Advanced AI according to 1,580 researchers: uncertain, unsafe, and sooner than we thought

Katja Grace, Julius Simonelli, Benjamin Weinstein-Raun, David Krueger, Nathan Young, Tim Duffy, Tom Adamczewski, Joseph Reeve, and Alexander Beck

September 2026

Abstract

In December 2024, we ran a survey about the future of AI, reaching 1,580 researchers internationally who had published at any of six leading artificial intelligence (AI) venues, corresponding to a 10% response rate. Respondents expressed substantial uncertainty about AI's consequences and timelines. This was the fourth iteration of the Expert Survey on Progress in AI extending a survey series that began in 2016. Significantly, timelines to human-level AI performance have shortened markedly between iterations of the survey, while researchers' deep concerns around the consequences of advanced AI have persisted or grown. In this survey, researchers assigned an 18% chance on average of future AI advances causing human extinction or similarly permanent and severe disempowerment of the species, and most researchers assigned at least a 10% chance to such outcomes. 72% favored greater prioritization of research aimed at minimizing AI risks. Among eleven different risks, researchers would direct the most concern toward AI making it easy to spread false information such as deepfakes, with 83% saying that deserved substantial or extreme concern over the next thirty years. Large majorities also said such concern is warranted by AI systems manipulating large-scale public opinion, AI enabling dangerous groups to develop powerful tools such as engineered viruses, authoritarian rulers using AI to control their populations, and AI worsening economic inequality.

1 Introduction

We report on the fourth iteration of the Expert Survey on Progress in AI (ESPAI): a survey that has been run periodically since 2016 with minimal modifications (Grace et al., 2018; Stein-Perlman et al., 2022; Grace et al., 2025). To our knowledge, this is the longest-running series of surveys of AI researchers, and the previous iteration, conducted in 2023, was the largest peer-reviewed survey of AI researchers ever (Grace et al., 2025). The surveys have been widely cited, including by TIME, Nature Briefing, The New York Times and The Economist (AI Impacts, 2024).

Previous iterations of the survey found highly uncertain timelines for human-level AI performance. We operationalize human-level AI performance in two ways—'high-level machine intelligence' (HLMI) and 'full automation of labor' (FAOL)—and the mean aggregate forecasts for both have assigned non-negligible probabilities to them occurring within ten years, and to them not having occurred within two centuries.

Across the survey series, timelines to human-level performance have repeatedly shortened faster than the passage of time. For instance, in the mean aggregate forecast, the year by which HLMI has a 50% chance of arriving moved up from 2061 in 2016 to 2059[1] (2022), 2047 (2023), and 2042 (2024). Over the course of eight years, the forecast horizon shrank from 45 years out to 18, getting about 3.4 years closer for each year that passed.

[1]Reported as 2060 in 2023; small changes to code and data cleaning between surveys shifted the aggregate slightly.

Meanwhile, the profile of perceived risks has remained stable. Since 2016, the median participant has placed a 5% chance on extremely bad long-run impacts on humanity if HLMI is built. And since 2022 (when we added the questions), the median participant has put at least 5% on human extinction or similarly permanent and severe disempowerment. In this survey, the median chance given to extinction or similar disempowerment in some form was 10%, with chances shifting slightly upward since the previous survey across the distribution (see Figure 1b).

Since the first survey in 2016, AI has gone from a relatively niche topic to powering the most valuable companies in the world. Public discussion of concerns about risk from AI has also exploded. In 2016 the non-negligible chances given to extremely negative impacts from advanced AI were a surprising finding, whereas by 2024 many highly prominent AI researchers had publicly signed the Center for AI Safety's Statement on AI Risk, stating, "Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war" (Center for AI Safety, 2023). So substantial concerns about AI risks among AI researchers are no longer so surprising, but they remain alarming, and this survey serves as evidence that these concerns are shared by a majority of AI researchers, not just a prominent minority.

At the same time, human-level AI, AI risks, and the risk of human extinction in particular remain somewhat controversial among researchers. For this reason, over the years this survey has taken particular care to reduce non-response bias and to assess the available evidence about whether non-response bias is likely to explain our principal qualitative findings, discussed in Appendix A.

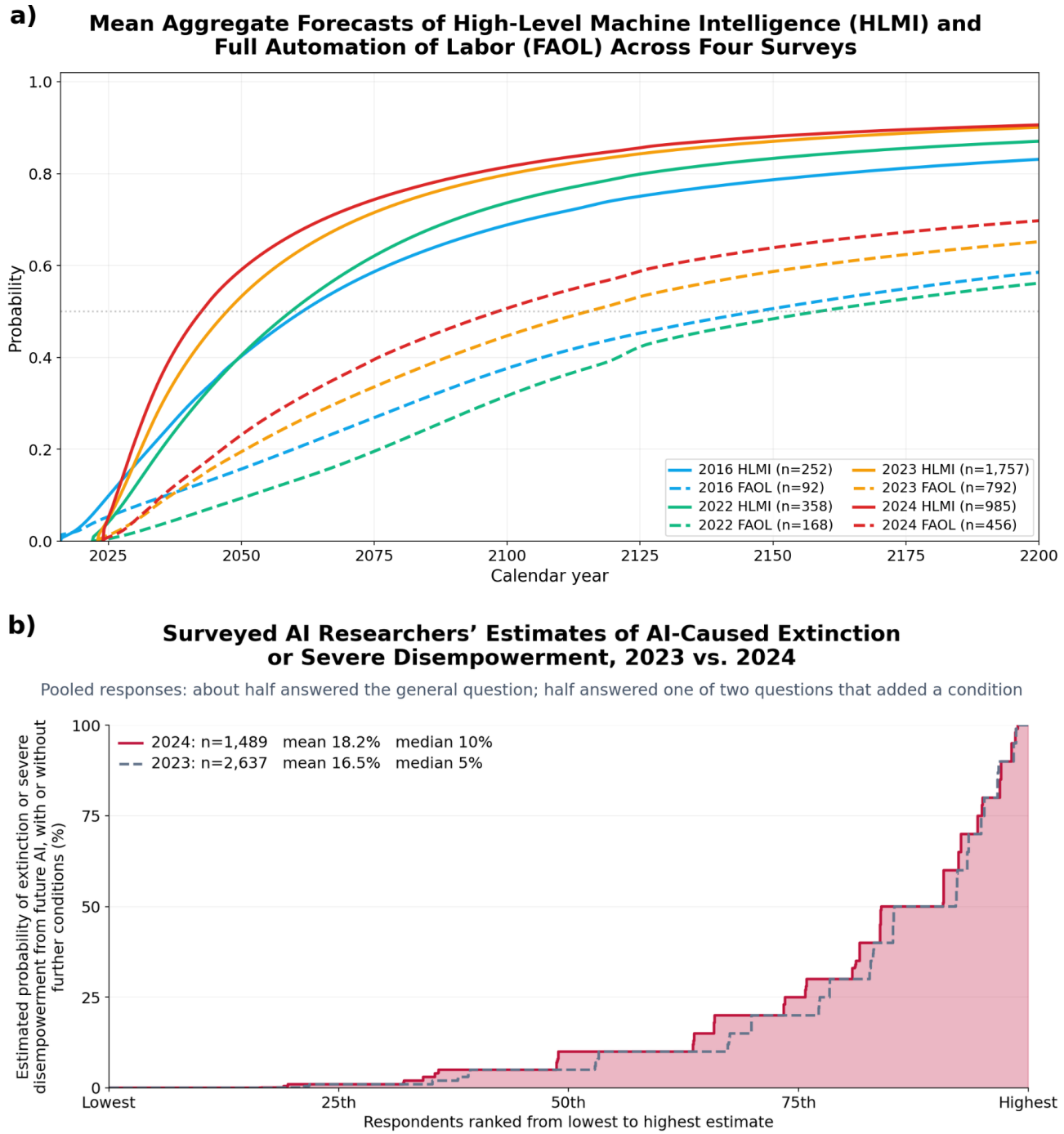


Figure 1: ***Notable results: a) forecast time to human-level AI performance continues to fall many years per year; b) 51% of researchers assign at least a 10% chance to human extinction or similarly permanent and severe disempowerment from advanced AI, with perceived risk inching upward since the last survey.*** *a) Each line represents a mean aggregate forecast for one of two framings of human-level AI performance in one of four surveys. See Section 3.2 for more details. b) Each vertical slice shows one respondent's probability of extinction or similarly permanent and severe disempowerment, with or without further conditions, with participants sorted by increasing pessimism. See Section 3.9 for further details.*

2 Methods

We largely followed the 2023 survey's data collection and analysis methods. The differences are listed in Appendix B.

2.1 Recruitment

In December 2024, we recruited participants who had published in 2023 at any of the following six top-tier AI venues:

- Conference on Neural Information Processing Systems (NeurIPS)
- International Conference on Machine Learning (ICML)
- International Conference on Learning Representations (ICLR)
- Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence (AAAI)
- Journal of Machine Learning Research (JMLR)
- International Joint Conference on Artificial Intelligence (IJCAI)

We sent emails to 19,874 addresses[2] and received 2,052 complete or partial responses for a response rate of 10%. We included submissions with at least one answered question; submissions that did not reach the final question were classified as partial.

After fielding the survey, we discovered that our collection process had erroneously included people outside of the intended group (mostly researchers publishing in workshops or venues connected with the intended venues). We thus excluded these participants from the main analysis. This left us with 1,580 respondents for the main analysis, with 1,502 reaching the final question and 78 dropping out earlier.

We collected the responses through the Qualtrics survey platform. The time frame for submitting responses was between December 9, 2024 and December 24, 2024. Participants were offered a choice between gift cards for $50 worth of purchases, money, or charitable donations.[3]

2.2 Question types

The full text of the survey is available in Appendix C.

Most primary questions in the survey solicited responses in one of three ways: on an ordered rating scale (see Figure 2); as a probability estimate; or as an estimate of a future year. A smaller number of primary questions asked for write-in responses or numerical estimates. For most primary questions, a random subset of participants received secondary questions asking for optional write-in responses about their interpretation of the question and which considerations were important in their answer.

2.3 Definitions and phrasing

As in 2023, we provided the following clarifications for ambiguous terms:

[2]Excluding those that bounced or were recognized by the survey platform as having failed before that, but including those that possibly weren't received due to spam filters for instance. Our understanding is that failed emails could include those of people who have previously unsubscribed from receiving emails from us with Qualtrics, and therefore might not be an arbitrary subset with respect to opinion on the survey.

[3]Available options differed by the participant's country, with a small number of participants from certain countries unable to receive any reward due to regulation.

Figure 2: ***A survey excerpt illustrating an ordered rating scale.***

- *High-level machine intelligence (HLMI)* was described as occurring "when unaided machines can accomplish every task better and more cheaply than human workers. Ignore aspects of tasks for which being a human is intrinsically advantageous, e.g. being accepted as a jury member. *Think feasibility, not adoption.*"
- An occupation was described as *fully automatable* "when unaided machines can accomplish it better and more cheaply than human workers. Ignore aspects of occupations for which being a human is intrinsically advantageous, e.g. being accepted as a jury member. *Think feasibility, not adoption.*"
- Correspondingly, *full automation of labor (FAOL)* was described as occurring "when all occupations are fully automatable. That is, when for any occupation, machines could be built to carry out the task better and more cheaply than human workers."
- *AI safety research* was described as "any AI-related research that, rather than being primarily aimed at improving the capabilities of AI systems, is instead primarily aimed at minimizing potential risks of AI systems (beyond what is already accomplished for those goals by increasing AI system capabilities)." Examples given for what AI safety research might include were:
  - Improving the human-interpretability of machine learning algorithms for the purpose of improving the safety and robustness of AI systems, not focused on improving AI capabilities
  - Research on long-term existential risks from AI systems
  - AI-specific formal verification research
  - Developing methodologies to identify, measure, and mitigate biases in AI models to ensure

fair and ethical decision-making.[4]

– Policy research about how to maximize the public benefits of AI

2.4 Framing effects

The way a question is framed can greatly influence the response (Tversky and Kahneman, 1981). To measure and control for framing effects, we posed different versions of several questions on the same topics to different randomized subsets of respondents.

All questions about how soon a milestone would be reached were framed in two ways: half of respondents were asked to estimate the probability that a milestone would be reached by a given year ("fixed-years framing"), while the other half were asked the year by which they estimated that the milestone would be reached with a specific probability ("fixed-probabilities framing"). To minimize confusion, each participant received the same framing throughout the survey. We have observed consistently different results from these framings across all editions of this survey, with the fixed-years framing producing generally later predictions.

Other topics with questions utilizing different framings include human-level AI, productivity gains from AI, and catastrophic outcomes.

A comparison of framing effects for key questions about human-level AI performance is in Appendix D.

2.5 Randomization of question blocks

In order to limit the length of the survey for each respondent, we randomly assigned participants to receive different subsets of the questions. They received both different blocks of related questions and sometimes different questions within a block. Figure 3 shows a schematic diagram of the randomization process.

[4]Half of respondents shown this definition received the additional AI-ethics example; half did not. A version with it was added to address a concern that the absence of examples from the field of AI ethics within this set of examples might suggest that it was intentionally excluded (then two versions were retained, to allow for continuity of comparison over time).

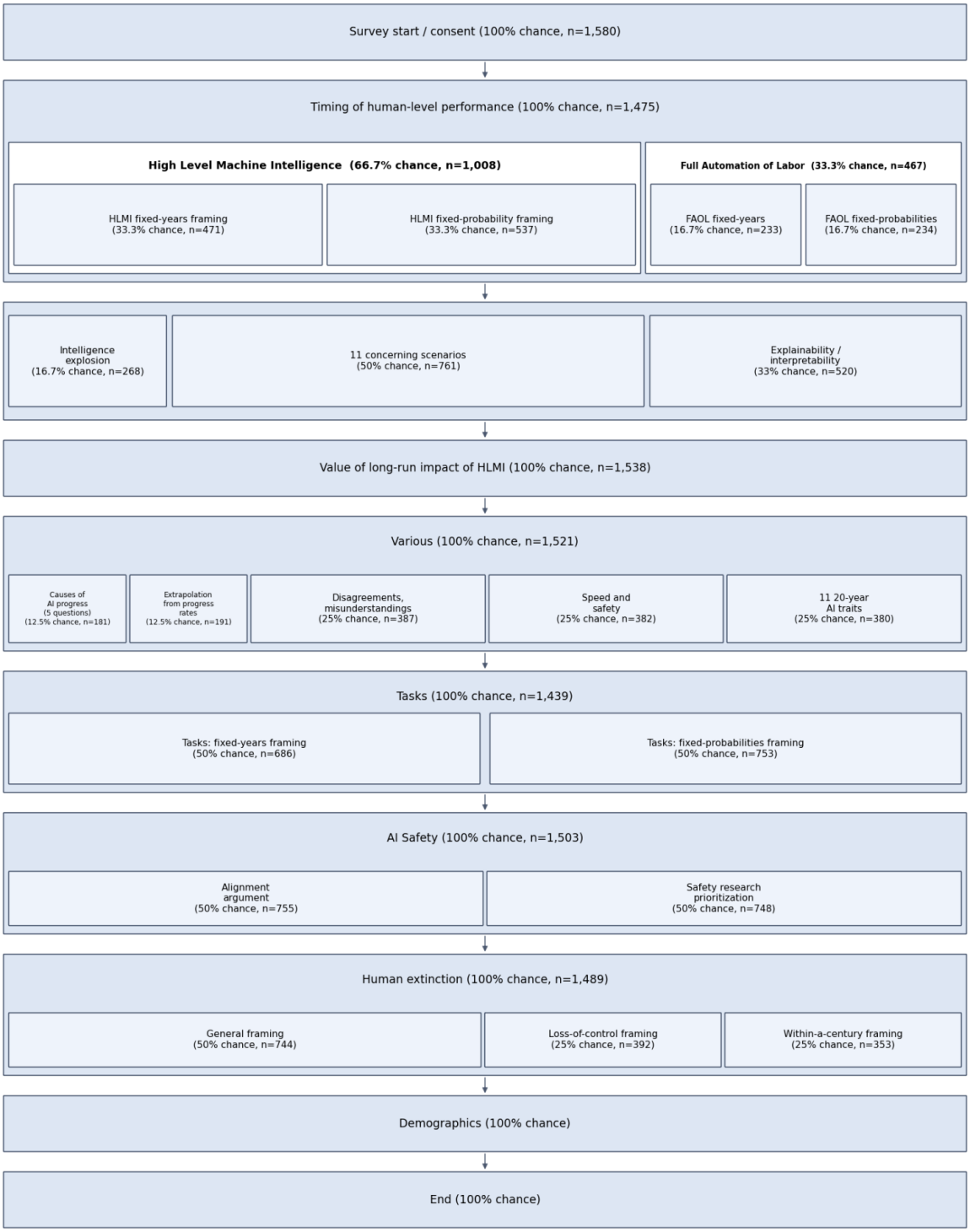


Figure 3: ***Randomized flow of participants through the survey.*** *Participants move from top to bottom, receiving one block of questions from among those in each horizontal set. Percentages indicate the chance of a participant receiving each block. Some blocks contain further randomization of primary questions within them that isn't shown, and most contain randomized insertion of secondary follow-up questions. Figures given for n are the number of valid responses to the question.*

2.6 Data cleaning

Our data cleaning was similar, but not identical, to that for the previous survey (Grace et al., 2025).

One substantial element was dealing with non-numeric answers given to quantitative questions (e.g. 'infinity', 'ten to twenty years'), since in most cases we allowed non-numeric inputs. We often converted those to numeric answers, where a clear interpretation was available and a programmatic parsing rule could handle that response format (e.g. changing '~40' to '40'), but excluded those that were too ambiguous or in a format our rules didn't cover. We converted answers such as 'never' and 'infinite' to 100 million years, to include them in analysis with approximately the intended implications (near-zero probability at every point within the coming centuries).

We removed 331 question responses across 111 participants where an event was given a higher probability of having occurred by an earlier date than by a later date. In most of these cases (243), the probabilities given at the three time horizons also added to 100%, suggesting these participants' misunderstanding of the question involved treating the probabilities as shares that must total 100%, rather than as cumulative probabilities at each time horizon. We further removed 37 responses that added to 100% but were not descending, where the participant had another answer that added to 100% and was descending, implying this type of misunderstanding. We retained 113 question responses that summed to 100% but where misunderstanding could not be confirmed via descending probabilities.

2.7 Analysis of timeline responses

We combined responses from fixed-years and fixed-probabilities framings described in Section 2.4 using gamma[5] mixture CDF aggregation, following the methodology of the 2023 survey report (Grace et al., 2025). This means we fit a gamma CDF to each respondent's three data points (where valid and indicating timelines in the future), then, at each time point, took both the arithmetic mean and the median of the fitted cumulative probabilities, producing a mean aggregate CDF and a median aggregate CDF from which to read percentiles.[6]

This approach permits extrapolation beyond the elicited horizons, and respondents expecting very late AI developments remain represented through fitted CDFs with long right tails rather than being dropped. It also treats both question framings symmetrically. Appendix E compares alternate aggregation methods.

Our approach does not permit including negative numbers of years in the aggregate, which is the natural way for participants to indicate that they think a milestone has already become feasible for AI. Thus our forecasts are made for the subset of people who say a level of AI performance is yet to become feasible. In this survey, this makes little difference, as only one person tried to enter negative numbers of years.

3 Results

We outline a selection of key results from the survey.

3.1 Time until tasks can be automated

All participants were assigned task-forecast questions, in one of the two framings as described in Section 2.4. Participants in the fixed-probabilities framing saw the following wording:

> How many years until you think the following AI tasks will be feasible with:

[5]A gamma distribution is a flexible, right-skewed probability distribution over positive values. Here, it is used as a smooth parametric approximation to each respondent's implied distribution over the timing of AI development.

[6]Small code changes between surveys sometimes shift the fitted curves slightly.

- a small chance (10%)?
- an even chance (50%)?
- a high chance (90%)?

  Let a task be 'feasible' if one of the best resourced labs could implement it in less than a year if they chose to. Ignore the question of whether they would choose to.

Participants in the fixed-years framing were instead asked how likely each task was to be feasible within 10, 20, and 50 years.

We then gave each participant a random four out of 39 task descriptions (per task, n=111-167, pooling both framings and aggregated as described in Section 2.7). Full descriptions of the tasks are available online: https://tinyurl.com/aitasks. Here are four example tasks:

> "Given a one-sentence description of the task, download and finetune an existing open source LLM, without more input from humans. The fine-tune must improve the performance of the LLM on some predetermined benchmark metric."
>
> "Routinely and autonomously prove mathematical theorems that are publishable in top mathematics journals today, including generating the theorems to prove."
>
> "Fold laundry as well and as fast as the median human clothing store employee."
>
> "Given a one-sentence description of the task and given the same information you would give a human to perform this task (such as information about the house), physically install the electrical wiring in a new home, without more input from humans."

Figure 4 shows mean aggregate forecasts for when these 39 tasks will become feasible with AI, alongside forecasts for the four occupations discussed in Section 3.2 and for HLMI and FAOL, all generated using the method described in Section 2.7 above. (Task labels are not full descriptions as seen by participants; these are linked in Appendix C.) Appendix F lists the years assigned 10%, 50%, and 90% probability for the 39 tasks.

Of the 39 tasks, 34 were forecast to have a roughly even chance of becoming feasible within 10 years (by 2034). These included independently building a payment processing site, generating fiction worthy of The New York Times Best Seller list, generating songs worthy of the US Top 40, providing high-quality phone banking services, independently fine-tuning open-source LLMs, adeptly folding laundry, replicating machine learning research, and racing through a city as a bipedal robot.

The five with later dates were all within mathematics, science, or engineering domains: discovering physics equations from simulation (2035), conducting ML research and writing a conference paper (2036), routinely generating and proving publishable math theorems (2039), installing electrical wiring in a new home (2040), and solving a long-standing math problem without human input, e.g., a Millennium Prize Problem (2053)[7].

According to forecasts, most tasks could also happen very soon: for all but one task, the forecast reached a 10% probability of becoming feasible by 2027 at the latest. The one remaining forecast, which reached a 10% chance in 2028, was for solving a long-standing math problem without human input.

We retained all task questions across survey iterations rather than removing tasks we judged already feasible. This judgment is often difficult, and retaining the questions lets us see how participants' answers change as capabilities develop. If participants believe a task has already become feasible, they can

[7]The task specifies 'without more input from humans', which would exclude the proposed September 2026 Navier-Stokes solution from satisfying this (OpenAI, 2026).

respond with zero years, a negative number of years, or a comment to that effect. (Zero years would be included in the aggregate, while the others would not.) However, even when estimating when a task would have a 10% chance of becoming feasible, no task received an answer of zero years from a majority of respondents; the highest proportion was 32% for the high-school history essay task. No one entered a negative value, although one person wrote that they would like to.

This suggests that participants might instead infer from the inclusion of a task that we are referring to a level of performance that has not yet been achieved, therefore imagining a different task than the one that participants imagined when the task was first included in the survey. This complicates interpretation of answers for some tasks assigned very short timelines, where it is plausible that the task is already feasible.

Forecasts for the 39 tasks were largely stable: 33 moved by less than two years in either direction. The largest exceptions were proving publishable mathematical theorems and autonomously conducting machine learning research, which both moved roughly seven years earlier.

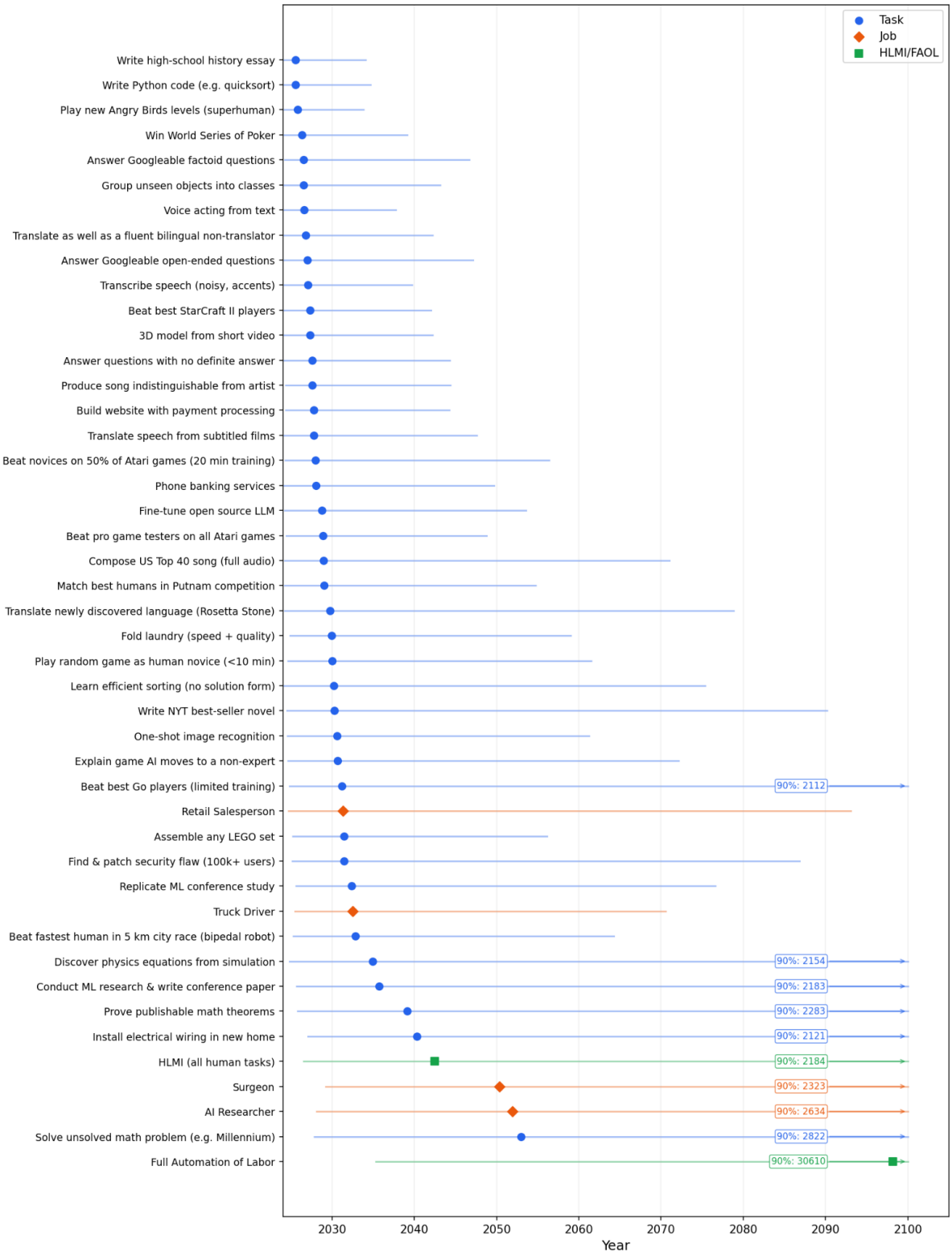


Figure 4: ***Mean aggregate forecasts of when AI milestones will become feasible (labels paraphrased).*** *Markers show the year with a 50% probability of the milestone having been reached, from the mean aggregate distribution; horizontal lines show the 10%–90% range. Milestones shown are from several imperfectly comparable question types. Task labels are shortened; full task descriptions include further details, available in Appendix C. This figure corresponds to Figure 1 in the 2023 survey report* (Grace et al., 2025).

3.2 Timing of human-level AI and automation of occupations

We asked participants two substantially different questions about when they expected human-level performance, focused on a definition in terms of either 'high-level machine intelligence' (HLMI) or 'full automation of labor' (FAOL) (see definitions above in Section 2.3). In both cases, participants had a fifty-fifty chance of receiving the question in a fixed-years framing or a fixed-probabilities framing (see Section 2.4 above). So for instance, participants had a one-third chance (n=537) of receiving this question:

> Say we have '**high-level machine intelligence**' when unaided machines can accomplish every task better and more cheaply than human workers. Ignore aspects of tasks for which being a human is intrinsically advantageous, e.g. being accepted as a jury member. *Think feasibility, not adoption.*
>
> For the purposes of this question, assume that human scientific activity continues without major negative disruption.
>
> How many years until you expect:
>
> a 10% probability of HLMI existing? _____ years
>
> a 50% probability of HLMI existing? _____ years
>
> a 90% probability of HLMI existing? _____ years

Another approximate third (n=471) saw the same question but in the fixed-years framing.

The FAOL question block was more involved, asking the participant to first make forecasts about when four specific occupations would become fully automatable, then to provide an existing human occupation that they expect to be among the final ones to be fully automatable, then to forecast that, and only then to forecast FAOL—the point where all occupations are fully automatable. (See Appendix C for a link to the full sequence of questions.) Participants also received this in fixed-years (n=233) or fixed-probabilities (n=234) formats.

As detailed in Section 2.7, we fit a gamma distribution to each valid response set. We then averaged the fitted distributions from both fixed-years and fixed-probabilities framings together, for each of HLMI and FAOL. Figure 5 shows the resulting mean aggregate distributions. It also shows median aggregate distributions: those resulting from the same process except showing the median probability assigned for each year by the set of individual fitted gamma distributions.

Figure 6 shows how the forecasts have changed over subsequent surveys.

In the mean aggregate forecast, HLMI reaches a 10% chance in 2027 and a 50% chance in 2042 (18 years from the survey). The latter represents a five-year drop since the 2023 survey a year earlier, following a 12-year drop over the year between previous surveys. The mean aggregate forecast for FAOL reaches a 10% chance in 2035 and a 50% chance in 2098 (74 years after the survey). The latter is a 17-year drop from the previous survey a year earlier (2115[8]), itself a 43-year drop from 2158 in the 2022 survey.

[8]Reported as 2116 in 2023; small changes to code and data cleaning between surveys shifted the aggregate slightly.

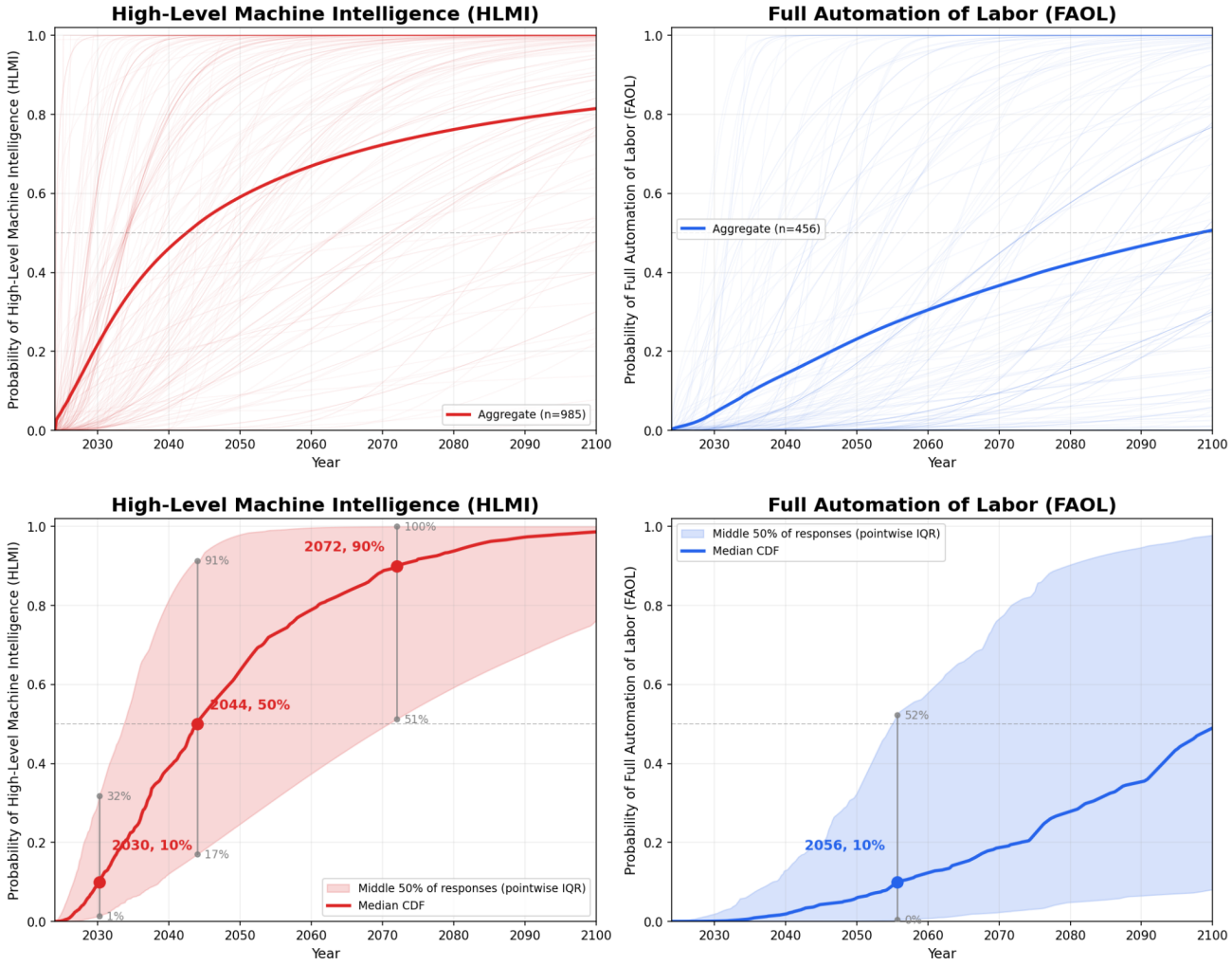


Figure 5: ***Aggregate probabilities of high-level machine intelligence (HLMI) and full automation of labor (FAOL) arriving by a particular year.*** *Thick lines represent the mean (top charts) or median (bottom charts) CDF over all respondents who gave valid future answers to the question; thin lines (top charts) show a random subset of 200 of these individual respondent gamma CDFs; shaded regions (bottom charts) show the central 50% of these individual CDFs.*

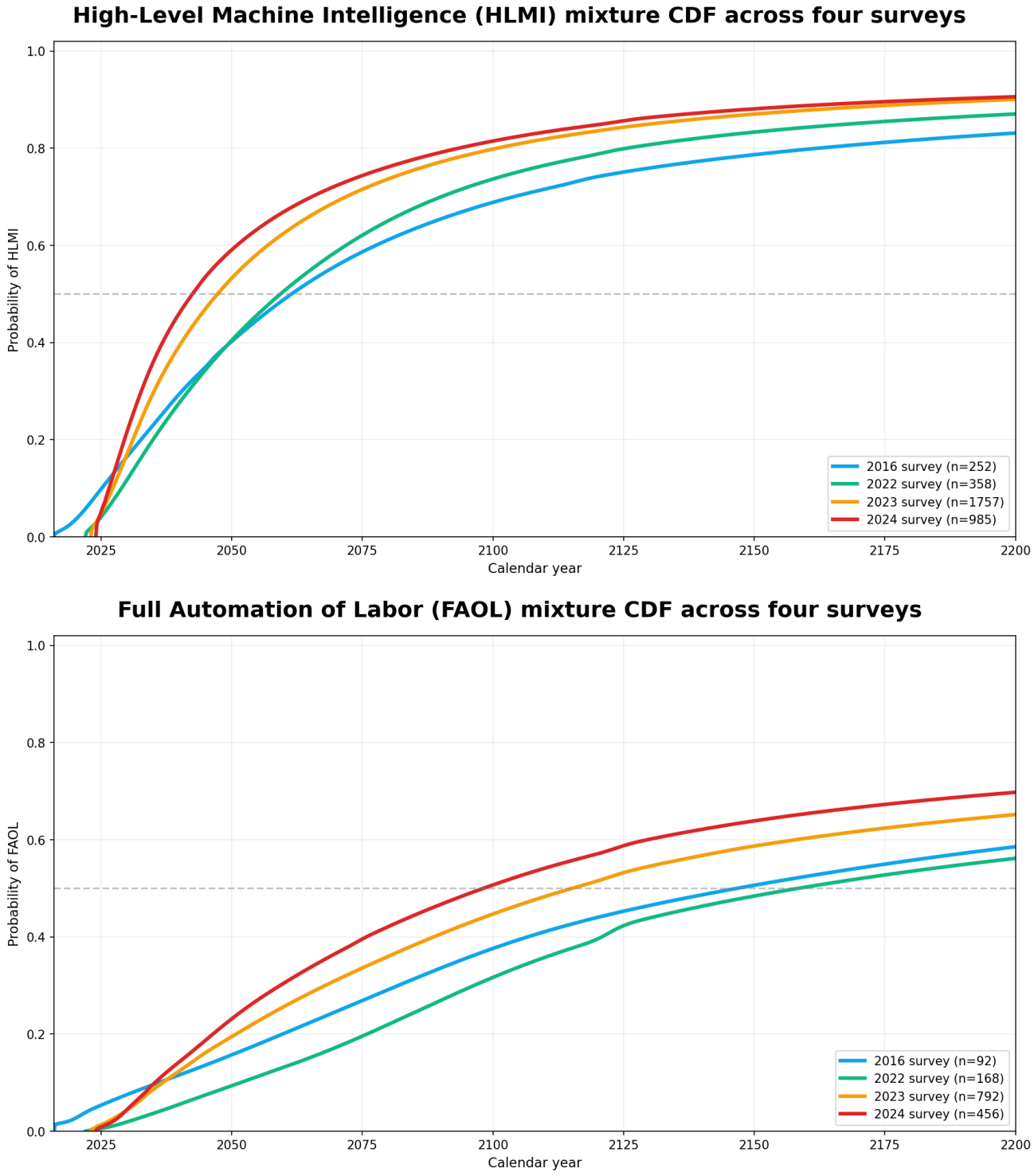


Figure 6: ***Mean aggregate forecasts of HLMI and FAOL over four surveys.*** *Lines represent an average over CDFs fitted to individual responses in each survey.*

As shown in Figure 5, respondents as a group have highly diverse views, and most individual respondents are uncertain over a diverse range of outcomes. The mean aggregate forecasts shown above put a ten percent chance on both HLMI and FAOL within around a decade, while also holding out substantial probability for these not having arrived in one hundred years.

Over past surveys, FAOL has consistently been forecast much later than HLMI. The reasons for this are unclear: one might imagine that if all human tasks are able to be automated, then depending on one's definitions, occupations would naturally be automatable already or in a short time. Yet we have consistently seen over fifty years of difference between the dates for an even chance of these two events. The questions differ in more ways than the definitions, so that may contribute. This trend continues here, with a gap of 56 years, the smallest to date.

Table 1: ***Comparison of human-level AI milestone forecasts between this survey and the previous one.*** *Years are dates where the mean aggregate CDF reaches the specified probability.*[9]

Outcome	2023 survey	2024 survey	Shift
FAOL: 10% probability	2037	2035	2 yr
FAOL: 50% probability	2115	2098	17 yr
HLMI: 10% probability	2027	2027	0 yr
HLMI: 50% probability	2047	2042	5 yr

3.3 Causes of AI progress

To elicit opinions about the relative importance of various causes of AI progress, we asked respondents to estimate how much less AI progress they would have expected with half as much of each of five factors. For instance, half as many researchers or half as much funding. For computing hardware, "half as much" referred to the cost decline on a logarithmic scale: respondents were asked to imagine a 5-fold rather than a 20-fold reduction in costs, approximately half as far on that scale. Each participant was asked about a random subset of three inputs. Here is one of the questions:

> The next questions ask about the sensitivity of progress in AI capabilities to changes in inputs.
>
> 'Progress in AI capabilities' is an imprecise concept, so we are asking about progress as you naturally conceive of it, and looking for approximate answers.
>
> Imagine that over the past decade, only **half as much researcher effort** had gone into AI research. For instance, if there were actually 1,000 researchers, imagine that there had been only 500 researchers (of the same quality).
>
> How much less progress in AI capabilities would you expect to have seen?
> *e.g. If you think progress is linear in the number of researchers, so 50% less progress would have been made, write '50'. If you think only 20% less progress would have been made write '20'.*
>
> _____% less

Results are shown in Figure 7. Opinions varied widely. Researcher effort was seen as least important and hardware cost reductions most important. Results were very similar to those in the previous survey.

[9]The aggregate forecast is built by fitting a gamma distribution to each respondent and averaging the CDFs.

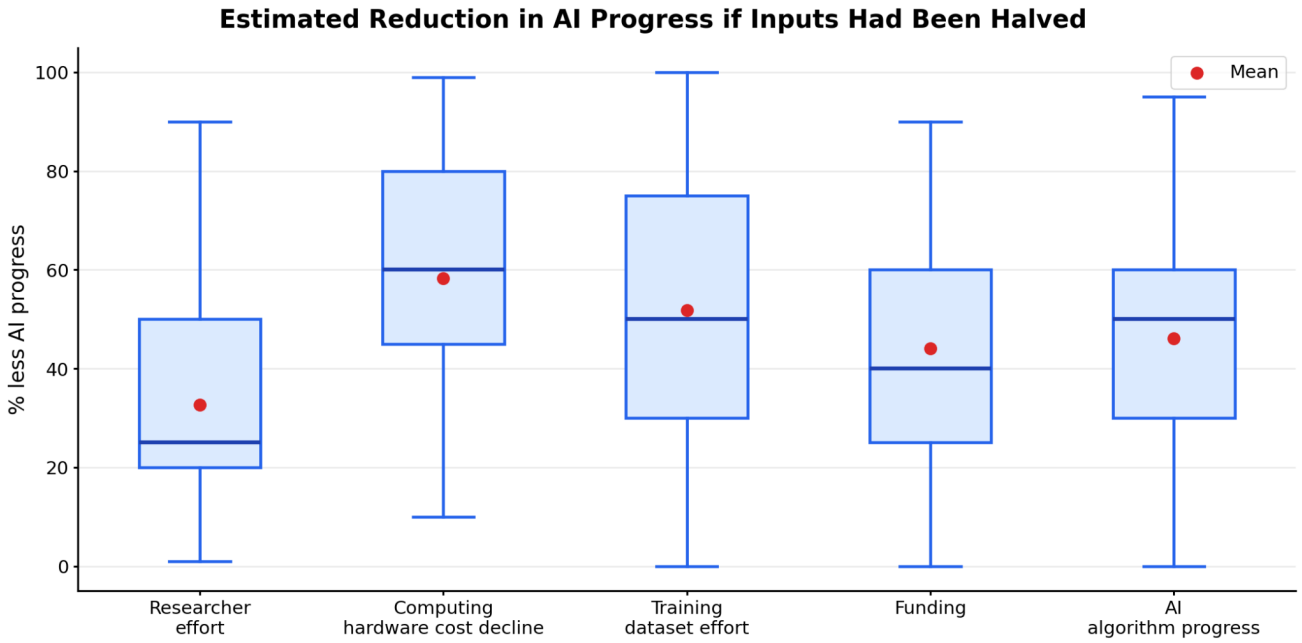


Figure 7: ***Comparing the causes of AI progress.*** *Values for each cause are given as percentages of true, historically observed AI progress. Higher values indicate greater importance to progress. Red dots are means; box shows IQR with median line. Corresponds to Figure 5 in the 2023 survey report* (Grace et al., 2025).

3.4 Chance of an intelligence explosion

We asked a sixth of respondents about the possibility of advanced AI speeding up technological progress in three different ways, directing two such questions to each participant. One of these questions asked about an argument for this outcome (n=189):

> Some people have argued the following:
>
> *If AI systems do nearly all research and development, improvements in AI will accelerate the pace of technological progress, including further progress in AI.*
>
> *Over a short period (less than 5 years), this feedback loop could cause technological progress to become more than an order of magnitude faster.*
>
> How likely do you find this argument to be broadly correct?

Answers were on a rating scale: Quite unlikely (0-20%), Unlikely (21-40%), About even chance (41-60%), Likely (61-80%), Quite likely (81-100%).

Their answers, as shown in Figure 8, were similar to those from previous years, with only a small fraction finding the argument quite likely (81-100%)—4%, down from 12%, 7%, and 9% in the 2016, 2022, and 2023 surveys, respectively.

Another question asks:

> Assume that HLMI will exist at some point.
>
> How likely do you then think it is that the rate of global technological improvement will dramatically increase (e.g. by a factor of ten) as a result of machine intelligence:
>
> Within **two years** of that point? _____% chance

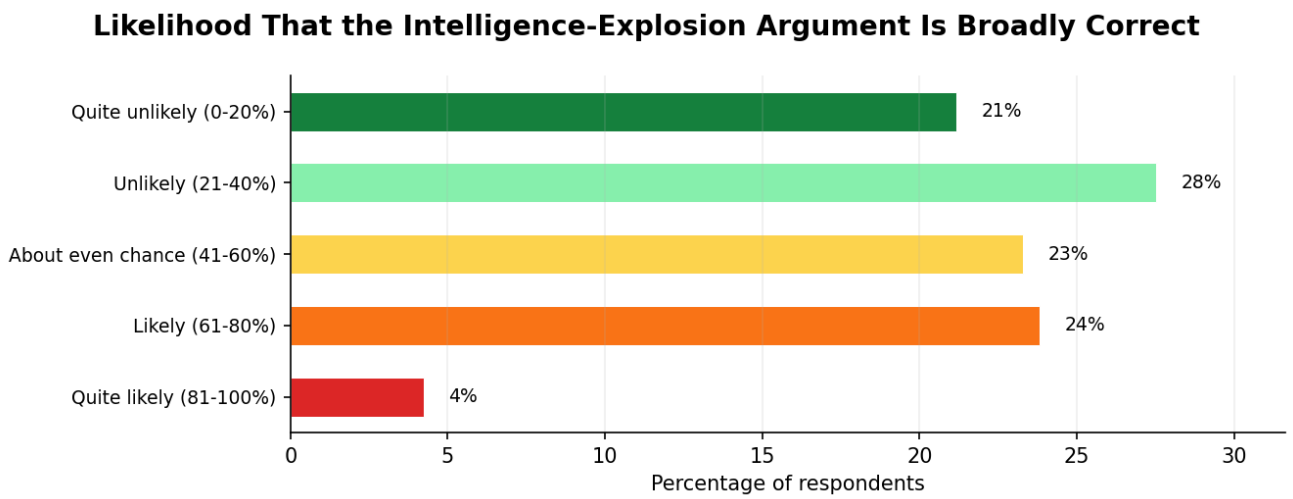


Figure 8: ***How likely participants find an argument that AI progress will create a feedback loop and accelerate AI development.***

Within **thirty years** of that point? _____% chance

The third question is similar except asking instead about machine intelligence that is vastly better than humans at all professions following HLMI.

Results are shown in Table 2. The median participant found both of these outcomes unlikely but not very unlikely in two years, and likely in thirty years. Median responses for both questions were identical to those in the 2023 survey.

Table 2: ***Probabilities of intelligence explosion.*** *Median responses were unchanged from 2023. Indeed, they have been stable across all editions of the survey, except for vastly superhuman AI within 30 years of HLMI, which was 50% in 2016 instead of 60%.*

Outcome	N	Mean	IQR	Median	2023 Median
Dramatic tech speedup within 2 years of HLMI	173	30.3%	10-50%	20%	20%
Dramatic tech speedup within 30 years of HLMI	172	71.6%	50-95%	80%	80%
Vastly superhuman AI within 2 years of HLMI	171	20.9%	1-30%	10%	10%
Vastly superhuman AI within 30 years of HLMI	172	59.6%	30-95%	60%	60%

3.5 Traits of AI systems in twenty years

We asked participants (n=374-380), "In 2044, how likely do you think the following will be for at least some state-of-the-art AI systems?" and offered them eleven socially consequential traits summarized in Figure 9.

Figure 10 shows the fraction of participants who rated each trait 'likely (60%-90%)' or 'very likely (>90%)', and the corresponding 2023 predictions for the year 2043. The fraction of people in this group inched up for almost all traits, with "Are able to cause important actions in the world similarly to a human e.g. run a business, run an advocacy campaign" seeing the largest growth.

"Take actions to attain power" remained the trait fewest people saw as 'likely' or 'very likely' to be present in AI systems twenty years in the future, though it attained a median response of "even chance", up from "unlikely" in the last survey.

3.6 True decision explanation prospects

We asked (n=520):

> For typical state-of-the-art AI systems in 2029, do you think it will be possible for users to know the true reasons for systems making a particular choice? By "true reasons" we mean the AI correctly explains its internal decision-making process in a way humans can understand. By "true reasons" we do not mean the decision itself is correct.

We gave participants five options, shown in Figure 11. The responses are very similar to those in the 2023 survey about 2028. Only 22% of participants considered this likely or very likely.

3.7 Scenarios deserving concern

We asked half of respondents (n=753-760), "How would you rate the level of concern these scenarios deserve over the next thirty years?" and offered eleven scenarios, and an option for a write-in 'other' scenario. They could choose between four levels of concern, shown in Figure 12 with the results.

All eleven scenarios were widely considered deserving of at least substantial concern, with the least—"Near-full automation of labor makes people struggle to find meaning in their lives."—still considered deserving of substantial or extreme concern by over a third of participants. The scenarios considered

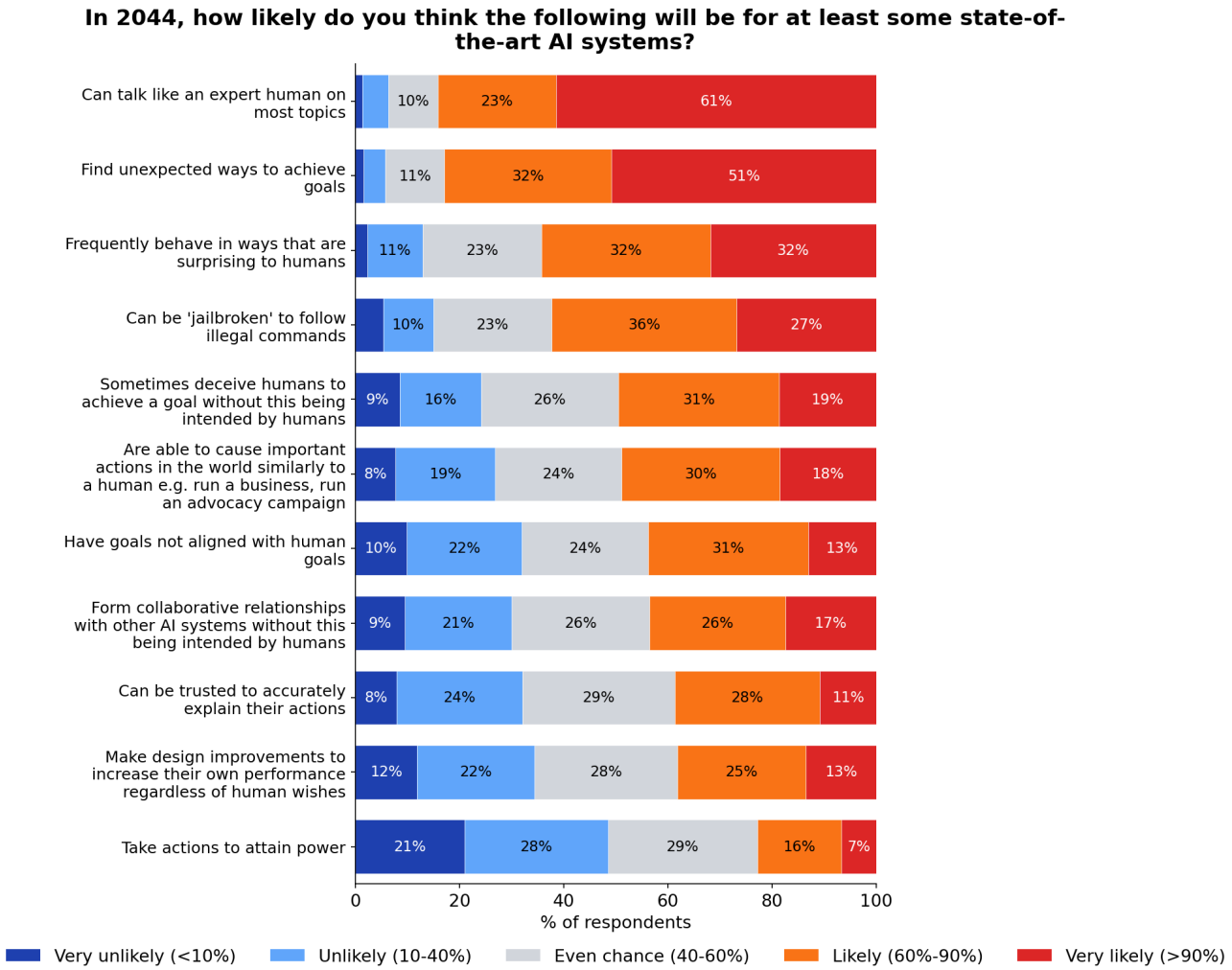


Figure 9: ***Mixed expectations of AI traits within twenty years.*** *Respondents were asked how likely each trait would be "for at least some state-of-the-art AI systems" in 2044. Trait descriptions are summarized. Corresponds to Figure 7 in the 2023 survey report* (Grace et al., 2025).

most deserving of concern were both around erosion of public epistemology: "AI makes it easy to spread false information, e.g. deepfakes" and "AI systems manipulate large-scale public opinion trends".

Opinions were fairly stable since the previous survey, with none gaining or losing more than four percentage points of participants who consider them deserving of at least substantial concern. However, the six scenarios rated highest by this measure in 2023 all moved downward slightly, and the five others all moved upward.

3.8 Impact of a future with HLMI

We asked all respondents (n=1538) the following, after reminding them of the HLMI definition:

> Assume for the purpose of this question that HLMI will at some point exist. How positive or negative do you expect the overall impact of this to be on humanity, in the long run?
>
> Please answer by saying how probable you find the following kinds of impact, with probabilities adding to 100%:

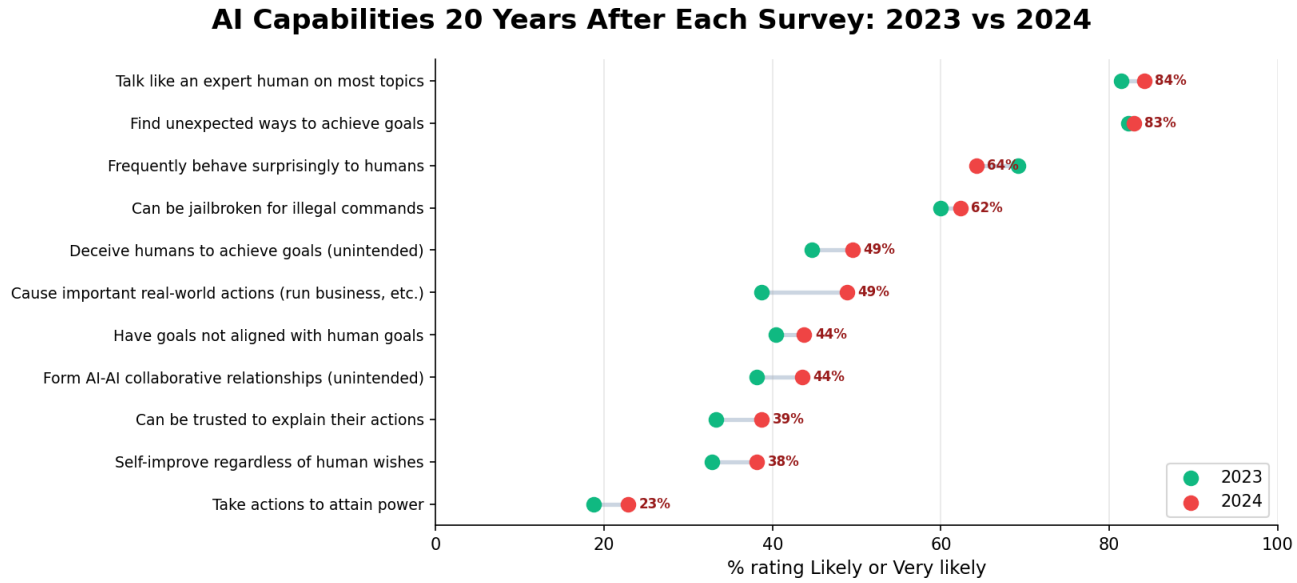


Figure 10: ***Growing anticipations of socially consequential AI traits.*** *Fraction of participants considering each trait 'likely (60%-90%)' or 'very likely (>90%)' for at least some state-of-the-art AI systems at a date twenty years later, for surveys in 2023 and 2024. Trait descriptions are summarized.*

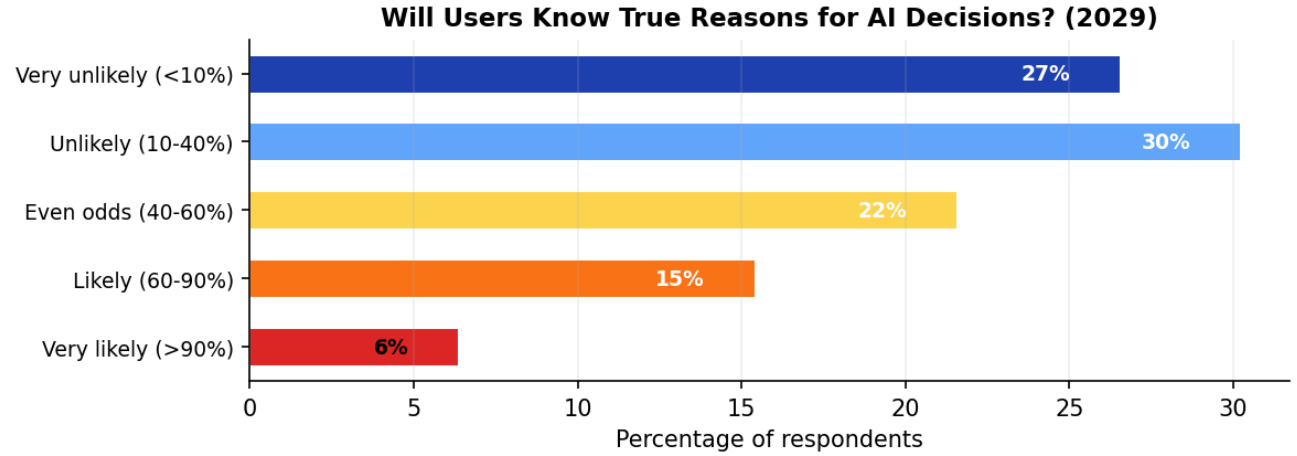


Figure 11: ***Pessimism about understanding AI decisions within five years.*** *Corresponds to Figure 8 in the 2023 survey report* (Grace et al., 2025).

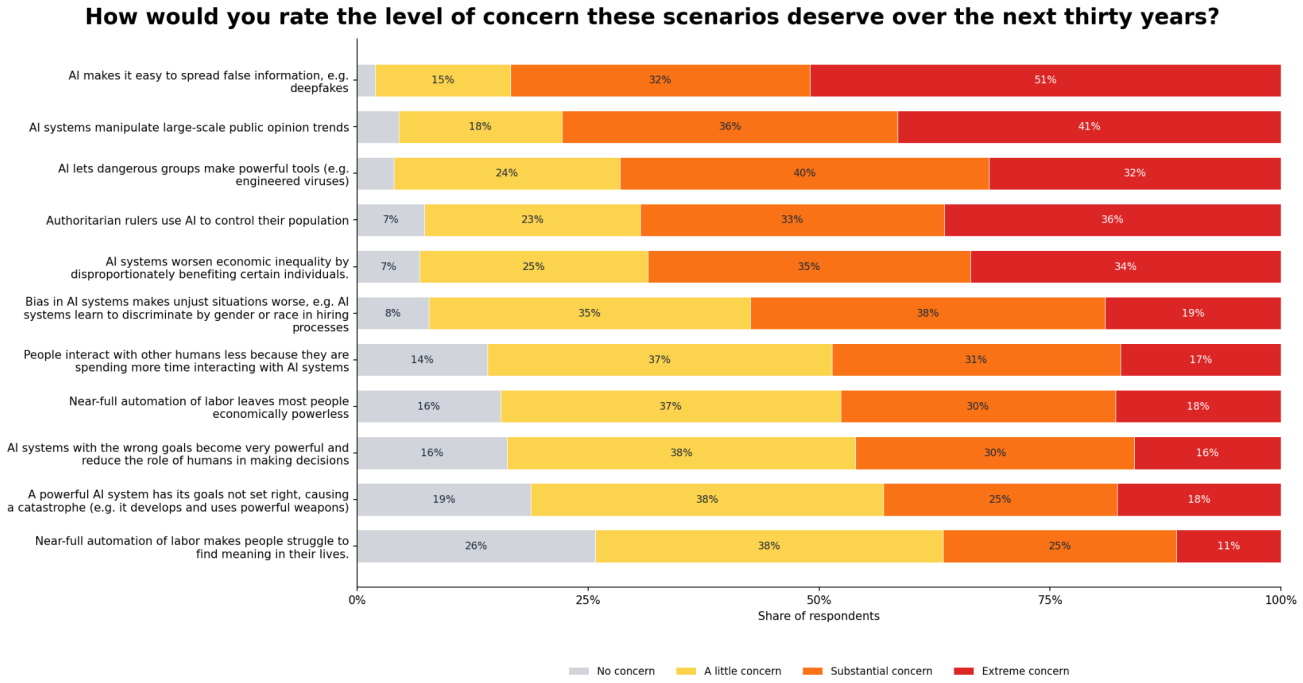


Figure 12: ***Level of concern deserved over the next 30 years. Labels are paraphrased; full descriptions include further details, available in Appendix C.*** *Corresponds to Figure 9 in the 2023 survey report* (Grace et al., 2025).

Extremely good (e.g. rapid growth in human flourishing) _____%

On balance good _____%

More or less neutral _____%

On balance bad _____%

Extremely bad (e.g. human extinction) _____%

Total [sum]%

The probabilities had to sum to 100% for the participant to enter the response, and we excluded one response because it contained a negative probability.

As in previous years, the range of responses was wide, with participants spread between fully expecting extremely good overall impact and fully expecting extremely bad overall impact. However, the population is not polarized: most participants were in agreement that the future impacts of HLMI looked very uncertain: 788 of 1,538 respondents (51.2%) assigned at least 5% probability to both extremely good and extremely bad overall impacts. Table 3 illustrates the range of answers. For two alternative visualizations of this data, see Appendix G.

On average participants only put 52% on overall impacts of HLMI being good, and reserved 28% for bad outcomes. This ambivalence is striking in light of the participant pool being people participating in AI research.

Table 3: ***Average predicted overall impacts of HLMI on humanity***

Outcome	Mean	Median
Extremely good (e.g. rapid growth in human flourishing)	23.9%	15.0%
On balance good	27.7%	25.0%
More or less neutral	20.7%	20.0%
On balance bad	17.8%	15.0%
Extremely bad (e.g. human extinction)	9.9%	5.0%

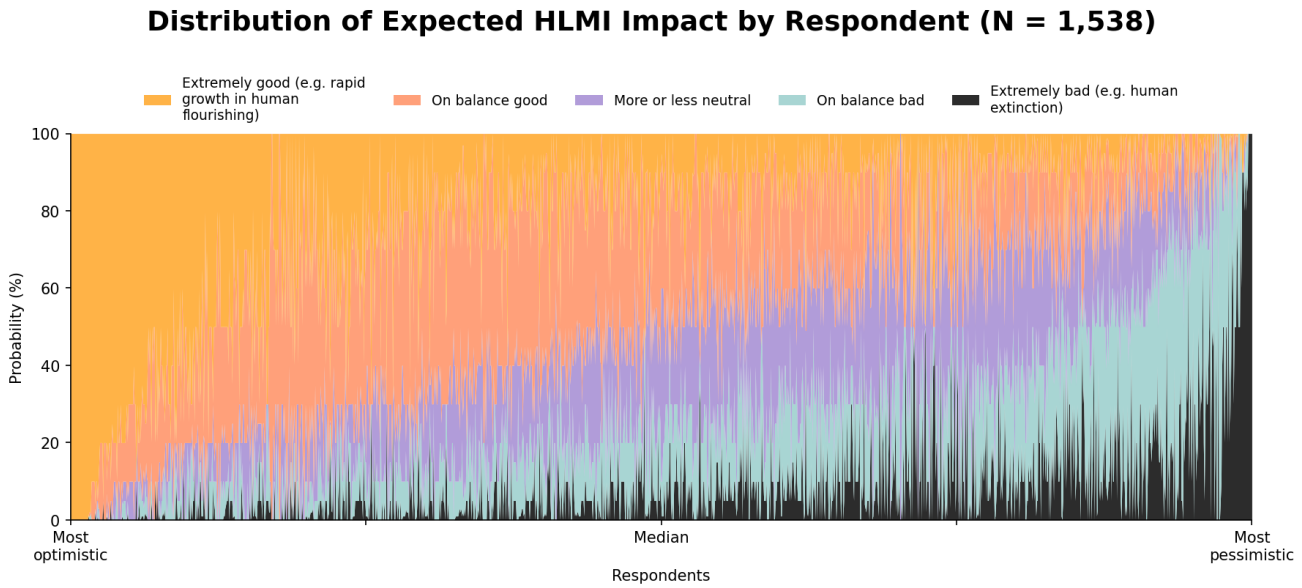


Figure 13: ***Predicted effects of HLMI on humanity.*** *Each vertical slice shows one respondent's probability allocation between five categories of overall impact on humanity in the long run from HLMI. Participants are sorted left to right by a weighted sum of probabilities corresponding to overall optimism.*

3.9 Extinction or severe disempowerment

To shed light on the substantial probabilities given to 'Extremely bad (e.g. human extinction)' in Section 3.8, in 2022 we added a question asking more directly about AI endangering humanity, alongside a variation specifying risks from losing control of advanced AI. In 2023, we added a further variation asking about risks within the next 100 years. These questions are as follows:

> What probability do you put on future AI advances causing human extinction or similarly permanent and severe disempowerment of the human species?
>
> _____%
>
> What probability do you put on human inability to control future advanced AI systems causing human extinction or similarly permanent and severe disempowerment of the human species?
>
> _____%
>
> What probability do you put on future AI advances causing human extinction or similarly permanent and severe disempowerment of the human species within the next 100 years?
>
> _____%

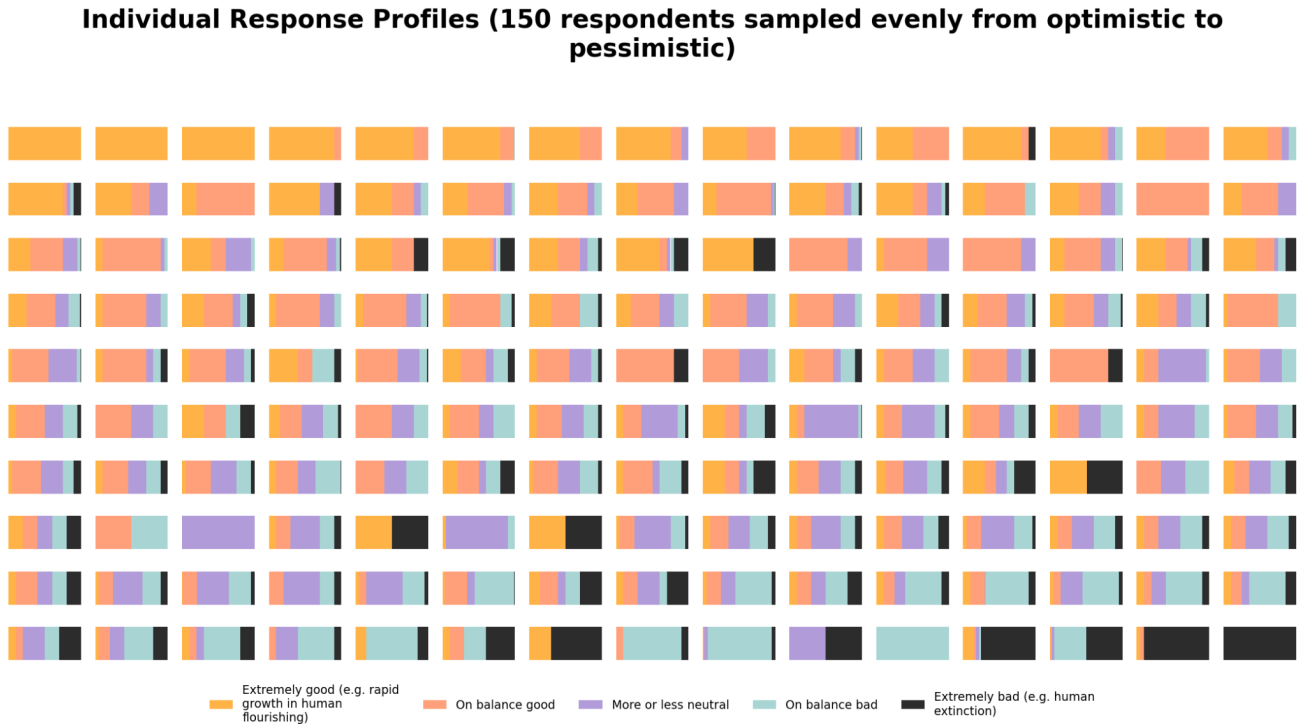


Figure 14: ***Individual response profiles for the overall impact of HLMI on humanity.*** *Each small rectangle represents one respondent's five-category probability distribution, sampled evenly from the most net-optimistic to the most net-pessimistic (weighting extremely good and extremely bad outcomes evenly). The chart illustrates the diversity of expert opinion.*

Each participant received one of these three questions, with the first twice as likely to be assigned as either of the others (n=744, 392, 353).

The distribution of responses can be seen in Figure 15.

Combining responses to the three questions[10] (n=1489), the median participant expects at least around a 10% chance of future AI leading to human extinction or similarly permanent and severe disempowerment of the human species. The group's average probability is at least around 18%. 81% of participants gave extinction or permanent and severe disempowerment at least a 1% chance, while 12% gave it zero chance. A third gave it at least a 20% chance. The middle 50% of estimates ranged from 1% to 25%. Figure 1b and Figure 15 show the distribution of expectations. Figure 16 shows how many participants gave at least 10% or 25% chances in different questions and years.

All three questions received similar answers, weakly suggesting that the risks people are imagining are generally within the coming century, and that 'human inability to control future advanced AI systems' is a reasonable description of a large part of the problem. (Only weakly because different participants answered each of these questions, and different question framings could prompt different answers that participants might not stably endorse if they considered all three questions side by side.)

Compared with the results of the 2023 survey (Grace et al., 2025) in Table 4, we see a slight increase in

[10] Additional constraints on two of the questions would make this a strict lower bound if participant probability assignments were assumed to be logically coherent, since participants could for instance put further probability on this outcome via issues other than 'inability to control' or in over one hundred years, respectively, but should not put less probability on extinction overall than a more specific version of extinction. However human probability assignment is fallible and in particular sometimes subject to the conjunction fallacy, so we instead treat this as an approximate lower bound.

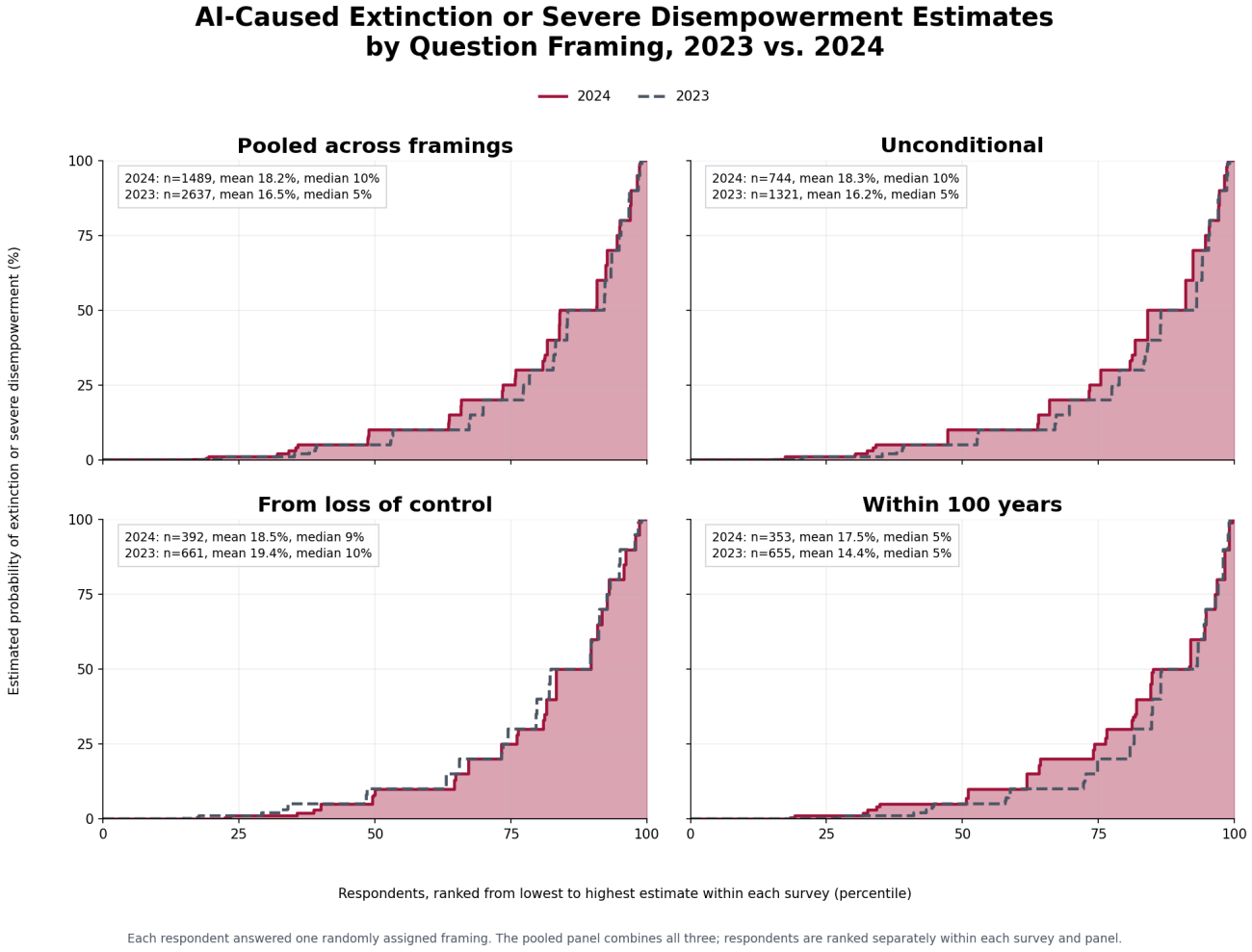
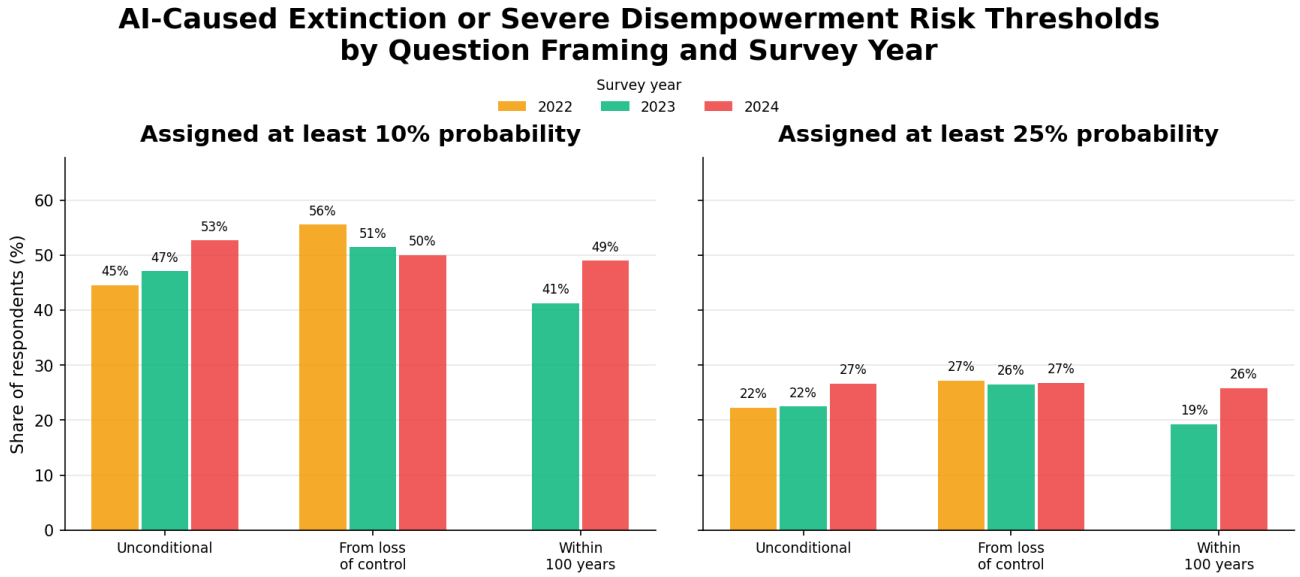


Figure 15: ***Distribution of chances assigned to human extinction or similarly permanent and severe disempowerment from AI.*** *Each vertical slice shows one respondent's probability of extinction or permanent and severe disempowerment, with participants sorted by decreasing optimism. A curve that is further up and to the left indicates a greater level of concern. Participants answered one of three versions of this question shown separately above; 'pooled' combines these answers. From this we can see that respondents were more concerned overall than in 2023, but less concerned about loss of control in particular.*



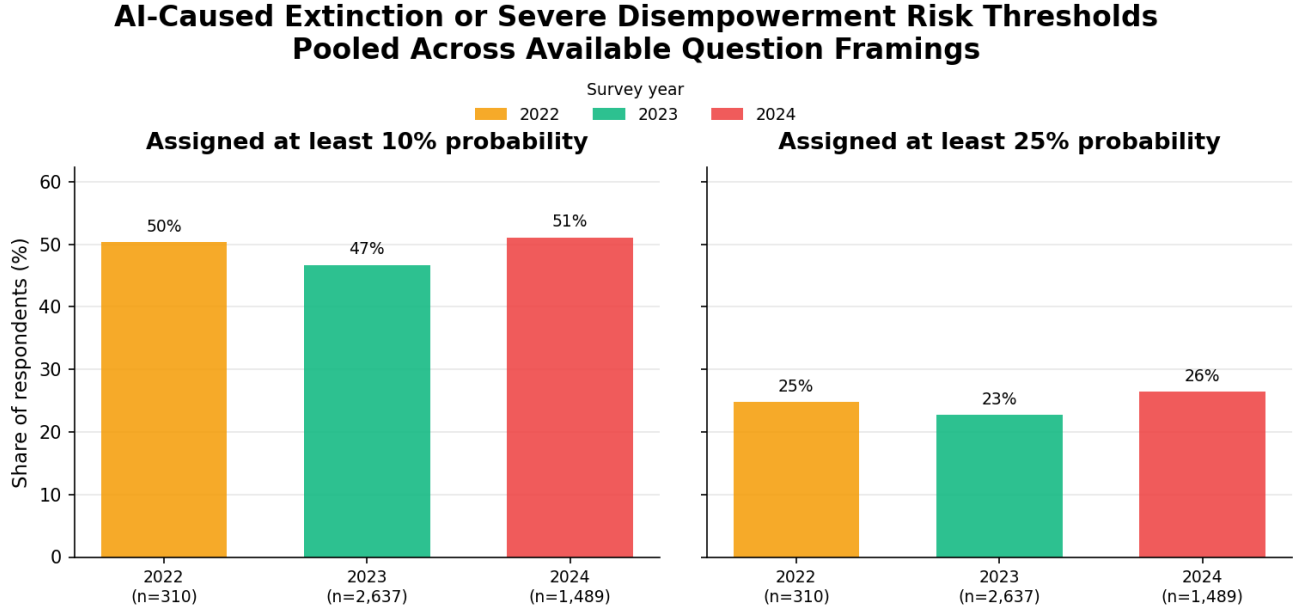


Figure 16: ***Fraction of respondents assigning at least 10% and at least 25% probability to extinction or similarly permanent and severe disempowerment.*** *Top: separated by question; bottom: pooled across the question variants available in each survey.*

the chances assigned to these catastrophic outcomes. For the first time in these surveys, the median AI researcher says human extinction or permanent and severe disempowerment from future AI has a 10% chance, given no further conditions. See Figure 17.

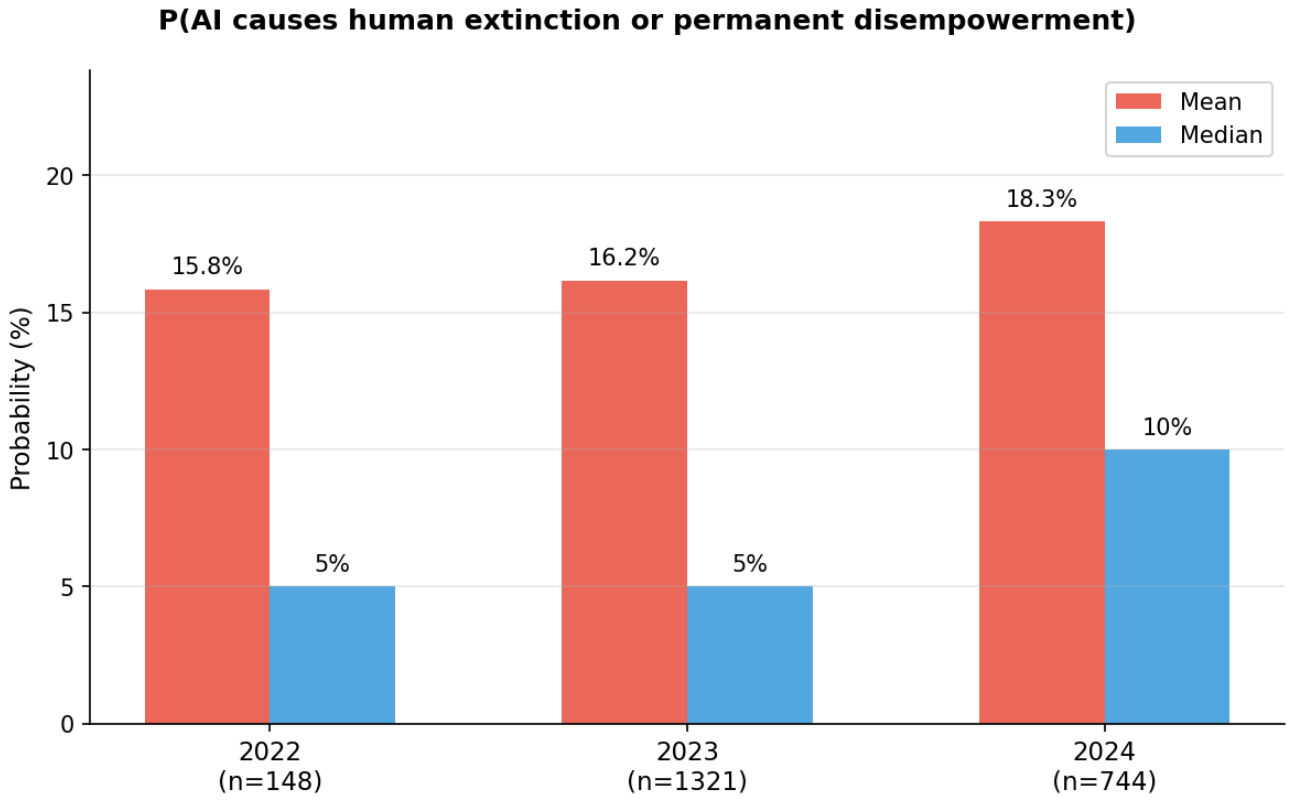


Figure 17: ***Mean and median probabilities given to AI causing human extinction or similarly permanent and severe disempowerment, without further conditions.*** *Both rose between the 2023 and 2024 surveys.*

Table 4: ***Extinction or permanent and severe disempowerment looks slightly more likely in 2024 than in 2023****. This corresponds to Figure 12 in the 2023 survey report* (Grace et al., 2025)*.*

Question	N	Mean	2023 Survey Mean	Change	Median	2023 Survey Median	Change	>=10%	>=25%
What probability do you put on future AI advances causing human extinction or similarly permanent and severe disempowerment of the human species?	744	18.3%	16.2%	+2.1 pp	10.0%	5.0%	+5.0 pp	52.7%	26.6%
What probability do you put on human inability to control future advanced AI systems causing human extinction or similarly permanent and severe disempowerment of the human species?	392	18.5%	19.4%	-0.9 pp	9.0%	10.0%	-1.0 pp	50.0%	26.8%
What probability do you put on future AI advances causing human extinction or similarly permanent and severe disempowerment of the human species within the next 100 years?	353	17.5%	14.4%	+3.1 pp	5.0%	5.0%	unchanged	49.0%	25.8%
Pooled (all variants combined)	1489	18.2%	16.5%	+1.7 pp	10.0%	5.0%	+5.0 pp	51.1%	26.5%

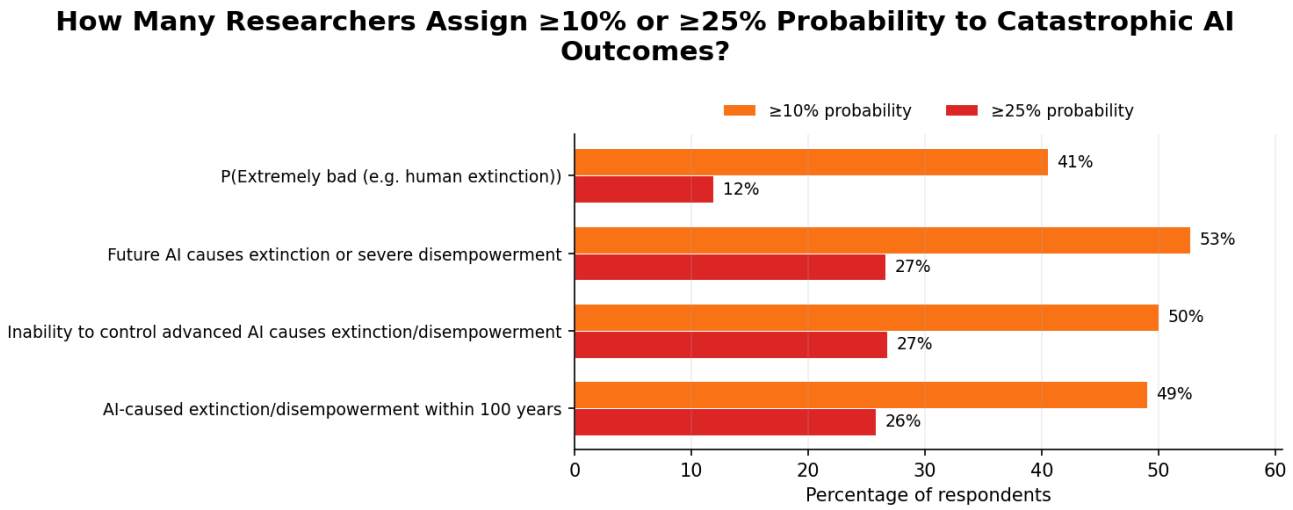


Figure 18: ***Large fractions of participants see serious chances of catastrophic outcomes from future AI advances.*** *A comparison of participants with high answers for three questions about extinction or disempowerment, and one discussed earlier about overall impacts of HLMI.*

Researchers assigned somewhat lower chances overall to an extremely bad outcome from HLMI than they did to human extinction or severe disempowerment from AI advances. See Figure 18 above. The two questions start from different assumptions, so they are not straightforwardly comparable. The HLMI question asks respondents to assume HLMI will at some point exist and to divide 100% across five categories of overall long-run impact, while the extinction questions make no such assumption and ask for a single probability about future AI advances in general. Differences in assumptions, scope, and response format could each produce a gap of this kind, so the comparison does not by itself establish that respondents consider extinction or severe disempowerment acceptable. That some fraction of researchers do not count such outcomes as extremely bad, at least under some envisaged conditions, remains a possibility these questions do not test; some researchers and AI company employees have expressed such views publicly (Torres, 2025; Sutton, 2023).

There were substantial geographical differences in expectations of extinction or disempowerment, as shown in Figure 19. Researchers who completed their undergraduate studies in Asia saw more risk than those who studied in North America or Europe.

Researchers who had thought more previously about the social impacts of smarter-than-human AI tended to judge the risks as larger than those who had thought less, as shown in Figure 20. However the differences were relatively small: all but the least attentive to the topic had the same median: 10%. This is some reassuring evidence against the median being affected by heavily involved people being more likely to take the survey.

3.10 Support for AI safety research

We asked half of participants (n=748) the following, after defining 'AI safety research' per Section 2.3:

> How much should society prioritize AI safety research, relative to how much it is currently prioritized?

Figure 21 shows the results. There was strong support for increased prioritization, with 72% of researchers saying society should prioritize AI safety research either "more" or "much more." These findings are very similar to those of the 2022 and 2023 surveys (and around 20 percentage points higher than those of the 2016 survey).

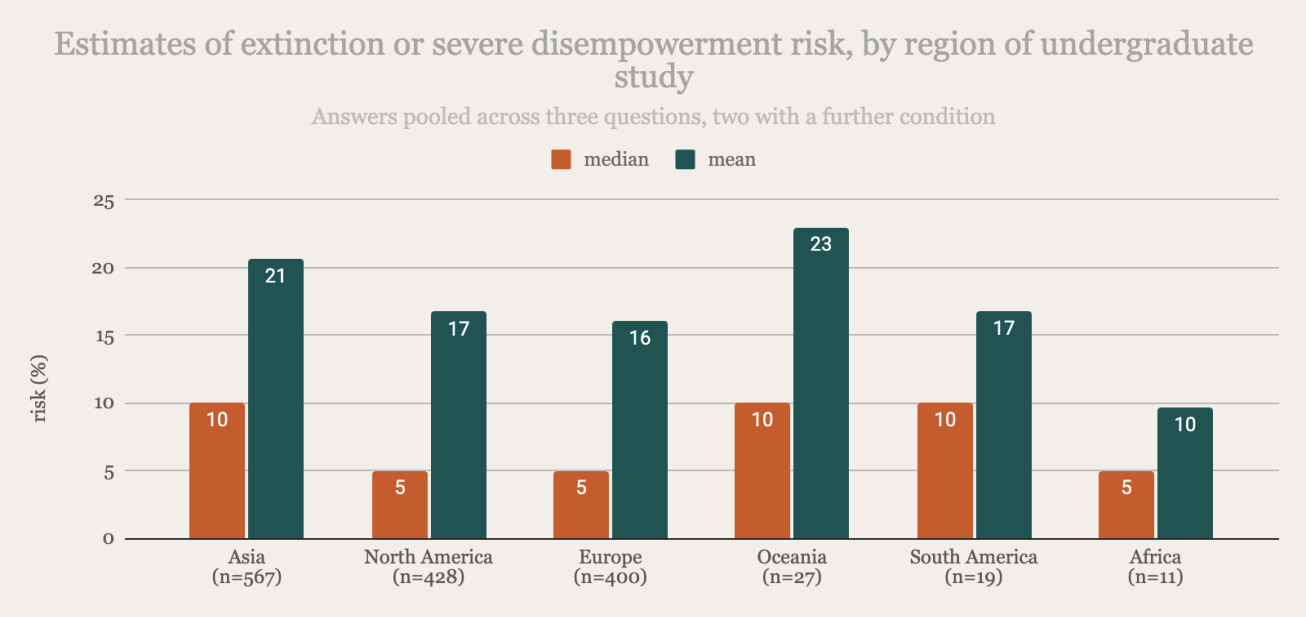


Figure 19: ***Researchers who did their undergraduate studies in different continents had different distributions of perception of extreme risk from AI. Note that for three regions n is very small.***

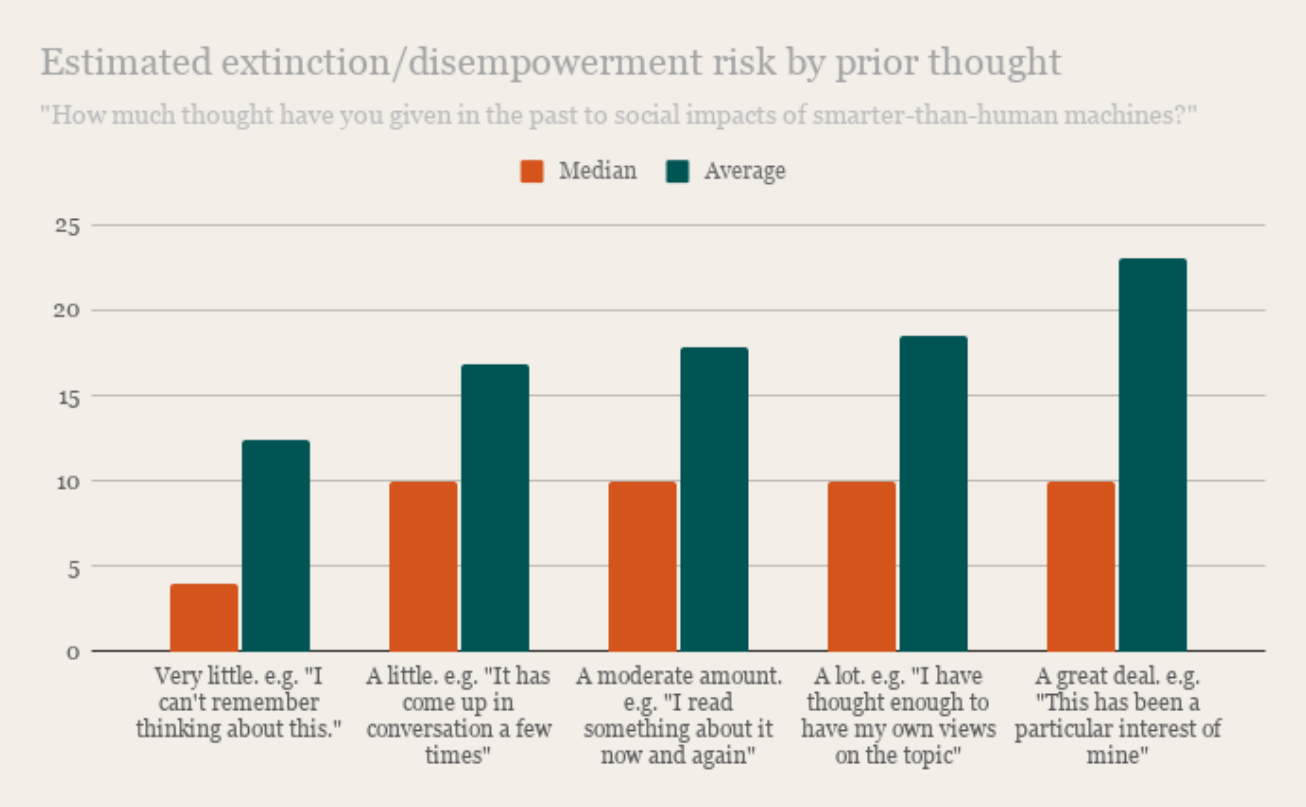


Figure 20: ***Researchers who had thought more about social impacts of smarter-than-human machines tended to see higher risks, but views were similar across the distribution.***

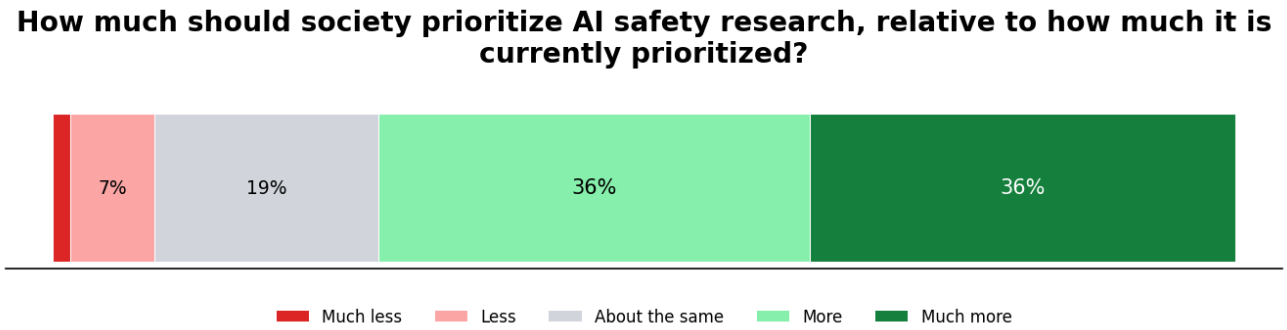


Figure 21: ***The median AI researcher still thinks AI safety should be prioritized "more."***

3.11 Stuart Russell's argument for AI risk

We presented half of participants with a summary by Stuart Russell of an argument for why highly advanced AI might pose a serious risk (Russell, 2014), available in the question list in Appendix C, then asked them three questions, each with a five-point rating scale (n=755, 753, 752):

> Do you think this argument points at an important problem?
>
> How valuable is it to work on this problem ***today***, compared to other problems in AI?
>
> How hard do you think this problem is compared to other problems in AI?

Responses are shown in Figure 22. Participants overwhelmingly found the problem important, but were spread out on the question of how valuable it is to work on today compared to other problems in AI. They tended to find it harder than other problems in AI, with very few considering it easier.

Only 12% of participants considered this among the most important problems in the field, and only 7% considered it much more valuable than other problems in AI to work on. This is interesting in combination with over half of participants indicating that they think AI poses at least a 10% chance of human extinction or disempowerment. Possibly Russell's argument does not capture much of what concerns them about future AI, or its apparent difficulty to resolve makes it less important or valuable, or participants consider other problems in AI even more pressing.

Responses in 2024 were broadly similar to those in 2023. Compared with 2016, respondents continued to rate the problem as more important, more valuable to work on today, and harder, although these measures had not all increased consistently across survey waves.

3.12 Socially beneficial rate of AI progress

We asked a quarter of participants (n=382):

> What rate of global AI progress over the next five years would make you feel most optimistic for humanity's future? Assume any change in speed affects all projects equally.

Results are shown in Figure 23. Participants could also opt for 'other', which 3% did.

Opinion was spread roughly a third each way between faster paces, slower paces, and the current speed. "Much slower" was the least popular answer, but saw substantial growth since 2023, increasing by around 80%.

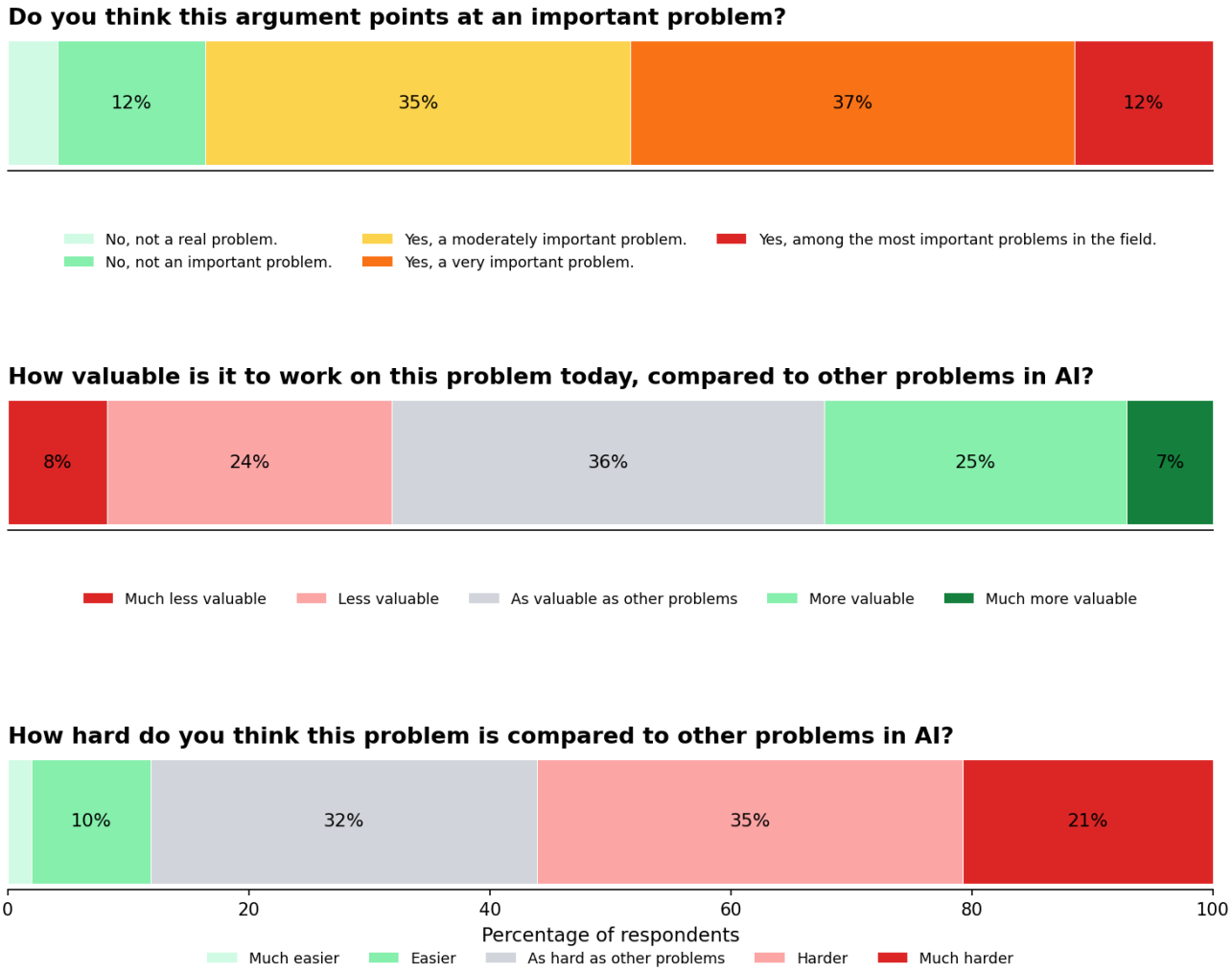


Figure 22: ***Views on Stuart Russell's argument for AI risk.*** *Participants read a summary of an argument before being asked about it. Corresponds to Figure 15 in the 2023 survey report* (Grace et al., 2025).

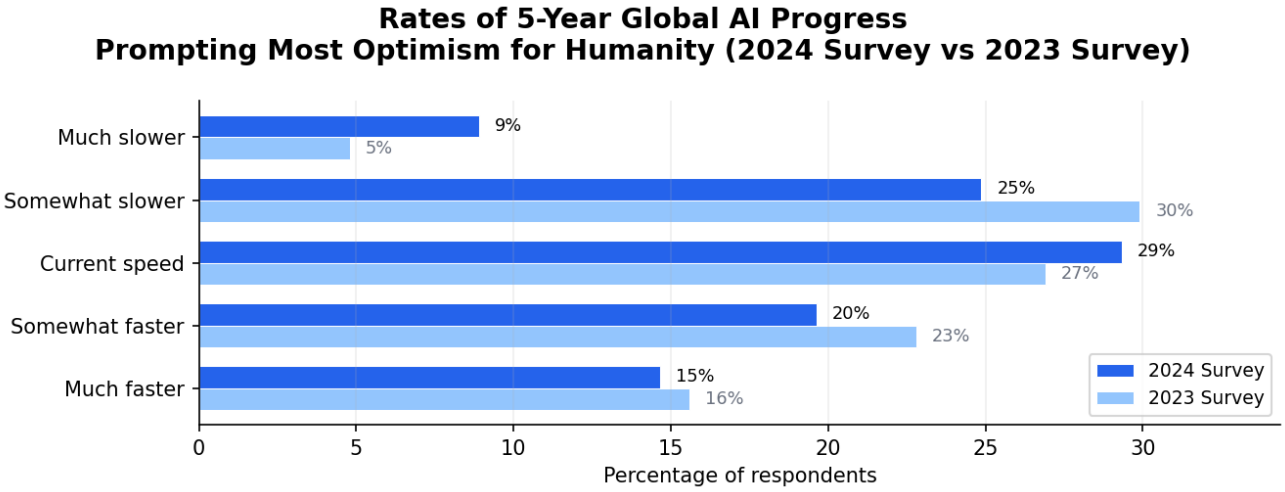


Figure 23: Respondents shifted slightly ***toward optimism for slower AI progress compared to the 2023 survey.***

3.13 Impressions of others' views

We asked a quarter of participants (n=385):

> To what extent do you think you disagree with the typical AI researcher about when HLMI will exist?

Results are below.

Table 5: ***Perceived disagreement with others on when HLMI will exist.***

Response	Count	Percentage
Not much	169	44%
A moderate amount	174	45%
A lot	42	11%

We asked a quarter of participants (n=386):

> To what extent do you think people's concerns about future risks from AI are due to misunderstandings of AI research?

Around three quarters of respondents answered 'somewhat' or 'to a large extent', though 'almost entirely' was much less popular. Responses here are very similar to those in the 2023 survey.

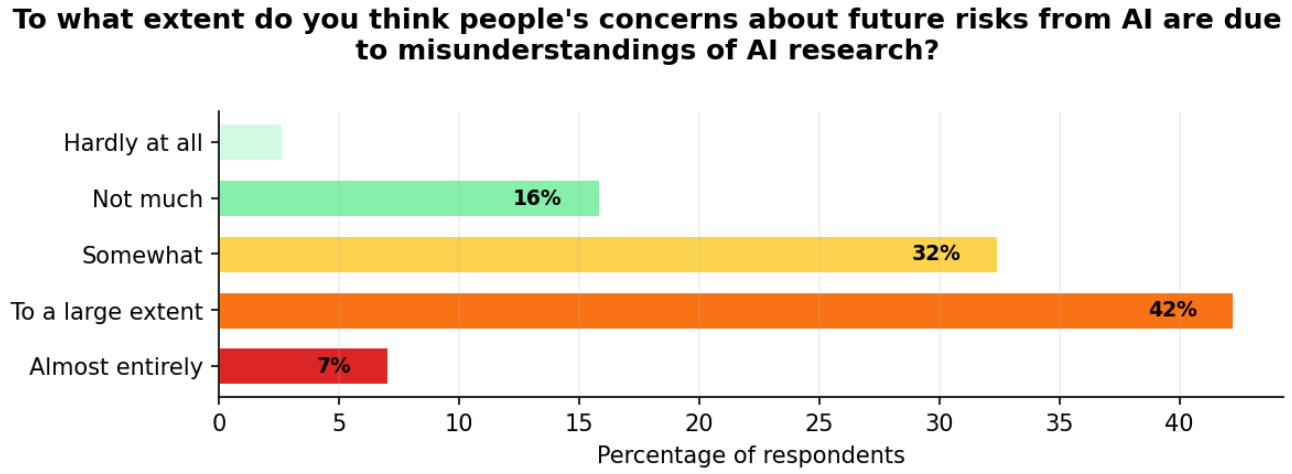


Figure 24: ***Extent of AI concerns attributed to misunderstanding by participants.***

4 Discussion

Here, we summarize some key results from the survey, note some related works, and discuss some limitations.

4.1 Summary of key results

4.1.1 Positive and negative outcomes from future AI systems

As in previous surveys, most researchers anticipated substantial risks from more advanced AI. The average probability assigned to human extinction or extreme disempowerment from future AI advances was more than one in six, and most researchers gave it at least one in ten. See Section 3.9. These numbers

are slightly up from the last survey. Researchers thought many other scenarios deserved high degrees of concern over the next 30 years, with AI misinformation and public opinion manipulation at the top, followed by misuse (e.g. bioweapons), authoritarianism, and economic inequality. See Section 3.7. On average researchers assigned a 10% chance to extremely bad long-run consequences for humanity overall if high-level machine intelligence is developed. Researchers still considered extremely good outcomes more likely than extremely bad outcomes, assigning them 24% (vs. 10%) on average. See Section 3.8.

72% of researchers thought society should support AI safety research more (36%) or much more (36%). See Section 3.10. Researchers were around evenly split regarding whether they would be most optimistic for humanity's future if the pace of global AI progress were faster (34%), slower (34%), or the current speed (29%). See Section 3.12. The fraction of researchers who would be most optimistic if progress were much slower nearly doubled from the 2023 survey, from 5% to 9%.

4.1.2 Timelines to AI milestones

Forecasts of broadly human-level AI capabilities moved earlier again in the 2024 survey, although they remained strongly dependent on how the question was asked. For machines able to perform every task better and more cheaply than human workers, the mean aggregate forecast reached 50% probability in 2042, compared with 2047 in the previous survey, assuming no major disruption to scientific activity. When respondents instead considered when all occupations would be fully automatable, the corresponding date was 2098, compared with 2115. Both questions concern what machines could do, rather than whether they would actually be deployed.

In the mean aggregate forecasts, all but five of the 39 narrow tasks we asked about were expected to be achievable by AI systems within ten years, and all but one were assigned a ten percent chance of becoming feasible by 2027 at the latest. Timelines were already short for most tasks, and changed little: 33 of the 39 moved by less than two years in either direction. See Section 3.1.

While researchers expected fast progress on AI, only 4% of those asked rated an argument for an intelligence explosion—that AI doing nearly all R&D could make technological progress more than an order of magnitude faster in under five years—as quite likely to be broadly correct, down from 7–12% in previous years. See Section 3.4.

As in previous surveys, researchers were not optimistic that we would understand the reasons for AI systems' decisions within the next five years. Only 22% of researchers found this likely or very likely (see Section 3.6). More researchers saw many important AI traits as likely or very likely 20 years out than they did in the survey a year earlier: being able to be jailbroken to take illegal actions, deceiving humans to achieve goals, taking important real-world actions, having goals not aligned with human goals, forming unintended collaborative relationships with other AIs, self-improving regardless of human wishes, and taking actions to attain power all seemed likely to more people. On the other hand, more researchers became optimistic since the previous survey that in 20 years' time, AI systems would be able to be trusted to explain their decisions, and that their behavior would not be frequently surprising, though both remain minority opinions. See Section 3.5.

4.2 Comparison with other surveys

The Expert Survey on Progress in AI is the longest-running series of surveys of experts in the field of AI. The 2023 iteration of this survey was the largest survey of AI researchers at the time, with 2,778 responses. It remains the largest peer-reviewed one, though a recent non-peer-reviewed survey focused on values and governance opinions had more participants (O'Donovan et al., 2025). Pew Research surveyed 1,013 AI experts in 2024, recruiting authors and presenters from 21 AI-related conferences (McClain et al., 2025).

4.3 Limitations

Previous versions of our survey have often been cited as evidence that AI researchers believe AI is likely to cause human extinction. However, we do not ask specifically about extinction, only "human extinction or similarly permanent and severe disempowerment." While this is still very concerning in any case, it is not clear from our work how many researchers are concerned about extinction, per se. Figure 12 suggests that researchers see other scenarios as more of a concern, at least in the next thirty years. Future work could seek to disentangle these concerns.

While we use the same methodology for identifying survey participants (researchers publishing in top-tier venues), this population changes from year to year. In recent years, frontier AI companies (e.g., OpenAI, Anthropic, Google DeepMind, xAI, Meta, DeepSeek) have grown dramatically, and many researchers have moved to industry; at the same time, AI companies have largely stopped publishing in the venues our survey draws on. As a result, our survey almost certainly underweights the perspective of researchers who work at AI companies (vs. academics), which would skew our results if there is a systematic difference between these populations. Anecdotally, it seems more common for researchers with shorter timelines to join frontier AI companies, suggesting that overall, researchers' timelines may have shrunk more than our results would indicate.

People make systematic, predictable errors when reasoning about probabilities (Tversky and Kahneman, 1983), and their performance depends heavily on how statistical information is presented (Gigerenzer et al., 2007; McDowell and Jacobs, 2017). These findings suggest caution in interpreting the numerical probabilities reported in our survey. We believe this effect is likely less severe in our population, however, as reasoning about probabilities is an integral part of AI and machine learning research.

Expertise, however, is a poor predictor of forecasting accuracy. Experts are frequently outperformed by simple statistical models, and their stated confidence intervals are typically far too narrow (Tetlock, 2017; Camerer and Johnson, 1991; Ben-David et al., 2013). While we consider the predictions of AI experts a topic of considerable interest, it is unclear how valuable their predictions are for accurately forecasting outcomes.

The survey we report on in this paper was conducted in December 2024, over 18 months ago. The field of AI moves fast, and so the opinions of AI researchers at the time of publication are likely to be substantially different in some ways from what we report.

Acknowledgements

Many thanks for help and support to Caitlin Elizondo, Robin Goins, Jimmy Rintjema, Rick Korzekwa, Joe Carlsmith, Will MacAskill, Josh Kalla, Michelle Hutchinson, Howie Lempel, and many others.

We are grateful to Coefficient Giving, Jaan Tallinn, Future of Life Institute (FLI) and others for funding this project.

Thanks also to all of the researchers who participated in this survey, including the following people who agreed to be recognized by name:

Solomon Teferra Abate, Tanishq Mathew Abraham, Rui Abreu, Linara Adilova, André S. Afonso, Joshua C. Agar, Areeb Ahmad, Atiyeh Ahmadi, Anirudh Ajith, Samuel Albanie, James Urquhart Allingham, Yiannis Aloimonos, Erik Altman, Philipp Altmann, Kenza Amara, Bang An, Pământ Andrei-Florin, Alessio Ansuini, Jacy Reese Anthis, Richard J. Antonello, Luca Arnaboldi, Sercan Ö. Arık, Vijay Arya, Matthew Ashman, Hassan Ashtiani, Jian Aspru-Gurizki, Peter Auer, Thomas Augustin, Achraf Azize, Jinheon Baek, Ali Baheri, Junwen Bai, Ziyi Bai, Luke Bailey, Wei Bao, Cristian-Paul Bara,

Carlo Alberto Barbano, Fazl Barez, Clark Barrett, Brian R. Bartoldson, Shreyas Basavatia, Naomi Bashkansky, Jatin Batra, Mathieu Bauchy, Jacob Beck, Catarina Belem, Renato Berlinghieri, Martino Bernasconi, Yash Bhalgat, Kush Bhatia, Robi Bhattacharjee, Jeremiah Birrell, Gilles Blanchard, Kai Blin, Niclas Boehmer, Ondrej Bohdal, Blai Bonet, Alexandre Bonlarron, Maxim Bonnaerens, Akhilan Boopathy, Jannis Born, Jörg Bornschein, Adrian G. Bors, Djallel Bouneffouf, Victor Boutin, Pascal Bouvry, Joseph S. Boyle, François Brémond, Edward De Brouwer, Anh Tuan Bui, Ethan Caballero, Chen Cai, HanQin Cai, Roberta Calegari, Marco C. Campi, Ivan B. Campos, Anjie Cao, Bochuan Cao, Yuan Cao, Roberto Capobianco, Benjamin Cappell, Davide Carbone, David Carlson, Thomas Carta, Stephen Casper, Matteo Castiglioni, Biswadeep Chakraborty, Jay Chan, Yash Chandak, Kartik Chandra, Patrick Chatain, Sotirios Chatzis, El Mahdi Chayti, Cangxiong Chen, Chacha Chen, Chao Chen, Di Chen, Feng Chen, Han Chen, Haoyu Chen, Jiayu Chen, Joya Chen, Pin-Yu Chen, Rui Chen, Tianlong Chen, Wei (Wayne) Chen, Weitao Chen, Weiyu Chen, Yixin Chen, Zhengzhang Chen, Zhuo Chen, Chao Ran Cheng, Tin Sum Cheng, Xuxin Cheng, Anshuman Chhabra, Chun-Wei Chiang, Eli Chien, Sungik Choi, Davin Choo, Leshem Choshen, Rishav Chourasia, Phillip Christoffersen, Ben Chugg, Neo Christopher Chung, Andrea Cini, Stéphan Clémençon, Jeff Clune, Simon Colton, Vincent Conitzer, Anthony C. Constantinou, James Cook, Michael Cook, Jan Corazza, Filip Cano Córdoba, Luca Cosmo, Garrison W. Cottrell, Evan Crothers, Arthur da Cunha, Michael Curry, Joshua Cutler, Edwige Cyffers, Mohammed Dabbah, Eduardo Dadalto, Wang-Zhou Dai, Mohamad H. Danesh, Hien Dang, David Danks, Progyan Das, Aritra Dasgupta, Mehdi Dastani, Arghya Datta, Guy Davidson, Oishi Deb, Nima Dehmamy, Arnauld Delaunoy, Théo Delemazure, Argyrios Deligkas, Weijian Deng, Victor Dheur, Xuan Di, Gayo Diallo, Justin Diamond, Xiao Lei Diao, Dongsheng Ding, Lei Ding, Benjamin Doerr, Gabriele Dominici, Hao Dong, Kefan Dong, Yijun Dong, Yiqi Dong, Diego Delle Donne, Bakhtiyar Doskenov, Jiawei Du, Xingbo Du, Xuefeng Du, Kinshuk Dua, Alain Oliviero Durmus, Greg Durrett, Sujan Dutta, Johan Edstedt, Eric Elmoznino, Eng, Rainer Engelken, Bernd Ensing, Tolga Ergen, Stefano Ermon, David Evans, Maciej Falkiewicz, Flint Xiaofeng Fan, Simin Fan, Xuhui Fan, Yewen Fan, Zhiwen Fan, Qingkai Fang, Yin Fang, Faramarz Fekri, Shangbin Feng, Yunhe Feng, Andrea Filippo Ferraris, Quentin Feuillade-Montíxi, Hamid Reza Feyzmahdavian, Mário Figueiredo, Giorgos Filandrianos, Ferdinando Fioretto, Tor Erlend Fjelde, Michele Flammini, Alessandro Fontanella, Vincent Fortuin, Zachary R Fox, Dennis Frauen, Kilian Freitag, Gideon Freund, Jason A. Fries, Haotian Fu, Jingwen Fu, Xinyu Fu, Michael B. Dela Fuente, Tomoki Fukai, Hiroki Furuta, Ido Galil, Quentin Gallouédec, Kanishk Gandhi, Bin-Bin Gao, Chi Gao, Han Gao, Shang Gao, Xin Gao, Simone Garatti, Gianmarco Genalti, Chuqin Geng, Walter Gerych, Fatemeh Ghofrani, Avrajit Ghosh, Soumitra Ghosh, Francesco Giannini, Micah Goldblum, Chen Gong, Mononito Goswami, Jaya Goyal, Alfredo De Goyeneche, Federica Granese, Eric Graves, Alessio Gravina, Caterina Graziani, Florian Grötschla, Jinjin Gu, Jiaqi Guan, Guanlin, Alejandro Guerra-Manzanares, Charles Guille-Escuret, Guodong Guo, Rucheng Guo, Shangmin Guo, Zhijiang Guo, Saumya Gupta, Benjamin Gutteridge, Prashnna Kumar Gyawali, Seungwoong Ha, Daniel Haider, MohammadTaghi Hajiaghayi, Faisal Hamman, Lewis Hammond, ByungOk Han, Jiyeon Han, Kai Han, Lu Han, Yujin Han, Derek Hansen, Philip Harris, Peter Hase, Ben Hayes, Jun He, Ruifei He, Yang He, Chinmay Hegde, Kristine Heiney, Vincent J. Hellendoorn, Ricardo Henao, Alvin Heng, Louis Hernandez, Chris Hicks, Torsten Hoefler, Bram Hoex, Janina Hoffmann, Feng Hong, Pengyu Hong, Benjamin Hoover, Bojian Hou, Edward S. Hu, Lijie Hu, Xixu Hu, Yining Hua, Brian R.Y. Huang, Haiwen Huang, Kuan-Wei Huang, Ruozi Huang, Shuokang Huang, Taoan Huang, Victoria Huang, Wei Huang, Weiran Huang, Yufei Huang, Itay Hubara, Noor Hussein, Sehyun Hwang, David Hyland, Alexey Ignatiev, Filip Ilievski, Francesco Innocenti, Nicole Inverardi, Brian Irwin, Kei Ishikawa, Roshni G. Iyer, Sam Ade Jacobs, Ayush Jain, Naman Jain, Palak Jain, Ajil Jalal, Steven James, Stanisław Jastrzębski, Jongheon Jeong, Aditi Jha, Xiaowei Jia, Ziyu Jia, Xiangjian Jiang, Yuan Jiang, Zhongli Jiang, Zixuan Jiang, Lei Jiao, Cheng-Bin Jin, Fabian Jogl, ST John, Nicholas A. G. Johnson, Cameron Jones, Chaitanya K. Joshi, Nikola Jovanović, Sékou-Oumar Kaba, Yoshiyuki Kabashima, Rishabh Kabra, Freddie Kalaitzis, Marcus Kalander, Iden Kalemaj, Uday Kamal, Rafiq Kamel, Michael Kamp, Sai Srinivas Kancheti, Melih Kandemir, Mintong

Kang, Antonia Karamolegkou, Alexander M. Kasprzyk, Tyler Kastner, Stephan P. Kaufhold, Zixuan Ke, Piotr Keller, Eoin M. Kenny, Ahmed Khaled, Shahbaz Khan, Muhammad Uzair Khattak, Stephen Kiilu, Doyoung Kim, Hoyoung Kim, Jeongsol Kim, Seunghoi Kim, Tae-Kyun Kim, Taesup Kim, Woo Chang Kim, Yubin Kim, Yaman Kindap, Robert Kirk, Varsha Kishore, W. Bradley Knox, Murat Kocaoglu, Rafal Kocielnik, Junpei Komiyama, Mark Kong, Egor V. Kostylev, Klemen Kotar, Yiwen Kou, Humaira Kousar, Vojtěch Kovařík, James Kozloski, Gabriel Kreiman, Ranjay Krishna, Satyapriya Krishna, Dmitrii Krylov, Jan Kulveit, Navdeep Kumar, Raunak Kumar, Zeb Kurth-Nelson, Temur Kutsia, Joe Kwon, Siddhartha Laghuvarapu, Thibault Lahire, Rudolf Laine, Avantika Lal, Florian Lalande, Alex Lamb, Christoph H. Lampert, Martin Lange, Max Langenkamp, Anders Lansner, Luca A. Lanzendörfer, Matti Lassas, Tiep Le, Changhyeon Lee, Donghwan Lee, Joowon Lee, Junghyun Lee, Junkyu Lee, Messi H.J. Lee, Seungwon Lee, Yunsung Lee, Felix Leeb, Joel Lehman, Gabor Lengyel, Borja González León, Andrew C. Li, Bing Li, Changjiang Li, Chengrui Li, Gang Li, Jiatong Li, Junyi Li, Kaican Li, Lixiang Li, Miqing Li, Mingsong Li, Sheng Li, Sida Li, Tongxin Li, Xia Li, Xiner Li, Xuan Li, Yin Li, Yuhang Li, Yunfan Li, Zexi Li, Xiang Lian, Luofeng Liao, Yi-Lun Liao, Derek Lim, Jihao Andreas Lin, Qihang Lin, Yen Ting Lin, David Lindner, Ori Linial, Viliam Lisý, Cheng-Hao Liu, Daochang Liu, Gaowen Liu, Han Liu, Jacqueline T. Liu, Jiacheng Liu, Jiaqi Liu, Jing Liu, Ollie Liu, Shicheng Liu, Shuai Liu, Tianle Liu, Tianyang Liu, Wei Liu, Weiming Liu, Xiyang Liu, Zhihan Liu, Morgan Livingston, Abraham Loeb, Aryo Lotfi, Adam S. Lowet, Aurélie Lozano, Jianglin Lu, Kai Lu, Wang Lu, Sitao Luan, Zheng Luan, G. Bruno De Luca, Martin Lukac, Fangzhou Luo, A. I. Lvovsky, Clare Lyle, Hanbaek Lyu, Hanjia Lyu, Haoyu Ma, Shaocong Ma, Shiqing Ma, Yinghao Ma, Yining Ma, Yuheng Ma, Zinnia Ma, Alaa Maalouf, Luo Mai, Aleksandar Makelov, Shreshth A. Malik, Debmalya Mandal, Laura Manduchi, Charles C. Margossian, Riccardo Marin, Ivan Marisca, Alex Markham, Samuele Marro, Sohir Maskey, Federico Matteucci, Katie Matton, Cynthia Matuszek, Tamás Matuszka, Vasilios Mavroudis, David Mayo, Michael T. McCann, Calvin McCarter, Jingbiao Mei, Alexander Meinke, William Merrill, Panayotis Mertikopoulos, Evi Micha, Julian Michael, Beren Millidge, Martin Renqiang Min, Koen Minartz, Aaron Mishkin, Panagiotis Misiakos, Surbhi Mittal, Munyque Mittelmann, Bruno Mlodozeniec, Sangwoo Mo, Mahdiyar Molahasani, Juston Moore, Adrià Marcos Morales, Alex Morehead, Lili Mou, Richard Moulange, Ramy Mounir, Jesse Mu, Kushin Mukherjee, Mark Niklas Müller, Nicolas M. Müller, Andrei Ioan Mureșan, James M. Murphy, Vít Musil, Mirco Mutti, Andrew J. Nam, Giung Nam, Amani Namboori, Neel Nanda, Andrea Nascetti, Arash Nasr-Esfahany, Emanuele Natale, Sriraam Natarajan, Utkarsh Nath, Carlos Navarrete, Siddharth Nayak, Christopher Nemeth, Robert Osazuwa Ness, Dung Thuy Nguyen, Hoang H. Nguyen, Quoc Viet Hung Nguyen, Thy Nguyen, Xuan Son Nguyen, Kim A. Nicoll, Ercong Nie, Zhijie Nie, Zehao Niu, Antonio Norelli, Igor Nunes, Ifeoma Nwogu, James Odgers, Caspar Oesterheld, Seong Joon Oh, Tuomas Oikarinen, Alex Olshevsky, Michael A. Osborne, Shreyas Padhy, Sebastian Pagel, Rasmus Pagh, Brooks Paige, Jiachun Pan, Mohit Pandey, Konstantinos Panousis, Stefano Panzeri, David Pape, Leonard Papenmeier, Orr Paradise, Joonsuk Park, Seongsik Park, Keith Y. Patarroyo, Zeel B Patel, Adrien Pavão, Hammond Pearce, Jaakko Peltonen, Abhirama Subramanyam Penamakuri, Fábio Perez, Fernando Perez-Cruz, Laurent Perrussel, Jan Peters, Bao Pham, Maurizio Pierini, Dario Piga, Silvia Pitis, Sergey N. Pozdnyakov, Akshara Prabhakar, Arnu Pretorius, Yanyi Pu, Yunze Qi, Wei Qian, Xiaoning Qian, Dong Qiao, Chao Qin, Wenjie Qiu, Yuyang Qiu, Jason Rader, Can Rager, Ahmad Rahimi, Mohammad Mahdi Rahimi, Md Musfiqur Rahman, Senthil Rajasekaran, Cyrus Rashtchian, Richa Rastogi, Udaybhan Rathore, Maya Ravichandran, Yi Ren, Zhiyuan Ren, Gavin Rens, Marcello Restelli, Oliver Richardson, Till Richter, Bastian Rieck, Franz Rieger, Jonathan Roberts, Nicholas Roberts, Alexander Robey, Iván F. Rodriguez, Cristóbal Rojas, Willem Röpke, Valentino Delle Rose, Benjamin Rosman, Saptarshi Roy, Shuvendu Roy, Tim G. J. Rudner, David Rügamer, Francisco J. R. Ruiz, Phillip Rust, Hobin Ryu, Shoham Sabach, Dimitris Sacharidis, Vin Sachidananda, Aadirupa Saha, Anwar Said, Shinsaku Sakaue, Dvir Samuel, Matthias Samwald, Pedro Sandoval-Segura, Karthik Abinav Sankararaman, Hrishav Sapkota, Aseem Saxena, Michael Saxon, Enrico Scala, Andreas Schlaginhaufen, Patrick Schnell, Philip Schniter, Martin Schrimpf, Nikolaj Schwartzbach, Jonas Schweisthal, Matthew Scott, Dino Sejdinovic,

Yusuke Sekikawa, Mariia Seleznova, Daniil Selikhanovych, Diksha Sethi, Naman Shah, Nisarg Shah, Rohin Shah, Yash Shah, Tamar Rott Shaham, Hatef Otroshi Shahreza, Deepak Kumar Sharma, Judy Hanwen Shen, Tony Shen, Zhu Shen, Daqian Shi, Wenqi Shi, Yucheng Shi, Zhenmei Shi, Zijing Shi, Shahaf Shperberg, Umer Siddique, Carlo Siebenschuh, Sandeep Silwal, Jisoo Sim, Guillem Simeon, Dario Simionato, Berfin Şimşek, Harvineet Singh, Vasu Singla, Gustav Šír, Joar Skalse, Alexander Soen, Chris Solomon, Ziyang Song, Daniel Sonntag, Neel Sortur, Alexandra Souly, Joshua Southern, Benjamin A. Spiegel, Srinath Sridhar, Gaurang Sriramanan, Trevor Standley, Kristoffer Stenso-Smidt, Simon Steshin, Xuan Su, Shreyas Subramanian, Sriram Ganapathi Subramanian, Anand Subramoney, Yang Sui, Bhavya Sukhija, Douglas Summers-Stay, Jiakai Sun, Ke Sun, Lichao Sun, Qingyao Sun, Rui Sun, Ruiyang Sun, Shao-Hua Sun, Yan Sun, Vijaykaarti Sundarapandiyan, Stanisław Szufa, Arush Tagade, Mohammad Sadegh Talebi, Xiaoyang Tan, Hongyao Tang, Jiajun Tang, Luyao Tang, Siyi Tang, Tianyi Tang, Zeyu Tang, Muhammad Faaiz Taufiq, Cem Tekin, Elizaveta Tennant, Lucile Ter-Minassian, Mikhail Terekhov, Kartik Thakral, Michael Thielscher, Zoran Tiganj, Berk Tinaz, Sidney Tio, Karn Tiwari, Massimiliano Todisco, Ljupčo Todorovski, Fabricio Arend Torres, Nikolaos Tsilivis, Nikita Tsoy, Murad Tukan, Kanji Uchino, Szilvia Ujváry, Daeho Um, Berk Ustun, Karthik Valmeekam, Alexandre Variengien, John J. Vastola, Sahil Verma, Claire Vernade, Raul Vicente, Nicholas Vincent, Jilles Vreeken, Lasse Vuursteen, Nikhil Vyas, Binxu Wang, Danqing Wang, Fangyin Wang, Hanjing Wang, Jindong Wang, Junxiang Wang, Limei Wang, Liyuan Wang, Qi (Cheems) Wang, Ting Wang, Xiaosen Wang, Yanbang Wang, Yancheng Wang, Yimu Wang, Yufei Wang, Zedong Wang, Zhe Wang, Zhenhailong Wang, Zhenzhen Wang, Zhichao Wang, Zhihao Wang, Zirui Wang, Joe Watson, Honghao Wei, Ran Wei, Zeming Wei, Ziquan Wei, Andre Wibisono, Patrick Wienhöft, Veit D. Wild, Jeffrey Willette, Daniel J. Williams, Michael Witbrock, Zach Wood-Doughty, Chao Wu, Fang Wu, Junlin Wu, Renzhi Wu, Shenghao Wu, Shirley Wu, Xiao Wu, Yi Wu, Ying Wu, Kacper Wyrwal, Tingsong Xiao, Binghui Xie, Yi Xie, Yuxi Xie, Deyi Xiong, Guojun Xiong, Cheng Xu, Haoran Xu, Hong Xu, Kaiwen Xu, Nuo Xu, Renzhe Xu, Shizhou Xu, Xiaojian Xu, Yanxun Xu, Zihao Xu, Wangqi Xue, Sangeeta Yadav, Hao Yan, Dingqi Yang, Shentao Yang, Yi-Rui Yang, Yuheng Yang, Zonghai Yao, Zeyuan Yin, Lance Ying, Jinsung Yoon, Yuning You, Zunzhi You, Kenny Young, Jiahao Yu, Jialin Yu, Qi Yu, Bo Yuan, Han Yuan, Jiakang Yuan, Jun Yuan, Xiao Yue, Zihao Yue, Sheung Man Yuen, EungGu Yun, Alp Yurtsever, Daniele Zambon, Riccardo Zamboni, Benito van der Zander, Santiago Zanella-Béguelin, Valentina Zantedeschi, Jakob Zeitler, Eric Zelikman, Shenglai Zeng, Daochen Zha, Huixin Zhan, Xianyuan Zhan, Cheng Zhang, Chuxu Zhang, Fuxiang Zhang, Hongming Zhang, Hongyang Zhang, Huan Zhang, Jianwei Zhang, Jinghan Zhang, Mengmi Zhang, Mingjian Zhang, Qi Zhang, Shuo Zhang, Tianrong Zhang, Xiang Zhang, Xiangyu Zhang, Xingyuan Zhang, Yan Zhang, Yanjia Zhang, Yivan Zhang, Yu Zhang, Yuheng Zhang, Yuhui Zhang, Yuyang Zhang, Zechen Zhang, Canche Zhao, Xuandong Zhao, Yang Zhao, Yu Zhao, Arman Zharmagambetov, Haotian Zheng, Yizhen Zheng, Andy Zhou, Doudou Zhou, Fan Zhou, Jin Peng Zhou, Joey Tianyi Zhou, Mingyuan Zhou, Yihan Zhou, Zhangpeng Zhou, Banghua Zhu, Beier Zhu, Shixiang Zhu, Xiatian Zhu, Zhenyu Zhu, Zhihui Zhu, Shlomo Zilberstein, Zhuofan Zong

Appendix A: Participation bias and question-level response bias

Reducing bias is an important issue for a survey such as this where opinion may vary substantially between different groups, and participation may also be motivated by expression of some opinions.

This survey involved substantial effort to reduce biases:

1. We disseminated the survey very widely and neutrally: we collected as many email addresses as we could for researchers publishing in the relevant venues, and wrote to each individually, rather than relying on sources of dissemination or encouragement that could induce a biased sample. We took substantial care to reduce minor sources of bias in this email collection procedure, such as from varying rates of name overlap among researchers from different cultures.
2. We sought a high response rate (and seem to have achieved a relatively high one, though comparison is difficult between surveys): we paid participants a substantial sum to participate (this was a large part of the financial expense of the survey), sent repeated reminders, and in previous years experimented to improve the appeal of the invitation.
3. We avoided methods of increasing the response rate that might induce bias, such as seeking endorsement from prominent researchers, or encouraging people to share the survey.
4. In particular, we kept the description of the topic vague in the invitation—"the future of AI progress"—to minimize the possibility of people deciding to participate according to their opinions on the narrower topics. (There was still some potential for this, for people who recognized or looked up the name of the survey, the sender (Katja Grace), or the organization (AI Impacts), or followed links in the email.)

For the 2023 ESPAI (Grace et al., 2025), as well as attempting to minimize bias in ways very similar to the above, we investigated the possibility of remaining bias, given the results we collected. We released a separate FAQ detailing this and responding to circulating misconceptions (Grace, 2025). For this iteration of the survey we have not repeated the investigation, but given the similarity of the survey results, it would be surprising if the situation were not very similar.

In 2023:

- Each normal question was answered by on average 96% of those who saw it, leaving little room for bias from people skipping questions.
- Of people who answered at least one question, 95% got to seeing the final demographics question at the end, leaving little room for bias from people dropping out after learning the more specific topics of the survey.
- When asked "How much thought have you given in the past to social impacts of smarter-than-human machines?" only 10.3% answered that they had 'a particular interest' in the topic, bounding the fraction of participants who might plausibly be from the AI Safety community or otherwise ideologically motivated to take the survey to contribute to perception of risks (a commonly proposed model of bias in this survey).
- For respondents who answered 'a little' or 'very little' to the above question—i.e. those who had at most discussed the topic a few times—the median probability of "human extinction or similarly permanent and severe disempowerment of the human species" from advanced AI (asked with or without further conditions) was 5%, the same as for the entire group. This means that even if more concerned people were overrepresented, that would not be affecting the notably high reported probabilities of extinction or similar disempowerment.

- Some demographic groups were more likely to answer than others, and we noted two where different groups had both different propensities to respond and different patterns of response among those who did. Women only participated at around 66% of the base rate, and generally expected less extreme positive or negative outcomes. However, since women were only around one tenth of the background population, the scale of error from it is limited. People who did undergraduate study in Asia were only 84% as likely as the base rate to respond, and in mean aggregate expected high-level machine intelligence earlier, and had higher median extinction or disempowerment numbers. These and other less notable patterns suggest various small biases from demographic variation in response rates, but no particular reason to expect these to systematically push results in a particular direction.

For further discussion and analysis, see (Grace, 2025).

Appendix B: ESPAI methodological changes between 2023 and 2024

Questionnaire:

- In questions referencing future years, year values were incremented by one, e.g. the question in Section 3.6 which asked about 2028 in 2023 became 2029.
- A potentially outdated elaboration in the one-shot learning task question was removed.
- The year in the survey title and the email address on the consent page were updated.

Communication:

- We made a variety of small changes in our communications (e.g. invitation emails, reminders, names and affiliations of people issuing invitations) to invited participants. These changes included wording, timing, and formatting. We also made some small structural changes (e.g. slight changes to the incentive payment process).
- In 2024 some participants received their invitations later on in the collection period due to incorrect omission from the original outreach list.

Data cleaning and analysis:

- Data cleaning was similar but not identical.
- Data cleaning in both years parsed some non-numeric answers into numeric answers (e.g. 'about 5' into '5') and excluded others, but cleaning in 2024 probably parsed a smaller set of possible non-numeric inputs.
- The code used for the analysis was not identical, though the intention was for it to perform the same function. Slight differences in algorithms for gamma fitting likely introduced immaterial differences.

Other:

- The 2023 ESPAI was approved by the Ethics Committee of the University of Bonn (248/23-EP); the 2024 survey did not undergo a comparable formal ethics review.
- In 2024 we used our list of previously contacted participants as a source for identifying emails for our invitation candidates. It is unclear whether we did this in 2023. This could introduce a slight bias toward AI researchers who were also identified in an earlier survey.
- In 2024 an error in compiling the invitation list caused many ineligible individuals to be invited to participate. Submissions were checked for eligibility, and responses from participants not meeting the criteria outlined in the Recruitment section were excluded.

Appendix C: Survey materials and external resources

Full survey text, as presented to participants: http://aiimpacts.org/wp-content/uploads/2026/09/2024-Expert-Survey-on-Progress-in-AI.pdf

This link also includes the full text of Stuart Russell's argument for why highly advanced AI might pose a serious risk.

Anonymized and cleaned response data: http://aiimpacts.org/wp-content/uploads/2026/09/2024-expert-survey-anonymized.csv

Invitation and reminder letters: 2024 survey emails, reminders and front matter

List of full task descriptions: https://tinyurl.com/aitasks

FAQ for 2023 ESPAI methodology (which closely matches the methodology for this survey): https://blog.aiimpacts.org/p/faq-expert-survey-on-progress-in

Appendix D: Framing effects

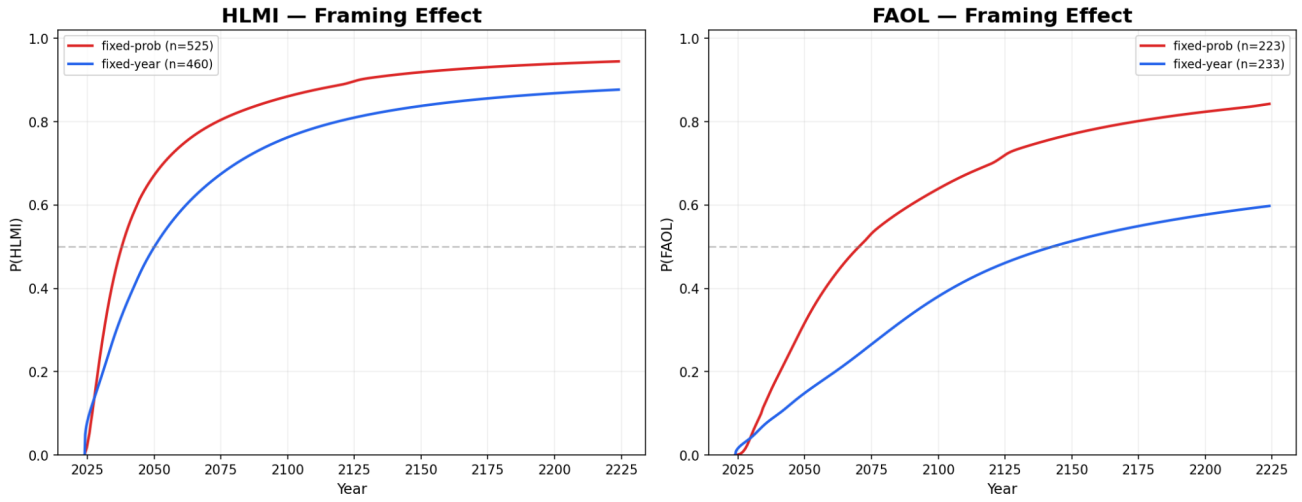


Figure 25: ***CDF comparison between the fixed-years framing and fixed-probabilities framing conditions for HLMI and FAOL.*** *Each curve is the mixture (mean) CDF for respondents assigned to that framing condition. The fixed-probabilities framing consistently produces earlier predictions than the fixed-years framing, and the HLMI framing consistently produces earlier predictions than the FAOL framing.*

Appendix E: Aggregation method comparison

Different choices in how individual survey responses are fitted and combined into aggregate forecasts can shift the headline numbers. Motivated by Adamczewski's reanalysis of ESPAI data (Adamczewski, 2024), this appendix compares three approaches on the key milestones:

- **Mean aggregation with mean-squared-error fitting** — the method used in past analyses of the three previous ESPAI surveys, and used most in this paper. Fit a gamma CDF to each respondent's three data points using mean squared error, then take the pointwise mean of all fitted CDFs. Because the method averages bounded probabilities rather than forecast years, each respondent's direct influence remains limited to their equal weight in the sample.
- **Median aggregation with mean-squared-error fitting** — same fitting, but take the pointwise median instead of mean. Less sensitive than the mean to a minority of extremely early or late forecasts. Adamczewski argues the median better represents the "typical expert". We report this for human-level AI forecasts in the main paper.
- **Mean aggregation with log-loss fitting** — fit using log-loss (cross-entropy) instead of MSE, then take the mean. Log-loss naturally weights errors near p=0 and p=1 more heavily than errors near p=0.5, matching the intuition that a five-percentage-point error at p=0.95 represents a much larger shift in belief than the same error at p=0.50.

Table 6: ***Headline 50% years generally change by a small number of years with the choice of mean or median, and less than three years with the choice of loss.*** *Switching from mean to median aggregation pushes the 50% year for HLMI from 2042 to 2044 and FAOL from 2098 to 2101. Switching from MSE to log-loss barely changes the aggregate; this is consistent with Adamczewski's finding that fitting choice has "hardly any impact." However, 50% years tend to be less affected than the tails, e.g., 10% years* (Adamczewski, 2024).

Milestone	Mix-Mean (MSE)	Mix-Median (MSE)	Mix-Mean (LogLoss)	N (MSE)	N (LogLoss)
HLMI	2042 (18.5 yr)	2044 (20.0 yr)	2043 (19.0 yr)	985	985
FAOL	2098 (74.2 yr)	2101 (77.1 yr)	2096 (72.0 yr)	456	456
Truck Driver	2033 (8.5 yr)	2034 (9.7 yr)	2033 (8.5 yr)	459	459
Surgeon	2050 (26.3 yr)	2053 (29.3 yr)	2051 (26.6 yr)	458	458
Retail Salesperson	2031 (7.3 yr)	2032 (8.3 yr)	2031 (7.4 yr)	460	460
AI Researcher	2052 (27.9 yr)	2055 (30.6 yr)	2052 (27.6 yr)	458	458

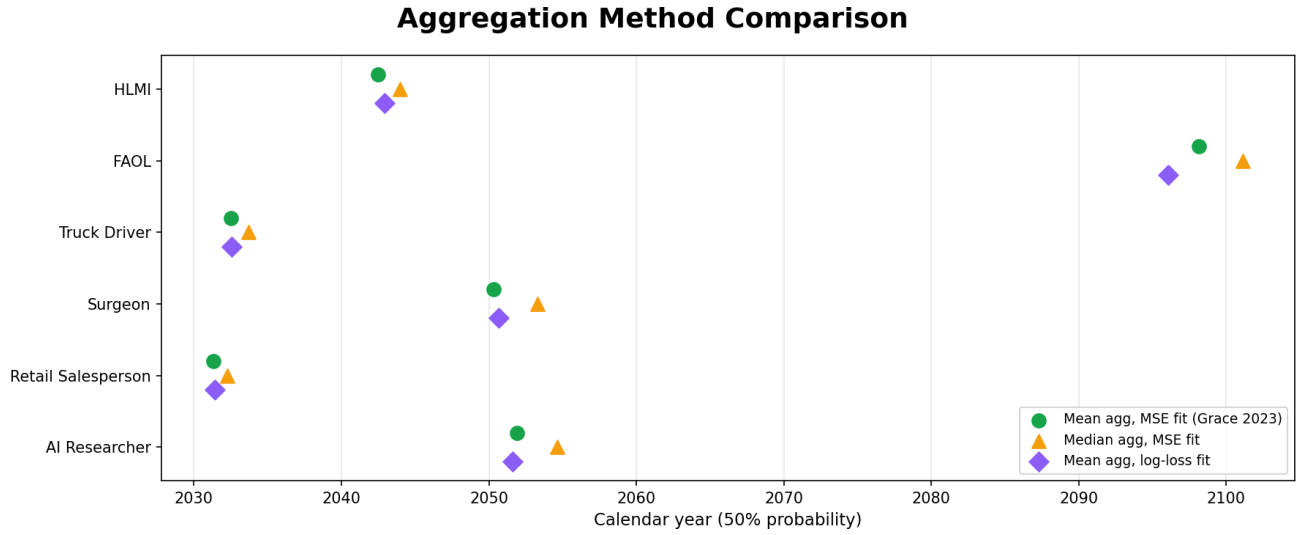


Figure 26: ***Regardless of loss, 50% year results with median aggregation are later than results with mean.***

Appendix F: Tabular data for AI milestone predictions

Table 7: ***Predicted attainment of AI milestones****, from Figure 4. Labels are short names; the full task descriptions given to participants are linked in Appendix C. Year values are years from the date of the survey (December 2024), with equivalent calendar years in parentheses. The 10%, 50% or 90% "prob year" is the year at which the mixture CDF reaches 10%, 50%, 90%.*

Milestone	N (fitted)	10% Prob Year	50% Prob Year	90% Prob Year
Write high-school history essay	129	0 yr (2024)	2 yr (2026)	10 yr (2034)
Write Python code (e.g. quicksort)	111	0 yr (2024)	2 yr (2026)	11 yr (2035)
Play new Angry Birds levels (superhuman)	167	0 yr (2024)	2 yr (2026)	10 yr (2034)
Win World Series of Poker	139	0 yr (2024)	2 yr (2026)	15 yr (2039)
Answer Googleable factoid questions	144	0 yr (2024)	3 yr (2027)	23 yr (2047)
Group unseen objects into classes	146	0 yr (2024)	3 yr (2027)	19 yr (2043)
Voice acting from text	141	0 yr (2024)	3 yr (2027)	14 yr (2038)
Translate as well as a fluent bilingual non-translator	158	0 yr (2024)	3 yr (2027)	18 yr (2042)
Answer Googleable open-ended questions	157	0 yr (2024)	3 yr (2027)	23 yr (2047)
Transcribe speech (noisy, accents)	135	0 yr (2024)	3 yr (2027)	16 yr (2040)
Beat best StarCraft II players	147	0 yr (2024)	3 yr (2027)	18 yr (2042)
3D model from short video	151	0 yr (2024)	3 yr (2027)	18 yr (2042)
Answer questions with no definite answer	146	0 yr (2024)	4 yr (2028)	20 yr (2044)
Produce song indistinguishable from artist	149	0 yr (2024)	4 yr (2028)	20 yr (2044)
Build website with payment processing	134	0 yr (2024)	4 yr (2028)	20 yr (2044)
Translate speech from subtitled films	120	0 yr (2024)	4 yr (2028)	24 yr (2048)
Beat novices on 50% of Atari games (20 min training)	148	0 yr (2024)	4 yr (2028)	32 yr (2056)
Phone banking services	143	0 yr (2024)	4 yr (2028)	26 yr (2050)
Fine-tune open source LLM	139	0 yr (2024)	5 yr (2029)	30 yr (2054)
Beat pro game testers on all Atari games	150	0 yr (2024)	5 yr (2029)	25 yr (2049)
Compose US Top 40 song (full audio)	148	0 yr (2024)	5 yr (2029)	47 yr (2071)
Match best humans in Putnam competition	158	0 yr (2024)	5 yr (2029)	31 yr (2055)
Translate newly discovered language (Rosetta Stone)	122	0 yr (2024)	6 yr (2030)	55 yr (2079)
Fold laundry (speed + quality)	162	1 yr (2025)	6 yr (2030)	35 yr (2059)
Play random game as human novice (<10 min)	140	1 yr (2025)	6 yr (2030)	38 yr (2062)
Learn efficient sorting (no solution form)	135	0 yr (2024)	6 yr (2030)	51 yr (2075)
Write NYT best-seller novel	155	0 yr (2024)	6 yr (2030)	66 yr (2090)
One-shot image recognition	134	1 yr (2025)	7 yr (2031)	37 yr (2061)
Explain game AI moves to a non-expert	150	1 yr (2025)	7 yr (2031)	48 yr (2072)
Beat best Go players (limited training)	148	1 yr (2025)	7 yr (2031)	88 yr (2112)
Assemble any LEGO set	141	1 yr (2025)	7 yr (2031)	32 yr (2056)
Find & patch security flaw (100k+ users)	135	1 yr (2025)	7 yr (2031)	63 yr (2087)
Replicate ML conference study	140	2 yr (2026)	8 yr (2032)	53 yr (2077)
Beat fastest human in 5 km city race (bipedal robot)	125	1 yr (2025)	9 yr (2033)	40 yr (2064)
Discover physics equations from simulation	132	1 yr (2025)	11 yr (2035)	130 yr (2154)
Conduct ML research & write conference paper	164	2 yr (2026)	12 yr (2036)	159 yr (2183)
Prove publishable math theorems	149	2 yr (2026)	15 yr (2039)	259 yr (2283)
Install electrical wiring in new home	130	3 yr (2027)	16 yr (2040)	97 yr (2121)
Solve unsolved math problem (e.g. Millennium)	130	4 yr (2028)	29 yr (2053)	798 yr (2822)

Appendix G: Supplemental figures

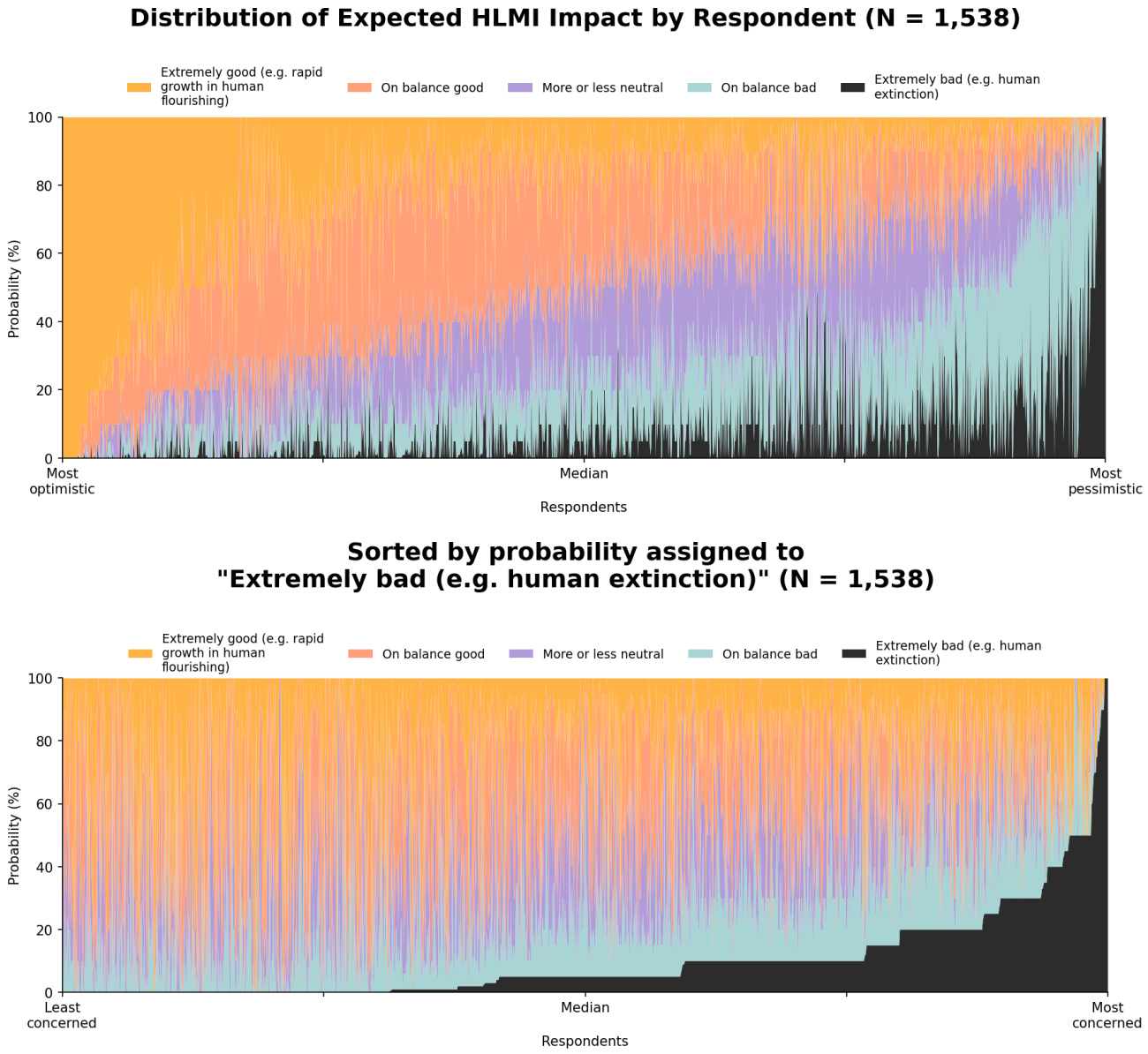


Figure 27: ***Predicted distribution of long-run impacts of HLMI on humanity*** *(from Section 3.8). Each vertical slice shows one respondent's probability allocation between five categories of overall impact on humanity in the long run from HLMI. The top version was in Section 3.8; the bottom version is the same except ordered by probability of extremely bad outcomes alone, to highlight the distribution of expectations on that. Note that the 'extremely good' and 'extremely bad' categories were presented with examples.*

References

Tom Adamczewski. How should we analyse survey forecasts of AI timelines? bayes.net, December 2024. URL https://bayes.net/espai/.

AI Impacts. List of AI Impacts' citations. AI Impacts Wiki, 2024. URL https://wiki.aiimpacts.org/selected_citations. Updated January 25, 2024. Accessed September 10, 2026.

Itzhak Ben-David, John R. Graham, and Campbell R. Harvey. Managerial miscalibration. *The Quarterly Journal of Economics*, 128(4):1547–1584, 2013. doi: 10.1093/qje/qjt023. URL https://doi.org/10.1093/qje/qjt023.

Colin F. Camerer and Eric J. Johnson. The process-performance paradox in expert judgment: How can experts know so much and predict so badly? In K. Anders Ericsson and Jacqui Smith, editors, *Toward a General Theory of Expertise: Prospects and Limits*, pages 195–217. Cambridge University Press, New York, 1991. URL https://authors.library.caltech.edu/records/qrp4n-9az74.

Center for AI Safety. Statement on AI extinction risk, May 2023. URL https://safe.ai/work/statement-on-ai-extinction-risk. Released May 30, 2023. Accessed September 10, 2026.

Gerd Gigerenzer, Wolfgang Gaissmaier, Elke Kurz-Milcke, Lisa M. Schwartz, and Steven Woloshin. Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest*, 8(2):53–96, 2007. doi: 10.1111/j.1539-6053.2008.00033.x. URL https://doi.org/10.1111/j.1539-6053.2008.00033.x.

Katja Grace. FAQ: Expert survey on progress in AI methodology. AI Impacts blog, October 2025. URL https://blog.aiimpacts.org/p/faq-expert-survey-on-progress-in.

Katja Grace, John Salvatier, Allan Dafoe, Baobao Zhang, and Owain Evans. Viewpoint: When will AI exceed human performance? Evidence from AI experts. *Journal of Artificial Intelligence Research*, 62: 729–754, 2018. doi: 10.1613/jair.1.11222. URL https://doi.org/10.1613/jair.1.11222.

Katja Grace, Julia Fabienne Sandkühler, Harlan Stewart, Benjamin Weinstein-Raun, Stephen Thomas, Zach Stein-Perlman, John Salvatier, Jan Brauner, and Richard C. Korzekwa. Thousands of AI authors on the future of AI. *Journal of Artificial Intelligence Research*, 84, 2025. doi: 10.1613/jair.1.19087. URL https://doi.org/10.1613/jair.1.19087.

Colleen McClain, Brian Kennedy, Jeffrey Gottfried, Monica Anderson, and Giancarlo Pasquini. How the U.S. public and AI experts view artificial intelligence. Pew Research Center, April 2025. URL https://www.pewresearch.org/internet/2025/04/03/how-the-us-public-and-ai-experts-view-artificial-intelligence/.

Michelle McDowell and Perke Jacobs. Meta-analysis of the effect of natural frequencies on Bayesian reasoning. *Psychological Bulletin*, 143(12):1273–1312, 2017. doi: 10.1037/bul0000126. URL https://doi.org/10.1037/bul0000126.

Cian O'Donovan, Sarp Gurakan, Xiaomeng Wu, Jack Stilgoe, Nicolas Bert, Ekaterina Dmitrichenko, Embla Gjørva, Stella Liu, Teddy Zamborsky, and Tianyu Zhao. Visions, values, voices: a survey of artificial intelligence researchers. Zenodo, April 2025. URL https://zenodo.org/records/15118399.

OpenAI. On the Navier–Stokes Millennium Prize Problem, September 2026. URL https://openai.com/index/navier-stokes-solution/.

Stuart Russell. Of myths and moonshine. Edge.org, November 2014. URL https://www.edge.org/conversation/the-myth-of-ai#26015. Contribution to the discussion of "The Myth of AI".

Zach Stein-Perlman, Benjamin Weinstein-Raun, and Katja Grace. 2022 expert survey on progress in AI. AI Impacts, August 2022. URL https://aiimpacts.org/2022-expert-survey-on-progress-in-ai/.

Richard S. Sutton. AI succession. Keynote, World Artificial Intelligence Conference, Shanghai, July 2023. URL http://incompleteideas.net/Talks/waic3.pdf. Slides; video at https://www.youtube.com/watch?v=NgHFMolXs3U. Accessed September 13, 2026.

Philip E. Tetlock. *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton University Press, revised edition, 2017. doi: 10.2307/j.ctt1pk86s8. URL https://www.jstor.org/stable/j.ctt1pk86s8. Originally published in 2005.

Émile Torres. Should humanity go extinct? Exploring the arguments for traditional and Silicon Valley pro-extinctionism. *The Journal of Value Inquiry*, December 2025. doi: 10.1007/s10790-025-10072-7. URL https://doi.org/10.1007/s10790-025-10072-7.

Amos Tversky and Daniel Kahneman. The framing of decisions and the psychology of choice. *Science*, 211(4481):453–458, 1981. doi: 10.1126/science.7455683. URL https://doi.org/10.1126/science.7455683.

Amos Tversky and Daniel Kahneman. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90(4):293–315, 1983. doi: 10.1037/0033-295X.90.4.293. URL https://doi.org/10.1037/0033-295X.90.4.293.