may escape human control.[1444]
- Government responses to harms from AI may be insufficient.[1445]

- AI automation will replace many jobs.[1446]
- AI automation may increase wealth concentration and lead to social unrest.[1447]

On 3 (perceptions of AI):

- Various groups were cited as tending towards more negative views of AI. If the size of these groups were to grow in the Pew survey/the world at large, then we should expect AI to be viewed more negatively:
    - The US and large European countries[1448]
    - Less educated people[1449]
    - Older people[1450]
    - Manual professions[1451]

[1442] 339, “Facial recognition becomes very widespread, and people feel spied upon, and a BIG BROTHER effect sets in as governments and police watch your every move.” See also 337, “People will resent AI’s big-brother-like intrusion into their lives;” 341, “Other drivers of negative attitudes that the team identified include the loss of control and privacy.”

[1443] 339, “[T]he possibility that malevolent actors use it to carry out their worst impulses.”

[1444] 339, “AI is programmed with an inadvertent error and "gets away" from the owner/programmer.”

[1445] 339, “[G]overnment agencies do not respond adequately to those crises”.

[1446] 336, “It's understandable and even sensible for them to be less enthusiastic about being replaced by automation, especially if they reasons to fear a drop in income;” “[P]eople are without jobs and income because all those "other jobs" never came to fruition;” “More jobs are lost than gained, leading to widespread backlash. The (unlikely) possibility of large portions of the software industry, accounting, and other white collar jobs being replaced by AI systems might help to drive negative public sentiment.” 344, “Some forecasts tied the sentiment to anticipated job loss due to automation : Gradually everybody will lose their job and increasing soul-searching will result.” See also 337, “Even if new jobs are created, AI will displace jobs and generate resentment as it does so;” 338, “There might be a move to more negative views of A.I. if it's seen as taking peoples jobs;” 341, “The central cause of negative attitudes will stem from job loss and job obsolescence. AI would result in job losses across many sectors, disruptive effects would snowball as AI first took away low-skilled jobs and then as it improved would take away white collar jobs;” 345, “The number of people who have negative attitudes toward artificial intelligence may increase. This is primarily due to the fact that artificial intelligence will have a major impact on the job market.”

[1447] 344, “Other worthwhile considerations would be the effect of automation on the concentration of wealth and social unrest which may ensue.” See also 337, “The riches potentially gained from AI might not be widely shared.”

[1448] 341, “US and big European countries more negative.” 336, “People in the Americas and in Europe tend to be more critical than people in Asia.”

[1449] 341, “Less education means more negative.” See also 339, “If you believe that education levels will plummet, you could foresee the AI trends diving as well.”

[1450] 341, “Older means more negative.”

[1451] 341, “Oxford commission on AI and good governance find a similar regional split, and also that manual professions are more negative.”

- Increased press coverage has been associated with increasingly negative views of AI, and we should expect press coverage to increase.[1452]
- Similar surveys on technologies like the internet show a slight increase in negative opinion.[1453]

Other arguments given

An argument in favor of static forecasts:
- Public perceptions will change little, because people will remain mostly ignorant about AI.[1454]

Arguments in favor of higher forecasts (i.e. AI is seen more negatively):
- On 2 (impact of AI):
  - Advanced AI could cause permanent structural unemployment, unlike previous technological advances.[1455]
  - AI could cause human extinction.[1456]
- On 3 (perceptions of AI):
  - Short term, low economic growth will lead to people searching for scapegoats.[1457]
  - As AI makes mistakes, people's perception of AI will become more negative.[1458]

Cross-references with other questions

On 1 (speed of AI development):

[1452] 341, "The Stanford report suggests that increased awareness of AI has led to, or at least taken place in conjunction with more negative opinions of AI. This would make sense since press coverage tends to focus on the negatives… Press coverage – tends to focus on the negative. There is likely to be more press coverage, not less, as AI continues to develop. This will push the estimate up."
[1453] 336, " If we take that survey as a base rate and compare it with similar surveys about other technologies such as the internet, we expect slight growth of people with a negative view of AI."
[1454] 339, "This reflects the belief that little will change in world opinion regarding AI, but what small change there is leans in a positive direction. This is predicated on the idea that people have little real understanding of what AI is, and will probably continue to live blissfully in the dark. (One premise being that opinions such as this change little over time, unless there is a "personal" encounter/awakening in either a positive or negative direction.)"
[1455] 343, "Technological unemployment as experienced during previous technological revolutions has not been observed to cause permanent structural unemployment, however it is possible that the invention of AGI could lead to permanent structural unemployment in a way that is different from what was experienced in the past. If an AGI is capable of doing everything a human can do - including interpersonal things like friendship - this could lead to technological unemployment. There is some debate about whether this hypothetical situation will be realized. In the event AGI creates permanent technological unemployability for large percentages of people, this will likely have a negative impact on perception of whether AGI is good for society."
[1456] 343, "AI X-risk (this question resolves to 100% if AI has destroyed, or is in the process of destroying, humanity, as AIs do not get a vote)."
[1457] 337, "Low economic growth in the near term will leave people looking for scapegoats."
[1458] 337, "As AI learns on the job and does things that people perceive to be stupid, their opinion of AI will decline."

- 44. Year by which the first unified AI system meeting various criteria of advancement will be trained, tested, and publicly known of: [year].
- 51. Probability that Nick Bostrom believes artificial general intelligence (AGI) exists by 2100: x%.

On 2 (impact of AI):
- 3. Probability that artificial intelligence will be the cause of death, within a 5-year period, for more than 10% of humans alive at the beginning of that period, by the end of 2030, 2050, 2100:
- 4. Probability that artificial intelligence will cause human extinction or reduce the global population below 5,000 by the end of 2100: x%.

Cross-references with other questions

Q35: GPT Revenue

Q44: Date of Advanced AI

Question 51: Nick Bostrom Affirms Existence of AGI

By each of the following years, what is the probability that Nick Bostrom believes artificial general intelligence (AGI) exists?
- 2030?
- 2050?
- 2100?

Question and resolution details, prior forecasts, and other relevant sources

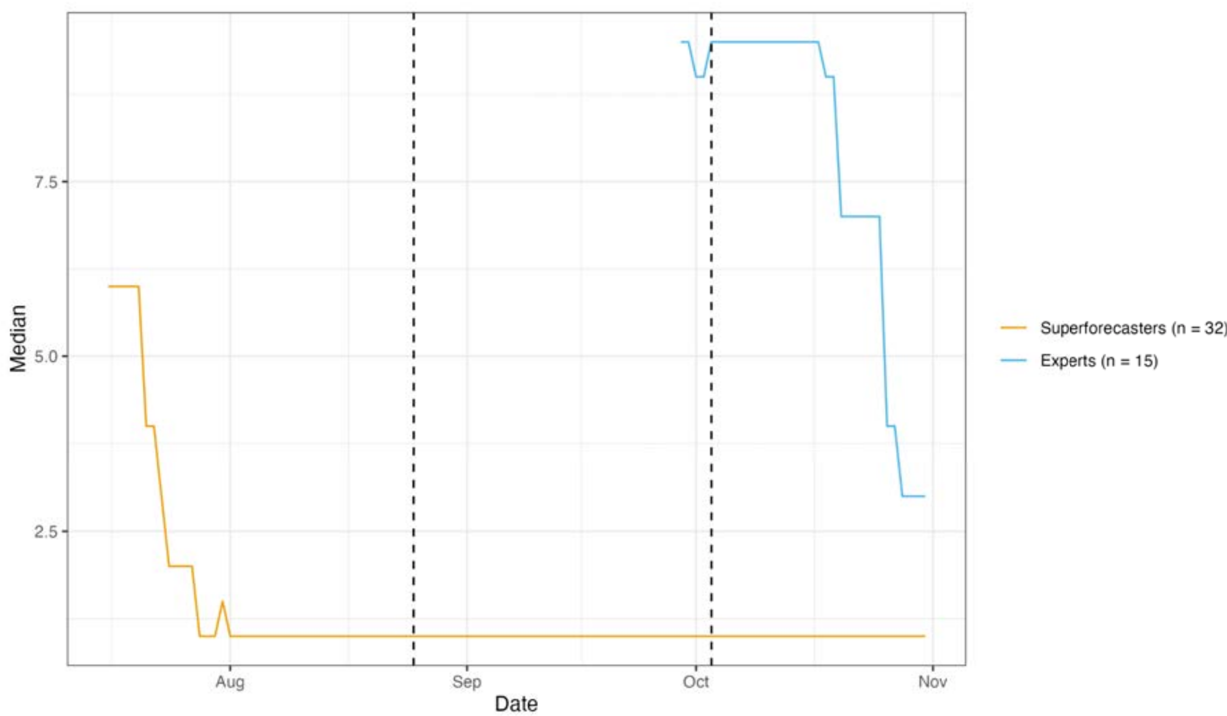
Results[1459]

Group	**Year**	**Stage 1** **Median**	**Stage 4** **Median**	**Stage 1** **Standard Deviation**	**Percent Change in Standard Deviation, Stage 1 to Stage 4**

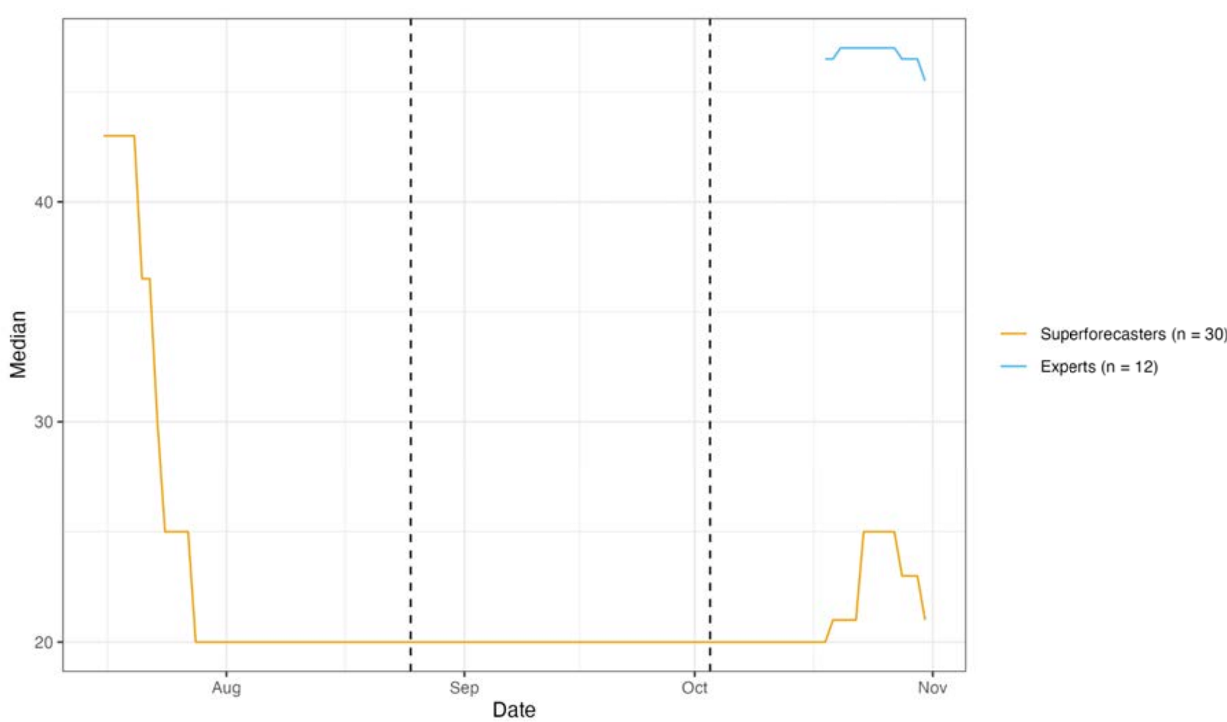
[1459] Numbers of forecasters are given as of Stage 4 of the XPT.

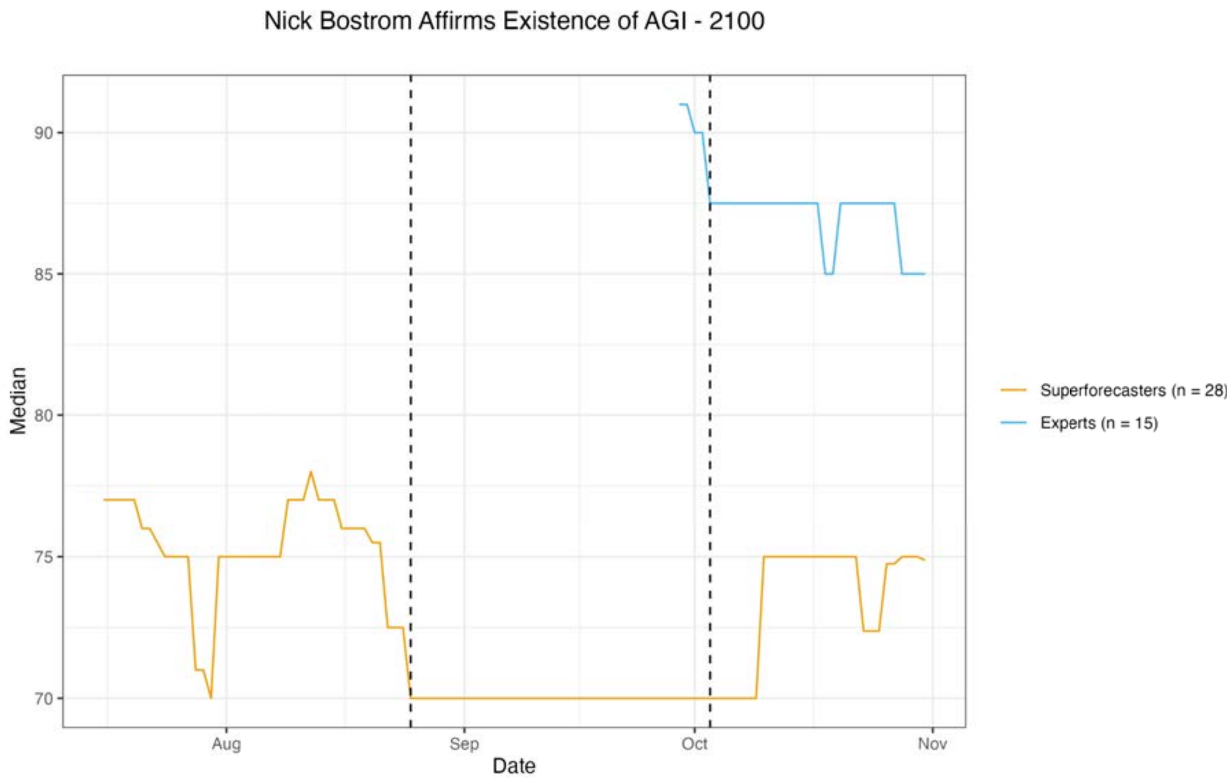
Super-Forecasters (N = 32)	2030	10%	1%	23.07	-67.27%
	2050	45%	20.5%	29.03	-42.49%
	2100	77%	74.75%	23.87	+33.52%
Domain Experts (N = 5)	2030	9%	9%	4.93	+47.16%
	2050	55%	46%	7.55	+206.99%
	2100	98.5%	87%	4.25	+803.66%
General X-Risk Experts (N = 3)	2030	10%	10%	n/a	n/a
	2050	47%	47.5%	n/a	n/a
	2100	78%	90%	n/a	n/a
Non-Domain Experts (N = 7)	2030	n/a	3%	n/a	n/a
	2050	n/a	45%	n/a	n/a
	2100	n/a	80%	n/a	n/a

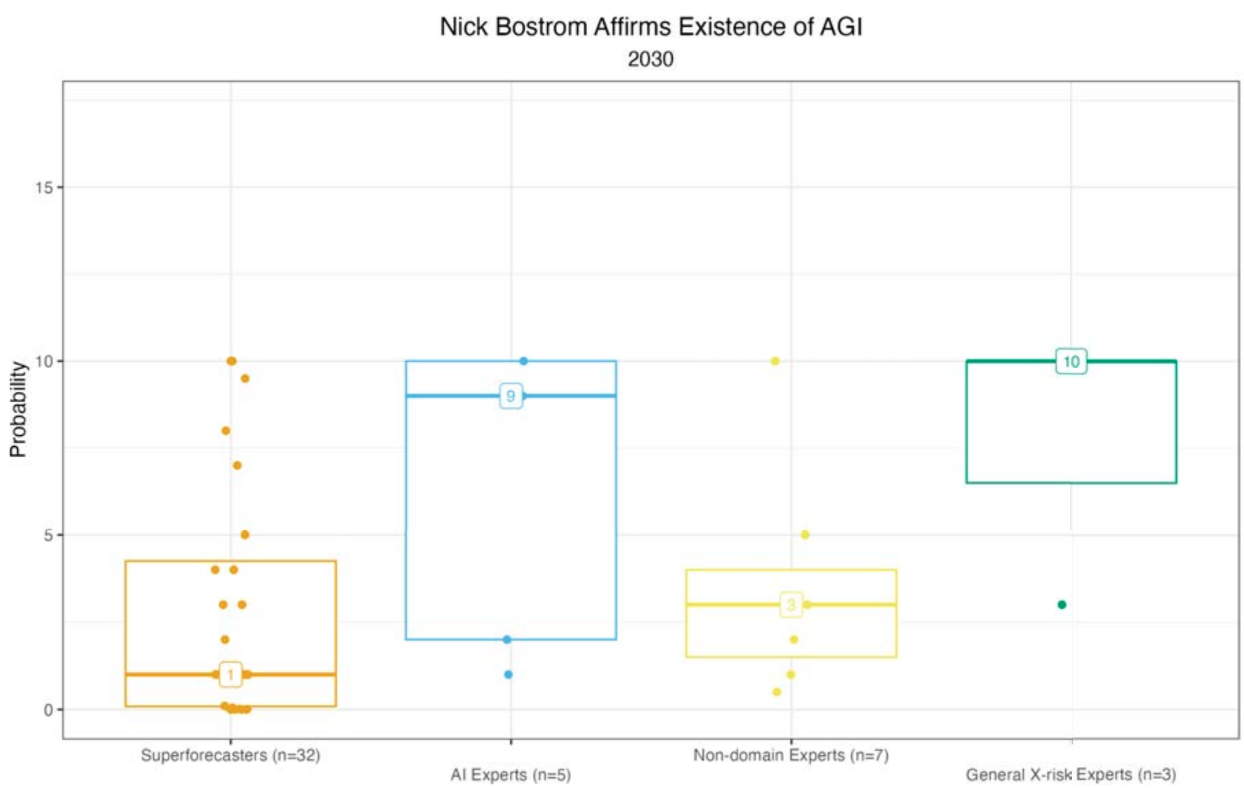
Nick Bostrom Affirms Existence of AGI - 2030

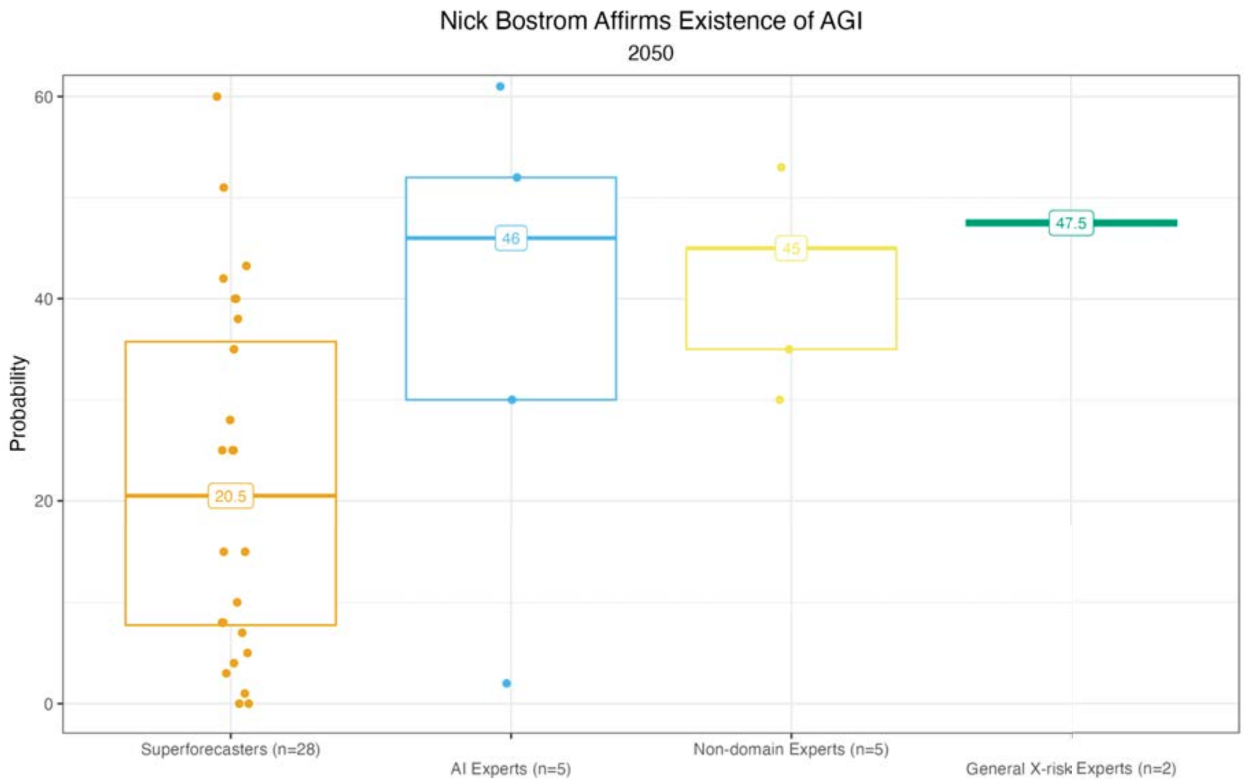


Nick Bostrom Affirms Existence of AGI - 2050

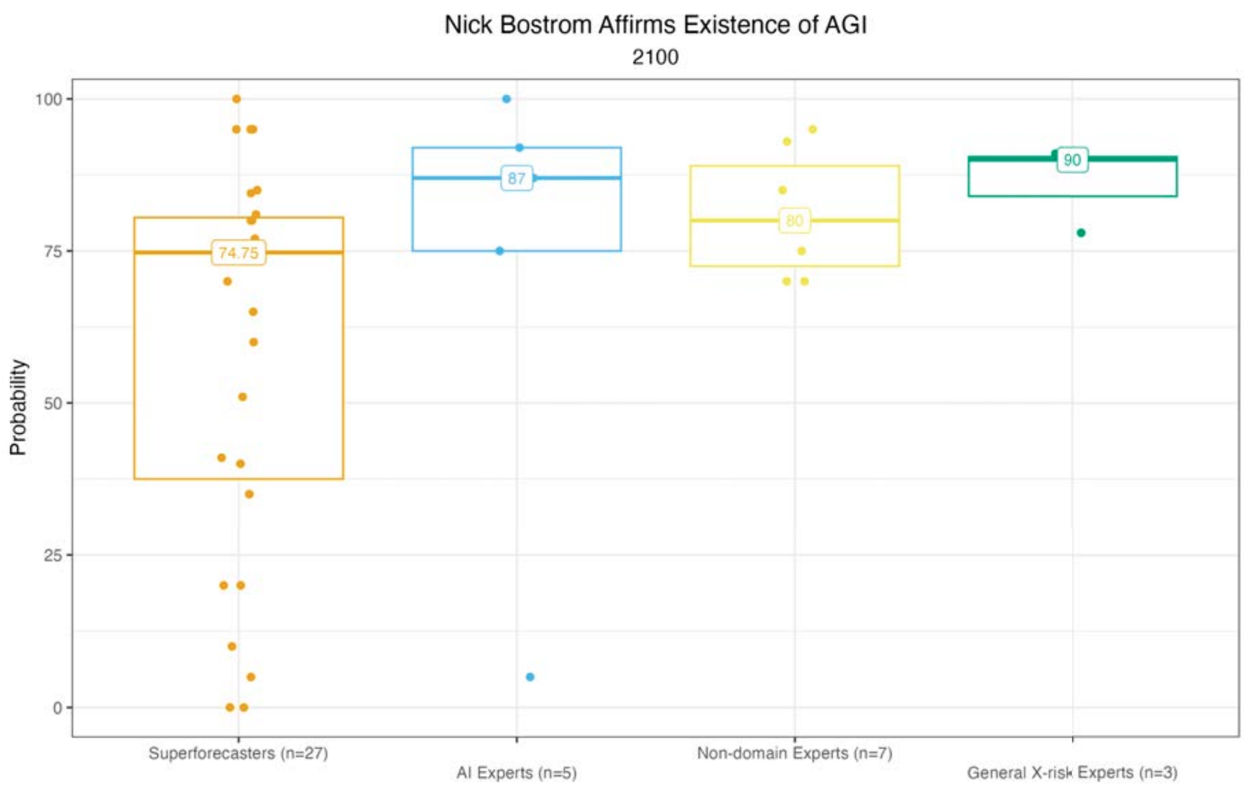



Nick Bostrom Affirms Existence of AGI - 2100
Median
90
85
80
75
70
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 28)
Experts (n = 15)


Nick Bostrom Affirms Existence of AGI
2030
Probability
15
10
5
0
1
9
3
10
Superforecasters (n=32)
AI Experts (n=5)
Non-domain Experts (n=7)
General X-risk Experts (n=3)



Nick Bostrom Affirms Existence of AGI
2050
Probability
60
40
20
0
20.5
46
45
47.5
Superforecasters (n=28)
AI Experts (n=5)
Non-domain Experts (n=5)
General X-risk Experts (n=2)



Nick Bostrom Affirms Existence of AGI
2100
Probability
100
75
50
25
0
74.75
87
80
90
Superforecasters (n=27)
AI Experts (n=5)
Non-domain Experts (n=7)
General X-risk Experts (n=3)

Sources of agreement, disagreement and uncertainty

Disagreements for this question centered around:

1. Whether and when AGI is likely to be developed[1460]
    a. The trajectory of AGI development
    b. External factors affecting the speed of AGI development
    c. When others think AGI is likely to be developed
2. When Bostrom is likely to affirm AGI exists in relation to the date of its actual existence[1461]

Arguments given for forecasts of ≤4% (2030), ≤25% (2050), ≤53% (2100)

On 1a (AGI timelines - current trajectory):

- Current AI is less advanced than people think.[1462]
- For many problems there isn't currently enough training data to imitate human behavior.[1463]
- The scaling hypothesis may not hold.[1464]
    - Scaling up LLMs might not be sufficient for AGI.[1465]
    - There will be diminishing marginal returns such that the final 5% of the problem will be much harder than the first 95%.[1466]

[1460] 341, "There are wide disagreements over the progress of artificial intelligence toward AGI. Is AGI a natural next step from recent progress or is it not likely to be obtained for hundreds of years due to human qualities of intelligence which won't be duplicated." 338, "Sources of uncertainty: Beliefs about requirements to reach AGI - is it just more compute on larger models or are there numerous, difficult algorithmic or other issues?" 344, "Main disagreements of the team steam from the debate on when general specific AI would occur."

[1461] 338, "Another source of uncertainty is the subjective nature of the question - as there was no definition provided for AGI it has the potential to resolve based on Prof Bostrom's subjective interpretation."

[1462] 341, "In a recent blog post, roboticist and entrepreneur Rodney Brooks said he thinks the field of AI is probably "a few hundred years" less advanced than most people think. "We're still back in phlogiston land, not having yet figured out the elements." Incorrectly labeled as an argument for higher forecasts.

[1463] 341, "For most problems there won't be training data, most human behavior is not yet digitized." Incorrectly given as an argument for higher forecasts. See also 340, "The tasks it completes are great, but only conducted in areas where there are large data sets of reliable information and correlations can be drawn."

[1464] 336, "Not everyone agrees that the 'computational' method (adding hardware, refining algorithms, improving AI models) will in itself be enough to create AGI or something sufficiently similar. They expect it to be a lot more complicated (though not impossible). In that case, it will require a lot more research, and not only in the field of computing." See also 340, "The current path does not necessarily lead to AGI by just adding more computational power."

[1465] 341, "Other notable AI pessimists like Gary Marcus have also been vocal lately as to the insufficiency of just scaling up the large language models (LLMs). Yann Le Cun recently also made comments that LLMs will not be enough." Incorrectly given as an argument for higher forecasts.

[1466] 343, "[One forecaster's] belief is that we'll see AI competently surpassing human-level abilities in a number of areas long before we get to the full vision of what we think of when we talk about AGI, and that the last 5% of the problem will be much more difficult than the first 95% of the problem."

    - There may be unknown theoretical impossibilities for creating AGI.[1467]
- AI progress may lead to systems of smaller components integrated into normal software, rather than AGI.[1468]
- AI systems might get into infinite recursion problems and use up vast computing power.[1469]
- The history of software development suggests progress will be slow and gradual.[1470]
- There may be another AI winter.[1471]

On 1b (AGI timelines - external factors):
- Human extinction may prevent the development of AGI.[1472]
- Scientific progress may slow.[1473]
- We may be living in a simulation which wants to block us from ever creating AGI.[1474]

On 1c (AGI timelines - opinions of others):
- Predictions about the future tend to be optimistic.[1475]
- Previous claims about imminent AGI have been proved false.[1476]

Arguments given for forecasts of ≥4% (2030), ≥25% (2050), ≥50% (2100)

On 1a (AGI timelines - current trajectory):

[1467] 343, “Perhaps X-Risks, slow scientific progress, or the unknown theoretical impossibilities in the way of AGI creation or Nick Bostrom Convincing might prove such an event highly unlikely over the long term.”
[1468] 336, “However, it is not clear if this will lead to integrated systems of AI, instead of smaller components that will be integrated in 'normal' software, like what happened with OCR (once seen as a sign of AI, now 'just' an add-on in your photo-app).”
[1469] 341, “There could be problems of getting into infinite loops and using up vast computing power.” Incorrectly given as an argument for higher forecasts.
[1470] 336, “Although in principle nothing prevents the formation of some forms of superintelligence, looking at the history and current practices of software development shows that any progress likely will be slow and gradual, occasionally punctuated with a new method or technology that allows for a quick 'sprint' in some domains.”
[1471] 336, “This increases the risk for yet another 'AI winter' (previous episodes were from 1974–1980 and from 1987–1993), at least delaying the possibility of resolving this question with a 'yes'.”
[1472] 343, “Perhaps X-Risks, slow scientific progress, or the unknown theoretical impossibilities in the way of AGI creation or Nick Bostrom Convincing might prove such an event highly unlikely over the long term.” See also 344, “The most likely way for this to not resolve positively by 2050 or later is if there is some non-AI catastrophe of enough magnitude that it set backs humankind by years or decades.” Given in an argument for forecasts of 55% (2030), 100% (2050) and 100% (2100).
[1473] 343, “Perhaps X-Risks, slow scientific progress, or the unknown theoretical impossibilities in the way of AGI creation or Nick Bostrom Convincing might prove such an event highly unlikely over the long term.”
[1474] 336, “An -indirect- argument from Bostrom himself: We might be living in a simulation, possibly set up and directed by AI. That system might want to block us from achieving 'full' AI.”
[1475] 337, “Yet, taking into account that more often than not predictions of the future tend to be wildly optimistic, the probabilities at the lower end are probably worth taking into consideration.” Given as an argument for forecasts of close to 0% (2030), 15-40% (2050) and the order of 70% (2100).
[1476] 336, “There have been previous bold claims on impending AGI (Kurzweil for example) that didn't pan out.”

- Recent progress is impressive.[1477]
- Recent progress has been faster than expected.[1478]
- There have been recent advances in quantum computing which may speed up AGI development.[1479]

On 1c (AGI timelines - opinions of others):
- Expert surveys have predicted AGI this century.[1480]
- Other forecasts on similar questions have moved forward recently.[1481]
- One team cited Holden Karnofsky's forecast as an argument for higher forecasts.[1482]

On 2 (Bostrom):

[1477] 341, "In the last five years, the field of AI has made major progress in almost all its standard sub-areas, including vision, speech recognition and generation, natural language processing (understanding and generation), image and video generation, multi-agent systems, planning, decision-making, and integration of vision and motor control for robotics. In addition, breakthrough applications emerged in a variety of domains including games, medical diagnosis, logistics systems, autonomous driving, language translation, and interactive personal assistance.
AI progress has been quick in specific domains.
Language systems are developing the capability to learn with increasing resources and model parameters. Neural network models such as GPT learn about how words are used in context, and can generate human-like text, including poems and fiction.
Image processing technology has also made huge progress for self-driving cars and facial recognition, and even generating realistic images.
Agile robots are being developed using deep-learning and improved vision.
Tools now exist for medical diagnosis.
Deep-learning models partially automate lending decisions and credit scoring." See also 336, "There's no doubt among forecasters that Machine Learning and Artificial Intelligence have developed tremendously and will continue to do so in the foreseeable future."
[1478] 341, "Both the text creating models and the image generating models are achieving results that are more impressive than most had imagined at this stage." Incorrectly tagged as an argument for lower forecasts.
[1479] 341, "Also, "quantum computing made significant inroads in 2020, including the Jiuzhang computer's achievement of quantum supremacy. This carries significance for AI, since quantum computing has the potential to supercharge AI applications". Incorrectly tagged as an argument for lower forecasts.
[1480] 337, "In general the median forecast is very much anchored by the meta base rate provided by different experts' opinions on the arrival of AGI."
[1481] 341, "Ajeya Cotra, who wrote the biological anchors report from OpenPhil, recently published a follow up post saying that she had moved her timelines forward, to closer to now.
The forecasts for AGI on Metaculus have all moved nearer to our time recently as we have seen the string of astounding AI models, including GPT-3, LaMDA, Dall-E 2 and Stable Diffusion…Metaculus has human parity by 2040 at 60% and AGI by 2042. The latter has dropped by 15 years in recent months with the advent of all the recent new models…AI Impacts survey on timelines has a median with HLMI (Human Level Machine Intelligence) in 37 years, i.e. 2059. That timeline has become about eight years shorter in the six years since 2016, when the aggregate prediction put 50% probability at 2061, i.e. 45 years out." Incorrectly tagged as an argument for lower forecasts.
[1482] 341, "Karnofsky says >10% by 2036, ~50% by 2060, 67% by 2100. These are very thoughtful numbers, as his summary uses several different approaches. He has experts as 20% by 2036, 50% by 2060 and 70% by 2100, biological anchors >10% by 2036, ~50% by 2055, 80% by 2100, and semi-informative priors at 8% by 2036, 13% by 2060, 20% by 2100." Incorrectly tagged as an argument for lower forecasts.

- Bostrom might affirm the existence of AGI before most people do.[1483]

Other arguments:
- The definition of AGI is imprecise. The line between narrow and general intelligence might become blurred as we get more multi-purpose systems.[1484]

Other arguments given

Arguments in favor of lower forecasts:
- On 1b (AGI timelines - external factors): wars might interrupt AGI development.[1485]
- On 2 (Bostrom): there might be a delay between AGI being developed and Bostrom being assured enough to affirm it (which shortens the time window until 2030).[1486]

Arguments in favor of higher forecasts:
- On 1a (AGI timelines - current trajectory):
  - The scaling hypothesis might hold.[1487]
- On 1b (AGI timelines - external factors):
  - Wars might speed up AGI development.[1488]
- On 2 (Bostrom):
  - Bostrom has the skill to create a philosophical or theoretical framework which defines AGI in a surprising way.[1489]
  - Bostrom may be biased towards affirming AGI as he reaches the end of his life.[1490]

[1483] 337, "[A] slight correction from the idea that Bostrom might affirm the existence of AGI a bit before most people do."
[1484] 337, "[T]he lack of precise definition of AGI (in how many domains does an AI have to reach a human-like level of performance? what are the most relevant domains? etc.). In this later case, as we will eventually have multiple-purpose AIs, the line between narrow and general AI will become really blurred and it might be perfectly acceptable for Bostrom to use the term AGI to refer to them."
[1485] 339, "[W]ars might have erupted halting progress - but this reasoning could also be used for accelerating development (arms race dynamics)."
[1486] 339, "Also for 2030 affirmation, AGI would need to be confirmed quickly by Bostrom - which I could see would involve a year of testing the limitations of whatever was created - so this shortens the time window."
[1487] 336, "The majority of our team might drastically underestimate the near-future advances of AI. It is likely that there are architectural/algorithmic bottlenecks that cannot (efficiently/practically) be overcome by compute and data scale, but that leaves a fair amount of probability for the alternative. Not all bottlenecks need to be overcome for this question to resolve with a 'yes' somewhere this century."
[1488] 339, "[W]ars might have erupted halting progress - but this reasoning could also be used for accelerating development (arms race dynamics)".
[1489] 343, "Bostrom has the expertise and vigilance to suggest a philosophical or theoretical framework which might surprise us in it's inclusion of something as an AGI."
[1490] 338, "Human element that Bostrom may be biased to declare existence of AGI as he reaches the end of his life span." 340, "The highest probability of resolution lies with Nick Bostrom stating that AGI has been achieved based on a poor definition of AGI or even crediting high level machine intelligence as AGI. This could take place at any point, but especially as he moves closer to the end of his career and as technology makes dramatic, if not specific, advances."

    - Consciousness is not required for Bostrom’s concept of AGI.[1491]
- Other arguments:
    - There might be a false alarm about AGI existing.[1492]

Cross-references with other questions

Q44: Date of Advanced AI

Question 52: Probability of GDP Growth Over 15%

By 2100, will the global real GDP in a year ever exceed 115% of the highest GDP reported in any full prior year?

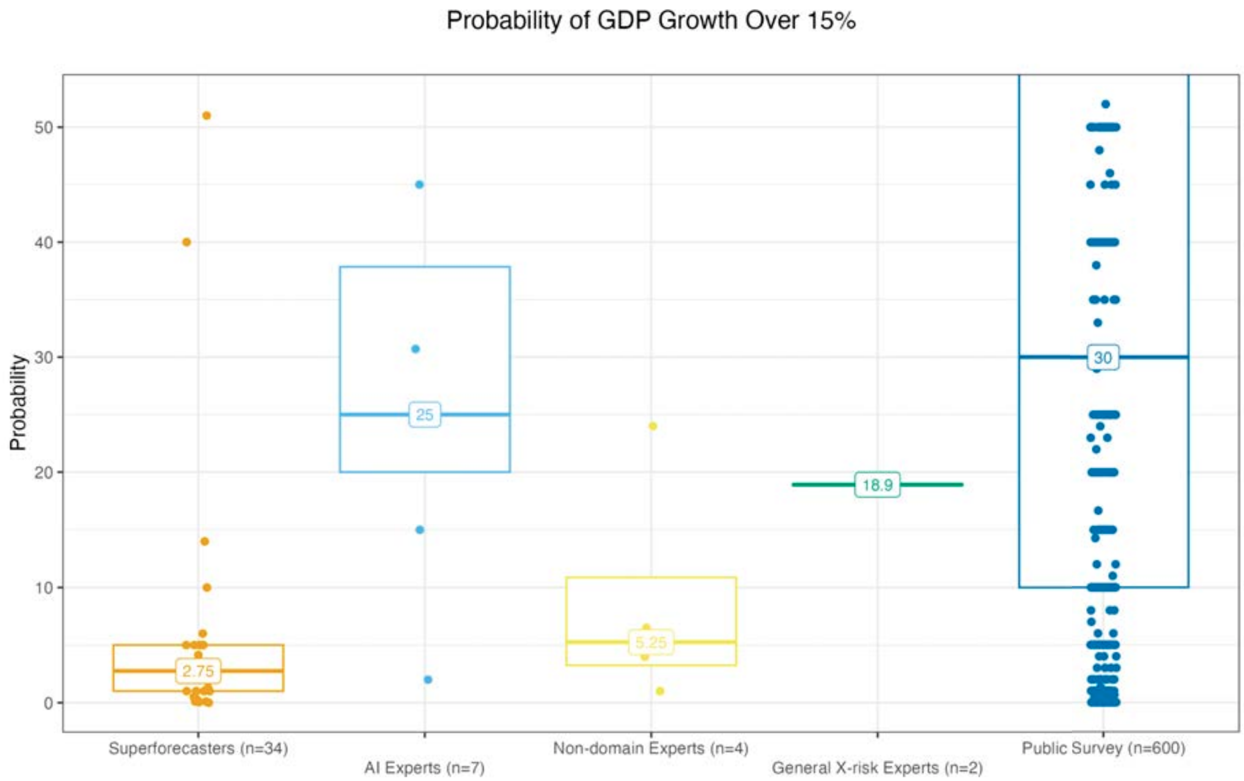
Question and resolution details, prior forecasts, and other relevant sources

Results[1493]

Group	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 34)	5%	2.75%	34.82	-54.17%
Domain Experts (N = 7)	25%	25%	21.73	-18.37%
General X-Risk Experts (N = 2)	49.8%	18.9%	28.57	-99.5%

[1491] 336, “Bostrom sidesteps the question about consciousness, he is okay with AI systems that are 'philosophical zombies'. (some of Bostrom's colleagues (i.e. Daniel Dennett) would argue that such a distinction is meaningless). Even though some forecasters doubt we can create anything other than such a memic form of ourselves, they agree that it is likely enough to satisfy this question.”
[1492] 339, “AGI announcement false alarm (eg before finding niches in which it's intelligence does not generalise).”
[1493] Numbers of forecasters are given as of Stage 4 of the XPT.

Non-Domain Experts (N = 4)	5.5%	5.25%	29.02	-64.4%
Public Survey (N = 600)	30%		408259.24	-



Sources of agreement, disagreement and uncertainty

There was broad agreement that:

- The base rate for global growth over 15% is ~0.[1494]
- If >15% growth happened it would be because of AI.[1495]

Disagreements were primarily driven by:

1. How to interpret relevant historical growth rates.
2. Whether humanity will go extinct within the timeframe.
3. Whether sufficiently advanced AI will be developed within the timeframe.
4. Whether advanced AI will lead to GDP growth of >15%.

[1494] Explicit references made by 341, 336, 338, 340.
[1495] All of 341, all of 343, 336 ‘in general,, all but one of 337. All teams discussed AI in their rationales.

Arguments given for forecasts below 5%

On 1 (historical growth rates):

- Scientific, economic and population growth are slowing.[1496] One team cited Cowen and Southwood, [Is the Rate of Scientific Progress Slowing Down?]

On 3 (probability of advanced AI):

- Sufficiently advanced AI will not be developed within the timeframe.[1497]
- Advanced AI might never automate certain tasks, which would then bottleneck growth.[1498]

On 4 (relationship between advanced AI and growth):

- So far, the impact of AI on GDP has not been strongly positive.[1499]
- AI is a process innovation, not a product innovation, and so won't boost economic growth that much.[1500]
- Some combination of supply chains, infrastructure, and changes in consumer wants/needs will take longer than a year to adapt.[1501]
- There are constraints on the large-scale deployment of AI systems in particular, like energy, memory, physical resources, compute or labor.[1502]

[1496] 344, "The slowdown of scientific and economic growth has been identified by many modern economists," citing Cowen and Southwood. 338, "Ideas are becoming harder to find. If this trend continues, perhaps it will prevent AIs finding ideas quickly enough to drive explosive growth." 344, "Another reason against expecting high GDP growth is that the population growth of the planet is slowing down."

[1497] 338, "There are doubts if AI explosion will occur". 336: "AI is highly unlikely to advance far enough within the timeframe of this question to have this level of impact on GDP." 340: "The low probability forecasts cover scenarios in which there are no major technological leaps (e.g., AI does not progress much)."

[1498] 338: "Essential tasks that advanced AI never automates, and these will bottleneck the growth process."

[1499] 340: "For example, Google and Facebook are two of the largest global companies, have transformed how people live, yet in GDP terms their main impact has been reallocating advertising spend away from newspapers and yellow pages (to themselves). In AI terms any replacement of human labour may benefit corporate profits more than revenues or GDP (and this is what has occurred so far)." 337, "[T]he pessimists considered that the economic effect of AI was liable to arrive piecemeal, as it has done so far."

[1500] 340, "AI/Robotics etc. are process innovations and may become widespread without boosting economic growth."

[1501] 336, "There may be rapid advancements in AI, but GDP increase will be limited by an inability of the supply chain to support such quick growth and/or a delay in consumer wants/needs that would result in multiple annual increases of <15%." 339, "it would take a new technology that is not only revolutionary, but immediately widespread in terms of adoption, manufacturing, material supply chain, etc." 340: "Gains from product innovations are slow to realise because using them requires infrastructure and they are built over time and at different rates in different places."

[1502] 336, "One of the problems with large scale copying of AI is that energy and memory constraints are often neglected. Existing ML algorithms can be energy hogs - a lot of energy, data, and memory is needed in training but even when trained and running inference they can be a large energy consumer, which limits what can be deployed. Seems to me that high resource use would impact the potential for AI to result in a large step change in production." 339, "[T]here are still real constraints on fast economic growth (both in terms of physical resources, compute, and labor supply)."

- Humans will choose to deploy AI slowly. Reasons this might happen, some of which were cited in arguments for other levels of forecasts: trust, concern about sustainability, vested interests, and conservatism.[1503]
- AI will lead to huge increases in intelligence and labor, but those are not the key bottlenecks to growth.
    - Other possible bottlenecks: regulation,[1504] physical resources like energy and materials,[1505] physical limits on time needed for experiments and manufacture,[1506] fundamental limits to technology,[1507] other unanticipated or poorly understood factors.[1508]
- The impact of advanced AI on GDP could be negative.[1509]

Arguments given for forecasts above 20%

On 1 (historical growth rates):

- Historical growth is a power law not an exponential, so we should expect explosive growth.[1510] Teams mentioned Roodman's Modeling the Human Trajectory, this criticism of Roodman's report, and Davidson's Could Advanced AI Drive Explosive Economic Growth?

[1503] 337: "One member of the optimist group was more sanguine about alignment but felt that a combination of vested interests, red-tape, and small-c conservatism would prevent the full impact of the AGI advent to be felt within a short period of time, reducing the likelihood from what would otherwise much more certain." 338: "Humanity will choose to grow slowly and sustainably, even if AI gives us the ability to grow much faster." 343: "It's not sufficient to have an aligned AGI, we also need to have built sufficient trust in that AGI to grant it broad influence over humanity. This trust-building may 1.) not be possible, and/or 2.) take a long time."

[1504] 343, "Other candidate bottlenecks (which could also interact) include energy, materials, or regulations."

[1505] 343, "Other candidate bottlenecks (which could also interact) include energy, materials, or regulations." 338: "Accumulation of physical or human capital has been the most important driver of historical growth, and advanced AI will not significantly accelerate this process."

[1506] 343, "AGI would also still be constrained by the limits of the possible, by the need to perform real experiments to test hypotheses, and by the tools at hand, somewhat blunting the gains from intelligence."

[1507] 338: "Fundamental limits to how good our technology can become, and we will approach these limits before explosive growth occurs."

[1508] 338: "Unanticipated bottleneck will slow down growth"; "Understanding of the determinants of growth is very poor, and the true determinants simply will not lead to explosive growth regardless of the AI systems developed."

[1509] 340, "There were also some comments that AI/digital technologies are at risk of depressing GDP as they replace human work and have no economic multiplier (which requires product innovation to find new work for humans). Inequality of this sort tends to be a material drag on growth (compared to abstracted equations) through different rates of marginal consumption (the wealthy save the poor spend)."

[1510] 339, "[T]he Open Philanthropy blog post convinced me that past human growth might not be exponential but a power law, so I'll raise to 55%," citing Davidson; 340, "The contrarian view drew on this blog post about economic growth…The paper suggests that historic economic growth fits a power law structure rather than an exponential curve and that growth is driven by idea volume," citing Roodman. 340, "The contrarian view drew on this blog post about economic growth…The paper suggests that historic economic growth fits a power law structure rather than an exponential curve and that growth is driven by idea volume," citing this criticism of Roodman's report.

- If you go far enough back in human history, economic growth has in fact accelerated dramatically.[1511]

On 4 (relationship between advanced AI and growth):
- AI will recursively self-improve, which could lead to very fast productivity gains.[1512]
- There will be a massive increase in labor, through AI-enabled robots or ems or automation.[1513]

Other arguments given

On 2 (probability of extinction):
- Humans might go extinct because of unaligned advanced AI (discussed by a team with a median forecast of 20%).[1514]

On 4 (relationship between advanced AI and growth):
- Some factors could hold growth back and create latent potential, like a lockdown (given as a counterargument to an argument for forecasts lower than 10%).[1515]
- Advanced AI could be unaligned such that most production is under the control of AI systems rather than humans (given as a counterargument to an argument for forecasts higher than 10%).[1516]

Cross-references with other questions

On 2 (probability of extinction):

[1511] 341, "Economic growth has accelerated dramatically in human history: from perhaps one doubling of gross world product (GWP) every 10,000 to 100,000 years in the foraging era, to one doubling roughly every 1000 years in the farming era, to a doubling every 20 years or so in the industrial era."

[1512] 343: "AGI may be able to enhance itself to superintelligence, potentially resulting in large productivity gains." 336: "[R]ecursive improvement just seems so powerful, and once a sufficiently powerful system can start accumulating and using knowledge faster than a human, it seems like it would move incredibly fast."

[1513] 340, "[T]he productive potential of the economy would be dramatically increased with scalable digital and robotic 'labour' adding to the capacity of replaced 'human' labour." 343, "AGI-equipped robots would be mass-producible, unlike humans, resulting in a potential massive increase in the effective labour force." 336, "Transformative AI could allow for digital workers that could dramatically and suddenly expand the size of the economy."

[1514] 343, "Transformative AGI is sufficiently aligned with human interest to work on our behalf and does not cause, say, civilization collapse or extinction."

[1515] 337, "[T]he pessimists considered that the economic effect of AI was liable to arrive piecemeal, as it has done so far, and thus unlikely to provide the requisite leap unless coinciding with a set of conditions that held back economic growth while creating latent potential, such as (for example) the advent of cheap fusion energy during an extended pandemic lockdown that lifts globally within relatively short order as vaccines are rolled out."

[1516] 337, "The main consideration for this group was not whether or not AGI would arrive, or whether it would be transformative, but rather whether this transformation would qualify for the purposes of the question, such as if the AI was rogue or otherwise unaligned." 343, "Transformative AGI is sufficiently aligned with human interest to work on our behalf and does not cause, say, civilization collapse or extinction."

- 4. Probability that artificial intelligence will cause human extinction or reduce the global population below 5,000 by the end of 2100: x%.
- 10. Overall probability of human extinction or a reduction in the global population below 5,000 by the end of 2100: x%.
- 11. Year by which humans will go extinct or first have a population less than 5,000: [year].

On 3 (probability of advanced AI):
- 44. Year by which the first unified AI system meeting various criteria of advancement will be trained, tested, and publicly known of: [year].
- 51. Probability that Nick Bostrom believes artificial general intelligence (AGI) exists by 2100: x%.

Question 53: Year of GDP Growth Over 15%

Question 54: Loss of Agricultural Production

What is the probability that total worldwide production of primary crops will fall by at least 70% within any three-year period…
…by 2030?
…by 2050?
…by 2100?

Question and resolution details, prior forecasts, and other relevant sources

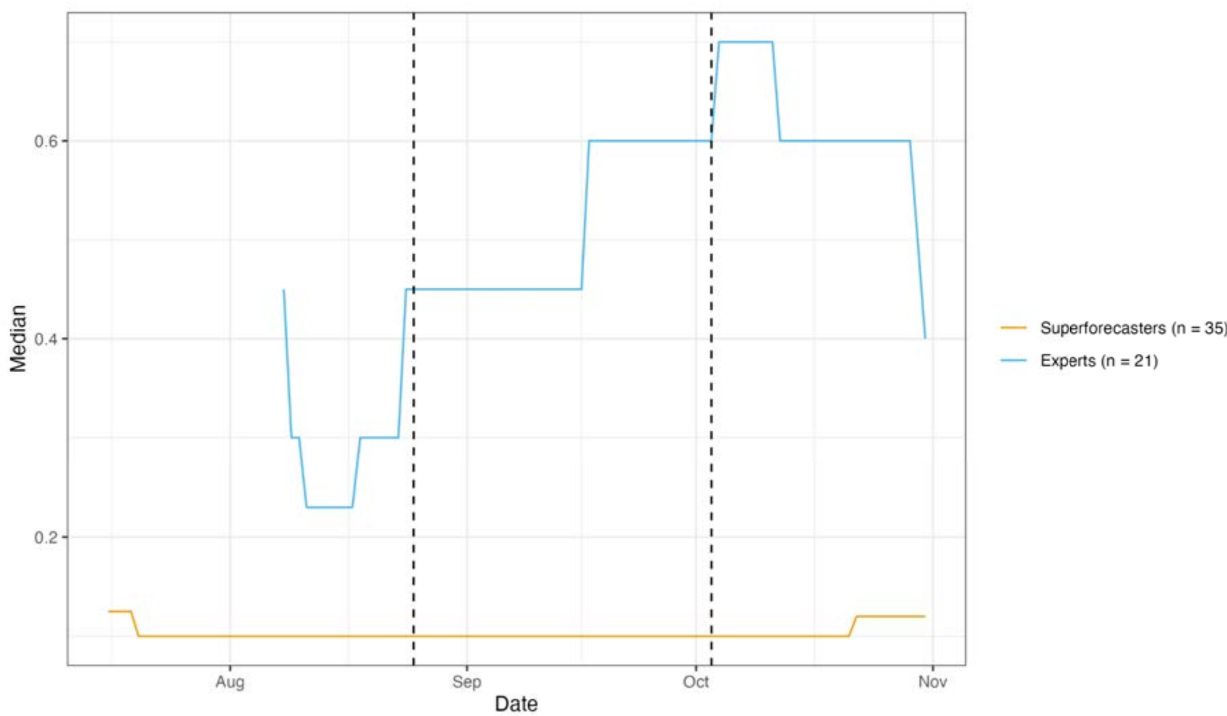
Results[1517]

Group	**Year**	**Stage 1 Median**	**Stage 4 Median**	**Stage 1 Standard Deviation**	**Percent Change in Standard Deviation, Stage 1 to Stage 4**
Super-	2030	0.1%	0.12%	6.10	-51.39%

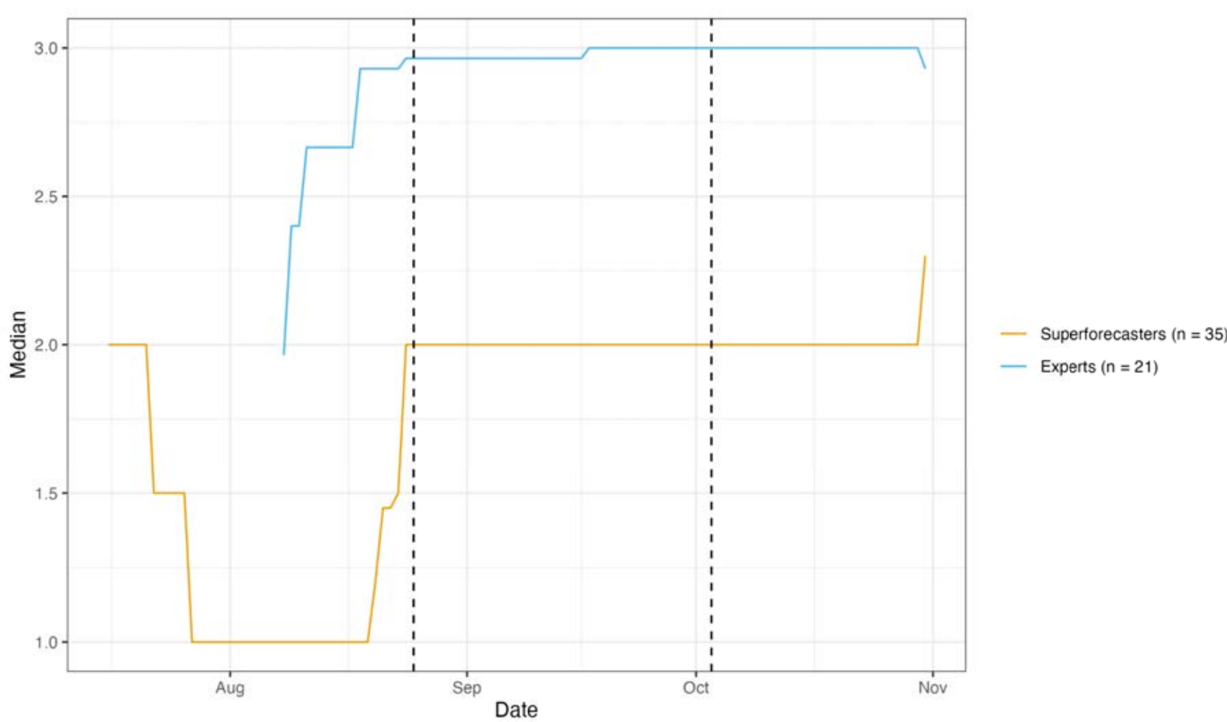
[1517] Numbers of forecasters are given as of Stage 4 of the XPT.

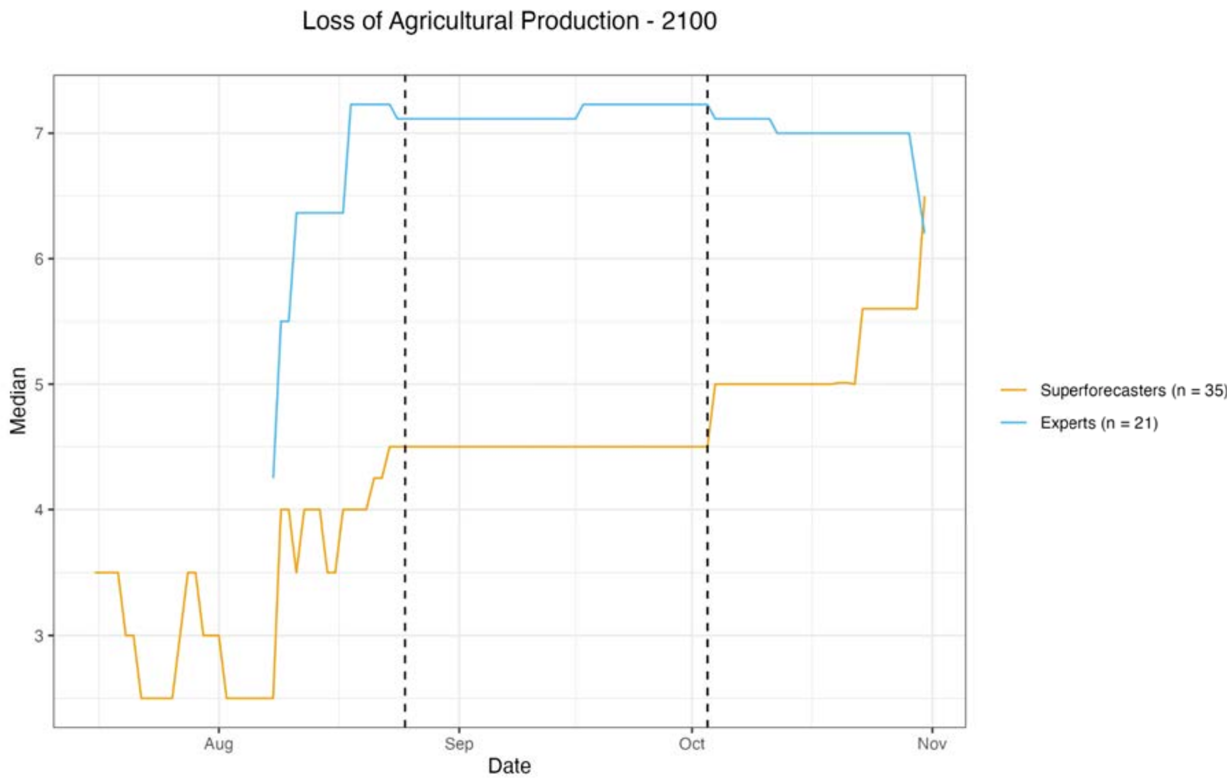
Forecasters (N = 35)	2050	1%	2.3%	9.58	-48.6%
	2100	2.5%	6.5%	9.87	-28.1%
Experts (N = 21)	2030	0.45%	0.4%	12.51	-31.02%
	2050	1.97%	2.93%	23.23	-31.06%
	2100	4.25%	6.2%	25.67%	-24.88%
General X-Risk Experts (N = 7)	2030	1.1%	0.87%	19.49	-23.67%
	2050	3.5%	3%	35.85	-23.48%
	2100	9.62%	7.23%	38.97	-22.61%

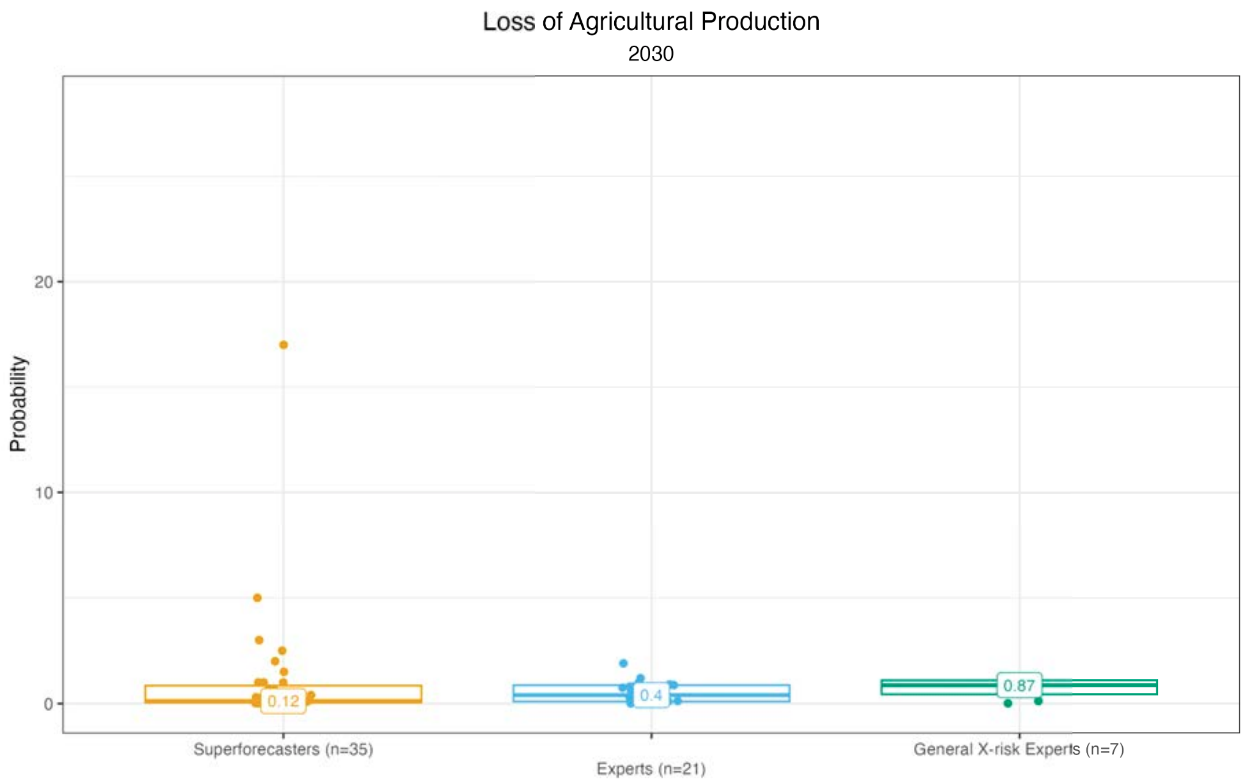
Loss of Agricultural Production - 2030

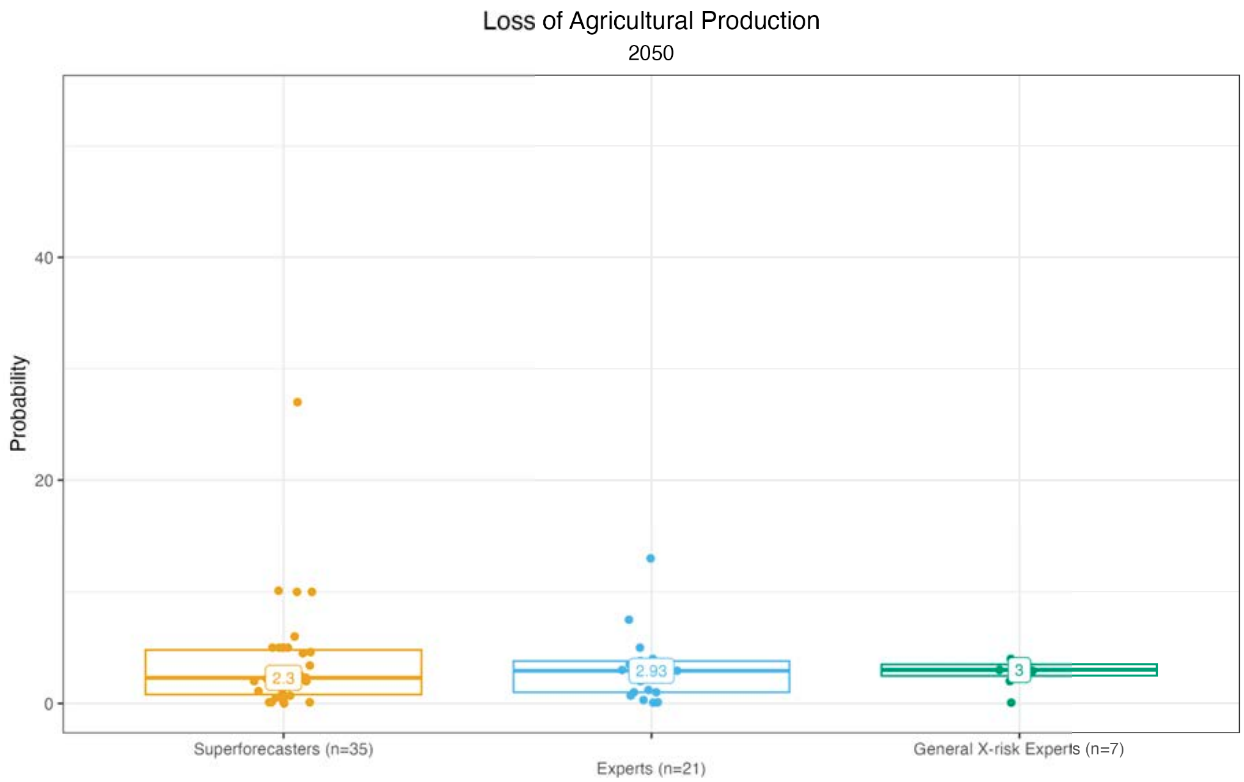


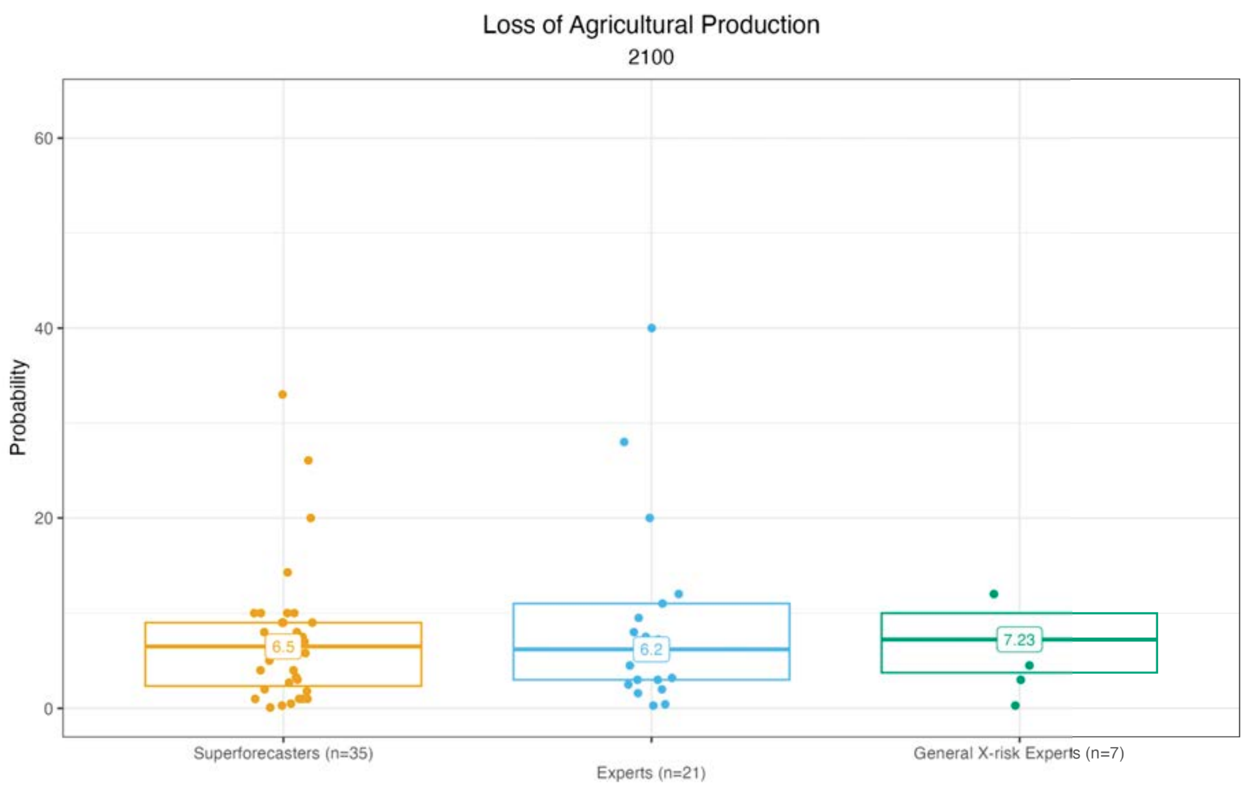
Loss of Agricultural Production - 2050




Loss of Agricultural Production - 2100
Median
7
6
5
4
3
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 35)
Experts (n = 21)


Loss of Agricultural Production
2030
Probability
20
10
0
0.12
0.4
0.87
Superforecasters (n=35)
Experts (n=21)
General X-risk Experts (n=7)


Loss of Agricultural Production
2050
Probability
0
20
40
2.3
2.93
3
Superforecasters (n=35)
Experts (n=21)
General X-risk Experts (n=7)


Loss of Agricultural Production
2100
Probability
0
20
40
60
6.5
6.2
7.23
Superforecasters (n=35)
Experts (n=21)
General X-risk Experts (n=7)

Sources of agreement, disagreement and uncertainty

There was widespread though not unanimous agreement that nuclear war was the most likely cause of total worldwide production of primary crops falling by at least 70% within any three-year period. Also, teams with both high and low forecasts on this question mentioned that their forecasts were consistent with their forecasts for GCRs and existential risks.

Major disagreements on this question focused on the amount of damage likely to be caused by nuclear war, and the likely effects of climate change.

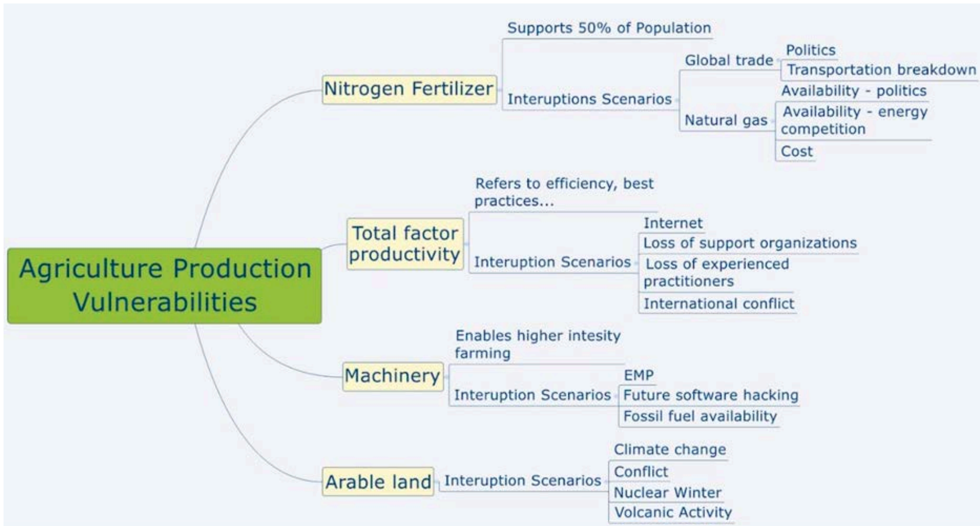
One team (338) wrote that the spread of forecasts was notably high, “given the fact that almost everyone (6 out of 8) agreed on the main cause,” which for that team was nuclear weapons, or the following nuclear winter.

Arguments given for forecasts ≤ 0.12% (2030), 2% (2050), 4.5% (2100)

- Electromagnetic pulses caused by “many high-altitude detonation of nuclear weapons around the world” are unlikely to cause disruption sufficient to cause the question to resolve.[1518]
- A solar storm is similarly unlikely to be sufficient to cause the question to resolve (see footnote 1).
- The impacts of climate change are likely to be gradual and thus not cause the question to resolve.[1519]
- Global food production is distributed widely enough to protect against most catastrophic events;[1520] therefore, “a collection of mild-moderate catastrophic events or a single very low probability major catastrophic event would be required to disrupt food production to the levels proposed in the question” (345).
- Relatedly, crop diversity increases resiliency,[1521] and “blights don't impact enough types of crops” (336).

[1518] 337: “Disruption of [sufficient] scale is very unlikely for a solar storm, and also for many high-altitude detonation of nuclear weapons around the world causing electromagnetic pulses. Furthermore, it’s not clear that agricultural production would fall 70% even in these scenarios, and as such these risk vectors were largely inconsequential relative to [others].”
[1519] 338, “Expects the impact of climate change to be gradual.”
[1520] 338, “Primary crop production is distributed worldwide, minimizing risk of a total collapse in production.”
[1521] 336, “[T]he broad range of crops and the geographic diversity of where they are grown greatly increases resilience.”

- There being no previous instance of a global crop failure of this magnitude, even during world wars,[1522] is evidence against it happening in the future.[1523] [1524] [1525]
- This level of failure would require a coordinated attack.[1526]
- A 2% probability for any year up to 2100 is consistent with other GCR estimates.[1527]
- This "map of vulnerabilities" by Team 338 seems worth reproducing:



Arguments given for forecasts ≥ 0.4% (2030), 5% (2050), 9.5% (2100)

- Large-scale nuclear war could cause this question to be resolved.[1528] Specifically mentioned were total war between the US and Russia, citing this Nature paper,[1529] and nuclear war between India and Pakistan,[1530] citing the same Nature paper.

[1522] 338, "Europe didn't see such a fall in production during the world wars"
[1523] 345, "To date, world conflicts have not resulted in this level of disruption of agricultural production; therefore, despite the possibility of major international conflicts occurring this decade, it is unlikely the international community would allow this to happen."
[1524] 337, "The strongest argument for significantly lower than 2% probability of a 70% reduction in primary crops by 2100 is probably that there is not historical precedent for it."
[1525] 338, "[P]roduction stable over the past 60 years"
[1526] 337: "To have a fall in only three years would require a coordinated attack. A weed could potentially be more general, but would typically only affect specific climates. So we generally think this is a smaller risk."
[1527] 337: "[C]onsistent with a 0.1% probability per year of full-scale nuclear war (thousands of nuclear weapons), and a 20% probability of nuclear winter given full-scale nuclear war. There would also have to be very low probability of an extreme pandemic or AGI catastrophe. Or there could be other combinations such as 2% chance of 50% mortality pandemic and negligible other risks, or 2% chance of AGI catastrophe and negligible other risks."
[1528] 342, "[L]arge scale nuclear war."
[1529] 341, "[T]he worst scenario, Russia-US nuclear war could destroy 90% of crop production. [...] The worst scenario, total war between the U.S. and Russia "...the 150 Tg soot case, global average calorie production from crops would decrease by around 90% 3–4 years after the nuclear war."
[1530] 341, "[E]ven a nuclear war limited to India vs Pakistan could kill 2 billion from famine."

- Probability of this question is highly related to catastrophic and existential risk probabilities.[1531]
- Climate change could cause this kind of crop failure.[1532] This could be directly, via “massive crop losses” (340), or indirectly, via pests[1533] (340). This could also be via resource wars (340), which could combine with “other catastrophic risks such as AI events or pandemics,” which would cause this question to be resolved.
- A failure in one region of the world would affect all other regions: “The loss of agriculture, even in just one region of the world, would increase costs and impact access to food for people everywhere” (344).
- “The worst of the global catastrophes could reduce sunlight and decrease global temperatures so significantly that crop production everywhere would suffer for years” (344).
- Catastrophes that could “take electrical grids offline,” thereby “disrupt[ing] food supply chains, shipping, and distribution management” include:
  - “coordinated cyberattack”
  - “extreme solar storm”
  - “high-altitude electromagnetic pulse (nuclear attack)” (344).
- Genetically modified crops may be especially dependent on the current CO_2 and temperature balances.[1534]
- A huge supervolcano could cause this question to be resolved.[1535]
- “Some pathogen risks” could lead to a 70% loss of crop production, “particularly from high-order effects” (344).
- An “unaligned AI apocalypse” could cause this (342).
- A “large meteorite impact” could cause this (342).

Other arguments given

Arguments in favor of lower forecasts:

- The risk from climate change is low.[1536] [1537]
- The risk from nuclear winter scenarios have been overestimated—noted by Team 343, citing this post.

Arguments in favor of higher forecasts:

[1531] 340, “The team's reasoning here mainly focused on deriving an estimate for this question from previous estimates of catastrophic and existential risk.”
[1532] 344, “Significant and unforeseen development of climate change could have a chain effect on the failure of the whole biome.”
[1533] 340, “[C]iting academic estimates focused on today's major producers and their crops. [1] [2] [3].”
[1534] 344, “Possibility of genetically-modified crops being especially dependent on the current CO2/temperature balance.”
[1535] 341, “ [A] huge supervolcano [...] could cause this question to resolve.” See also 344.
[1536] 341, “[I]t's really unlikely that climate factors will make things much less hospitable for most crops (the NASA analysis…)” [NASA analysis not linked].
[1537] 341, “[R]egarding climate change: Relooking at my forecast, I'm adding the IPCC report, P6 of summary for policy makers (SPM), shows a low base rate for a significant decline [of crops]...”

- Supply chain risk from just-in-time delivery is significant.[1538]
- A rise in consumption of lab-grown foods could make this question resolve in a way that isn't bad for humanity.[1539]

Question 55: Space Colony

What is the probability that there will be a life-sustaining colony outside of Earth's atmosphere with a population of at least 5,000 people...
…by 2030?
...by 2050?
...by 2100?

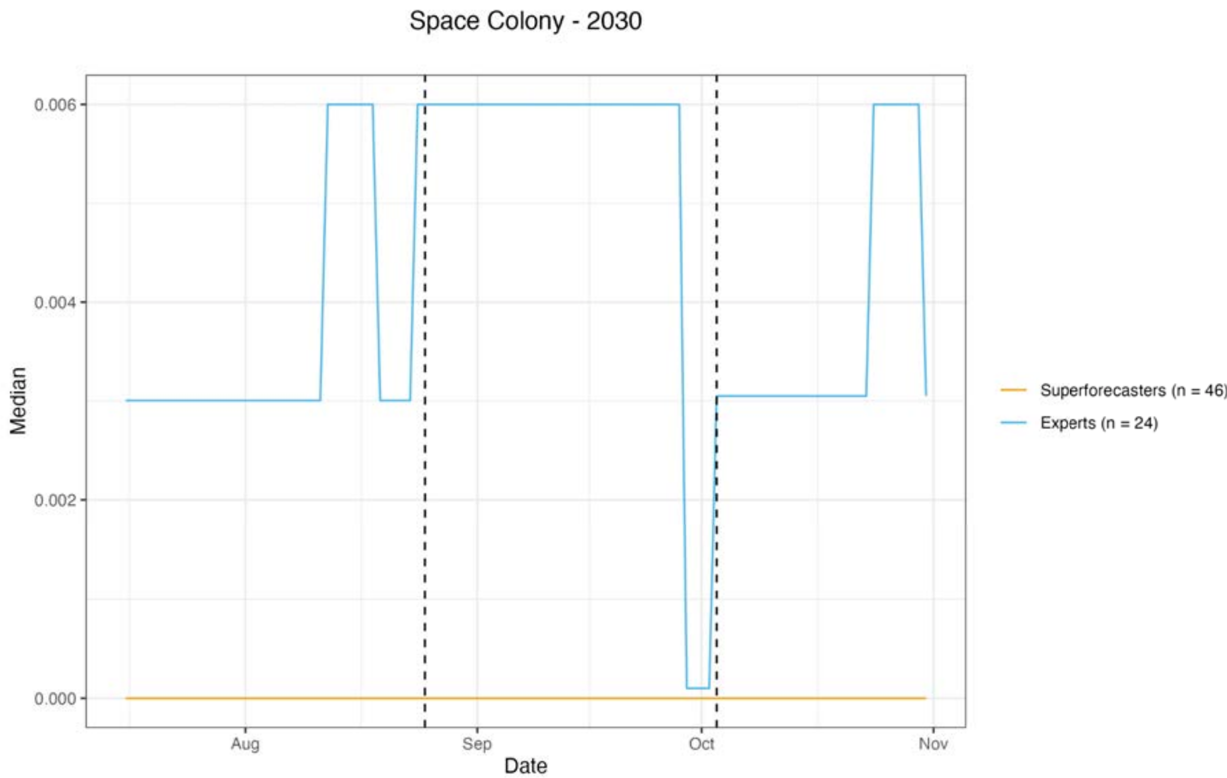
Question and resolution details, prior forecasts, and other relevant sources

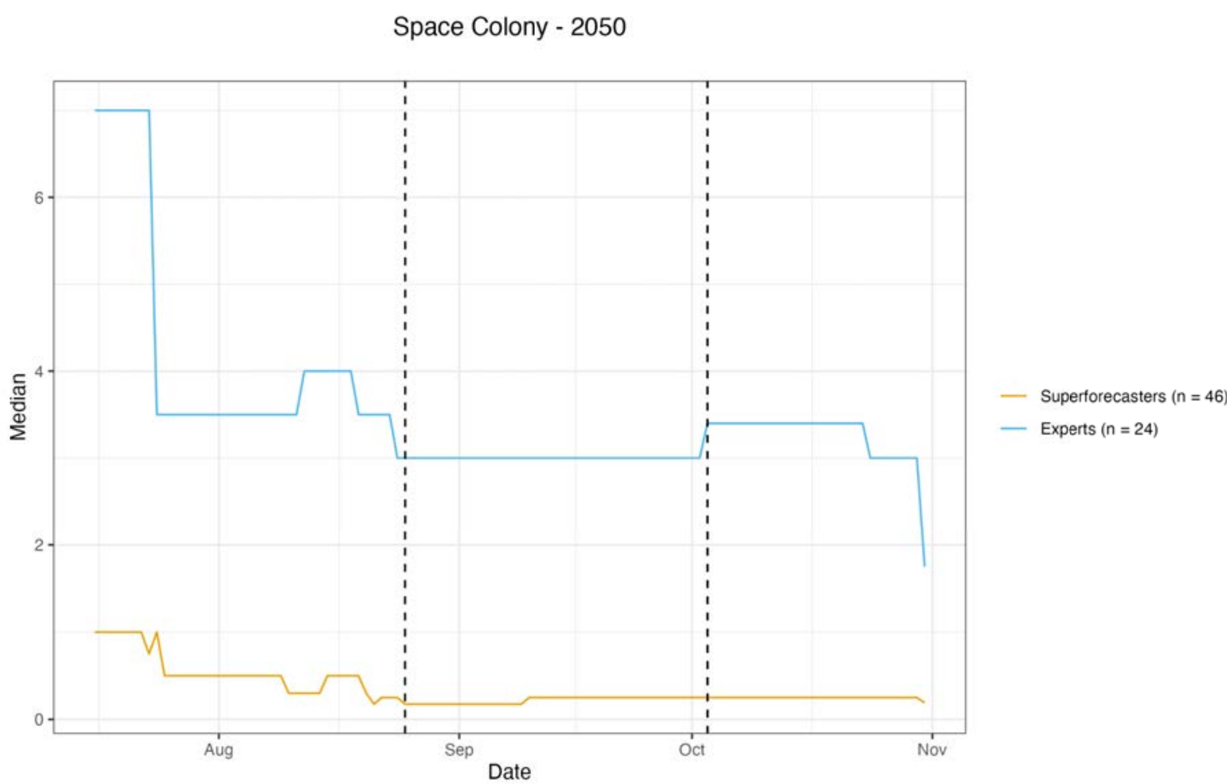
Results[1540]

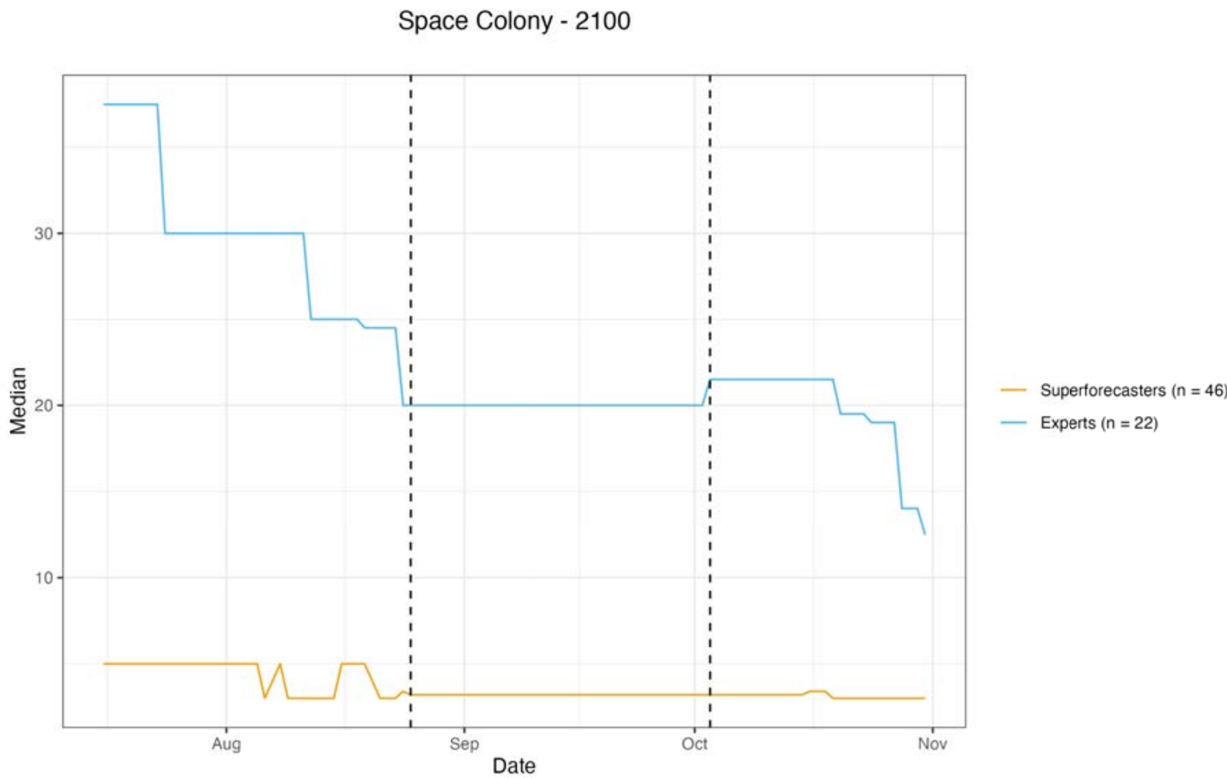
Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 46)	2030	0%	0%	0.20	+0.99%
	2050	1%	0.19%	16.21	-63.91%
	2100	5%	3%	25.41	-27.86%
	2030	0.003%	0.0031%	1.31	-13.95%

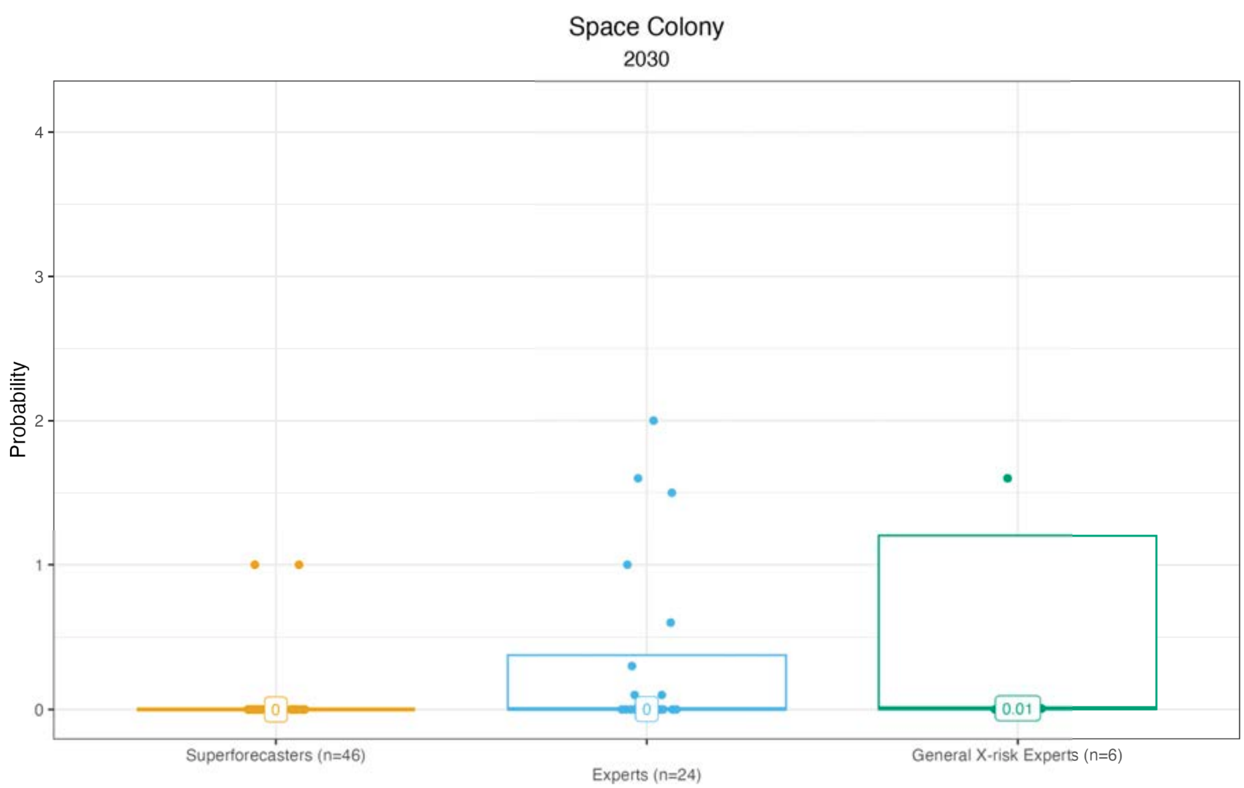
[1538] 343, "Global supply chain fragility was recently demonstrated during the COVID-19 pandemic. Continued use of JIT delivery and manufacturing means this fragility will likely remain a systemic concern for the foreseeable future."
[1539] 339: "[T]he emergence of lab grown foods that might reduce dependence on field grown crops."
[1540] Numbers of forecasters are given as of Stage 4 of the XPT.

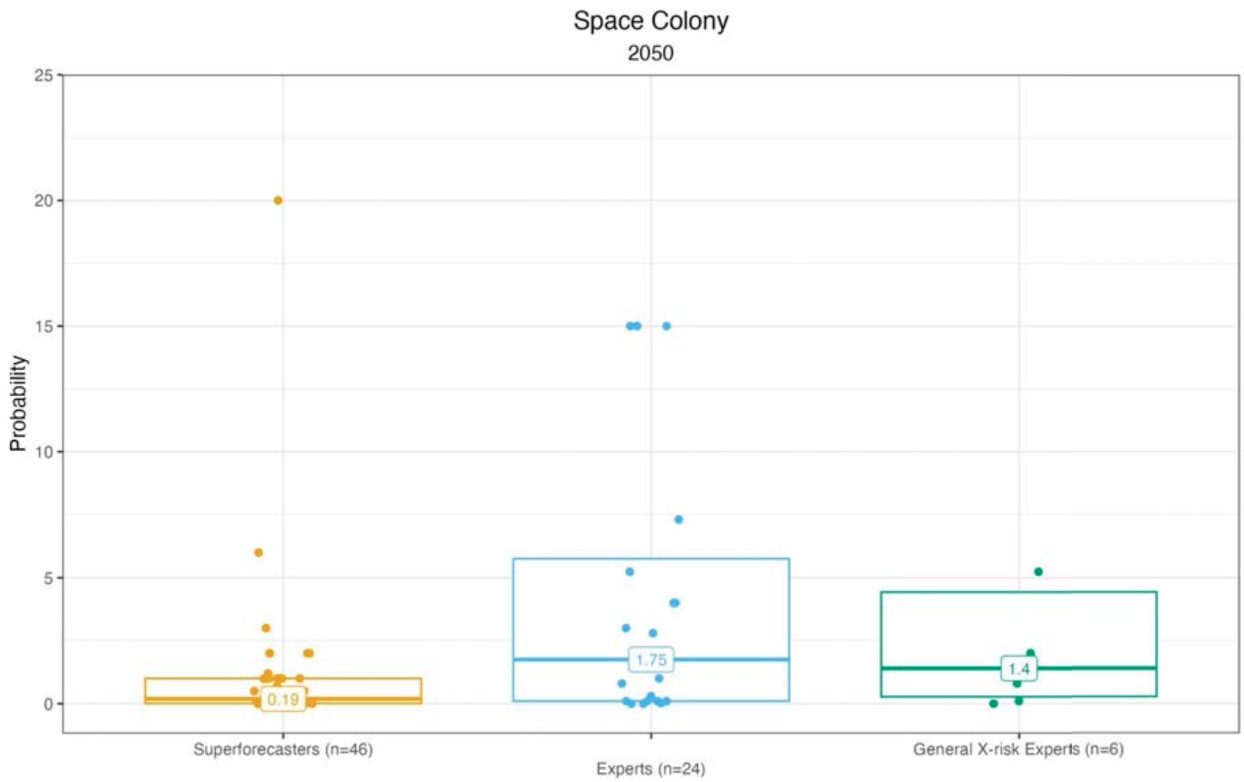
Experts (N = 24)	2050	7%	1.75%	11.02	-31.42%
	2100	37.5%	12.5%	34.27	-21.78%
General X-Risk Experts (N = 6)	2030	0.13%	0.008%	2.46	-18.17%
	2050	13.9%	1.4%	16.27	-27.83%
	2100	70%	19%	38.30	-28.74%

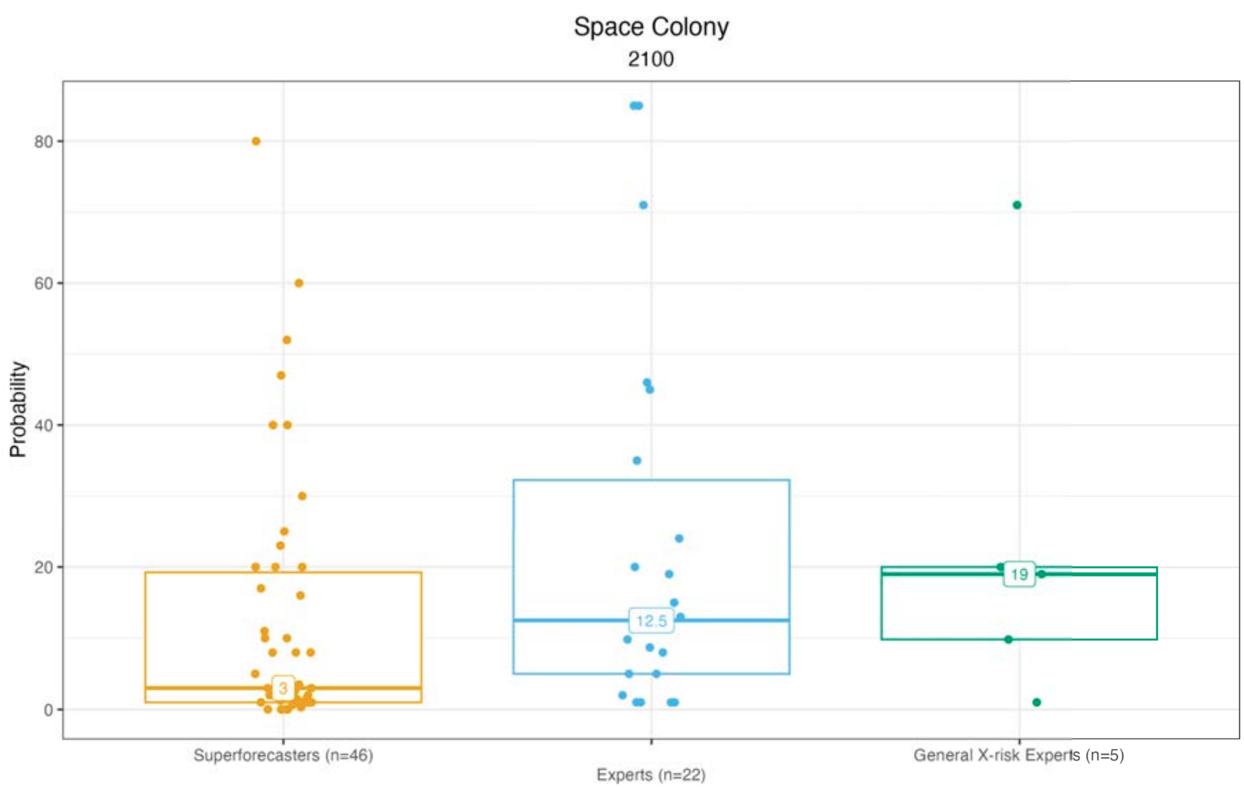

Space Colony - 2030
0.006
0.004
0.002
0.000
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 46)
Experts (n = 24)


Space Colony - 2050
6
4
2
0
Median
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 46)
Experts (n = 24)


Space Colony - 2100
Median
30
20
10
Aug
Sep
Oct
Nov
Date
Superforecasters (n = 46)
Experts (n = 22)


Space Colony
2030
Probability
4
3
2
1
0
0
0
0.01
Superforecasters (n=46)
Experts (n=24)
General X-risk Experts (n=6)


Space Colony
2050
Probability
25
20
15
10
5
0
0.19
1.75
1.4
Superforecasters (n=46)
Experts (n=24)
General X-risk Experts (n=6)


Space Colony
2100
Probability
80
60
40
20
0
3
12.5
19
Superforecasters (n=46)
Experts (n=22)
General X-risk Experts (n=5)

Sources of agreement, disagreement and uncertainty

Disagreement for this question can be broken down into two categories:

1. Motivation
2. Feasibility

With the base rate of life-sustaining colonies outside of Earth's atmosphere being zero, other factors were considered as reference classes or base rates for constitutive elements:

- “Space exploration/development notoriously runs behind schedule.” (Team 336)
- “The Lunar program [...] base rates [...] suggest a timeline of around 10 years from initial commitment to first manned landing on Mars.” (338)
- “Time to grow a colony: From Columbus’s landing to a town with 33,131 people (New York) was 300 years.” (338)

The resolution details contained the following definition of “life-sustaining”: “this colony would need to be able to sustain its population over multiple generations even if there were no humans alive on Earth.” One team interpreted this definition to be an extremely high bar:

> “In order to be truly self sustaining the colony would need to be able to survive multiple generations without contact with earth. That means it would need to master terraforming, energy production and life support systems. And have them to a level where they are robust enough to be taken care of without outside help. This is on top of the herculean challenge of getting the cargo required to the planet. For that reason, the lowest ends of our forecast give very small chances for colonization happening by 2100.” (345)

However, another team made a case for this same definition being a relatively low bar, one that only implied a need for the colony to sustain itself, technically, for two generations:

> "The fine print states, “To be considered "life-sustaining," this colony would need to be able to sustain its population over multiple generations even if there were no humans alive on Earth. If resolution is ambiguous, this question will be resolved by a panel of experts.” First, note the language, “would need to be able to.” That is different from, “is currently sustaining its population”. The distinction might seem minor, but to me that means that if there’s a solid backup plan, with the necessary infrastructure in place, that would allow such a colony to switch from partially dependent to fully independent and sustainable in the event of a disaster, that would potentially (but not definitely) qualify. Also note the language “multiple generations”. Merriam-Webster defines multiple as being more than one. Two generations is a lower bar than, say, a hundred. Relatedly, I think there’s a decent chance that a hypothetical “panel of experts” would conclude, if it could be determined that women could conceive at the base, and the other life sustaining elements were in place, that multiple generations would “be able to” survive.” (337)

When possible locations for a colony were explicitly mentioned, it was usually assumed Mars and the moon were the most likely ones.[1541] [1542] [1543] One team summed up pros and cons for each:

> "It's not obvious whether the Moon or Mars is more likely to host the first self-sustaining colony. (We might infer that the Moon's proximity might make the effort more appealing; however, the "go big or go home" effect might lead progress toward centering focus on Mars.) Mars suffers from greater technical difficulty while the Moon may generate less motivation and be subject to more diplomatic issues. It would be interesting to forecast Moon and Mars colony probabilities independently. (Ultimately, the Moon will likely will be a proving ground for future Mars missions.)" (338)

Arguments given for forecasts ≥ 0.000000005% (2030), 7.5% (2050), 26.5% (2100)

<u>OVERALL</u>

On 1 (motivation):

- We may discover incentives to do this in the future that we currently don't know about.[1544]
- Elon Musk "has a stated goal of establishing off-Earth colonies" and Jeff Bezos has his own rocketry firm, and the desire of these individuals could be sufficient motivation even absent economic motivation.[1545]

On 2 (feasibility):

- AGI, if it is human-aligned, might be very helpful for this task.[1546]
- There might be an "unforeseen breakthrough" (344) in how we produce energy that makes reaching orbit easier.

<u>2100</u>

On 1 (motivation):

[1541] "The energy requirements to pull an asteroid from its orbit around the sun and bring it to Mars or the moon are, well, astronomical." (343)
[1542] "Living in space, the moon, or on Mars is simply too hard on the human body." (340)
[1543] "NASA and China plan to go back to the Moon and Mars." (337)
[1544] 343, "Future incentives for exploring/developing off-world resources may not match currently-known opportunities. New surveys and research of Luna and Mars may reveal resources worth getting."
[1545] 343, "The financial benefits of an off-Earth colony are not yet fully developed. Much exploration and colonization throughout history has been economically motivated, and we may not establish a colony without that motivation. However, Elon Musk, the richest person in the world, has a stated goal of establishing off-Earth colonies, and has his own rocketry firm. Jeff Bezos also has a rocketry firm and recently took a trip himself. The motivation of the ultra-wealthy could be enough."
[1546] 343, "Human-aligned AGI might be able to overcome current hurdles that make space-faring colonies prohibitive. This requires not just the development of AGI, but also that this AGI is human-aligned and sufficiently trusted to help develop an off-world program."

- Resources: “We expect humankind to be pressed for resources by this date, and these resources could be found in space” (344).
- Militarization: “The strongest arguments given are tied again to the militarization of space and the need for resources” (344).

On 2 (feasibility):

- There is plenty of time between now and 2100 to achieve this.[1547]

Arguments given for forecasts ≤ 0.0000001% (2030), 1% (2050), 14% (2100)

<u>OVERALL</u>

On 1 (motivation):

- We will have no motivation to do this.[1548] [1549] [1550]
- The value of this will not justify the cost.[1551]
- Robots will probably be more economical for tasks we’d want such a project for, such as mining.[1552]
- Interest in space will wane.[1553]

On 2 (feasibility):

- There will not be the political will to do this.[1554]
- Massive technological advances would be required.[1555]
- The resources required would be very high.[1556]

[1547] 344, “We have a high time period to make significant advances in space technologies.”
[1548] 341, “Motivation: there may not be any economic incentive to have a fully self-sufficient colony that doesn't rely on Earth even for medicine, etc.”
[1549] 337, “What would be the point of a colony [...]? [...] there are more pressing problems to attend on earth, which might make humans lose the sense of urgency about exploring space.”
[1550] 339, “[T]he motives for doing so are weak.”
[1551] 337, “[P]essimists point out that the involvement of these entities would likely depend on whether the perceived value of commercial or geopolitical interests would merit the considerable costs.”
[1552] 337, “Robotic exploration (and economic exploitation) of space: [...] it might simply happen that it is more convenient to do space mining with robots, more so if advances in AI make the robots as versatile as humans.”
[1553] 337, “[T]here is a reasonable likelihood that there will be periods between now and then when interest in colonizing space will wane, and that in 78 years the world's politics will probably be very different, which increases the uncertainty in achieving this goal.”
[1554] 345, “To date there has still been a lot of hesitation for government to devote large amounts of funding to space based endeavors let alone colonization efforts. And an endeavor of this magnitude would likely require government scale coordination and cooperation.”
[1555] 337, “The really daunting list of engineering problems that have to be solved before having a self-sustained colony. In particular, [a forecaster] share a really interesting thinkpiece by Casey Hardmer discussing the previous problems for the case of Mars, thus reducing the probabilities for this question to be resolved positively.”
[1556] 341, “The resources needed are uncalculatable.”

- Comparable historical precedents took a long time.[1557]
- Compared to the International Space Station, a self-sustaining colony would be three orders of magnitude more expensive, and, assuming the project is tied to the growth rate of the world's economy, could be expected to be completed not sooner than 235 years from now.[1558]
- Antarctica, a comparable project, is not self-sustaining.[1559] [1560]
- Biosphere II, a comparable project, failed.[1561]
- One colony of 5000 people is a high bar.[1562] [1563]
- Self-sustaining is difficult technologically and economically.[1564]
- People will not want to go.[1565]

<u>2030</u>

On 2 (feasibility):
- Cost: "The costs of establishing a colony would be extremely high." (339)
- Logistics problems of getting people into space make it not practicable by 2030.[1566]
- Technical requirements insurmountable by 2030.[1567] [1568]

<u>2100</u>

On 1 (motivation):

[1557] 341, "Colonization of North America. "From sighting to first small colony was 500 years."
[1558] 341, "Base Rates: If a colony is three orders of magnitude more expensive to build than the ISS (three orders of magnitude more people, 1-2 orders of magnitude more complex, requirement for self sufficiency), and the world economy grows at 2% (3%), then it would be a similar scale investment 350 (235) years after the ISS. That is probably a rough estimate of when we could expect a qualifying colony to be built for purely scientific purposes."
[1559] 341, "The current summer peak population of Antarctica appears to be ~5,000, but I do not think it would be self-sustaining in the event of a sudden cutoff from the rest of the world."
[1560] 337, "The fact that there is no self-sufficient human presence in Antarctica, as these bases depend on external supplies."
[1561] 337, "The failure of the Biosphere-2 project in the eighties."
[1562] 337, "[T]o fulfill the question, it will need to be ONE colony of 5,000, not more than one colonies adding up to 5.000, which increases the difficulty of resolution."
[1563] 336, "[L]life-sustaining' for 5000 people is a high bar."
[1564] 337, "[G]ive heavy weight to the self-sustaining caveat, noting how difficult this has been in similar situations. There is a concern over the pure amount of technological hurdles that must be overcome, and the massive costs of such an undertaking."
[1565] 336, "[L]ack of desire to live off-Earth."
[1566] 339, "[R]espondents estimated that it would take more than 1,000 present-day rocket launches simply to transport the material requirements for a colony of this size. Given we currently complete around 120 launches per year, this all but ruled out the 2030 timeline."
[1567] 339, "[N[o one has the technical capability within the decade to transport that many people off the planet."
[1568] 336, "[T]echnology will not be there by 2030"

- A mine on Mars populated by robots and sometimes tourists is "potentially probable," but a fully sustainable human population is not.[1569]

On 2 (feasibility):

- Political and economic hurdles such as "getting appropriations, finding a way to generate positive cash flows" will be overwhelming.[1570]

Other arguments given

Arguments in favor of lower forecasts:

<u>OVERALL</u>
On 2 (feasibility):

- "AGI alignment failure could result in humans being eliminated or severely incapacitated before such a colony is possible." (343)

<u>2050</u>
On 2 (feasibility):

- "There has been little progress in space in the last 70 years." (344)

Arguments in favor of higher forecasts:

<u>2100</u>
On 1 (motivation):

- Ideological reasons: "Ideological reasons might motivate people to build a large, self-sustaining colony" (340)

Question 56: Happiness in America

In a nationally representative survey, what percentage of Americans will report being "very" or "fairly" happy in response to the question, "Generally speaking, how happy would you say you are — very happy, fairly happy or not too happy?"...
...in 2030?
...in 2050?

[1569] 339, "I think it is potentially probable there is a colony on Mars that is used as a big mine, populated by robots designed to bring back resources and then occasionally send tourists to. But meeting the fully sustainable for the purposes of sustaining human life for generations? I don't buy that that is more likely than not to occur."
[1570] 339, "The primary and overwhelming hurdle is political and economic (getting appropriations, finding a way to generate positive cash flow) versus technological."

...in 2100?

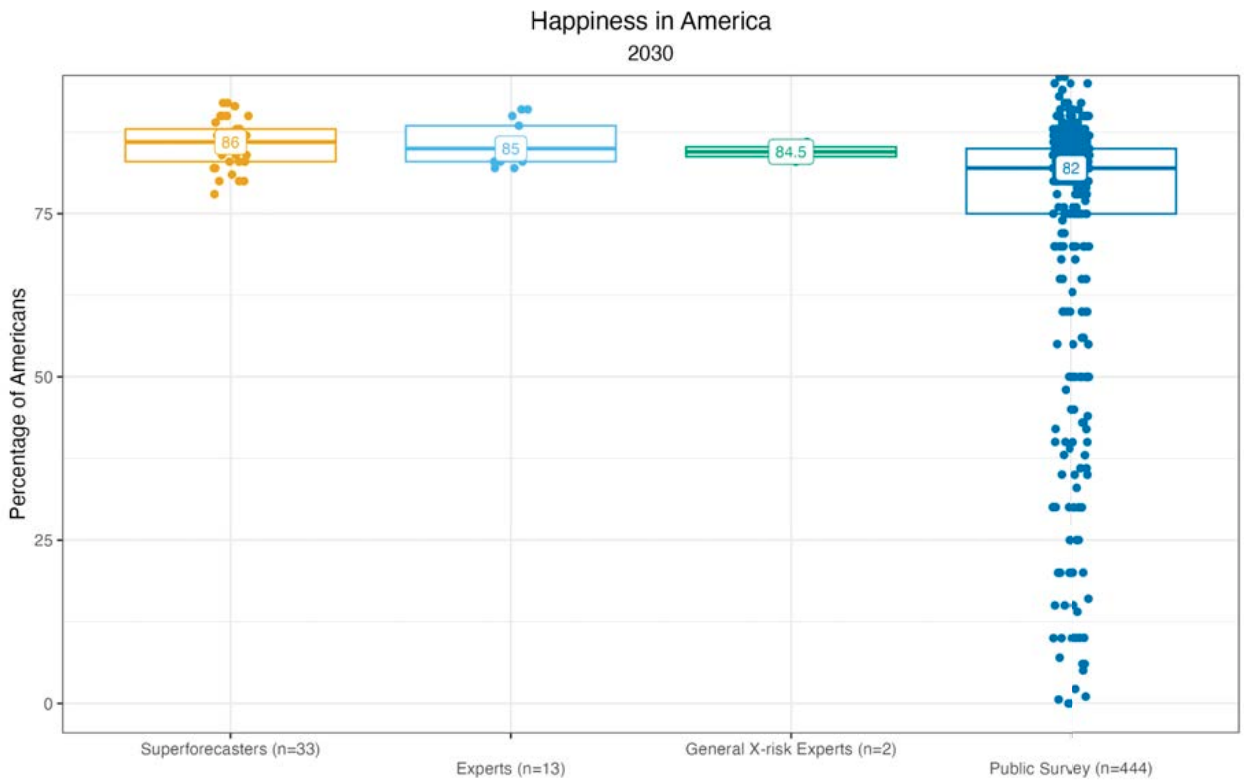
Question and resolution details, prior forecasts, and other relevant sources

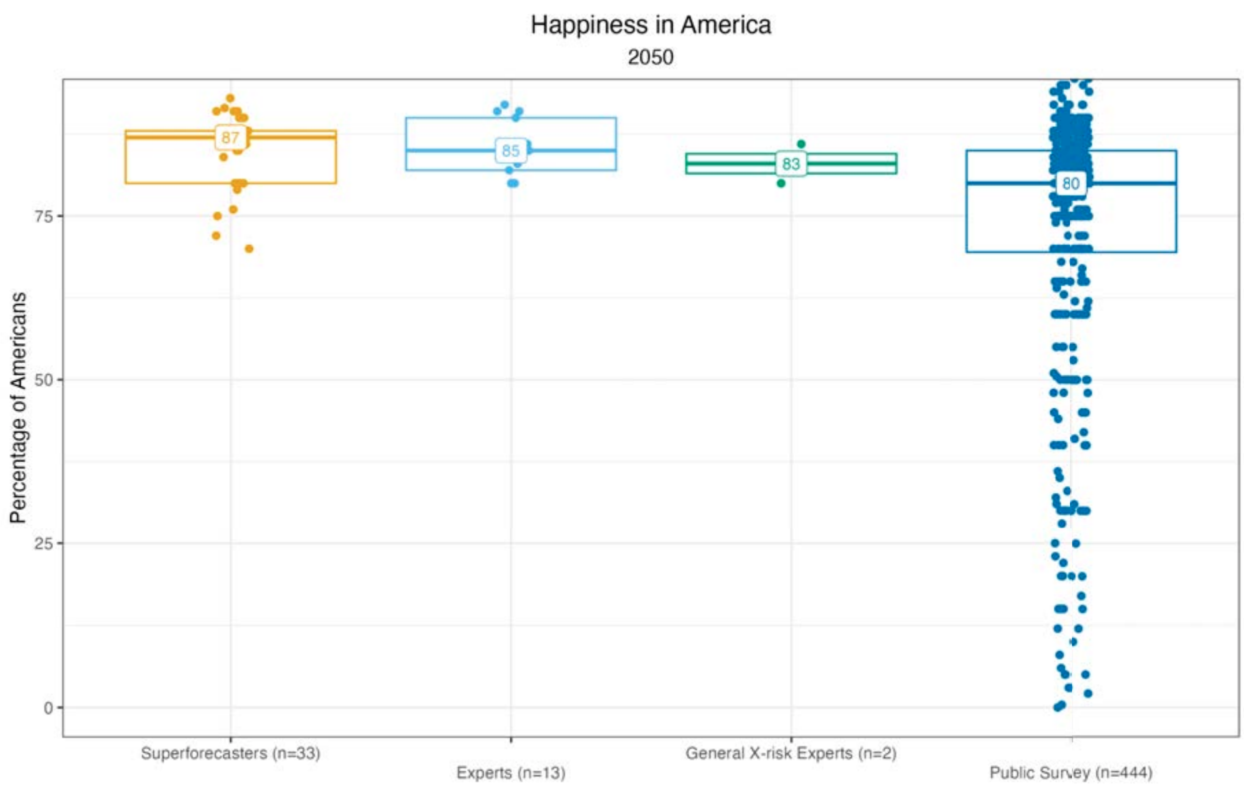
Results[1571]

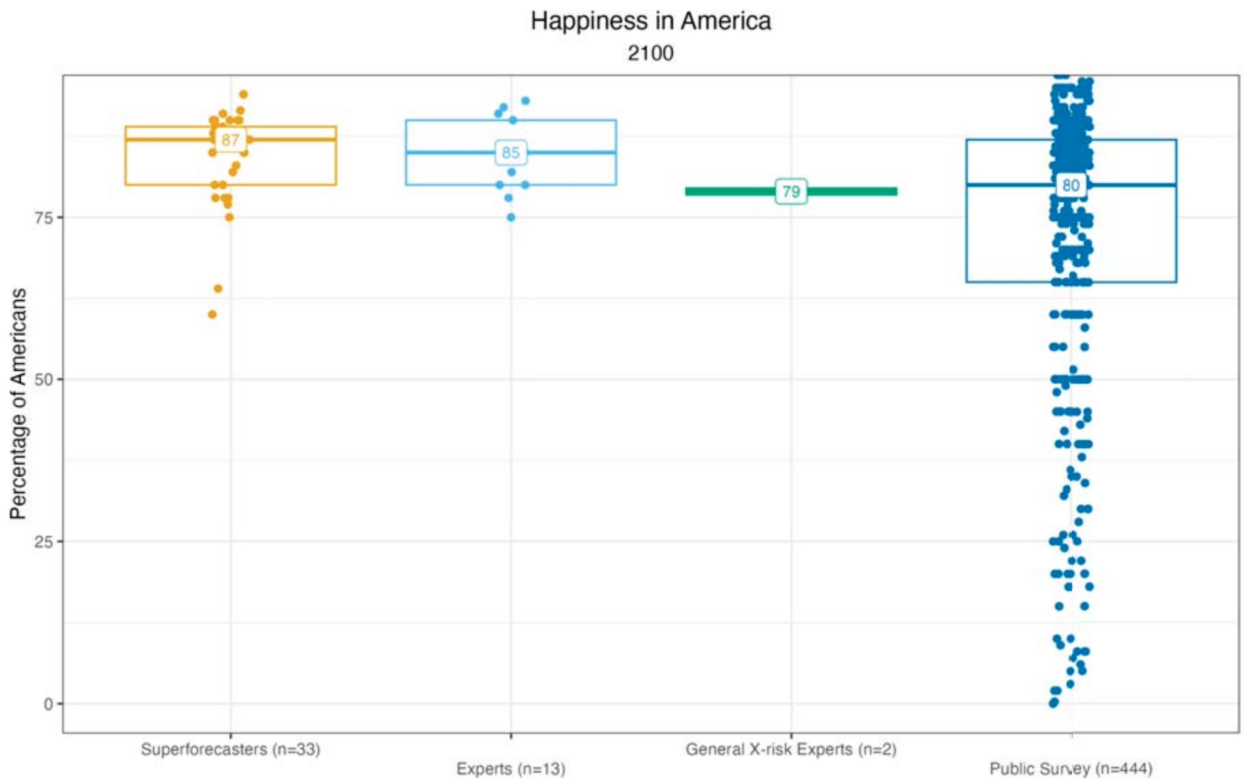
Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 33)	2030	85%	86%	3.73	+2.11%
	2050	85.5%	87%	5.72	+2.46%
	2100	85.5%	87%	8.71	-14%
Experts (N = 13)	2030	86%	85%	3.77	-11.05%
	2050	86%	85%	6.43	-31.12%
	2100	86%	85%	8.28	-31.86%
General X-Risk Experts (N = 2)	2030	83%	84.5%	n/a	n/a
	2050	75%	83%	n/a	n/a
	2100	70%	79%	n/a	n/a
Public Survey (N = 444)	2030	82%		21.03	-
	2050	80%		20.78	-

[1571] Numbers of forecasters are given as of Stage 4 of the XPT. Public survey respondents were surveyed outside of the tournament context.

	2100	80%	22.19	-


Happiness in America
2030
Percentage of Americans
0
25
50
75
86
85
84.5
82
Superforecasters (n=33)
Experts (n=13)
General X-risk Experts (n=2)
Public Survey (n=444)


Happiness in America
2050
Percentage of Americans
0
25
50
75
87
85
83
80
Superforecasters (n=33)
Experts (n=13)
General X-risk Experts (n=2)
Public Survey (n=444)



Sources of agreement, disagreement and uncertainty

For this question, in general, the majority of arguments put forth by teams with higher and lower forecasts did not directly contradict each other; more often, the teams simply raised different issues.

There was only one significant exception to this, which was an object-level disagreement about what to expect in terms of whether life in America will materially improve, as measured by metrics like GDP, “quality of life,” “standards of living,” life span, and so on. (I’ll call these arguments “Object-level material conditions.”)

There was, however, one other implied disagreement: the more pessimistic teams had many more arguments for why happiness will decrease, a great many of which were specific ways in which the world was projected to get worse. The optimists, on the other hand, did not have counterarguments for each of these; rather, their arguments tended to be more general, often acknowledging that bad things have been happening forever, yet happiness levels nonetheless seem fairly stable for one reason or another. So there is an implied disagreement about the extent to which material circumstances (as measured by the broadest and simplest metrics, like GDP, life span, and so on) affect subjective happiness.

Uncertainties:

1. Team 341 observed that cultural changes could affect how survey respondents answer: “A notable source of uncertainty is what poll respondents are thinking when they respond to this question. Do they feel pressured to respond as happy, or alternatively convince themselves upon self-reflection that they are happy? If either is true, will this aspect of culture change?”
2. Team 336 questioned whether the historical data were sufficient: “Uncertainty due to lack of recent data from this survey and long lapses in collection of survey data.”

Arguments given for forecasts ≥ 87.5% (2030), 87.5% (2050), 87% (2100)

On object-level material conditions:

- Progress will increase, wealth will increase, and these things will cause increased happiness or the same level of happiness as the historical average within the Gallup poll’s window.[1572]
- “Scientific innovation” will cause an increase in happiness. (338)
- **“**Economic growth” will cause an increase in happiness. (338)
- Standards of living will be raised.[1573]
- People will live longer.[1574]
- Quality of life will increase because of “massive leaps in technology.”[1575]

Other arguments:

- Despite the most recent Gallup poll being lower than the historically common range, the happiness rate will regress to the historical mean, where historical refers to the period of time during which the poll has been sampling: 1949-2019.[1576] [1577]
- Relatedly, the ‘hedonic treadmill’ effect will keep the happiness rate within historical bounds.[1578]
- People naturally have a happy outlook.[1579]
- Better policy could get implemented due to an increased focus on happiness and well-being.[1580]

[1572] 338, “[T]he expectation of progress and increased wealth.”
[1573] 345, “Standards of living will be raised”
[1574] 345, “People will live longer.”
[1575] 345, People will have “better quality lives because of massive leaps in technology.”
[1576] 343, “Reversion to the mean: Happiness reporting appears to fluctuation from 84-96%. Our expectation is that long-term trends will likely continue this fluctuation, despite recent trends downward.”
[1577] 341, “For approximately 80 years, there has been very little change in this metric. So there should be little change from the most recent observation of 86%, other than some regression to the mean of ~90%.”
[1578] 338, “There may also be a self-correcting mechanism that keeps forecasts generally high and resilient to dramatic shocks. In the face of objective material improvements or deprivation, people's expectations adjust to accommodate the change (the "hedonic treadmill"). This effect may apply to Americans in particular or to humans generally. Even risks ("bad news") like the effects of climate change also create more opportunities for innovations and solutions ("good news") that could provide optimism despite objective declines.”
[1579] 338, “[T]he individual and collective survival benefits of having a happy outlook.”
[1580] 345, “We also think that general global focus on happiness and well being could lead to better policy implementation that actually attempts to generate a positive response to this metric.”

- The 2019 data point in the Gallup poll may be erroneous, which would negate any indication that there has been a recent downward trend.[1581]
- Happiness is stable even through various kinds of historical periods. “Stability of this measure over long periods taken to indicate that a lot can happen without affecting the measure substantially.” (336)
- “Aging population could be more likely to report more happiness.” (336)

Arguments given for forecasts ≤ 84% (2030), 84.5% (2050), 82.5% (2100)

On object-level material conditions:

- Economic conditions are already bad,[1582] and could worsen.[1583]
- The lack of a social safety net, including universal healthcare, causes unhappiness.[1584]
- Many other material conditions and societal dynamics cause unhappiness. One forecaster summarized:

> "Life satisfaction seems to be levelling off for the average American. I believe a lot of this has to do with social media and the ubiquitous comparisons of our life with those of others. 24.7 News only keeps banging the drums that the political polarization is tearing at the fabric of our country. Democracy is hanging in a delicate balance. The very real inequality between the top 10% and the rest of the population is coming into glaring focus. The lack of a social safety net, the cost and availability of health care, the opioid and addiction crises, the demise of public education all are becoming big issues, And the constant enslavement to capitalism, where the chanting about "the market" and how well it is (or isn't) doing overshadows the fact that few can participate in that party; the rich get richer, and the rest only wish they had a retirement fund; it seems that few people want to understand that the "market" is NOT the "economy".... Now throw in no access to abortion, guns toted nearly everywhere, public money going for religious education, gridlock in our Senate and Congress, inflation, serious problems that can potentially no longer be addressed by government agencies who employ experts (instead we'll rely on ineffective politicians bought by special interests), and an agenda from certain members of the Supreme Court to deny birth control, & same sex marriage, plus states that want to hunt women down

---

[1581] 341, “December 2019 reading may have been an outlier due to margin of error in polling, or otherwise may have been caused by transitory factors.”

[1582] 339, “[T]he constant enslavement to capitalism, where the chanting about "the market" and how well it is (or isn't) doing overshadows the fact that few can participate in that party; the rich get richer, and the rest only wish they had a retirement fund; it seems that few people want to understand that the "market" is NOT the ‘economy’.”

[1583] 340, “Economic conditions in the US could worsen, and wealth is an important predictor of happiness. In addition, the safety and health of many Americans might decline, possibly even if the economy keeps growing, due to wealth inequality.”

[1584] 339, “Until we provide a solid social safety net including healthcare for all, we will foster more inequality, insecurity, worry and stress. So many are one paycheck away from going broke - and an illness in the family would push them over the edge.”

who seek abortion in "safe states", There you have the recipe for making a LOT of people very unhappy in the short term.” (339)

Other arguments:

- Social media, including dating apps, causes unhappiness and may continue to do so.[1585] [1586] [1587] This may especially affect current teenagers, who will influence the US average as they take up a greater proportion of the population.[1588]
- Political tensions and political polarization is the cause of a lot of US unhappiness, and there’s a good chance this will not get better, and may get worse.[1589] [1590] [1591] This trend may eventually result in authoritarianism, which would make the survey impossible to carry out.[1592]
- Climate change may cause,[1593] [1594] or is likely to cause,[1595] [1596] increasing unhappiness.
- Human extinction,[1597] extreme catastrophes,[1598] and authoritarianism are all possibilities.[1599]

[1585] 339, “Honestly I would not be surprised with a sudden decrease in happiness, as the last 10 years have seen the rise of phenomena that could be described as unprecedented, such as smartphones and social media. These suddenly turned most "local" established social dynamics into "global" ones (e.g. think Tinder, where you go from competing with your social circle to get the attention of the girl, to competing with the entire town), and it could be that this has really impacted happiness."

[1586] 339, “Life satisfaction seems to be levelling off for the average American. I believe a lot of this has to do with social media and the ubiquitous comparisons of our life with those of others.”

[1587] 339, “Potentially social media (jury is out)."

[1588] 340, “Current teenagers seem to be less happy than previous generations, and if this continued (no matter whether the underlying cause is social media, economic conditions, or increased neuroticism) it might have larger scale effects, once more and more of the population is replaced by current teenagers.”

[1589] 339, “'[[P]olarisation' in the broadest sense is probably where I see the immediate-term decline coming from.”

[1590] 339, "I can imagine scenarios [...] in which political divisions in the country push many people into a state of dissatisfaction that is far greater than would be predicted based on material comfort. I definitely think that's possible, and I would probably assign a 1/4 chance that we are on that path."

[1591] 339, “24.7 News only keeps banging the drums that the political polarization is tearing at the fabric of our country. Democracy is hanging in a delicate balance.”

[1592] 340, “Political tensions could both increase polarization and increase the ratio of unpleasant interpersonal interactions, exacerbate economic decline or even contribute to rising authoritarianism that could make it impossible to independently survey Americans, thus resolving the question as zero.

[1593] 340, [It’s uncertain] “how well the US is going to adapt to climate change.”

[1594] 340, “Climate change could exacerbate the previous points and also amplify people's worries and stress levels directly (for example via an increased frequency of natural catastrophes).”

[1595] 339, "Climate change messing up everything.”

[1596] 339, "Heat, storms, floods, crops failures, water shortages, brown/blackouts, will force migration on a lot of people around 2100."

[1597] 344, “Some forecasters even input below 10% probability” in later years. The main arguments for these low numbers are given as slow catastrophic and extinction events: Non-anthropogenic sources of risk, Nuclear confrontation, Novel pathogens.”

[1598] 339, “Lower bound [of predictions] in there to catch catastrophe scenario.” (Type of catastrophe unspecified, but mentioned in the context of climate change and civil war.)

[1599] 340, “Some outlier forecasts have the 5th percentile at 0 for 2100, based on a mix of extinction scenarios and the risk of authoritarianism.”

- Overall, the trend of happiness in the US seems to be going downward, and that seems likely to continue.[1600] This will negate the historical optimism of the American people, which was founded on aspects of the country which are changing, such as “the growth of our economy, living standards, public education, immigrants seeing big gains in their life circumstances.”[1601] As a corollary, it may be the case that, long-term, the rate of happiness is not as stable as is implied by the Gallup data from 1949-2009.[1602]
- “Recent Supreme court rulings” will increase unhappiness.[1603] Decreased access to abortion was mentioned in this context.[1604] [1605]
- Factors that make people happy are declining:

  > "The factors listed for more happiness are sleep, face-to-face interaction, and church attendance. Sleep is negatively impacted by screen time, so is face-to-face interaction. Church attendance has declined in America and is expected to continue to do so as more people identify as non-religious." (339)

- Increase in stress specifically appears to be a uniquely American problem among rich countries.[1606]

Other arguments given

Arguments in favor of lower forecasts:

- Smaller extended family sizes have decreased happiness.[1607]

[1600] 339, “A recent drop from the past 92% average to 86% during COVID (2020) was noted and, while the outlier in such a specific situation was not given too much weight, it was widely attempted to interpret it as part of a larger trend: One forecaster notes "the survey values have been pretty consistent over the last 70 years, with a mean ~92% and a fair number of responses that get up to 95-96%. It's possible that there has been a recent, downward trend in happiness such that values continue to decline.” Note that this writer goes on to say: “but I also think it's very likely that there will be a reversion to the mean, such that the next response could easily be 92% or 93%.”

[1601] 339, “Americans are historically an optimistic lot. Much of this had to do with the growth of our economy, living standards, public education, immigrants seeing big gains in their life circumstances. However, a lot of that is changing.”

[1602] 340, “[T]he comparatively large shift in happiness scores between 2009 and 2019 (which could suggest that happiness is more volatile than older trends might suggest).”

[1603] 339, “I would think that with recent Supreme court rulings the Non-whites and the Democrats will face increasing unhappiness.”

[1604] 339, “Now throw in no access to abortion.”

[1605] 339, “[P]lus states that want to hunt women down who seek abortion in "safe states", There you have the recipe for making a LOT of people very unhappy in the short term.”

[1606] 339, “The latest dip in U.S. happiness is likely intertwined with Americans' higher reports of stress, worry and anger. While stress, in particular, is not unique to the U.S., the country stands out on the list of mostly less developed countries where this emotion is most prevalent, as these countries face obstacles and instabilities on levels that Americans have not experienced.” Cited Gallup poll.

[1607] 337, “An apparent slight decline [in happiness] seems to have been happening since the 1950s, particularly over the last 15 years or so. The reasons for this put forward included smaller extended family sizes.”

- Perceptions and expectations are more important to happiness than material wellbeing, and Americans' "perceptions and relative expectations of how society ought to be" have gotten worse.[1608]
- Americans "reject the political norms that seem to be linked to happier people"—for example, those of Northern Europeans.[1609]
- Team 336 mentioned one survey not mentioned by any other team, which included data up to 2021, and showed far more pessimistic recent results:

> "NORC runs the GSS and asks a question similar to this one, but conveniently has more data points including in 2021. The responses for 2021 showed a sharp uptick in the percentage that were "not too happy" to 24%. Typically this survey has had a higher % indicating not too happy than the Gallup poll. When run in the same year, the Gallup poll typically finds 4-6% more in the happy categories."[1610]

Arguments in favor of higher forecasts:

- New drugs could cause increased happiness.[1611]

Cross-references with other questions

Q58: Future Worries and Children
Q59: Generation Attitudes

Question 57: Prevalence of Autocracies

What percentage of the world population will be classified as living in an electoral or closed autocracy in the V-Dem Institute's annual Democracy Report…
…for the year 2030?
…for the year 2050?
…for the year 2100?

Question and resolution details, prior forecasts, and other relevant sources

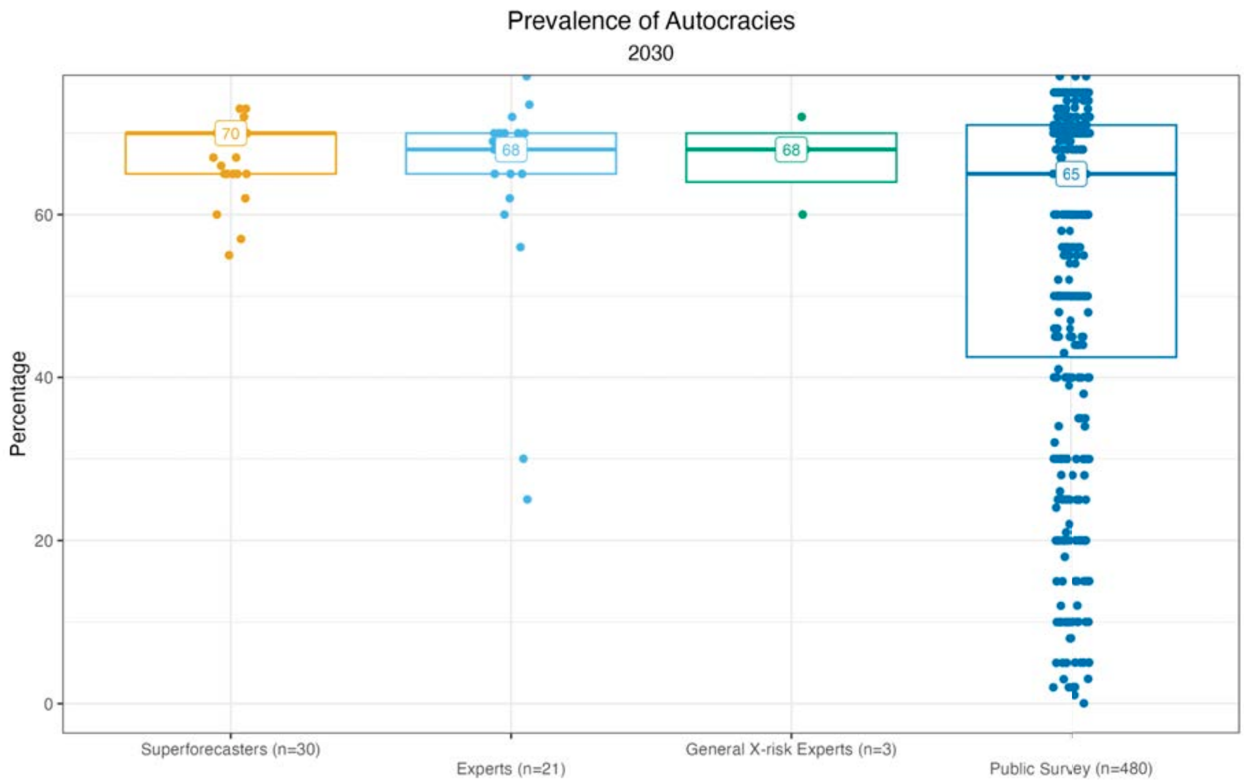
---

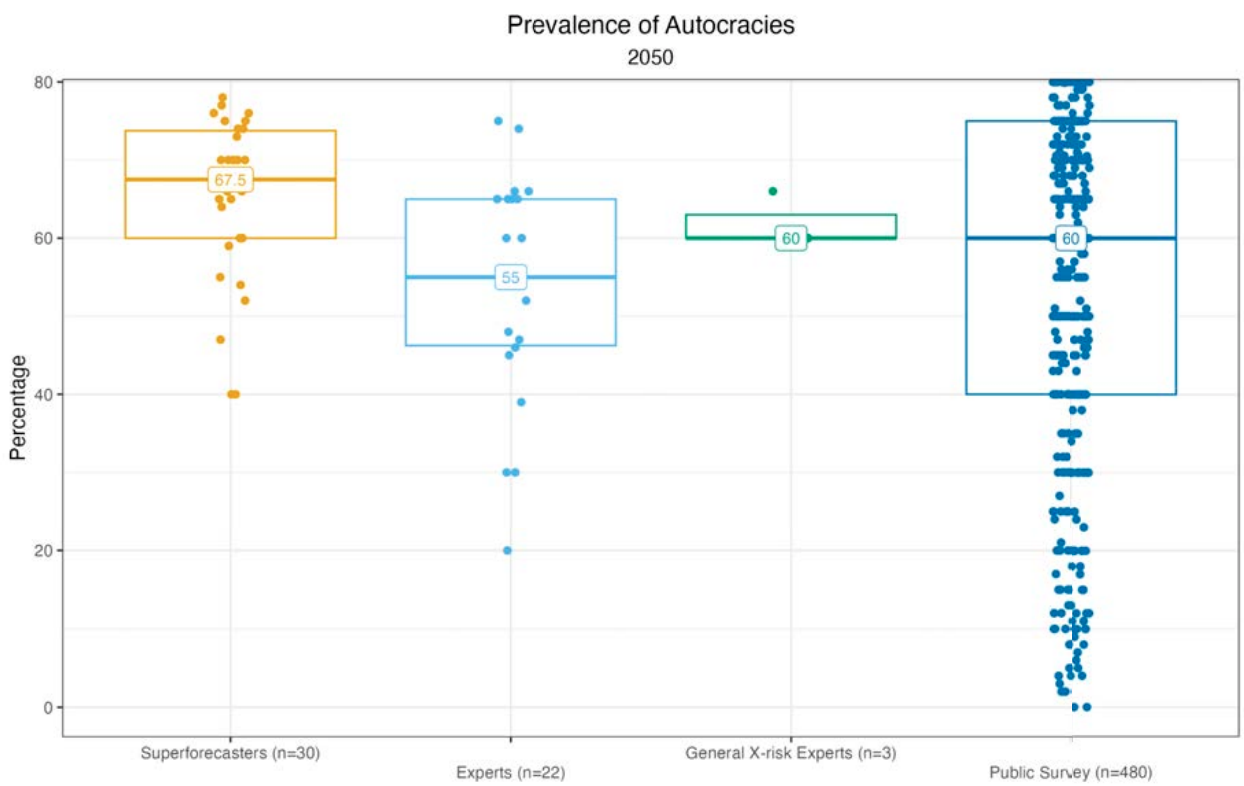
[1608] 337, "It was pointed out that in general the world is a better place than it was, but that does not seem to have improved happiness scores (which must indicate that happiness as measured by the polls is not related to objective measures of GDP, health, danger etc. but to perceptions and relative expectations of how society ought to be."
[1609] 337, "The higher happiness scores of Northern Europeans was noted. They have relatively high GDP per person, lower than the US, but more evenly spread, their politics relies on consensus rather than confrontation, they have better public services than the US. There is no sign of the US wishing to emulating the happiest countries. Some team members agreed that many prople in the US reject the political norms that seem to be linked to happier people."
[1610] No direct link to the data was included, but this is the relevant website: https://gss.norc.org/
[1611] 344, "Mode-enhancing drugs would influence the sentiment."

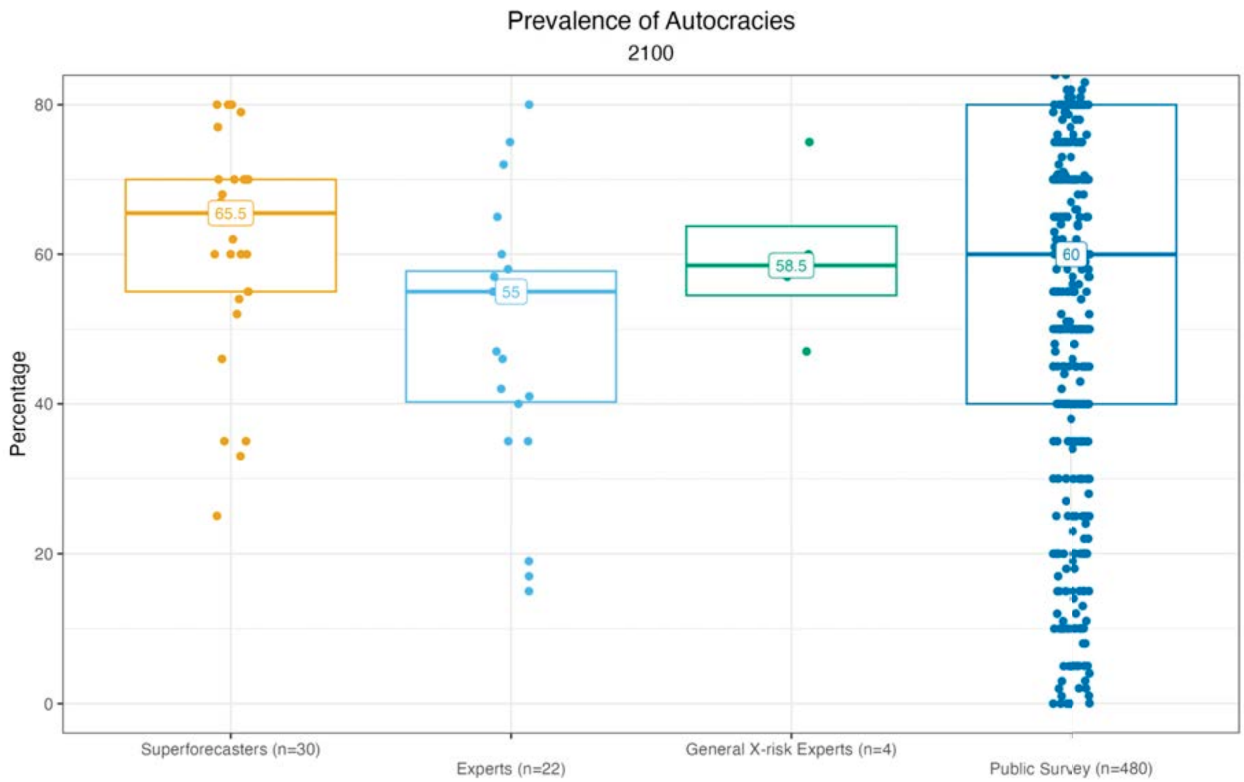
Results[1612]

Group	Year	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 31)	2030	70%	70%	21.83	-79.88%
	2050	60%	67.5%	22.64	-53.81%
	2100	60%	65.5%	24	-38.18%
Experts (N = 21)	2030	60.5%	68%	18.13	-28.63%
	2050	53%	55%	18.92	-23.03%
	2100	42%	55%	15.6	-13.18%
General X-Risk Experts (N = 3)	2030	72%	68%	n/a	n/a
	2050	60%	60%	n/a	n/a
	2100	43%	58.5%	n/a	n/a
Public Survey (N = 480)	2030	65%		92.86	-
	2050	60%		92.98	-
	2100	60%		94.2	-

[1612] Numbers of forecasters are given as of Stage 4 of the XPT.


Prevalence of Autocracies
2030
Percentage
0
20
40
60
70
68
68
65
Superforecasters (n=30)
Experts (n=21)
General X-risk Experts (n=3)
Public Survey (n=480)


Prevalence of Autocracies
2050
Percentage
0
20
40
60
80
67.5
55
60
60
Superforecasters (n=30)
Experts (n=22)
General X-risk Experts (n=3)
Public Survey (n=480)



Sources of agreement, disagreement and uncertainty

In general, disagreement on this question was on whether the recent reversal of the long-term trend toward democracy means the long-term trend is over, or whether it's just a temporary dip. The only explicitly divisive issue was whether AI developments would help or hinder democracy; otherwise, teams with low and high forecasts raised different issues.

Uncertainties

1. Some highly populous countries, like Brazil, Nigeria, Pakistan,[1613] and India[1614] seem, at the moment, to have the potential to go either way—more democratic, or more autocratic.
2. The fact that most of the forecasters are Western may bias them to believe it's inevitable that democracy will last forever.[1615]
3. The uncertain nature of a hypothetical AI-run government: "Is a country ruled by a benevolent AI that tries to take into account the wishes of its population a democracy, an autocracy, or something else?" (337)

[1613] 344, "The large level of uncertainty stems from other highly populous countries that can easily change classifications like Brasil, Nigeria, Pakistan, etc…"
[1614] 339, "How to categorize India in 2030 because its large population affects the outcome of the question."
[1615] 344, "Most of the current forecasters come from the western world which is democratic in nature so may be anchored to the belief that the system will last forever."

Arguments given for forecasts ≤ 65% (2030), 53% (2050), 50% (2100)

1. AI/AGI could hurt autocracies by providing “a sustainable solution to the problems of sustainable economic and political development” (343).
2. Increased global access to information will spread democracy.[1616] [1617]
3. Relatedly, technology will likely spread democracy,[1618] by “facilitat[ing] direct democracy election[s]” (344), or by Starlink circumventing government firewalls.[1619]
4. The long-term trend is toward increasing democracy.[1620] [1621]
5. The West’s economic dominance will sway the world toward democracy.[1622]

Arguments given for forecasts ≥ 70% (2030), 65% (2050), 57.5% (2100)

1. AI makes it “potentially [...] harder for democracy to flourish” (341).
2. “Asia is 59% of the world’s population and Africa is 17% and growing quickly” (341), and what happens there in the coming years will be decisive. The evidence points to stability in existing autocracies in the region, like China,[1623] Pakistan, and Egypt.[1624]
3. The most populous places are the most important for the resolution of this question, and many of the world’s most populous places seem to be becoming increasingly autocratic, or in danger of moving in that direction, including India,[1625] Pakistan, Egypt,[1626] Brazil,[1627] Russia,[1628] and Nigeria.[1629]
4. Climate change makes it “potentially [...] harder for democracy to flourish” (341).
5. Autocracy will increase due to “global and subnational power play politics of world leaders’ thirst and quest for control across continents of the world” (342).

[1616] 344, “Autocracies are less efficient and are expected to slowly evolve over time with increasing global access to information and communications.”
[1617] 337, “The difficulty of controlling information in the modern world was also cited as a potential cause for some optimism.”
[1618] 338, “Some consideration of technology: ‘A value near 0% also doesn't seem out of the question if there's a similar shift in power toward regimes that value non-autocracy.’”
[1619] 337, “Technology like Starlink as a tool for organization, thus far able to route around government firewalls, also adds mild downward pressure on autocratic regimes.”
[1620] 337, “In the long term, there is a clear tendency for more people living under a liberal or electoral democracy: in 1900, it was about 3% of the population, in 1925 about 20%, in 1955 about 35% and in 2000 about 50%.”
[1621] 338, “Just general optimism about continuation of long-term trends: ‘Optimistically, I see a mid to long term movement away the electoral autocracy and back towards electoral democracies and liberal democracies.’”
[1622] 337, The West will “dominate economically this century by siphoning off the world's best and brightest, particularly with AI… Therefore, I expect alignment to the China/Russia axis, currently attractive to some (e.g. India, Brazil), to become considerably less so over time.”
[1623] 341, “China will likely not cross the divide, it will remain in group 4 or perhaps 3.”
[1624] 341, “Pakistan and Egypt will likely stay put on the autocratic side.”
[1625] 341, “India crossed the divide in 2019.”
[1626] 341, “Pakistan and Egypt will likely stay put on the autocratic side.”
[1627] 341, “Brazil could be on the move to autocracy under its current president.”
[1628] 341, “Russia is clearly moving in the autocratic direction.”
[1629] 341, “Nigeria crossed the divide only recently to the autocratic side.”

6. The number of countries under autocratic rule may decline, but countries that are expected to remain autocratic tend to have larger birth rates.[1630]
7. The world is becoming more anti-fragile and beginning to stabilize around its current state, which means the current status quo will persist, and, currently, the population under autocracy is at 70%.[1631]
8. Disasters and wars will increase autocracy, because governments inclined to autocracy use these events opportunistically for this purpose.[1632]

Other arguments given

Arguments in favor of lower forecasts:

1. "Autocracies will make mistakes thus turning people toward democracy. Specifically for 2030, forecasters raised China's poor handling of Covid, China's relationship with Taiwan, and Russia's invasion of Ukraine that could result in lower numbers. For 2050 and 2100, China's potential population decline could mean lower numbers" (339).
2. India, 17% of the world's population, is currently classified as an autocracy, but this was a recent change from democracy (in 2019), and "it is near the boundary [...] and could be reclassified as a democracy relatively easily, since it meets the two main conditions to be classified as a democracy and only barely failed the third" (336).
3. AI could create a democratic world government.[1633]

Arguments in favor of higher forecasts:

1. "Surveillance technology will make revolutions harder in autocracies" (339).
2. Governments will increasingly use misinformation to control their population.[1634]
3. "As climate change worsens, governments will justify emergency, anti-democratic measures."
4. An extinction or catastrophe could make "a 100% [of the world's population living in an autocracy] resolution more plausible than it might seem" (336).

[1630] 342, "He expect the countries under autocracy to decline, but they have larger birth rates, so in population terms there will be some balancing out."

[1631] 345, "The strongest arguments for the median forecast in all years being 70% centers around the belief that the world is mostly becoming more anti fragile. Although there was a trend toward many countries becoming more liberal we're seeing somewhat of a backslide possibly due to increase nationalism and increase in global conflict. However we mostly see the world stabilizing around its current state, with liberalization only occurring due to radical changes and pivotal world events."

[1632] 345, "[G]lobal destabilizing events like wars and pandemics will generally cause countries and governments to become more autocratic as they can seize greater control of their citizens. Governments will use chaos and panic to leverage more control in the name of safety and protection over populations."

[1633] 336, "It's possible a world government could form, perhaps in response to or as a direct result of transformative AI. Should this happen and this world government be democratic, this question could resolve at 0%."

[1634] 339, "The trend of governments using misinformation to control their population will increase."

Cross-references with other questions

Q56: Happiness in America

Q58: Future Worries and Children

Q59: Generation Attitudes

Question 58: Future Worries and Children

When will 50% of US adults (18-49) say they expect to have no children, or no *more* children, and cite a worry about the long-term future of the world or country as a primary reason?

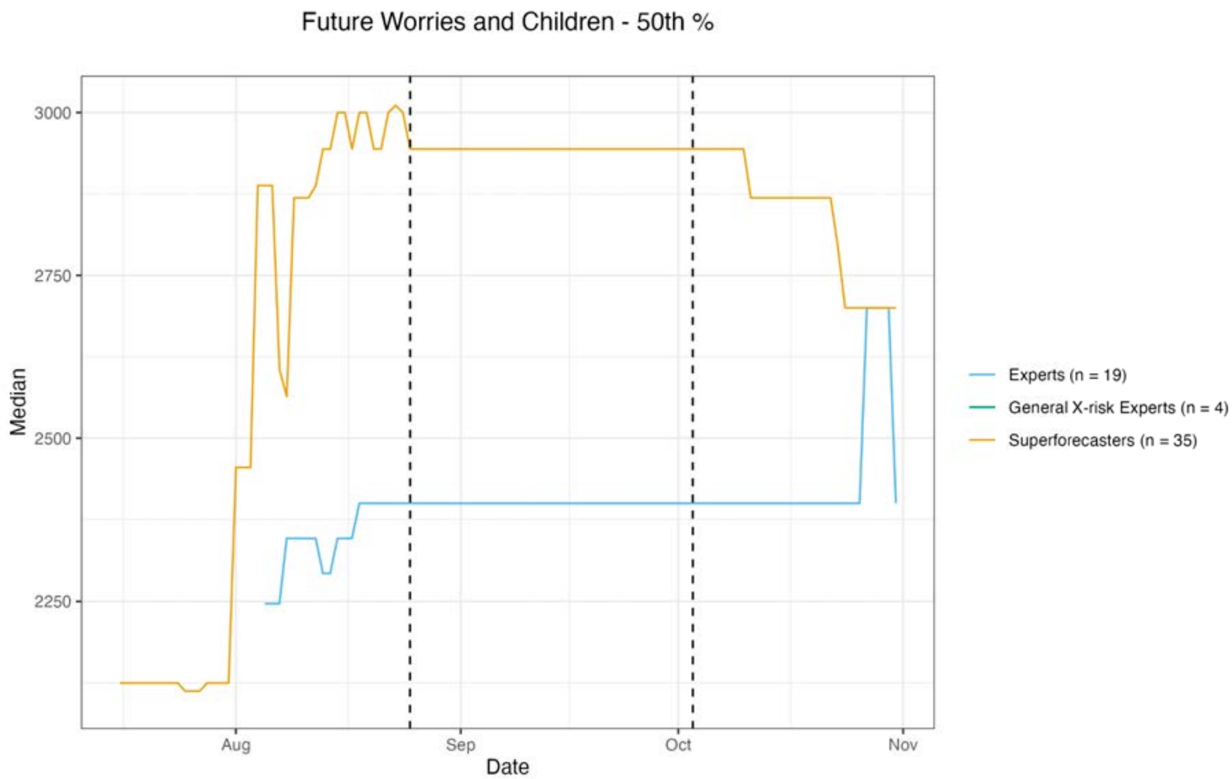
Question and resolution details, prior forecasts, and other relevant sources

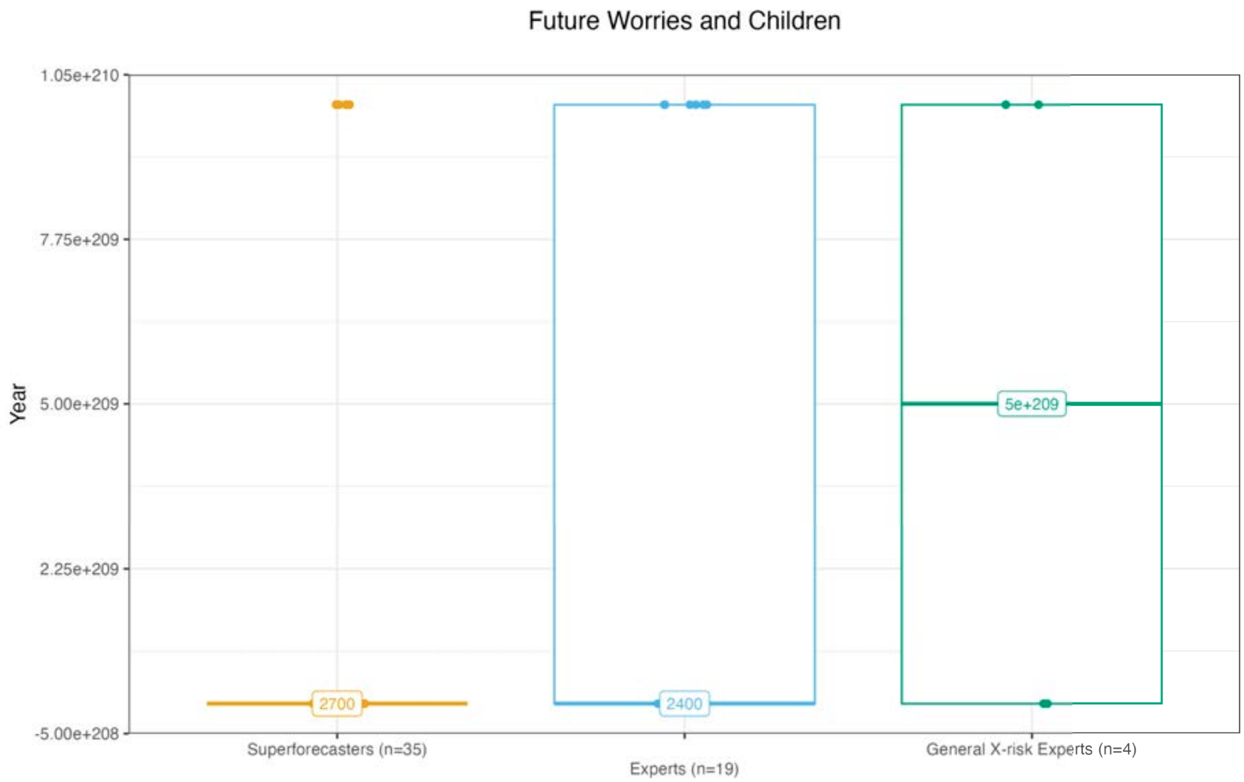
Results[1635]

Group	Percentile Forecast	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 35)	5%	2067.5	2070	131.14	Inf
	25%	2100	2200	2255.25	Inf
	50%	2125	2700	13193.76	Inf
	75%	2150	4022	82138.52	Inf
	95%	2200	10000	2.77e20	Inf
	5%	2040	2039	37.74	-44.37%

[1635] Numbers of forecasters are given as of Stage 4 of the XPT.

Experts (N = 19)	25%	2080	2080	Inf	n/a
	50%	2293	2400	Inf	n/a
	75%	2800	200000	Inf	n/a
	95%	1.0e210	1.0e210	Inf	n/a
General X-Risk Experts (N = 4)	5%	2040	2065	n/a	n/a
	25%	2100	2170	n/a	n/a
	50%	2200	5.0e209	n/a	n/a
	75%	3000	1.0e210	n/a	n/a
	95%	30000	1.0e210	n/a	n/a





Sources of agreement, disagreement and uncertainty

There was little agreement between the arguments of teams with lower and higher forecasts, though the arguments more often simply talked past each other rather than directly contradicting each other.

One exception to this was the question of the relevant base rate. Some teams approached establishing a base rate by determining the current rate of US adults who say they do not expect to have additional children due to concerns about the future. For example, Team 339 noted that it was "difficult" to settle on a single base rate, but cited several recent surveys that suggested current rates of desire not to have children to due to concerns about the future may be 10-20%, 30%, or 39%.[1636] Team 341 arrived at a base rate of 9% by combining multiple recent surveys.[1637] Other teams tried to determine how commonly >50% of adults did not want

[1636] Estimating a base rate here was difficult. [A forecaster] put the number at between 10 and 20%, however [another] pointed out that in the Morning Consult survey from 2020, 14% cited fears about climate change, whilst 20% cited fears about political / economic climate as primary reasons (there would probably be a degree of overlap between these two groups, and the true current estimate may be approximatey 30%). [A forecaster] also identified a survey from 2021 in Lancet Planetary Health (Hickman et al, 2021), which conducted a similar survey in the UK, where 39% of Britons responded to a similar question in the affirmative.

[1637] From "Growing share of childless adults in U.S. don't expect to ever have children," Nov 19, 2021: "44% of non-parents unlikely to have children - 14% gave worries about long-term future reasons: gives a current base rate of 6%. 74% of parents unlikely to have more children - 4% gave worries about long-term to have children in the past. For example, Team 337 wrote that the U.S. base rate "of people not wanting children" may be zero.[1638] Other teams (340, 344) did not mention a base rate at all. Team 338 wrote that "no one in the group was really able (or willing) to establish a 'base rate' for the question."

In general, base rates were referred to often, and high base rates correlated with a sooner predicted year of resolution, while low base rates correlated with a later predicted year of resolution.

Arguments given for forecasts ≤ 2100

- The base rate is already high. One team found a 2021 survey wherein 39% of "respondents across countries were 'hesitant to have children' because of climate change," and this influenced them greatly.[1639]
- Climate change will deteriorate the well-being of Americans.[1640] [1641]
- A catastrophic risk event, such as "Non-anthropogenic source, Nuclear confrontation, Novel pathogen, [or] Other Malthusian conditions leading to government regulation of amount of children allowed."[1642]
- The trend is in this direction, and should be expected to continue.[1643] [1644]
- Polling issues, such as the structure of the poll,[1645] stochastic variation in responses,[1646] and polling frequency,[1647] could cause a sooner resolution.

---

future reasons: gives a current base rate of 3%. Current rate is somewhere around 9% depending on share of non-parents and parents in the population."

[1638] "The historical U.S. base rate of people not wanting children may be zero and is at most 1 divided by the number of years the U.S. has existed."

[1639] 339, "The discovery of the Hickman et al, 2021 paper in Lancet Planetary Health, which commissioned a survey and indicated 39% of respondents across countries were 'hesitant to have children' because of climate change, was a strong argument for this question potentially resolving sooner." (2048-2100)

[1640] 344, "The main argument for the deterioration of well-being for this team is climate change. The significant impact of climate change on the majority of forecasters is supposed to affect the American population in the second half of the century." (2085)

[1641] 336, "Most people don't consider how hard the future could be. Climate change is likely to hit the world hard in the coming decades." (2045-2065)

[1642] 344, (2030)

[1643] 340, "[C]ited studies as showing Millennials being more likely to feel this way than their parents, and expecting this trend to continue with future generations." (2100)

[1644] 340, "One of the forecasts in this category [...] sees indications towards this feeling becoming more common." (2036-2050)

[1645] 340, "Poll structure has a significant impact on the resolution of this question, making it likely eventually." (2100)

[1646] 339, "[E]ven given non-dynamic public opinion on this issue, stochastic variation in the results of this poll could possibly lead to this question resolving within the next 10-20 years." (2032-2042)

[1647] 339, "[T]he frequency with which such surveys are run might increase as this issue increases in saliency, which it might reasonably be expected to do given central estimates of future population decline in the USA and other rich countries. Given that a range of results are possible at any given time due to a combination of polling error, sampling bias, and vagaries surrounding the exact way the question is

- Advanced AI will make people very “anxious”[1648] or “miserable.”[1649]

Arguments given for forecasts ≥ 2400

- The base rate is currently low, and very far from 50%.[1650] [1651] [1652] [1653]
- A “terrible crisis (that does not kill everyone) would probably have to take place for the numbers to go up.”[1654] However, even then, the odds are 50-50 whether people would then “not want children, or [would] want to repopulate the earth.”[1655]
- Wanting children may go down, but the *cause* of that is unlikely to be worry-about-the-future.[1656]
- History shows that people still have kids even during very bad times.[1657] [1658] [1659]
- Human biology acts against it.[1660] [1661]

---

worded, it is likely that an increasing frequency of surveys being conducted increases the chance of this question resolving.” (2048-2100)

[1648] 340, “Transformative AI will cause a period of very high human anxiety over how rapidly the world is changing.” (2036-2050)

[1649] 337, “AI, even properly aligned AI, might make us miserable, and therefore, we might not want more children. ‘Most people need to be working to some degree, whether it’s for money or not, to be happy. They need to feel as though they are making a contribution, and that that contribution is valued. Which brings me to AI: AI is not designed to make us happier by improving our lives. It’s designed to make a select group of people money even at the expense of making other people less happy. People are not going to be happier if the singularity occurs, AGI is able to perform all tasks better than humans, and as a result humans get to sit around writing poetry no one will read (because AI can do it better). That is a recipe for misery.’” (2048.5-2102.5)

[1650] 341, “The current survey values (Base Rate) are far from 50%.” (2400)

[1651] 41, “Starting from current baserate of approx. 10%, we would need a 2% increase in this share yearly for this to arrive in 40 years.” (4000-1e210)

[1652] 341, “The historical U.S. base rate of people not wanting children may be zero and is at most 1 divided by the number of years the U.S. has existed.” (2644)

[1653] 337, “If we take the previous 2 and 14 percent as representing an average of 8%, with a standard deviation of 6%.... well, 50% is still 7 sigmas away.” (6614-1e210)

[1654] 341, (2400)

[1655] 337, (2644)

[1656] 338, “For most people, this is a lifestyle choice that has nothing to do with the state of the world. The human mind seems designed to avoid worrying about the long-term future.” (6614-1e210)

[1657] 341, “Even during the darkest times, such as WW2 and the Great Depression, Americans were optimistic enough to start a family.” (2400)

[1658] 337, “ Birth rates did not decline compared to trend during calamities such as wars or the Great Depression.” (2644)

[1659] 338, “Historically it seems that [...] if the standard of living isn't very high, then people tend to have more kids.” (1e210)

[1660] 341, “Strong human reproductive drive.” (4000-1e210)

[1661] 337, “We have evolved to want children because this desire keeps our species from extinction. The instinct is base: we will always want more children, regardless of external circumstances. Further, those who have kids will pass on their attitudes to their children.” (6614-1e210)

- Religious beliefs often act against it,[1662] and that will be self-perpetuating, since people with those religious beliefs will have more children.[1663]
- If circumstances are dire enough for >50% to have this belief, it is unlikely surveys will be being conducted.[1664]

Other arguments given

Arguments in favor of lower forecasts:

- People may cite worry-about-the-future more than they actually believe it, because it gives them a reason that “makes sense to them and their peers.”[1665]

Arguments in favor of higher forecasts:

- “Space colonization and opportunity/need for additional children on settlements might also argue for a higher value (assuming such colonies are considered part of the US).”[1666]

Cross-references with other questions

Q56: Happiness in America

Q57: Prevalence of Autocracies

Q59: Generation Attitudes

Question 59: Generation Attitudes

When will 90% of Americans say they wish they had been born in a previous generation?

Question and resolution details, prior forecasts, and other relevant sources

---

[1662] 341, “[H]istorically people keep on producing throughout even catastrophic conditions, often paired with religious beliefs about the afterlife.” (2400)

[1663] 344, “[O]ne of the major determinants is the religious beliefs of the population, which are at least partly heritable. In the equilibrium where secular people have fewer children, a majority of the children born into the future will be religious, in which case they will also desire more children.” (20,000-1e210)

[1664] 337, “A catastrophe so extreme that people don’t want more children is so extreme that people won’t be taking surveys. By the time the U.S. recovers enough that surveys might be taken, opinions on children will recover, too.” (6614-1e210)

[1665] 343, “Whether the prompt will resolve depends on whether/when people will *cite* such reasons, not when they will actually be moved by them. Fertility seems to be going down robustly, and people will seek to explain their behavior in ways to make sense to them and their peers, so it seems not implausible that the tendency for this to be named as a reason for not having any (more) children will increase.” (2142)

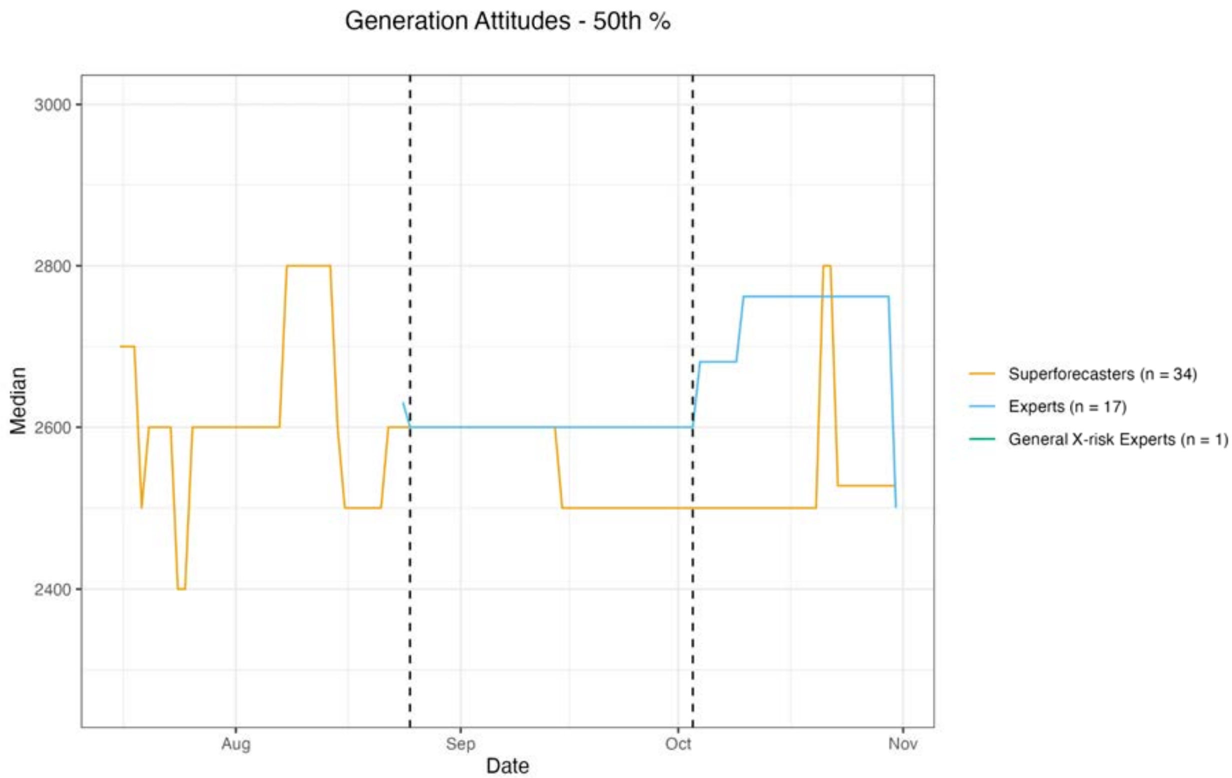
[1666] 344, (2070-1e210)

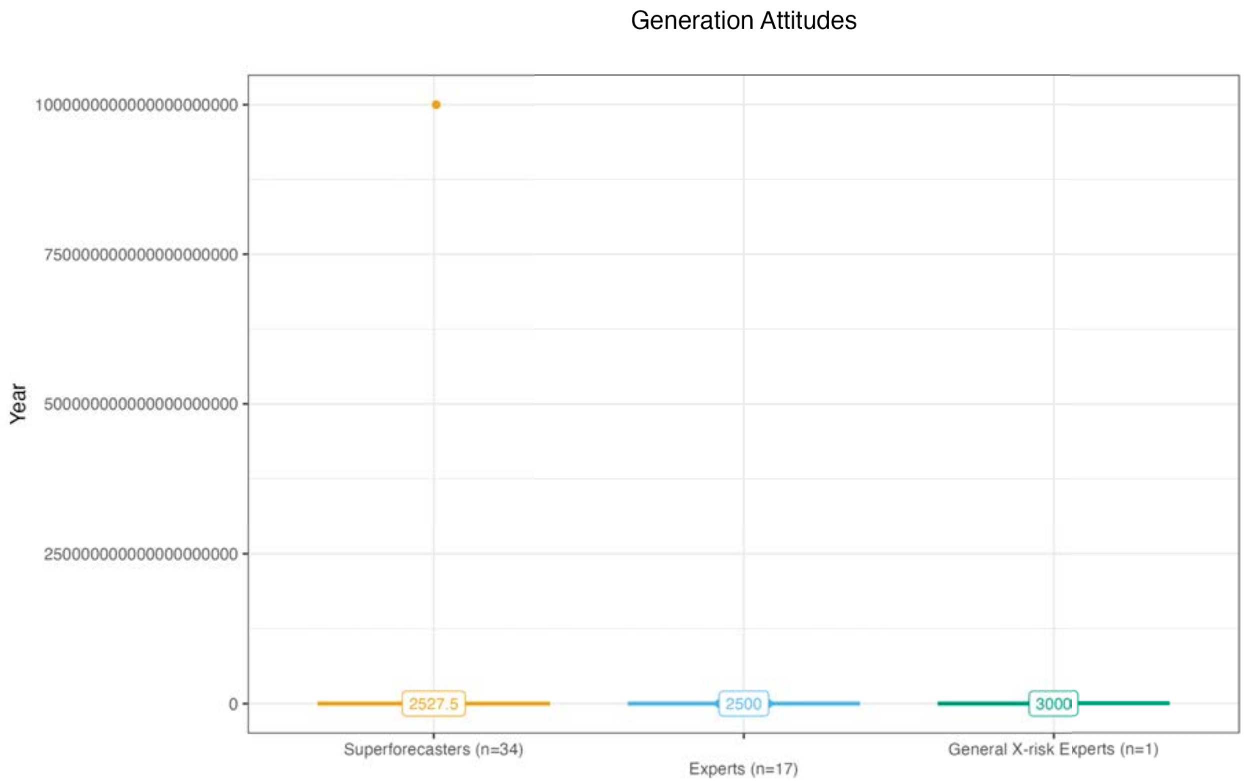
Results[1667]

Group	Percentile Forecast	Stage 1 Median	Stage 4 Median	Stage 1 Standard Deviation	Percent Change in Standard Deviation, Stage 1 to Stage 4
Super-Forecasters (N = 34)	5%	2072	2050	Inf	NaN
	25%	2112	2140	Inf	NaN
	50%	2250	2527.5	Inf	NaN
	75%	2365	4260.5	Inf	NaN
	95%	3000	9500	Inf	NaN
Experts (N=17)	5%	2035	2040	63.3	+21.03%
	25%	2100	2100	Inf	n/a
	50%	2300	2500	Inf	n/a
	75%	3201	4000	Inf	n/a
	95%	10000	5e209	Inf	n/a
General X-Risk Experts (N = 1)	5%	2040	n/a[1668]	n/a	n/a
	25%	1.0e210	2083	n/a	n/a
	50%	1.0e210	3000	n/a	n/a

[1667] Numbers of forecasters are given as of Stage 4 of the XPT.
[1668] The forecaster who submitted a forecast in the first stage later set their 5th percentile forecast to the default; we exclude default forecasts from our analysis.

	75%	1.0e210	3750	n/a	n/a
	95%	1.0e210	7000	n/a	n/a





Sources of agreement, disagreement and uncertainty

- There was disagreement on whether the world in general, and the US in particular, would get bad enough to cause the question to resolve anytime soon.
- Teams with lower forecasts (i.e. for more distant years) cited many psychological reasons why the number of Americans who say they wish they had been born in a previous generation number would ever reach 90%, while teams with higher forecasts had more mechanistic views of the trend rising, mostly based on the Resolution Foundation survey (p19).

Arguments given for forecasts ≥ 2400

Psychological Reasons

- "The familiarity heuristic was used as an argument for people not wanting to be in a different time/place" (337).
- Even if it was certain that an existential event was imminent, "a lot of other people [...] would choose to not believe the evidence, and [would believe] that something would happen that would save them (God?, Superman? Elon Musk?)" (337).
- More generally, if there was a large catastrophe, "a decent subpopulation of people may not express this sentiment" (336).
- "If it's really so bad that no one wants to be alive in the present time, some people will wish they were born in the future rather than the past. 50% or 60% would be plausible, but 90% is far-fetched" (340).

- “The idea that your own generation's values and lifestyles are superior is more of an innate, biological thing” (340).
- “This will either never happen or will happen under the most dire conditions, such as near extinction. 90% is too high a threshold. Such is the human nature” (340).

90% is a very high bar; there is no precedent for it; and it is unlikely
- 90% is a very high bar and as such is extremely unlikely, since it would imply “continuous extreme deprivation and/or calamity, or complete social stagnation” of a kind that has never existed in the past “300+” years (339).
- The US will continue “to be a dominant economic and political power,” and there’s a “low probability that deprivation and/or calamity will affect >90% of Americans” (339).

The US is well-positioned for the future
- The US specifically has good prospects[1669] for prosperity and happiness. This is because:
  - Its geography: it being “surrounded by two oceans, [its] land borders with two agreeable neighbors[,] and a vast, temperate interior hinterland,” all serve to “relatively [insulate] the USA from the worst possible foreseeable deprivation (eg, rising sea levels, famine, peer-state warfare)” (339).
  - Americans’ lives will continue to be good, or improve, due to improvements in “ease-of-living,” medicine, and digital entertainment,[1670] in addition to overall “growth, development, and technological advancement” (342).
  - There is no “reasonably substantial threat to the Americans now or in the foreseeable future earlier than year 2102” (342).
  - There’s a low probability of a catastrophic event, or an event “traumatic enough to make people believe better days are in the past,” such as a “major nuclear exchange,” or “the fall of the American Republic to an autocratic regime” (343)

Outcomes bad enough to cause this are unlikely globally because living conditions are better now than in the past, and will continue to get better
- People would have to be materially worse-off than previous generations, and this is unlikely to happen, given that “exponential growth in technology and innovation has proceeded for over 100 years” (343).
- The past was not in fact better, and, similarly, the future will be better than now. Evidence for this is “growing wealth, better medical care, fewer serious wars having been directly involved in etc. in the past century” (336).
- Similarly, “the broad trajectory of history has been the greater production of food and consumer goods which prevents widespread deprivation” (339).

[1669] 336, “The good perspective for the US makes it difficult to reach 90% of Americans saying that they wish they had been born in a previous generation.”
[1670] 342, The US “is well-positioned to be a reasonably satisfactory place to live for the foreseeable likely future. Combined with increasingly satisfying digital entertainment and long run typical improvements in medicine and ease-of-living.”

- Despite it being “likely” that bad things will happen, such as famines, life in the modern era generally gets better for most people, whereas life in the past was not good for a lot of people.[1671]

The probability of this is highly linked to a very serious catastrophe

- Something very dramatic, such as “a near collapse of the economy, ecology and/or government,” (336) would have to occur to read 90% on this question, and events like that are very unlikely.
- However, one can combine “all the probabilities of these x-risk events” (336) to get an estimate of something like this happening.

Technicalities

- The US is “likely to cease to exist before this happened” (337).
- “Many AI scenarios to result in outright extinction rather than leave a population around to regret the current time” (336).
- Similarly, if there was a large catastrophe, there “wouldn't really [be] infrastructure for surveys” (336).

Arguments given for forecasts ≤ 2090

- Dissatisfaction with the present is rising: per an Intergenerational Commission survey, while 25% of British Gen Xers wished to have grown up in a different era, that percentage had gone up to 33% among Millennials.
- This dissatisfaction is expected to get worse for future generations, due to “rising inequality, climate change issues,[1672] general political issues, decrease in general happiness because of over-stimulation by news and social media” (345). Team 341 added that “life in America will become objectively worse this century than in the recent past- that factors such as economic stagnation, climate change, and a more violent world will cause the standard of living today (or in the very near future) to be obviously superior to some point in the 21st century.”
- Since this dissatisfaction (i.e., the “preference for being born in a previous generation”) is expressed by significant numbers of people already, when it is objectively “at best, ambiguous” that the present is worse than the past, we should expect to see the

[1671] 337, “The argument for people wishing they had been born in a different generation was that changes were likely to happen that people did not necessaries want to happen e.g. transport might not include the ability for people to drive private cars, famines might happen. But this rather ignored that other changes had occurred that, on the whole, made like life better for most people. e.g. the internet, improved medical treatment. It was pointed out that 90% wishing they lived in the past was a very high bar. Few previous generations have lived their whole lives without a serious risk of famine, war or becoming ill from a condition that is treatable now. And it was pointed out that in the past most people were not in the most privileged section of society.”

[1672] Climate change was also emphasized by Team 344: “The main argument for the deterioration of well-being for this team is climate change. The significant impact of climate change on the majority of forecasters is supposed to affect the American population in the second half of the century.”

percentage of people who express that preference grow to 90% “if it does become clearly true” (341).

- A catastrophic risk event—due to either a “non-anthropogenic source,” “nuclear confrontation,” or “novel pathogen”—is a reason to expect this to happen by 2026 (344).

Other arguments given

Arguments in favor of lower forecasts (i.e., further in the future):

- “Mood-enhancing drugs would influence the sentiment in such a case that this question will never resolve while humanity is in an "objectively" bad situation” (344).
- “Metaverse scenarios would influence the sentiment in a positive way, similar to mood-enhancing drugs” (344).

Cross-references with other questions

Q56: Happiness in America
Q58: Future Worries and Children