January 2019

# Artificial Intelligence: American Attitudes and Trends

Baobao Zhang & Allan Dafoe

Center for the Governance of AI
Future of Humanity Institute
University of Oxford

Future of Humanity Institute
UNIVERSITY OF OXFORD

UNIVERSITY OF OXFORD

# Artificial Intelligence: American Attitudes and Trends

Baobao Zhang and Allan Dafoe

*January 2019*

## Contents

# Acknowledgements

## Primary researchers

**Baobao Zhang**

Research Affiliate, Center for the Governance of AI, Future of Humanity Institute, University of Oxford

PhD Candidate, Department of Political Science, Yale University

**Allan Dafoe**

Director, Center for the Governance of AI, Future of Humanity Institute, University of Oxford

Associate Professor and Senior Research Fellow in the International Politics of AI, University of Oxford

## Editing and design

For useful feedback we would like to thank: Miles Brundage, Jack Clark, Kanta Dihal, Jeffrey Ding, Carrick Flynn, Ben Garfinkel, Rose Hadshar, Tim Hwang, Katelynn Kyker, Jade Leung, Luke Muehlhauser, Cullen O'Keefe, Michael Page, William Rathje, Carl Shulman, Brian Tse, Remco Zwetsloot, and the YouGov Team (Marissa Shih and Sam Luks). In particular, we are grateful for Markus Anderljung's insightful suggestions and detailed editing.

Copy editor: Steven Van Tassell

Cover design: Laura Pomarius

Web design: Baobao Zhang

Research assistants: Will Marks and Catherine Peng

## Funders

The research was funded by the Ethics and Governance of Artificial Intelligence Fund and Good Ventures.

## For media or other inquiries

Baobao Zhang

+1 (813) 368-9992

Email: surveys@governance.ai

Website: www.governance.ai

We invite suggestions for questions and partnership opportunities for future survey waves.

## Recommended citation

Zhang, Baobao and Allan Dafoe. "Artificial Intelligence: American Attitudes and Trends." Oxford, UK: Center for the Governance of AI, Future of Humanity Institute, University of Oxford, 2019.

# 1 Executive summary

Advances in artificial intelligence (AI)[1] could impact nearly all aspects of society: the labor market, transportation, healthcare, education, and national security. AI's effects may be profoundly positive, but the technology entails risks and disruptions that warrant attention. While technologists and policymakers have begun to discuss AI and applications of machine learning more frequently, public opinion has not shaped much of these conversations. In the U.S., public sentiments have shaped many policy debates, including those about immigration, free trade, international conflicts, and climate change mitigation. As in these other policy domains, we expect the public to become more influential over time. It is thus vital to have a better understanding of how the public thinks about AI and the governance of AI. Such understanding is essential to crafting informed policy and identifying opportunities to educate the public about AI's character, benefits, and risks.

In this report, we present the results from an extensive look at the American public's attitudes toward AI and AI governance. As the study of the public opinion toward AI is relatively new, we aimed for breadth over depth, with our questions touching on: workplace automation; attitudes regarding international cooperation; the public's trust in various actors to develop and regulate AI; views about the importance and likely impact of different AI governance challenges; and historical and cross-national trends in public opinion regarding AI. Our results provide preliminary insights into the character of U.S. public opinion regarding AI. However, our findings raise more questions than they answer; they are more suggestive than conclusive. Accordingly, **we recommend caution** in interpreting the results; we confine ourselves to primarily reporting the results. More work is needed to gain a deeper understanding of public opinion toward AI.

Supported by a grant from the Ethics and Governance of AI Fund, we intend to conduct more extensive and intensive surveys in the coming years, including of residents in Europe, China, and other countries. We welcome collaborators, especially experts on particular policy domains, on future surveys. Survey inquiries can be emailed to surveys@governance.ai.

This report is based on findings from a nationally representative survey conducted by the Center for the Governance of AI, housed at the Future of Humanity Institute, University of Oxford, using the survey firm YouGov. The survey was conducted between June 6 and 14, 2018, with a total of 2,000 American adults (18+) completing the survey. The analysis of this survey was pre-registered on the Open Science Framework. Appendix A provides further details regarding the data collection and analysis process.

## 1.1 Select results

Below we highlight some results from our survey[2]:

- Americans express mixed support for the development of AI. After reading a short explanation, a substantial minority (41%) somewhat support or strongly support the development of AI, while a smaller minority (22%) somewhat or strongly opposes it.
- Demographic characteristics account for substantial variation in support for developing AI. Substantially more support for developing AI is expressed by college graduates (57%) than those with high school or less education (29%); by those with larger reported household incomes, such as those earning over $100,000 annually (59%), than those earning less than $30,000 (33%); by those with computer science or programming experience (58%) than those without (31%); by men (47%) than women (35%). These differences are not easily explained away by other characteristics (they are robust to our multiple regression).
- The overwhelming majority of Americans (82%) believe that robots and/or AI should be carefully managed. This figure is comparable to with survey results from EU respondents.
- Americans consider all of the thirteen AI governance challenges presented in the survey to be important for governments and technology companies to manage carefully. The governance challenges perceived to be the most likely to impact people around the world within the next decade *and* rated the highest in issue importance were[3]:
  1. Preventing AI-assisted surveillance from violating privacy and civil liberties

[1] We define AI as machine systems capable of sophisticated (intelligent) information processing. For other definitions, see Footnote 2 in Dafoe (2018).

[2] These results are presented roughly in the order in which questions were presented to respondents.

[3] Giving equal weight to the likelihood and the rated importance of the challenge.

2. Preventing AI from being used to spread fake and harmful content online
3. Preventing AI cyber attacks against governments, companies, organizations, and individuals
4. Protecting data privacy

- We also asked the above question, but focused on the likelihood of the governance challenge impacting solely Americans (rather than people around the world). Americans perceive that all of the governance challenges presented, except for protecting data privacy and ensuring that autonomous vehicles are safe, are slightly more likely to impact people around the world than to impact Americans within the next 10 years.

- Americans have discernibly different levels of trust in various organizations to develop and manage[4] AI for the best interests of the public. Broadly, the public puts the most trust in university researchers (50% reporting "a fair amount of confidence" or "a great deal of confidence") and the U.S. military (49%); followed by scientific organizations, the Partnership on AI, technology companies (excluding Facebook), and intelligence organizations; followed by U.S. federal or state governments, and the UN; followed by Facebook.

- Americans express mixed support (1) for the U.S. investing more in AI military capabilities *and* (2) for cooperating with China to avoid the dangers of an AI arms race. Providing respondents with information about the risks of a U.S.-China AI arms race slightly decreases support for the U.S. investing more in AI military capabilities. Providing a pro-nationalist message or a message about AI's threat to humanity failed to affect Americans' policy preferences.

- The median respondent predicts that there is a 54% chance that high-level machine intelligence will be developed by 2028. We define *high-level machine intelligence* as when machines are able to perform almost all tasks that are economically relevant today better than the median human (today) at each task. See Appendix B for a detailed definition.

- Americans express weak support for developing high-level machine intelligence: 31% of Americans support while 27% oppose its development.

- Demographic characteristics account for substantial variation in support for developing high-level machine intelligence. There is substantially more support for developing high-level machine intelligence by those with larger reported household incomes, such as those earning over $100,000 annually (47%) than those earning less than $30,000 (24%); by those with computer science or programming experience (45%) than those without (23%); by men (39%) than women (25%). These differences are not easily explained away by other characteristics (they are robust to our multiple regression).

- There are more Americans who think that high-level machine intelligence will be harmful than those who think it will be beneficial to humanity. While 22% think that the technology will be "on balance bad," 12% think that it would be "extremely bad," leading to possible human extinction. Still, 21% think it will be "on balance good," and 5% think it will be "extremely good."

## 1.2 Reading notes

- In all tables and charts, results are weighted to be representative of the U.S. adult population, unless otherwise specified. We use the weights provided by YouGov.

- Wherever possible, we report the margins of error (MOEs), confidence regions, and error bars at the 95% confidence level.

- For tabulation purposes, percentage points are rounded off to the nearest whole number in the figures. As a result, the percentages in a given figure may total slightly higher or lower than 100%. Summary statistics that include two decimal places are reported in Appendix B.

[4] Our survey asked separately about trust in 1) building and 2) managing the development and use of AI. Results are similar and are combined here.

# 2 General attitudes toward AI

## 2.1 More Americans support than oppose developing AI

We measured respondents' support for the further development of AI after providing them with basic information about the technology. Respondents were given the following definition of AI:

> Artificial Intelligence (AI) refers to computer systems that perform tasks or make decisions that usually require human intelligence. AI can perform these tasks or make these decisions without explicit human instructions. Today, AI has been used in the following applications: [five randomly selected applications]

Each respondent viewed five applications randomly selected from a list of 14 that included translation, image classification, and disease diagnosis. Afterward, respondents were asked how much they support or oppose the development of AI. (See Appendix B for the list of the 14 applications and the survey question.)

Americans express mixed support for the development of AI, although more support than oppose the development of AI, as shown in Figure 2.1. A substantial minority (41%) somewhat or strongly supports the development of AI. A smaller minority (22%) somewhat or strongly oppose its development. Many express a neutral attitude: 28% of respondents state that they neither support nor oppose while 10% indicate they do not know.

Our survey results reflect the cautious optimism that Americans express in other polls. In a recent survey, 51% of Americans indicated that they support continuing AI research while 31% opposed it (Morning Consult 2017). Furthermore, 77% of Americans expressed that AI would have a "very positive" or "mostly positive" impact on how people work and live in the next 10 years, while 23% thought that AI's impact would be "very negative" or "mostly negative" (Northeastern University and Gallup 2018).

## 2.2 Support for developing AI is greater among those who are wealthy, educated, male, or have experience with technology

We examined support for developing AI by 11 demographic subgroup variables, including age, gender, race, and education. (See Appendix A for descriptions of the demographic subgroups.) We performed a multiple linear regression to predict support for developing AI using all these demographic variables.

Support for developing AI varies greatly between demographic subgroups, with gender, education, income, and experience being key predictors. As seen in Figure 2.2, a majority of respondents in each of the following four subgroups express support for developing AI: those with four-year college degrees (57%), those with an annual household income above $100,000 (59%), those who have completed a computer science or engineering degree (56%), and those with computer science or programming experience (58%). In contrast, women (35%), those with a high school degree or less (29%), and those with an annual household income below $30,000 (33%), are much less enthusiastic about developing AI. One possible explanation for these results is that subgroups that are more vulnerable to workplace automation express less enthusiasm for developing AI. Within developed countries, women, those with low levels of education, and low-income workers have jobs that are at higher risk of automation, according to an analysis by the Organisation for Economic Co-operation and Development (Nedelkoska and Quintini 2018).

We used a multiple regression that includes all of the demographic variables to predict support for developing AI. The support for developing AI outcome variable was standardized, such that it has mean 0 and unit variance.

Significant predictors of *support* for developing AI include:

- Being a Millennial/post-Millennial (versus being a Gen Xer or Baby Boomer)
- Being a male (versus being a female)
- Having graduated from a four-year college (versus having a high school degree or less)
- Identifying as a Democrat (versus identifying as a Republican)
- Having a family income of more than $100,000 annually (versus having a family income of less than $30,000 annually)
- Not having a religious affiliation (versus identifying as a Christian)

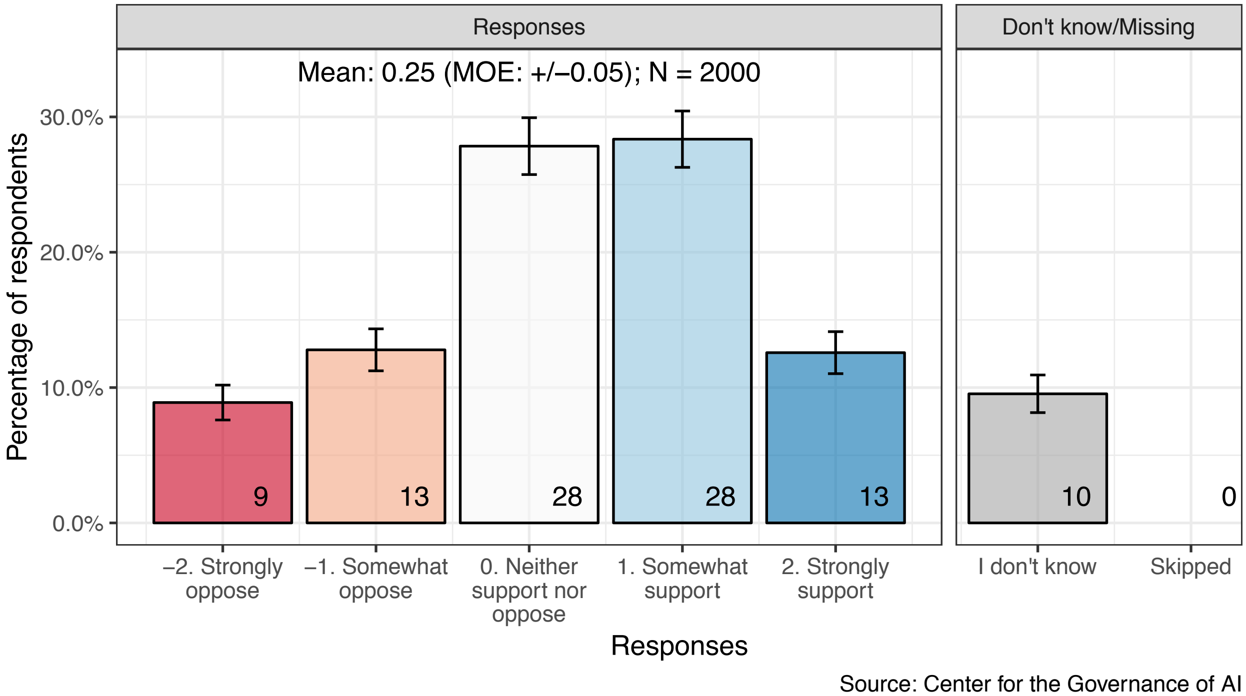


Figure 2.1: Support for developing AI

- Having CS or programming experience (versus not having such experience)

Some of the demographic differences we observe in this survey are in line with existing public opinion research. Below we highlight three salient predictors of support for AI based on the existing literature: gender, education, and income.

Around the world, women have viewed AI more negatively than men. Fifty-four percent of women in EU countries viewed AI positively, compared with 67% of men (Eurobarometer 2017). Likewise in the U.S., 44% of women perceived AI as unsafe – compared with 30% of men (Morning Consult 2017). This gender difference could be explained by the fact that women have expressed higher distrust of technology than men do. In the U.S., women, compared with men, were more likely to view genetically modified foods or foods treated with pesticides as unsafe to eat, to oppose building more nuclear power plants, and to oppose fracking (Funk and Rainie 2015).

One's level of education also predicts one's enthusiasm toward AI, according to existing research. Reflecting upon their own jobs, 32% of Americans with no college education thought that technology had increased their opportunities to advance – compared with 53% of Americans with a college degree (Smith and Anderson 2016). Reflecting on the economy at large, 38% of those with post-graduate education felt that automation had helped American workers while only 19% of those with less than a college degree thought so (Graham 2018). A similar trend holds in the EU: those with more years of education, relative to those with fewer years, were more likely to value AI as good for society and less likely to think that AI steals people's jobs (Eurobarometer 2017).

Another significant demographic divide in attitudes toward AI is income: low-income respondents, compared with high-income respondents, view AI more negatively. For instance, 40% of EU residents who had difficulty paying their bills "most of the time" hold negative views toward robots and AI, compared with 27% of those who "almost never" or "never" had difficulty paying their bills (Eurobarometer 2017). In the U.S., 19% of those who made less than $50,000 annually think that they are likely to lose their job to automation – compared with only 8% of Americans who made more than $100,000 annually (Graham 2018). Furthermore, Americans' belief that AI will help the economy, as well as their support for AI research is positively correlated with their income (Morning Consult 2017).

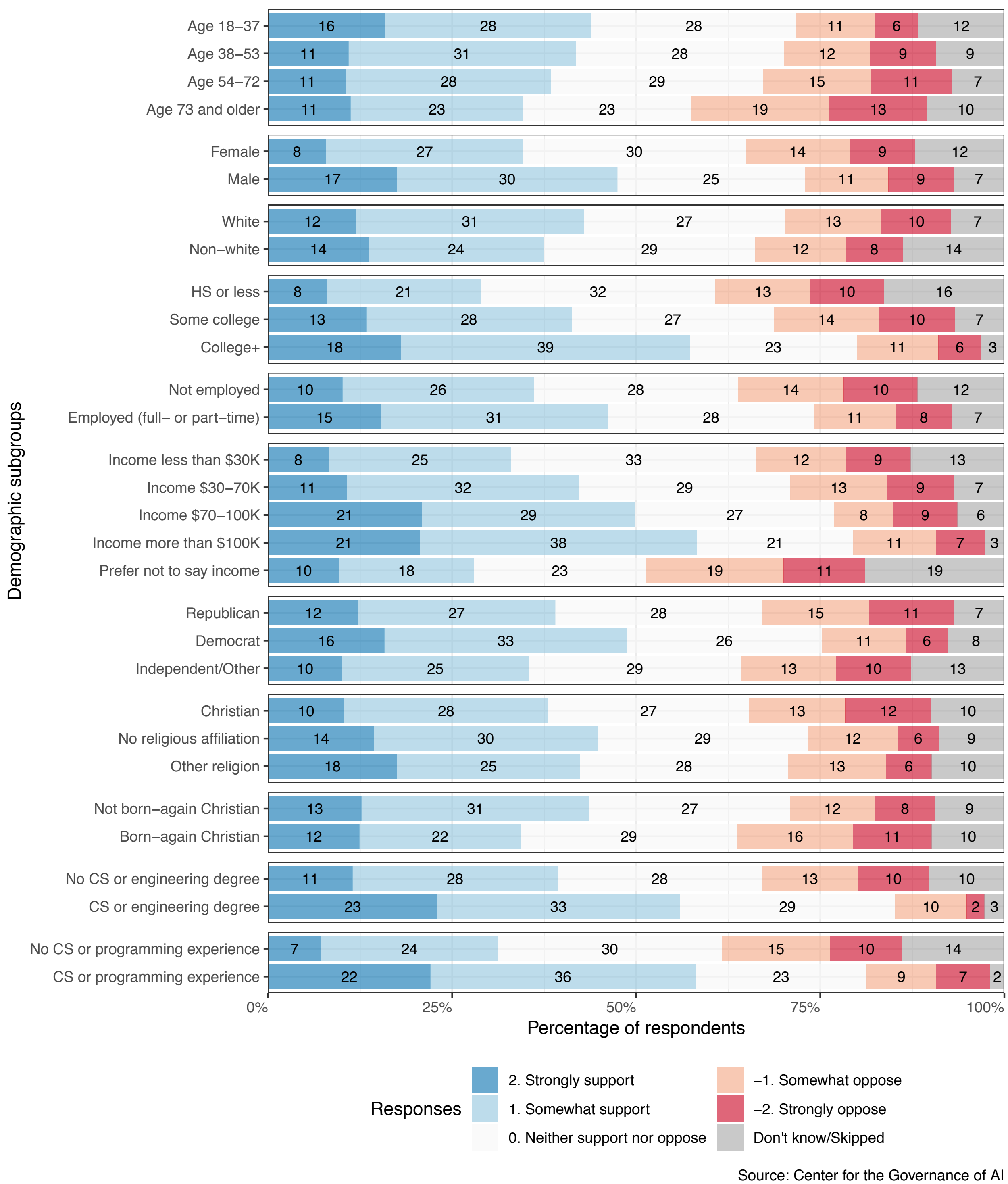


Figure 2.2: Support for developing AI across demographic characteristics: distribution of responses

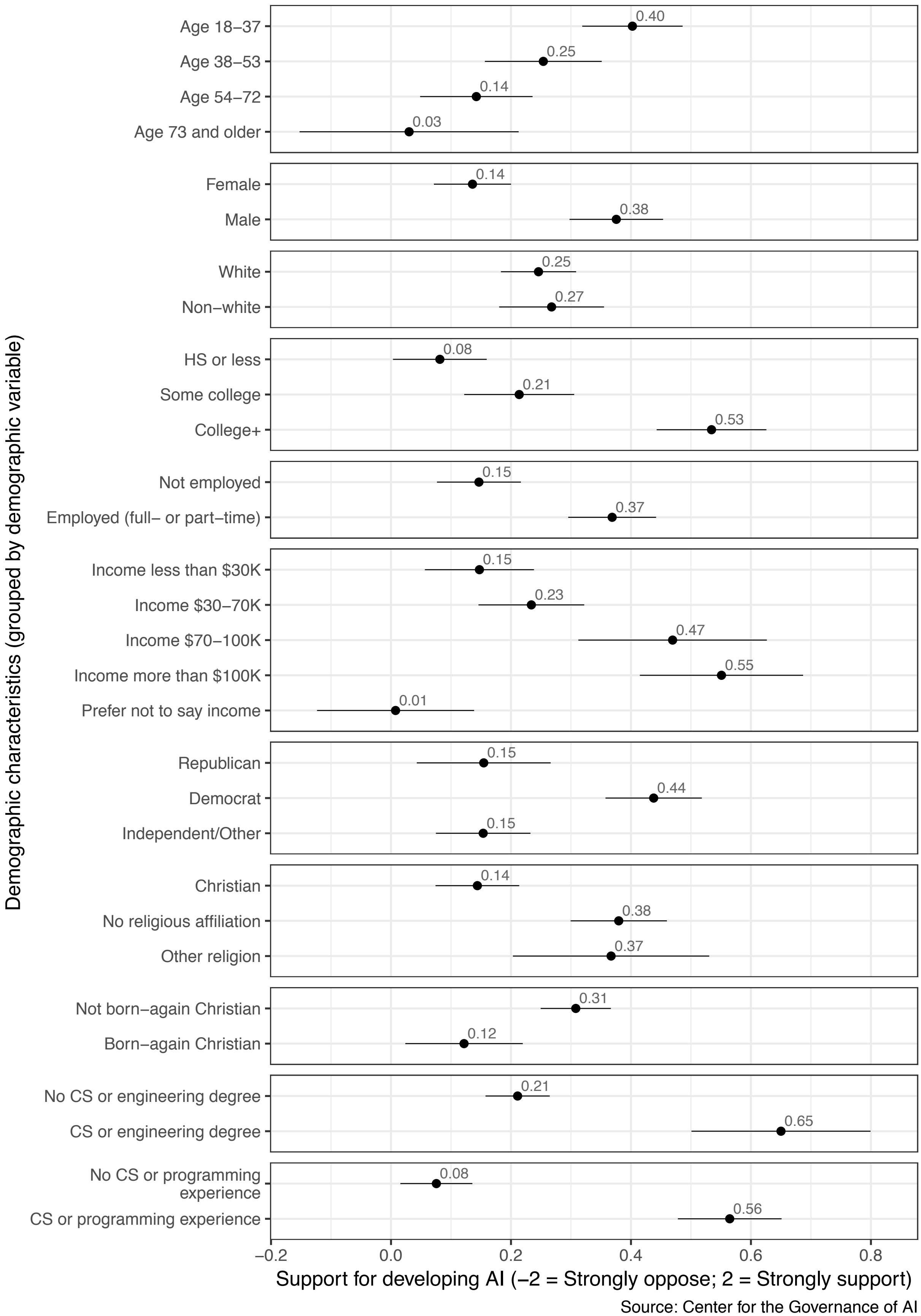


Figure 2.3: Support for developing AI across demographic characteristics: average support across groups

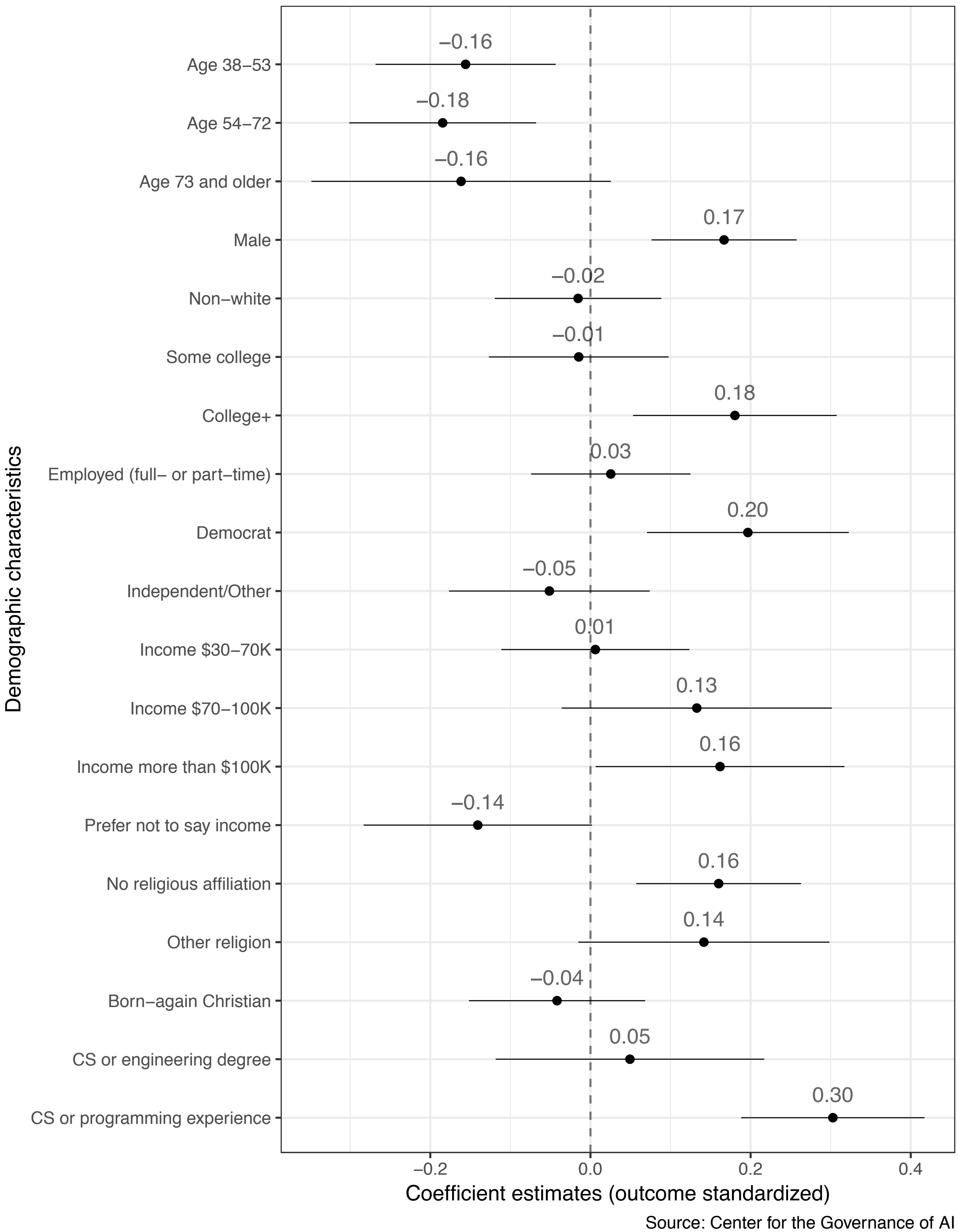


Figure 2.4: Predicting support for developing AI using demographic characteristics: results from a multiple linear regression that includes all demographic variables

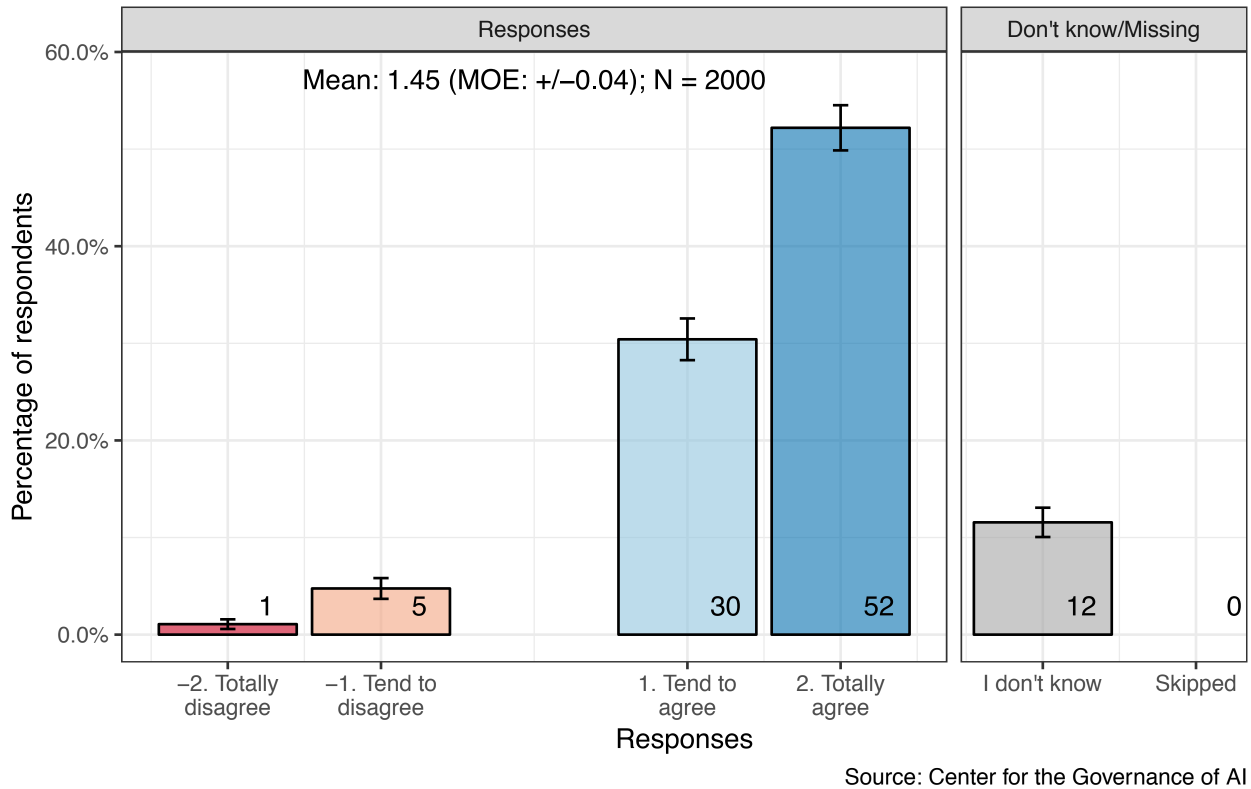


Figure 2.5: Agreement with statement that AI and/or robots should be carefully managed

## 2.3 An overwhelming majority of Americans think that AI and robots should be carefully managed

To compare Americans' attitudes with those of EU residents, we performed a survey experiment that replicated a question from the 2017 Special Eurobarometer #460. (Details of the survey experiment are found in Appendix B.) The original question asked respondents to what extent they agree or disagree with the following statement:

> Robots and artificial intelligence are technologies that require careful management.

We asked a similar question except respondents were randomly assigned to consider one of these three statements:

- AI and robots are technologies that require careful management.
- AI is a technology that requires careful management.
- Robots are technologies that require careful management.

Our respondents were given the same answer choices presented to the Eurobarometer subjects.

The overwhelming majority of Americans – more than eight in 10 – agree that AI and/or robots should be carefully managed, while only 6% disagree, as seen in Figure 2.5.[5] We find that variations in the statement wording produce minor differences, statistically indistinguishable from zero, in responses.

Next, we compared our survey results with the responses from the 2017 Special Eurobarometer #460 by country (Eurobarometer 2017). For the U.S., we used all the responses to our survey question, unconditional on the experimental condition, because the variations in question-wording do not affect responses.

[5]These percentages that we discuss here reflect the average response across the three statements. See Appendix B for the topline result for each statement.

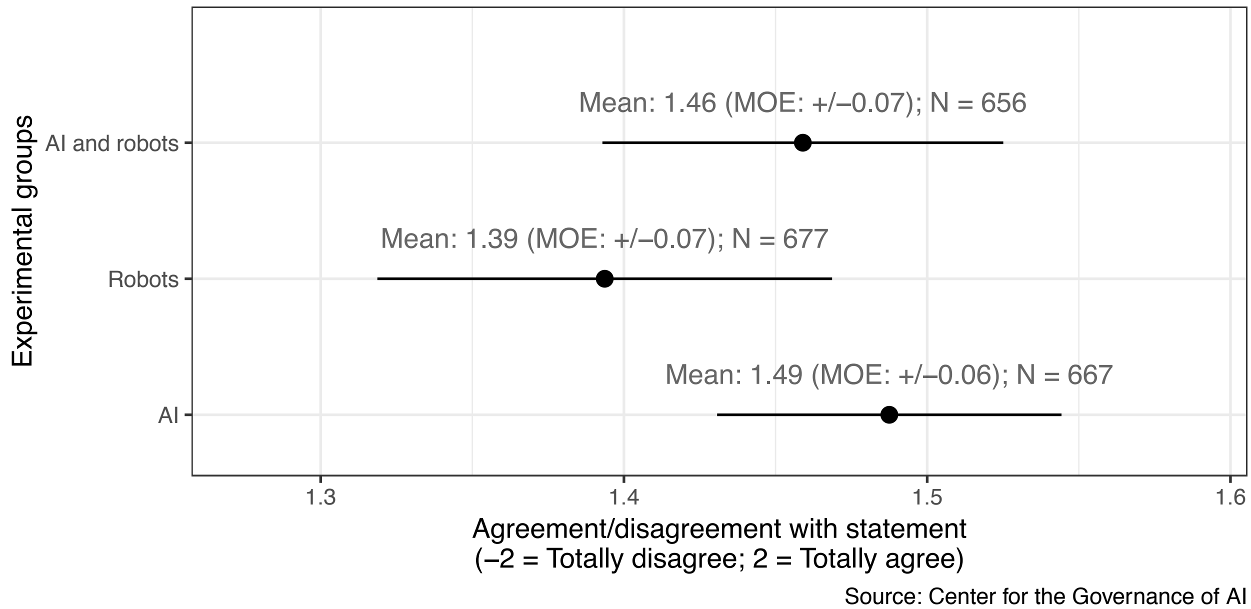


Figure 2.6: Agreement with statement that AI and/or robots should be carefully managed by experimental condition

The percentage of those in the U.S. who agree with the statement (82%) is not far off from the EU average (88% agreed with the statement). Likewise, the percentage of Americans who disagree with the statement (6% disagree) is comparable with the EU average (7% disagreed). The U.S. ranks among the lowest regarding the agreement with the statement in part due to the relatively high percentage of respondents who selected the "don't know" option.

## 2.4 Harmful consequences of AI in the context of other global risks

At the beginning of the survey, respondents were asked to consider five out of 15 potential global risks (the descriptions are found in Appendix B). The purpose of this task was to compare respondents' perception of AI as a global risk with their notions of other potential global risks. The global risks were selected from the Global Risks Report 2018, published by the World Economic Forum. We edited the description of each risk to be more comprehensible to non-expert respondents while preserving the substantive content. We gave the following definition for a global risk:

> A "global risk" is an uncertain event or condition that, if it happens, could cause significant negative impact for at least 10 percent of the world's population. That is, at least 1 in 10 people around the world could experience a significant negative impact.[6]

After considering each potential global risk, respondents were asked to evaluate the likelihood of it happening globally within 10 years, as well as its impact on several countries or industries.

We use a scatterplot (Figure 2.8 to visualize results from respondents' evaluations of global risks. The *x*-axis is the perceived likelihood of the risk happening globally within 10 years. The *y*-axis is the perceived impact of the risk. The mean perceived likelihood and impact is represented by a dot. The corresponding ellipse contains the 95% confidence region.

In general, Americans perceive all these risks to be impactful: on average they rate each as having between a moderate (2) and severe (3) negative impact if they were to occur. Americans perceive the use of weapons of mass destruction to be the most impactful – at the "severe" level (mean score 3.0 out of 4). Although they do not think this risk as likely as other risks, they still assign it an average of 49% probability of occurring within 10 years. Risks in the upper-right quadrant are

[6]Our definition of global risk borrowed from the Global Challenges Foundation's definition: "an uncertain event or condition that, if it happens, can cause a significant negative impact on at least 10% of the world's population within the next 10 years" (Cotton-Barratt et al. 2016).

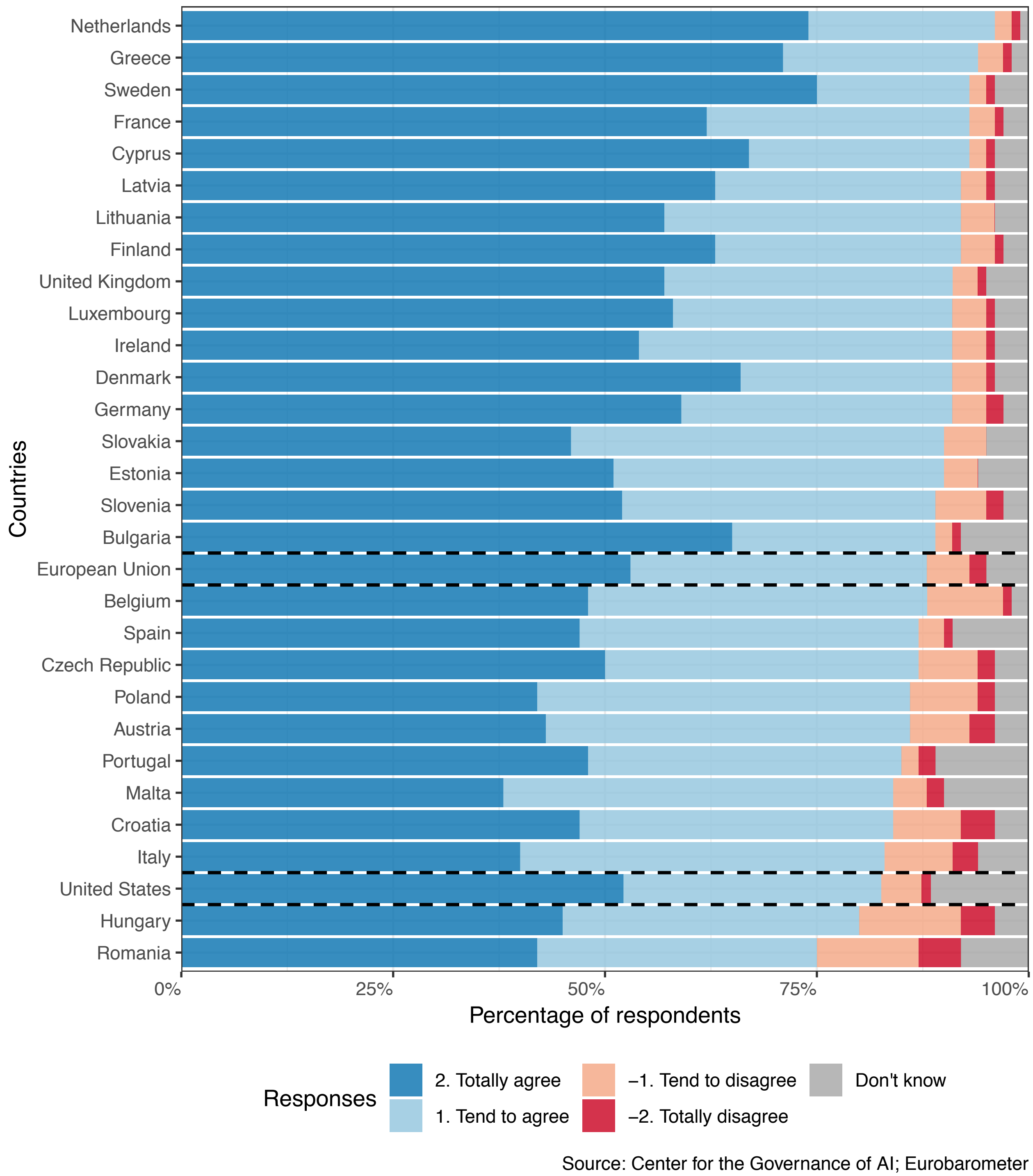


Figure 2.7: Agreement with statement that robots and AI require careful management (EU data from 2017 Special Eurobarometer #460)

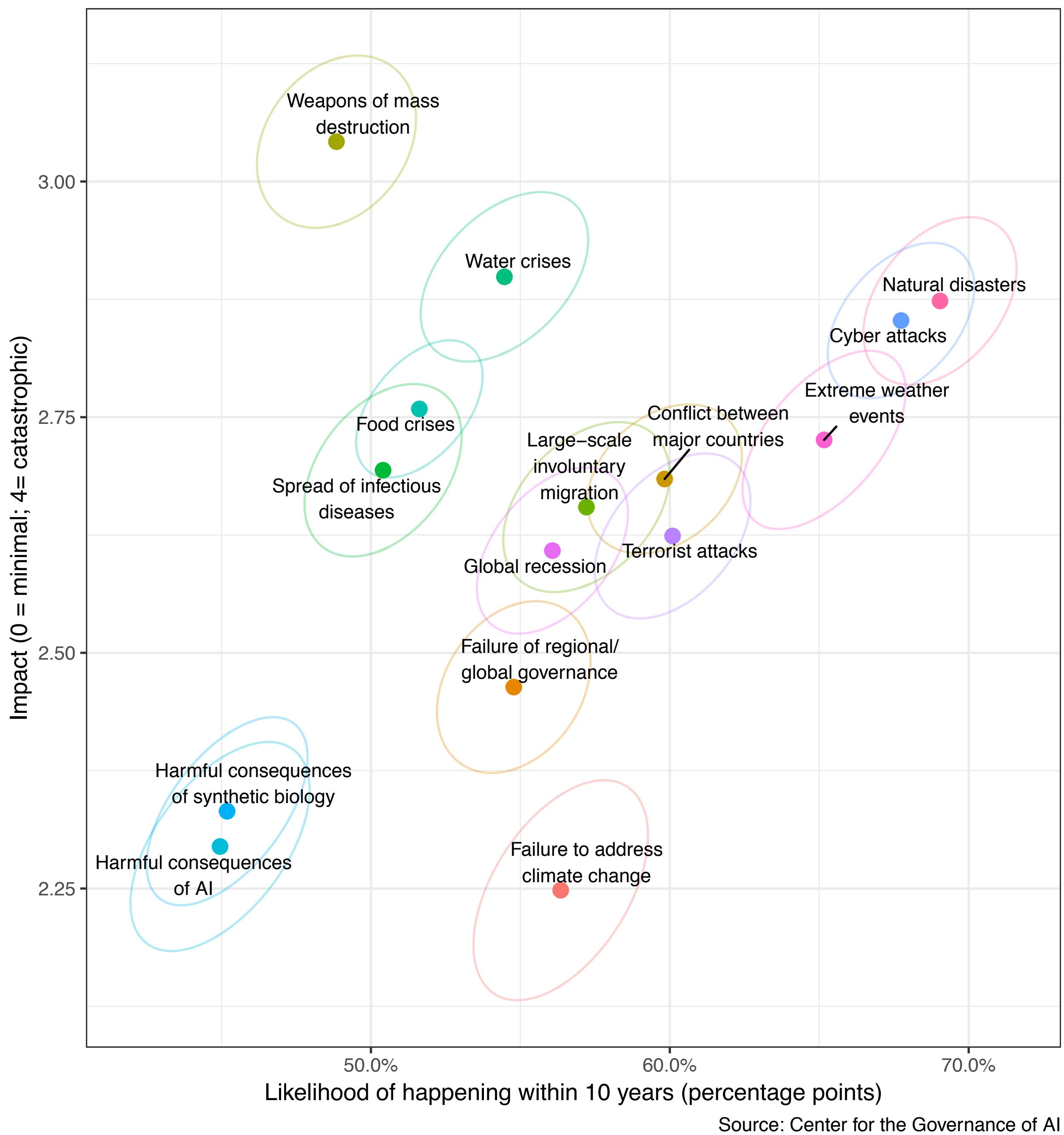


Figure 2.8: The American public's perceptions of 15 potential global risks

perceived to be the most likely as well as the most impactful. These include natural disasters, cyber attacks, and extreme weather events.

The American public and the nearly 1,000 experts surveyed by the World Economic Forum share similar views regarding most of the potential global risks we asked about (World Economic Forum 2018). Both the public and the experts rank extreme weather events, natural disasters, and cyber attacks as the top three most likely global risks; likewise, both groups consider weapons of mass destruction to be the most impactful. Nevertheless, compared with experts, Americans offer a lower estimate of the likelihood and impact of the failure to address climate change.

The American public appears to over-estimate the likelihoods of these risks materializing within 10 years. The mean responses suggest (assuming independence) that about eight (out of 15) of these global risks, which would have a significant negative impact on at least 10% of the world's population, will take place in the next 10 years. One explanation for this is that it arises from the broad misconception that the world is in a much worse state than it is in reality (Pinker 2018; Rosling, Rönnlund, and Rosling 2018). Another explanation is that it arises as a byproduct of respondents interpreting "significant negative impact" in a relatively minimal way, though this interpretation is hard to sustain given the mean severity being between "moderate" and "severe." Finally, this result may be because subjects centered their responses within the distribution of our response options, the middle value of which was the 40-60% option; thus, the likelihoods should not be interpreted literally in the absolute sense.

The adverse consequences of AI within the next 10 years appear to be a relatively low priority in respondents' assessment of global risks. It – along with adverse consequences of synthetic biology – occupy the lower left quadrant, which contains what are perceived to be lower-probability, lower-impact risks.[7] These risks are perceived to be as impactful (within the next 10 years) as the failure to address climate change, though less probable. One interpretation of this is that the average American simply does not regard AI as posing a substantial global risk. This interpretation, however, would be in tension with some expert assessment of catastrophic risks that suggests unsafe AI could pose significant danger (World Economic Forum 2018; Sandberg and Bostrom 2008). The gap between experts and the public's assessment suggests that this is a fruitful area for efforts to educate the public.

Another interpretation of our results is that Americans do have substantial concerns about the long-run impacts of advanced AI, but they do not see these risks as likely in the coming 10 years. As support for this interpretation, we later find that 12% of American's believe the impact of high-level machine intelligence will be "extremely bad, possibly human extinction," and 21% that it will be "on balance bad." Still, even though the median respondent expects around a 54% chance of high level machine intelligence within 10 years, respondents may believe that the risks from high level machine intelligence will manifest years later. If we assume respondents believe global catastrophic risks from AI only emerge from high-level AI, we can infer an implied global risk, conditional on high-level AI (within 10 years), of 80%. Future work should try to unpack and understand these beliefs.

## 2.5 Americans' understanding of key technology terms

We used a survey experiment to understand how the public understands the terms *AI*, *automation*, *machine learning*, and *robotics*. (Details of the survey experiment are found in Appendix B.) We randomly assigned each respondent one of these terms and asked them:

> In your opinion, which of the following technologies, if any, uses [artificial intelligence (AI)/automation/machine learning/robotics]? Select all that apply.

Because we wanted to understand respondents' perceptions of these terms, we did not define any of the terms. Respondents were asked to consider 10 technological applications, each of which uses AI or machine learning.

Though the respondents show at least a partial understanding of the terms and can identify their use within the considered technological applications correctly, the respondents underestimate the prevalence of AI, machine learning, and robotics in everyday technological applications, as reported in Figure 2.9. (See Appendix C for details of our statistical analysis.)

[7] The World Economic Forum's survey asked experts to evaluate the "adverse consequences of technological advances," defined as "[i]ntended or unintended adverse consequences of technological advances such as artificial intelligence, geo-engineering and synthetic biology causing human, environmental and economic damage." The experts considered these "adverse consequences of technological advances" to be less likely and lower-impact, compared with other potential risks.

Among those assigned the term *AI*, a majority think that virtual assistants (63%), smart speakers (55%), driverless cars (56%), social robots (64%), and autonomous drones use AI (54%). Nevertheless, a majority of respondents assume that Facebook photo tagging, Google Search, Netflix or Amazon recommendations, or Google Translate do not use AI.

Why did so few respondents consider the products and services we listed to be applications of AI, automation, machine learning, or robotics?

A straightforward explanation is that inattentive respondents neglect to carefully consider or select the items presented to them (i.e., non-response bias). Even among those assigned the term *robotics*, only 62% selected social robots and 68% selected industrial robots. Our analysis (found in Appendix C) confirms that respondent inattention, defined as spending too little or too much time on the survey, predicts non-response to this question.

Another potential explanation for the results is that the American public – like the public elsewhere – lack awareness of AI or machine learning. As a result, the public does not know that many tech products and services use AI or machine learning. According to a 2017 survey, nearly half of Americans reported that they were unfamiliar with AI (Morning Consult 2017). In the same year, only 9% of the British public said they had heard of the term "machine learning" (Ipsos MORI 2018). Similarly, less than half of EU residents reported hearing, reading, or seeing something about AI in the previous year (Eurobarometer 2017).

Finally, the so-called "AI effect" could also explain the survey result. The AI effect describes the phenomenon that the public does not consider an application that uses AI to utilize AI once that application becomes commonplace (McCorduck 2004). Because 85% of Americans report using digital products that deploy AI (e.g., navigation apps, video or music streaming apps, digital personal assistants on smartphones, etc.) (Reinhart 2018), they may not think that these everyday applications deploy AI.

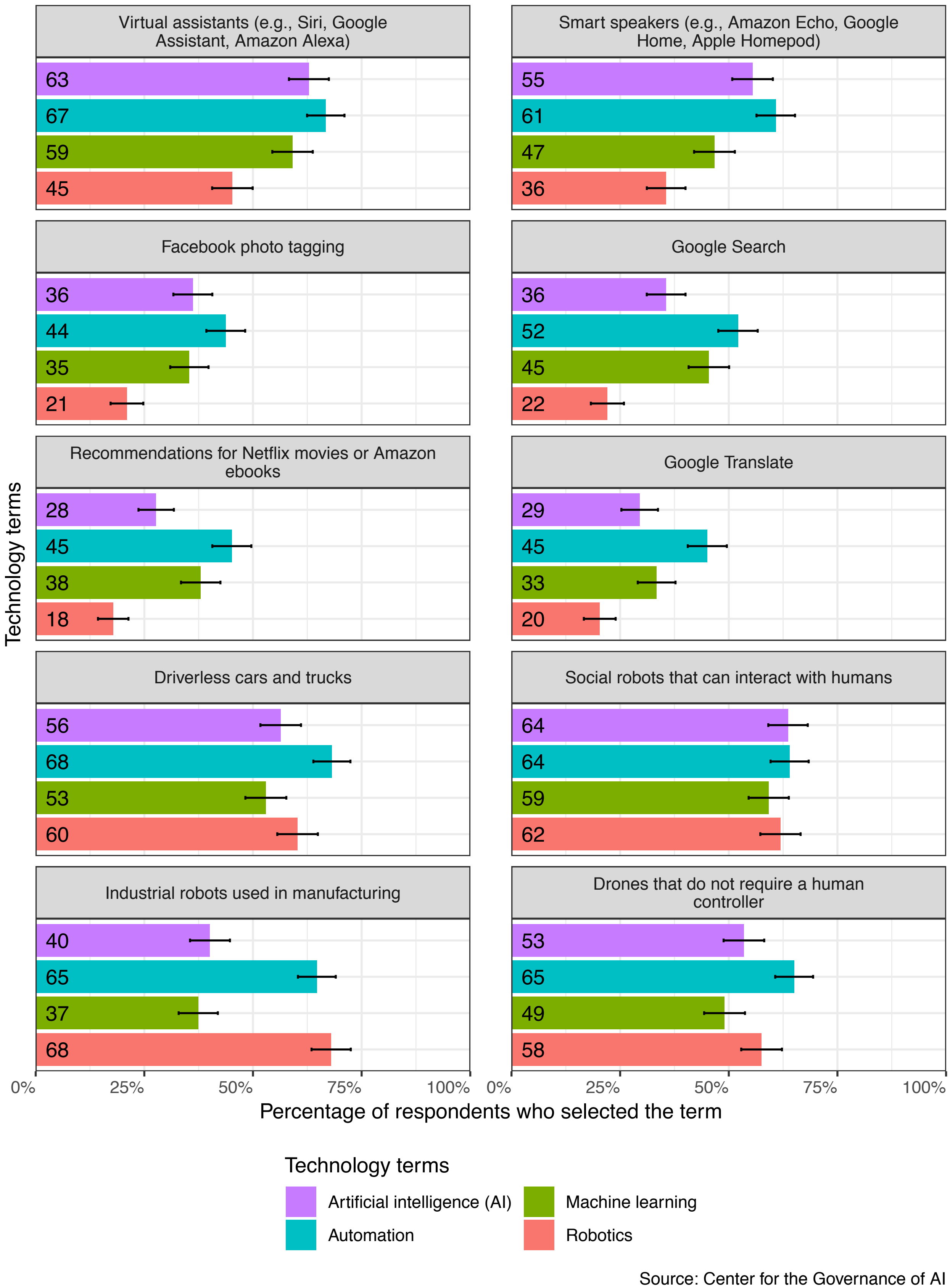


Figure 2.9: What applications or products that the public thinks use AI, automation, machine learning, or robotics

# 3 Public opinion on AI governance

## 3.1 Americans consider many AI governance challenges to be important; prioritize data privacy and preventing AI-enhanced cyber attacks, surveillance, and digital manipulation

We sought to understand how Americans prioritize policy issues associated with AI. Respondents were asked to consider five AI governance challenges, randomly selected from a test of 13 (see Appendix B for the text); the order these five were to each respondent was also randomized.

After considering each governance challenge, respondents were asked how likely they think the challenge will affect large numbers of people 1) in the U.S. and 2) around the world within 10 years.

We use scatterplots to visualize our survey results. In Figure 3.1, the *x*-axis is the perceived likelihood of the problem happening to large numbers of people in the U.S. In Figure 3.2, the *x*-axis is the perceived likelihood of the problem happening to large numbers of people around the world. The *y*-axes on both Figure 3.1 and 3.2 represent respondents' perceived issue importance, from 0 (not at all important) to 3 (very important). Each dot represents the mean perceived likelihood and issue importance, and the correspondent ellipse represents the 95% bivariate confidence region.

Americans consider all the AI governance challenges we present to be important: the mean perceived issues importance of each governance challenge is between "somewhat important" (2) and "very important" (3), though there is meaningful and discernible variation across items.

The AI governance challenges Americans think are most likely to impact large numbers of people, and are important for tech companies and governments to tackle, are found in the upper-right quadrant of the two plots. These issues include data privacy as well as AI-enhanced cyber attacks, surveillance, and digital manipulation. We note that the media have widely covered these issues during the time of the survey.

There are a second set of governance challenges that are perceived on average, as about 7% less likely, and marginally less important. These include autonomous vehicles, value alignment, bias in using AI for hiring, the U.S.-China arms race, disease diagnosis, and technological unemployment. Finally, the third set of challenges are perceived on average another 5% less likely, and about equally important, including criminal justice bias and critical AI systems failures.

We also note that Americans predict that all of the governance challenges mentioned in the survey, besides protecting data privacy and ensuring the safety of autonomous vehicles, are more likely to impact people around the world than to affect people in the U.S. While most of the statistically significant differences are substantively small, one difference stands out: Americans think that autonomous weapons are 7.6 percentage points more likely to impact people around the world than Americans. (See Appendix C for details of these additional analyses.)

We want to reflect on one result. "Value alignment" consists of an abstract description of alignment problem and a reference to what sounds like individual level harms: "while performing jobs [they could] unintentionally make decisions that go against the values of its human users, such as physically harming people." "Critical AI systems failures," by contrast, references military or critical infrastructure uses, and unintentional accidents leading to "10 percent or more of all humans to die." The latter was weighted as less important than the former: we interpret this as a probability weighted assessment of importance, since presumably the latter, were it to happen, is much more important. We thus think the issue importance question should be interpreted in a way that down-weights low probability risks. This perspective also plausibly applies to the "impact" measure for our global risks analysis, which placed "harmful consequences of synthetic biology" and "failure to address climate change" as less impactful than most other risks.

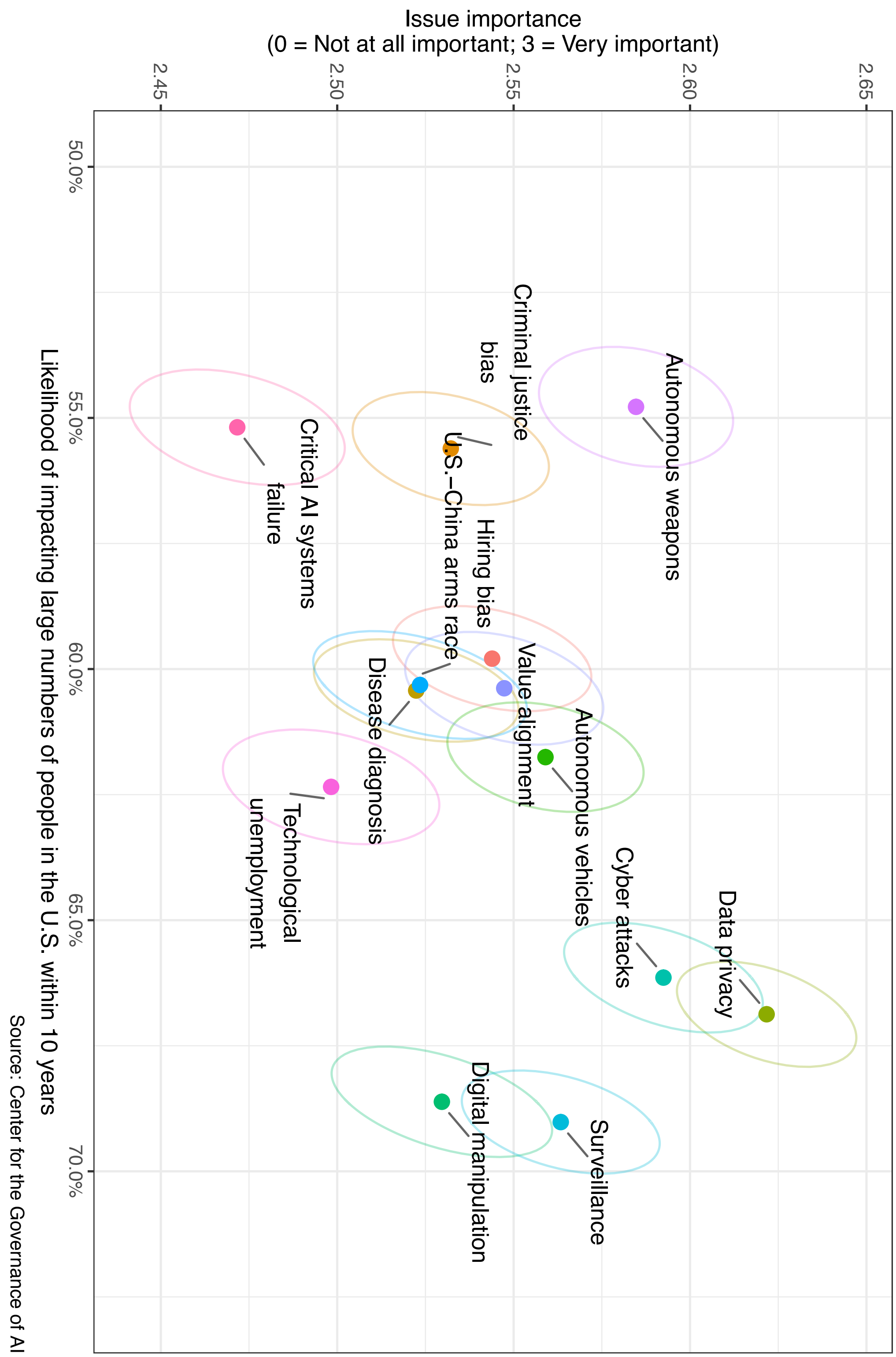


Figure 3.1: Perceptions of AI governance challenges in the U.S.

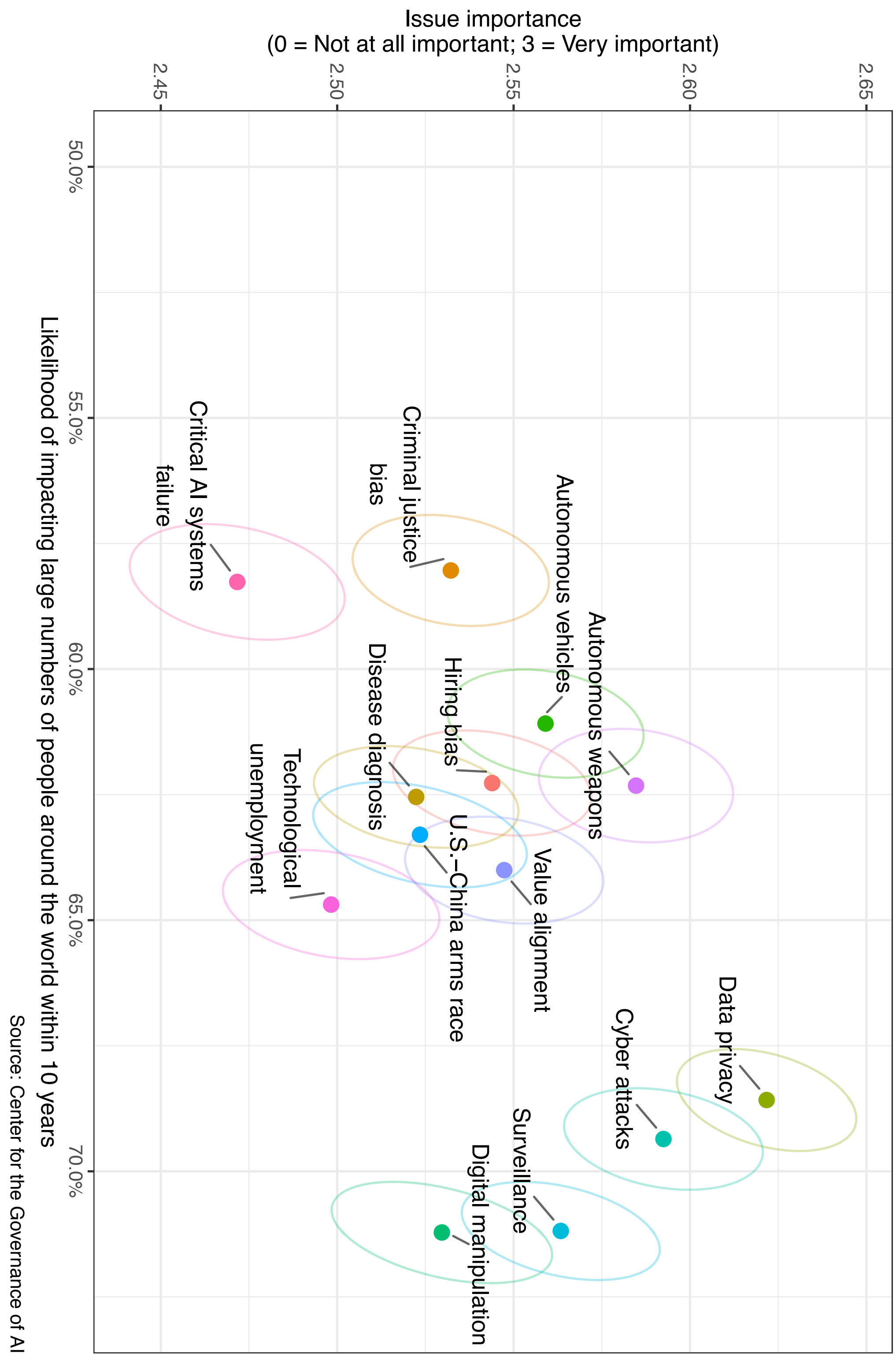


Figure 3.2: Perceptions of AI governance challenges around the world

## 3.2 Americans who are younger, who have CS or engineering degrees express less concern about AI governance challenges

We performed further analysis by calculating the percentage of respondents in each subgroup who consider each governance challenge to be "very important" for governments and tech companies to manage. (See Appendix C for additional data visualizations.) In general, differences in responses are more salient across demographic subgroups than across governance challenges. In a linear multiple regression predicting perceived issue importance using demographic subgroups, governance challenges, and the interaction between the two, we find that the stronger predictors are demographic subgroup variables, including age group and having CS or programming experience.

Two highly visible patterns emerge from our data visualization. First, a higher percentage of older respondents, compared with younger respondents, consider nearly all AI governance challenges to be "very important." As discussed previously, we find that older Americans, compared with younger Americans, are less supportive of developing AI. Our results here might explain this age gap: older Americans see each AI governance challenge as substantially more important than do younger Americans. Whereas 85% of Americans older than 73 consider each of these issues to be very important, only 40% of Americans younger than 38 do.

Second, those with CS or engineering degrees, compared with those who do not, rate all AI governance challenges as less important. This result could explain our previous finding that those with CS or engineering degrees tend to exhibit greater support for developing AI.[8]

## 3.3 Americans place the most trust in the U.S. military and universities to build AI; trust tech companies and non-governmental organizations more than the government to manage the technology

Respondents were asked how much confidence they have in various actors to develop AI. They were randomly assigned five actors out of 15 to evaluate. We provided a short description of actors that are not well-known to the public (e.g., NATO, CERN, and OpenAI).

Also, respondents were asked how much confidence, if any, they have in various actors to manage the development and use of AI in the best interests of the public. They were randomly assigned five out of 15 actors to evaluate. Again, we provided a short description of actors that are not well-known to the public (e.g., AAAI and Partnership on AI). Confidence was measured using the same four-point scale described above.[9]

Americans do not express great confidence in most actors to develop or to manage AI, as reported in Figures 3.4 and 3.5. A majority of Americans do not have a "great deal" or even a "fair amount" of confidence in any institution, except university researchers, to develop AI. Furthermore, Americans place greater trust in tech companies and non-governmental organizations (e.g., OpenAI) than in governments to manage the development and use of the technology.

University researchers and the U.S. military are the most trusted groups to develop AI: about half of Americans express a "great deal" or even a "fair amount" of confidence in them. Americans express slightly less confidence in tech companies, non-profit organizations (e.g., OpenAI), and American intelligence organizations. Nevertheless, opinions toward individual actors within each of these groups vary. For example, while 44% of Americans indicated they feel a "great deal" or even a "fair amount" of confidence in tech companies, they rate Facebook as the least trustworthy of all the actors. More than four in 10 indicate that they have no confidence in the company.[10]

---

[8] In Table C.15, we report the results of a saturated linear model using demographic variables, governance challenges, and the interaction between these two types of variables to predict perceived issue importance. We find that those who are 54-72 or 73 and older, relative to those who are below 38, view the governance issues as more important (two-sided $p$-value $< 0.001$ for both comparisons). Furthermore, we find that those who have CS or engineering degrees, relative to those who do not, view the governance challenges as less important (two-sided $p$-value $< 0.001$).

[9] The two sets of 15 actors differed slightly because for some actors it seemed inappropriate to ask one or the other question. See Appendix B for the exact wording of the questions and descriptions of the actors.

[10] Our survey was conducted between June 6 and 14, 2018, shortly after the fallout of the Facebook/Cambridge Analytica scandal. On April 10-11, 2018, Facebook CEO Mark Zuckerberg testified before the U.S. Congress regarding the Cambridge Analytica data leak. On May 2, 2018, Cambridge Analytica announced its shutdown. Nevertheless, Americans' distrust of the company existed before the Facebook/Cambridge Analytica scandal. In a pilot survey that we conducted on Mechanical Turk during July 13-14, 2017, respondents indicated a substantially lower level of confidence in Facebook, compared with other actors, to develop and regulate AI.

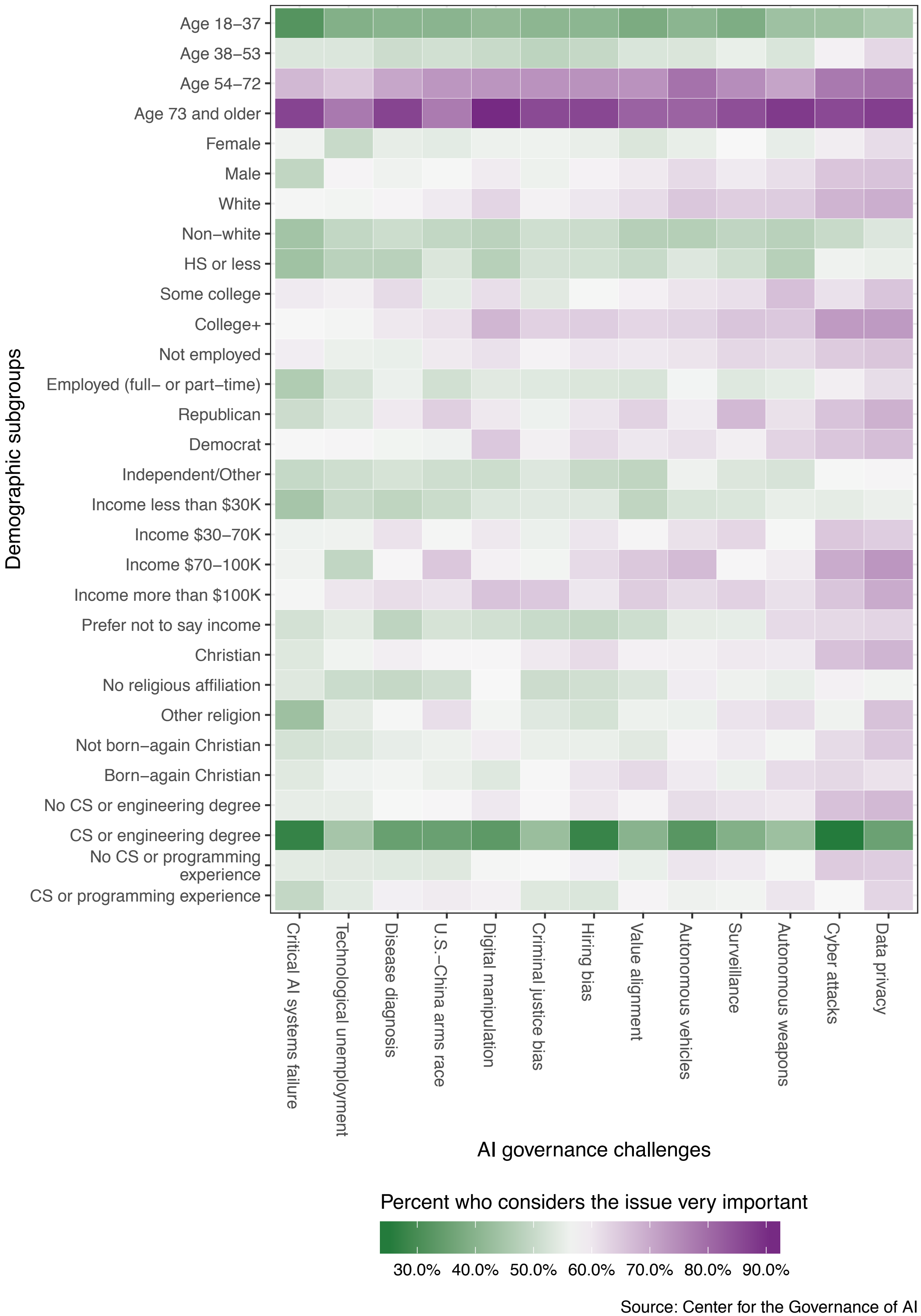


Figure 3.3: AI governance challenges: issue importance by demographic subgroups

The results on the public's trust of various actors to manage the develop and use of AI provided are similar to the results discussed above. Again, a majority of Americans do not have a "great deal" or even a "fair amount" of confidence in any institution to manage AI. In general, the public expresses greater confidence in non-governmental organizations than in governmental ones. Indeed, 41% of Americans express a "great deal" or even a "fair amount" of confidence in "tech companies," compared with 26% who feel that way about the U.S. federal government. But when presented with individual big tech companies, Americans indicate less trust in each than in the broader category of "tech companies." Once again, Facebook stands out as an outlier: respondents give it a much lower rating than any other actor. Besides "tech companies," the public places relatively high trust in intergovernmental research organizations (e.g., CERN), the Partnership on AI, and non-governmental scientific organizations (e.g., AAAI). Nevertheless, because the public is less familiar with these organizations, about one in five respondents give a "don't know" response.

Mirroring our findings, recent survey research suggests that while Americans feel that AI should be regulated, they are unsure *who* the regulators should be. When asked who "should decide how AI systems are designed and deployed," half of Americans indicated they do not know or refused to answer (West 2018a). Our survey results seem to reflect Americans' general attitudes toward public institutions. According to a 2016 Pew Research Center survey, an overwhelming majority of Americans have "a great deal" or "a fair amount" of confidence in the U.S. military and scientists to act in the best interest of the public. In contrast, public confidence in elected officials is much lower: 73% indicated that they have "not too much" or "no confidence" (Funk 2017). Less than one-third of Americans thought that tech companies do what's right "most of the time" or "just about always"; moreover, more than half think that tech companies have too much power and influence in the U.S. economy (Smith 2018). Nevertheless, Americans' attitude toward tech companies is not monolithic but varies by company. For instance, our research findings reflect the results from a 2018 survey, which reported that a higher percentage of Americans trusted Apple, Google, Amazon, Microsoft, and Yahoo to protect user information than trust Facebook to do so (Ipsos and Reuters 2018).

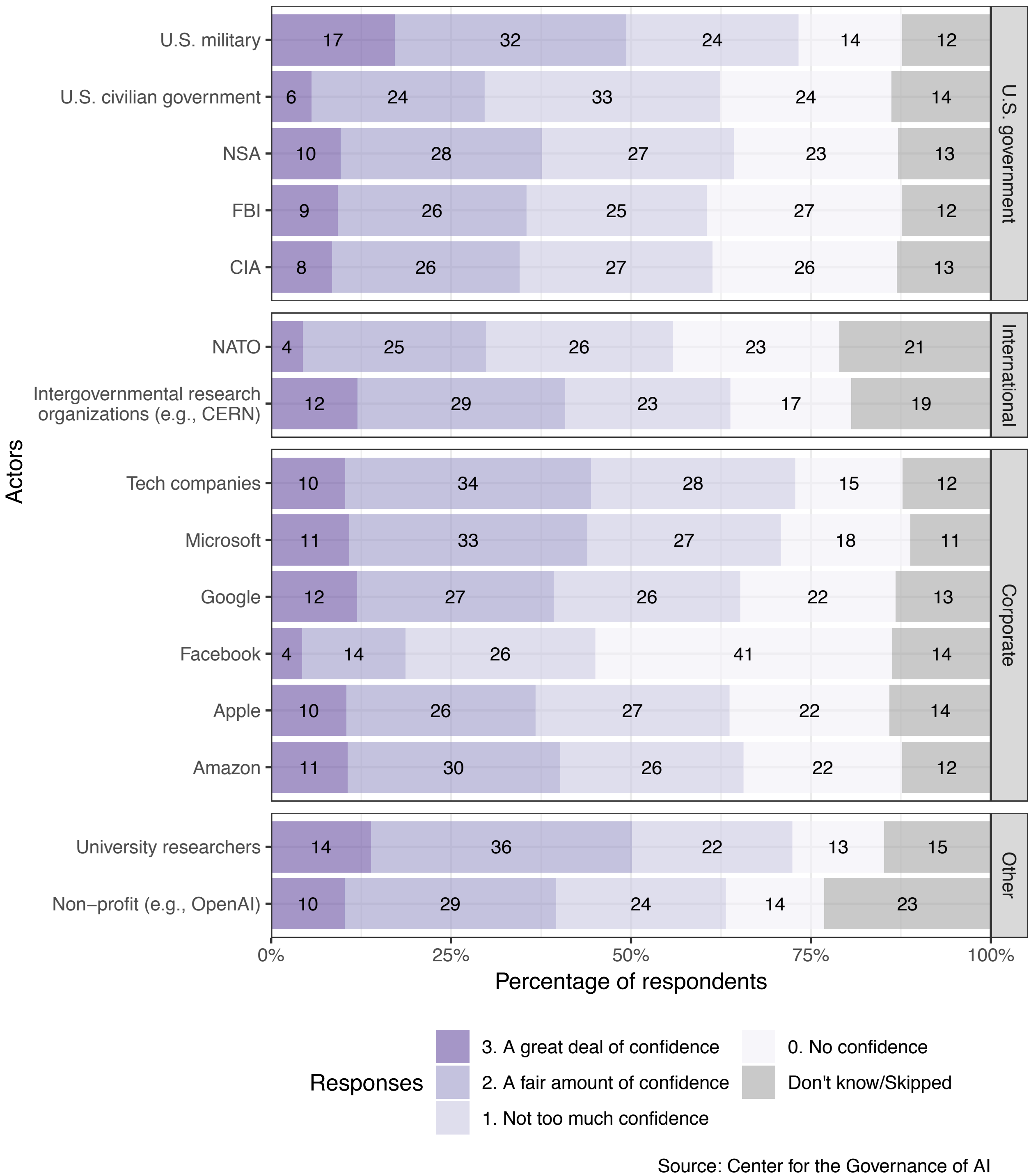


Figure 3.4: Trust in various actors to develop AI: distribution of responses

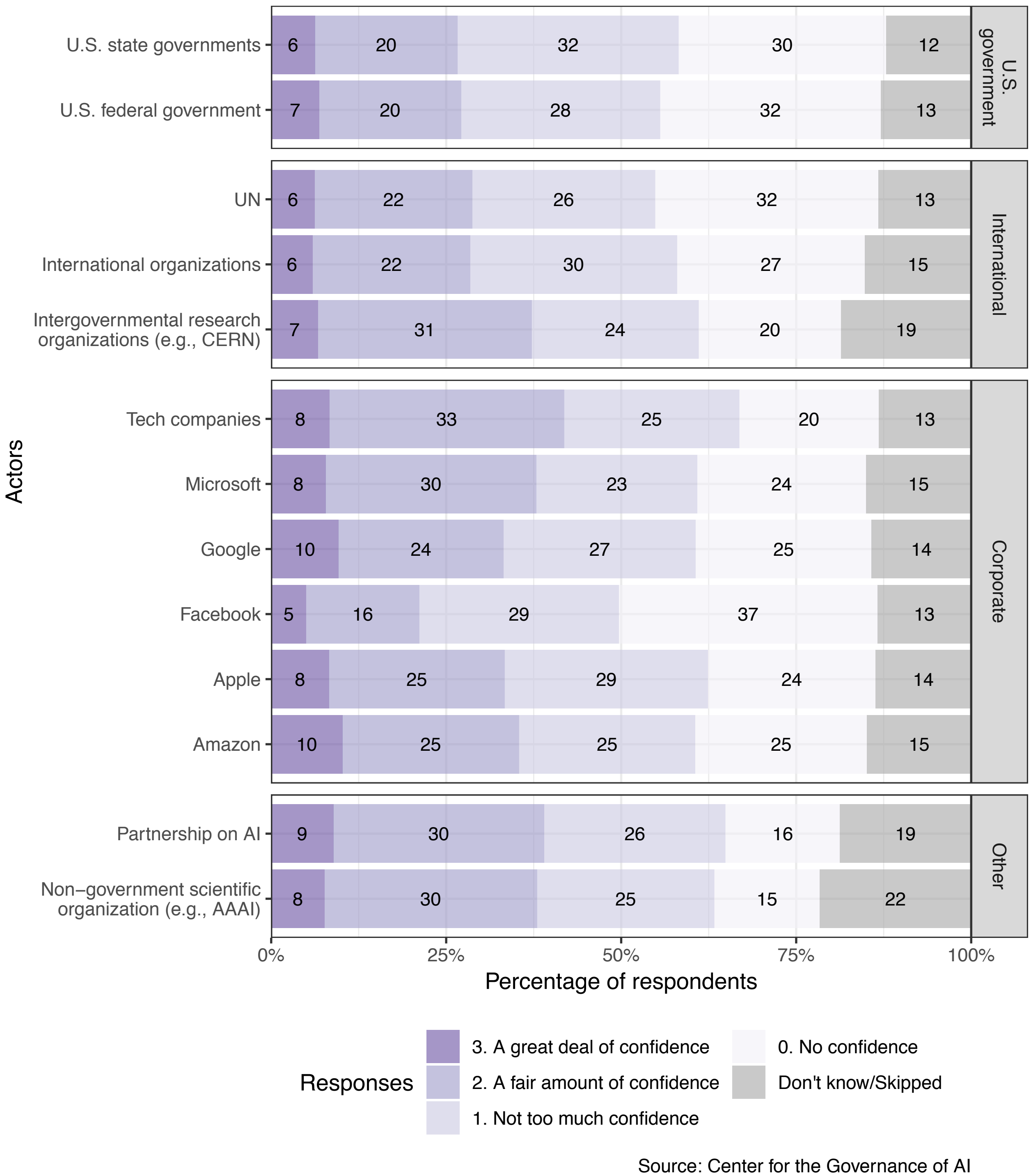


Figure 3.5: Trust in various actors to manage AI: distribution of responses

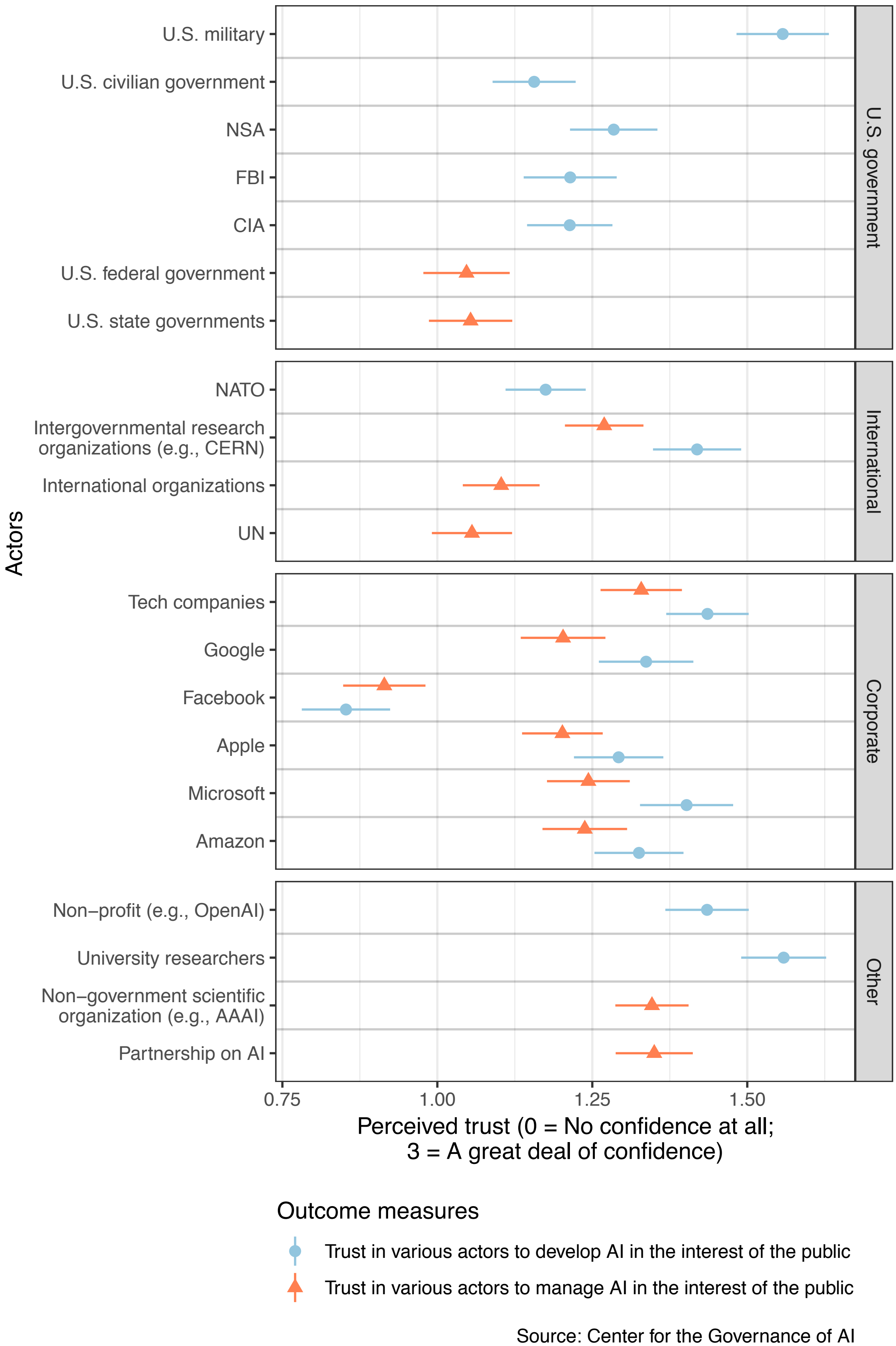


Figure 3.6: Trust in various actors to develop and manage AI in the interest of the public

# 4 AI policy and U.S.-China relations

## 4.1 Americans underestimate the U.S. and China's AI research and development

In this survey experiment, we asked respondents to consider either the U.S. or China's status in AI research and development (R&D). (Details of the survey experiment are found in Appendix B.) Respondents were asked the following:

> Compared with other industrialized countries, how would you rate [the U.S./China] in AI research and development?

By almost any metric of absolute achievement (not per-capita achievement), the U.S. and China are the world leaders in the research and development of AI. The U.S. and China led participation in the 2017 AAAI Conference, one of the important ones in the field of AI research; 34% of those who presented papers had a U.S. affiliation while 23% had a Chinese affiliation (Goldfarb and Trefler 2018). The U.S. and China also have the highest percentage of the world's AI companies, 42% and 23%, respectively (IT Juzi and Tencent Institute 2017). Most clearly, the U.S. and China have the largest technology companies focused on developing and using AI (Google, Facebook, and Amazon in the U.S.; Tencent, Alibaba, and Baidu in China).

Yet, only a minority of the American public thinks the U.S. or China's AI R&D is the "best in the world," as reported in Figure 4.1. Our survey result seems to reflect the gap between experts and the public's perceptions of U.S.'s scientific achievements in general. While 45% of scientists in the American Association for the Advancement of Science think that scientific achievements in the U.S. are the best in the world, only 15% of the American public express the same opinion (Funk and Rainie 2015).

According to our survey, there is not a clear perception by Americans that the U.S. has the best AI R&D in the world. While 10% of Americans believe that the U.S. has the best AI R&D in the world, 7% think that China does. Still, 36% of Americans believe that the U.S.'s AI R&D is "above average" while 45% think China's is "above average." Combining these into a single measure of whether the country has "above average" or "best in the world" AI R&D, Americans do not perceive the U.S. to be superior, and the results lean towards the perception that China is superior. Note that we did not ask for a direct comparison, but instead asked each respondent to evaluate one country independently on an absolute scale Appendix C.

Our results mirror those from a recent survey that finds that Americans think that China's AI capability will be on par with the U.S.'s in 10 years (West 2018b). The American public's perceptions could be caused by media narratives that China is catching up to the U.S. in AI capability (Kai-Fu 2018). Nevertheless, another study suggests that although China has greater access to big data than the U.S., China's AI capability is about half of the U.S.'s (Ding 2018). Exaggerating China's AI capability could exacerbate growing tensions between the U.S. and China (Zwetsloot, Toner, and Ding 2018). As such, future research should explore how factual – non-exaggerated – information about American and Chinese AI capabilities influences public opinions.

## 4.2 Communicating the dangers of a U.S.-China arms race requires explaining policy trade-offs

In this survey experiment, respondents were randomly assigned to consider different arguments about a U.S.-China arms race. (Details of the survey experiment are found in Appendix B.) All respondents were given the following prompt:

> Leading analysts believe that an AI arms race is beginning, in which the U.S. and China are investing billions of dollars to develop powerful AI systems for surveillance, autonomous weapons, cyber operations, propaganda, and command and control systems.

Those in the treatment condition were told they would read a short news article. The three treatments were:

1. **Pro-nationalist treatment**: The U.S. should invest heavily in AI to stay ahead of China; quote from a senior National Security Council official
2. **Risks of arms race treatment**: The U.S.-China arms race could increase the risk of a catastrophic war; quote from Elon Musk

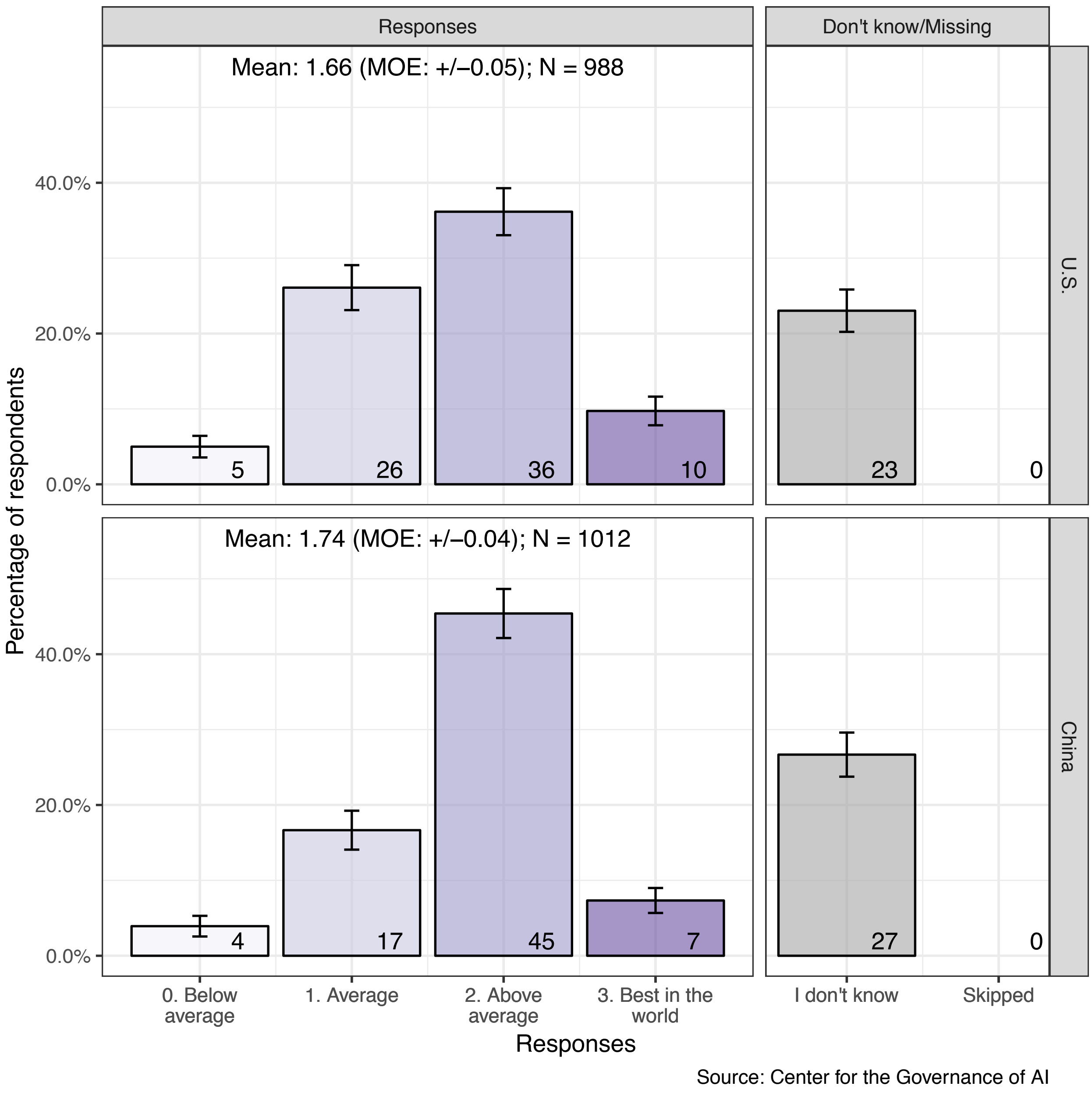


Figure 4.1: Comparing Americans' perceptions of U.S. and China's AI research and development quality

3. **One common humanity treatment**: The U.S.-China arms race could increase the risk of a catastrophic war; quote from Stephen Hawking about using AI for the good of all people rather than destroying civilization

Respondents were asked to consider two statements and indicate whether they agree or disagree with them:

- The U.S. should invest more in AI military capabilities to make sure it doesn't fall behind China's, even if doing so may exacerbate the AI arms race.
- The U.S. should work hard to cooperate with China to avoid the dangers of an AI arms race, even if doing so requires giving up some of the U.S.'s advantages. Cooperation could include collaborations between American and Chinese AI research labs, or the U.S. and China creating and committing to common safety standards for AI.

Americans, in general, weakly agree that the U.S. should invest more in AI military capabilities *and* cooperate with China to avoid the dangers of an AI arms race, as seen in Figure 4.2. Many respondents do not think that the two policies are mutually exclusive. The correlation between responses to the two statements, unconditional on treatment assignment, is only -0.05. In fact, 29% of those who agree that the U.S. and China should cooperate also agree that the U.S. should invest more in AI military capabilities. (See Figure C.2 for the conditional percentages.)

Respondents assigned to read about the risks of an arms race (Treatment 2) indicate significantly higher agreement with the pro-cooperation statement (Statement 2) than the investing in AI military capabilities statement (Statement 1), according to Figure 4.4. Those assigned to Treatment 2 are more likely to view the two statements as mutually exclusive. In contrast, respondents assigned to the other conditions indicate similar levels of agreement with both statements.

After estimating the treatment effects, we find that the experimental messages do little to change the respondents' preferences. Treatment 2 is the one exception. Treatment 2 decreases respondents' agreement with the statement that the U.S. should invest more in AI military capabilities by 27%, as seen in Figure 4.3. Future research could focus on testing more effective messages, such as op-eds (Coppock et al. 2018) or videos (Paluck et al. 2015), which explains that U.S.'s investment in AI for military use will decrease the likelihood of cooperation with China.

## 4.3 Americans see the potential for U.S.-China cooperation on some AI governance challenges

We examined issue areas where Americans perceive likely U.S.-China cooperation. Each respondent was randomly assigned to consider three out of five AI governance challenges. For each challenge, the respondent was asked, "For the following issues, how likely is it that the U.S. and China can cooperate?". (See Appendix B for the question text.)

On each of these AI governance issues, Americans see some potential for U.S.-China cooperation, according to Figure 4.5. U.S.-China cooperation on value alignment is perceived to be the most likely (48% mean likelihood). Cooperation to prevent AI-assisted surveillance that violates privacy and civil liberties is seen to be the least likely (40% mean likelihood) – an unsurprising result since the U.S. and China take different stances on human rights.

Despite current tensions between Washington and Beijing, the Chinese government, as well as Chinese companies and academics, have signaled their willingness to cooperate on some governance issues. These include banning the use of lethal autonomous weapons (Kania 2018), building safe AI that is aligned with human values (China Institute for Science and Technology Policy at Tsinghua University 2018), and collaborating on research (News 2018). Most recently, the major tech company Baidu became the first Chinese member of the Partnership on AI, a U.S.-based multi-stakeholder organization committed to understanding and discussing AI's impacts (Cadell 2018).

In the future, we plan to survey Chinese respondents to understand how they view U.S.-China cooperation on AI and what governance issues they think the two countries could collaborate on.

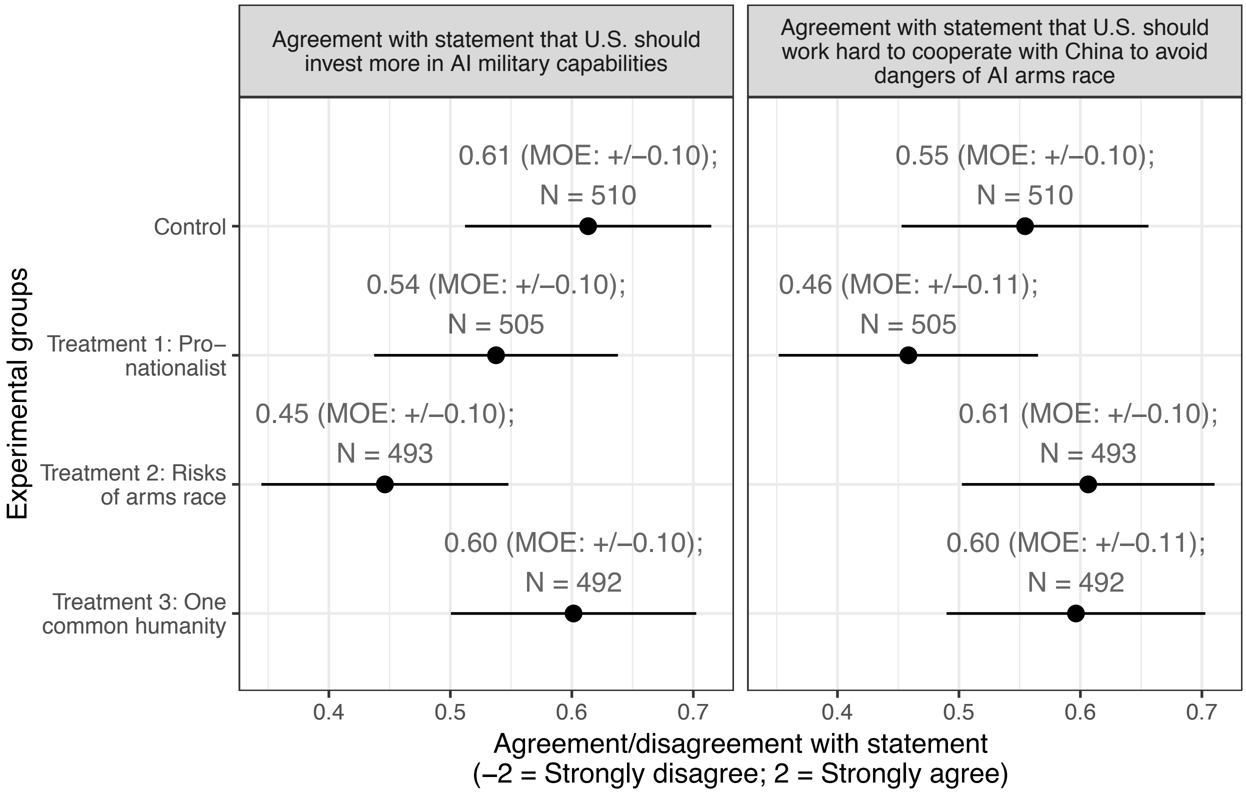


Figure 4.2: Responses from U.S.-China arms race survey experiment

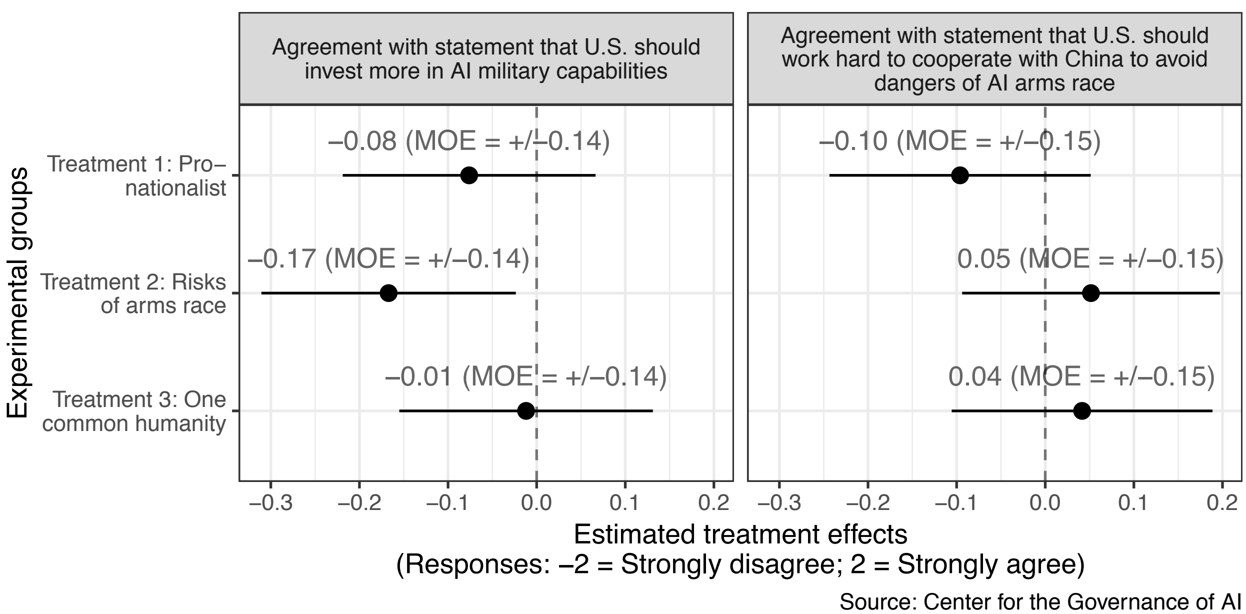


Figure 4.3: Effect estimates from U.S.-China arms race survey experiment

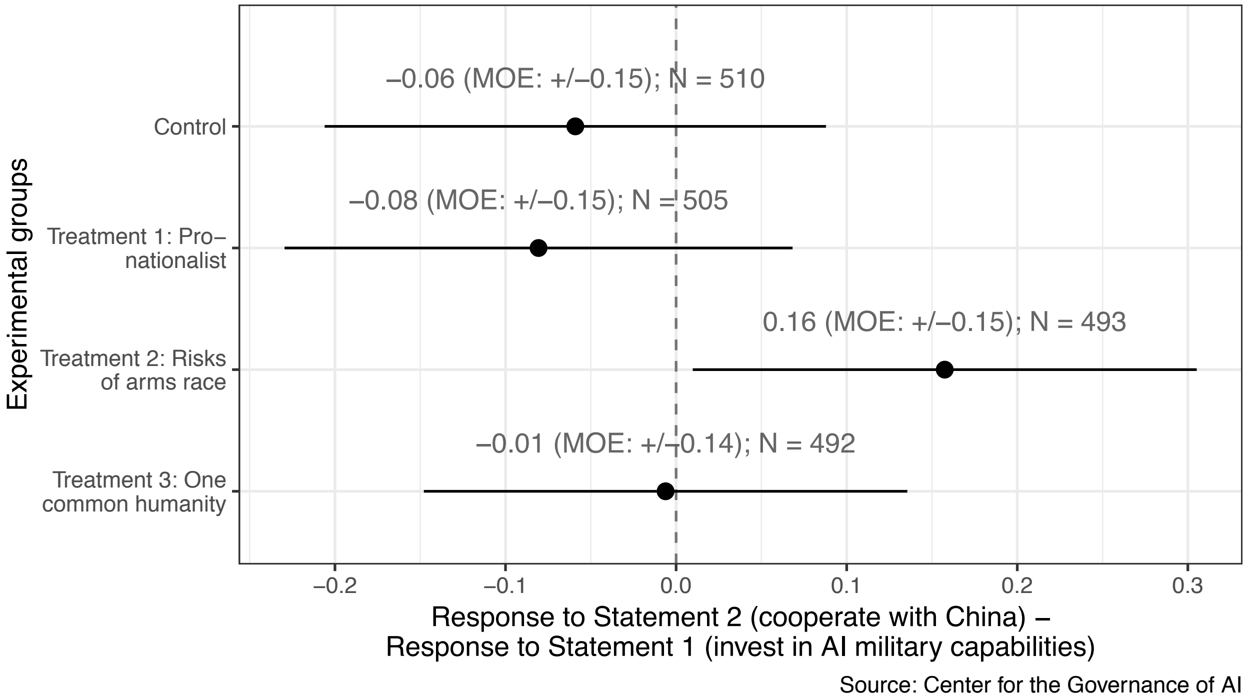


Figure 4.4: Difference in response to the two statements by experimental group

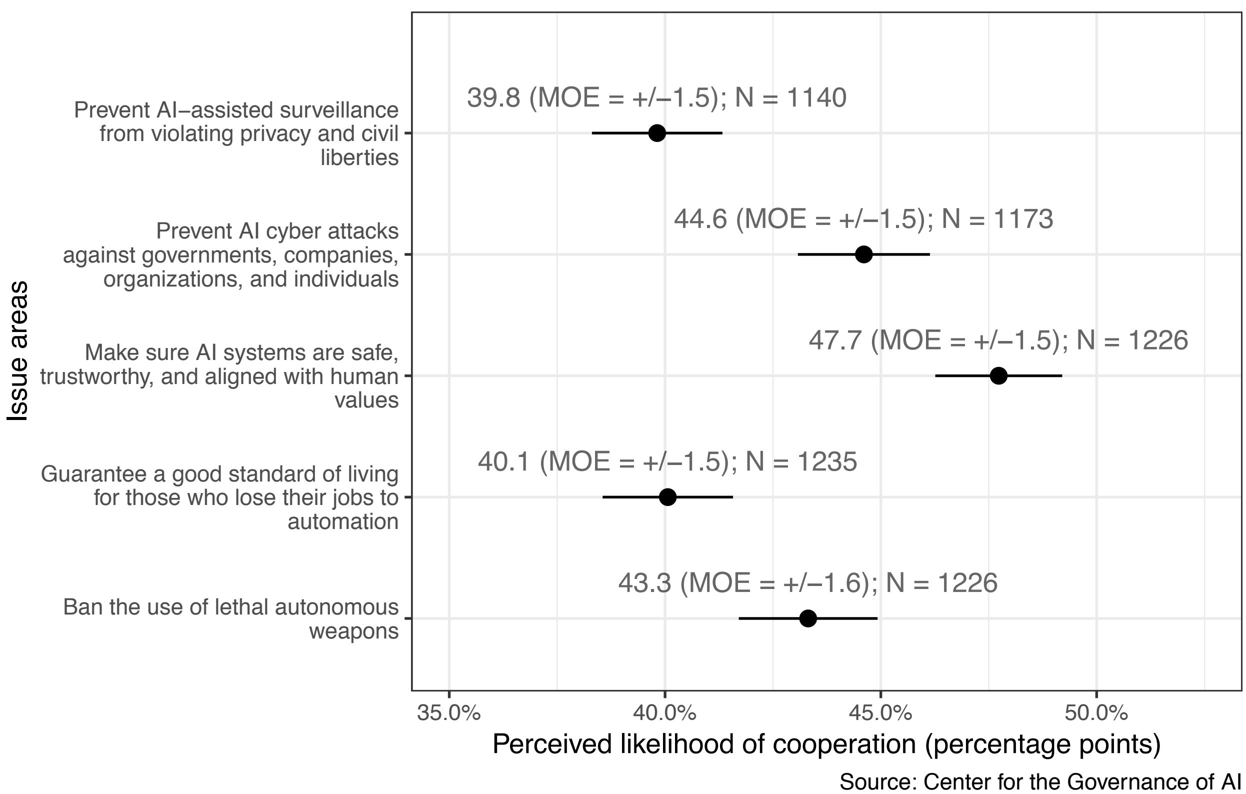


Figure 4.5: Issue areas for possible U.S.-China cooperation

# 5 Trend across time: attitudes toward workplace automation

Survey questions measuring Americans' perceptions of workplace automation have existed since the 1950s. Our research seeks to track changes in these attitudes across time by connecting past survey data with original, contemporary survey data.

## 5.1 Americans do not think that labor market disruptions will increase with time

American government agencies, think tanks, and media organizations began conducting surveys to study public opinion about technological unemployment during the 1980s when unemployment was relatively high. Between 1983 and 2003, the U.S. National Science Foundation (NSF) conducted eight surveys that asked respondents the following:

> In general, computers and factory automation will create more jobs than they will eliminate. Do you strongly agree, agree, disagree, or strongly disagree?

Our survey continued this time trend study by posing a similar – but updated – question (see Appendix B):

> Do you strongly agree, agree, disagree, or strongly disagree with the statement below?
>
> In general, automation and AI will create more jobs than they will eliminate.

Our survey question also addressed the chief ambiguity of the original question: lack of a future time frame. We used a survey experiment to help resolve this ambiguity by randomly assigning respondents to one of four conditions. We created three treatment conditions with the future time frames of 10 years, 20 years, and 50 years, as well as a control condition that did not specify a future time frame.

On average, Americans disagree with the statement more than they agree with it, although about a quarter of respondents in each experimental group give "don't know" responses. Respondents' agreement with the statement seems to increase slightly with the future time frame, but formal tests in Apppendix C reveal that there exist no significant differences between the responses to the differing future time frames. This result is puzzling from the perspective that AI and robotics will increasingly automate tasks currently done by humans. Such a view would expect more *disagreement* with the statement as one looks further into the future. One hypothesis to explain our results is that respondents believe the disruption from automation is destabilizing in the upcoming 10 years but eventually institutions will adapt and the labor market will stabilize. This hypothesis is consistent with our other finding that the median American predicts a 54% chance of high-level machine intelligence being developed within the next 10 years.

## 5.2 Extending the historical time trend

The percentage of Americans that disagrees with the statement that automation and AI will create more jobs than they destroy is similar to the historical rate of disagreement with the same statement about computers and factory automation. Nevertheless, the percentage who agree with the statement has decreased by 12 percentage points since 2003 while the percentage who responded "don't know" has increased by 18 percentage points since 2003, according to Figure 5.2.

There are three possible reasons for these observed changes. First, we have updated the question to ask about "automation and AI" instead of "computers and factory automation." The technologies we asked about could impact a wider swath of the economy; therefore, respondents may be more uncertain about AI's impact on the labor market. Second, there is a difference in survey mode between the historical data and our data. The NSF surveys were conducted via telephone while our survey is conducted online. Some previous research has shown that online surveys, compared with telephone surveys, produce a greater percentage of "don't know" responses (Nagelhout et al. 2010; Bronner and Kuijlen 2007). But, other studies have shown that online surveys cause no such effect (Shin, Johnson, and Rao 2012; Bech and Kristensen 2009). Third, the changes in the responses could be due to the actual changes in respondents' perceptions of workplace automation over time.

[12] Note that our survey asked respondents this question with the time frames 10, 20 and 50 years, whereas the NSF surveys provided no time frame.

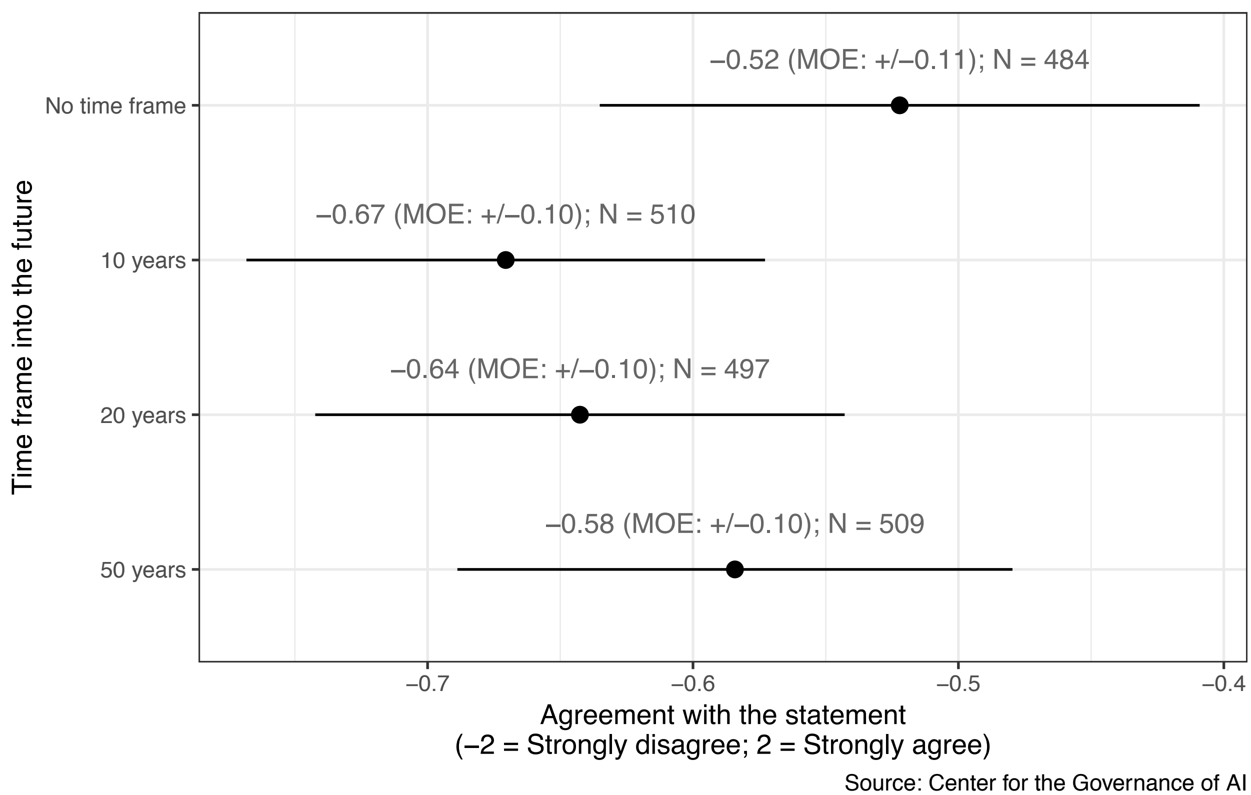


Figure 5.1: Agreement with the statement that automation and AI will create more jobs than it will eliminate

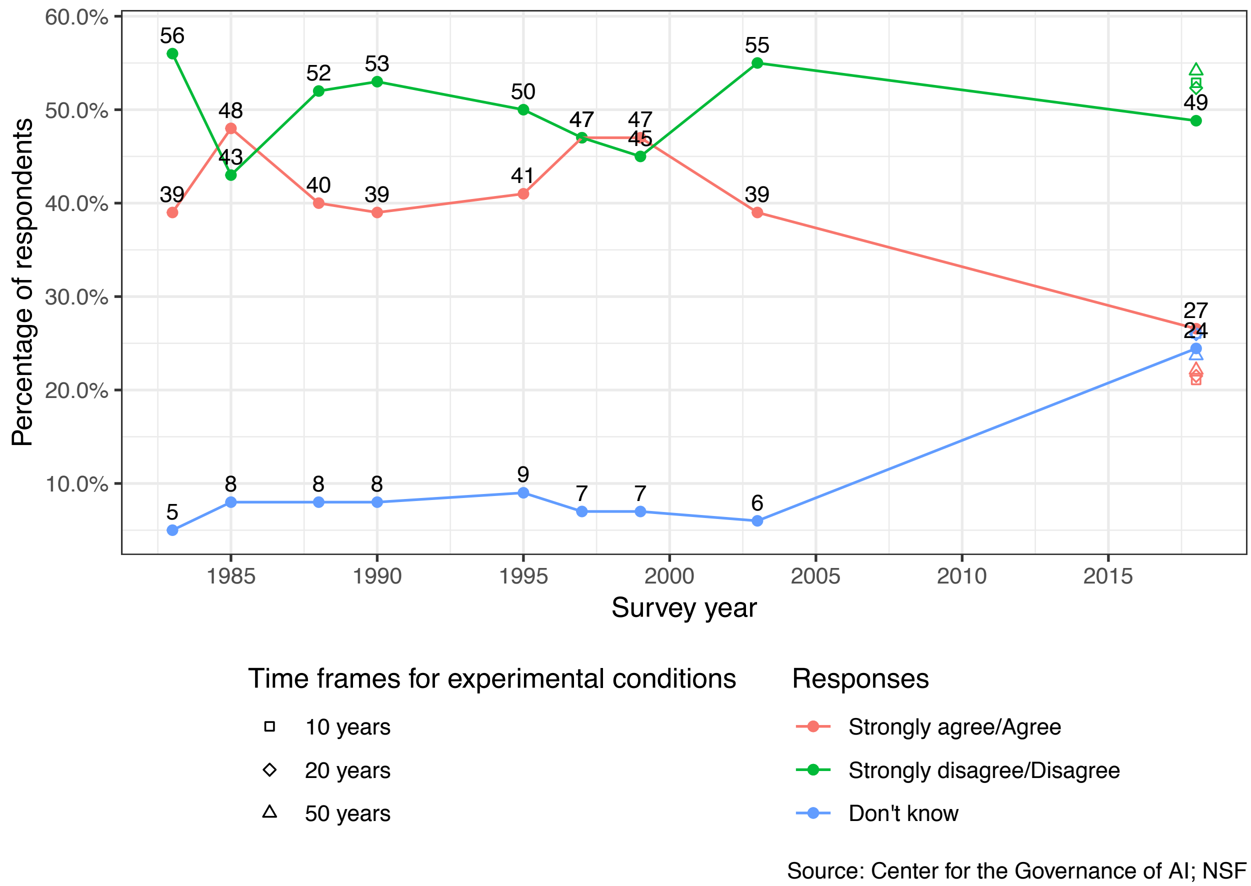


Figure 5.2: Response to statement that automation will create more jobs than it will eliminate[12] (data from before 2018 from National Science Foundation surveys)

# 6 High-level machine intelligence

## 6.1 The public predicts a 54% likelihood of high-level machine intelligence within 10 years

Respondents were asked to forecast when high-level machine intelligence will be developed. High-level machine intelligence was defined as the following:

> We have high-level machine intelligence when machines are able to perform almost all tasks that are economically relevant today better than the median human (today) at each task. These tasks include asking subtle common-sense questions such as those that travel agents would ask. For the following questions, you should ignore tasks that are legally or culturally restricted to humans, such as serving on a jury.[13]

Respondents were asked to predict the probability that high-level machine intelligence will be built in 10, 20, and 50 years.

We present our survey results in two ways. First, we show the summary statistics in a simple table. Next, to compare the public's forecasts with forecasts made by AI researchers in 2016 (Grace et al. 2018), we aggregated the respondents' forecasts using the same method. Note that Grace et al. (2018) gave a stricter definition of high-level machine intelligence that involved machines being better than all humans at all tasks.[14]

Table 6.1: Summary statistics of high-level machine intelligence forecast

| Year | Respondent type | 25th percentile | Median | Mean | 75th percentile | *N* |
|---|---|---|---|---|---|---|
| 2028 | All respondents | 30% | 54% | 54% | 70% | 2000 |
| 2038 | All respondents | 50% | 70% | 70% | 88% | 2000 |
| 2068 | All respondents | 70% | 88% | 80% | 97% | 2000 |
| 2028 | No CS or engineering degree | 30% | 54% | 55% | 70% | 1805 |
| 2038 | No CS or engineering degree | 50% | 70% | 71% | 88% | 1805 |
| 2068 | No CS or engineering degree | 70% | 88% | 81% | 98% | 1805 |
| 2028 | CS or engineering degree | 30% | 50% | 48% | 70% | 195 |
| 2038 | CS or engineering degree | 50% | 70% | 67% | 88% | 195 |
| 2068 | CS or engineering degree | 50% | 73% | 69% | 97% | 195 |

Respondents predict that high-level machine intelligence will arrive fairly quickly. The median respondent predicts a likelihood of 54% by 2028, a likelihood of 70% by 2038, and a likelihood of 88% by 2068, according to Table 6.1.

These predictions are considerably sooner than the predictions by experts in two previous surveys. In Müller and Bostrom (2014), expert respondents predict a 50% probability of high-level human intelligence being developed by 2040-2050 and 90% by 2075. In Grace et al. (2018), experts predict that there is a 50% chance that high-level machine intelligence will be built by 2061. Plotting the public's forecast with the expert forecast from Grace et al. (2018), we see that the public predicts high-level machine intelligence arriving much sooner than experts forecast. Employing the same aggregation method used in Grace et al. (2018), Americans predict that there is a 50% chance that high-level machine intelligence will be developed by 2026.

Results in Walsh (2018) also show that the non-experts (i.e., readers of a news article about AI) are more optimistic in their predictions of high-level machine intelligence compared with experts. In Walsh's study, the median AI expert predicted a 50% probability of high-level machine intelligence by 2061 while the median non-expert predicted a 50% probability by 2039. In our survey, respondents with CS or engineering degrees, compared with those who do not, provide a somewhat longer timeline for the arrival of high-level machine intelligence, according to Table 6.1. Nevertheless, those with CS or

[13] Note that our definition of high-level machine intelligence is equivalent to what many would consider human-level machine intelligence. Details of the question are found in Appendix B.

[14] In Grace et al. (2018), each respondent provides three data points for their forecast, and these are fitted to the Gamma CDF by least squares to produce the individual cumulative distribution function (CDFs). Each "aggregate forecast" is the mean distribution over all individual CDFs (also called the "mixture" distribution). The confidence interval is generated by bootstrapping (clustering on respondents) and plotting the 95% interval for estimated probabilities at each year. Survey weights are not used in this analysis due to problems incorporating survey weights into the bootstrap.

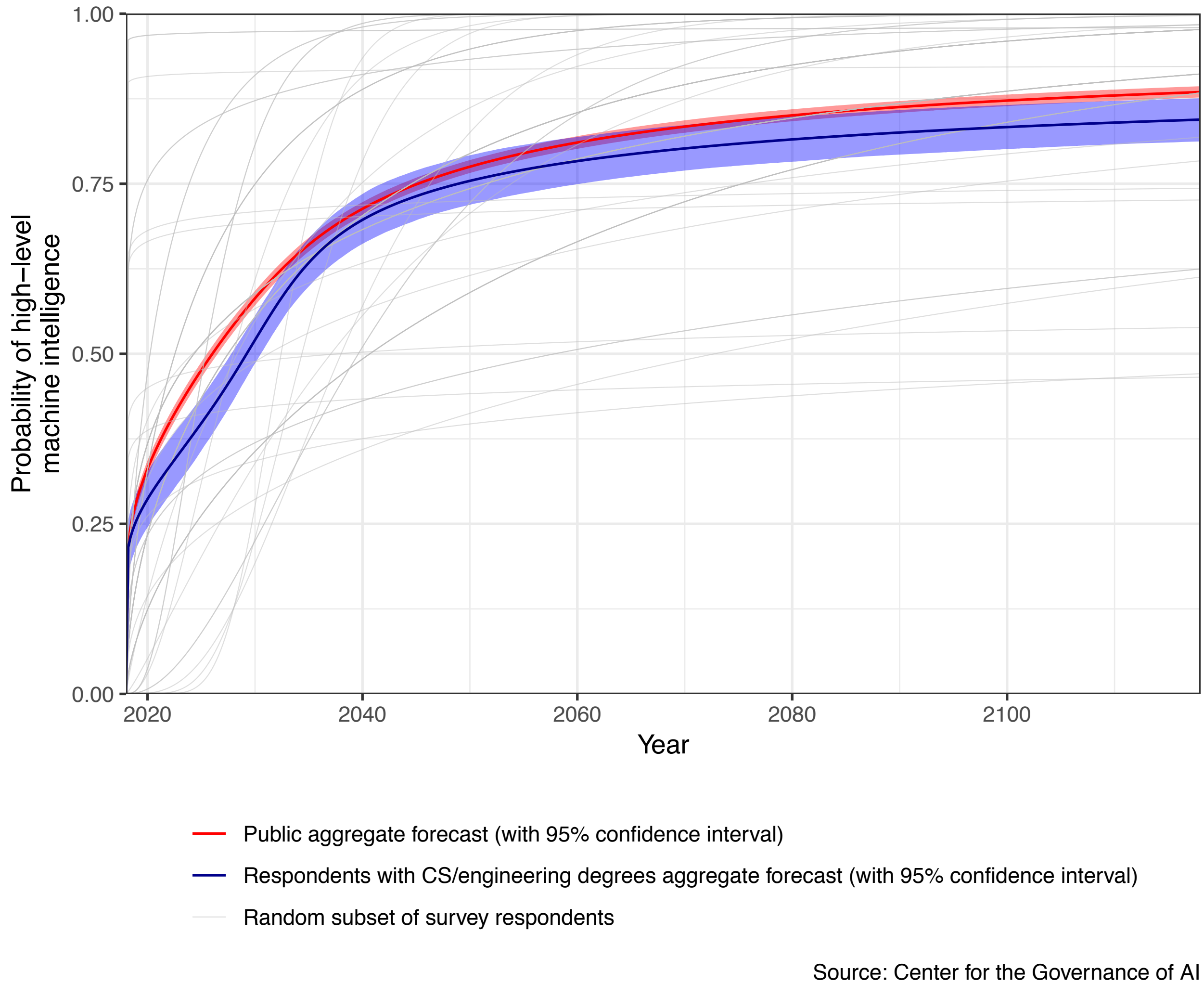


Figure 6.1: The American public's forecasts of high-level machine intelligence timelines

engineering degrees in our sample provide forecasts are more optimistic than those made by experts from Grace et al. (2018); furthermore, their forecasts show considerable overlap with the overall public forecast (see Figure 6.1).

The above differences could be due to different definitions of high-level machine intelligence presented to respondents. However, we suspect that it is not the case for the following reasons. (1) These differences in timelines are larger, more significant than we think could be reasonably attributed to beliefs about these different levels of intelligence. (2) We found similar results using the definition in Grace et al. (2018), on a (different) sample of the American public. In a pilot survey conducted on Mechanical Turk during July 13-14, 2017, we asked American respondents about human-level AI, defined as the following:

> Human-level artificial intelligence (human-level AI) refers to computer systems that can operate with the intelligence of an average human being. These programs can complete tasks or make decisions as successfully as the average human can.

In this pilot study, respondents also provided forecasts that are more optimistic than the projections by AI experts. The respondents predict a median probability of 44% by 2027, a median probability of 62% by 2037, and a median probability of 83% by 2067.

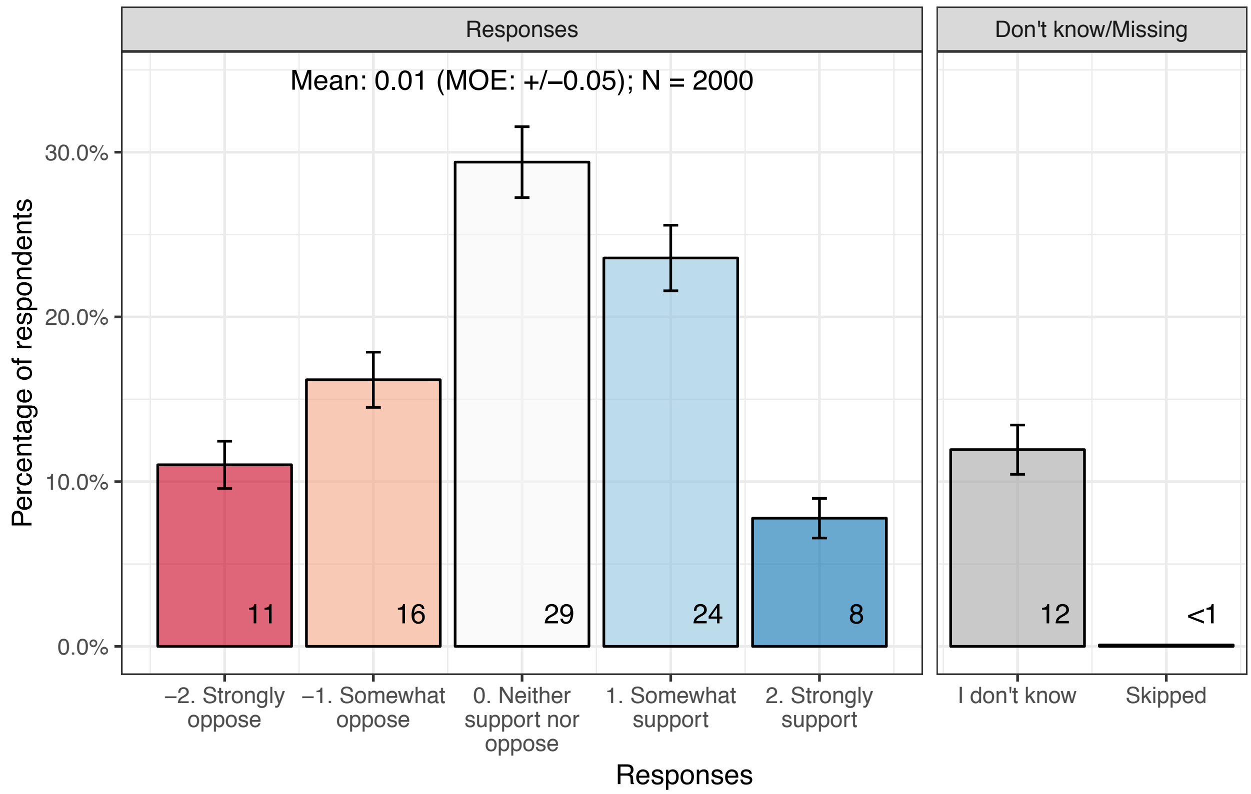


Figure 6.2: Support for developing high-level machine intelligence

## 6.2 Americans express mixed support for developing high-level machine intelligence

Respondents were asked how much they support or oppose the development of high-level machine intelligence. (See Appendix B for the question text.) Americans express mixed support for developing high-level machine intelligence, much like how they feel about developing AI. About one-third of Americans (31%) somewhat or strongly support the development of high-level machine intelligence, while 27% somewhat or strongly oppose it.[15] Many express a neutral attitude: 29% state that they neither support nor oppose, while 12% indicate they don't know.

The correlation between support for developing AI and support for developing high-level machine intelligence is 0.61. The mean level of support for developing high-level machine intelligence, compared with the mean level of support for developing AI, is 0.24 points (MOE = +/- 0.04) lower on a five-point scale (two-sided $p$-value $< 0.001$), according to Table C.31.

## 6.3 High-income Americans, men, and those with tech experience express greater support for high-level machine intelligence

Support for developing high-level machine intelligence varies greatly between demographic subgroups, although only a minority in each subgroup supports developing the technology. Some of the demographic trends we observe regarding support for developing AI also are evident regarding support for high-level machine intelligence. Men (compared with women), high-income Americans (compared with low-income Americans), and those with tech experience (compared with those without) express greater support for high-level machine intelligence.

[15] The discrepancy between this figure and the percentages in Figure 6.2 is due to rounding. According to Table B.129, 7.78% strongly support and 23.58% somewhat support; therefore, 31.36% – rounding to 31% – of respondents either support or somewhat support.

We used a multiple regression that includes all of the demographic variables to predict support for developing high-level machine intelligence. The support for developing AI outcome variable was standardized, so it has mean 0 and unit variance.

Significant predictors correlated with support for developing high-level machine intelligence include:

- Being male (versus being female)
- Identifying as a Republican (versus identifying as an Independent or "other")[16]
- Having a family income of more than $100,000 annually (versus having a family income of less than $30,000 annually)
- Having CS or programming experience (versus not having such experience)

This last result about women less supportive of developing high-level machine intelligence than men is noteworthy as it speaks to the contrary claim sometimes made that it is primarily men who are concerned about the risks from advanced AI. Men are argued to be disproportionately worried about human-level AI because of reasons related to evolutionary psychology (Pinker 2018) or because they have the privilege of not confronting the other harms from AI, such as biased algorithms (Crawford 2016).

We also performed the analysis above but controlling for respondents' support for developing AI (see Appendix). Doing so allows us to identify subgroups those attitudes toward AI diverges from their attitudes toward high-level machine intelligence. In this secondary analysis, we find that being 73 or older is a significant predictor of *support* for developing high-level machine intelligence. In contrast, having a four-year college degree is a significant predictor of *opposition* to developing high-level machine intelligence. These are interesting inversions of the bivariate association, where older and less educated respondents were more concerned about AI; future work could explore this nuance.

## 6.4 The public expects high-level machine intelligence to be more harmful than good

This question sought to quantify respondents' expected outcome of high-level machine intelligence. (See Appendix B for the question text.) Respondents were asked to consider the following:

> Suppose that high-level machine intelligence could be developed one day. How positive or negative do you expect the overall impact of high-level machine intelligence to be on humanity in the long run?

Americans, on average, expect that high-level machine intelligence will have a harmful impact on balance. Overall, 34% think that the technology will have a harmful impact; in particular, 12% said it could be extremely bad, leading to possible human extinction. More than a quarter of Americans think that high-level machine intelligence will be good for humanity, with 5% saying it will be extremely good. Since forecasting the impact of such technology on humanity is highly uncertain, 18% of respondents selected "I don't know." The correlation between one's expected outcome and one's support for developing high-level machine intelligence is 0.69.

A similar question was asked to AI experts in Grace et al. (2018); instead of merely selecting one expected outcome, the AI experts were asked to predict the likelihood of each outcome. In contrast to the general public, the expert respondents think that high-level machine intelligence will be more beneficial than harmful.[17] Although they assign, on average, a 27% probability of high-level machine intelligence of being extremely good for humanity, they also assign, on average, a 9% probability of the technology being extremely bad, including possibly causing human extinction.

[16] In the survey, we allowed those who did not identify as Republican, Democrat, or Independent to select "other." The difference in responses between Republicans and Democrats is not statistically significant at the 5% level. Nevertheless, we caution against over-interpreting these results related to respondents' political identification because the estimated differences are substantively small while the correlating confidence intervals are wide.

[17] To make the two groups' results more comparable, we calculated the expected value of the experts' predicted outcomes so that it is on the same -2 to 2 scale as the public's responses. To calculate this expected value, we averaged the sums of each expert's predicted likelihoods multiplied by the corresponding outcomes; we used the same numerical outcome as described in the previous subsection. The expected value of the experts' predicted outcomes is 0.08, contrasted with the public's average response of -0.17.

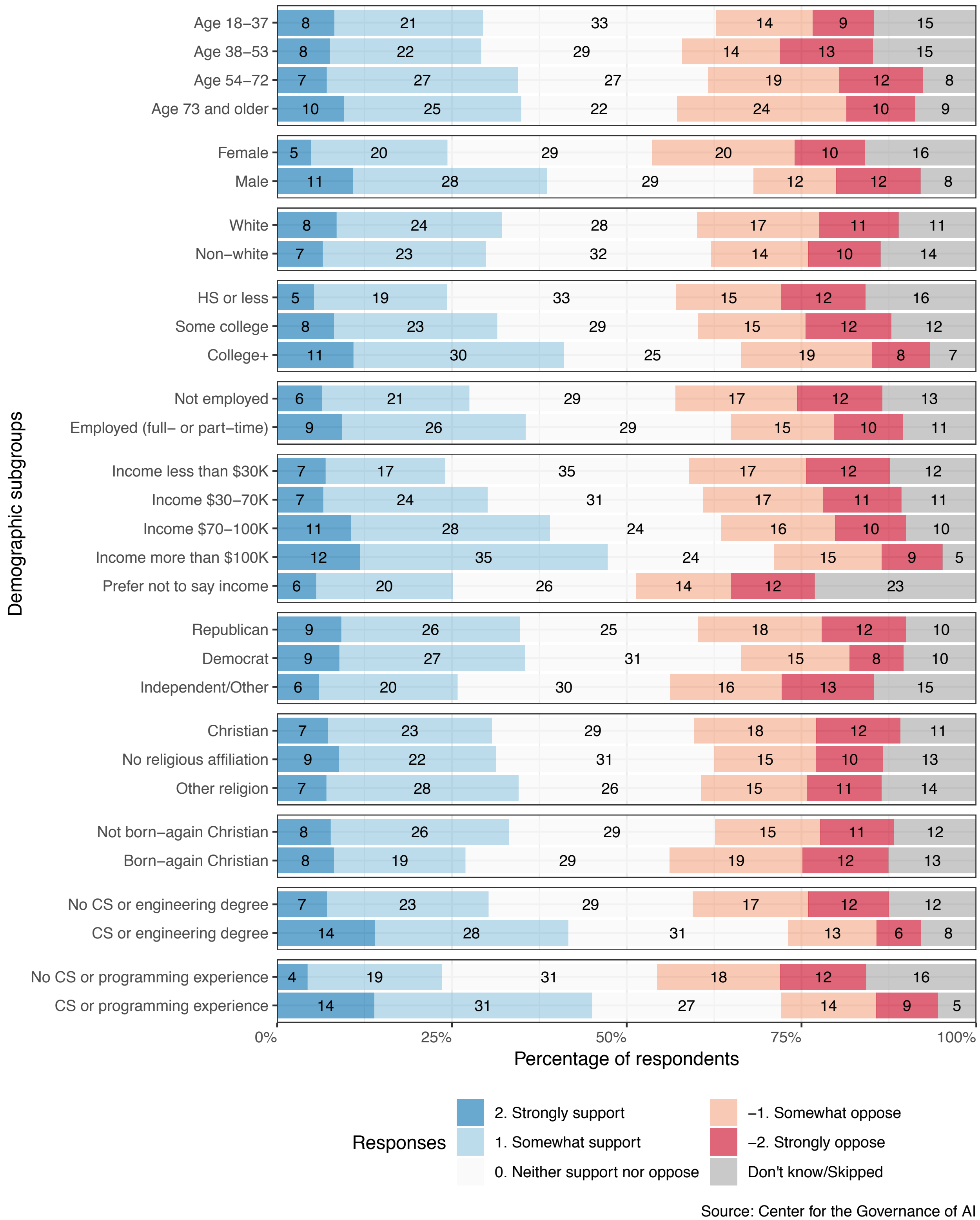


Figure 6.3: Support for developing high-level machine intelligence across demographic characteristics: distribution of responses

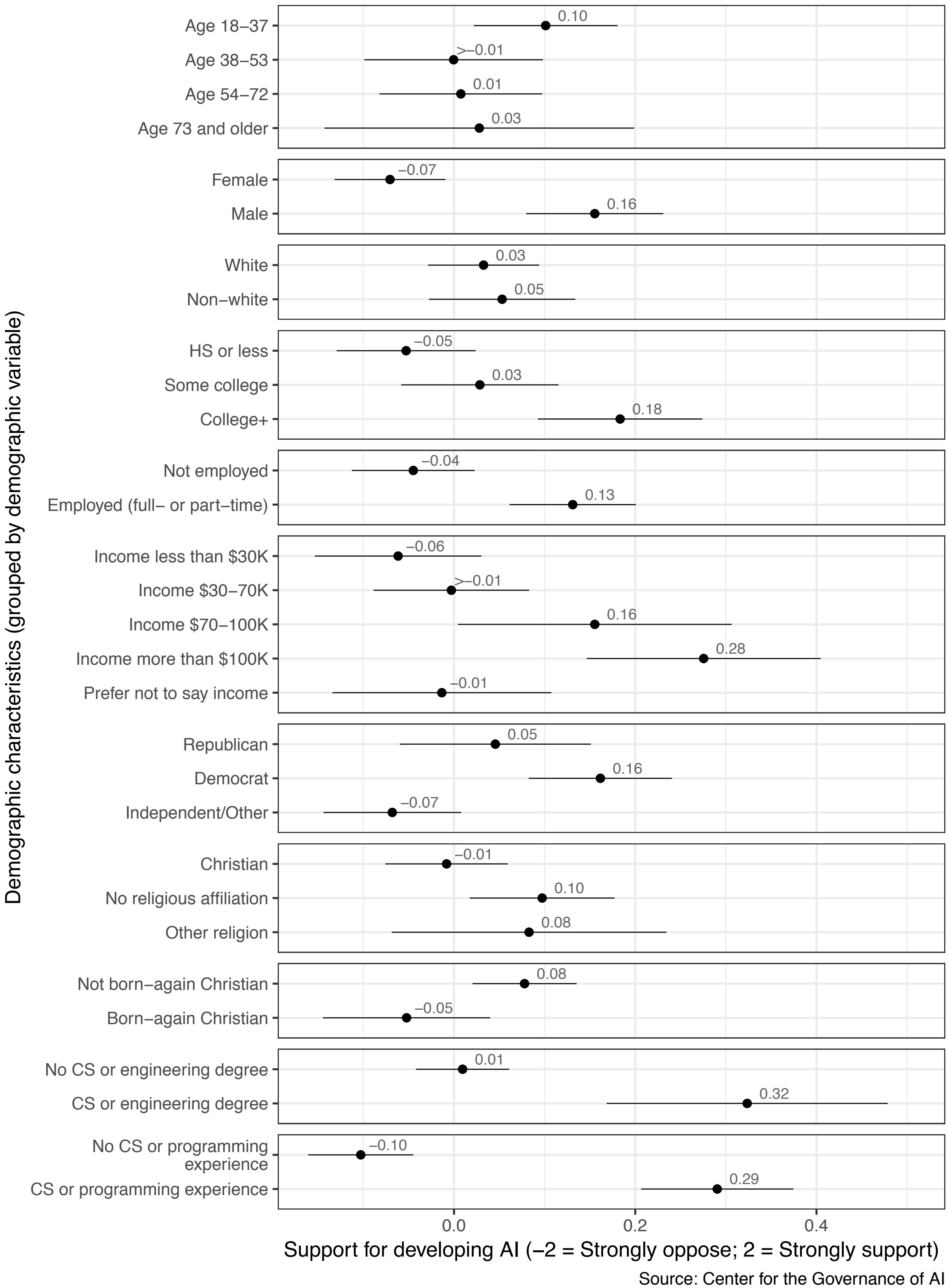


Figure 6.4: Support for developing high-level machine intelligence across demographic characteristics: average support across groups

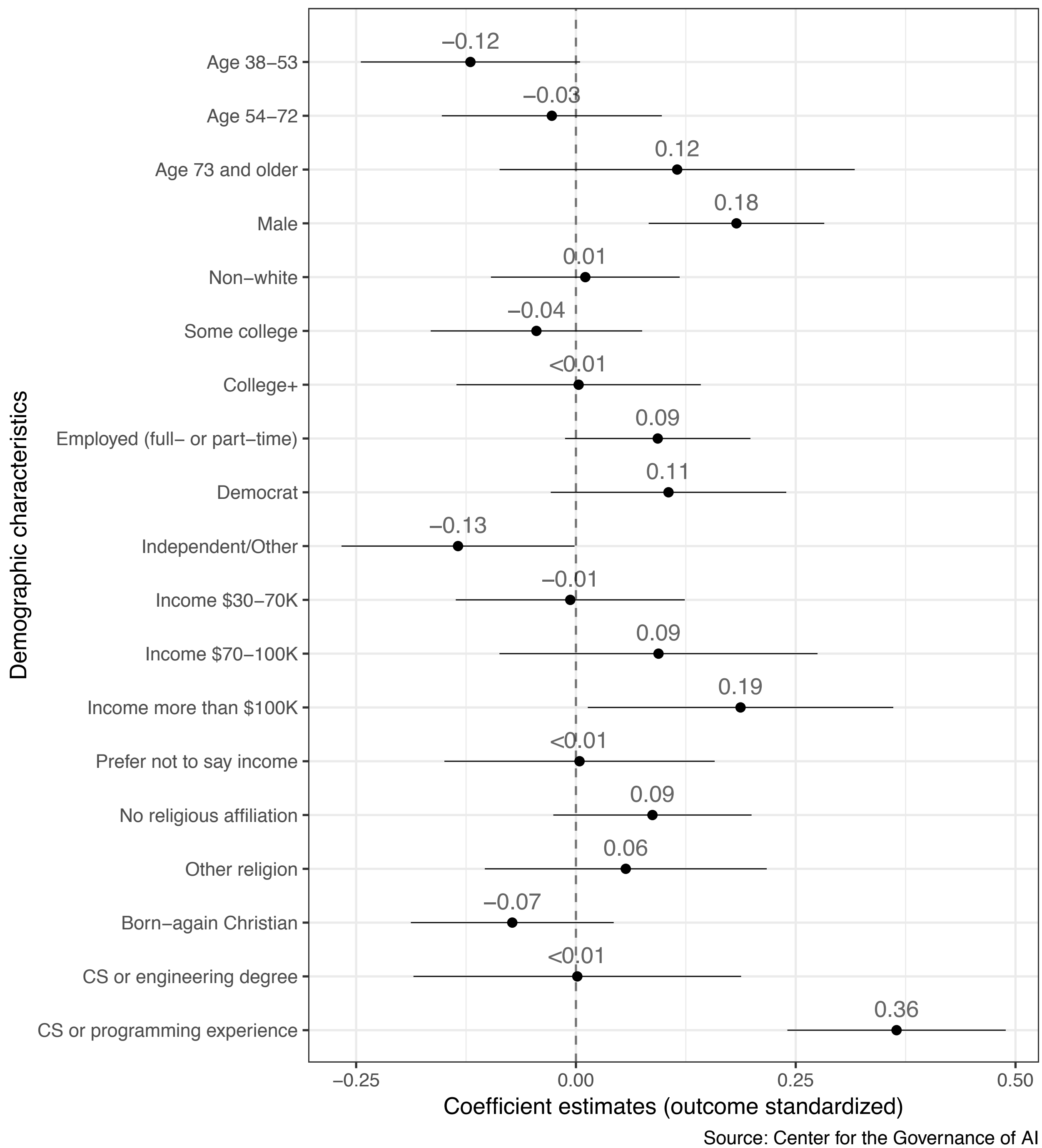


Figure 6.5: Predicting support for developing high-level machine intelligence using demographic characteristics: results from a multiple linear regression that includes all demographic variables

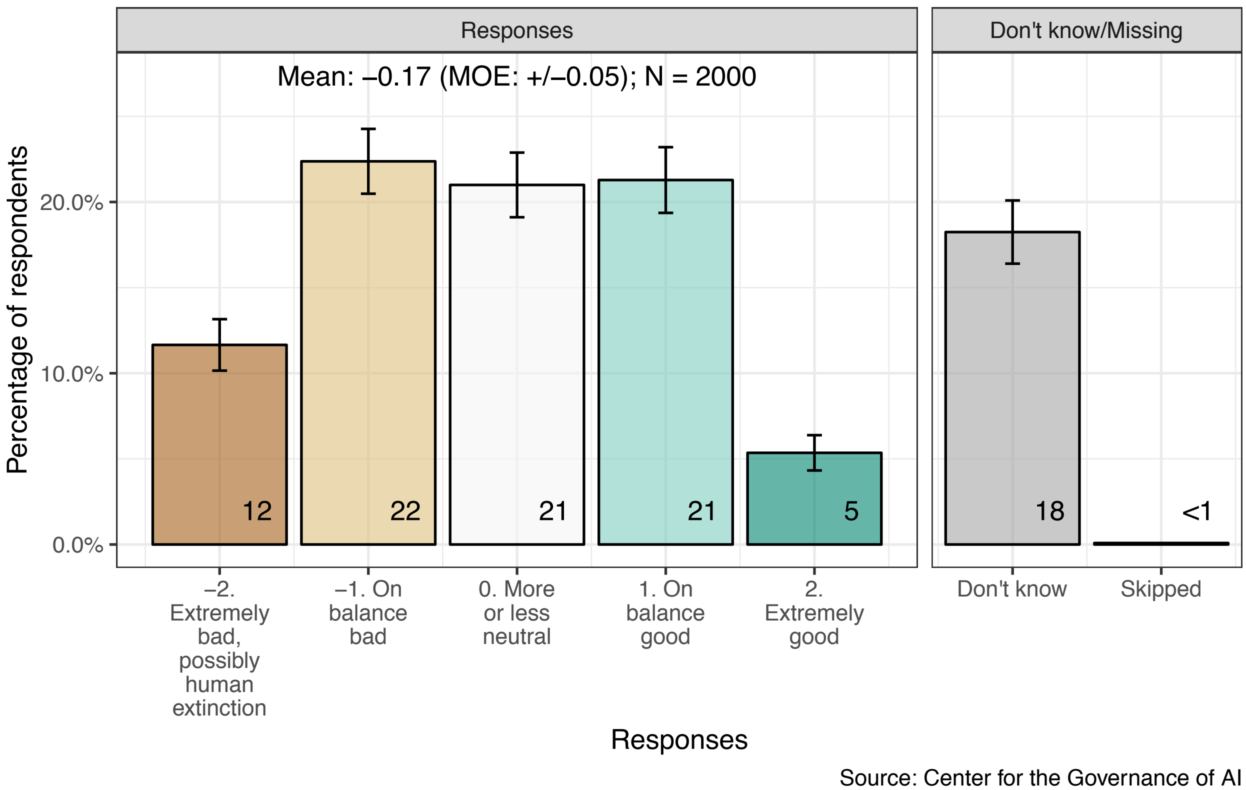


Figure 6.6: Expected positive or negative impact of high-level machine intelligence on humanity

# A Appendix A: Methodology

## A.1 YouGov sampling and weights

YouGov interviewed 2,387 respondents who were then matched down to a sample of 2,000 to produce the final dataset. The respondents were matched to a sampling frame on gender, age, race, and education. The frame was constructed by stratified sampling from the full 2016 American Community Survey (ACS) one-year sample with selection within strata by weighted sampling with replacements (using the person weights on the public use file).

The matched cases were weighted to the sampling frame using propensity scores. The matched cases and the frame were combined and a logistic regression was estimated for inclusion in the frame. The propensity score function included age, gender, race/ethnicity, years of education, and geographic region. The propensity scores were grouped into deciles of the estimated propensity score in the frame and post-stratified according to these deciles.

The weights were then post-stratified on 2016 U.S. presidential vote choice, and a four-way stratification of gender, age (four-categories), race (four-categories), and education (four-categories), to produce the final weight.

## A.2 Demographic subgroups

We use the following demographic subjects in our analysis:

- Age group as defined by Pew Research Center: Millennial/post-Millennial adults (born after 1980; ages 18-37 in 2018), Gen Xers (born 1965-1980; ages 38-53 in 2018), Baby Boomers (born 1946-1964; ages 54-72 in 2018), Silents/Greatest Generation (1945 and earlier; ages 73 and over in 2018)
- Gender: male, female
- Race: white, non-white
- Level of education: graduated from high school or less, some college (including two-year college), graduated from a four-year college or more
- Employment status: employed (full- or part-time), not employed
- Annual household income: less than $30,000, $30,000-70,000, $70,000-100,000, more than $100,000, prefer not to say
- Political party identification: Democrats (includes those who lean Democrat), Republicans (includes those who lean Republican), Independents/Others
- Religion: Christian, follow other religions, non-religious
- Identifies as a born-again Christian: yes, no
- Completed a computer science or engineering degree in undergraduate or graduate school: yes, no
- Has computer science or programming experience: yes, no

We report the unweighted sample sizes of the demographic subgroups in Table A.1.

Table A.1: Size of demographic subgroups

| Demographic subgroups | Unweighted sample sizes |
|---|---:|
| Age 18-37 | 702 |
| Age 38-53 | 506 |
| Age 54-72 | 616 |
| Age 73 and older | 176 |
| Female | 1048 |
| Male | 952 |
| White | 1289 |
| Non-white | 711 |
| HS or less | 742 |
| Some college | 645 |
| College+ | 613 |

| Demographic subgroups | Unweighted sample sizes |
|---|---|
| Not employed | 1036 |
| Employed (full- or part-time) | 964 |
| Income less than $30K | 531 |
| Income $30-70K | 626 |
| Income $70-100K | 240 |
| Income more than $100K | 300 |
| Prefer not to say income | 303 |
| Republican | 470 |
| Democrat | 699 |
| Independent/Other | 831 |
| Christian | 1061 |
| No religious affiliation | 718 |
| Other religion | 221 |
| Not born-again Christian | 1443 |
| Born-again Christian | 557 |
| No CS or engineering degree | 1805 |
| CS or engineering degree | 195 |
| No CS or programming experience | 1265 |
| CS or programming experience | 735 |

## A.3 Analysis

We pre-registered the analysis of this survey on Open Science Framework. Pre-registration increases research transparency by requiring researchers to specify their analysis before analyzing the data (Nosek et al. 2018). Doing so prevents researchers from misusing data analysis to come up with statistically significant results when they do not exist, otherwise known as $p$-hacking.

Unless otherwise specified, we performed the following procedure:

- Survey weights provided by YouGov were used in our primary analysis. For transparency, Appendix B contains the unweighted topline results, including raw frequencies.
- For estimates of summary statistics or coefficients, "don't know" or missing responses were re-coded to the weighted overall mean, unconditional on treatment conditions. Almost all questions had a "don't know" option. If more than 10% of the variable's values were don't know" or missing, we included a (standardized) dummy variable for "don't know"/missing in the analysis. For survey experiment questions, we compared "don't know"/missing rates across experimental conditions. Our decision was informed by the Standard Operating Procedures for Don Green's Lab at Columbia University (Lin and Green 2016).
- Heteroscedasticity-consistent standard errors were used to generate the margins of error at the 95% confidence level. We report cluster-robust standard errors whenever there is clustering by respondent. In figures, each error bar shows the 95% confidence intervals. Each confidence ellipse shows the 95% confidence region of the bivariate means assuming the two variables are distributed multivariate normal.
- In regression tables, * denotes $p < 0.05$, ** denotes $p < 0.01$, and *** denotes $p < 0.001$.

## A.4 Data sharing

We plan to make our survey data, as well as the R and Markdown code that produced this report, publicly available through the Harvard Dataverse six months after the publication of this report.

# B Appendix B: Topline questionnaire

Below, we present the survey text as shown to respondents. The numerical codings are shown in parentheses following each answer choice.

In addition, we report the topline results: percentages weighted to be representative of the U.S. adult population, the unweighted raw percentages, and the raw frequencies. Note that in all survey experiments, respondents were randomly assigned to each experimental group with equal probability.

## B.1 Global risks

[All respondents were presented with the following prompt.]

We want to get your opinion about global risks. A "global risk" is an uncertain event or condition that, if it happens, could cause a significant negative impact for at least 10 percent of the world's population. That is at least 1 in 10 people around the world could experience a significant negative impact.

You will be asked to consider 5 potential global risks.

[Respondents were presented with five items randomly selected from the list below. One item was shown at a time.]

- **Failure to address climate change**: Continued failure of governments and businesses to pass effective measures to reduce climate change, protect people, and help those impacted by climate change to adapt.
- **Failure of regional or global governance**: Regional organizations (e.g., the European Union) or global organizations (e.g., the United Nations) are unable to resolve issues of economic, political, or environmental importance.
- **Conflict between major countries**: Disputes between major countries that lead to economic, military, cyber, or societal conflicts.
- **Weapons of mass destruction**: Use of nuclear, chemical, biological or radiological weapons, creating international crises and killing large numbers of people.
- **Large-scale involuntary migration**: Large-scale involuntary movement of people, such as refugees, caused by conflict, disasters, environmental or economic reasons.
- **Rapid and massive spread of infectious diseases**: The uncontrolled spread of infectious diseases, for instance as a result of resistance to antibiotics, that leads to widespread deaths and economic disruptions.
- **Water crises**: A large decline in the available quality and quantity of fresh water that harms human health and economic activity.
- **Food crises**: Large numbers of people are unable to buy or access food. Harmful consequences of artificial intelligence (AI): Intended or unintended consequences of artificial intelligence that causes widespread harm to humans, the economy, and the environment.
- **Harmful consequences of synthetic biology**: Intended or unintended consequences of synthetic biology, such as genetic engineering, that causes widespread harm to humans, the economy, and the environment.
- **Large-scale cyber attacks**: Large-scale cyber attacks that cause large economic damages, tensions between countries, and widespread loss of trust in the internet.
- **Large-scale terrorist attacks**: Individuals or non-government groups with political or religious goals that cause large numbers of deaths and major material damage.
- **Global recession**: Economic decline in several major countries that leads to a decrease in income and high unemployment.
- **Extreme weather events**: Extreme weather events that cause large numbers of deaths as well as damage to property, infrastructure, and the environment.
- **Major natural disasters**: Earthquakes, volcanic activity, landslides, tsunamis, or geomagnetic storms that cause large numbers of deaths as well as damage to property, infrastructure, and the environment.

QUESTION:

**What is the likelihood of [INSERT GLOBAL RISK] happening globally within the next 10 years?** Please use the slider to indicate your answer. 0% chance means it will certainly not happen and 100% chance means it will certainly happen.

ANSWER CHOICES:[18]

- Very unlikely: less than 5% chance (2.5%)
- Unlikely: 5-20% chance (12.5%)
- Somewhat unlikely: 20-40% chance (30%)
- Equally likely as unlikely: 40-60% chance (50%)
- Somewhat likely: 60-80% chance (70%)
- Likely: 80-95% chance (87.5%)
- Very likely: more than 95% chance (97.5%)
- I don't know

QUESTION:

If [INSERT GLOBAL RISK] were to happen, what would be the size of the negative impact for several countries or industries within the next 10 years?

ANSWER CHOICES:

- Minimal (0)
- Minor (1)
- Moderate (2)
- Severe (3)
- Catastrophic (4)
- I don't know

Table B.1: Likelihood - Failure to address climate change; $N = 666$

| Responses | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 10.53 | 10.21 | 68 |
| Unlikely 5-20% | 6.87 | 6.46 | 43 |
| Somewhat unlikely 20-40% | 11.61 | 11.41 | 76 |
| Equally likely as unlikely 40-60% | 18.44 | 18.62 | 124 |
| Somewhat likely 60-80% | 15.81 | 15.77 | 105 |
| Likely 80-95% | 13.47 | 13.81 | 92 |
| Very likely > 95% | 16.00 | 16.37 | 109 |
| I don't know | 7.17 | 7.21 | 48 |
| Skipped | 0.10 | 0.15 | 1 |

Table B.2: Likelihood - Failure of regional/global governance; $N = 652$

| Responses | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 5.40 | 5.52 | 36 |
| Unlikely 5-20% | 7.99 | 7.98 | 52 |
| Somewhat unlikely 20-40% | 12.14 | 12.42 | 81 |
| Equally likely as unlikely 40-60% | 24.71 | 24.39 | 159 |
| Somewhat likely 60-80% | 17.80 | 18.10 | 118 |
| Likely 80-95% | 11.54 | 11.96 | 78 |
| Very likely > 95% | 8.86 | 9.51 | 62 |
| I don't know | 10.96 | 9.66 | 63 |
| Skipped | 0.58 | 0.46 | 3 |

[18] For this and other questions that ask respondents about likelihoods, each multiple-choice answer was coded to the mean value across the probabilities in the answer's range.

Table B.3: Likelihood - Conflict between major countries; $N = 625$

| Responses | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 3.47 | 3.36 | 21 |
| Unlikely 5-20% | 6.45 | 7.04 | 44 |
| Somewhat unlikely 20-40% | 10.68 | 10.40 | 65 |
| Equally likely as unlikely 40-60% | 22.16 | 20.64 | 129 |
| Somewhat likely 60-80% | 22.46 | 23.36 | 146 |
| Likely 80-95% | 13.92 | 14.24 | 89 |
| Very likely > 95% | 12.21 | 12.80 | 80 |
| I don't know | 8.49 | 8.00 | 50 |
| Skipped | 0.16 | 0.16 | 1 |

Table B.4: Likelihood - Weapons of mass destruction; $N = 645$

| Responses | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 7.05 | 6.67 | 43 |
| Unlikely 5-20% | 13.71 | 13.80 | 89 |
| Somewhat unlikely 20-40% | 15.19 | 15.04 | 97 |
| Equally likely as unlikely 40-60% | 24.33 | 24.19 | 156 |
| Somewhat likely 60-80% | 17.15 | 17.36 | 112 |
| Likely 80-95% | 9.26 | 9.15 | 59 |
| Very likely > 95% | 6.44 | 6.98 | 45 |
| I don't know | 6.87 | 6.82 | 44 |
| Skipped | 0 | 0 | 0 |

Table B.5: Likelihood - Large-scale involuntary migration; $N = 628$

| Responses | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 6.70 | 6.53 | 41 |
| Unlikely 5-20% | 7.83 | 7.32 | 46 |
| Somewhat unlikely 20-40% | 11.57 | 11.62 | 73 |
| Equally likely as unlikely 40-60% | 18.65 | 18.31 | 115 |
| Somewhat likely 60-80% | 20.91 | 21.34 | 134 |
| Likely 80-95% | 13.63 | 14.01 | 88 |
| Very likely > 95% | 12.31 | 13.06 | 82 |
| I don't know | 8.27 | 7.64 | 48 |
| Skipped | 0.12 | 0.16 | 1 |

Table B.6: Likelihood - Spread of infectious diseases; $N = 620$

| Responses | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 4.76 | 4.03 | 25 |
| Unlikely 5-20% | 13.12 | 13.06 | 81 |
| Somewhat unlikely 20-40% | 17.24 | 17.58 | 109 |
| Equally likely as unlikely 40-60% | 22.76 | 23.39 | 145 |
| Somewhat likely 60-80% | 17.55 | 17.58 | 109 |
| Likely 80-95% | 10.07 | 10.00 | 62 |
| Very likely > 95% | 6.94 | 6.94 | 43 |

| Responses | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| I don't know | 7.46 | 7.26 | 45 |
| Skipped | 0.12 | 0.16 | 1 |

Table B.7: Likelihood - Water crises; $N = 623$

| Responses | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 6.37 | 6.10 | 38 |
| Unlikely 5-20% | 9.71 | 10.43 | 65 |
| Somewhat unlikely 20-40% | 13.22 | 13.64 | 85 |
| Equally likely as unlikely 40-60% | 21.23 | 21.03 | 131 |
| Somewhat likely 60-80% | 20.26 | 19.26 | 120 |
| Likely 80-95% | 11.04 | 10.91 | 68 |
| Very likely > 95% | 10.83 | 11.72 | 73 |
| I don't know | 7.33 | 6.90 | 43 |
| Skipped | 0 | 0 | 0 |

Table B.8: Likelihood - Food crises; $N = 1073$

| Responses | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 6.29 | 5.96 | 64 |
| Unlikely 5-20% | 12.53 | 11.65 | 125 |
| Somewhat unlikely 20-40% | 14.49 | 14.82 | 159 |
| Equally likely as unlikely 40-60% | 22.53 | 22.55 | 242 |
| Somewhat likely 60-80% | 16.90 | 17.24 | 185 |
| Likely 80-95% | 10.46 | 10.90 | 117 |
| Very likely > 95% | 9.38 | 10.07 | 108 |
| I don't know | 7.36 | 6.71 | 72 |
| Skipped | 0.08 | 0.09 | 1 |

Table B.9: Likelihood - Harmful consequences of AI; $N = 573$

| Responses | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 11.26 | 11.34 | 65 |
| Unlikely 5-20% | 16.43 | 16.06 | 92 |
| Somewhat unlikely 20-40% | 15.95 | 15.53 | 89 |
| Equally likely as unlikely 40-60% | 19.36 | 20.07 | 115 |
| Somewhat likely 60-80% | 11.56 | 11.34 | 65 |
| Likely 80-95% | 8.30 | 8.03 | 46 |
| Very likely > 95% | 7.71 | 7.85 | 45 |
| I don't know | 9.43 | 9.77 | 56 |
| Skipped | 0 | 0 | 0 |

Table B.10: Likelihood - Harmful consequences of synthetic biology; $N$ = 630

| Responses | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 9.92 | 9.68 | 61 |
| Unlikely 5-20% | 15.66 | 15.08 | 95 |
| Somewhat unlikely 20-40% | 15.06 | 15.24 | 96 |
| Equally likely as unlikely 40-60% | 23.48 | 22.86 | 144 |
| Somewhat likely 60-80% | 12.32 | 12.86 | 81 |
| Likely 80-95% | 7.47 | 7.62 | 48 |
| Very likely > 95% | 6.04 | 6.19 | 39 |
| I don't know | 10.06 | 10.48 | 66 |
| Skipped | 0 | 0 | 0 |

Table B.11: Likelihood - Cyber attacks; $N$ = 650

| Responses | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 2.04 | 2.15 | 14 |
| Unlikely 5-20% | 4.28 | 3.69 | 24 |
| Somewhat unlikely 20-40% | 7.74 | 7.85 | 51 |
| Equally likely as unlikely 40-60% | 15.78 | 16.15 | 105 |
| Somewhat likely 60-80% | 22.66 | 21.85 | 142 |
| Likely 80-95% | 16.44 | 16.62 | 108 |
| Very likely > 95% | 22.40 | 23.54 | 153 |
| I don't know | 8.53 | 8.00 | 52 |
| Skipped | 0.12 | 0.15 | 1 |

Table B.12: Likelihood - Terrorist attacks; $N$ = 635

| Responses | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 5.21 | 4.88 | 31 |
| Unlikely 5-20% | 4.53 | 4.88 | 31 |
| Somewhat unlikely 20-40% | 12.43 | 11.81 | 75 |
| Equally likely as unlikely 40-60% | 19.47 | 19.21 | 122 |
| Somewhat likely 60-80% | 22.28 | 22.52 | 143 |
| Likely 80-95% | 15.74 | 15.43 | 98 |
| Very likely > 95% | 12.45 | 12.91 | 82 |
| I don't know | 7.89 | 8.35 | 53 |
| Skipped | 0 | 0 | 0 |

Table B.13: Likelihood - Global recession; $N$ = 599

| Responses | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 4.17 | 3.67 | 22 |
| Unlikely 5-20% | 7.34 | 7.18 | 43 |
| Somewhat unlikely 20-40% | 12.68 | 12.85 | 77 |
| Equally likely as unlikely 40-60% | 23.43 | 24.21 | 145 |
| Somewhat likely 60-80% | 23.83 | 23.04 | 138 |
| Likely 80-95% | 10.80 | 10.85 | 65 |

| Responses | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very likely > 95% | 8.34 | 8.68 | 52 |
| I don't know | 9.41 | 9.52 | 57 |
| Skipped | 0 | 0 | 0 |

Table B.14: Likelihood - Extreme weather events; $N$ = 613

| Responses | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 3.52 | 3.10 | 19 |
| Unlikely 5-20% | 5.64 | 5.22 | 32 |
| Somewhat unlikely 20-40% | 8.77 | 8.81 | 54 |
| Equally likely as unlikely 40-60% | 20.12 | 18.76 | 115 |
| Somewhat likely 60-80% | 18.09 | 18.27 | 112 |
| Likely 80-95% | 13.02 | 14.03 | 86 |
| Very likely > 95% | 24.95 | 25.45 | 156 |
| I don't know | 5.89 | 6.36 | 39 |
| Skipped | 0 | 0 | 0 |

Table B.15: Likelihood - Natural disasters; $N$ = 637

| Responses | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 2.47 | 2.51 | 16 |
| Unlikely 5-20% | 4.10 | 4.08 | 26 |
| Somewhat unlikely 20-40% | 7.32 | 7.06 | 45 |
| Equally likely as unlikely 40-60% | 17.63 | 17.74 | 113 |
| Somewhat likely 60-80% | 19.43 | 19.15 | 122 |
| Likely 80-95% | 18.12 | 18.05 | 115 |
| Very likely > 95% | 25.73 | 26.37 | 168 |
| I don't know | 5.21 | 5.02 | 32 |
| Skipped | 0 | 0 | 0 |

Table B.16: Size of negative impact - Failure to address climate change; $N$ = 666

| Responses | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Minimal | 13.46 | 13.96 | 93 |
| Minor | 11.26 | 10.96 | 73 |
| Moderate | 23.37 | 23.27 | 155 |
| Severe | 28.41 | 28.08 | 187 |
| Catastrophic | 14.26 | 14.56 | 97 |
| I don't know | 9.13 | 9.01 | 60 |
| Skipped | 0.10 | 0.15 | 1 |

Table B.17: Size of negative impact - Failure of regional/global governance; $N$ = 652

| Responses | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Minimal | 6.04 | 5.98 | 39 |
| Minor | 6.09 | 5.67 | 37 |
| Moderate | 28.68 | 28.99 | 189 |
| Severe | 33.21 | 34.05 | 222 |
| Catastrophic | 10.76 | 10.89 | 71 |
| I don't know | 15.12 | 14.26 | 93 |
| Skipped | 0.10 | 0.15 | 1 |

Table B.18: Size of negative impact - Conflict between major countries; $N$ = 625

| Responses | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Minimal | 1.18 | 0.96 | 6 |
| Minor | 4.94 | 4.80 | 30 |
| Moderate | 28.81 | 28.16 | 176 |
| Severe | 38.23 | 38.56 | 241 |
| Catastrophic | 14.80 | 16.00 | 100 |
| I don't know | 11.89 | 11.36 | 71 |
| Skipped | 0.14 | 0.16 | 1 |

Table B.19: Size of negative impact - Weapons of mass destruction; $N$ = 645

| Responses | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Minimal | 2.28 | 2.17 | 14 |
| Minor | 4.99 | 4.19 | 27 |
| Moderate | 13.57 | 13.49 | 87 |
| Severe | 31.05 | 31.01 | 200 |
| Catastrophic | 38.06 | 39.38 | 254 |
| I don't know | 10.05 | 9.77 | 63 |
| Skipped | 0 | 0 | 0 |

Table B.20: Size of negative impact - Large-scale involuntary migration; $N$ = 628

| Responses | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Minimal | 2.07 | 2.07 | 13 |
| Minor | 8.67 | 8.28 | 52 |
| Moderate | 25.63 | 25.96 | 163 |
| Severe | 35.31 | 36.15 | 227 |
| Catastrophic | 18.14 | 17.83 | 112 |
| I don't know | 9.99 | 9.55 | 60 |
| Skipped | 0.19 | 0.16 | 1 |

Table B.21: Size of negative impact - Spread of infectious diseases; $N$ = 620

| Responses | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Minimal | 2.72 | 2.58 | 16 |
| Minor | 6.03 | 5.65 | 35 |
| Moderate | 26.86 | 28.06 | 174 |
| Severe | 32.00 | 32.58 | 202 |
| Catastrophic | 20.50 | 20.48 | 127 |
| I don't know | 11.88 | 10.65 | 66 |
| Skipped | 0 | 0 | 0 |

Table B.22: Size of negative impact - Water crises; $N$ = 623

| Responses | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Minimal | 1.72 | 1.93 | 12 |
| Minor | 4.42 | 4.65 | 29 |
| Moderate | 19.92 | 19.42 | 121 |
| Severe | 36.71 | 36.44 | 227 |
| Catastrophic | 27.24 | 28.25 | 176 |
| I don't know | 10.00 | 9.31 | 58 |
| Skipped | 0 | 0 | 0 |

Table B.23: Size of negative impact - Food crises; $N$ = 1073

| Responses | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Minimal | 2.55 | 2.61 | 28 |
| Minor | 7.22 | 6.99 | 75 |
| Moderate | 22.81 | 22.37 | 240 |
| Severe | 33.93 | 34.67 | 372 |
| Catastrophic | 24.04 | 24.88 | 267 |
| I don't know | 9.38 | 8.39 | 90 |
| Skipped | 0.08 | 0.09 | 1 |

Table B.24: Size of negative impact - Harmful consequences of AI; $N$ = 573

| Responses | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Minimal | 7.54 | 7.50 | 43 |
| Minor | 14.82 | 13.79 | 79 |
| Moderate | 27.77 | 27.92 | 160 |
| Severe | 20.46 | 21.82 | 125 |
| Catastrophic | 14.62 | 14.31 | 82 |
| I don't know | 14.79 | 14.66 | 84 |
| Skipped | 0 | 0 | 0 |

Table B.25: Size of negative impact - Harmful consequences of synthetic biology; $N = 630$

| Responses | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Minimal | 6.77 | 6.67 | 42 |
| Minor | 11.95 | 11.59 | 73 |
| Moderate | 28.40 | 27.94 | 176 |
| Severe | 26.03 | 26.03 | 164 |
| Catastrophic | 11.15 | 11.90 | 75 |
| I don't know | 15.70 | 15.87 | 100 |
| Skipped | 0 | 0 | 0 |

Table B.26: Size of negative impact - Cyber attacks; $N = 650$

| Responses | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Minimal | 1.19 | 1.23 | 8 |
| Minor | 4.46 | 4.46 | 29 |
| Moderate | 21.43 | 21.23 | 138 |
| Severe | 38.26 | 37.69 | 245 |
| Catastrophic | 23.01 | 24.46 | 159 |
| I don't know | 11.66 | 10.92 | 71 |
| Skipped | 0 | 0 | 0 |

Table B.27: Size of negative impact - Terrorist attacks; $N = 635$

| Responses | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Minimal | 2.61 | 2.68 | 17 |
| Minor | 6.11 | 6.14 | 39 |
| Moderate | 29.29 | 29.45 | 187 |
| Severe | 33.69 | 33.70 | 214 |
| Catastrophic | 15.97 | 15.91 | 101 |
| I don't know | 12.32 | 12.13 | 77 |
| Skipped | 0 | 0 | 0 |

Table B.28: Size of negative impact - Global recession; $N = 599$

| Responses | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Minimal | 2.71 | 2.67 | 16 |
| Minor | 5.94 | 5.68 | 34 |
| Moderate | 29.89 | 29.72 | 178 |
| Severe | 35.49 | 36.23 | 217 |
| Catastrophic | 14.63 | 14.52 | 87 |
| I don't know | 11.35 | 11.19 | 67 |
| Skipped | 0 | 0 | 0 |

Table B.29: Size of negative impact - Extreme weather events; $N = 613$

| Responses | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Minimal | 2.54 | 2.45 | 15 |
| Minor | 6.69 | 6.53 | 40 |
| Moderate | 25.94 | 26.43 | 162 |
| Severe | 32.50 | 31.97 | 196 |
| Catastrophic | 22.79 | 23.00 | 141 |
| I don't know | 9.56 | 9.62 | 59 |
| Skipped | 0 | 0 | 0 |

Table B.30: Size of negative impact - Natural disasters; $N = 637$

| Responses | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Minimal | 1.29 | 1.26 | 8 |
| Minor | 5.86 | 5.81 | 37 |
| Moderate | 22.26 | 23.08 | 147 |
| Severe | 36.41 | 36.11 | 230 |
| Catastrophic | 27.47 | 27.32 | 174 |
| I don't know | 6.72 | 6.44 | 41 |
| Skipped | 0 | 0 | 0 |

## B.2 Survey experiment: what the public considers AI, automation, machine learning, and robotics

[Respondents were randomly assigned to one of the four questions. The order of answer choices was randomized, except that "None of the above" was always shown last.]

QUESTIONS:

- In your opinion, which of the following technologies, if any, uses **artificial intelligence** (AI)? Select all the apply.
- In your opinion, which of the following technologies, if any, uses **automation**? Select all that apply.
- In your opinion, which of the following technologies, if any, uses **machine learning**? Select all that apply.
- In your opinion, which of the following technologies, if any, uses **robotics**? Select all that apply.

ANSWER CHOICES:

- Virtual assistants (e.g., Siri, Google Assistant, Amazon Alexa)
- Smart speakers (e.g., Amazon Echo, Google Home, Apple Homepod)
- Facebook photo tagging
- Google Search
- Recommendations for Netflix movies or Amazon ebooks
- Google Translate
- Driverless cars and trucks
- Social robots that can interact with humans
- Industrial robots used in manufacturing
- Drones that do not require a human controller
- None of the above

Table B.31: Artificial intelligence (AI); $N = 493$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|

| | | | |
|---|---|---|---|
| Virtual assistants (e.g., Siri, Google Assistant, Amazon Alexa) | 62.87 | 64.30 | 317 |
| Smart speakers (e.g., Amazon Echo, Google Home, Apple Homepod) | 55.46 | 56.19 | 277 |
| Facebook photo tagging | 36.16 | 36.51 | 180 |
| Google Search | 35.59 | 36.51 | 180 |
| Recommendations for Netflix movies or Amazon ebooks | 27.73 | 29.01 | 143 |
| Google Translate | 29.49 | 30.02 | 148 |
| Driverless cars and trucks | 56.38 | 57.20 | 282 |
| Social robots that can interact with humans | 63.63 | 64.10 | 316 |
| Industrial robots used in manufacturing | 40.11 | 40.16 | 198 |
| Drones that do not require a human controller | 53.48 | 52.74 | 260 |

Table B.32: Automation; $N = 513$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Virtual assistants (e.g., Siri, Google Assistant, Amazon Alexa) | 66.75 | 67.06 | 344 |
| Smart speakers (e.g., Amazon Echo, Google Home, Apple Homepod) | 60.81 | 61.01 | 313 |
| Facebook photo tagging | 43.74 | 45.42 | 233 |
| Google Search | 52.12 | 53.80 | 276 |
| Recommendations for Netflix movies or Amazon ebooks | 45.13 | 46.39 | 238 |
| Google Translate | 45.06 | 46.39 | 238 |
| Driverless cars and trucks | 68.16 | 68.62 | 352 |
| Social robots that can interact with humans | 64.00 | 64.72 | 332 |
| Industrial robots used in manufacturing | 64.70 | 65.11 | 334 |
| Drones that do not require a human controller | 65.04 | 65.69 | 337 |

Table B.33: Machine learning; $N = 508$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Virtual assistants (e.g., Siri, Google Assistant, Amazon Alexa) | 59.10 | 60.43 | 307 |
| Smart speakers (e.g., Amazon Echo, Google Home, Apple Homepod) | 46.70 | 46.65 | 237 |

| | | | |
|---|---|---|---|
| Facebook photo tagging | 35.37 | 36.81 | 187 |
| Google Search | 45.42 | 46.26 | 235 |
| Recommendations for Netflix movies or Amazon ebooks | 37.97 | 38.19 | 194 |
| Google Translate | 33.40 | 34.06 | 173 |
| Driverless cars and trucks | 52.96 | 54.33 | 276 |
| Social robots that can interact with humans | 59.19 | 59.45 | 302 |
| Industrial robots used in manufacturing | 37.41 | 37.80 | 192 |
| Drones that do not require a human controller | 49.03 | 49.41 | 251 |

Table B.34: Robotics; $N = 486$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Virtual assistants (e.g., Siri, Google Assistant, Amazon Alexa) | 45.27 | 46.30 | 225 |
| Smart speakers (e.g., Amazon Echo, Google Home, Apple Homepod) | 35.59 | 36.83 | 179 |
| Facebook photo tagging | 21.00 | 21.40 | 104 |
| Google Search | 22.07 | 23.25 | 113 |
| Recommendations for Netflix movies or Amazon ebooks | 17.84 | 18.31 | 89 |
| Google Translate | 20.30 | 21.19 | 103 |
| Driverless cars and trucks | 60.26 | 61.93 | 301 |
| Social robots that can interact with humans | 61.89 | 63.17 | 307 |
| Industrial robots used in manufacturing | 67.99 | 69.75 | 339 |
| Drones that do not require a human controller | 57.55 | 59.05 | 287 |

## B.3 Knowledge of computer science (CS)/technology

QUESTION:

What is your knowledge of computer science/technology? (Select all that apply.)

ANSWER CHOICES:

- I have taken at least one college-level course in computer science.
- I have a computer science or engineering undergraduate degree.
- I have a graduate degree in computer science or engineering.
- I have programming experience.
- I don't have any of the educational or work experiences described above.

Table B.35: Computer science/technology background; $N = 2000$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Took at least one college-level course in CS | 24.73 | 25.05 | 501 |
| CS or engineering undergraduate degree | 7.12 | 7.30 | 146 |
| CS or engineering graduate degree | 3.85 | 3.75 | 75 |
| Have programming experience | 10.88 | 11.10 | 222 |
| None of the above | 63.68 | 63.20 | 1264 |

## B.4 Support for developing AI

[All respondents were presented with the following prompt.]

Next, we would like to ask you questions about your attitudes toward artificial intelligence.

Artificial Intelligence (AI) refers to computer systems that perform tasks or make decisions that usually require human intelligence. AI can perform these tasks or make these decisions without explicit human instructions. Today, AI has been used in the following applications:

[Respondents were shown five items randomly selected from the list below.]

- Translate over 100 different languages
- Predict one's Google searches
- Identify people from their photos
- Diagnose diseases like skin cancer and common illnesses
- Predict who are at risk of various diseases
- Help run factories and warehouses
- Block spam email
- Play computer games
- Help conduct legal case research
- Categorize photos and videos
- Detect plagiarism in essays
- Spot abusive messages on social media
- Predict what one is likely to buy online
- Predict what movies or TV shows one is likely to watch online

QUESTION:

**How much do you support or oppose the development of AI?**

ANSWER CHOICES:

- Strongly support (2)
- Somewhat support (1)
- Neither support nor oppose (0)
- Somewhat oppose (-1)
- Strongly oppose (-2)
- I don't know

Table B.36: Support for developing AI; $N = 2000$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Strongly support | 12.58 | 12.65 | 253 |
| Somewhat support | 28.36 | 28.65 | 573 |
| Neither support nor oppose | 27.84 | 27.60 | 552 |
| Somewhat oppose | 12.79 | 12.75 | 255 |

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Strongly oppose | 8.90 | 9.05 | 181 |
| I don't know | 9.54 | 9.30 | 186 |
| Skipped | 0 | 0 | 0 |

## B.5 Survey experiment: AI and/or robots should be carefully managed

QUESTION:

Please tell me to what extent you agree or disagree with the following statement.

[Respondents were presented with one statement randomly selected from the list below.]

- AI and robots are technologies that require careful management.
- AI is a technology that requires careful management.
- Robots are technologies that require careful management.

ANSWER CHOICES:

- Totally agree (2)
- Tend to agree (1)
- Tend to disagree (-1)
- Totally disagree (-2)
- I don't know

Table B.37: Responses to statement - AI and robots; $N = 656$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Totally agree | 51.41 | 53.20 | 349 |
| Tend to agree | 30.09 | 28.96 | 190 |
| Tend to disagree | 4.79 | 3.81 | 25 |
| Totally disagree | 0.59 | 0.76 | 5 |
| I don't know | 13.12 | 13.26 | 87 |
| Skipped | 0 | 0 | 0 |

Table B.38: Responses to statement - AI; $N = 667$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Totally agree | 53.54 | 53.67 | 358 |
| Tend to agree | 30.85 | 30.13 | 201 |
| Tend to disagree | 3.67 | 3.90 | 26 |
| Totally disagree | 0.80 | 0.90 | 6 |
| I don't know | 11.14 | 11.39 | 76 |
| Skipped | 0 | 0 | 0 |

Table B.39: Responses to statement - Robots; $N = 677$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Totally agree | 51.66 | 52.44 | 355 |
| Tend to agree | 30.31 | 31.31 | 212 |
| Tend to disagree | 5.76 | 5.17 | 35 |

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Totally disagree | 1.81 | 1.48 | 10 |
| I don't know | 10.46 | 9.60 | 65 |
| Skipped | 0 | 0 | 0 |

## B.6 Trust of actors to develop AI

QUESTION:

**How much confidence, if any, do you have in each of the following to develop AI in the best interests of the public?**

[Respondents were shown five items randomly selected from the list below. We included explainer text for actors not well known to the public; respondents could view the explainer text by hovering their mouse over the actor's name. The items and the answer choices were shown in a matrix format.]

- The U.S. military
- The U.S. civilian government
- National Security Agency (NSA)
- Federal Bureau of Investigation (FBI)
- Central Intelligence Agency (CIA)
- North Atlantic Treaty Organization (NATO)
- Explainer text for NATO: NATO is a military alliance that includes 28 countries including most of Europe, as well as the U.S. and Canada.
- An international research organization (e.g., CERN)
- Explainer text for CERN: The European Organization for Nuclear Research, known as CERN, is a European research organization that operates the largest particle physics laboratory in the world.
- Tech companies
- Google
- Facebook
- Apple
- Microsoft
- Amazon
- A non-profit AI research organization (e.g., OpenAI)
- Explainer text for OpenAI: Open AI is an AI non-profit organization with backing from tech investors that seeks to develop safe AI. University researchers

ANSWER CHOICES:

- A great deal of confidence (3)
- A fair amount of confidence (2)
- Not too much confidence (1)
- No confidence (0)
- I don't know

Table B.40: U.S. military; $N = 638$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| A great deal of confidence | 17.16 | 17.08 | 109 |
| A fair amount of confidence | 32.19 | 30.88 | 197 |
| Not too much confidence | 23.92 | 24.14 | 154 |
| No confidence | 14.40 | 14.89 | 95 |
| I don't know | 12.33 | 13.01 | 83 |
| Skipped | 0 | 0 | 0 |

Table B.41: U.S. civilian government; $N = 671$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| A great deal of confidence | 5.59 | 5.66 | 38 |
| A fair amount of confidence | 24.04 | 24.29 | 163 |
| Not too much confidence | 32.77 | 33.23 | 223 |
| No confidence | 23.80 | 23.40 | 157 |
| I don't know | 13.79 | 13.41 | 90 |
| Skipped | 0 | 0 | 0 |

Table B.42: NSA; $N = 710$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| A great deal of confidence | 9.63 | 9.30 | 66 |
| A fair amount of confidence | 28.04 | 26.90 | 191 |
| Not too much confidence | 26.65 | 26.76 | 190 |
| No confidence | 22.82 | 24.37 | 173 |
| I don't know | 12.87 | 12.68 | 90 |
| Skipped | 0 | 0 | 0 |

Table B.43: FBI; $N = 656$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| A great deal of confidence | 9.26 | 9.60 | 63 |
| A fair amount of confidence | 26.20 | 25.46 | 167 |
| Not too much confidence | 25.07 | 25.15 | 165 |
| No confidence | 27.10 | 27.44 | 180 |
| I don't know | 12.25 | 12.20 | 80 |
| Skipped | 0.14 | 0.15 | 1 |

Table B.44: CIA; $N = 730$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| A great deal of confidence | 8.43 | 8.77 | 64 |
| A fair amount of confidence | 26.10 | 25.07 | 183 |
| Not too much confidence | 26.80 | 26.99 | 197 |
| No confidence | 25.61 | 26.30 | 192 |
| I don't know | 12.93 | 12.74 | 93 |
| Skipped | 0.13 | 0.14 | 1 |

Table B.45: NATO; $N = 695$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| A great deal of confidence | 4.40 | 4.17 | 29 |
| A fair amount of confidence | 25.41 | 24.75 | 172 |
| Not too much confidence | 25.98 | 26.62 | 185 |
| No confidence | 23.13 | 24.03 | 167 |

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| I don't know | 21.08 | 20.43 | 142 |
| Skipped | 0 | 0 | 0 |

Table B.46: Intergovernmental research organizations (e.g., CERN); $N$ = 645

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| A great deal of confidence | 11.97 | 12.25 | 79 |
| A fair amount of confidence | 28.87 | 28.84 | 186 |
| Not too much confidence | 22.94 | 22.64 | 146 |
| No confidence | 16.85 | 16.59 | 107 |
| I don't know | 19.37 | 19.69 | 127 |
| Skipped | 0 | 0 | 0 |

Table B.47: Tech companies; $N$ = 674

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| A great deal of confidence | 10.28 | 10.83 | 73 |
| A fair amount of confidence | 34.15 | 34.57 | 233 |
| Not too much confidence | 28.40 | 27.15 | 183 |
| No confidence | 14.91 | 15.13 | 102 |
| I don't know | 12.15 | 12.17 | 82 |
| Skipped | 0.12 | 0.15 | 1 |

Table B.48: Google; $N$ = 645

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| A great deal of confidence | 11.91 | 11.47 | 74 |
| A fair amount of confidence | 27.35 | 26.82 | 173 |
| Not too much confidence | 25.92 | 26.67 | 172 |
| No confidence | 21.56 | 21.40 | 138 |
| I don't know | 13.00 | 13.33 | 86 |
| Skipped | 0.26 | 0.31 | 2 |

Table B.49: Facebook; $N$ = 632

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| A great deal of confidence | 4.29 | 3.96 | 25 |
| A fair amount of confidence | 14.35 | 13.45 | 85 |
| Not too much confidence | 26.40 | 27.22 | 172 |
| No confidence | 41.27 | 42.88 | 271 |
| I don't know | 13.44 | 12.18 | 77 |
| Skipped | 0.25 | 0.32 | 2 |

Table B.50: Apple; $N = 697$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| A great deal of confidence | 10.41 | 10.76 | 75 |
| A fair amount of confidence | 26.29 | 26.26 | 183 |
| Not too much confidence | 27.00 | 27.98 | 195 |
| No confidence | 22.20 | 21.81 | 152 |
| I don't know | 13.84 | 12.91 | 90 |
| Skipped | 0.26 | 0.29 | 2 |

Table B.51: Microsoft; $N = 597$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| A great deal of confidence | 10.85 | 10.89 | 65 |
| A fair amount of confidence | 33.08 | 32.66 | 195 |
| Not too much confidence | 26.89 | 27.14 | 162 |
| No confidence | 17.99 | 17.76 | 106 |
| I don't know | 11.05 | 11.39 | 68 |
| Skipped | 0.14 | 0.17 | 1 |

Table B.52: Amazon; $N = 685$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| A great deal of confidence | 10.60 | 10.95 | 75 |
| A fair amount of confidence | 29.53 | 29.34 | 201 |
| Not too much confidence | 25.51 | 25.40 | 174 |
| No confidence | 22.02 | 22.19 | 152 |
| I don't know | 12.34 | 12.12 | 83 |
| Skipped | 0 | 0 | 0 |

Table B.53: Non-profit (e.g., OpenAI); $N = 659$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| A great deal of confidence | 10.19 | 10.17 | 67 |
| A fair amount of confidence | 29.40 | 30.35 | 200 |
| Not too much confidence | 23.57 | 23.98 | 158 |
| No confidence | 13.65 | 13.66 | 90 |
| I don't know | 23.04 | 21.70 | 143 |
| Skipped | 0.13 | 0.15 | 1 |

Table B.54: University researchers; $N = 666$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| A great deal of confidence | 13.86 | 14.11 | 94 |
| A fair amount of confidence | 36.29 | 36.04 | 240 |
| Not too much confidence | 22.27 | 22.82 | 152 |
| No confidence | 12.75 | 12.31 | 82 |

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| I don't know | 14.70 | 14.56 | 97 |
| Skipped | 0.14 | 0.15 | 1 |

## B.7 Trust of actors to manage AI

QUESTION:

**How much confidence, if any, do you have in each of the following to manage the development and use of AI in the best interests of the public?**

[Respondents were shown five items randomly selected from the list below. We included explainer text for actors not well known to the public; respondents could view the explainer text by hovering their mouse over the actor's name. The items and the answer choices were shown in a matrix format.]

- U.S. federal government
- U.S. state governments
- International organizations (e.g., United Nations, European Union)
- The United Nations (UN)
- An intergovernmental research organization (e.g., CERN)
- Explainer text for CERN: The European Organization for Nuclear Research, known as CERN, is a European research organization that operates the largest particle physics laboratory in the world.
- Tech companies
- Google
- Facebook
- Apple
- Microsoft
- Amazon
- Non-government scientific organizations (e.g., AAAI)
- Explainer text for AAAI: Association for the Advancement of Artificial Intelligence (AAAI) is a non-government scientific organization that promotes research in, and responsible use of AI.
- Partnership on AI, an association of tech companies, academics, and civil society groups

ANSWER CHOICES:

- A great deal of confidence (3)
- A fair amount of confidence (2)
- Not too much confidence (1)
- No confidence (0)
- I don't know

Table B.55: U.S. federal government; $N = 743$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| A great deal of confidence | 6.86 | 6.59 | 49 |
| A fair amount of confidence | 20.26 | 20.19 | 150 |
| Not too much confidence | 28.44 | 28.67 | 213 |
| No confidence | 31.50 | 32.44 | 241 |
| I don't know | 12.68 | 11.84 | 88 |
| Skipped | 0.25 | 0.27 | 2 |

Table B.56: U.S. state governments; $N = 713$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| A great deal of confidence | 6.25 | 6.45 | 46 |
| A fair amount of confidence | 20.39 | 19.21 | 137 |
| Not too much confidence | 31.57 | 32.12 | 229 |
| No confidence | 29.65 | 30.72 | 219 |
| I don't know | 11.69 | 11.22 | 80 |
| Skipped | 0.45 | 0.28 | 2 |

Table B.57: International organizations; $N = 827$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| A great deal of confidence | 5.94 | 5.80 | 48 |
| A fair amount of confidence | 22.48 | 21.77 | 180 |
| Not too much confidence | 29.58 | 29.87 | 247 |
| No confidence | 26.81 | 27.45 | 227 |
| I don't know | 14.81 | 14.87 | 123 |
| Skipped | 0.38 | 0.24 | 2 |

Table B.58: UN; $N = 802$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| A great deal of confidence | 6.23 | 6.61 | 53 |
| A fair amount of confidence | 22.49 | 21.57 | 173 |
| Not too much confidence | 26.14 | 26.18 | 210 |
| No confidence | 31.90 | 31.55 | 253 |
| I don't know | 12.64 | 13.59 | 109 |
| Skipped | 0.60 | 0.50 | 4 |

Table B.59: Intergovernmental research organizations (e.g., CERN); $N = 747$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| A great deal of confidence | 6.69 | 7.10 | 53 |
| A fair amount of confidence | 30.51 | 29.72 | 222 |
| Not too much confidence | 23.89 | 24.10 | 180 |
| No confidence | 20.32 | 20.21 | 151 |
| I don't know | 18.36 | 18.61 | 139 |
| Skipped | 0.22 | 0.27 | 2 |

Table B.60: Tech companies; $N = 758$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| A great deal of confidence | 8.33 | 8.44 | 64 |
| A fair amount of confidence | 33.50 | 32.98 | 250 |
| Not too much confidence | 25.07 | 26.12 | 198 |

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| No confidence | 19.88 | 20.45 | 155 |
| I don't know | 12.81 | 11.74 | 89 |
| Skipped | 0.41 | 0.26 | 2 |

Table B.61: Google; $N = 767$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| A great deal of confidence | 9.61 | 9.13 | 70 |
| A fair amount of confidence | 23.60 | 23.86 | 183 |
| Not too much confidence | 27.44 | 27.77 | 213 |
| No confidence | 25.13 | 25.03 | 192 |
| I don't know | 13.75 | 13.95 | 107 |
| Skipped | 0.47 | 0.26 | 2 |

Table B.62: Facebook; $N = 741$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| A great deal of confidence | 4.99 | 4.45 | 33 |
| A fair amount of confidence | 16.18 | 16.19 | 120 |
| Not too much confidence | 28.50 | 28.21 | 209 |
| No confidence | 36.95 | 38.46 | 285 |
| I don't know | 13.14 | 12.42 | 92 |
| Skipped | 0.24 | 0.27 | 2 |

Table B.63: Apple; $N = 775$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| A great deal of confidence | 8.25 | 8.39 | 65 |
| A fair amount of confidence | 25.10 | 24.90 | 193 |
| Not too much confidence | 29.08 | 28.65 | 222 |
| No confidence | 23.91 | 24.52 | 190 |
| I don't know | 13.55 | 13.42 | 104 |
| Skipped | 0.12 | 0.13 | 1 |

Table B.64: Microsoft; $N = 771$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| A great deal of confidence | 7.79 | 7.78 | 60 |
| A fair amount of confidence | 30.11 | 29.83 | 230 |
| Not too much confidence | 22.98 | 23.48 | 181 |
| No confidence | 24.10 | 24.38 | 188 |
| I don't know | 14.68 | 14.14 | 109 |
| Skipped | 0.35 | 0.39 | 3 |

Table B.65: Amazon; $N = 784$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| A great deal of confidence | 10.19 | 10.33 | 81 |
| A fair amount of confidence | 25.22 | 24.87 | 195 |
| Not too much confidence | 25.20 | 25.38 | 199 |
| No confidence | 24.53 | 24.87 | 195 |
| I don't know | 14.87 | 14.54 | 114 |
| Skipped | 0 | 0 | 0 |

Table B.66: Non-government scientific organization (e.g., AAAI); $N = 792$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| A great deal of confidence | 7.64 | 7.83 | 62 |
| A fair amount of confidence | 30.32 | 30.05 | 238 |
| Not too much confidence | 25.37 | 26.39 | 209 |
| No confidence | 15.03 | 14.65 | 116 |
| I don't know | 21.46 | 20.83 | 165 |
| Skipped | 0.19 | 0.25 | 2 |

Table B.67: Partnership on AI; $N = 780$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| A great deal of confidence | 8.89 | 9.23 | 72 |
| A fair amount of confidence | 30.12 | 29.49 | 230 |
| Not too much confidence | 25.89 | 26.79 | 209 |
| No confidence | 16.33 | 15.77 | 123 |
| I don't know | 18.64 | 18.59 | 145 |
| Skipped | 0.12 | 0.13 | 1 |

## B.8 AI governance challenges

We would like you to consider some potential policy issues related to AI. Please consider the following:

[Respondents were shown five randomly-selected items from the list below, one item at a time. For ease of comprehension, we include the shorten labels used in the figures in square brackets.]

- **[Hiring bias] Fairness and transparency in AI used in hiring**: Increasingly, employers are using AI to make hiring decisions. AI has the potential to make less biased hiring decisions than humans. But algorithms trained on biased data can lead to lead to hiring practices that discriminate against certain groups. Also, AI used in this application may lack transparency, such that human users do not understand what the algorithm is doing, or why it reaches certain decisions in specific cases.
- **[Criminal justice bias] Fairness and transparency in AI used in criminal justice**: Increasingly, the criminal justice system is using AI to make sentencing and parole decisions. AI has the potential to make less biased hiring decisions than humans. But algorithms trained on biased data could lead to discrimination against certain groups. Also, AI used in this application may lack transparency such that human users do not understand what the algorithm is doing, or why it reaches certain decisions in specific cases.
- **[Disease diagnosis] Accuracy and transparency in AI used for disease diagnosis**: Increasingly, AI software has been used to diagnose diseases, such as heart disease and cancer. One challenge is to make sure the AI can correctly diagnose those who have the disease and not mistakenly diagnose those who do not have the disease. Another

challenge is that AI used in this application may lack transparency such that human users do not understand what the algorithm is doing, or why it reaches certain decisions in specific cases.

- **[Data privacy] Protect data privacy**: Algorithms used in AI applications are often trained on vast amounts of personal data, including medical records, social media content, and financial transactions. Some worry that data used to train algorithms are not collected, used, and stored in ways that protect personal privacy.
- **[Autonomous vehicles] Make sure autonomous vehicles are safe**: Companies are developing self-driving cars and trucks that require little or no input from humans. Some worry about the safety of autonomous vehicles for those riding in them as well as for other vehicles, cyclists, and pedestrians.
- **[Ditigal manipulation] Prevent AI from being used to spread fake and harmful content online**: AI has been used by governments, private groups, and individuals to harm or manipulate internet users. For instance, automated bots have been used to generate and spread false and/or harmful news stories, audios, and videos.
- **[Cyber attacks] Prevent AI cyber attacks against governments, companies, organizations, and individuals**: Computer scientists have shown that AI can be used to launch effective cyber attacks. AI could be used to hack into servers to steal sensitive information, shut down critical infrastructures like power grids or hospital networks, or scale up targeted phishing attacks.
- **[Surveillance] Prevent AI-assisted surveillance from violating privacy and civil liberties**: AI can be used to process and analyze large amounts of text, photo, audio, and video data from social media, mobile communications, and CCTV cameras. Some worry that governments, companies, and employers could use AI to increase their surveillance capabilities.
- **[U.S.-China arms race] Prevent escalation of a U.S.-China AI arms race**: Leading analysts believe that an AI arms race is beginning, in which the U.S. and China are investing billions of dollars to develop powerful AI systems for surveillance, autonomous weapons, cyber operations, propaganda, and command and control systems. Some worry that a U.S.-China arms race could lead to extreme dangers. To stay ahead, the U.S. and China may race to deploy advanced military AI systems that they do not fully understand or can control. We could see catastrophic accidents, such as a rapid, automated escalation involving cyber and nuclear weapons.
- **[Value alignment] Make sure AI systems are safe, trustworthy, and aligned with human values**: As AI systems become more advanced, they will increasingly make decisions without human input. One potential fear is that AI systems, while performing jobs they are programmed to do, could unintentionally make decisions that go against the values of its human users, such as physically harming people.
- **[Autonomous weapons] Ban the use of lethal autonomous weapons (LAWs)**: Lethal autonomous weapons (LAWs) are military robots that can attack targets without control by humans. LAWs could reduce the use of human combatants on the battlefield. But some worry that the adoption of LAWs could lead to mass violence. Because they are cheap and easy to produce in bulk, national militaries, terrorists, and other groups could readily deploy LAWs.
- **[Technological unemployment] Guarantee a good standard of living for those who lose their jobs to automation**: Some forecast that AI will increasingly be able to do jobs done by humans today. AI could potentially do the jobs of blue-collar workers, like truckers and factory workers, as well as the jobs of white-collar workers, like financial analysts or lawyers. Some worry that in the future, robots and computers can do most of the jobs that are done by humans today.
- **[Critical AI systems failure] Prevent critical AI systems failures**: As AI systems become more advanced, they could be used by the military or in critical infrastructure, like power grids, highways, or hospital networks. Some worry that the failure of AI systems or unintentional accidents in these applications could cause 10 percent or more of all humans to die.

QUESTION:

**In the next 10 years, how likely do you think it is that this AI governance challenge will impact large numbers of people in the U.S.?**

ANSWER CHOICES:

- Very unlikely: less than 5% chance (2.5%)
- Unlikely: 5-20% chance (12.5%)
- Somewhat unlikely: 20-40% chance (30%)
- Equally likely as unlikely: 40-60% chance (50%)
- Somewhat likely: 60-80% chance (70%)

- Likely: 80-95% chance (87.5%)
- Very likely: more than 95% chance (97.5%)
- I don't know

QUESTION:

**In the next 10 years, how likely do you think it is that this AI governance challenge will impact large numbers of people around the world?**

ANSWER CHOICES:

- Very unlikely: less than 5% chance (2.5%)
- Unlikely: 5-20% chance (12.5%)
- Somewhat unlikely: 20-40% chance (30%)
- Equally likely as unlikely: 40-60% chance (50%)
- Somewhat likely: 60-80% chance (70%)
- Likely: 80-95% chance (87.5%)
- Very likely: more than 95% chance (97.5%)
- I don't know

QUESTION:

**In the next 10 years, how important is it for tech companies and governments to carefully manage the following challenge?**

ANSWER CHOICES:

- Very important (3)
- Somewhat important (2)
- Not too important (1)
- Not at all important (0)
- I don't know

Table B.68: Likelihood in the US - Hiring bias; $N = 760$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 2.57 | 2.63 | 20 |
| Unlikely 5-20% | 6.07 | 6.18 | 47 |
| Somewhat unlikely 20-40% | 10.86 | 10.92 | 83 |
| Equally likely as unlikely 40-60% | 22.27 | 22.50 | 171 |
| Somewhat likely 60-80% | 23.34 | 22.89 | 174 |
| Likely 80-95% | 12.39 | 12.76 | 97 |
| Very likely > 95% | 9.86 | 9.61 | 73 |
| I don't know | 12.35 | 12.37 | 94 |
| Skipped | 0.29 | 0.13 | 1 |

Table B.69: Likelihood in the US - Criminal justice bias; $N = 778$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 4.94 | 4.50 | 35 |
| Unlikely 5-20% | 8.76 | 8.61 | 67 |
| Somewhat unlikely 20-40% | 13.25 | 12.85 | 100 |
| Equally likely as unlikely 40-60% | 21.23 | 21.08 | 164 |
| Somewhat likely 60-80% | 17.13 | 17.22 | 134 |
| Likely 80-95% | 12.28 | 12.60 | 98 |
| Very likely > 95% | 9.05 | 9.64 | 75 |

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| I don't know | 12.90 | 12.98 | 101 |
| Skipped | 0.45 | 0.51 | 4 |

Table B.70: Likelihood in the US - Disease diagnosis; $N = 767$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 2.79 | 2.61 | 20 |
| Unlikely 5-20% | 4.73 | 4.95 | 38 |
| Somewhat unlikely 20-40% | 10.18 | 9.52 | 73 |
| Equally likely as unlikely 40-60% | 23.12 | 23.21 | 178 |
| Somewhat likely 60-80% | 20.50 | 19.95 | 153 |
| Likely 80-95% | 13.43 | 13.95 | 107 |
| Very likely > 95% | 9.72 | 10.17 | 78 |
| I don't know | 13.62 | 13.69 | 105 |
| Skipped | 1.91 | 1.96 | 15 |

Table B.71: Likelihood in the US - Data privacy; $N = 807$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 2.75 | 2.11 | 17 |
| Unlikely 5-20% | 4.53 | 4.58 | 37 |
| Somewhat unlikely 20-40% | 7.52 | 7.19 | 58 |
| Equally likely as unlikely 40-60% | 16.10 | 15.86 | 128 |
| Somewhat likely 60-80% | 18.81 | 19.33 | 156 |
| Likely 80-95% | 17.00 | 16.36 | 132 |
| Very likely > 95% | 20.59 | 21.69 | 175 |
| I don't know | 10.87 | 10.78 | 87 |
| Skipped | 1.84 | 2.11 | 17 |

Table B.72: Likelihood in the US - Autonomous vehicles; $N = 796$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 3.65 | 3.64 | 29 |
| Unlikely 5-20% | 5.80 | 5.90 | 47 |
| Somewhat unlikely 20-40% | 10.93 | 10.43 | 83 |
| Equally likely as unlikely 40-60% | 16.17 | 16.33 | 130 |
| Somewhat likely 60-80% | 23.62 | 23.62 | 188 |
| Likely 80-95% | 15.78 | 15.45 | 123 |
| Very likely > 95% | 12.29 | 12.94 | 103 |
| I don't know | 10.89 | 10.68 | 85 |
| Skipped | 0.87 | 1.01 | 8 |

Table B.73: Likelihood in the US - Digital manipulation; $N = 741$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 2.79 | 2.83 | 21 |

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Unlikely 5-20% | 3.24 | 3.10 | 23 |
| Somewhat unlikely 20-40% | 8.12 | 7.69 | 57 |
| Equally likely as unlikely 40-60% | 13.81 | 14.30 | 106 |
| Somewhat likely 60-80% | 16.58 | 16.33 | 121 |
| Likely 80-95% | 17.74 | 18.08 | 134 |
| Very likely > 95% | 23.45 | 23.62 | 175 |
| I don't know | 12.49 | 12.15 | 90 |
| Skipped | 1.77 | 1.89 | 14 |

Table B.74: Likelihood in the US - Cyber attacks; $N = 745$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 3.36 | 2.42 | 18 |
| Unlikely 5-20% | 4.28 | 3.89 | 29 |
| Somewhat unlikely 20-40% | 8.44 | 8.59 | 64 |
| Equally likely as unlikely 40-60% | 15.45 | 15.84 | 118 |
| Somewhat likely 60-80% | 19.22 | 19.46 | 145 |
| Likely 80-95% | 15.96 | 15.30 | 114 |
| Very likely > 95% | 20.52 | 21.21 | 158 |
| I don't know | 9.70 | 10.47 | 78 |
| Skipped | 3.07 | 2.82 | 21 |

Table B.75: Likelihood in the US - Surveillance; $N = 784$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 2.70 | 2.42 | 19 |
| Unlikely 5-20% | 2.92 | 2.81 | 22 |
| Somewhat unlikely 20-40% | 6.19 | 6.38 | 50 |
| Equally likely as unlikely 40-60% | 15.23 | 15.05 | 118 |
| Somewhat likely 60-80% | 18.95 | 18.75 | 147 |
| Likely 80-95% | 16.03 | 15.69 | 123 |
| Very likely > 95% | 23.52 | 24.23 | 190 |
| I don't know | 12.15 | 12.12 | 95 |
| Skipped | 2.32 | 2.55 | 20 |

Table B.76: Likelihood in the US - U.S.-China arms race; $N = 766$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 3.24 | 3.26 | 25 |
| Unlikely 5-20% | 5.98 | 6.01 | 46 |
| Somewhat unlikely 20-40% | 10.01 | 10.84 | 83 |
| Equally likely as unlikely 40-60% | 18.74 | 18.41 | 141 |
| Somewhat likely 60-80% | 20.08 | 19.71 | 151 |
| Likely 80-95% | 13.17 | 12.79 | 98 |
| Very likely > 95% | 10.62 | 11.36 | 87 |
| I don't know | 15.17 | 14.62 | 112 |
| Skipped | 3.00 | 3.00 | 23 |

Table B.77: Likelihood in the US - Value alignment; $N = 783$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 3.78 | 4.21 | 33 |
| Unlikely 5-20% | 7.30 | 6.90 | 54 |
| Somewhat unlikely 20-40% | 9.01 | 9.07 | 71 |
| Equally likely as unlikely 40-60% | 20.34 | 19.54 | 153 |
| Somewhat likely 60-80% | 19.26 | 19.28 | 151 |
| Likely 80-95% | 13.66 | 13.79 | 108 |
| Very likely > 95% | 12.96 | 13.67 | 107 |
| I don't know | 12.43 | 12.26 | 96 |
| Skipped | 1.26 | 1.28 | 10 |

Table B.78: Likelihood in the US - Autonomous weapons; $N = 757$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 6.22 | 5.94 | 45 |
| Unlikely 5-20% | 10.36 | 9.38 | 71 |
| Somewhat unlikely 20-40% | 12.75 | 12.68 | 96 |
| Equally likely as unlikely 40-60% | 18.91 | 19.02 | 144 |
| Somewhat likely 60-80% | 15.72 | 15.72 | 119 |
| Likely 80-95% | 11.44 | 11.76 | 89 |
| Very likely > 95% | 10.72 | 11.23 | 85 |
| I don't know | 11.99 | 12.29 | 93 |
| Skipped | 1.89 | 1.98 | 15 |

Table B.79: Likelihood in the US - Technological unemployment; $N = 738$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 3.08 | 2.98 | 22 |
| Unlikely 5-20% | 5.80 | 5.69 | 42 |
| Somewhat unlikely 20-40% | 11.00 | 11.11 | 82 |
| Equally likely as unlikely 40-60% | 17.74 | 17.62 | 130 |
| Somewhat likely 60-80% | 17.16 | 17.75 | 131 |
| Likely 80-95% | 14.86 | 14.91 | 110 |
| Very likely > 95% | 15.75 | 15.99 | 118 |
| I don't know | 12.84 | 12.20 | 90 |
| Skipped | 1.75 | 1.76 | 13 |

Table B.80: Likelihood in the US - Critical AI systems failure; $N = 778$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 6.98 | 6.43 | 50 |
| Unlikely 5-20% | 7.94 | 7.58 | 59 |
| Somewhat unlikely 20-40% | 12.26 | 12.98 | 101 |
| Equally likely as unlikely 40-60% | 20.36 | 20.31 | 158 |
| Somewhat likely 60-80% | 15.59 | 15.42 | 120 |
| Likely 80-95% | 12.25 | 11.83 | 92 |
| Very likely > 95% | 9.36 | 10.15 | 79 |

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| I don't know | 14.85 | 14.78 | 115 |
| Skipped | 0.41 | 0.51 | 4 |

Table B.81: Likelihood around the world - Hiring bias; $N = 760$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 2.95 | 3.03 | 23 |
| Unlikely 5-20% | 5.47 | 5.00 | 38 |
| Somewhat unlikely 20-40% | 8.54 | 8.55 | 65 |
| Equally likely as unlikely 40-60% | 20.23 | 21.45 | 163 |
| Somewhat likely 60-80% | 21.55 | 21.32 | 162 |
| Likely 80-95% | 13.68 | 13.55 | 103 |
| Very likely > 95% | 12.20 | 12.11 | 92 |
| I don't know | 15.04 | 14.61 | 111 |
| Skipped | 0.35 | 0.39 | 3 |

Table B.82: Likelihood around the world - Criminal justice bias; $N = 778$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 4.44 | 4.24 | 33 |
| Unlikely 5-20% | 8.06 | 7.71 | 60 |
| Somewhat unlikely 20-40% | 10.96 | 10.80 | 84 |
| Equally likely as unlikely 40-60% | 19.17 | 19.41 | 151 |
| Somewhat likely 60-80% | 18.29 | 18.25 | 142 |
| Likely 80-95% | 13.09 | 13.62 | 106 |
| Very likely > 95% | 9.38 | 9.90 | 77 |
| I don't know | 16.38 | 15.94 | 124 |
| Skipped | 0.23 | 0.13 | 1 |

Table B.83: Likelihood around the world - Disease diagnosis; $N = 767$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 2.31 | 2.35 | 18 |
| Unlikely 5-20% | 4.18 | 4.17 | 32 |
| Somewhat unlikely 20-40% | 9.93 | 9.13 | 70 |
| Equally likely as unlikely 40-60% | 21.28 | 20.99 | 161 |
| Somewhat likely 60-80% | 20.47 | 20.47 | 157 |
| Likely 80-95% | 15.00 | 15.38 | 118 |
| Very likely > 95% | 10.94 | 11.47 | 88 |
| I don't know | 15.80 | 15.91 | 122 |
| Skipped | 0.09 | 0.13 | 1 |

Table B.84: Likelihood around the world - Data privacy; $N = 807$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 2.86 | 2.23 | 18 |

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Unlikely 5-20% | 2.92 | 2.60 | 21 |
| Somewhat unlikely 20-40% | 8.32 | 8.30 | 67 |
| Equally likely as unlikely 40-60% | 13.79 | 14.75 | 119 |
| Somewhat likely 60-80% | 19.07 | 18.84 | 152 |
| Likely 80-95% | 18.43 | 18.22 | 147 |
| Very likely > 95% | 21.09 | 21.81 | 176 |
| I don't know | 13.34 | 13.01 | 105 |
| Skipped | 0.19 | 0.25 | 2 |

Table B.85: Likelihood around the world - Autonomous vehicles; $N = 796$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 3.77 | 3.52 | 28 |
| Unlikely 5-20% | 5.25 | 5.65 | 45 |
| Somewhat unlikely 20-40% | 12.37 | 11.68 | 93 |
| Equally likely as unlikely 40-60% | 16.74 | 17.21 | 137 |
| Somewhat likely 60-80% | 21.09 | 21.11 | 168 |
| Likely 80-95% | 14.13 | 14.45 | 115 |
| Very likely > 95% | 12.04 | 12.19 | 97 |
| I don't know | 13.99 | 13.57 | 108 |
| Skipped | 0.63 | 0.63 | 5 |

Table B.86: Likelihood around the world - Digital manipulation; $N = 741$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 1.98 | 2.16 | 16 |
| Unlikely 5-20% | 1.67 | 1.48 | 11 |
| Somewhat unlikely 20-40% | 7.34 | 7.29 | 54 |
| Equally likely as unlikely 40-60% | 12.68 | 12.96 | 96 |
| Somewhat likely 60-80% | 17.18 | 17.00 | 126 |
| Likely 80-95% | 21.22 | 21.73 | 161 |
| Very likely > 95% | 22.31 | 22.00 | 163 |
| I don't know | 15.24 | 14.98 | 111 |
| Skipped | 0.39 | 0.40 | 3 |

Table B.87: Likelihood around the world - Cyber attacks; $N = 745$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 1.08 | 1.21 | 9 |
| Unlikely 5-20% | 4.95 | 4.03 | 30 |
| Somewhat unlikely 20-40% | 4.76 | 5.10 | 38 |
| Equally likely as unlikely 40-60% | 16.95 | 16.64 | 124 |
| Somewhat likely 60-80% | 18.94 | 19.73 | 147 |
| Likely 80-95% | 19.13 | 19.06 | 142 |
| Very likely > 95% | 20.57 | 20.40 | 152 |
| I don't know | 13.20 | 13.42 | 100 |
| Skipped | 0.42 | 0.40 | 3 |

Table B.88: Likelihood around the world - Surveillance; $N = 784$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 1.26 | 1.40 | 11 |
| Unlikely 5-20% | 3.55 | 3.19 | 25 |
| Somewhat unlikely 20-40% | 5.12 | 5.36 | 42 |
| Equally likely as unlikely 40-60% | 14.26 | 14.41 | 113 |
| Somewhat likely 60-80% | 18.90 | 19.13 | 150 |
| Likely 80-95% | 20.30 | 19.77 | 155 |
| Very likely > 95% | 22.62 | 22.70 | 178 |
| I don't know | 13.93 | 13.90 | 109 |
| Skipped | 0.07 | 0.13 | 1 |

Table B.89: Likelihood around the world - U.S.-China arms race; $N = 766$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 3.21 | 3.13 | 24 |
| Unlikely 5-20% | 4.61 | 4.83 | 37 |
| Somewhat unlikely 20-40% | 7.70 | 7.83 | 60 |
| Equally likely as unlikely 40-60% | 19.50 | 19.19 | 147 |
| Somewhat likely 60-80% | 20.71 | 20.76 | 159 |
| Likely 80-95% | 14.99 | 14.75 | 113 |
| Very likely > 95% | 12.46 | 12.92 | 99 |
| I don't know | 16.61 | 16.32 | 125 |
| Skipped | 0.22 | 0.26 | 2 |

Table B.90: Likelihood around the world - Value alignment; $N = 783$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 2.70 | 2.94 | 23 |
| Unlikely 5-20% | 4.66 | 4.60 | 36 |
| Somewhat unlikely 20-40% | 8.80 | 8.81 | 69 |
| Equally likely as unlikely 40-60% | 19.92 | 19.41 | 152 |
| Somewhat likely 60-80% | 18.97 | 18.77 | 147 |
| Likely 80-95% | 15.57 | 15.33 | 120 |
| Very likely > 95% | 14.93 | 15.71 | 123 |
| I don't know | 14.44 | 14.43 | 113 |
| Skipped | 0 | 0 | 0 |

Table B.91: Likelihood around the world - Autonomous weapons; $N = 757$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 3.72 | 3.70 | 28 |
| Unlikely 5-20% | 7.04 | 5.42 | 41 |
| Somewhat unlikely 20-40% | 9.42 | 9.64 | 73 |
| Equally likely as unlikely 40-60% | 17.23 | 17.44 | 132 |
| Somewhat likely 60-80% | 16.08 | 15.85 | 120 |
| Likely 80-95% | 16.35 | 17.04 | 129 |
| Very likely > 95% | 14.87 | 15.19 | 115 |

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| I don't know | 15.20 | 15.59 | 118 |
| Skipped | 0.09 | 0.13 | 1 |

Table B.92: Likelihood around the world - Technological unemployment; $N = 738$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 2.76 | 2.57 | 19 |
| Unlikely 5-20% | 4.92 | 4.47 | 33 |
| Somewhat unlikely 20-40% | 8.31 | 8.81 | 65 |
| Equally likely as unlikely 40-60% | 18.36 | 18.16 | 134 |
| Somewhat likely 60-80% | 19.90 | 21.00 | 155 |
| Likely 80-95% | 14.78 | 14.50 | 107 |
| Very likely > 95% | 16.71 | 16.67 | 123 |
| I don't know | 13.77 | 13.41 | 99 |
| Skipped | 0.51 | 0.41 | 3 |

Table B.93: Likelihood around the world - Critical AI systems failure; $N = 778$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 5.36 | 5.27 | 41 |
| Unlikely 5-20% | 8.07 | 7.97 | 62 |
| Somewhat unlikely 20-40% | 10.75 | 10.41 | 81 |
| Equally likely as unlikely 40-60% | 18.03 | 17.87 | 139 |
| Somewhat likely 60-80% | 16.71 | 16.84 | 131 |
| Likely 80-95% | 13.09 | 13.11 | 102 |
| Very likely > 95% | 11.23 | 11.83 | 92 |
| I don't know | 16.76 | 16.71 | 130 |
| Skipped | 0 | 0 | 0 |

Table B.94: Issue importance - Hiring bias; $N = 760$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very important | 56.86 | 57.11 | 434 |
| Somewhat important | 22.11 | 22.76 | 173 |
| Not too important | 6.56 | 6.05 | 46 |
| Not at all important | 1.50 | 1.58 | 12 |
| I don't know | 12.98 | 12.50 | 95 |
| Skipped | 0 | 0 | 0 |

Table B.95: Issue importance - Criminal justice bias; $N = 778$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very important | 56.08 | 56.68 | 441 |
| Somewhat important | 21.78 | 22.49 | 175 |

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Not too important | 6.65 | 5.91 | 46 |
| Not at all important | 1.83 | 1.67 | 13 |
| I don't know | 13.66 | 13.24 | 103 |
| Skipped | 0 | 0 | 0 |

Table B.96: Issue importance - Disease diagnosis; $N = 767$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very important | 55.60 | 56.98 | 437 |
| Somewhat important | 22.37 | 21.25 | 163 |
| Not too important | 6.68 | 6.91 | 53 |
| Not at all important | 1.98 | 1.83 | 14 |
| I don't know | 13.26 | 12.91 | 99 |
| Skipped | 0.11 | 0.13 | 1 |

Table B.97: Issue importance - Data privacy; $N = 807$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very important | 63.65 | 64.93 | 524 |
| Somewhat important | 17.65 | 17.10 | 138 |
| Not too important | 4.76 | 4.71 | 38 |
| Not at all important | 1.71 | 1.36 | 11 |
| I don't know | 12.05 | 11.65 | 94 |
| Skipped | 0.19 | 0.25 | 2 |

Table B.98: Issue importance - Autonomous vehicles; $N = 796$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very important | 58.70 | 59.55 | 474 |
| Somewhat important | 22.36 | 21.73 | 173 |
| Not too important | 6.13 | 6.28 | 50 |
| Not at all important | 1.44 | 1.63 | 13 |
| I don't know | 11.15 | 10.55 | 84 |
| Skipped | 0.22 | 0.25 | 2 |

Table B.99: Issue importance - Digital manipulation; $N = 741$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very important | 57.66 | 58.30 | 432 |
| Somewhat important | 18.75 | 18.08 | 134 |
| Not too important | 6.25 | 6.48 | 48 |
| Not at all important | 3.11 | 2.97 | 22 |
| I don't know | 14.16 | 14.04 | 104 |
| Skipped | 0.08 | 0.13 | 1 |

Table B.100: Issue importance - Cyber attacks; $N = 745$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very important | 62.12 | 61.21 | 456 |
| Somewhat important | 17.80 | 18.39 | 137 |
| Not too important | 7.07 | 7.38 | 55 |
| Not at all important | 1.14 | 1.07 | 8 |
| I don't know | 11.88 | 11.95 | 89 |
| Skipped | 0 | 0 | 0 |

Table B.101: Issue importance - Surveillance; $N = 784$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very important | 58.54 | 58.80 | 461 |
| Somewhat important | 19.33 | 19.26 | 151 |
| Not too important | 6.40 | 6.63 | 52 |
| Not at all important | 1.73 | 1.66 | 13 |
| I don't know | 13.93 | 13.52 | 106 |
| Skipped | 0.07 | 0.13 | 1 |

Table B.102: Issue importance - U.S.-China arms race; $N = 766$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very important | 55.88 | 55.74 | 427 |
| Somewhat important | 19.44 | 19.71 | 151 |
| Not too important | 7.07 | 7.57 | 58 |
| Not at all important | 2.38 | 2.35 | 18 |
| I don't know | 15.13 | 14.49 | 111 |
| Skipped | 0.10 | 0.13 | 1 |

Table B.103: Issue importance - Value alignment; $N = 783$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very important | 56.46 | 56.45 | 442 |
| Somewhat important | 20.49 | 20.95 | 164 |
| Not too important | 6.69 | 6.64 | 52 |
| Not at all important | 1.56 | 1.66 | 13 |
| I don't know | 14.80 | 14.30 | 112 |
| Skipped | 0 | 0 | 0 |

Table B.104: Issue importance - Autonomous weapons; $N = 757$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very important | 58.32 | 57.73 | 437 |
| Somewhat important | 20.00 | 19.55 | 148 |
| Not too important | 5.52 | 5.94 | 45 |
| Not at all important | 1.23 | 1.45 | 11 |

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| I don't know | 14.94 | 15.32 | 116 |
| Skipped | 0 | 0 | 0 |

Table B.105: Issue importance - Technological unemployment; $N = 738$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very important | 54.12 | 54.34 | 401 |
| Somewhat important | 22.07 | 22.49 | 166 |
| Not too important | 6.50 | 6.91 | 51 |
| Not at all important | 2.83 | 2.44 | 18 |
| I don't know | 14.39 | 13.69 | 101 |
| Skipped | 0.09 | 0.14 | 1 |

Table B.106: Issue importance - Critical AI systems failure; $N = 778$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very important | 52.63 | 53.86 | 419 |
| Somewhat important | 21.10 | 20.44 | 159 |
| Not too important | 7.98 | 8.10 | 63 |
| Not at all important | 2.93 | 2.44 | 19 |
| I don't know | 15.36 | 15.17 | 118 |
| Skipped | 0 | 0 | 0 |

## B.9 Survey experiment: comparing perceptions of U.S. vs. China AI research and development

[Respondents were presented with one randomly-selected question from the two below.]

QUESTIONS:

- Compared with other industrialized countries, how would you rate the U.S. in AI research and development?
- Compared with other industrialized countries, how would you rate China in AI research and development?

ANSWER CHOICES:

- Best in the world (3)
- Above average (2)
- Average (1)
- Below average (0)
- I don't know

Table B.107: Perceptions of research and development - U.S.; $N = 988$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Best in the world | 9.73 | 10.02 | 99 |
| Above average | 36.16 | 37.55 | 371 |
| Average | 26.09 | 24.70 | 244 |
| Below average | 4.99 | 4.96 | 49 |
| I don't know | 23.03 | 22.77 | 225 |
| Skipped | 0 | 0 | 0 |

Table B.108: Perceptions of research and development - China; $N = 1012$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Best in the world | 7.33 | 7.41 | 75 |
| Above average | 45.40 | 46.64 | 472 |
| Average | 16.66 | 15.81 | 160 |
| Below average | 3.93 | 3.66 | 37 |
| I don't know | 26.68 | 26.48 | 268 |
| Skipped | 0 | 0 | 0 |

## B.10 Survey experiment: U.S.-China arms race

[All respondents were presented with the following prompt.]

We want to understand your thoughts on some important issues in the news today. Please read the short news article below.

Leading analysts believe that an "AI arms race" is beginning, in which the U.S. and China are investing billions of dollars to develop powerful AI systems for surveillance, autonomous weapons, cyber operations, propaganda, and command and control systems.

[Respondents were randomly assigned to one of the four experimental groups listed below.]

### B.10.1 Control

[No additional text.]

### B.10.2 Nationalism treatment

Some leaders in the U.S. military and tech industry argue that the U.S. government should invest much more resources in AI research to ensure that the U.S.'s AI capabilities stay ahead of China's. Furthermore, they argue that the U.S. government should partner with American tech companies to develop advanced AI systems, particularly for military use.

According to a leaked memo produced by a senior National Security Council official, China has "assembled the basic components required for winning the Al arms race…Much like America's success in the competition for nuclear weapons, China's 21st Century Manhattan Project sets them on a path to getting there first."

### B.10.3 War risks treatment

Some prominent thinkers are concerned that a U.S.-China arms race could lead to extreme dangers. To stay ahead, the U.S. and China may race to deploy advanced military AI systems that they do not fully understand or can control. We could see catastrophic accidents, such as a rapid, automated escalation involving cyber and nuclear weapons.

"Competition for AI superiority at [the] national level [is the] most likely cause of World War Three," warned Elon Musk, the CEO of Tesla and SpaceX.

### B.10.4 Common humanity treatment

Some prominent thinkers are concerned that a U.S.-China arms race could lead to extreme dangers. To stay ahead, the U.S. and China may race to deploy advanced military AI systems that they do not fully understand or can control. We could see catastrophic accidents, such as a rapid, automated escalation involving cyber and nuclear weapons.

"Unless we learn how to prepare for, and avoid, the potential risks, AI could be the worst event in the history of our civilization. It brings dangers, like powerful autonomous weapons," warned the late Stephen Hawking, one of the world's most prominent physicists. At the same time, he said that with proper management of the technology, researchers "can create AI for the good of the world."

[The order of the next two questions is randomized.]

QUESTION:

**How much do you agree or disagree with the following statement?**

> The U.S. should invest more in AI military capabilities to make sure it doesn't fall behind China's, even if doing so may exacerbate the arms race. For instance, the U.S. could increase AI research funding for the military and universities. It could also collaborate with American tech companies to develop AI for military use.

ANSWER CHOICES:

- Strongly agree (2)
- Somewhat agree (1)
- Neither agree nor disagree (0)
- Somewhat disagree (-1)
- Strongly disagree (-2)
- I don't know

QUESTION:

**How much do you agree or disagree with the following statement?**

> The U.S. should work hard to cooperate with China to avoid the dangers of an AI arms race, even if doing so requires giving up some of the U.S.'s advantages. Cooperation could include collaborations between American and Chinese AI research labs, or the U.S. and China creating and committing to common safety standards.

ANSWER CHOICES:

- Strongly agree (2)
- Somewhat agree (1)
- Neither agree nor disagree (0)
- Somewhat disagree (-1)
- Strongly disagree (-2)
- I don't know

Table B.109: Responses to statement that U.S. should invest more in AI military capabilities - Control; $N = 510$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Strongly agree | 23.38 | 24.31 | 124 |
| Somewhat agree | 25.99 | 25.88 | 132 |
| Neither agree nor disagree | 23.48 | 22.75 | 116 |
| Somewhat disagree | 8.88 | 8.82 | 45 |
| Strongly disagree | 4.93 | 4.71 | 24 |
| I don't know | 13.34 | 13.53 | 69 |
| Skipped | 0 | 0 | 0 |

Table B.110: Responses to statement that U.S. should invest more in AI military capabilities - Treatment 1: Pro-nationalist; $N = 505$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Strongly agree | 20.88 | 20.40 | 103 |
| Somewhat agree | 26.89 | 27.52 | 139 |
| Neither agree nor disagree | 21.79 | 22.18 | 112 |
| Somewhat disagree | 11.69 | 12.28 | 62 |
| Strongly disagree | 5.30 | 5.35 | 27 |
| I don't know | 13.45 | 12.28 | 62 |
| Skipped | 0 | 0 | 0 |

Table B.111: Responses to statement that U.S. should invest more in AI military capabilities - Treatment 2: Risks of arms race; $N = 493$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Strongly agree | 18.26 | 19.07 | 94 |
| Somewhat agree | 27.85 | 27.38 | 135 |
| Neither agree nor disagree | 21.69 | 20.28 | 100 |
| Somewhat disagree | 12.87 | 13.79 | 68 |
| Strongly disagree | 6.88 | 6.90 | 34 |
| I don't know | 12.45 | 12.58 | 62 |
| Skipped | 0 | 0 | 0 |

Table B.112: Responses to statement that U.S. should invest more in AI military capabilities - Treatment 3: One common humanity; $N = 492$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Strongly agree | 22.38 | 20.53 | 101 |
| Somewhat agree | 27.29 | 27.85 | 137 |
| Neither agree nor disagree | 24.37 | 23.98 | 118 |
| Somewhat disagree | 6.73 | 7.11 | 35 |
| Strongly disagree | 6.17 | 6.91 | 34 |
| I don't know | 13.07 | 13.62 | 67 |
| Skipped | 0 | 0 | 0 |

Table B.113: Responses to statement that U.S. should work hard to cooperate with China to avoid dangers of AI arms race - Control; $N = 510$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Strongly agree | 22.34 | 22.55 | 115 |
| Somewhat agree | 26.16 | 26.27 | 134 |
| Neither agree nor disagree | 22.02 | 20.59 | 105 |
| Somewhat disagree | 8.29 | 9.02 | 46 |
| Strongly disagree | 7.38 | 7.45 | 38 |
| I don't know | 13.59 | 13.92 | 71 |
| Skipped | 0.21 | 0.20 | 1 |

Table B.114: Responses to statement that U.S. should work hard to cooperate with China to avoid dangers of AI arms race - Treatment 1: Pronationalist; $N = 505$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Strongly agree | 18.51 | 18.81 | 95 |
| Somewhat agree | 27.35 | 28.12 | 142 |
| Neither agree nor disagree | 20.08 | 20.99 | 106 |
| Somewhat disagree | 10.09 | 9.90 | 50 |
| Strongly disagree | 8.45 | 7.92 | 40 |
| I don't know | 15.51 | 14.26 | 72 |
| Skipped | 0 | 0 | 0 |

Table B.115: Responses to statement that U.S. should work hard to cooperate with China to avoid dangers of AI arms race - Treatment 2: Risks of arms race; $N = 493$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Strongly agree | 24.97 | 25.96 | 128 |
| Somewhat agree | 25.32 | 25.15 | 124 |
| Neither agree nor disagree | 21.53 | 20.49 | 101 |
| Somewhat disagree | 9.83 | 9.94 | 49 |
| Strongly disagree | 5.84 | 5.68 | 28 |
| I don't know | 12.51 | 12.78 | 63 |
| Skipped | 0 | 0 | 0 |

Table B.116: Responses to statement that U.S. should work hard to cooperate with China to avoid dangers of AI arms race - Treatment 3: One common humanity; $N = 492$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Strongly agree | 23.63 | 24.19 | 119 |
| Somewhat agree | 27.52 | 28.46 | 140 |
| Neither agree nor disagree | 21.31 | 20.33 | 100 |
| Somewhat disagree | 8.50 | 7.32 | 36 |
| Strongly disagree | 6.72 | 6.91 | 34 |
| I don't know | 12.31 | 12.80 | 63 |
| Skipped | 0 | 0 | 0 |

## B.11 Issue areas for possible U.S.-China cooperation

QUESTION:

**For the following issues, how likely is it that the U.S. and China can cooperate?**

[Respondents were presented with three issues from the list below. All three issues were presented on the same page; the order that they appeared was randomized.]

- Prevent AI cyber attacks against governments, companies, organizations, and individuals.
- Prevent AI-assisted surveillance from violating privacy and civil liberties.

- Make sure AI systems are safe, trustworthy, and aligned with human values.
- Ban the use of lethal autonomous weapons.
- Guarantee a good standard of living for those who lose their jobs to automation.

ANSWER CHOICES:

- Very unlikely: less than 5% chance (2.5%)
- Unlikely: 5-20% chance (12.5%)
- Somewhat unlikely: 20-40% chance (30%)
- Equally likely as unlikely: 40-60% chance (50%)
- Somewhat likely: 60-80% chance (70%)
- Likely: 80-95% chance (87.5%)
- Very likely: more than 95% chance (97.5%)
- I don't know

Table B.117: Likelihood of cooperation with China - Prevent AI cyber attacks against governments, companies, organizations, and individuals; $N = 1173$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely 5% | 9.20 | 8.95 | 105 |
| Unlikely 5-20% | 10.26 | 10.49 | 123 |
| Somewhat unlikely 20-40% | 17.56 | 17.22 | 202 |
| Equally likely as unlikely 40-60% | 23.55 | 23.36 | 274 |
| Somewhat likely 60-80% | 13.77 | 13.73 | 161 |
| Likely 80-95% | 6.98 | 7.25 | 85 |
| Very likely > 95% | 4.14 | 4.18 | 49 |
| I don't know | 14.45 | 14.75 | 173 |
| Skipped | 0.08 | 0.09 | 1 |

Table B.118: Likelihood of cooperation with China - Prevent AI-assisted surveillance from violating privacy and civil liberties; $N = 1140$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely 5% | 12.43 | 12.37 | 141 |
| Unlikely 5-20% | 12.78 | 13.33 | 152 |
| Somewhat unlikely 20-40% | 19.48 | 19.74 | 225 |
| Equally likely as unlikely 40-60% | 21.93 | 20.70 | 236 |
| Somewhat likely 60-80% | 10.59 | 10.79 | 123 |
| Likely 80-95% | 4.02 | 4.12 | 47 |
| Very likely > 95% | 3.82 | 4.12 | 47 |
| I don't know | 14.87 | 14.74 | 168 |
| Skipped | 0.08 | 0.09 | 1 |

Table B.119: Likelihood of cooperation with China - Make sure AI systems are safe, trustworthy, and aligned with human values; $N = 1226$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely 5% | 6.34 | 6.53 | 80 |
| Unlikely 5-20% | 9.07 | 8.97 | 110 |
| Somewhat unlikely 20-40% | 16.79 | 16.88 | 207 |
| Equally likely as unlikely 40-60% | 26.32 | 25.53 | 313 |

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Somewhat likely 60-80% | 14.84 | 14.85 | 182 |
| Likely 80-95% | 7.35 | 7.26 | 89 |
| Very likely > 95% | 5.77 | 5.87 | 72 |
| I don't know | 13.38 | 13.95 | 171 |
| Skipped | 0.14 | 0.16 | 2 |

Table B.120: Likelihood of cooperation with China - Ban the use of lethal autonomous weapons; $N = 1226$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely 5% | 12.28 | 12.32 | 151 |
| Unlikely 5-20% | 11.14 | 10.85 | 133 |
| Somewhat unlikely 20-40% | 14.03 | 14.03 | 172 |
| Equally likely as unlikely 40-60% | 23.98 | 23.65 | 290 |
| Somewhat likely 60-80% | 10.15 | 10.60 | 130 |
| Likely 80-95% | 6.67 | 6.93 | 85 |
| Very likely > 95% | 5.69 | 5.46 | 67 |
| I don't know | 15.91 | 15.99 | 196 |
| Skipped | 0.14 | 0.16 | 2 |

Table B.121: Likelihood of cooperation with China - Guarantee a good standard of living for those who lose their jobs to automation; $N = 1235$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely 5% | 13.19 | 13.36 | 165 |
| Unlikely 5-20% | 13.01 | 13.28 | 164 |
| Somewhat unlikely 20-40% | 18.26 | 18.46 | 228 |
| Equally likely as unlikely 40-60% | 22.81 | 22.19 | 274 |
| Somewhat likely 60-80% | 9.46 | 9.39 | 116 |
| Likely 80-95% | 5.08 | 5.18 | 64 |
| Very likely > 95% | 4.27 | 4.53 | 56 |
| I don't know | 13.78 | 13.44 | 166 |
| Skipped | 0.14 | 0.16 | 2 |

## B.12 Trend across time: job creation or job loss

QUESTION:

**How much do you agree or disagree with the following statement?**

[Respondents were presented with one statement randomly selected from the list below.]

- In general, automation and AI will create more jobs than they will eliminate.
- In general, automation and AI will create more jobs than they will eliminate in 10 years.
- In general, automation and AI will create more jobs than they will eliminate in 20 years.
- In general, automation and AI will create more jobs than they will eliminate in 50 years.

ANSWER CHOICES:

- Strongly agree (2)

- Agree (1)
- Disagree (-1)
- Strongly disagree (-2)
- I don't know

Table B.122: Responses to statement that automation and AI will create more jobs than they will eliminate - No time frame; $N = 484$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Strongly agree | 6.37 | 6.82 | 33 |
| Agree | 20.19 | 18.18 | 88 |
| Disagree | 27.39 | 28.10 | 136 |
| Strongly disagree | 21.43 | 22.31 | 108 |
| Don't know | 24.45 | 24.38 | 118 |
| Skipped | 0.17 | 0.21 | 1 |

Table B.123: Responses to statement that automation and AI will create more jobs than they will eliminate - 10 years; $N = 510$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Strongly agree | 3.40 | 3.53 | 18 |
| Agree | 17.67 | 18.04 | 92 |
| Disagree | 30.03 | 29.02 | 148 |
| Strongly disagree | 22.85 | 23.92 | 122 |
| Don't know | 26.04 | 25.49 | 130 |
| Skipped | 0 | 0 | 0 |

Table B.124: Responses to statement that automation and AI will create more jobs than they will eliminate - 20 years; $N = 497$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Strongly agree | 3.69 | 4.02 | 20 |
| Agree | 17.82 | 17.10 | 85 |
| Disagree | 31.02 | 30.99 | 154 |
| Strongly disagree | 21.31 | 21.73 | 108 |
| Don't know | 25.98 | 25.96 | 129 |
| Skipped | 0.18 | 0.20 | 1 |

Table B.125: Responses to statement that automation and AI will create more jobs than they will eliminate - 50 years; $N = 509$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Strongly agree | 6.77 | 6.48 | 33 |
| Agree | 15.37 | 15.52 | 79 |
| Disagree | 35.35 | 35.56 | 181 |
| Strongly disagree | 18.82 | 18.27 | 93 |
| Don't know | 23.69 | 24.17 | 123 |
| Skipped | 0 | 0 | 0 |

## B.13 High-level machine intelligence: forecasting timeline

QUESTION:

The following questions ask about high-level machine intelligence. We have high-level machine intelligence when machines are able to perform almost all tasks that are economically relevant today better than the median human (today) at each task. These tasks include asking subtle common-sense questions such as those that travel agents would ask. For the following questions, you should ignore tasks that are legally or culturally restricted to humans, such as serving on a jury.

**In your opinion, how likely is it that high-level machine intelligence will exist in 10 years? 20 years? 50 years? For each prediction, please use the slider to indicate the percent chance that you think high-level machine intelligence will exist. 0% chance means it will certainly not exist. 100% chance means it will certainly exist.**

______ In 10 years?

______ In 20 years?

______ In 50 years?

ANSWER CHOICES:

- Very unlikely: less than 5% chance (2.5%)
- Unlikely: 5-20% chance (12.5%)
- Somewhat unlikely: 20-40% chance (30%)
- Equally likely as unlikely: 40-60% chance (50%)
- Somewhat likely: 60-80% chance (70%)
- Likely: 80-95% chance (87.5%)
- Very likely: more than 95% chance (97.5%)
- I don't know

Table B.126: Forecasting high-level machine intelligence - 10 years; $N = 2000$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 4.46 | 4.50 | 90 |
| Unlikely 5-20% | 8.19 | 8.20 | 164 |
| Somewhat unlikely 20-40% | 14.84 | 14.75 | 295 |
| Equally likely as unlikely 40-60% | 20.34 | 19.95 | 399 |
| Somewhat likely 60-80% | 21.08 | 21.25 | 425 |
| Likely 80-95% | 10.69 | 10.65 | 213 |
| Very likely > 95% | 7.40 | 7.85 | 157 |
| I don't know | 12.91 | 12.75 | 255 |
| Skipped | 0.09 | 0.10 | 2 |

Table B.127: Forecasting high-level machine intelligence - 20 years; $N = 2000$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 1.52 | 1.45 | 29 |
| Unlikely 5-20% | 2.73 | 2.95 | 59 |
| Somewhat unlikely 20-40% | 6.26 | 5.85 | 117 |
| Equally likely as unlikely 40-60% | 16.83 | 16.40 | 328 |
| Somewhat likely 60-80% | 18.17 | 18.65 | 373 |
| Likely 80-95% | 22.25 | 22.25 | 445 |
| Very likely > 95% | 17.91 | 18.30 | 366 |
| I don't know | 14.18 | 14.00 | 280 |

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Skipped | 0.15 | 0.15 | 3 |

Table B.128: Forecasting high-level machine intelligence - 50 years; $N$ = 2000

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Very unlikely < 5% | 2.28 | 2.30 | 46 |
| Unlikely 5-20% | 1.66 | 1.55 | 31 |
| Somewhat unlikely 20-40% | 2.75 | 2.75 | 55 |
| Equally likely as unlikely 40-60% | 10.08 | 9.90 | 198 |
| Somewhat likely 60-80% | 12.33 | 12.20 | 244 |
| Likely 80-95% | 14.43 | 14.50 | 290 |
| Very likely > 95% | 40.86 | 41.15 | 823 |
| I don't know | 15.52 | 15.55 | 311 |
| Skipped | 0.09 | 0.10 | 2 |

## B.14 Support for developing high-level machine intelligence

QUESTION:

**How much do you support or oppose the development of high-level machine intelligence?**

ANSWER CHOICES:

- Strongly support
- Somewhat support
- Neither support nor oppose
- Somewhat oppose
- Strongly oppose
- I don't know

Table B.129: Support for developing high-level machine intelligence; $N$ = 2000

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Strongly support | 7.78 | 8.10 | 162 |
| Somewhat support | 23.58 | 23.30 | 466 |
| Neither support nor oppose | 29.40 | 28.75 | 575 |
| Somewhat oppose | 16.19 | 16.60 | 332 |
| Strongly oppose | 11.02 | 11.10 | 222 |
| I don't know | 11.94 | 12.05 | 241 |
| Skipped | 0.09 | 0.10 | 2 |

## B.15 Expected outcome of high-level machine intelligence

QUESTION:

Suppose that high-level machine intelligence could be developed one day. How positive or negative do you expect the overall impact of high-level machine intelligence to be on humanity in the long run?

ANSWER CHOICES:

- Extremely good
- On balance good
- More or less neutral
- On balance bad
- Extremely bad, possibly human extinction
- I don't know

Table B.130: Expected outcome of high-level machine intelligence; $N = 2000$

| Answer choices | Percentages (weighted) | Percentages (unweighted) | Raw frequencies |
|---|---|---|---|
| Extremely good | 5.35 | 5.45 | 109 |
| On balance good | 21.28 | 21.25 | 425 |
| More or less neutral | 21.00 | 21.10 | 422 |
| On balance bad | 22.38 | 23.10 | 462 |
| Extremely bad, possibly human extinction | 11.66 | 11.55 | 231 |
| Don't know | 18.25 | 17.45 | 349 |
| Skipped | 0.09 | 0.10 | 2 |

# C Appendix C: Additional data analysis results

## C.1 Support for developing AI

Table C.1 shows the regression results used to produce Figure 2.4.

Table C.1: Predicting support for developing AI using demographic characteristics: results from a multiple linear regression that includes all demographic variables; outcome standardized to have mean 0 and unit variance

| Variables | Coefficients (SEs) |
|---|---|
| (Intercept) | -0.27 (0.09)** |
| Age 38-53 | -0.16 (0.06)** |
| Age 54-72 | -0.18 (0.06)** |
| Age 73 and older | -0.16 (0.10) |
| Male | 0.17 (0.05)*** |
| Non-white | -0.02 (0.05) |
| Some college | -0.01 (0.06) |
| College+ | 0.18 (0.06)** |
| Employed (full- or part-time) | 0.03 (0.05) |
| Democrat | 0.20 (0.06)** |
| Independent/Other | -0.05 (0.06) |
| Income $30-70K | 0.01 (0.06) |
| Income $70-100K | 0.13 (0.09) |
| Income more than $100K | 0.16 (0.08)* |
| Prefer not to say income | -0.14 (0.07) |
| No religious affiliation | 0.16 (0.05)** |
| Other religion | 0.14 (0.08) |
| Born-again Christian | -0.04 (0.06) |
| CS or engineering degree | 0.05 (0.09) |
| CS or programming experience | 0.30 (0.06)*** |
| $N$ = 2000 | $F(19,1980) = 11.75$; $p$-value: $<0.001$ |

## C.2 Survey experiment and cross-national comparison: AI and/or robots should be carefully managed

We present the percentage of "don't know" or missing responses to the survey question (see Appendix B for the survey question text). Regression analysis shows that the varying the term used (i.e., AI, AI and robots, and robots) does not change responses to the statement that such technologies should be carefully managed. This finding is robust to a regression where we controlled for "don't know" or missing responses. In Table C.6, we present the distribution of responses to the statement by country.

Table C.2: Survey experiment attrition check: agreement with statement that AI and/or robots should be carefully managed

| Experimental condition | Percent DK/missing | Percent DK | Percent missing |
|---|---|---|---|
| AI | 11.39 | 11.39 | 0 |
| AI and robots | 13.26 | 13.26 | 0 |
| Robots | 9.60 | 9.60 | 0 |

Table C.3: Survey experiment attrition check: agreement with statement that AI and/or robots should be carefully managed

| Variables | Coefficients (SEs) |
|---|---|
| (Intercept) | 0.11 (0.01)*** |
| AI and robots | 0.02 (0.02) |
| Robots | -0.01 (0.02) |
| $N$ = 2000 | $F$(2, 1997) = 1.03; $p$-value: 0.359 |

Table C.4: Survey experiment results: agreement with statement that AI and/or robots should be carefully managed

| Variables | Coefficients (SEs) |
|---|---|
| (Intercept) | 1.49 (0.03)*** |
| AI and robots | -0.03 (0.04) |
| Robots | -0.09 (0.05) |
| $N$ = 2000 | $F$(2, 1997) = 1.92; $p$-value: 0.146 |

Table C.5: Survey experiment results: agreement with statement that AI and/or robots should be carefully managed (controlling for DK/missing responses)

| Variables | Coefficients (SEs) |
|---|---|
| (Intercept) | 1.46 (0.03)*** |
| AI and robots | 0.03 (0.05) |
| Robots | -0.07 (0.05) |
| $N$ = 2000 | $F$(5, 1994) = 0.91; $p$-value: 0.471 |

Table C.6: Distribution of responses to statement that AI and robots should be carefully managed by country (in percentages); EU countries data from Eurobarometer

| Countries | Totally disagree | Tend to disagree | Tend to agree | Totally agree | Don’t know |
|---|---|---|---|---|---|
| Austria | 3 | 7 | 43 | 43 | 4 |
| Belgium | 1 | 9 | 40 | 48 | 2 |
| Bulgaria | 1 | 2 | 24 | 65 | 8 |
| Croatia | 4 | 8 | 37 | 47 | 4 |
| Cyprus | 1 | 2 | 26 | 67 | 4 |
| Czech Republic | 2 | 7 | 37 | 50 | 4 |
| Denmark | 1 | 4 | 25 | 66 | 4 |
| Estonia | 0 | 4 | 39 | 51 | 6 |
| European Union | 2 | 5 | 35 | 53 | 5 |
| Finland | 1 | 4 | 29 | 63 | 3 |
| France | 1 | 3 | 31 | 62 | 3 |
| Germany | 2 | 4 | 32 | 59 | 3 |
| Greece | 1 | 3 | 23 | 71 | 2 |
| Hungary | 4 | 12 | 35 | 45 | 4 |
| Ireland | 1 | 4 | 37 | 54 | 4 |
| Italy | 3 | 8 | 43 | 40 | 6 |

| Countries | Totally disagree | Tend to disagree | Tend to agree | Totally agree | Don't know |
|---|---|---|---|---|---|
| Latvia | 1 | 3 | 29 | 63 | 4 |
| Lithuania | 0 | 4 | 35 | 57 | 4 |
| Luxembourg | 1 | 4 | 33 | 58 | 4 |
| Malta | 2 | 4 | 46 | 38 | 10 |
| Netherlands | 1 | 2 | 22 | 74 | 1 |
| Poland | 2 | 8 | 44 | 42 | 4 |
| Portugal | 2 | 2 | 37 | 48 | 11 |
| Romania | 5 | 12 | 33 | 42 | 8 |
| Slovakia | 0 | 5 | 44 | 46 | 5 |
| Slovenia | 2 | 6 | 37 | 52 | 3 |
| Spain | 1 | 3 | 40 | 47 | 9 |
| Sweden | 1 | 2 | 18 | 75 | 4 |
| United Kingdom | 1 | 3 | 34 | 57 | 5 |
| United States | 1 | 5 | 30 | 52 | 12 |

## C.3 Harmful consequences of AI in the context of other global risks

Table C.7 summarizes responses to 15 potential global risks.

Table C.7: Summary statistics: the American public's perceptions of 15 potential global risks

| Potential risks | Mean perceived likelihood | Mean perceived impact | *N* |
|---|---|---|---|
| Failure to address climate change | 56% | 2.25 | 666 |
| Failure of regional/global governance | 55% | 2.46 | 652 |
| Conflict between major countries | 60% | 2.68 | 625 |
| Weapons of mass destruction | 49% | 3.04 | 645 |
| Large-scale involuntary migration | 57% | 2.65 | 628 |
| Spread of infectious diseases | 50% | 2.69 | 620 |
| Water crises | 54% | 2.90 | 623 |
| Food crises | 52% | 2.76 | 1073 |
| Harmful consequences of AI | 45% | 2.29 | 573 |
| Harmful consequences of synthetic biology | 45% | 2.33 | 630 |
| Cyber attacks | 68% | 2.85 | 650 |
| Terrorist attacks | 60% | 2.62 | 635 |
| Global recession | 56% | 2.61 | 599 |
| Extreme weather events | 65% | 2.73 | 613 |
| Natural disasters | 69% | 2.87 | 637 |

## C.4 Survey experiment: what the public considers AI, automation, machine learning, and robotics

We formally tested whether or not respondents think AI, automation, machine learning, and robotics are used in different applications. (See Appendix B for the survey question text.) For each technological application, we used an $F$-test to test whether any of terms randomly assigned to the respondents affect respondents' selecting that application. Because we ran 10 $F$-tests, we used the Bonferroni correction to control the familywise error rate. The Bonferroni correction rejected the null hypothesis at alpha level $\alpha/10$, instead of $\alpha$. For instance, to test whether the $F$-static is significant at the 5% level, we set the alpha level at $\alpha/10 = 0.005$. Our results (in Table C.8) show that except for social robots, respondents think that AI, automation, machine learning, and robotics are used in each of the applications presented in the survey.

Table C.8: Respondents distinguish between AI, automation, machine learning, and robotics

| Technological applications | *F*-statistic | *p*-value | Significant |
|---|---|---|---|
| Virtual assistants (e.g., Siri, Google Assistant, Amazon Alexa) | $F(3, 1996) = 18.12$ | <0.001 | Yes |
| Smart speakers (e.g., Amazon Echo, Google Home, Apple Homepod) | $F(3, 1996) = 24.76$ | <0.001 | Yes |
| Facebook photo tagging | $F(3, 1996) = 20.22$ | <0.001 | Yes |
| Google Search | $F(3, 1996) = 37.30$ | <0.001 | Yes |
| Recommendations for Netflix movies or Amazon ebooks | $F(3, 1996) = 33.69$ | <0.001 | Yes |
| Google Translate | $F(3, 1996) = 24.62$ | <0.001 | Yes |
| Driverless cars and trucks | $F(3, 1996) = 9.08$ | <0.001 | Yes |
| Social robots that can interact with humans | $F(3, 1996) = 1.05$ | 0.369 | No |
| Industrial robots used in manufacturing | $F(3, 1996) = 55.72$ | <0.001 | Yes |
| Drones that do not require a human controller | $F(3, 1996) = 9.68$ | <0.001 | Yes |

Next, we investigated the problem of respondents not selecting technological applications where it would be logical to pick them (e.g., not selecting industrial robots or social robots when presented with the term "robotics"). Our regression analysis shows that this type of non-response is correlated with respondents' inattention.

We used two measures as a proxy for inattention:

1. time to complete the survey
2. the absolute deviation from the median time to complete the survey.

Because the distribution of completion times is heavily skewed right, we used absolute deviation from the median, as opposed to the mean. The median is 13 minutes whereas the mean is 105 minutes. We incorporated the second measure because we suspected that people who took very little time *or* a very long time to complete the survey were inattentive.

We used three outcomes that measured non-response:

1. the number of items selected
2. not selecting "none of the above"
3. selecting items containing the word "robots" for respondents assigned to consider "robotics"

Using multiple regression, we showed that inattention predicts non-response measured by the three outcomes above (see Tables C.9, C.10, and C.11).

Table C.9: Correlation between survey completion time and number of selected items

| Variables | Coefficients (SEs) |
|---|---|
| (Intercept) | 3.58 (0.17)*** |
| Survey completion time (min) | 0.14 (0.01)*** |
| Absolute deviation from median survey completion time (min) | -0.14 (0.01)*** |
| Term: automation | 0.98 (0.22)*** |
| Term: machine learning | -0.09 (0.22) |
| Term: Robotics | -0.51 (0.20)* |
| $N = 2000$ | $F(5, 1994) = 47.47$; *p*-value: <0.001 |

Table C.10: Correlation between survey completion time and not selecting 'none of the above'

| Variables | Coefficients (SEs) |
|---|---|
| (Intercept) | 0.79 (0.02)*** |

| Variables | Coefficients (SEs) |
|---|---|
| Survey completion time (min) | 0.01 (<0.01)*** |
| Absolute deviation from median survey completion time (min) | -0.01 (<0.01)*** |
| Term: automation | 0.05 (0.02)* |
| Term: machine learning | -0.04 (0.02) |
| Term: Robotics | 0.04 (0.02) |
| $N$ = 2000 | $F$(5, 1994) = 13.16; $p$-value: <0.001 |

Table C.11: Correlation between survey completion time and selecting 'robots' when assigned the term 'robotics'

| Variables | Coefficients (SEs) |
|---|---|
| (Intercept) | 0.87 (0.06)*** |
| Survey completion time (min) | 0.06 (0.01)*** |
| Absolute deviation from median survey completion time (min) | -0.06 (0.01)*** |
| $N$ = 486 | $F$(2, 483) = 50.55; $p$-value: <0.001 |

## C.5 AI governance challenges: prioritizing governance challenges

We compared respondents' perceived likelihood of each governance challenge impacting large numbers of people in the U.S. with respondents' perceived likelihood of each governance challenge impacting large numbers of people around the world. (See Appendix B for the survey question text.) For each governance challenge, we used linear regression to estimate the difference between responses to the U.S. question and the world question.

Because we ran 13 tests, we used the Bonferroni correction to control the familywise error rate. In our case, the Bonferroni correction rejected the null hypothesis at alpha level $\alpha/13$, instead of $\alpha$. To test whether the differences are significant at the 5% level, we set the alpha level at $\alpha/13 = 0.004$. According to Table C.12, Americans perceive that all governance challenges, except for protecting data privacy and ensuring safe autonomous vehicles, are more likely to impact people around the world than in the U.S. specifically. In particular, Americans think that autonomous weapons are 7.6 percentage points more likely to impact people around the world than in the U.S.

Table C.12: Comparing perceived likelihood: in U.S. vs. around the world; each difference is the U.S. mean likelihood subtracted from the world mean likelihood

| Governance challenge | U.S. mean likelihood | Difference (SE) | $p$-value | Significant |
|---|---|---|---|---|
| Hiring bias | 59.8 | 2.5 (0.8) | 0.001 | Yes |
| Criminal justice bias | 55.6 | 2.5 (0.8) | 0.003 | Yes |
| Disease diagnosis | 60.4 | 2.1 (0.6) | 0.001 | Yes |
| Data privacy | 66.9 | 1.7 (0.6) | 0.010 | No |
| Autonomous vehicles | 61.8 | -0.7 (0.8) | 0.401 | No |
| Digital manipulation | 68.6 | 2.6 (0.7) | <0.001 | Yes |
| Cyber attacks | 66.2 | 3.2 (0.9) | <0.001 | Yes |
| Surveillance | 69.0 | 2.2 (0.7) | 0.002 | Yes |
| U.S.-China arms race | 60.3 | 3.0 (0.7) | <0.001 | Yes |
| Value alignment | 60.4 | 3.6 (0.7) | <0.001 | Yes |
| Autonomous weapons | 54.7 | 7.6 (0.8) | <0.001 | Yes |
| Technological unemployment | 62.3 | 2.3 (0.7) | <0.001 | Yes |
| Critical AI systems failure | 55.2 | 3.1 (0.8) | <0.001 | Yes |

To highlight the differences between the responses of demographic subgroups regarding issue importance, we created an additional graph (Figure C.1). Here, we subtracted the overall mean of perceived issue importance across all responses from each subgroup-governance challenge mean.[19] Table C.15 shows the results from a saturated regression predicting perceived issue importance using demographic variables, AI governance challenge, and interactions between the two types of variables.

Table C.13: Perception of AI governance challenges in the U.S.: summary statistics table

| Governance challenge | Mean likelihood | Mean issue importance | Product of likelihood and issue importance |
|---|---|---|---|
| Surveillance | 69% | 2.56 | 1.77 |
| Data privacy | 67% | 2.62 | 1.75 |
| Digital manipulation | 69% | 2.53 | 1.74 |
| Cyber attacks | 66% | 2.59 | 1.71 |
| Autonomous vehicles | 62% | 2.56 | 1.58 |
| Technological unemployment | 62% | 2.50 | 1.56 |
| Value alignment | 60% | 2.55 | 1.54 |
| Disease diagnosis | 60% | 2.52 | 1.52 |
| U.S.-China arms race | 60% | 2.52 | 1.52 |
| Hiring bias | 60% | 2.54 | 1.52 |
| Autonomous weapons | 55% | 2.58 | 1.42 |
| Criminal justice bias | 56% | 2.53 | 1.41 |
| Critical AI systems failure | 55% | 2.47 | 1.36 |

Table C.14: Perception of AI governance challenges in the world: summary statistics table

| Governance challenge | Mean likelihood | Mean issue importance | Product of likelihood and issue importance |
|---|---|---|---|
| Surveillance | 71% | 2.56 | 1.82 |
| Digital manipulation | 71% | 2.53 | 1.80 |
| Cyber attacks | 69% | 2.59 | 1.80 |
| Data privacy | 69% | 2.62 | 1.80 |
| Value alignment | 64% | 2.55 | 1.63 |
| Technological unemployment | 65% | 2.50 | 1.62 |
| Autonomous weapons | 62% | 2.58 | 1.61 |
| U.S.-China arms race | 63% | 2.52 | 1.60 |
| Hiring bias | 62% | 2.54 | 1.58 |
| Disease diagnosis | 63% | 2.52 | 1.58 |
| Autonomous vehicles | 61% | 2.56 | 1.56 |
| Criminal justice bias | 58% | 2.53 | 1.47 |
| Critical AI systems failure | 58% | 2.47 | 1.44 |

Table C.15: Results from a saturated regression predicting perceived issue importance using demographic variables, AI governance challenge, and interactions between the two types of variables; the coefficients for the interactions variables are not shown due to space constraints

| Variables | Coefficient (SEs) |
|---|---|
| (Intercept) | 2.25 (0.11)*** |

[19] Note that the perceived issue importance was measured on a four-point scale, where 0 meant "not at all important" and 3 meant "very important." We only mean-centered the outcomes; we did not standardize such that the outcomes have unit variance.

| Variables | Coefficient (SEs) |
|---|---|
| Age 38-53 | 0.11 (0.07) |
| Age 54-72 | 0.35 (0.06)*** |
| Age 73 and older | 0.44 (0.07)*** |
| Male | 0.02 (0.05) |
| Non-white | -0.01 (0.05) |
| Some college | 0.03 (0.07) |
| College+ | 0.15 (0.07)* |
| Employed (full- or part-time) | -0.09 (0.06) |
| Income $30-70K | 0.09 (0.08) |
| Income $70-100K | 0.13 (0.10) |
| Income more than $100K | -0.01 (0.10) |
| Prefer not to say income | 0.04 (0.08) |
| Democrat | 0.13 (0.07) |
| Independent/Other | 0.14 (0.07) |
| No religious affiliation | -0.04 (0.06) |
| Other religion | -0.05 (0.08) |
| Born-again Christian | 0.07 (0.07) |
| CS or engineering degree | -0.35 (0.10)*** |
| CS or programming experience | -0.01 (0.07) |
| Criminal justice bias | 0.05 (0.13) |
| Disease diagnosis | -0.06 (0.14) |
| Data privacy | 0.16 (0.13) |
| Autonomous vehicles | -0.07 (0.14) |
| Digital manipulation | -0.14 (0.15) |
| Cyber attacks | 0.05 (0.14) |
| Surveillance | <0.01 (0.15) |
| U.S.-China arms race | 0.04 (0.13) |
| Value alignment | -0.06 (0.13) |
| Autonomous weapons | 0.06 (0.14) |
| Technological unemployment | -0.12 (0.14) |
| Critical AI systems failure | -0.27 (0.15) |
| $N$ = 10000 observations, 2000 respondents | $F(259,1999) = 3.36$; $p$-value: <0.001 |

## C.6 Trust in various actors to develop and manage AI in the interest of the public

Table C.16 displays the mean level of trust the public expresses in various actors to develop and manage AI in the interest of the public.

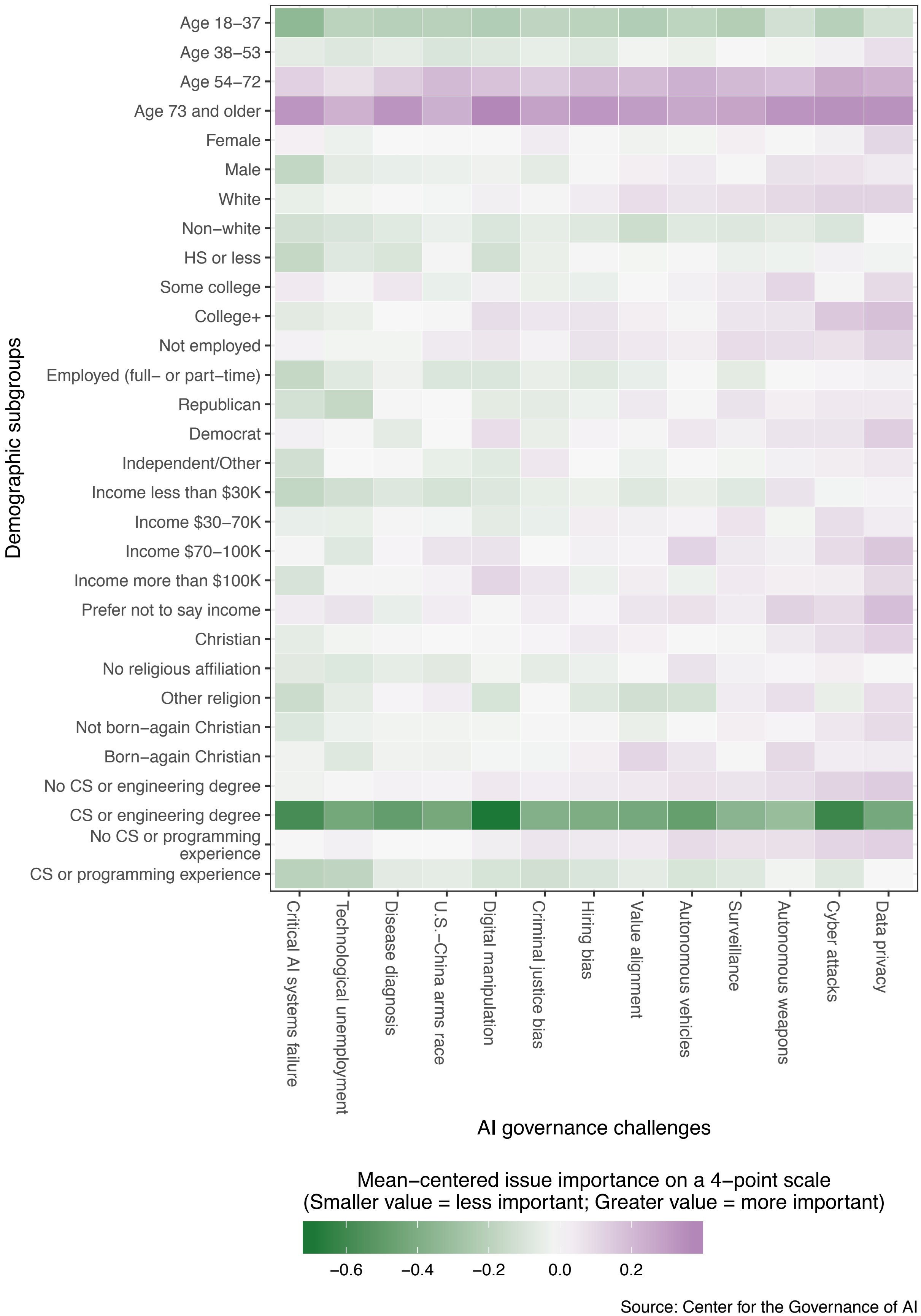


Figure C.1: AI governance challenges: issue importance by demographic subgroups

## C.7 Survey experiment: comparing perceptions of U.S. vs. China AI research and development

A substantial percentage of respondents selected "I don't know" when answering this survey question. (See Appendix B for the survey question text.) Our regression analysis shows that there is a small but statistically significant difference between respondents' perception of R&D in the U.S. as compared to in China, as seen in Tables C.19 and C.20.

Table C.17: Survey experiment attrition check: comparing U.S. and China's AI research and development

| Experimental condition | Percent DK/missing | Percent DK | Percent missing |
|---|---|---|---|
| China | 26.48 | 26.48 | 0 |
| U.S. | 22.77 | 22.77 | 0 |

Table C.18: Survey experiment attrition check: comparing U.S. and China's AI research and development

| Variables | Coefficients (SEs) |
|---|---|
| (Intercept) | 0.27 (0.01)*** |
| U.S. | -0.04 (0.02) |
| $N = 2000$ | $F(1, 1998) = 3.12$; $p$-value: 0.078 |

Table C.19: Survey experiment results: comparing U.S. and China's AI research and development

| Variables | Coefficients (SEs) |
|---|---|
| (Intercept) | 1.74 (0.02)*** |
| U.S. | -0.08 (0.03)* |
| $N = 2000$ | $F(1, 1998) = 6.58$; $p$-value: 0.01 |

Table C.20: Survey experiment results: comparing U.S. and China's AI research and development (controlling for DK/missing responses)

| Variables | Coefficients (SEs) |
|---|---|
| (Intercept) | 1.74 (0.02)*** |
| U.S. | -0.08 (0.03)** |
| $N = 2000$ | $F(3, 1996) = 6.14$; $p$-value: $<0.001$ |

Table C.16: Trust in various actors to develop and manage AI in the interest of the public: mean responses

| Actors | Trust to develop AI | Trust to manage AI |
|---|---|---|
| U.S. military | 1.56 (MOE: +/-0.07); N = 638 | |
| U.S. civilian government | 1.16 (MOE: +/-0.07); N = 671 | |
| NSA | 1.28 (MOE: +/-0.07); N = 710 | |
| FBI | 1.21 (MOE: +/-0.08); N = 656 | |
| CIA | 1.21 (MOE: +/-0.07); N = 730 | |
| U.S. federal government | | 1.05 (MOE: +/-0.07); N = 743 |
| U.S. state governments | | 1.05 (MOE: +/-0.07); N = 713 |
| NATO | 1.17 (MOE: +/-0.06); N = 695 | |
| Intergovernmental research organizations (e.g., CERN) | 1.42 (MOE: +/-0.07); N = 645 | 1.27 (MOE: +/-0.06); N = 747 |
| International organizations | | 1.10 (MOE: +/-0.06); N = 827 |
| UN | | 1.06 (MOE: +/-0.06); N = 802 |
| Tech companies | 1.44 (MOE: +/-0.07); N = 674 | 1.33 (MOE: +/-0.07); N = 758 |
| Google | 1.34 (MOE: +/-0.08); N = 645 | 1.20 (MOE: +/-0.07); N = 767 |
| Facebook | 0.85 (MOE: +/-0.07); N = 632 | 0.91 (MOE: +/-0.07); N = 741 |
| Apple | 1.29 (MOE: +/-0.07); N = 697 | 1.20 (MOE: +/-0.07); N = 775 |
| Microsoft | 1.40 (MOE: +/-0.08); N = 597 | 1.24 (MOE: +/-0.07); N = 771 |
| Amazon | 1.33 (MOE: +/-0.07); N = 685 | 1.24 (MOE: +/-0.07); N = 784 |
| Non-profit (e.g., OpenAI) | 1.44 (MOE: +/-0.07); N = 659 | |
| University researchers | 1.56 (MOE: +/-0.07); N = 666 | |
| Non-government scientific organization (e.g., AAAI) | | 1.35 (MOE: +/-0.06); N = 792 |
| Partnership on AI | | 1.35 (MOE: +/-0.06); N = 780 |

## C.8 Survey experiment: U.S.-China arms race

We checked that "don't know" or missing responses to both statements are not induced by the information treatments. (See Appendix B for the survey experiment text.) Next, we examined the correlation between responses to the two statements using a 2D bin count graph. The overall Pearson correlation coefficient is -0.05 but there exists considerable variation by experimental condition.

Table C.21: Survey experiment attrition check: agreement with statement that U.S. should invest more in AI military capabilities

| Experimental condition | Percent DK/missing | Percent DK | Percent missing |
|---|---|---|---|
| Control | 13.53 | 13.53 | 0 |
| Treatment 1: Pro-nationalist | 12.28 | 12.28 | 0 |
| Treatment 2: Risks of arms race | 12.58 | 12.58 | 0 |
| Treatment 3: One common humanity | 13.62 | 13.62 | 0 |

Table C.22: Survey experiment attrition check: agreement with statement that U.S. should invest more in AI military capabilities

| Variables | Coefficients (SEs) |
|---|---|
| (Intercept) | 0.13 (0.02)*** |
| Treatment 1: Pro-nationalist | <0.01 (0.02) |
| Treatment 2: Risks of arms race | -0.01 (0.02) |
| Treatment 3: One common humanity | >-0.01 (0.02) |
| $N$ = 2000 | $F$(3, 1996) = 0.08; $p$-value: 0.972 |

Table C.23: Survey experiment attrition check: agreement with statement that U.S. should work hard to cooperate with China to avoid dangers of AI arms race

| Experimental condition | Percent DK/missing | Percent DK | Percent missing |
|---|---|---|---|
| Control | 14.12 | 13.92 | 0.2 |
| Treatment 1: Pro-nationalist | 14.26 | 14.26 | 0.0 |
| Treatment 2: Risks of arms race | 12.78 | 12.78 | 0.0 |
| Treatment 3: One common humanity | 12.80 | 12.80 | 0.0 |

Table C.24: Survey experiment attrition check: agreement with statement that U.S. should work hard to cooperate with China to avoid dangers of AI arms race

| Variables | Coefficients (SEs) |
|---|---|
| (Intercept) | 0.14 (0.02)*** |
| Treatment 1: Pro-nationalist | 0.02 (0.02) |
| Treatment 2: Risks of arms race | -0.01 (0.02) |
| Treatment 3: One common humanity | -0.02 (0.02) |
| $N$ = 2000 | $F$(3, 1996) = 0.76; $p$-value: 0.516 |

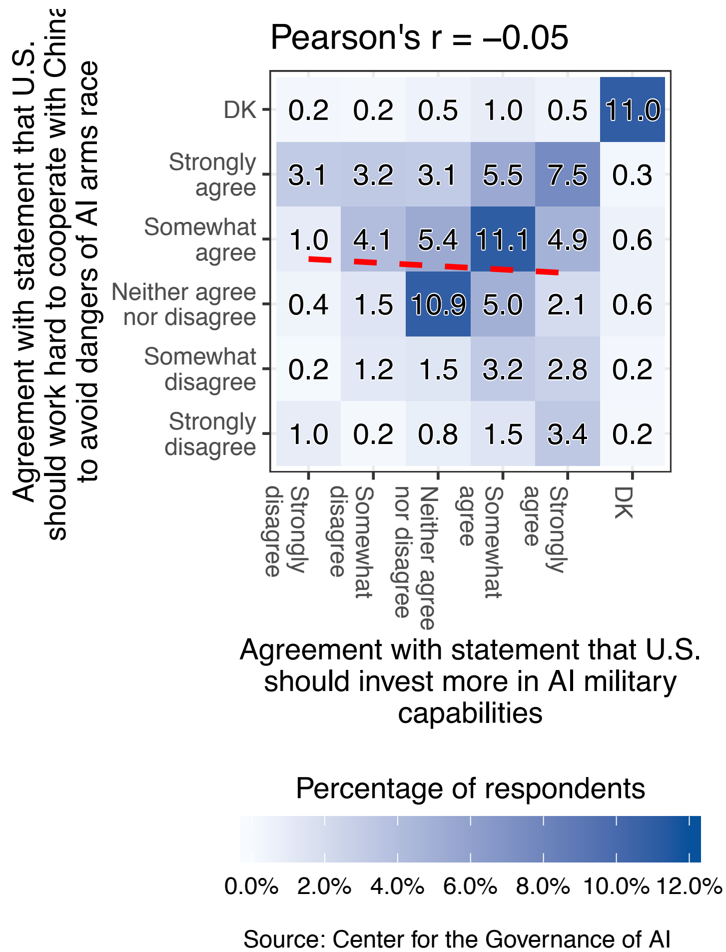


Figure C.2: Correlation between responses to the two statements from survey experiment

Table C.25: Correlation between responses to the two statements

| Experimental condition | Pearson correlation |
|---|---|
| Overall | -0.05 |
| Control | -0.06 |
| Treatment 1: Pro-nationalist | -0.03 |
| Treatment 2: Risks of arms race | -0.12 |
| Treatment 3: One common humanity | -0.01 |

## C.9 Trend across time: job creation or job loss

There are many "don't know" responses to this survey question (see Appendix B for the survey question text). Nevertheless, "don't know" or missing responses are not affected by the experimental future time framing. $F$-tests reveal that there are no differences in responses to the three future time frames, as seen in Table C.30.

Table C.26: Survey experiment attrition check: future time frame

| Experimental condition | Percent DK/missing | Percent DK | Percent missing |
|---|---|---|---|
| No time frame | 24.59 | 24.38 | 0.21 |
| 10 years | 25.49 | 25.49 | 0.00 |
| 20 years | 26.16 | 25.96 | 0.20 |
| 50 years | 24.17 | 24.17 | 0.00 |

Table C.27: Survey experiment attrition check: future time frame

| Variables | Coefficients (SEs) |
|---|---|
| (Intercept) | 0.25 (0.02)*** |
| 10 years | 0.01 (0.03) |
| 20 years | 0.02 (0.03) |
| 50 years | -0.01 (0.03) |
| $N$ = 2000 | $F(3, 1996) = 0.34$; $p$-value: 0.795 |

Table C.28: Survey experiment results: future time frame

| Variables | Coefficients (SEs) |
|---|---|
| (Intercept) | -0.52 (0.06)*** |
| 10 years | -0.15 (0.08) |
| 20 years | -0.12 (0.08) |
| 50 years | -0.06 (0.08) |
| $N$ = 2000 | $F(3, 1996) = 1.48$; $p$-value: 0.219 |

Table C.29: Survey experiment results: future time frame (controlling for DK/missing responses)

| Variables | Coefficients (SEs) |
|---|---|
| (Intercept) | -0.52 (0.06)*** |
| 10 years | -0.15 (0.08) |

| Variables | Coefficients (SEs) |
|---|---|
| 20 years | -0.12 (0.08) |
| 50 years | -0.06 (0.08) |
| $N$ = 2000 | $F$(7, 1992) = 1.68; $p$-value: 0.108 |

Table C.30: Testing coefficients for time frames are equivalent

| Tests | $F$-statistic | $p$-value |
|---|---|---|
| 10 years = 20 years | $F$(1, 1992) = 0.15 | 0.70 |
| 10 years = 50 years | $F$(1, 1992) = 1.41 | 0.24 |
| 20 years = 50 years | $F$(1, 1992) = 0.63 | 0.43 |

## C.10 High-level machine intelligence: forecasting timeline

Figure C.3 displays the mean predicted the likelihood of high-level machine intelligence for each year by demographic subgroup. Figure C.4 displays the median predicted probability of high-level machine intelligence for each year by demographic subgroup.

## C.11 Support for developing high-level machine intelligence

We examined the correlation between support for developing AI and support for developing high-level machine intelligence using a 2D bin count graph. The overall Pearson correlation coefficient is 0.61, according to Figure C.5.

The mean level of support for developing high-level machine intelligence, compared with the mean level of support for developing AI, is 0.24 points (MOE = +/- 0.04) lower on a five-point scale (two-sided $p$-value $< 0.001$), as shown in Table C.31.

Table C.32 displays the regression results used to produce Figure 6.5.

To identify subgroups that have diverging attitudes toward high-level machine intelligence versus AI, we performed multiple regression using both the demographic subgroups variables *and* respondents' support for developing AI as predictors. The support for developing high-level machine intelligence outcome variable was standardized such that it has mean 0 and unit variance. The results are shown in Table C.33.

After controlling for one's support for developing AI, significant predictors correlated with support for developing high level machine intelligence, including:

- Being a member of the Silent Generation (versus being a Millennial/post-Millennial)
- Having CS or programming experience (versus not having such experience)
- Having a high school degree or less (versus having at least a four-year college degree)

Table C.31: Difference between support for developing AI and support for developing high-level machine intelligence

| Variables | Coefficients (SEs) |
|---|---|
| (Intercept) | 0.25 (0.03)*** |
| High-level machine intelligence | -0.24 (0.02)*** |
| $N$ = 2000 | |

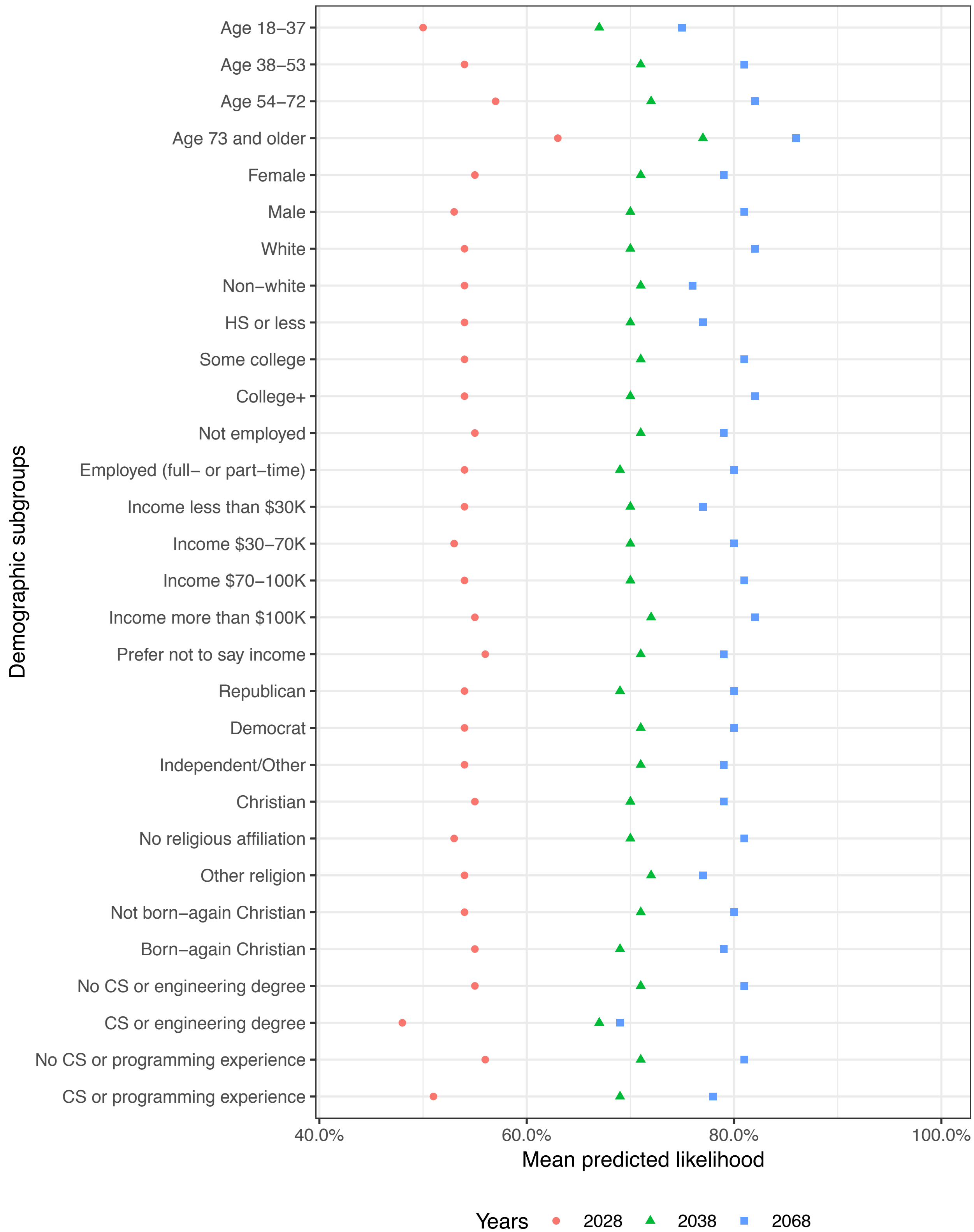


Figure C.3: Mean predicted likelihood of high-level machine intelligence for each year by demographic subgroup

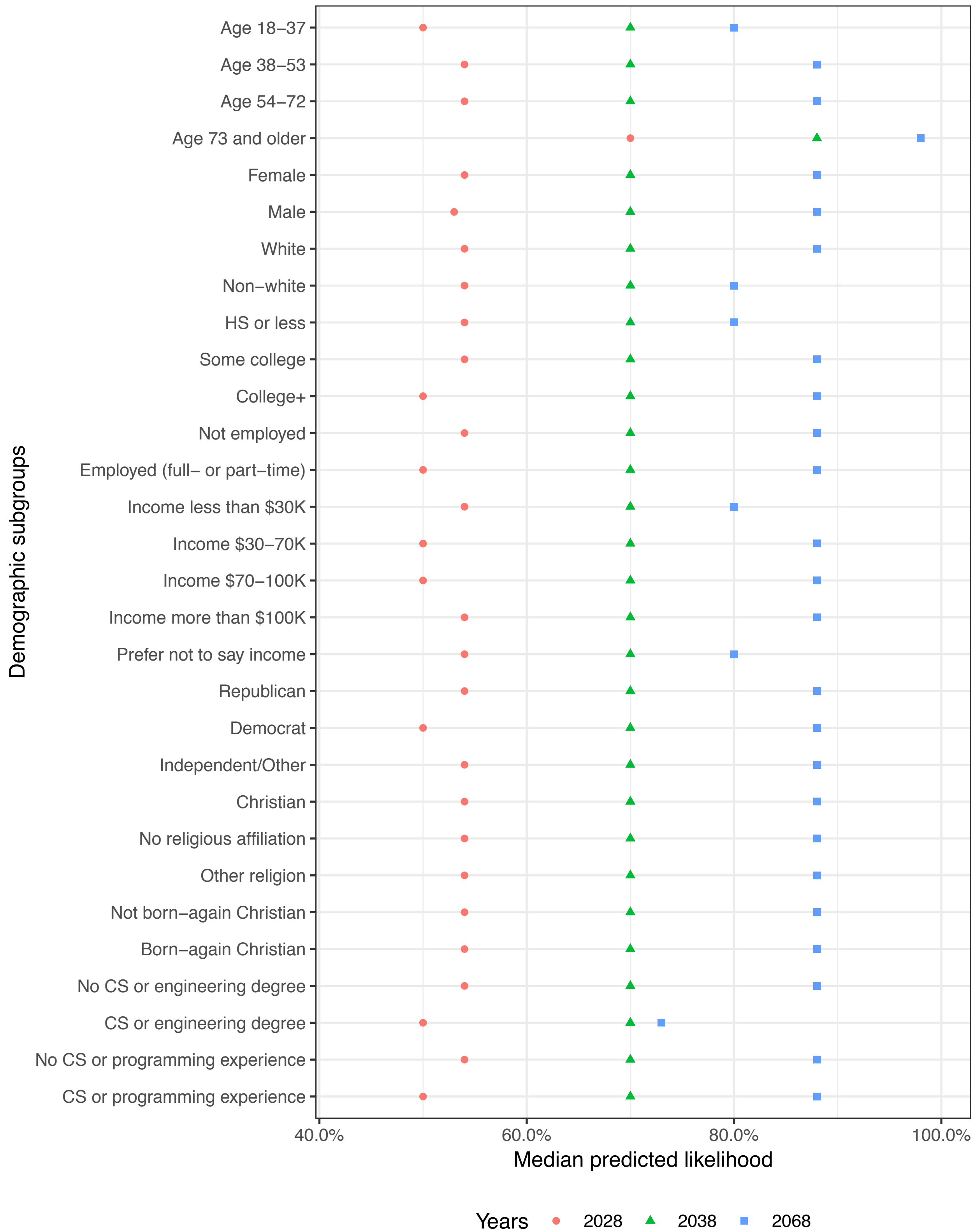


Figure C.4: Median predicted likelihood of high-level machine intelligence for each year by demographic subgroup

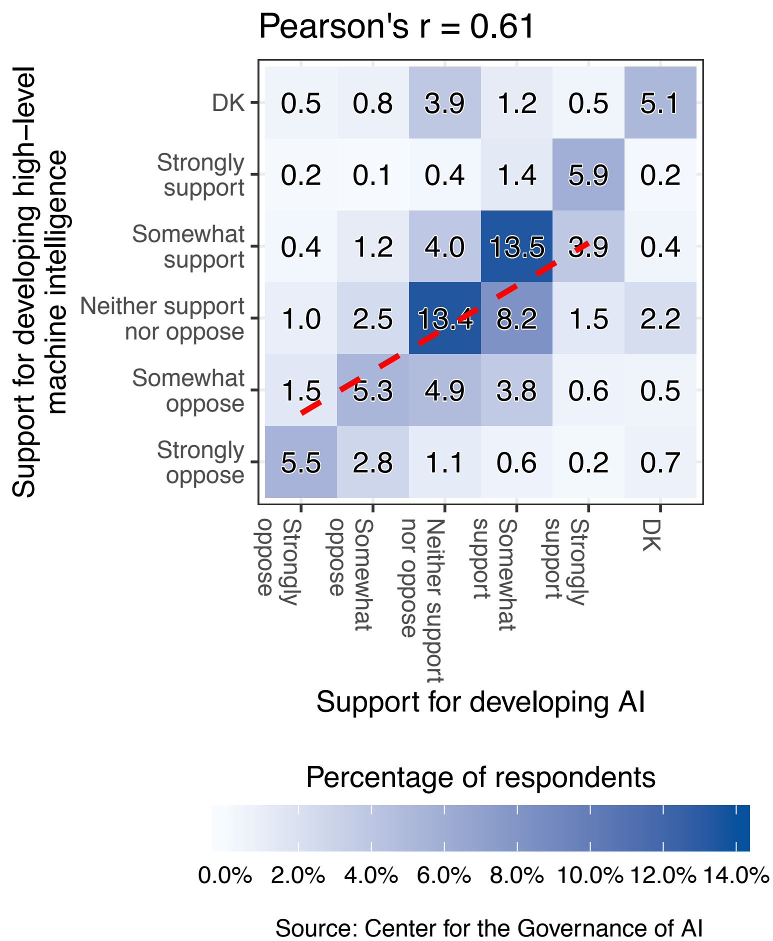


Figure C.5: Correlation between support for developing AI and support for developing high-level machine intelligence

Table C.32: Predicting support for developing high-level machine intelligence using demographic characteristics: results from a multiple linear regression that includes all demographic variables; outcome standardized to have mean 0 and unit variance

| Variables | Coefficients (SEs) |
|---|---|
| (Intercept) | -0.25 (0.09)** |
| Age 38-53 | -0.12 (0.06) |
| Age 54-72 | -0.03 (0.06) |
| Age 73 and older | 0.12 (0.10) |
| Male | 0.18 (0.05)*** |
| Non-white | 0.01 (0.05) |
| Some college | -0.04 (0.06) |
| College+ | <0.01 (0.07) |
| Employed (full- or part-time) | 0.09 (0.05) |
| Democrat | 0.11 (0.07) |
| Independent/Other | -0.13 (0.07)* |
| Income $30-70K | -0.01 (0.07) |
| Income $70-100K | 0.09 (0.09) |
| Income more than $100K | 0.19 (0.09)* |
| Prefer not to say income | <0.01 (0.08) |
| No religious affiliation | 0.09 (0.06) |
| Other religion | 0.06 (0.08) |
| Born-again Christian | -0.07 (0.06) |
| CS or engineering degree | <0.01 (0.10) |
| CS or programming experience | 0.36 (0.06)*** |
| $N$ = 2000 | $F$(19,1980) = 7.27; $p$-value: <0.001 |

Table C.33: Predicting support for developing high-level machine intelligence using demographic characteristics: results from a multiple linear regression that includes all demographic variables and respondents' support for developing AI; outcome standardized to have mean 0 and unit variance

| Variables | Coefficients (SEs) |
|---|---|
| (Intercept) | -0.23 (0.08)** |
| Age 38-53 | -0.02 (0.05) |
| Age 54-72 | 0.09 (0.05) |
| Age 73 and older | 0.22 (0.09)* |
| Male | 0.08 (0.04) |
| Non-white | 0.02 (0.05) |
| Some college | -0.04 (0.05) |
| College+ | -0.11 (0.06) |
| Employed (full- or part-time) | 0.08 (0.04) |
| Democrat | -0.02 (0.06) |
| Independent/Other | -0.10 (0.05) |
| Income $30-70K | -0.01 (0.06) |
| Income $70-100K | 0.01 (0.07) |
| Income more than $100K | 0.08 (0.07) |
| Prefer not to say income | 0.09 (0.07) |
| No religious affiliation | -0.02 (0.05) |
| Other religion | -0.03 (0.07) |
| Born-again Christian | -0.05 (0.05) |

| Variables | Coefficients (SEs) |
|---|---|
| CS or engineering degree | -0.03 (0.07) |
| CS or programming experience | 0.17 (0.05)*** |
| Support for developing AI | 0.58 (0.02)*** |
| $N$ = 2000 | $F(20,1979) = 54.15$; $p$-value: $<0.001$ |

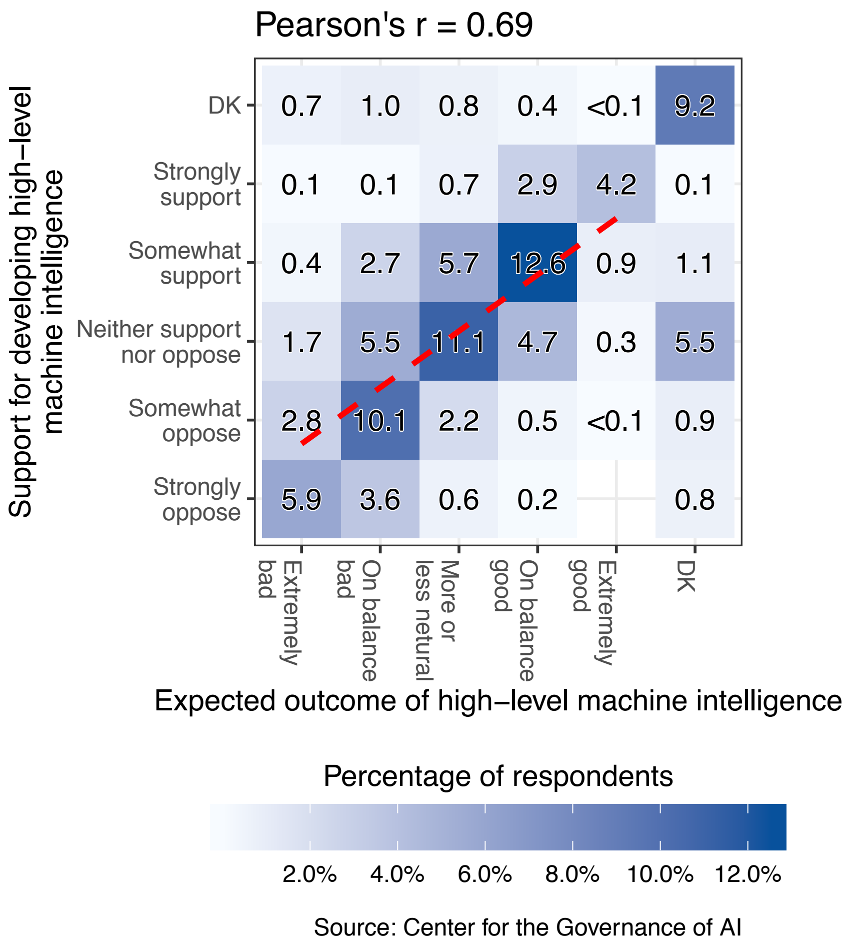


Figure C.6: Correlation between expected outcome and support for developing high-level machine intelligence

## C.12 Expected outcome of high-level machine intelligence

We examined the correlation between respondents' expected outcome of high-level machine intelligence and support for developing high-level machine intelligence using a 2D bin count graph. The overall Pearson correlation coefficient is 0.69, as seen in Figure C.6.

# About us

## About the Center for the Governance of AI

The Center for the Governance of AI, housed at the Future of Humanity Institute, University of Oxford, strives to help humanity capture the benefits and mitigate the risks of artificial intelligence. Our focus is on the political challenges arising from transformative AI: advanced AI systems whose long-term impacts may be as profound as the industrial revolution. The Center seeks to guide the development of AI for the common good by conducting research on important and neglected issues of AI governance, and advising decision makers on this research through policy engagement.

The Center produces research which is foundational to the field of AI governance, for example mapping crucial considerations to direct the research agenda, or identifying distinctive features of the transition to transformative AI and corresponding policy considerations. Our research also addresses more immediate policy issues, such as malicious use and China's AI strategy. Our work takes a cross-disciplinary approach, looking at transformative AI through the lenses of e.g. international security, the history of technology development, law and public opinion.

In addition to research, the Center for the Governance of AI is active in international policy circles, and actively advises governments and industry leaders on AI strategy. The Center for the Governance of AI researchers has spoken at the NIPS and AAAI/ACM conferences, and at events involving the German Federal Foreign Office, the European Commission, the European Parliament, the UK House of Lords, the U.S. Congress, and others.

The Center's papers and reports are available at https://www.governance.ai.

## About the Future of Humanity Institute

The Future of Humanity Institute, University of Oxford, is a multidisciplinary research institute at the University of Oxford. Academics at FHI bring the tools of mathematics, philosophy and social sciences to bear on big-picture questions about humanity and its prospects. The Institute is led by Founding Director Professor Nick Bostrom. Humanity has the potential for a long and flourishing future. Our mission is to shed light on crucial considerations that might shape that future. More information at http://www.fhi.ox.ac.uk/.

## References

Bech, Mickael, and Morten Bo Kristensen. 2009. "Differential Response Rates in Postal and Web-Based Surveys in Older Respondents." *Survey Research Methods* 3 (1): 1–6.

Bronner, Fred, and Ton Kuijlen. 2007. "The Live or Digital Interviewer - a Comparison Between CASI, CAPI and CATI with Respect to Differences in Response Behaviour." *International Journal of Market Research* 49 (2): 167–90.

Cadell, Cate. 2018. "Search engine Baidu becomes first China firm to join U.S. AI ethics group." Reuters. https://perma.cc/DC63-9CFT.

China Institute for Science and Technology Policy at Tsinghua University. 2018. "China AI Development Report 2018." China Institute for Science; Technology Policy at Tsinghua University.

Coppock, Alexander, Emily Ekins, David Kirby, and others. 2018. "The Long-Lasting Effects of Newspaper Op-Eds on Public Opinion." *Quarterly Journal of Political Science* 13 (1): 59–87.

Cotton-Barratt, Owen, Sebastian Farquhar, John Halstead, Stefan Schubert, and Andrew Snyder-Beattie. 2016. "Global Catastrophic Risks 2016." Technical report. Global Challenges Foundation; the Global Priorities Project. https://perma.cc/54M8-JWSS.

Crawford, Kate. 2016. "Artificial Intelligence's White Guy Problem." *New York Times*, June 25. https://perma.cc/6C9S-5FNP.

Dafoe, Allan. 2018. "AI Governance: A Research Agenda." Governance of AI Program, Future of Humanity Institute, University of Oxford. https://perma.cc/L35A-7W2X.

Ding, Jeffrey. 2018. "Deciphering China's AI Dream." Technical report. Governance of AI Program, Future of Humanity Institute, University of Oxford. https://perma.cc/9BED-QGTU.

Eurobarometer. 2017. "Special Eurobarometer 460: Attitudes Towards the Impact of Digitisation and Automation on Daily Life." Eurobarometer. https://perma.cc/9FRT-ADST.

Funk, Cary. 2017. "Real Numbers: Mixed Messages About Public Trust in Science." *Issues in Science and Technology* 34 (1). https://perma.cc/UF9P-WSRL.

Funk, Cary, and Lee Rainie. 2015. "Public and Scientists' Views on Science and Society." Survey report. Pew Research Center. https://perma.cc/9XSJ-8AJA.

Goldfarb, Avi, and Daniel Trefler. 2018. "AI and International Trade." In *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press. https://www.nber.org/chapters/c14012.pdf.

Grace, Katja, John Salvatier, Allan Dafoe, Baobao Zhang, and Owain Evans. 2018. "Viewpoint: When Will AI Exceed Human Performance? Evidence from AI Experts." *Journal of Artificial Intelligence Research* 62: 729–54. https://doi.org/10.1613/jair.1.11222.

Graham, Edward. 2018. "Views on Automation's U.S. Workforce Impact Highlight Demographic Divide." Morning Consult. https://perma.cc/544D-WRUM.

Ipsos and Reuters. 2018. "Facebook Poll 3.23.18." Survey report. Ipsos; Reuters. https://perma.cc/QGH5-S3KE.

Ipsos MORI. 2018. "Public Views of Machine Learning: Findings from Public Research and Engagement Conducted on Behalf of the Royal Society." Survey report. The Royal Society. https://perma.cc/79FE-TEHH.

IT Juzi and Tencent Institute. 2017. "2017 China-U.S. AI Venture Capital State and Trends Research Report." Technical report. IT Juzi; Tencent Institute. https://perma.cc/AMN6-DZUV.

Kai-Fu, Lee. 2018. *AI Superpowers: China, Silicon Valley and the New World Order*. New York: Houghton Mifflin Harcourt.

Kania, Elsa. 2018. "China's Strategic Ambiguity and Shifting Approach to Lethal Autonomous Weapons Systems." Lawfare. https://perma.cc/K68L-YF7Q.

Lin, Winston, and Donald P Green. 2016. "Standard Operating Procedures: A Safety Net for Pre-Analysis Plans." *PS:*

*Political Science & Politics* 49 (3): 495–500.

McCorduck, Pamela. 2004. *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence*. New York: A K Peters/CRC Press.

Morning Consult. 2017. "National Tracking Poll 170401." Survey report. Morning Consult. https://perma.cc/TBJ9-CB5K.

Müller, Vincent C, and Nick Bostrom. 2014. "Future Progress in Artificial Intelligence: A Poll Among Experts." *AI Matters* 1 (1). ACM: 9–11. https://perma.cc/ZY22-E5FF.

Nagelhout, Gera E, Marc C Willemsen, Mary E Thompson, Geoffrey T Fong, Bas Van den Putte, and Hein de Vries. 2010. "Is Web Interviewing a Good Alternative to Telephone Interviewing? Findings from the International Tobacco Control (Itc) Netherlands Survey." *BMC Public Health* 10 (1): 351.

Nedelkoska, Ljubica, and Glenda Quintini. 2018. "Automation, Skills Use and Training." Working Papers No. 202. Organisation for Economic Co-operation; Development. https://doi.org/10.1787/2e2f4eea-en.

News, Bloomberg. 2018. "China Calls for Borderless Research to Promote AI Development." Bloomberg News. https://www.bloomberg.com/news/articles/2018-09-17/china-calls-for-borderless-research-to-promote-ai-development.

Northeastern University and Gallup. 2018. "Optimism and Anxiety: Views on the Impact of Artificial Intelligence and Higher Education's Response." Survey report. Northeastern University; Gallup. https://perma.cc/57NW-XCQN.

Nosek, Brian A, Charles R Ebersole, Alexander C DeHaven, and David T Mellor. 2018. "The Preregistration Revolution." *Proceedings of the National Academy of Sciences* 115 (11): 2600–2606.

Paluck, Elizabeth Levy, Paul Lagunes, Donald P Green, Lynn Vavreck, Limor Peer, and Robin Gomila. 2015. "Does Product Placement Change Television Viewers' Social Behavior?" *PloS One* 10 (9): e0138610.

Pinker, Steven. 2018. *Enlightenment Now: The Case for Reason, Science, Humanism, and Progress*. New York: Penguin.

Reinhart, RJ. 2018. "Most Americans Already Using Artificial Intelligence Products." Survey report. Gallup. https://perma.cc/RVY5-WP9W.

Rosling, Hans, Anna Rosling Rönnlund, and Ola Rosling. 2018. *Factfulness: Ten Reasons We're Wrong About the World–and Why Things Are Better Than You Think*. New York: Flatiron Books.

Sandberg, Anders, and Nick Bostrom. 2008. "Global Catastrophic Risks Survey." Technical report #2008-1. Future of Humanity Institute, Oxford University. https://perma.cc/TA97-KD3Z.

Shin, Eunjung, Timothy P Johnson, and Kumar Rao. 2012. "Survey Mode Effects on Data Quality: Comparison of Web and Mail Modes in a Us National Panel Survey." *Social Science Computer Review* 30 (2): 212–28.

Smith, Aaron. 2018. "Public Attitudes Toward Technology Companies." Survey report. Pew Research Center. https://perma.cc/KSN6-6FRW.

Smith, Aaron, and Monica Anderson. 2016. "Automation in Everyday Life." Pew Research Center. https://perma.cc/WU6B-63PZ.

Walsh, Toby. 2018. "Expert and Non-Expert Opinion About Technological Unemployment." *International Journal of Automation and Computing* 15 (5): 637–42. https://doi.org/10.1007/s11633-018-1127-x.

West, Darrell M. 2018a. "Brookings Survey Finds Divided Views on Artificial Intelligence for Warfare, but Support Rises If Adversaries Are Developing It." Survey report. Brookings Institution. https://perma.cc/3NJV-5GV4.

———. 2018b. "Brookings survey finds worries over AI impact on jobs and personal privacy, concern U.S. will fall behind China." Survey report. Brookings Institution. https://perma.cc/HY38-3GCC.

World Economic Forum. 2018. "The Global Risks Report 2018: 13th Edition." World Economic Forum. https://perma.cc/8XM8-LKEN.

Zwetsloot, Remco, Helen Toner, and Jeffrey Ding. 2018. "Beyond the Ai Arms Race." *Foreign Affairs* November. https://perma.cc/CN76-6LG8.