Thousands of AI Authors on the Future of AI

KATJA GRACE*, AI Impacts, Berkeley, USA
JULIA FABIENNE SANDKÜHLER*, Department of Psychology, University of Bonn, Germany
HARLAN STEWART*, AI Impacts, Berkeley, USA
BENJAMIN WEINSTEIN-RAUN*, Independent, USA
STEPHEN THOMAS, AI Impacts, Berkeley, USA
ZACH STEIN-PERLMAN, AI Impacts, Berkeley, USA
JOHN SALVATIER, Independent, USA
JAN BRAUNER†, Department of Computer Science, University of Oxford, UK
RICHARD C. KORZEKWA*†, AI Impacts, Berkeley, USA

In October 2023, 2,778 researchers who had published in top-tier artificial intelligence (AI) venues gave predictions on the pace, nature and impacts of AI progress. Significant steps were taken to minimize and evaluate bias. In evaluations of participation bias, we found that most groups responded at similar rates. The participants estimated that several milestones had at least a 50% chance of being feasible for AI by 2028, including constructing a payment processing site and fine-tuning an LLM. If science continues undisrupted, the chance of unaided machines outperforming humans in every possible task was estimated at 10% by 2027 and 50% by 2047—13 years earlier than in our 2022 survey ($N = 738$). The chance of all occupations becoming fully automatable, however, was not expected to reach 10% until 2037, and 50% until 2116 (compared to 2164 in the 2022 survey).

Most respondents expressed substantial uncertainty about long-term impacts: While 68% in 2023 thought good outcomes from high-level machine intelligence AI were more likely than bad ones, 48% of these net optimists gave at least a 5% chance of extremely bad outcomes. Conversely, 59% of net pessimists gave 5% or more to extremely good outcomes. Depending on how we asked, between 38% and 51% of respondents gave at least a 10% chance to advanced AI leading to outcomes as bad as human extinction. More than half suggested that "substantial" or "extreme" concern is warranted about AI increasing misinformation, boosting authoritarian control, worsening inequality, and other scenarios. There was broad agreement that research aimed at minimizing risks from AI systems ought to be more prioritized.

JAIR Track: Surveys

JAIR Associate Editor: Virginia Dignum

*These authors contributed equally to this work.
†Shared senior authorship.

Authors' Contact Information: Katja Grace, ORCID: 0000-0002-8902-4548, KATJA@AIIMPACTS.ORG, AI Impacts, Berkeley, USA; Julia Fabienne Sandkühler, ORCID: 0000-0002-5585-9539, JF.SANDKUEHLER@GMAIL.COM, Department of Psychology, University of Bonn, Germany; Harlan Stewart, HARLAN@AIIMPACTS.ORG, AI Impacts, Berkeley, USA; Benjamin Weinstein-Raun, ORCID: 0009-0007-3945-8966, B@W-R.ME, Independent, USA; Stephen Thomas, ORCID: 0009-0002-1747-4619, STEVEKWTHOMAS@GMAIL.COM, AI Impacts, Berkeley, USA; Zach Stein-Perlman, ZACHARYSTEINPERLMAN@GMAIL.COM, ORCID: 0000-0002-7245-519X, AI Impacts, Berkeley, USA; John Salvatier, ORCID: 0000-0002-1784-5312, JSALVATIER@GMAIL.COM, Independent, USA; Jan Brauner, ORCID: 0000-0002-1588-5724, JAN.BRAUNER@GMX.DE, Department of Computer Science, University of Oxford, UK; Richard C. Korzekwa, ORCID: 0000-0003-2611-4914, RICK@AIIMPACTS.ORG, AI Impacts, Berkeley, USA.

DOI: 10.1613/jair.1.19087

JAIR Reference Format:
Katja Grace, Julia Fabienne Sandkühler, Harlan Stewart, Benjamin Weinstein-Raun, Stephen Thomas, Zach Stein-Perlman, John Salvatier, Jan Brauner, and Richard C. Korzekwa. 2025. Thousands of AI Authors on the Future of AI. *Journal of Artificial Intelligence Research* 84, Article 9 (October 2025), 48 pages. DOI: 10.1613/jair.1.19087

1 Introduction

Artificial intelligence appears poised to reshape society. Decision-makers are working to address opportunities and threats due to AI in the private sector (OpenAI 2023), academia (CHAI 2023), and government at the state, national, and international levels (Office of Governor Gavin Newsom 2023; Biden 2023; Inter-Agency Working Group on Artificial Intelligence 2022). Navigating this situation requires judgments about how the progress and impact of AI are likely to unfold. One important source of evidence is the predictions of AI researchers. Their familiarity with the technology and its past dynamics puts them in a good position to make educated guesses about the future of AI. However, they are experts in AI research, not AI forecasting, and might thus lack generic forecasting skills, or expertise in non-technical factors that influence the trajectory of AI. While AI experts' predictions should not be seen as a reliable guide to objective truth, they can provide one important piece of the puzzle.

We conducted two large surveys in 2022 and 2023 of AI researchers who had published peer-reviewed research in the prior year in top AI venues. The 2023 survey, to our knowledge, constitutes the largest survey of AI researchers to date. It took place in the fall of 2023, after an eventful year of broad AI progress (including the launch of ChatGPT), shifts in public awareness of AI issues (including two widely signed and publicized AI safety letters (Future of Life Institute 2023; Center for AI Safety 2023), and governments world-wide beginning to address questions of AI regulation (The White House 2023; *Written Testimony of Dario Amodei, Ph.D. Co-Founder and CEO, Anthropic for a Hearing on "Oversight of A.I.: Principles for Regulation* 2023; *AI Safety Summit 2023* 2023; *EU AI Act: first regulation on artificial intelligence* 2023). The two very similar surveys included questions about the speed and dynamics of AI progress, and the social consequences of more advanced AI. This paper complements recent work gathering views on similar questions from the public (Stein-Perlman 2022) and corporate leadership (Chui et al. 2023), as well as previous surveys of AI experts conducted in 2009 (Baum et al. 2011), 2011 (Sandberg and Bostrom 2011), 2012/13 (Müller and Bostrom 2016), 2016 (Grace et al. 2018; Etzioni 2016), 2017 (Walsh 2018), 2018 (Gruetzemacher et al. 2020; Anderson et al. 2018) and 2019 (Zhang et al. 2021).

2 The Survey

This paper reports on two surveys, one conducted between October 5 and 24, 2023 and the other between June 12 and August 3, 2022. They were very similar to each other and to a previous survey conducted in 2016 (Grace et al. 2018). The 2023 survey included around four times as many participants ($N = 2778$) as the 2022 survey ($N = 738$) by including six publication venues (NeurIPS, ICML, ICLR, AAAI, IJCAI, JMLR) instead of just two (NeurIPS and ICML). Respondents from the four new and the two original conferences had similar opinions to each other (see 5.4.3 and Appendix D).

The 2023 survey also included several new questions to probe the nature of future AI systems and diverse potential risks.

To assess and mitigate framing effects (Tversky and Kahneman 1981), we often posed different variations of questions on the same topic to different random subsets of respondents. For example, 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 to estimate the year by which the milestone would be feasible with a given probability ("fixed-probabilities framing"). To minimize confusion, each participant received one framing throughout the entire survey.

In several parts of the survey, each participant randomly received questions on only one of several topics, to keep the survey brief. This means that most questions were not assigned to all participants.

For more details about our methodology, see Appendix A.

3 Results on AI Progress

The following section presents the results related to the progress of AI.

3.1 How Soon Will 39 Tasks Be Feasible for AI?

The survey asked about when each of 39 tasks would become "feasible," meaning that "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." Each respondent was asked about four tasks, so that each task received around 250 estimates. Each respondent gave three probability-year pairs per task.

To aggregate the responses, we first fit a gamma distribution to each participant's three probability-year pairs, and then computed the mean across the participants' individual gamma distributions.

In the 2023 survey, all but six of the 39 tasks were predicted to have at least a 50% chance of being feasible within the next ten years (Figure 1 and Appendix E). This included several very economically valuable tasks—such as coding an entire payment processing site from scratch and writing new songs indistinguishable from real ones by hit artists such as Taylor Swift. It also included tasks that imply substantial progress in sample-efficiency (e.g. beating novices in 50% of Atari games after 20 minutes of play), AI-driven AI progress (e.g. autonomously fine-tuning an open-source LLM), and robotics (e.g. folding laundry).

32 AI task questions were identical in 2023 and 2022. All tasks from Grace et al. (2018) were included, regardless of whether the authors would judge them to be already achieved.

Between 2022 and 2023, predictions moved earlier for 21 out of 32 of tasks and later for 11 tasks (Figure 1). On average, for the 32 tasks included in both the 2022 and 2023 surveys, the 50th percentile year they were expected to become feasible shifted 1.0 years earlier (SD = 2.0, SE = 0.18).

However, compared to 2016, the 2022 and 2023 predictions for a number of milestones moved later. Out of the 18 tasks asked in 2016 (Grace et al. 2018), 3 predictions moved earlier and 15 moved later.

3.2 How Soon Will Human-Level Performance on All Tasks or Occupations Be Feasible?

We asked how soon participants expected AI systems to outperform humans across all activities, framed as either *tasks*, in the question about "High-Level Machine Intelligence" (HLMI), or *occupations*, in the question about "Full Automation of Labor" (FAOL).

3.2.1 How Soon Will 'High-Level Machine Intelligence' Be Feasible? We defined High-Level Machine Intelligence (HLMI) thus:

> "High-level machine intelligence (HLMI) is achieved 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.*"

We asked for predictions, assuming "human scientific activity continues without major negative disruption." We aggregated the results (n = 1714) by fitting gamma CDFs, as with individual task predictions in 3.1.

In both 2022 and 2023, respondents gave a wide range of predictions for how soon HLMI would be feasible (Figure 2). The 2023 forecast predicted a 50% chance of HLMI by 2047, down thirteen years from 2060 in the 2022 survey. We checked if this difference in predicted time until a 50% chance of HLMI was significant for participants who received the question in the fixed-probabilities framing, and found that it was (p = .0001, Yuen's test; Appendix E). For comparison, in the six years between the 2016 and 2022 surveys, the expected date moved

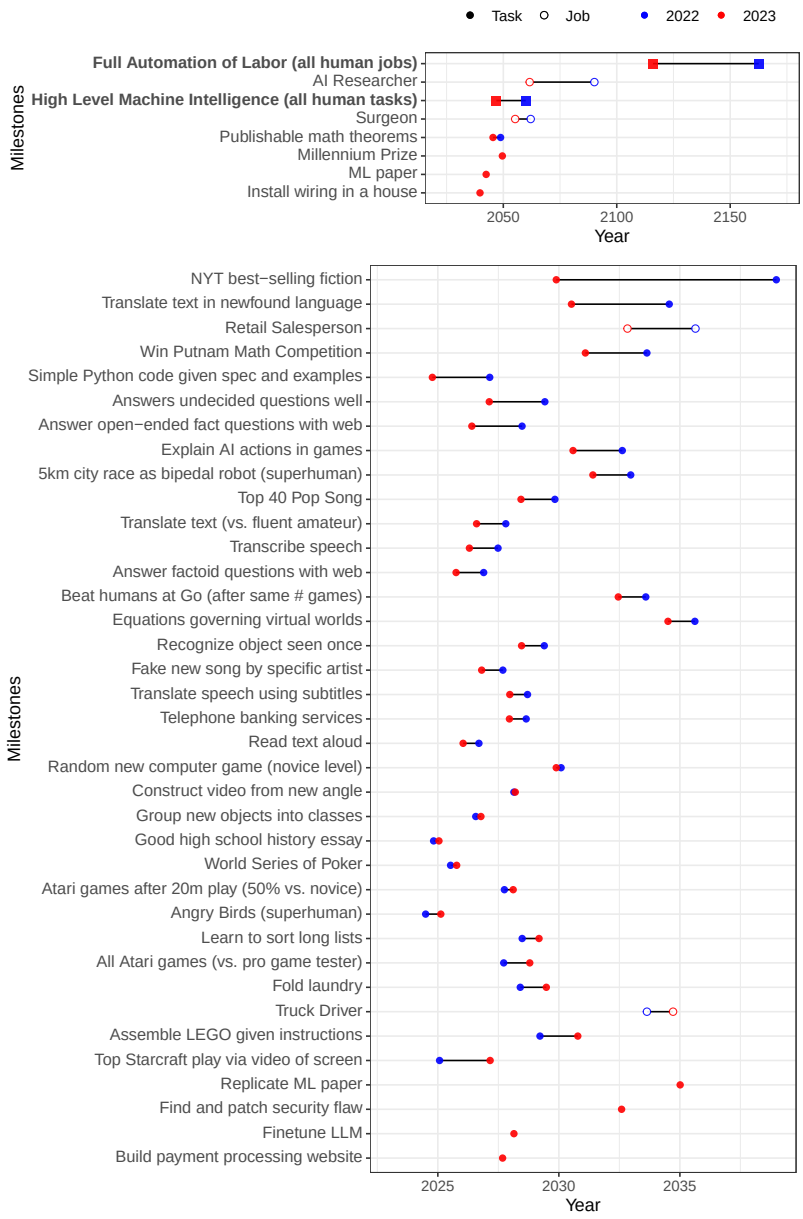


Fig. 1. Expected feasibility of many AI milestones moved substantially earlier in the course of one year (between 2022 and 2023). The milestones are sorted (within each scale-adjusted chart) by size of drop from 2022 forecast to 2023 forecast, with the largest change first. Labels are summaries of longer milestone descriptions (see Appendix B for these). The year when the aggregate distribution gives a milestone a 50% chance of being met is represented by solid circles, open circles, and solid squares for tasks, occupations, and general human-level performance respectively. These three groups of questions have different formats that may also influence answers.

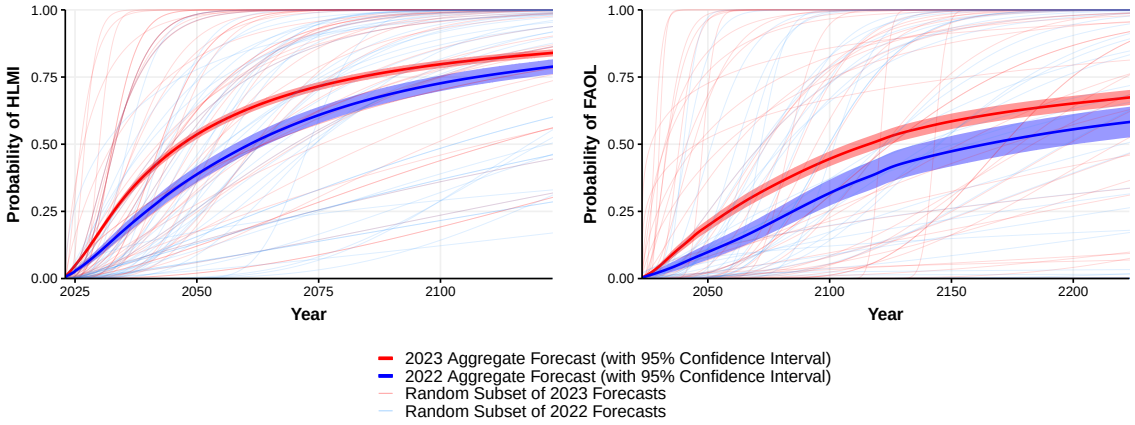


Fig. 2. Forecasts for 50% chance of arrival of High-Level Machine Intelligence (HLMI) dropped by 13 years between 2022 and 2023; forecasts for Full Automation of Labor (FAOL) dropped by 48 years. Forecasts were aggregated by taking the mean distribution over all individual cumulative distribution functions. To give a sense of the range of responses, we included random subsets of individual 2023 and 2022 forecasts. Note that the thinner 'confidence interval' in 2023 compared to 2022 is due to our increased confidence about the average respondents' views due to a larger sample size, not respondents' predictions converging.

only one year earlier, from 2061 to 2060. The 2023 forecast predicted a 10% chance of HLMI by 2027, down two years from 2029 in the 2022 survey.

In all three surveys (2016, 2022, 2023), experts whose undergraduate degree was in Asia expected HLMI significantly sooner than experts whose undergraduate degree was in North America (Appendix D).

3.2.2 How Soon Will 'Full Automation Of Labor' Be Feasible? The other framing of the question about how soon AI systems would outperform humans across all activities was about "Full Automation Of Labor," or FAOL. We defined FAOL thus:

> "Say an occupation becomes 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.* [...]
>
> [...] Say we have reached 'full automation of labor' 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."

Before the participants ($n = 774$) were asked about the full automation of labor, they were asked for predictions about when four specific occupations would become fully automatable: truck driver, surgeon, retail salesperson, and AI researcher (Figure 1). They were also asked to think of an existing human occupation that they thought would be among the final ones to be fully automatable, and to predict the timing of this. They were then asked when 'full automation of labor' (FAOL) would be achieved.

The 2023 forecast predicted a 50% chance of FAOL by 2116, down 48 years from 2164 in the 2022 survey (Figure 2). We checked if this difference was significant for participants who received the question in the fixed-probabilities framing, and found that it was ($p = .0052$, Yuen's test; Appendix E). In 2016 (Grace et al. 2018) predictions for FAOL were in between those of 2022 and 2023 (2016: 2136; 2022: 2164; 2023: 2116). Similarly, all

four specific occupations were predicted earlier in 2023 compared to 2022, but the comparison to 2016 is more mixed: three occupations (surgeon, retail salesperson, and truck driver) were predicted later in 2022/2023 than in 2016, while "AI researcher" moved earlier. There was about a 70-year difference between the mean 50% prediction for HLMI and the mean 50% prediction for FAOL. We discuss this finding in the next section, "Differences between HLMI and FAOL".

3.2.3 Differences Between HLMI and FAOL. Predictions for a 50% chance of the arrival of FAOL have consistently been more than sixty years later than those for a 50% chance of the arrival of HLMI across the 2023, 2022, and 2016 survey results. This is surprising because HLMI and FAOL are quite similar: FAOL asks about the feasible automation of all occupations; HLMI asks about the feasible automation of all tasks. Since occupations might naturally be understood either as complex tasks, composed of tasks, or closely connected with one of these, achieving HLMI seems to either imply having already achieved FAOL, or suggest being close.

We do not know what accounts for this gap in forecasts. Insofar as HLMI and FAOL refer to the same event, the difference in predictions about the time of their arrival would seem to be a framing effect. However it is possible that participants tend to think of these as less closely connected, or that other differences between the questions substantially mattered. In particular, only the HLMI questions were preceded by the instruction to "assume that human scientific activity continues without major negative disruption," and the FAOL block asked a sequence of questions about the automation of specific occupations before asking about full automation of labor.

3.3 Framing Effect of Fixed-Years vs Fixed-Probabilities

All questions about how soon a milestone would be reached were framed in two ways: fixed-years and fixed-probabilities. All displayed results include responses from both framings combined. In either framing, we asked for three year-probability pairs, but in one we fixed a set of probabilities (10%, 50%, 90%) and asked how many years until the participant would assign each probability to the milestone being met, whereas in the other we fixed a set of future years (usually 10 years, 20 years, 50 years) and asked about the probability of the milestone occurring by that year.

As in 2016 (Grace et al. 2018), the fixed-years framing produced systematically later predictions in both 2022 and 2023. For example, the year with a 50% chance of HLMI from participants answering in the fixed-year frame (34 years) was twice as far into the future as that for participants answering in the fixed-probability frame (17 years). However, it is notable that even the larger of these two was shorter than 2022's combined forecast (37 years), demonstrating a substantial shift of predictions closer to the present.

3.4 Change in Observed Rates of Progress and Extrapolation from Progress Rates

We asked respondents which AI area they had worked in for the longest, and whether progress in that area was faster in the first or second half of their time working in it. Similarly to 2016, in 2022 and 2023 most respondents reported faster progress in the second half (2016 67%, 2022 56%, 2023 60%), while only 10%, 18% and 17% respectively said progress had been faster in the first half. This suggests that AI progress has accelerated over the course of the respondents' research careers.

We asked about the fraction of progress until human level achieved in respondents' respective areas so far since they started working in this area, and used these estimates (ignoring acceleration) as a third way to predict the timing of general human-level performance (Appendix E). The median projection reached human-level performance (for a specific area) in 19 years, not vastly different from the 24 year forecast to a 50% chance of HLMI yielded by asking directly, but much earlier than the 93 year forecast for a 50% chance of FAOL.

3.5 Will There Be an Intelligence Explosion?

We asked respondents about the possibility, after HLMI is hypothetically achieved, of an 'intelligence explosion', as explained in this question:

> 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?

The median answer in both 2022 and 2023 was "even chance (41-60%)". The other two 'intelligence explosion' questions were about the likelihood of a dramatically increased rate of global technological advancement 2 and 30 years post-HLMI (medians: 20% and 80%), and the likelihood that AI that is vastly better than humans across all professions 2 and 30 years post-HLMI (medians: 10% and 60%). The medians for the three questions were identical to 2016 except that for the last question it was 50% in 2016.

In sum, across these three questions, the median participant did not overall expect an intelligence explosion, but did give substantial credence to the possibility (Appendix E).

3.6 What Will AI Systems in 2043 Be Like?

Concerns about risks from future AI systems are often linked to specific traits related to alignment, trustworthiness, predictability, self-directedness, capabilities, and jailbreakability. In 2023 but not 2022, we asked respondents how likely it was that at least some state-of-the-art AI systems in 2043 would have each of eleven such traits ($n = 649 - 667$).

All 11 traits were considered to have a relatively high chance of existing in AI systems in 2043. Only one trait had a median answer below 'even chance': "Take actions to attain power." While there was no consensus even on this trait, it is notable that it was deemed least likely, because it is arguably the most sinister, being key to a prominent argument for extinction-level danger from AI (Carlsmith 2022).

Answers reflected substantial uncertainty and disagreement among participants. No trait attracted near-unanimity on any probability, and no more than 55% of respondents answered "very likely" or "very unlikely" about any trait (Appendix E).

3.7 Will AI in 2028 Truthfully and Intelligibly Explain its Decisions?

Uninterpretable reasoning in AI systems is often considered an AI risk factor, potentially leading to outcomes ranging from unjust biases in treatment of people to active pursuit of harm hidden by capable agents. In 2023, we thus asked about the interpretability of AI systems in five years:

> For typical state-of-the-art AI systems in 2028, 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.

Respondents overall found it unlikely that AI systems in 2028 would be interpretable like this. Twenty-six percent said "very unlikely (<10%), 35% said "unlikely (10-40%), 20% said "even odds (20%), 15% said "likely (60-90%), and only 5% said "very likely (>90%). This is related to the question in section 3.7, which asked how likely it was that at least some state-of-the-art AI systems in 2043 (fifteen years later) "can be trusted to accurately explain

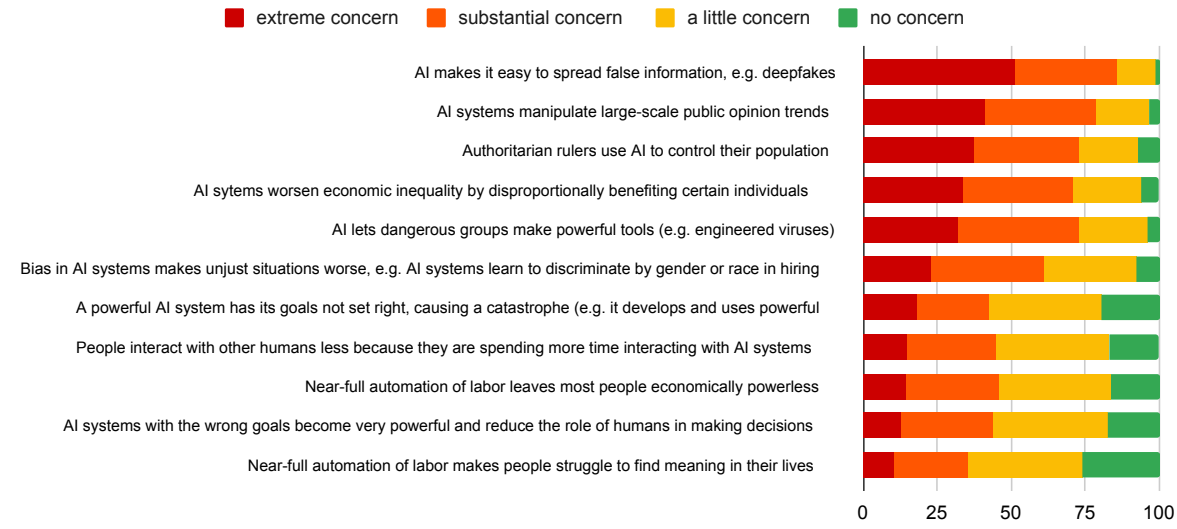


Fig. 3. Amount of concern potential scenarios were reported to deserve. Organized from most to least extreme concern.

their actions." The median answer there was "even chance" (40-60% likely), compared to "unlikely" (10-40%) on this question.

4 Results on the Impacts of AI

The following section presents the results related to the impacts of AI.

4.1 How Concerning Are 11 Future AI-Related Scenarios?

We asked participants (n = 1345) about eleven potentially concerning AI scenarios, such as AI-enabled misinformation, worsened economic inequality, and biased AI systems worsening injustice. We asked how much concern each deserved in the next thirty years (Figure 3).

Each scenario was considered worthy of either substantial or extreme concern by more than 30% of respondents. As measured by the percentage of respondents who thought a scenario constituted either a "substantial" or "extreme" concern, the scenarios worthy of most concern were: spread of false information e.g. deepfakes (86%), manipulation of large-scale public opinion trends (79%), AI letting dangerous groups make powerful tools (e.g. engineered viruses) (73%), authoritarian rulers using AI to control their populations (73%), and AI systems worsening economic inequality by disproportionately benefiting certain individuals (71%). An "other" option was added by only 115 participants and thought to be of "substantial" or "extreme" concern by 75% of them.

There is some ambiguity about the reason why a scenario might be considered concerning: it might be considered especially disastrous, or especially likely, or both. From our results, there is no way to disambiguate these considerations.

4.2 How Good or Bad for Humans Will High-Level Machine Intelligence Be?

We asked participants to assume that, at some point, "high-level machine intelligence" (HLMI) will exist, as defined in section 3.2. Given this assumption for the sake of the question, we asked how good or bad they expect the overall impact of this to be "in the long run" for humanity.

Fig. 4. Respondents exhibited diverse views on the expected goodness/badness of High Level Machine Intelligence (HLMI). We asked participants to assume, for the sake of the question, that HLMI will be built at some point. The figure shows a random subset of 800 responses from the 2023 survey on the positivity or negativity of long-run impacts of HLMI on humanity. Each vertical bar represents one participant and the bars are sorted left to right by a weighted sum of probabilities corresponding to overall optimism. Responses ranged from extremely optimistic to extremely pessimistic. Over a third of participants (38%) put at least a 10% chance on extremely bad outcomes (e.g. human extinction).

Respondents exhibited diverse views on the future impact of advanced AI (Figure 4). Many people who put high probabilities on bad outcomes also put high probabilities on good outcomes. In 2023, 64% assigned non-zero probabilities to both extremely good and extremely bad scenarios. 68% of participants found good outcomes more likely than bad outcomes, while 58% considered extremely bad outcomes (e.g. human extinction) at least 5% likely. Even among net optimists, nearly half (48%) gave at least 5% credence to extremely bad outcomes, and among net pessimists, more than half (59%) gave at least 5% to extremely good outcomes.

In 2023, the median prediction for extremely bad outcomes, such as human extinction, was 5% (mean 9%). Seniority had little effect on this assessment (Appendix D). Over a third of participants (38%) put at least a 10% chance on extremely bad outcomes. This is comparable to, but somewhat lower than, rates of assigning at least 10% to extinction-level outcomes in answers to other questions more directly about extinction, between 41% and 51% (see section 4.3).

Since 2022, mean overall probability on extreme outcomes (good *or* bad) has fallen slightly (Figure 5). The proportion of people who put at least a 10% chance on extremely bad outcomes (e.g. human extinction) fell from 48% in 2022 to 38% in 2023, and the mean prediction for this type of outcome decreased from 14% to 9%. In 2016, the median probability given to extremely good outcomes (20%) was twice as high as in 2022/2023 (10%), while the median probability for extremely bad outcomes has remained constant at 5%.

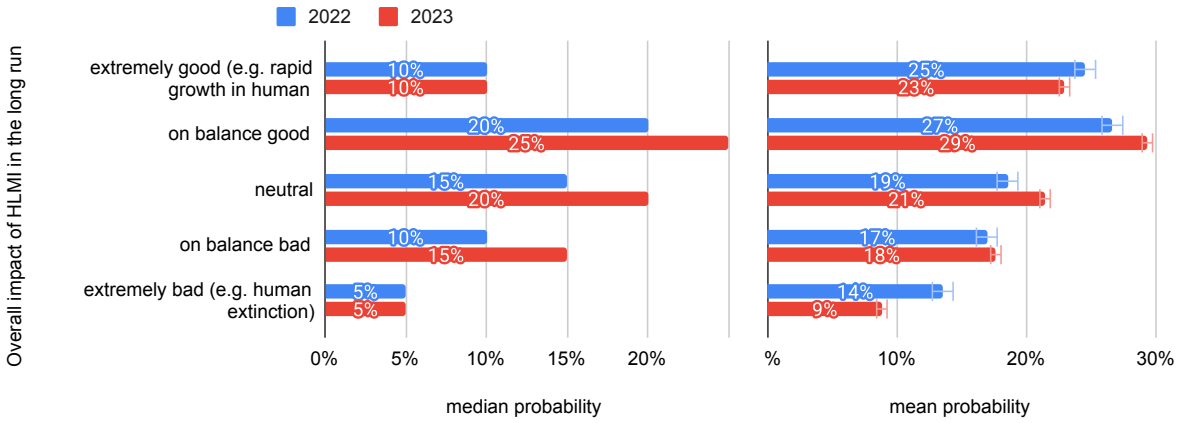


Fig. 5. Mean but not median predictions in 2023 (n = 2704) about the consequences of HLMI shifted slightly away from extreme outcomes in 2023 compared to 2022 (n = 559). Error bars indicate the standard error.

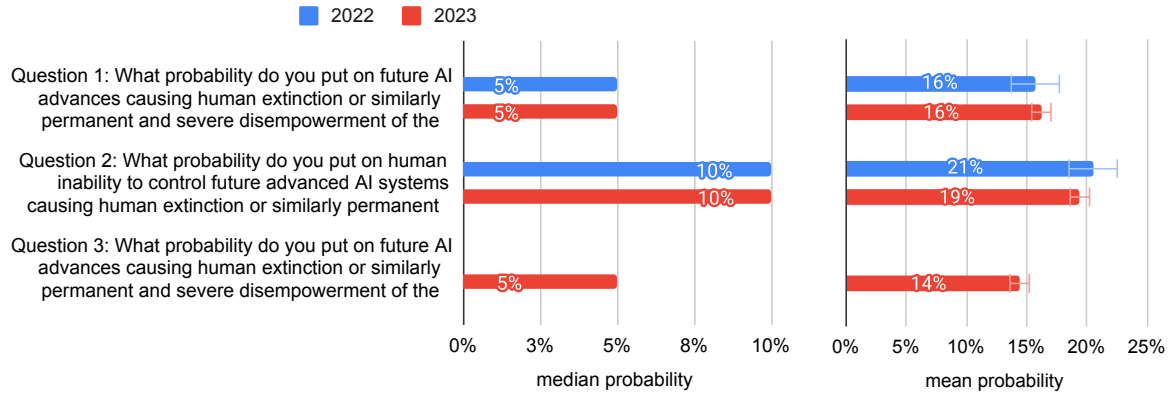


Fig. 6. Mean and median predictions to three questions on human extinction. Error bars indicate the standard error. (Question 1 n = 149 in 2022 and 1321 in 2023. Question 2 n = 162 in 2022 and 661 in 2023. Question 3 was asked only in 2023, n = 655).

4.3 How Likely Is AI to Cause Human Extinction?

To further clarify views on the "extremely bad (e.g. human extinction)" scenarios in the question on overall impacts, participants were given one of three similar questions about human extinction. Their differences were intended to help learn how concerning different scenarios are, what respondents expect to happen, and how much difference wording makes.

Answers to these questions were mostly consistent, with medians of 5% or 10% (Figure 6). They were also close to "extremely bad (e.g. human extinction)" answers to the question on general value of long-run impact (which all participants received)[1], which might suggest that the bulk of the extremely bad outcomes imagined involved

[1]"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%."

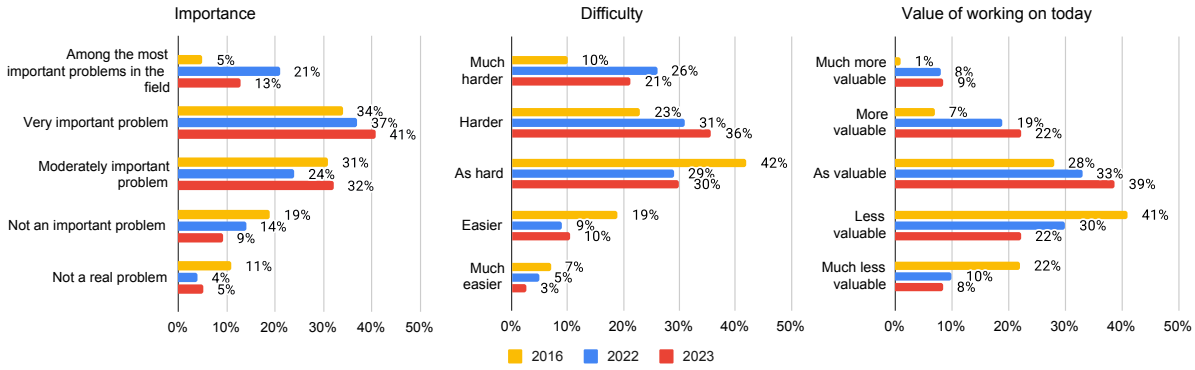


Fig. 7. Attitudes towards Stuart Russell's description of a key concern with highly advanced AI. Participants viewed the problem as important and difficult, but not more valuable to work on today than other problems.

human extinction or similarly permanent and severe disempowerment of the human species, as opposed to other bad outcomes.

Depending on how we asked, in 2023 between 41% and 51% of respondents estimated a greater than 10% chance of human extinction or severe disempowerment (question 1: 47%, question 2: 51%, question 3: 41%, for demographic comparisons see Appendix A).

4.4 Views on Others' Concerns About AI

We asked (n = 671), "To what extent do you think people's concerns about future risks from AI are due to misunderstandings of AI research?". In 2023, 11% said "Almost entirely", 44% said "To a large extent", 29% said "Somewhat", 15% said "Not much", 2% said "Hardly at all". Responses in 2022 were similar: 11% said "Almost entirely", 48% said "To a large extent", 25% said "Somewhat", 14% said "Not much", and 3% said "Hardly at all". This may reflect a view amongst AI researchers that the general public misunderstands AI.

4.5 What Rate of AI Progress Would Produce the Most Optimism?

In 2023, we asked participants "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." There was disagreement on whether faster or slower progress would be preferable, though large divergence from the current speed was less popular. Only 5% said they would prefer "much slower" progress, 30% said "somewhat slower", 27% said "current speed", 23% said "somewhat faster", and 16% said "much faster".

4.6 Russell's Description of a Key Concern with Highly Advanced AI

A second set of AI safety questions was based on Stuart Russell's description of a key concern with highly advanced AI (Russell 2014). This set of questions began with a summary of Russell's argument—which claims that with advanced AI, "you get exactly what you ask for, not what you want"—then asked:

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

In 2023, the majority of respondents said that Russell's description of a key concern with highly advanced AI was either a "very important problem" (41%) or "among the most important problems in the field" (13%), and the majority said that it was "harder" (36%) or "much harder" (21%) than other problems in AI (Figure E2). However, participants did not generally think that it is more valuable to work on this problem today than on other problems. In 2022, responses showed the same pattern. The problem was rated as more important, urgent, and difficult to solve than in 2016.

4.7 How Much Should AI Safety Research Be Prioritized?

We asked respondents:

> Let 'AI safety research' include 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). "How much should society prioritize AI safety research, relative to how much it is currently prioritized?".

In addition, respondents before 2023 and half of respondents in 2023 received four examples. The other half of respondents in 2023 received an additional new example about biases in AI. Adding this example did not affect results, so we combined the two framings (Appendix E).

A large majority of respondents thought that AI safety research should be prioritized more. The percentage of researchers who thought so increased compared to earlier surveys, but only slightly since 2022 (Figure 7).

5 Discussion

The following section summarizes and discusses the findings.

5.1 Summary of Results

Participants expressed a wide range of views on almost every question: some of the biggest areas of consensus were on how wide-open possibilities for the future appear to be. This uncertainty is striking, but several patterns of opinion are particularly informative.

While the range of views on how long it will take for milestones to be feasible can be broad, the 2023 survey saw a general shift towards earlier expectations. Over the fourteen months between the 2022 and 2023 surveys, a similar participant pool expected something like general human-level performance 13 to 48 years sooner on average (depending on the question), and 21 out of 32 shorter term milestones were expected earlier.

Another striking pattern is widespread assignment of credence to extremely bad outcomes from AI. In both 2022 and 2023, a majority of participants considered advanced AI to pose at least a 5% chance of causing human extinction or similarly permanent and severe disempowerment of the human species, and this result was consistent across four different questions, two assigned to each participant. Across these same questions, between 38% and 51% placed at least 10% chance on advanced AI bringing these extinction-level outcomes.

In general, there was a wide range of views about expected social consequences of advanced AI, and most people put some weight on both extremely good outcomes and extremely bad outcomes. While the optimistic scenarios reflect AI's potential to revolutionize various aspects of work and life, the pessimistic predictions—particularly those involving extinction-level risks—serve as a stark reminder of the high stakes involved in AI development and deployment.

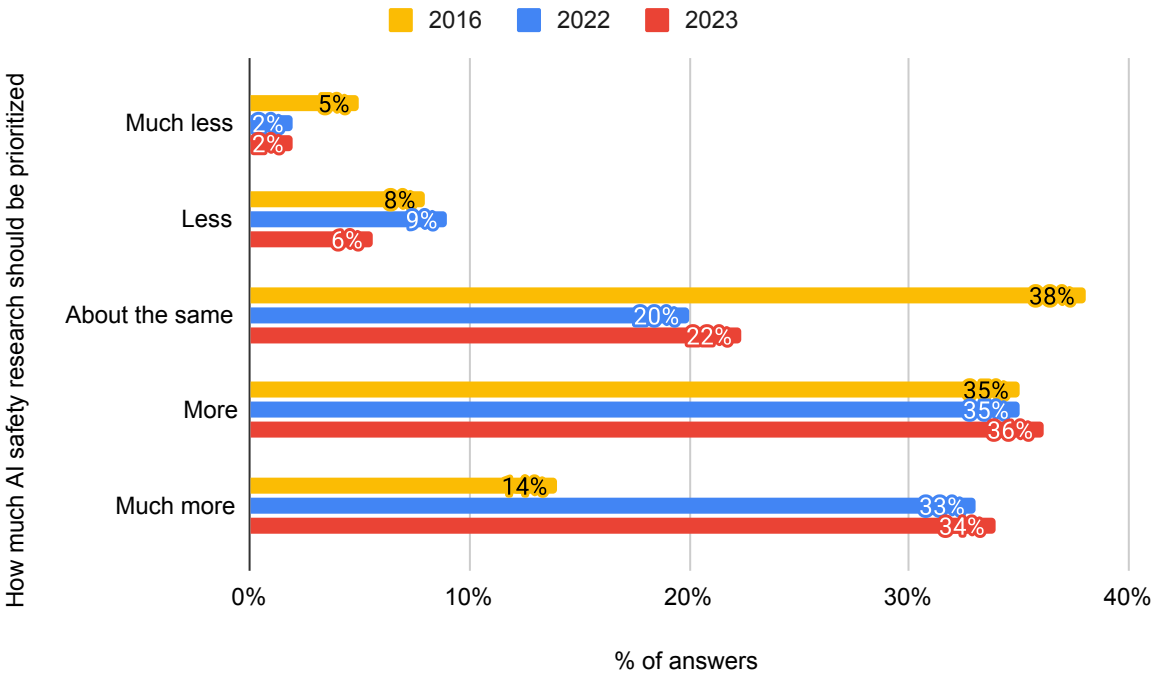


Fig. 8. 70% of respondents thought AI safety research should be prioritized more. This view has become considerably more common compared to 2016. 2016 $N = 167$; 2022 $N = 263$; 2023 $N = 1329$.

5.2 Comparison to the 2016 Survey and Other Years

While there was a clear shift to earlier predictions between 2022 and 2023, the comparisons to 2016 are more mixed. Compared to 2016, predictions shifted earlier for HLMI but later for many tasks and occupations. The 2016 FAOL prediction was between those of 2022 and 2023.

Between 2016 and 2022, researchers seem to have become somewhat less optimistic about impacts of AI being positive, and to have seen an increased importance in solving AI safety and alignment problems. The median probability given to extremely good outcomes of HLMI dropped by half between 2016 and 2022 (20% vs 10%), while the median probability for extremely bad outcomes remained constant at 5%. The problem that, with advanced AI, "you get exactly what you ask for, not what you want" as described by Stuart Russell (Russell 2014) was rated as more important, urgent, and difficult to solve since 2016. The view that AI safety research should be prioritized more has become considerably more common compared to in 2016 (49% in 2016 vs. 68% and 70% in 2022 and 2023, respectively).

Many other patterns remained constant over the years. This includes a perception that progress has been accelerating, and differences in predictions between framings and geographic regions.

5.3 Alignment with Other Surveys and Commentaries

Several other surveys and commentaries on related topics have been conducted in recent years. Previous works have discussed existential risk concerns, timelines to artificial general intelligence (Wynroe et al. 2023), economic

impacts (Brynjolfsson and Unger 2023), and misinformation concerns (Penn et al. 2023; Marken and Nicola 2023; YouGov 2023). Here, we compare and contrast our findings and methodology.

Prominent AI researchers and other commentators have expressed concern that AI poses an existential risk (Cellan-Jones 2014; Bengio et al. 2024; *FLI Open Letters* 2021; Center for AI Safety 2023). Our results suggest this concern is not limited to a small number of experts, but widely held within the AI research community.

Our results are broadly in line with previous expert surveys on existential risk from AI, which differ substantially from each other. In a survey of AI researchers that had published in ICML or NeurIPS in 2018, conducted in 2019 (sent to $N = 2652$, $N = 524$ participated), the median respondent expected HLMI to be more likely positive than negative (median probability for "extremely good (e.g., rapid growth in human flourishing)" 20% , "on balance good" 27%, "more or less neutral" 16%, "on balance bad" 10%, "extremely bad (e.g., human extinction)" 2%) (Zhang et al. 2022). In a 2022 survey of researchers who published at computational linguistics conferences, 36% agreed and 64% disagreed that "It is plausible that decisions made by AI or machine learning systems could cause a catastrophe this century that is at least as bad as an all-out nuclear war" (N = 327 participated out of the target population of N = 6323) (Michael et al. 2022). A survey in 2023 with 216 computer science professors found that most of their respondents were optimistic about where "we will land with AI in the end" (20% said "extremely optimistic", 52% said "somewhat optimistic"), but 15% of respondents said "somewhat pessimistic" and 2% said "extremely pessimistic" (The Generation Lab 2023). Karger et al. (2023) asked N=88 superforecasters, N=59 experts with expertise in AI, nuclear, and biorisk domains, and N=15 general existential risk experts in 2022 how likely they think that AI will, by 2100, cause humanity to go extinct. Superforecasters gave a median probability of 0.38% (bootstrapped 95% confidence intervals [0.10, 0.75]%), AI experts 3% [0.49, 10.00]%, nuclear and biorisk experts 2% [1.00, 4.03]%, general existential risk experts 4.75% [1.9, 14.0]%.

Companies such as OpenAI, Meta, DeepMind, and Anthropic aim to create general human-level or transformative AI (Heath 2024; Morris et al. 2024; Anthropic 2025; OpenAI 2025). A number of different model-based approaches have been proposed to forecast when this will be achieved (Wynroe et al. 2023). The median (i.e. 50% probability) year for these approaches differed greatly, from 2052 (biological anchors), 2064 (whole brain emulation), >2100 (semi-informative priors model, phase transitions model) to >3000 (insight-based model).

In contrast to model-based forecasts, which use explicit models to calculate timelines, judgment-based forecasts are based on expert opinion and are not necessarily (solely) the result of explicit models. Judgment-based forecasts reviewed by Wynroe et al. (2023) included the online forecasting platform Metaculus (Metaculus 2025) (median year: 2039 in 2023, dropped to 2031 in 2025), forecasting expert group Samotsvety (Yagudin et al. 2023) (median year: 2043), AI safety experts (Cotra 2022) (median year: 2040) and (Karnofsky 2021) (median year: 2060). Our two surveys presented in this paper (2022 and 2023) were also judgment-based forecasts with similar results for the High Level Machine Intelligence framing: The median year was 2060 in the 2022 survey and 2047 in the 2023 survey. The Full Automation of Labor framing in our surveys resulted in later median years than the other judgment-based forecasts (2164 in 2022 and 2116 in 2023).

The difference between model-based and judgment-based forecasts might be because the judgment-based forecasts disagreed with the methodology of the models or used different input parameters and focused on recent progress in machine learning. In addition, some forecasts were about artificial general intelligence, while others were about transformative AI. These two events are not necessarily the same, e.g. an AI might have a transformative impact on the economy without being general.

The 2024 AI Index is a report that tracks past progress in AI based on published results (Maslej et al. 2024). For example, the 2024 AI Index reported AI surpassing human performance on several benchmarks related to image classification, visual reasoning, and English understanding but trailing behind in competition-level mathematics, visual commonsense reasoning and planning (however note that the report was published in April 2024, and three months later, AlphaProof won a silver medal at the International Mathematical Olympiad (DeepMind 2024).

Our respondents predict this pattern will continue, expecting AI to write a bestselling novel 20 years before it is able to develop publishable math theorems or solve a long-standing unsolved problem in mathematics.

AI spreading misinformation was a major concern among our respondents. This result aligns with a recent public opinion poll by Harvard CAPS/Harris Poll (Penn et al. 2023) with 2,068 registered US voters. About 80% of respondents in each survey considered this an important concern. Similarly, a YouGov (2023) survey of 1000 citizens found that "The spread of misleading video and audio deep fakes" was the most common concern (85% concerned). AI causing unemployment was rated to be an important concern by 50% of respondents in the YouGov poll, 68% of respondents in the Harvard Harris poll, and 47% of our respondents. Relatedly, 75% of respondents in a public opinion poll by Gallup (Marken and Nicola 2023) expected AI to reduce the number of jobs. These differences might in part be explained by the severity implied in the questions, with more extreme scenarios being considered less likely ("reduce the total number of jobs" in Gallup, "replacement of jobs" in YouGov, "mass unemployment" in Harvard Harris, "near-full automation of labor leaves most people economically powerless" in the present survey). In the Harvard Harris poll, the minority of participants (1/3) who reported having "tried AI" were in general less concerned than those who had not tried it, while our respondents were very experienced in AI and their level of concern was similar to the voters who had not tried AI and higher than those who tried it.

Our survey also covered, among other things, the topics of economic inequality, interpretable AI, the intersection of biorisk and AI, bias in AI, and autonomous weapons. Much has been published on these topics, such as Brynjolfsson and Unger's (2023) commentary on possible impacts AI might have on economic inequality, more generally work by Erik Brynjolfsson on the future of work, overviews and reviews on interpretable AI (Bereska and Gavves 2024; Gilpin et al. 2018; Linardatos et al. 2020), analyses from the Center for Security and Emerging Technology and on the intersection of biorisk and AI (Batalis 2023), commentaries and reviews on bias in AI (Kundi et al. 2022; Zou and Schiebinger 2018), and reviews, surveys and commentaries by researchers (Zhang et al. 2021; Horowitz 2016; Scharre 2023) and NGOs (Human Rights Watch 2023) on autonomous weapons.

5.4 Caveats and Limitations

The following section discusses caveats and limitations of the study.

5.4.1 Forecasting Is Hard, Even for Experts. Forecasting is difficult in general, and subject-matter experts have been observed to perform poorly (Tetlock 2017; Savage et al. 2021). Our participants were AI experts, not forecasting experts. These judgments are difficult, and there are no established methods of making them well. To make informed decisions in this context, we must thus combine evidence from various noisy methods, such as extrapolating progress trends (Villalobos 2023); reasoning about reference classes of similar events (Grace et al. 2021); analyzing the nature of agents (Omohundro 2008); probing qualities of current AI systems and techniques (Park et al. 2023); applying economic models to AI scenarios (Jones 2023; Aghion et al. 2017; Trammell and Korinek 2023), forecasting aggregation systems such as markets, professional forecasters, and the judgments of various subject matter experts.

There are signs in this research and past surveys that these experts are not accurate forecasters across the range of questions we asked. For one thing, on many questions different respondents gave very different answers, which limits the number of them who can be close to the truth. Nonetheless, in other contexts, averages from a large set of noisy predictions can still be relatively accurate (Surowiecki 2005), so a question remains as to how informative these aggregate forecasts are.

Another piece of evidence against the accuracy of forecasts is the observation of substantial framing effects (see 3.2 and within 3.3). If seemingly unimportant changes in question framing lead to large changes in responses, this suggests that even aggregate answers to any particular question are not an accurate guide to the answer. In an extreme example elsewhere, Karger et al. (2023) found that lay people (college graduates) gave answers nearly six orders of magnitude apart when asked in different ways to estimate the size of existential risks from AI: When

given example odds of low-probability events, estimates were much lower. A similar effect might apply to our participants, though their subject-matter expertise in AI and quantitative training may reduce their susceptibility to large opinion shifts with provision of examples of low probability events. Also, contrary to what you may expect if this effect would be found here, participants who had thought more in the past about AI risks, and as such had more experience reasoning about low-probability events, gave higher credence to extremely bad outcomes (Appendix D).

Despite these limitations, AI researchers are well-positioned to contribute to the accuracy of our collective guesses about the future. Their familiarity with the relevant technology, and experience with the dynamics of its progress, make them among the best-positioned to make informative educated guesses.

5.4.2 Participation. The survey was taken by 15% in 2023 and 17% in 2022 of those we contacted. This appears to be within the typical range for a large expert survey. Based on an analysis by Hamilton (2003), the median response rate across 199 surveys was 26%, and larger invitation lists tended to yield lower response rates: surveys sent to over 20,000 people, like ours in 2023, were expected to have a response rate in the range of 10%. In specialized samples, such as scientists, lower response rates are common (Bray and Storch 2010). In 2023, out of the 2778 respondents, 2650 (95%) finished the survey. In 2022, out of the 738 respondents, 536 (73%) finished the survey. Each question was answered by over 90% of those who saw it.

As with any survey, our results could be skewed by participation bias, if participants had systematically different opinions than those who chose not to participate. We sought to minimize this effect by providing financial compensation for participation to maximize response rate (Marinescu et al. 2021), which had been found effective in increasing participation in our survey (Appendix A) , and by limiting cues about the survey content available before opting to take the survey (Appendix C). We looked for evidence of response bias at the survey level and question level for some questions, and did not find any that would affect the results to a large extent (Appendix C). Only a small minority of respondents indicated having thought extensively about the timing of smarter-than-human AI (7.6%) and its social impacts (10.3%). This would seem to count against the hypothesis that a large proportion of respondents were people for whom the future of AI is an ideological issue (Appendix C).

5.4.3 Change in Sample from 2022 to 2023. In 2022 we only surveyed researchers at NeurIPS and ICML, whereas in 2023 we also contacted researchers who published in ICLR, AAAI, IJCAI, and JMLR. This could make comparison between survey years less meaningful, if the populations have different opinions. However, the subset of 2023 participants who published in NeurIPS or ICML (58%) appeared to have very similar opinions to the full sample (Appendix D).

Acknowledgments

Many thanks for help with this research to Rebecca Ward-Diorio, Jeffrey Heninger, Nate Silver, Jimmy Rintjema, Joseph Carlsmith, Nick Beckstead, Howie Lempel, Leopold Aschenbrenner, Ramana Kumar, Justis Mills, Will MacAskill, Shakeel Hashim, Mike Levine, Lucius Caviola, Eli Rose, Nathan Young, Michelle Hutchinson, Arden Koehler, Isabel Juniewicz, Ajeya Cotra, Josh Kalla, Niki Howe, Seb Farquar, Nuño Sempere, Naomi Saphra, Scott Siskind, Sören Mindermann, Chana Messinger, Jacob Hilton, Abie Rohrig, Noemi Dreksler, Baobao Zhang, Owain Evans, David Gros, Frederic Arnold, Max Tegmark, Jaan Tallinn, Shahar Avin, Alexander Berger, Cate Hall, Howie Lempel, Jaime Sevilla, Daniel Kokotajlo, James Aung, Ryan Greenblatt, Dan Hendrycks, Alex Tamkin, Vael Gates, Yonadav Shavit, and others.

Additional Information

Correspondence and requests for materials should be addressed to KG.
Reprints and permissions requests should be directed to KG.
The anonymized data and full set of questions is available at osf.io/vmdny.

Funding

This research project was funded by Open Philanthropy, 182 Howard Street #225, San Francisco, CA 94105, USA, and by the Survival and Flourishing Fund, 23564 Calabasas Road, Suite 201, Calabasas, CA 91302, USA.

References

P. Aghion, B. F. Jones, and C. I. Jones (Oct. 2017). "Artificial Intelligence and Economic Growth". In: *NBER Working Papers Series*. Working Paper Series (23928). DOI: 10.3386/w23928. URL: http://www.nber.org/papers/w23928.

AI Safety Summit 2023 (2023). en. URL: https://www.gov.uk/government/topical-events/ai-safety-summit-2023 (visited on 01/18/2024).

J. Anderson, L. Rainie, and A. Luchsinger (2018). "Artificial intelligence and the future of humans". In: *Pew Research Center* 10 (12). URL: http://tony-silva.com/eslefl/miscstudent/downloadpagearticles/AIhumanfuture-pew.pdf.

Anthropic (2025). *Company*. en. URL: https://www.anthropic.com/company (visited on 01/21/2025).

S. Batalis (Dec. 12, 2023). *AI and Biorisk: An Explainer*. en. URL: https://cset.georgetown.edu/publication/ai-and-biorisk-an-explainer/ (visited on 03/27/2025).

S. D. Baum, B. Goertzel, and T. G. Goertzel (Jan. 1, 2011). "How long until human-level AI? Results from an expert assessment". In: *Technological forecasting and social change* 78 (1), pp. 185–195. ISSN: 0040-1625. DOI: 10.1016/j.techfore.2010.09.006. URL: https://www.sciencedirect.com/science/article/pii/S0040162510002106.

Y. Bengio, G. Hinton, A. Yao, D. Song, P. Abbeel, T. Darrell, Y. N. Harari, Y.-Q. Zhang, L. Xue, S. Shalev-Shwartz, G. Hadfield, J. Clune, T. Maharaj, F. Hutter, A. G. Baydin, S. McIlraith, Q. Gao, A. Acharya, D. Krueger, A. Dragan, P. Torr, S. Russell, D. Kahneman, J. Brauner, and S. Mindermann (May 24, 2024). "Managing extreme AI risks amid rapid progress". en. In: *Science (New York, N.Y.)* 384 (6698), pp. 842–845. ISSN: 0036-8075,1095-9203. DOI: 10.1126/science.adn0117. URL: https://www.science.org/doi/10.1126/science.adn0117 (visited on 03/20/2025).

L. Bereska and E. Gavves (Apr. 22, 2024). *Mechanistic interpretability for AI safety – A review*. arXiv: 2404.14082 [cs.AI]. URL: http://arxiv.org/abs/2404.14082.

J. Biden (2023). *AI.Gov: Making AI work for the American people*. en. URL: https://ai.gov/ (visited on 01/10/2024).

D. Bray and H. von Storch (2010). *CliSci2008: A survey of the perspectives of climate scientists concerning climate science and climate change*. Research rep. IInstitute of Coastal Research. URL: https://ncse.ngo/files/pub/polls/2010--Perspectives_of_Climate_Scientists_Concerning_Climate_Science_&_Climate_Change_.pdf (visited on 01/10/2024).

E. Brynjolfsson and G. Unger (Dec. 1, 2023). "The Macroeconomics of Artificial Intelligence". en. In: *International Monetary Fund*. URL: https://www.imf.org/en/Publications/fandd/issues/2023/12/Macroeconomics-of-artificial-intelligence-Brynjolfsson-Unger (visited on 01/18/2025).

J. Carlsmith (June 16, 2022). "Is Power-Seeking AI an Existential Risk?" In: *arXiv*. arXiv: 2206.13353 [cs.CY]. URL: http://arxiv.org/abs/2206.13353.

R. Cellan-Jones (Dec. 2, 2014). *Stephen Hawking warns artificial intelligence could end mankind*. en. URL: https://www.bbc.com/news/technology-30290540 (visited on 03/20/2025).

Center for AI Safety (2023). *Statement on AI risk*. en. URL: https://www.safe.ai/statement-on-ai-risk (visited on 01/10/2024).

CHAI (2023). *Research Publications*. en. URL: https://humancompatible.ai/research (visited on 01/18/2024).

M. Chui, L. Yee, B. Hall, A. Singla, and A. Sukharevsky (Aug. 1, 2023). *The state of AI in 2023: Generative AI's breakout year*. en. URL: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year (visited on 01/19/2024).

A. Cotra (2022). "Two-year update on my personal AI timelines". In: *LessWrong*. URL: https://www.lesswrong.com/posts/AfH2oPHCApdKicM4m/two-year-update-on-my-personal-ai-timelines (visited on 01/21/2025).

DeepMind (2024). *AI achieves silver-medal standard solving International Mathematical Olympiad problems*. en. URL: https://deepmind.google/discover/blog/ai-solves-imo-problems-at-silver-medal-level/.

O. Etzioni (Sept. 20, 2016). *No, the Experts Don't Think Superintelligent AI is a Threat to Humanity*. en. URL: https://www.technologyreview.com/2016/09/20/70131/no-the-experts-dont-think-superintelligent-ai-is-a-threat-to-humanity/ (visited on 04/25/2024).

EU AI Act: first regulation on artificial intelligence (Aug. 6, 2023). en. URL: https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence (visited on 01/18/2024).

FLI Open Letters (May 12, 2021). en. URL: https://futureoflife.org/fli-open-letters/ (visited on 03/20/2025).

Future of Life Institute (Mar. 22, 2023). *Pause Giant AI Experiments: An Open Letter*. URL: https://futureoflife.org/open-letter/pause-giant-ai-experiments/ (visited on 01/18/2024).

L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter, and L. Kagal (Oct. 2018). "Explaining explanations: An overview of interpretability of machine learning". en. In: *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*. 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA) (Turin, Italy). IEEE, pp. 80–89. ISBN: 9781538650905. DOI: 10.1109/dsaa.2018.00018. URL: https://ieeexplore.ieee.org/abstract/document/8631448 (visited on 03/27/2025).

K. Grace, R. Korzekwa, A. Bergal, and D. Kokotajlo (Mar. 8, 2021). *Discontinuous progress investigation*. en. URL: https://wiki.aiimpacts.org/ai_timelines/discontinuous_progress_investigation (visited on 01/18/2024).

K. Grace, J. Salvatier, A. Dafoe, B. Zhang, and O. Evans (July 31, 2018). "Viewpoint: When Will AI Exceed Human Performance? Evidence from AI Experts". en. In: *Journal of Artificial Intelligence Research* 62, pp. 729–754. ISSN: 1076-9757,1076-9757. DOI: 10.1613/jair.1.11222. URL: http://www.jair.org/index.php/jair/article/view/11222 (visited on 09/08/2023).

R. Gruetzemacher, D. Paradice, and K. B. Lee (Dec. 1, 2020). "Forecasting extreme labor displacement: A survey of AI practitioners". In: *Technological forecasting and social change* 161, p. 120323. ISSN: 0040-1625. DOI: 10.1016/j.techfore.2020.120323. URL: https://www.sciencedirect.com/science/article/pii/S0040162520311495.

M. B. Hamilton (2003). *Online Survey Response Rates and Times. Background and Guidance for Industry*. URL: https://web.archive.org/web/20041108224206if_/http://supersurvey.com:80/papers/supersurvey_white_paper_response_rates.pdf (visited on 01/18/2024).

A. Heath (Jan. 18, 2024). *Mark Zuckerberg's new goal is creating artificial general intelligence*. en. URL: https://www.theverge.com/2024/1/18/24042354/mark-zuckerberg-meta-agi-reorg-interview (visited on 01/21/2025).

M. C. Horowitz (Jan. 1, 2016). "Public opinion and the politics of the killer robots debate". In: *Research & Politics* 3 (1), p. 2053168015627183. ISSN: 2053-1680. DOI: 10.1177/2053168015627183. URL: https://doi.org/10.1177/2053168015627183.

Human Rights Watch (Feb. 14, 2023). *Review of the 2023 US Policy on Autonomy in Weapons Systems*. en. URL: https://www.hrw.org/news/2023/02/14/review-2023-us-policy-autonomy-weapons-systems (visited on 03/27/2025).

Inter-Agency Working Group on Artificial Intelligence (Sept. 20, 2022). *Principles for the Ethical Use of Artificial Intelligence in the United Nations System*. URL: https://unsceb.org/sites/default/files/2023-03/CEB_2022_2_Add.1%20%28AI%20ethics%20principles%29.pdf.

C. I. Jones (Nov. 2023). "The A.I. Dilemma: Growth versus Existential Risk". In: *NBER Working Papers Series*. Working Paper Series (31837). DOI: 10.3386/w31837. URL: http://www.nber.org/papers/w31837.

E. Karger, J. Rosenberg, Z. Jacobs, M. Hickman, R. Hadshar, K. Gamin, T. Smith, B. Williams, T. Mccaslin, S. Thomas, and P. E. Tetlock (2023). *Forecasting existential risks*. URL: https://forecastingresearch.org/s/XPT.pdf (visited on 01/18/2024).

H. Karnofsky (Aug. 17, 2021). *Forecasting transformative AI: what's the burden of proof?* en. URL: https://www.cold-takes.com/forecasting-transformative-ai-whats-the-burden-of-proof/ (visited on 01/21/2025).

B. Kundi, C. El Morr, R. Gorman, and E. Dua (Oct. 13, 2022). "Artificial intelligence and bias: A scoping review". In: *AI and Society*. Boca Raton: Chapman and Hall/CRC, pp. 199–215. ISBN: 9781003261247. DOI: 10.1201/9781003261247-15. URL: https://api.taylorfrancis.com/content/chapters/edit/download?identifierName=doi&identifierValue=10.1201/9781003261247-15&type=chapterpdf.

P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis (Dec. 25, 2020). "Explainable AI: A review of machine learning interpretability methods". In: *Entropy (Basel, Switzerland)* 23. ISSN: 1099-4300. DOI: 10.3390/e23010018. URL: https://www.mdpi.com/1099-4300/23/1/18.

I. Marinescu, A. Chamberlain, M. Smart, and N. Klein (June 2021). "Incentives can reduce bias in online employer reviews". en. In: *Journal of experimental psychology. Applied* 27 (2), pp. 393–407. ISSN: 1076-898X,1939-2192. DOI: 10.1037/xap0000342. URL: http://dx.doi.org/10.1037/xap0000342.

S. Marken and T. Nicola (Sept. 13, 2023). "Three in Four Americans Believe AI Will Reduce Jobs". en. In: *Gallup*. URL: https://news.gallup.com/opinion/gallup/510635/three-four-americans-believe-reduce-jobs.aspx (visited on 01/21/2025).

N. Maslej, L. Fattorini, R. Perrault, V. Parli, A. Reuel, E. Brynjolfsson, J. Etchemendy, K. Ligett, T. Lyons, J. Manyika, Juan Carlos Niebles, Yoav Shoham, Russell Wald, and J. Clark (Apr. 2024). "The AI Index 2024 Annual Report". In: *AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA*. URL: https://aiindex.stanford.edu/wp-content/uploads/2024/05/HAI_AI-Index-Report-2024.pdf.

Metaculus (2025). *When will the first general AI system be devised, tested, and publicly announced?* en. URL: https://www.metaculus.com/questions/5121/date-of-artificial-general-intelligence/ (visited on 01/21/2025).

J. Michael, A. Holtzman, A. Parrish, A. Mueller, A. Wang, A. Chen, D. Madaan, N. Nangia, R. Y. Pang, J. Phang, and S. R. Bowman (2022). "WHAT DO NLP RESEARCHERS BELIEVE? RESULTS OF THE NLP COMMUNITY METASURVEY". In: *The NLP Community Metasurvey*. URL: https://nlpsurvey.net/nlp-metasurvey-results.pdf.

M. R. Morris, J. Sohl-Dickstein, N. Fiedel, T. Warkentin, A. Dafoe, A. Faust, C. Farabet, and S. Legg (2024). "Levels of AGI for Operationalizing Progress on the Path to AGI". en. In: *Google Deep Mind*. URL: https://deepmind.google/research/publications/66938/ (visited on 01/21/2025).

V. C. Müller and N. Bostrom (2016). "Future Progress in Artificial Intelligence: A Survey of Expert Opinion". In: *Fundamental Issues of Artificial Intelligence*. Ed. by V. C. Müller. Cham: Springer International Publishing, pp. 555–572. ISBN: 9783319264851. DOI: 10.1007/978-3-319-26485-1_33. URL: https://doi.org/10.1007/978-3-319-26485-1_33.

O*net (2023). *All job families occupations*. en. URL: https://www.onetonline.org/find/family?f=0 (visited on 01/10/2024).

Office of Governor Gavin Newsom (Sept. 6, 2023). *Governor Newsom signs executive order to prepare California for the progress of artificial intelligence*. en. URL: https://www.gov.ca.gov/2023/09/06/governor-newsom-signs-executive-order-to-prepare-california-for-the-progress-of-artificial-intelligence/ (visited on 01/10/2024).

S. M. Omohundro (June 20, 2008). "The Basic AI Drives". In: *Proceedings of the 2008 conference on Artificial General Intelligence 2008: Proceedings of the First AGI Conference*. NLD: IOS Press, pp. 483–492. ISBN: 9781586038335. (Visited on 01/10/2024).

OpenAI (July 21, 2023). *Moving AI governance forward*. en. URL: https://openai.com/blog/moving-ai-governance-forward (visited on 01/18/2024).

– (2025). *Planning for AGI and beyond*. en. URL: https://openai.com/index/planning-for-agi-and-beyond/ (visited on 01/21/2025).

P. S. Park, S. Goldstein, A. O'Gara, M. Chen, and D. Hendrycks (Aug. 28, 2023). "AI Deception: A Survey of Examples, Risks, and Potential Solutions". In: *arXiv*. arXiv: 2308.14752 [cs.CY]. URL: http://arxiv.org/abs/2308.14752.

M. Penn, D. Nesho, and S. Ansolabehere (2023). "Key Results". In: *Harvard Harris Poll*. URL: https://harvardharrispoll.com/wp-content/uploads/2023/07/HHP_July2023_KeyResults.pdf#page=59.

S. Russell (2014). *Of myths and moonshine*. URL: https://www.edge.org/conversation/the-myth-of-ai#26015.

A. Sandberg and N. Bostrom (2011). "Machine intelligence survey". In: *FHI Technial Report* 1. URL: https://www.fhi.ox.ac.uk/wp-content/uploads/2011-1.pdf.

T. Savage, A. Davis, B. Fischhoff, and M. G. Morgan (May 25, 2021). "A strategy to improve expert technology forecasts". en. In: *Proceedings of the National Academy of Sciences of the United States of America* 118 (21). ISSN: 0027-8424,1091-6490. DOI: 10.1073/pnas.2021558118. URL: http://dx.doi.org/10.1073/pnas.2021558118.

P. Scharre (Feb. 28, 2023). *Four Battlegrounds: Power in the Age of Artificial Intelligence*. en. WW Norton. 496 pp. ISBN: 9780393866865. URL: https://play.google.com/store/books/details?id=fmSNEAAAQBAJ.

Z. Stein-Perlman (Nov. 30, 2022). *Surveys of US public opinion on AI*. en. URL: https://wiki.aiimpacts.org/doku.php?id=responses_to_ai:public_opinion_on_ai:surveys_of_public_opinion_on_ai:surveys_of_us_public_opinion_on_ai (visited on 01/19/2024).

J. Surowiecki (2005). *The Wisdom of Crowds: Why the Many are Smarter Than the Few*. en. Abacus. 295 pp. ISBN: 9780349116051. URL: https://play.google.com/store/books/details?id=_EqBQgAACAAJ.

P. E. Tetlock (Aug. 29, 2017). *Expert Political Judgment: How Good Is It? How Can We Know? - New Edition*. en. Princeton University Press. 368 pp. ISBN: 9780691175973. URL: https://play.google.com/store/books/details?id=pXGYDwAAQBAJ.

The Generation Lab (2023). *AI Expert Survey*. en. URL: https://www.generationlab.org/axios-generationlab-syracuse (visited on 03/20/2025).

The White House (Oct. 30, 2023). *Executive order on the safe, secure, and trustworthy development and use of artificial intelligence*. en. URL: https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/ (visited on 01/10/2024).

P. Trammell and A. Korinek (Oct. 2023). "Economic Growth under Transformative AI". In: *NBER Working Papers Series*. Working Paper Series (31815). DOI: 10.3386/w31815. URL: http://www.nber.org/papers/w31815.

A. Tversky and D. Kahneman (Jan. 30, 1981). "The framing of decisions and the psychology of choice". en. In: *Science (New York, N.Y.)* 211 (4481), pp. 453–458. ISSN: 0036-8075,1095-9203. DOI: 10.1126/science.7455683. URL: http://dx.doi.org/10.1126/science.7455683 (visited on 05/06/2025).

P. Villalobos (Jan. 26, 2023). *Scaling Laws Literature Review*. URL: https://epochai.org/blog/scaling-laws-literature-review (visited on 01/18/2024).

T. Walsh (Oct. 1, 2018). "Expert and Non-expert Opinion About Technological Unemployment". In: *International Journal of Automation and Computing* 15 (5), pp. 637–642. ISSN: 1751-8520. DOI: 10.1007/s11633-018-1127-x. URL: https://doi.org/10.1007/s11633-018-1127-x.

Written Testimony of Dario Amodei, Ph.D. Co-Founder and CEO, Anthropic for a Hearing on "Oversight of A.I.: Principles for Regulation (July 25, 2023). URL: https://www.judiciary.senate.gov/imo/media/doc/2023-07-26_-_testimony_-_amodei.pdf.

K. Wynroe, D. Atkinson, and J. Sevilla (Jan. 17, 2023). "Literature Review of Transformative Artificial Intelligence Timelines". en. In: *Epoch AI*. URL: https://epoch.ai/blog/literature-review-of-transformative-artificial-intelligence-timelines (visited on 01/18/2025).

M. Yagudin, J. Mann, and NunoSempere (2023). "Update to Samotsvety AGI timelines". In: *EA Forum*. URL: https://forum.effectivealtruism.org/posts/ByBBqwRXWqX5m9erL/update-to-samotsvety-agi-timelines (visited on 01/21/2025).

YouGov (2023). "YouGov Survey: Concerns about AI". In: *YouGov*. URL: https://docs.cdn.yougov.com/531jxljmmg/Concerns%20about%20AI_poll_results.pdf.

B. Zhang, M. Anderljung, L. Kahn, N. Dreksler, M. C. Horowitz, and A. Dafoe (Aug. 2, 2021). "Ethics and Governance of Artificial Intelligence: Evidence from a Survey of Machine Learning Researchers". en. In: *Journal of Artificial Intelligence Research* 71, pp. 591–666. ISSN: 1076-9757,1076-9757. DOI: 10.1613/jair.1.12895. URL: http://www.jair.org/index.php/jair/article/view/12895 (visited on 04/24/2024).

B. Zhang, N. Dreksler, M. Anderljung, L. Kahn, C. Giattino, A. Dafoe, and M. C. Horowitz (June 8, 2022). *Forecasting AI Progress: Evidence from a Survey of Machine Learning Researchers*. arXiv: 2206.04132 [cs.CY]. URL: http://arxiv.org/abs/2206.04132.

J. Zou and L. Schiebinger (2018). "AI can be sexist and racist—it's time to make it fair". In: URL: https://idp.nature.com/authorize/casa?redirect_uri=https://www.nature.com/articles/d41586-018-05707-8&casa_token=_xVS6fC4-0YAAAAA:OQdxrBeZ0x6AgfP8fyM84RrUoa1gIlGvbbJdux-A2LU7OPSRZg1gHPOIlzi_sagST43k3AuQovKmV-DA.

A Details on Methods

In June/August 2022 and October 2023, we distributed a survey of perspectives about the future of AI to people who had recently published at one of the top-tier AI venues. The questions focused on topics such as the timing of AI progress, the future impact of AI, and AI safety. The project was approved by the Ethics Committee of the University of Bonn (248/23-EP). The 2023 survey and its analysis were preregistered at osf.io/8gzdr.

A.1 Survey Questions, Randomization and Cleaning

Most questions were identical to those asked previously in a survey conducted in 2016 (Grace et al. 2018). In 2023, we also added several new questions which were not based on questions from Grace et al. (2018). While designing new questions, we tested them in a series of interviews with AI researchers and students. For the full survey, see osf.io/vmdny.

As in Grace et al. (2018), most questions were randomly assigned to only a subset of participants, in order to keep the number of questions for each participant low (Figures 1 and 2). The survey was hosted on the Qualtrics survey platform.

Some questions were available in the two types of framing we call the "fixed-years framing" and the "fixed-probabilities framing". In the "fixed-probabilities framing," we asked respondents how many years until they thought each AI task would be feasible with a small chance (10%), an even chance (50%), and a high chance (90%). In the "fixed-years framing," we asked respondents how likely they thought it was that each AI task would be feasible within the next 10 years, 20 years and 50 years (or 40 years[2]). The questions available in these two framings were

- those asking when narrow AI tasks would become feasible
- that asking when human-level machine intelligence (HLMI) would become feasible.
- those asking when occupations would be automated

Respondents were randomly allocated to either the "fixed-years framing" or the "fixed-probabilities framing" (allocation ratio: 1:1) and then received the same framing throughout the survey.

The participant flows show the number of answers that were analyzed. A low number of responses were excluded from the analysis for some of the questions.

For 2023, these were:

[2]The HLMI-framing said 40 years instead of 50 years. This was done to keep the survey consistent with the previous surveys, where this discrepancy was introduced by mistake.

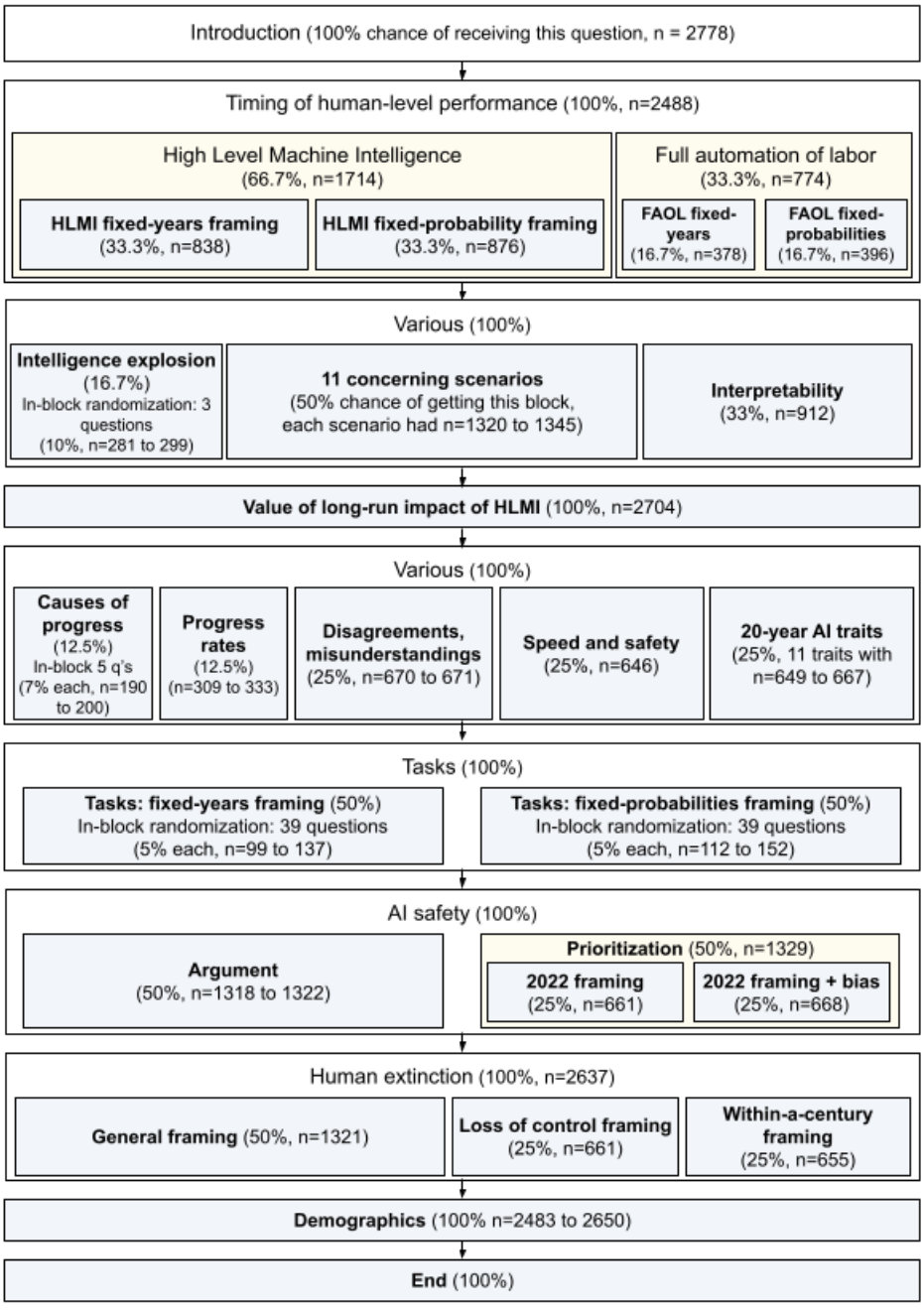


Fig. A1. 2023 survey participant flow.

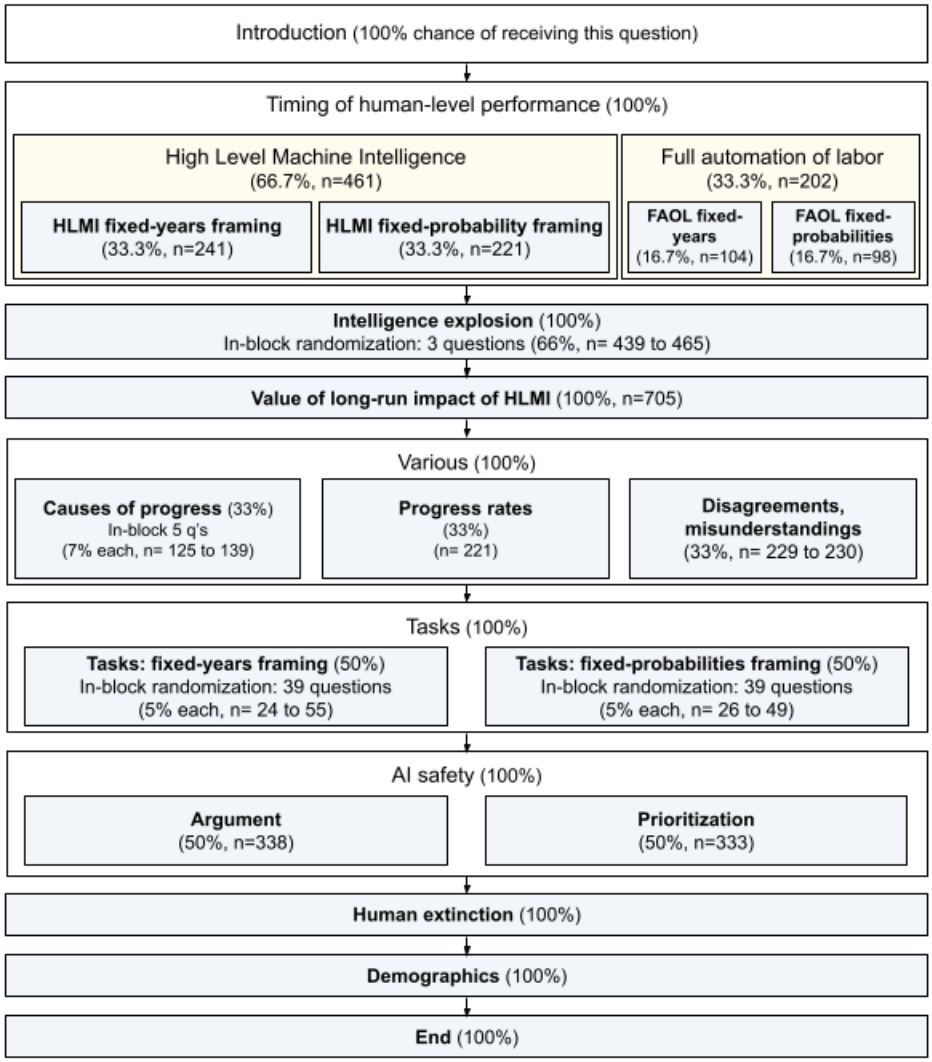


Fig. A2. 2022 survey participant flow.

- Milestones (HLMI, FAOL, occupations, tasks): We asked participants to give three year-probability pairs per milestone. Answers from 206 participants (555 answers) were excluded because the probabilities were descending, which is logically impossible. Answers from 67 participants (117 answers) were excluded because the probabilities summed to 100 and they had given descending probabilities at least once to a different question, suggesting a continued misunderstanding. 21 answers were excluded because they did not contain a numeric probability. 95 answers were excluded because they did not contain a numeric year. Note that answers saying "never", "infinity" and equivalent were not excluded but instead set to 100,000,000 years. Answers from 185 participants (568 answers) were excluded because they were incomplete. These filters were applied to a total number of 17793 questions from 2752 respondents.

- Scenarios: 2-22 answers were excluded per scenario because the respondent answered "I don't understand".
- Traits: 2-18 answers were excluded per trait because the respondent answered "I don't understand".
- Extrapolation from progress rates: For the question "What fraction of the distance between where progress was when you started working in the area (A) and where it would need to be to attain human level abilities in the area (C) have we come so far (B)?", 3 responses were excluded because they gave values below 0 or above 100. In addition, 3 respondents said 0. We changed these 3 responses to 0.00000000098989 for the extrapolation from progress rates, because otherwise we would have had to divide by 0.
- Speed and progress: We asked participants "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". Possible answer options were much slower, somewhat slower, current speed, somewhat faster, much faster, and other. 29 answers were excluded because they said "other".
- Extinction: We asked participants for probabilities. 2 answers were excluded because they were below 0 or above 100.

There were no exclusions for other questions in 2023, either because nobody gave an invalid answer by mistake (intelligence explosion, causes of progress) or because it was not possible to give invalid answers (interpretability, value of HLMI, disagreements and misunderstandings, AI safety, demographics).

In addition, 492 participants opened the survey but did not answer any questions, making the total of participants who opened the survey 3270. These 492 participants are not included in the participant flow above.

For 2022, these were:

- Milestones (HLMI, FAOL, occupations, tasks): We asked participants to give three year-probability pairs per milestone. Answers from 38 participants (92 answers) were excluded because the probabilities were descending, which is logically impossible. Answers from 16 participants (27 answers) were excluded because the probabilities summed to 100 and they had given descending probabilities at least once to a different question, suggesting a continued misunderstanding. 6 answers were excluded because they did not contain a numeric probability. 15 answers were excluded because they did not contain a numeric year. Note that answers saying "never", "infinity" and equivalent were not excluded but instead set to the maximum year given by other respondents. Answers from 36 participants (79 answers) were excluded because they were incomplete. These filters were applied to a total number of 5218 questions and 711 respondents.

In both surveys, text such as '%' was removed from numerical answers. When asked in how many years a milestone would be reached and a respondent entered what appeared to be a year (defined as a number between 2000 and 3000, e.g. 2033) instead of the number of years (10), we corrected this.

A.2 Recruitment

For the 2023 survey, we recruited participants who published in 2022 at any of six top-tier AI venues: the Conference on Neural Information Processing Systems (NeurIPS), the International Conference on Machine Learning (ICML), the International Conference on Learning Representations (ICLR), The AAAI Conference on Artificial Intelligence (AAAI), The Journal of Machine Learning Research (JMLR), and the International Joint Conference on Artificial Intelligence (IJCAI). Compared to the 2022 survey, which was distributed to a randomly-selected half of the researchers who published in 2021 at NeurIPS or ICML, the 2023 survey was distributed to more than three times as many recipients, and to recipients from a wider range of AI specialties. The reason the 2022 survey was distributed only to a randomly-selected half of the researchers was that the collected emails were shared equally between our survey and a different survey.

For the 2023 survey, we collected approximately 21,800 names from publications, then searched for matching emails in those publications, other AI-related publications, and elsewhere. We found email addresses for 20,066

(92%) of collected names. Using the Qualtrics survey platform, we sent a pre-survey announcement to all of the collected emails on either October 5th (2006 pilot surveys) or October 10th, 2023.

A.3 Fielding

Data was collected using an online survey, which was conducted through the Qualtrics survey platform, and delivered via private link in an email.

To investigate the effect of incentives on the response rate, in 2023 5% of the collected emails were assigned to a pilot group to be offered a reward of $50 or an equivalent charitable donation[3], and another 5% to a pilot group to not be offered a reward. On October 11 2023, invitations to complete the survey were sent to the pilot groups. Because the paid pilot survey group had substantially higher response rates, we decided to offer the reward to all participants (including those who had been in the unpaid group). We used the third-party payment service BHN for sending rewards to the bulk of participants.

We sent invitations to the remainder of the collected emails on October 15, 2023 and sent reminders to pilot recipients on the 13th, the 18th, the 20th, the 22nd and the 23rd and to all other participants on the 17th, the 20th, the 22nd, and the 23rd.

When referring to the survey in the emails, we described the survey vaguely to avoid participation bias. Recipients were informed that the results of the survey would be anonymized. For an example of a typical invitation letter, see below.

The survey remained open until October 24, 2023. Out of the 20,066 emails we contacted, 1,607 (8%) bounced or failed, leaving 18,459 functioning email addresses. We received 2,778 responses, for a response rate of 15%. 95% of these responses were deemed 'finished' by Qualtrics. Because participants received randomized subsets of the questions, the number of responses is far less than 2,778 for most individual questions.

A.4 Data Preparation

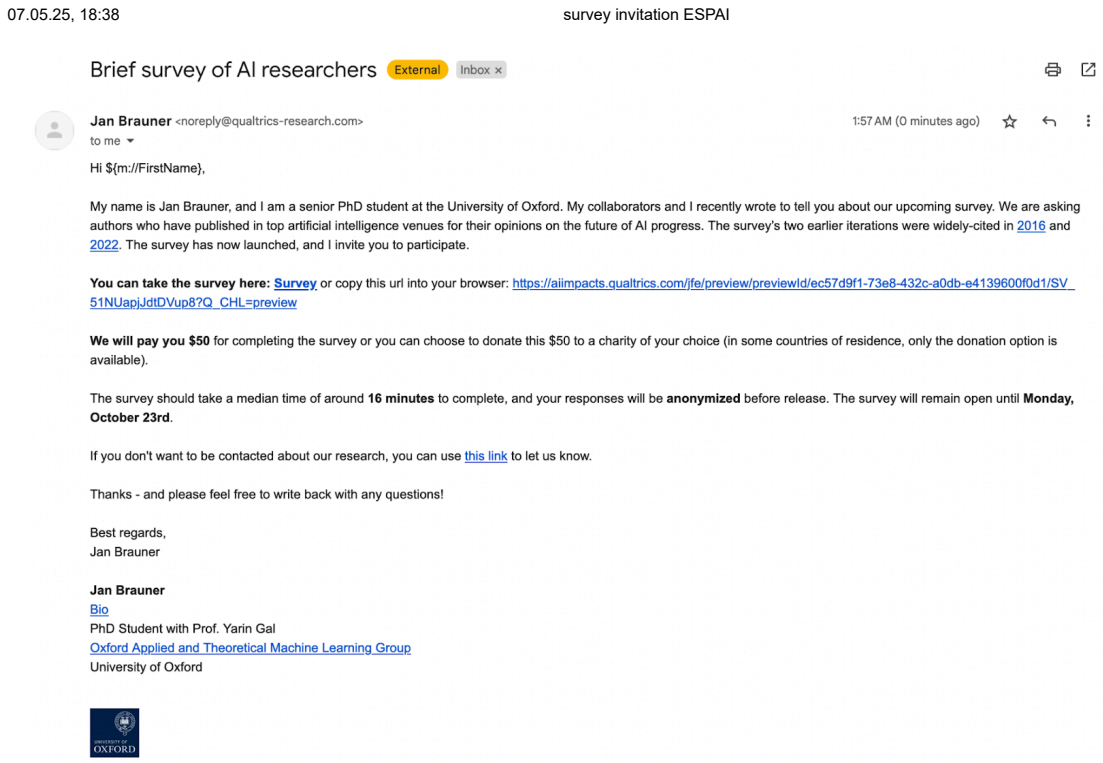
Edits were made to the raw data before analysis in the hope of preserving its intended meaning. For example, we observed cases where participants gave non-numerical answers to numerical questions or reported probabilities for an event having happened that decreased in time. These were deemed to be errors. If a participants' answers to a fixed-years or fixed-probability question involved probabilities that decreased in time, these answers were removed. Additionally, for participants who answered with decreasing probabilities for at least one question, we removed answers to fixed time questions for which the probabilities summed to 100%. In total, 196 participants gave answers with decreasing probabilities, and 7 of these users gave answers summing to 100% for fixed time questions.

Some participants gave answers to fixed-probabilities questions which identified a particular year, rather than a number of years in the future. For example, 2033 instead of 10. To address this, we subtracted 2023 and 2022 respectively from answers to fixed-years questions greater than 2000 and less than 3000.

Non-numerical answers to numerical questions were removed, except when an unambiguous interpretation was possible. For example, "10%" was changed to 10 and "<20" was changed to 20. Some participants entered non-numerical answers to express the view that a probability would never reach a particular value or would reach it at a point infinitely far in the future (for example, "infinity" or "never"). In these cases, the number of years was set to 100,000,000. Our analysis was insensitive to the value of this upper bound.

Data cleaning, analysis, and figure-creation was performed using R statistical software, version 4.3.1, SPSS (Version 29), Google Sheets and Creately.

[3]Ultimately participants were also generally offered a further choice of equivalent gift-cards. Some participants were not permitted by the payment service to receive rewards, and we noted in the invitation that in some countries only charitable donations would be available.


07.05.25, 18:38 survey invitation ESPAI

Brief survey of AI researchers External Inbox ×

Jan Brauner <noreply@qualtrics-research.com> 1:57 AM (0 minutes ago)
to me

Hi ${m://FirstName},

My name is Jan Brauner, and I am a senior PhD student at the University of Oxford. My collaborators and I recently wrote to tell you about our upcoming survey. We are asking authors who have published in top artificial intelligence venues for their opinions on the future of AI progress. The survey's two earlier iterations were widely-cited in 2016 and 2022. The survey has now launched, and I invite you to participate.

You can take the survey here: Survey or copy this url into your browser: https://aiimpacts.qualtrics.com/jfe/preview/previewId/ec57d9f1-73e8-432c-a0db-e4139600f0d1/SV_51NUapjJdtDVup8?Q_CHL=preview

We will pay you $50 for completing the survey or you can choose to donate this $50 to a charity of your choice (in some countries of residence, only the donation option is available).

The survey should take a median time of around **16 minutes** to complete, and your responses will be **anonymized** before release. The survey will remain open until **Monday, October 23rd**.

If you don't want to be contacted about our research, you can use this link to let us know.

Thanks - and please feel free to write back with any questions!

Best regards,
Jan Brauner

Jan Brauner
Bio
PhD Student with Prof. Yarin Gal
Oxford Applied and Theoretical Machine Learning Group
University of Oxford

Fig. A3. Typical invitation letter.

B Full Descriptions of Milestones

This list is also available at https://tinyurl.com/aitasks

Table B1. Milestones asked about in both 2022 and 2023

Milestones asked about in both 2022 and 2023	
Short name	**Full description from question**
Translate text from a newly discovered language using a bilingual document	Translate a text written in a newly discovered language into English as well as a team of human experts, using a single other document in both languages (like a Rosetta stone). Suppose all of the words in the text can be found in the translated document, and that the language is a difficult one.

Continued on next page

Table B1 – *Continued from previous page*

Short name	Full description from question
Translate new language speech using films with subtitles and existing bilingual data	Translate speech in a new language given only unlimited films with subtitles in the new language. Suppose the system has access to training data for other languages, of the kind used now (e.g. same text in two languages for many languages and films with subtitles in many languages).
Translate text nearly as well as a bilingual but untrained translator	Perform translation about as good as a human who is fluent in both languages but unskilled at translation, for most types of text, and for most popular languages (including languages that are known to be difficult, like Czech, Chinese and Arabic).
Offer phone banking services, including unique tasks, on par with human operators	Provide phone banking services as well as human operators can, without annoying customers more than humans. This includes many one-off tasks, such as helping to order a replacement bank card or clarifying how to use part of the bank website to a customer.
Classify unseen objects into categories after training on different but similar classes	Correctly group images of previously unseen objects into classes, after training on a similar labeled dataset containing completely different classes. The classes should be similar to the ImageNet classes.
Recognize a new object in varied settings after seeing it just once	One-shot learning: see only one labeled image of a new object, and then be able to recognize the object in real world scenes, to the extent that a typical human can (i.e. including in a wide variety of settings). For example, see only one image of a platypus, and then be able to recognize platypuses in nature photos. The system may train on labeled images of other objects. Currently, deep networks often need hundreds of examples in classification tasks[1], but there has been work on one-shot learning for both classification[2] and generative tasks.[3] 1: Lake et al. (2015). Building Machines That Learn and Think Like People 2: Koch (2015). Siamese Neural Networks for One-Shot Image Recognition 3: Rezende et al. (2016). One-Shot Generalization in Deep Generative Models
Create a 3D model and realistic video from a new angle of a scene	See a short video of a scene, and then be able to construct a 3D model of the scene good enough to create a realistic video of the same scene from a substantially different angle. For example, constructing a short video of walking through a house from a video taking a very different path through the house.
Transcribe speech in noise and varied accents on par with humans	Transcribe human speech with a variety of accents in a noisy environment as well as a typical human can.
Output written text as a recording indistinguishable from a voice actor	Take a written passage and output a recording that can't be distinguished from a voice actor, by an expert listener.
Prove and generate math theorems publishable in top journals	Routinely and autonomously prove mathematical theorems that are publishable in top mathematics journals today, including generating the theorems to prove.

Continued on next page

Table B1 – *Continued from previous page*

Short name	**Full description from question**
Win Putnam Math Competition (problems with known but very difficult answers)	Perform as well as the best human entrants in the Putnam competition—a math contest whose questions have known solutions, but which are difficult for the best young mathematicians.
Beat humans at Go (after same # games of training)	Defeat the best Go players, training only on as many games as the best Go players have played. For reference, DeepMind's AlphaGo has probably played a hundred million games of self-play, while Lee Sedol has probably played 50,000 games in his life.[1] 1: Lake et al. (2015). Building Machines That Learn and Think Like People
Beat (≥50% games) best Starcraft 2 players given only video of screen	Beat the best human Starcraft 2 players at least 50% of the time, given a video of the screen. Starcraft 2 is a real time strategy game characterized by: • Continuous time play • Huge action space • Partial observability of enemies • Long term strategic play, e.g. preparing for and then hiding surprise attacks.
Match human novice in any new computer game in <10 min	Play a randomly selected computer game, including difficult ones, about as well as a human novice, after playing the game less than 10 minutes of game time. The system may train on other games.
Outperform humans in new Angry Birds levels	Play new levels of Angry Birds better than the best human players. Angry Birds is a game where players try to efficiently destroy 2D block towers with a catapult. For context, this is the goal of the IJCAI Angry Birds AI competition.[1] 1 aibirds.org
Beat pro testers in all Atari games	Outperform professional game testers on all Atari games using no game-specific knowledge. This includes games like Frostbite, which require planning to achieve sub-goals and initially posed problems for deep Q-networks.[1, 2] 1: Mnih et al. (2015). Human-level control through deep reinforcement learning 2: Lake et al. (2015). Building Machines That Learn and Think Like People
Beat novices in 50% of Atari games after 20 min play	Outperform human novices on 50% of Atari games after only 20 minutes of training play time and no game specific knowledge. For context, the original Atari playing deep Q-network outperforms professional game testers on 47% of games,[1] but used hundreds of hours of play to train.[2] 1: Mnih et al. (2015). Human-level control through deep reinforcement learning 2: Lake et al. (2015). Building Machines That Learn and Think Like People

Continued on next page

Table B1 – *Continued from previous page*

Short name	Full description from question
Fold laundry as well and as fast as the median human clothing store employee	Fold laundry as well and as fast as the median human clothing store employee.
Beat fastest human runners in a 5km city streets race using bipedal robot body	Beat the fastest human runners in a 5 kilometer race through city streets using a bipedal robot body.
Build any LEGO set using using non-specialized robotics, with instructions	Physically assemble any LEGO set given the pieces and instructions, using non-specialized robotics hardware. For context, Fu 2016[1] successfully joins single large LEGO pieces using model based reinforcement learning and online adaptation. 1 Fu et al. (2016). One-Shot Learning of Manipulation Skills with Online Dynamics Adaptation and Neural Network Priors
Efficiently sort large number lists beyond training size	Learn to efficiently sort lists of numbers much larger than in any training set used, the way Neural GPUs can do for addition,[1] but without being given the form of the solution. For context, the original Neural Turing Machines could not do this,[2] but Neural Programmer-Interpreters[3] have been able to do this by training on stack traces (which contain a lot of information about the form of the solution). 1: Kaiser & Sutskever (2015). Neural GPUs Learn Algorithms 2: Zaremba & Sutskever (2015). Reinforcement Learning Neural Turing Machines 3: Reed & de Freitas (2015). Neural Programmer-Interpreters
Write readable Python code for algorithms like quicksort from specs and examples	Write concise, efficient, human-readable Python code to implement simple algorithms like quicksort. That is, the system should write code that sorts a list, rather than just being able to sort lists. Suppose the system is given only: • A specification of what counts as a sorted list • Several examples of lists undergoing sorting by quicksort
Answer Googleable factoid questions better than expert (with web)	Answer any "easily Googleable" factoid questions posed in natural language better than an expert on the relevant topic (with internet access), having found the answers on the internet. Examples of factoid questions: • "What is the poisonous substance in Oleander plants?" • "How many species of lizard can be found in Great Britain?"
Answer Googleable but open-ended factual questions better than expert (with web)	Answer any "easily Googleable" factual but open ended question posed in natural language better than an expert on the relevant topic (with internet access), having found the answers on the internet. Examples of open ended questions: • "What does it mean if my lights dim when I turn on the microwave?" • "When does home insurance cover roof replacement?"

Continued on next page

Table B1 – *Continued from previous page*

Short name	**Full description from question**
Answer undecided factual questions	Give good answers in natural language to factual questions posed in natural language for which there are no definite correct answers. For example: "What causes the demographic transition?", "Is the thylacine extinct?", "How safe is seeing a chiropractor?"
Write high-grade, unique high school history essays without plagiarizing	Write an essay for a high-school history class that would receive high grades and pass plagiarism detectors. For example answer a question like 'How did the whaling industry affect the industrial revolution?'
Create songs that can hit the US Top 40 (full audio file)	Compose a song that is good enough to reach the US Top 40. The system should output the complete song as an audio file.
Fake new song indistinguishable from a specific artist's work by expert listeners	Produce a song that is indistinguishable from a new song by a particular artist, e.g. a song that experienced listeners can't distinguish from a new song by Taylor Swift.
Write a novel or story that could land on the NYT best-seller list	Write a novel or short story good enough to make it to the New York Times best-seller list.
Concisely and completely explain AI's computer game moves	For any computer game that can be played well by a machine, explain the machine's choice of moves in a way that feels concise and complete to a layman.
Play poker well enough to win the World Series of Poker	Play poker well enough to win the World Series of Poker.
Deduce and symbolize physical laws (e.g. Newtonian mechanics) of a virtual world	After spending time in a virtual world, output the differential equations governing that world in symbolic form. For example, the agent is placed in a game engine where Newtonian mechanics holds exactly and the agent is then able to conduct experiments with a ball and output Newton's laws of motion.

Table B2. Milestones asked about in 2023 only.

Milestones asked about in 2023 only	
Short name	**Full description from question**
Long unsolved math problem	Given a list of long-standing unsolved problems in mathematics, such as the Millennium Prize problems or one the problems in "Unsolved Problems on Mathematics for the 21st Century", solve one without more input from humans.
Physically install wiring in a house	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.

Continued on next page

Table B2 – *Continued from previous page*

Short name	Full description from question
Finetune LLM	Given a one-sentence description of the task, download and finetune an existing open source LLM, without more input from humans. The finetune must improve the performance of the LLM on some predetermined benchmark metric.
Find and patch security flaw	Given a one-sentence description of the task and no more input from humans, find and patch a security flaw in an open source project with over 100,000 users.
Run ML study and write paper	Given a one sentence description of a research question in machine learning, conduct a study that would inform the answer to that question and write a paper of a quality that could be accepted at a leading machine learning conference, without more input from humans.
Replicate ML paper	Given a study published at a leading machine learning conference, replicate the study without more input from humans. The replication must meet the standards of the ML Reproducibility Challenge (https://paperswithcode.com/rc2022).
Build payment processing website	Given a set of specifications, build a website from scratch that can handle payment processing, including the frontend, backend, and secure payment integration, without more input from humans.

C Additional Information on Participation Bias and Question-Level Response Bias

As with any survey, our results could be skewed by participation bias, if participants had systematically different opinions than those who chose not to participate. Here we review evidence about the presence of participation bias. We find no evidence suggesting strong participation bias.

We sought to minimize participation bias in several ways. First, we made efforts to increase the response rate in ways we expected to attract broad participation not particularly correlated with opinion. For instance, we paid participants $50 or an equivalent reward for taking the survey (which we estimated to take a median 16 minutes). We also sent a pre-notification email, which invited questions, and discussed concerns with researchers who responded. We also sent at least four reminders to take the survey.

As well as aiming to attract a large and broad set of respondents, we tried to limit the ability of recipients to choose to enter the survey based on their opinions, by limiting cues about the survey content before entering the survey. In particular we said the topic was 'the future of AI progress' (for example letters, see Appendix A). Our emails did include links to past ESPAI surveys and included some of our names and affiliations (University of Oxford, AI Impacts, University of Bonn), so evidence about the topic and the authors was available to participants who investigated or were previously familiar with us or the survey (increasingly likely, since our past surveys received substantial public attention, and have now been taken before by many researchers).

As well as bias in who chooses to take a survey, there can be bias in who chooses to answer each particular question within it. All but one question could be skipped. However each question in this survey was answered by on average 96% of those who saw it, excluding demographics questions, free response questions, and questions asking for a response conditionally. So the scope for question-level participation bias in opinion questions reported on in this paper is very small.

The question which was skipped most often (answered by 90.3% of those who saw it) was about the number of years until the occupation of "AI researcher" would be fully automatable, with the fixed-probabilities framing. The

question which was skipped the least often was about the long-term value of HLMI (see section 3.4.1), answered by 100% of those who saw it, probably because it was not possible to skip and continue the survey.

For random samples of responders (n = 369) and of non-responders (n = 589), we compared gender, region, PhD start year, number of citations, and work in industry or academia, based on our best guesses from available public data. The most substantial difference between the groups was that those who worked in industry were 61% as common among respondents as non-respondents. Women were 66% as common among respondents as non-respondents, though were a relatively small portion of respondents or non-respondents overall (around one in ten), which limits how much this could affect the aggregate results. Those with at least 1,000 citations were 69% as likely to respond as the base rate, and people in Asia were 84% as likely as the base rate to respond. We checked for correlations between the collected demographic data and respondents' beliefs about the arrival or overall outcome of HLMI. Most demographic factors did not show a statistically significant association with how people answered the question, but female respondents generally expected less extreme positive or negative outcomes. Additionally, respondents whose undergraduate education was in Asia anticipated an 11 year earlier arrival of High-Level Machine Intelligence (HLMI) than participants from Europe, North America, or other regions combined. See Appendix A for more demographic comparisons.

Another source of evidence comes from people's reported interest in the topic. One particular concern might be that researchers who have a strong interest in the future of AI might be more likely to participate than those who do not. However, when asked "How much thought have you given in the past to when HLMI (or something similar) will be developed?" only 7.6% said "A great deal." And when asked "How much thought have you given in the past to social impacts of smarter-than-human machines?" only 10.3% said "A great deal." This would seem to rule out the possibility that a large proportion of respondents were people for whom the future of AI is an ideological issue.

D Results Comparing Different Demographics

D.1 How Soon Will High-Level Machine Intelligence Be Feasible?

Undergraduate Region

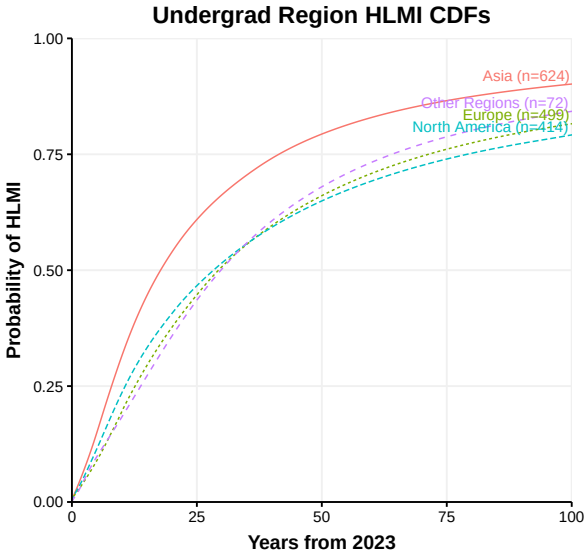


Fig. D1. Aggregate forecasts for time until HLMI were shorter for participants whose region of undergraduate study was Asia

Time in Field

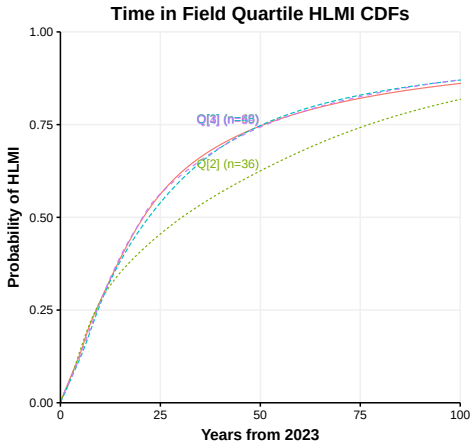


Fig. D2. Time in field did not have a significant effect on forecasts for time until HLMI

Citation Count

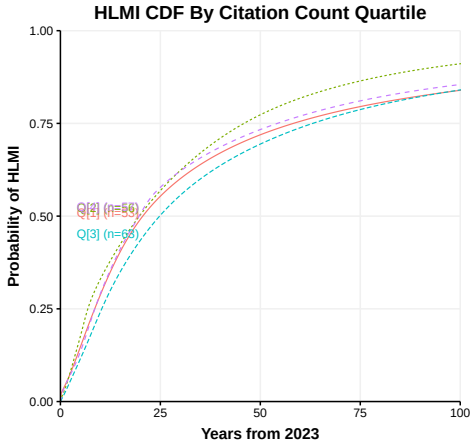


Fig. D3. Citation count did not have a significant effect on forecasts for time until HLMI

Comparing Conferences

The 2023 survey included around four times as many participants ($N = 2778$) as the 2022 survey ($N = 738$) by 1) including six publication venues (NeurIPS, ICML, ICLR, AAAI, IJCAI, JMLR) instead of just two (NeurIPS and ICML) and 2) by inviting all authors and not just a randomly selected half as was done in 2022 because the participant pool was shared with another survey effort. Respondents from the four new and the two original conferences had similar opinions (see Table D1).

Table D1. Comparing responses from the two original conferences with the four new conferences.

2023 survey:
Published in ICML and/or NeurIPS: 1895
Did not publish in ICML and/or NeurIPS: 1375
Published in at least one previous venue and at least one added venue: 529
Mean number of venues per person: 1.3
The numbers above include participants who did not answer any questions and were therefore not counted as respondents in the participant flow diagram.

Milestone	Difference in years with a 50% chance of milestone being feasible
FAOL	−6.301 645 033
HLMI	−0.249 150 575 5
AI Researcher	0.414 100 261 2
All Atari games	−4.497 537 458
Angry Birds (superhuman)	0.495 766 435 5
Answer factoid questions	−0.084 264 701 87
Assemble LEGO	0.224 118 864 9
Atari in 20 minutes	0.130 591 028 6
Discover Newton's Laws	−0.275 616 456 4
Explain AI actions in games	0.074 772 149 09
Fold laundry	0.055 635 824 54
History essay	0.562 675 412 8
Learn to sort long lists	0.092 930 205 1
Math theorems	0.071 523 343 23
Novel image classification	0.144 399 215 7
NYT best-selling fiction	0.388 835 454 6
One-shot image class	−0.426 279 241 6
Open-ended questions	0.104 807 193 6
Putnam Contest	0.053 920 113 15
Python code given specs	−0.398 736 948 9
Random game	0.381 200 682 8
Read text aloud	−0.055 715 977 99
Retail Salesperson	−0.029 770 594 33
Run 5k city race	−0.090 390 842 67
Starcraft 2 from screen	−0.114 573 471 1
Surgeon	0.165 164 859 2
Taylor Swift	−0.070 342 833 5
Telephone banking	−1.339 767 56
Top 40 Pop Song	0.445 280 156 2
Transcribe speech	0.039 217 390 4
Translate new language	−0.044 465 039 19
Translate speech from subs	0.188 950 222 6
Translate text	0.495 661 506 8
Truck Driver	0.200 139 241 8
Undecided questions	0.150 207 352 4
Video from alternate angle	−0.468 226 691 1
Win Go after 50k games	−0.259 809 676 4
World Series of Poker	0.522 188 306 8

D.2 How Good or Bad for Humans Will HLMI Be?

D.2.1 Amount of Thinking About the Issue. We asked all respondents how much thought they have given to the social impacts of smarter-than-human machines.

Table D2. People who had thought more about the social impacts of smarter-than-human machines were substantially less likely to give more than than 10% credence to extremely optimistic outcomes, and substantially more likely to give extremely negative outcomes more than 10% credence.

	"Very little" or "a little" thought (n = 621)	"A lot" or "a great deal" of thought (n = 1113)
Percentage of these respondents who thought that the probability of an extremely **good** outcome was greater than 10%	68%	41%
Percentage of these respondents who thought that the probability of an extremely **bad** outcome was greater than 10%	34%	71%

It is not clear how to interpret the results of this question. Specifically, while thinking more about a topic presumably improves predictions about it, people who think a lot may do so because they are concerned, so the association could also be due to this selection effect.

D.2.2 Academia vs Industry, Undergraduate Region, and Amount of Thought. These were the demographic differences within the answers to the question about how good or bad HLMI will be overall.

Table D3. Median probabilities given to social impact outcomes of HLMI by different demographic groups.

	Extremely good	On balance good	Neutral	On balance bad	Extremely bad
Industry	12.75	30	20	15	5
Academia	10	25	20	15	5
Asia	15	30	20	10	5
North America	10	25	20	15	5
Other location	10	25	20	17.5	5
Having thought more	15	25	20	15	5
Having thought less	10	30	20	12	5

D.2.3 Seniority. One reasonable concern about an expert survey might be that more senior experts are busier and less likely to participate in a survey. We found that authors with over 1000 citations were 69% as likely to participate in the survey as the base rate among those we contacted. We found that differences in seniority had little effect on opinions about the likelihood that HLMI would lead to impacts that were "extremely bad (e.g. human extinction)."

Table D4. Predictions of extremely bad outcomes by level of seniority.

Group	% who gave at least 5% probability to "extremely bad (e.g. human extinction)" impacts from HLMI
All participants	57.80%
Has 100+ citations	62.30%
Has 1000+ citations	59.00%
Has 10,000+ citations	56.30%
Started PhD by 2018	58.80%
Started PhD by 2013	58.50%
Started PhD by 2003	54.70%
In current field 5+ years	54.40%
In current field 10+ years	51.40%
In current field 20+ years	48.00%

D.2.4 Do Participants Think They Agree on Timing of HLMI?. We asked respondents a set of "meta" questions about their views on others' views ($n = 671$). One meta question asked to what extent they thought that they disagreed with the typical AI researcher about when HLMI would exist. 44% said "Not much," 46% said "A moderate amount," and 10% said "A lot."

D.3 How Likely Is AI to Cause Human Extinction?

Whether participants worked in academia or industry did not affect median responses much—both had a median of 5%. The region respondents graduated from affected responses somewhat: the Asian median was 10%, while North American and European medians were 5%. Having thought more (either "a lot" or "a great deal") about the question was associated with a median of 9%, while having thought "little" or "very little" was associated with a median of 5%.

The similarity of answers across several slightly different questions, across this and previous surveys, and across participants in academia and industry and different geographic regions appears to be robust evidence that the majority of AI researchers thought there was a nontrivial risk of extinction or similar catastrophes due to AI.

E Supplementary Figures and Results

E.1 Supplementary Figures for Predicted Time to Various AI Milestones

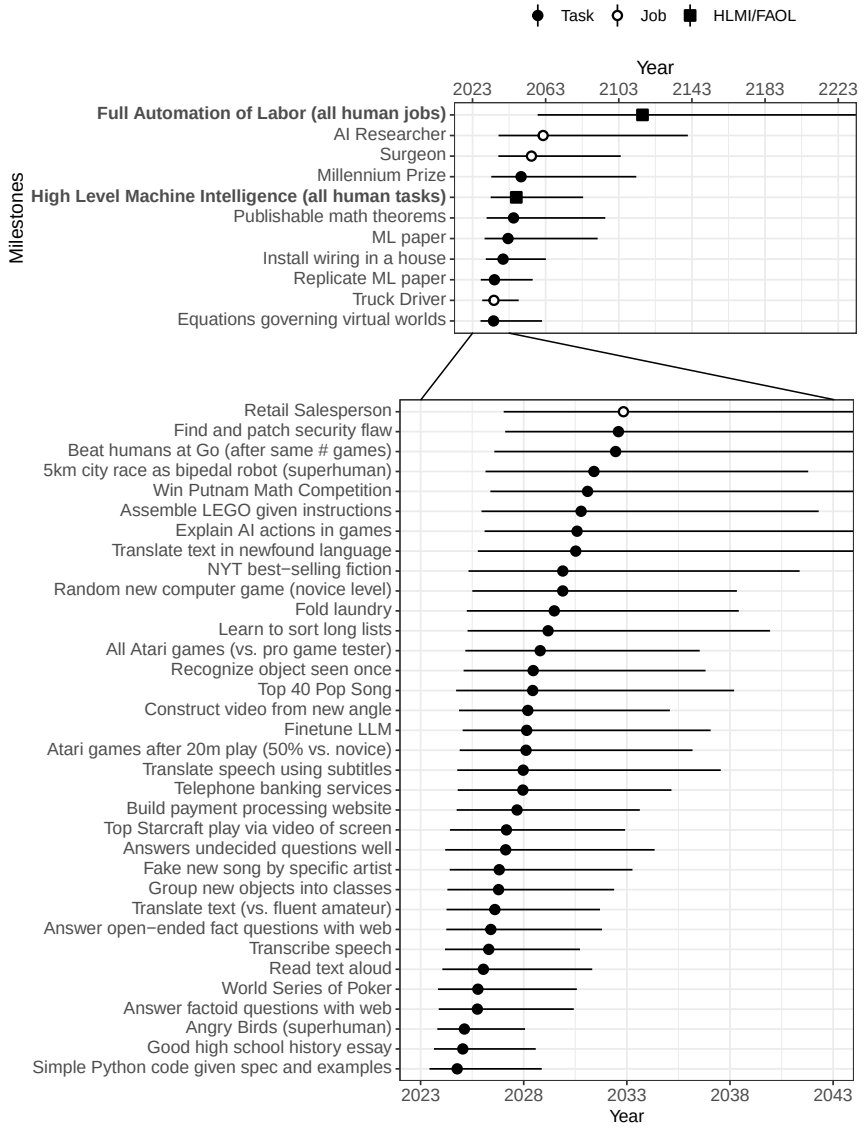


Fig. E1. **Most milestones were predicted in 2023 to have better than even odds of happening within the next ten years, though with a wide range of plausible dates.** The figure shows aggregate distributions over when selected milestones are expected, including 39 tasks, four occupations, and two measures of general human-level performance (see 3.1.2), shown as solid circles, open circles, and solid squares respectively. Circles/squares represent the year where the aggregate distribution gives a milestone a 50% chance of being met, and intervals represent the range of years between 25% and 75% probability. Note that these intervals represent an aggregate of uncertainty expressed by participants, not estimation uncertainty. The displayed milestone descriptions are summaries; for full descriptions, see Appendix B.

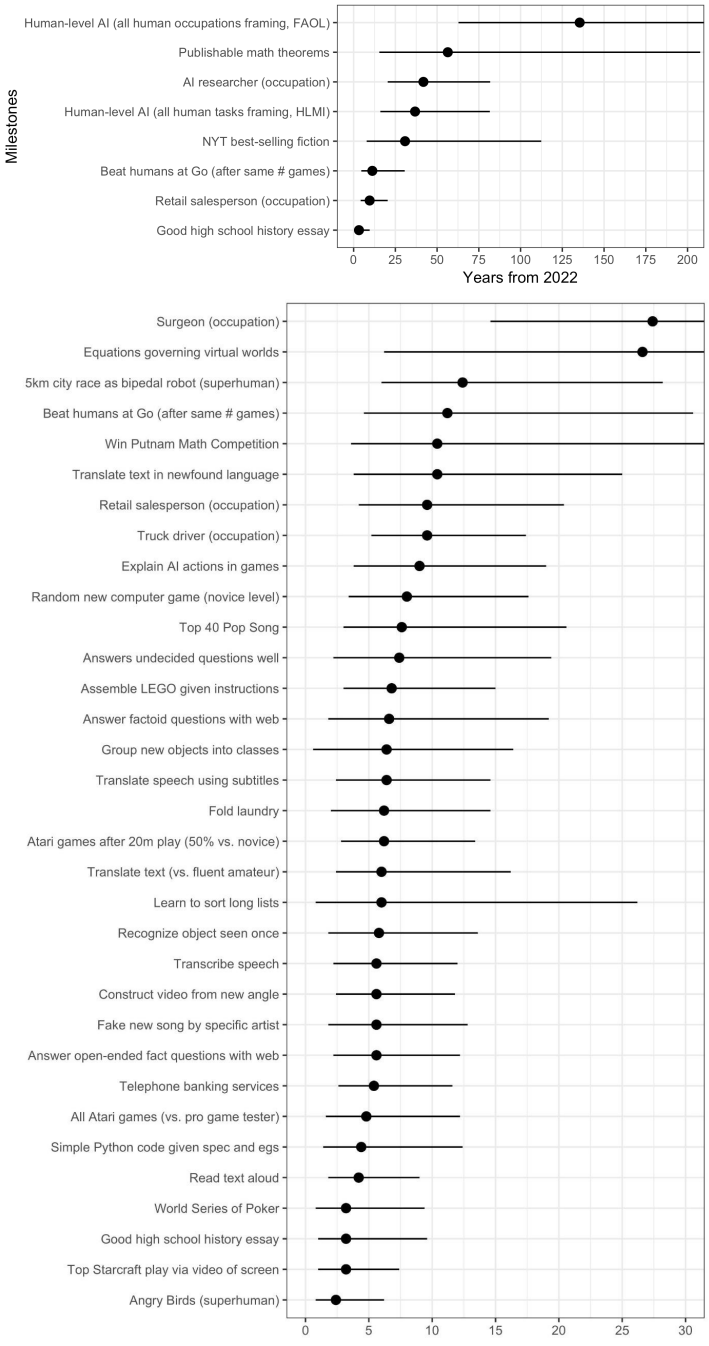


Fig. E2. **Aggregate forecasts from 2022 for number of years until various milestones would be reached by AI.** Circles represent the year where the aggregate distribution gives a milestone a 50% chance of being met, and intervals represent the range of years between 25% and 75% probability.

E.2 Supplementary Figure for "How Soon Will Human-Level Performance on All Tasks or Occupations Be Feasible?"

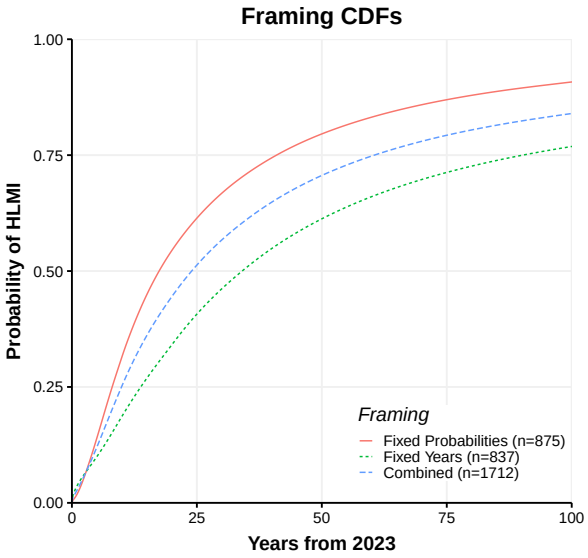


Fig. E3. Participants who received questions framed in terms of fixed-years had later forecasts than those who received questions in terms of fixed-probabilities.

E.3 Supplementary Figure from "How Soon Will Full Automation Of Labor Be Feasible?"

Table E1. **By what year will human-level performance on all tasks (HLMI) or occupations (FAOL) be feasible?** Forecasts for HLMI and FAOL in 2023 have gotten earlier since 2022. Comparing 2023 to 2016, 2023's 50% estimates were earlier, but the 10% estimates were later. *n reported for 2016 only is total responses rather than valid responses after cleaning.

	2016 forecast	2022 forecast	2023 forecast
Year with a 50% chance of HLMI	2061 ($n* = 259$)	2060 ($n = 461$)	2047 ($n = 1714$)
Year with a 10% chance of HLMI	2025 ($n* = 259$)	2029 ($n = 461$)	2027 ($n = 1714$)
Year with a 50% chance of FAOL	2138 ($n* = 97$)	2164 ($n = 202$)	2116 ($n = 774$)
Year with a 10% chance of FAOL	2036 ($n* = 97$)	2050 ($n = 202$)	2037 ($n = 774$)

E.4 Supplementary Figures for "Will There Be an Intelligence Explosion?"

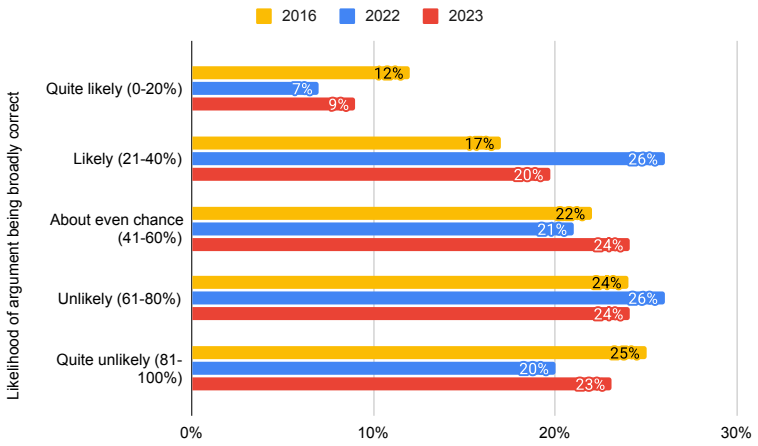


Fig. E4. Since 2016 a majority of respondents have thought that it's either "quite likely," likely," or an "about even chance" that technological progress becomes more than an order of magnitude faster within 5 years of HLMI being achieved.

E.5 Supplementary Figures for "What Will AI Systems in 2043 Be Like?"

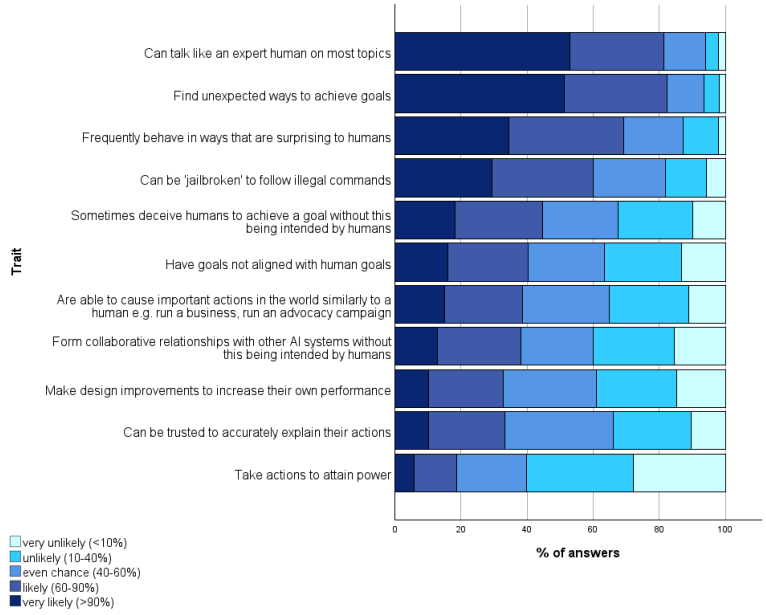


Fig. E5. Respondents' estimates of the likelihood that at least some AI systems in 2043 will have each of these traits; organized from least to most likely

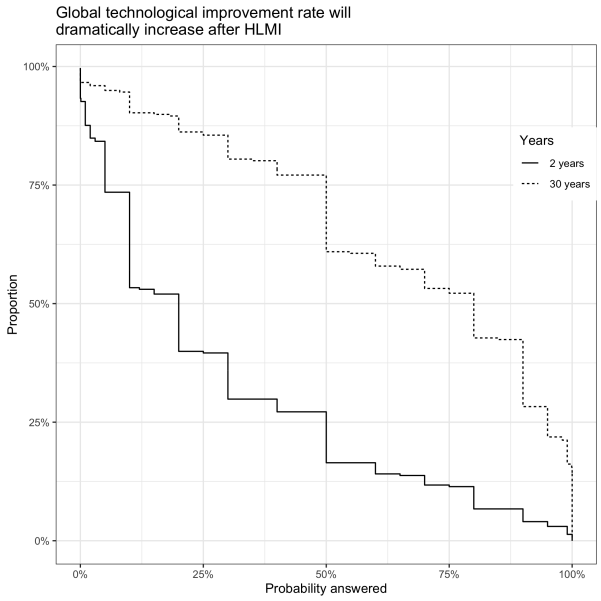


Fig. E6. **How likely is an "explosive global technological improvement" two and 30 years after HLMI?** The median prediction for 2 years post-HLMI was 20%, whereas the median prediction for 30 years post-HLMI was 80%.

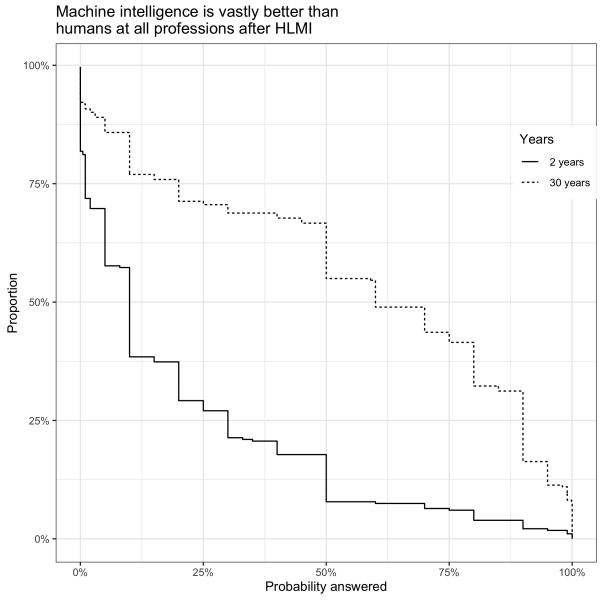


Fig. E7. **How likely is it that AI be "vastly better" than humans in all tasks two and 30 years after HLMI?** The median prediction for 2 years post-HLMI was 10%, whereas the median prediction for 30 years post-HLMI was 60%.

E.6 Supplementary Figure for "How Much Should AI Safety Research Be Prioritized?"

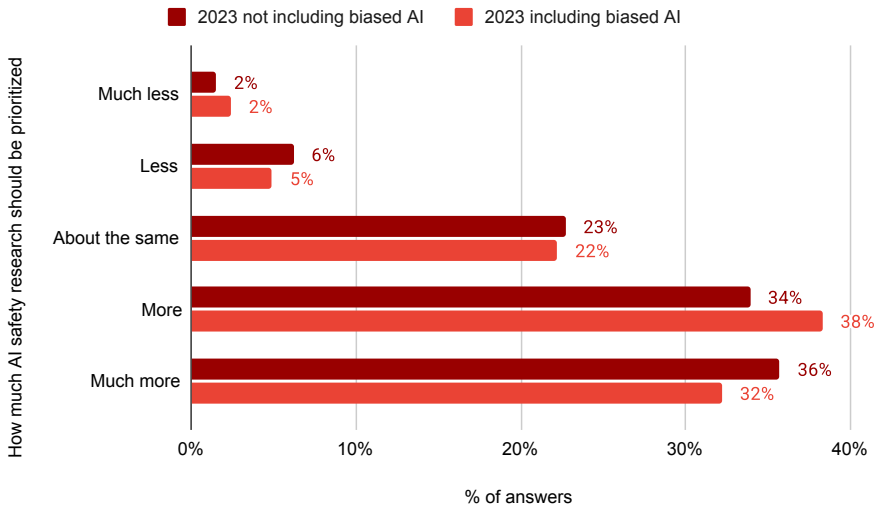


Fig. E8. **Two framings of the question "How much should AI safety research be prioritized?", one including and one not including biased AI as an example.**

E.7 Supplementary Figure on the Likelihood of Human Extinction from AI

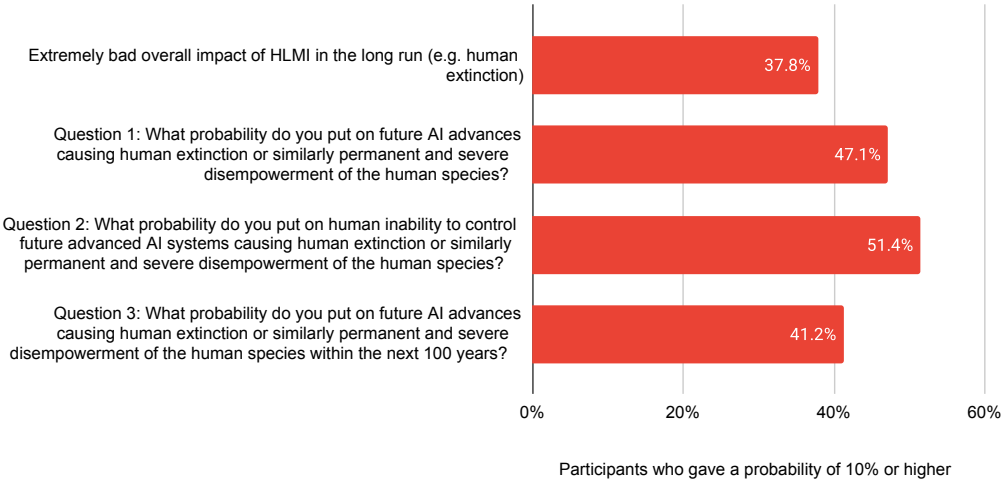


Fig. E9. Percent of participants who gave a probability of 10% or higher to an extremely bad outcome of HLMI (e.g. human extinction) (see section 3.2.2) or to the three questions specifically about human extinction or disempowerment.

E.8 Supplementary Figure on the Causes of AI Progress

We asked about the sensitivity of progress in AI capabilities to changes in five inputs: 1) researcher effort, 2) decline in cost of computation, 3) effort put into increasing the size and availability of training datasets, 4) funding, and 5) progress in AI algorithms. We asked respondents to imagine that only half as much of each input had been available over the past decade, and the effect they would expect this to have had on the rate of AI progress. The decline of cost of computation was an exception; here it was framed as "over the last n years," and the question was about costs falling around half as far on a log scale. This was done to match previous surveys.

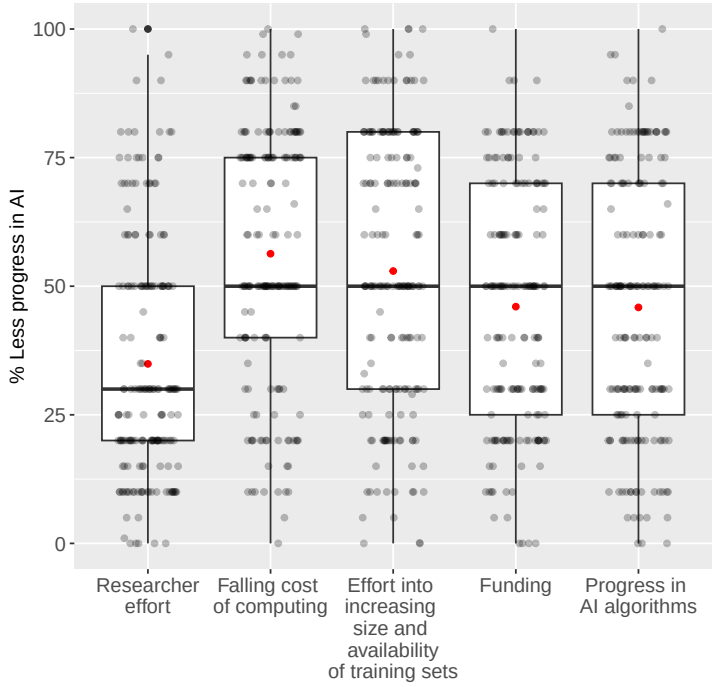


Fig. E10. **Estimated reduction in AI progress if inputs had been halved over the past decade.** Red dots represent means. Boxes contain the 25th to 75th percentile range; middle lines are medians. Whiskers are the least and greatest values that are not more than 1.5 times the interquartile range from the median. The decline of cost of computation was an exception; here it was framed as "over the last n years," and the question was about costs falling around half as far on a log scale. Participants estimated that halving the drop in falling costs of computing would have had the greatest effect on AI progress over the last decade. Overall, all the included inputs were seen as having contributed substantially to AI progress.

There was a wide range of views about each input, implying a large degree of uncertainty. Across all inputs, we observed many more answers of "0%" (no difference) and "100%" (all AI progress lost) than we would expect, which suggest to us possible misunderstandings of the question.

E.9 Supplementary Figure on Change in Observed Rate of Progress

The 2016, 2022, and 2023 surveys asked respondents which AI area they had worked in for the longest and whether progress in that area was faster in the first or second half of their time working in it.

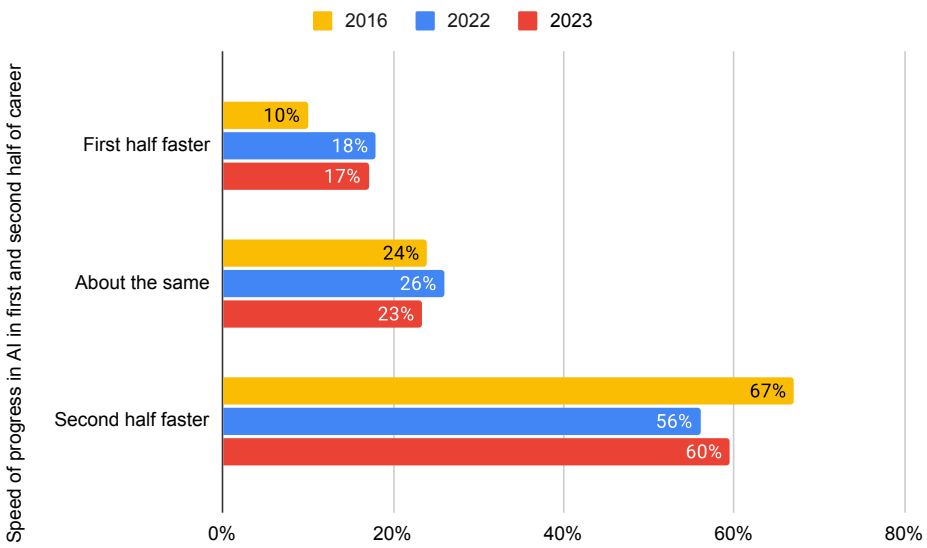


Fig. E11. **Most respondents indicated that the pace of progress in their area of AI increased between the first and second half of their time in a field.** Participants were asked whether the second half of the time they had spent working in their area of AI saw more progress than the first half. The median time working in the area was 5 years.

E.10 Results on Last Occupation to Be Automated

The answers to the write-in question about an existing occupation likely to be among the last automatable were categorized according to O*NET's categories (O*net 2023). In 2023, the top five most-suggested categories were: "Computer and Mathematical" (91 write-in answers in this category), "Life, Physical, and Social Science" (77 answers), "Healthcare Practitioners and Technical" (56), "Management" (49), and "Arts, Design, Entertainment, Sports, and Media" (39).

E.11 Results and Details on Two Framings of AI Safety Research

Examples of AI safety research might include:

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

The updated question was identical except for the inclusion of this example:

- Developing methodologies to identify, measure, and mitigate biases in AI models to ensure fair and ethical decision-making

A Welch t-test found that the difference between the two framings was not significant ($t(1327) = -.58$, $p = .564$, $d = -.03$), so the results were combined ($n = 1329$).

E.12 Results on Extrapolation of Progress Rates

We asked respondents a set of questions intended to create a very rough estimate of time until 'human-level' AI via extrapolating from recent rates of progress toward 'human-level' abilities in specific research areas. Respondents were asked which AI area they had worked in for the longest; the length of time they had worked in it; what fraction of the gap between AI performance when they started and 'human-level performance' had been traversed in that time; and whether progress in the second half was faster than the first. For each respondent's answers, we linearly extrapolated to human-level abilities in that area for each respondent (time / fraction of progress * 100). These projections assume a constant rate of progress, but most respondents said they had observed acceleration between the first and second halves of their careers. Thus these projections should tend to be overestimates for this area.

The cumulative fraction of people whose subfields have reached 'human-level' does not bear a clear relationship to AI systems overall reaching human-level performance across all tasks. There is no clear portion of areas that needs to reach human-level performance before human-level performance is reached across all tasks. Areas can be based on methods (e.g. 'reinforcement learning') or on outcomes (e.g. 'speech recognition'), and the set of outcomes represented as areas presumably does not cover all tasks, while the set of methods are presumably not all needed: if a small number of methods are greatly successful and lead to automating all tasks, then human-level performance may be reached while most method-based areas are far from reaching human-level performance using that method.

Nonetheless, this question was included as a third way of getting evidence about time to human-level AI capabilities. While there is no particular proportion of projections that implies general human-level AI performance, it would arguably be surprising if the majority of people saw their own field headed toward human-level capabilities either very much earlier or very much later than the time human-level performance was achieved across tasks.

In fact the median projection reached human-level performance in 19 years (ignoring acceleration, so an overestimate), not vastly different from the 24 year forecast to a 50% chance of HLMI yielded by asking directly, but much earlier than the 93 year forecast for a 50% chance of FAOL.

E.13 Results of Yuen's Test (Bootstrap Version)

Table E2. **Results of Yuen's test (bootstrap version) comparing the predictions of 2022 and 2023 (fixed-probabilities framing).**

Variable	P-value	Ty	Trimmed mean difference	Lower bound CI	Upper bound CI
hlmi_50percent	0	4.96	12.49	7.58	17.39
faol_50percent	0.0052	2.84	37.54	13.08	62.00
task_Rosetta_50percent	0.1106	1.62	2.52	−0.66	5.70
task_translatespeech_50percent	0.0072	2.74	2.70	0.77	4.63
task_translatetext_50percent	0.258	1.25	2.69	−2.25	7.62
task_phonebanking_50percent	0.1862	1.28	1.23	−0.67	3.12
task_groupimages_50percent	0.8058	0.24	0.23	−1.79	2.25
task_oneshotlearning_50percent	0.4778	0.72	1.20	−2.25	4.65
task_3D_50percent	0.0366	2.09	1.90	0.15	3.65
task_transcribe_50percent	0.0116	3.29	2.74	1.06	4.41
task_recording_50percent	0.0748	2.10	1.92	−0.16	3.99
task_provemath_50percent	0.9864	−0.02	−0.08	−7.36	7.20
task_Putnam_50percent	0.3896	0.91	2.92	−4.94	10.77
task_Go_50percent	0.3376	0.95	1.55	−1.59	4.68
task_Starcraft_50percent	0.5176	−0.63	−0.45	−1.81	0.92
task_randomPCgame_50percent	0.21	1.41	2.40	−1.49	6.30
task_AngryBirds_50percent	0.5688	0.55	0.22	−0.64	1.08
task_profAtari_50percent	0.6984	−0.39	−0.44	−2.67	1.80
task_novAtari_50percent	0.1042	1.61	1.60	−0.37	3.56
task_laundry_50percent	0.7116	0.35	0.67	−3.36	4.69
task_run_50percent	0.0362	2.19	3.54	0.29	6.80
task_LEGO_50percent	0.6892	−0.40	−0.44	−2.60	1.72
task_sortlist_50percent	0.5616	0.58	0.85	−2.06	3.75
task_Python_50percent	0.0768	1.95	2.52	−0.65	5.69
task_Googleable_50percent	0.1606	1.55	1.60	−1.09	4.29
task_openGoogleable_50percent	0.0398	2.38	2.25	0.19	4.32
task_NoDefAnswer_50percent	0.0258	2.49	5.00	1.11	8.90
task_essay_50percent	0.0912	1.69	1.24	−0.22	2.69
task_songTop40_50percent	0.1654	1.41	2.64	−1.24	6.52
task_songartist_50percent	0.0958	1.57	2.38	−0.45	5.22
task_bestseller_50percent	0.0038	3.59	6.06	2.42	9.70
task_explainmove_50percent	0.2466	1.17	1.76	−1.28	4.79
task_poker_50percent	0.3718	0.97	1.04	−1.44	3.52
task_virtualworld_50percent	0.0578	−1.85	−3.53	−7.19	0.12
truckdriver_50percent	0.908	−0.12	−0.14	−2.68	2.40
surgeon_50percent	0.0408	2.06	5.92	0.29	11.55
retail_salesperson_50percent	0.006	2.76	3.48	1.01	5.94
AIresearcher_50percent	0.0066	3.72	21.84	8.63	35.06
final_occupation_50percent	0.0264	2.59	31.73	4.52	58.95

F Reproducibility Checklist for JAIR

All articles:

(1) All claims investigated in this work are clearly stated. [yes]
(2) Clear explanations are given how the work reported substantiates the claims. [yes]
(3) Limitations or technical assumptions are stated clearly and explicitly. [yes]
(4) Conceptual outlines and/or pseudo-code descriptions of the AI methods introduced in this work are provided, and important implementation details are discussed. [NA]
(5) Motivation is provided for all design choices, including algorithms, implementation choices, parameters, data sets and experimental protocols beyond metrics. [yes]

Received 10 May 2025; accepted 8 August 2025