

The 2018 U.S. AI Public Opinion Survey

Baobao Zhang and Allan Dafoe

June 27, 2018

Contents

Introduction	2
Acknowledgments	2
Executive summary	2
Analysis of survey data	3
General questions	3
Survey experiment: what the public considers AI, automation, machine learning, and robotics	3
Procedure	3
Results and discussion	4
Support for developing AI	5
Procedure	5
Results and discussions	6
Demographic predictors of support for developing AI	6
Survey experiment: AI and/or robots should be carefully managed	9
Procedure	9
Results and discussion	10
Comparison with Eurobarometer data	11
AI governance challenges	12
Prioritizing governance challenges	12
Procedure	12
Results and discussion	13
Demographic predictors of issue importance	15
Trust of actors to develop and manage AI	19
Procedure	19
Results and discussion	19
AI policy and China	21
Survey experiment: comparing perceptions of U.S. vs. China AI research and development . .	21
Procedure	21
Results and discussion	21
Survey experiment: U.S.-China arms race	23
Procedure	23
Results and discussion	23
Issue areas for possible U.S.-China cooperation	26
Procedure	26
Results and discussion	27
Trend across time: job creation or job loss	27
Survey experiment: future time frame	28
Procedure	28
Results and discussion	28
Extending the time trend	30
High-level machine intelligence	31
Forecasting timeline	31
Results and discussion	31
Support for developing high-level machine intelligence	32

Procedure	32
Results and discussion	32
Demographic predictors of support for developing high-level machine intelligence . . .	33
Expected outcome of high-level machine intelligence	36
Procedure	36
Results and discussion	36
Comparing perceptions of global risks	37
Procedure	37
Results and discussion	37
Appendix: survey test questions	39
AI governance challenges	39
Description of actors	40
U.S.-China arms race survey experiment treatments	40
Nationalism treatment	40
War risks treatment	40
Common humanity treatment	41
Global risks	41
Appendix: methodology	41
References	42

Introduction

This report is based on findings from a nationally representative survey conducted by the Governance of AI Program at the University of Oxford using the survey firm YouGov. The survey was conducted between June 6 and June 14, 2018; 2,000 American adults (18+) completed the survey. The Methodology section in the Appendix provides further details regarding the details of the data collection and analysis process.¹

Acknowledgments

We would like to thank the following for giving us useful feedback on this project: Michael Page, Jeffrey Ding, Brian Tse, Remco Zwetsloot, and the YouGov Team (Marissa Shih and Sam Luks).

Executive summary

This report explores Americans' attitudes towards artificial intelligence (AI), an emerging technology that could impact many aspects of society. We highlight some key findings from our study:

- In Americans' minds, the terms AI, machine learning, automation, and robotics have distinct meanings. The public's understanding of these terms has nuance, but is often inaccurate. For instance, a majority of Americans do not think Facebook photo tagging, Google Search, Netflix or Amazon recommendations, or Google Translate use AI or machine learning.
- Americans express a mixed opinion about developing AI. A substantial minority of Americans (41%) support developing AI while a smaller minority (22%) oppose developing AI.

¹Unless otherwise specified: Heteroscedasticity-consistent standard errors are reported in parentheses after point estimates. 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.

- Demographic subgroups that indicate greater support for developing AI include younger Americans, males, Democrats, those with high income, those with computer science or programming experience, and those who do not identify as Christian.
- An overwhelming majority of Americans (82%) agree that robots and/or AI should be carefully managed. Americans' response is on par with that of European respondents.
- Americans consider all major governance challenges associated with AI to be important for governments and tech companies to manage. Governance challenges perceived to be the most likely to impact Americans within the next decade and most important include:
 1. protecting data privacy
 2. preventing AI cyber attacks against governments, companies, organizations, and individuals
 3. preventing AI-assisted surveillance from violating privacy and civil liberties
 4. preventing AI from being used to spread fake and harmful content online
- Americans do not have a great deal of confidence in the U.S. government, international organizations, corporations or other actors to develop or manage AI in the best interest of the public. The U.S. military and university researchers are perceived to be the most trusted groups to develop AI. The public places the most trust in Partnership on AI and non-governmental scientific organizations (e.g., AAAI) to manage the development and use of AI.
- Americans, in general, are more supportive of the U.S. investing more in AI military capabilities *and* of 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 decreases support for the U.S. investing more in AI military capabilities by 27%. But other experimental information treatments fail to change Americans' policy preferences.
- Respondents predict that high-level machine intelligence will arrive sooner than AI researchers forecast. The median respondent predicts there is 54% probability of high-level machine intelligence by 2028, 70% probability by 2038, and 88% probability by 2068.
- Americans express mixed support for developing high-level machine intelligence. About one-third of Americans (32%) somewhat or strongly support the development of AI while 27% somewhat or strongly oppose the development of AI.
- Americans are ambivalent about high-level machine intelligence, and lean towards thinking it will be on balance harmful. Twenty-two percent think the technology will be "on balance bad," and 12% that it would be "extremely bad, possibly human extinction." By contrast, 21% think it will be "on balance good," and 5% think it will be "extremely good."

Analysis of survey data

General questions

The first set of questions in the survey helps us explain Americans' most basic understanding of AI and their general attitudes towards the technology.

Survey experiment: what the public considers AI, automation, machine learning, and robotics

Procedure

We use a survey experiment to understand what the public considers AI, automation, machine learning, and robotics. Each respondent is randomly assigned to consider one of these terms; he or she is asked:

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

Respondents consider the following technologies:

- 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

Results and discussion

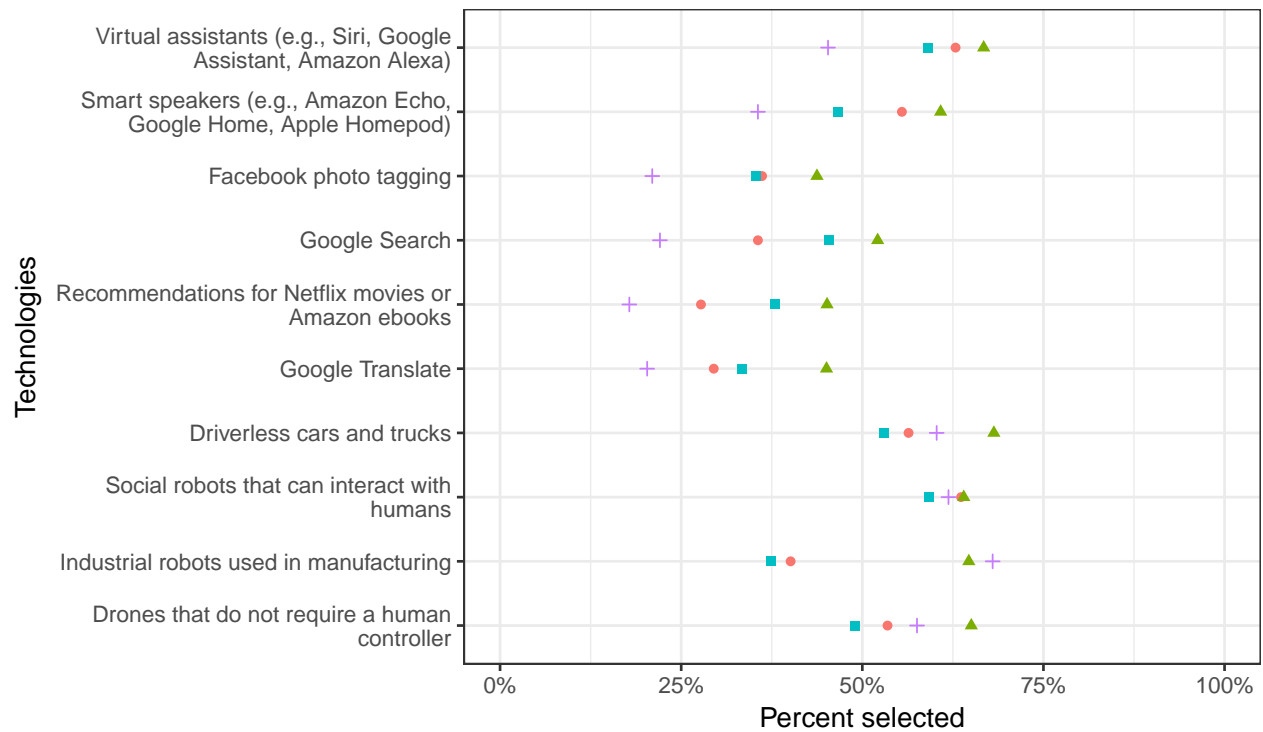
The figure below shows the percent of respondents who selected each technology conditional on what term is randomly assigned to them.

Americans exhibit nuanced perceptions of what they consider AI, automation, machine learning, and robotics. In general, respondents underestimate what prevalence of AI, machine learning, and robotics in everyday technological applications. For almost all technologies, respondents assigned the term “automation” are more likely to select the technology than those assigned to the other terms.

Among those assigned the term “AI,” a majority think that virtual assistants (63%), smart speakers (55%), driverless cars (56%), and autonomous drones use AI (54%). A majority of respondents mistakenly assume that Facebook photo tagging, Google Search, Netflix or Amazon recommendations, or Google Translate do not use AI or machine learning. Curiously, among those assigned the term “robotics,” only 62% think social robots involve robotics, and only 70% think industrial robots involve robotics.²

²One hypothesis to explain this result is that some of this is due to inattentiveness, where respondents did not select all items than they would if they gave it greater consideration.

What the public considers AI, automation, machine learning, and robotics



Terms randomly assigned to respondents

- Artificial intelligence (AI)
- ▲ Automation
- Machine learning
- + Robotics

Support for developing AI

Procedure

Respondents are given the following definition of AI:

Artificial Intelligence (AI) refers to computer systems that perform tasks or makes 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 views five applications randomly selected from a list of 14:

- 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

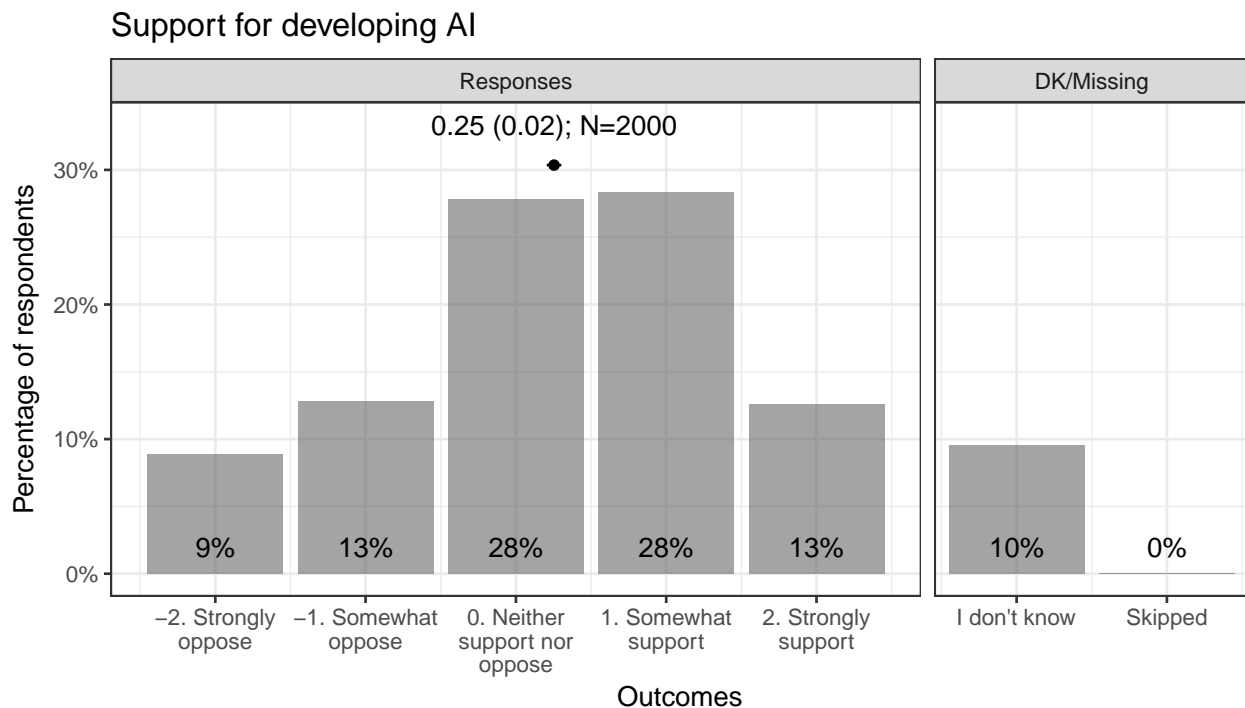
Afterward, respondents are asked how much they support or oppose the development of AI. The responses are coded in the following way:

- -2 = Strongly oppose
- -1 = Somewhat oppose
- 0 = Neither support nor oppose
- 1 = Somewhat support
- 2 = Strongly support

Results and discussions

The distribution of responses is displayed in the figure below. The mean outcome and the corresponding standard error are also presented.

Americans express mixed support for the development of AI, although more of them support than oppose the development of AI. A substantial minority (41%) somewhat or strongly support the development of AI. A smaller minority (22%) somewhat or strongly oppose the development of AI. Many express a neutral attitude: 28% state that they neither support nor oppose while 10% indicate they don't know.



Demographic predictors of support for developing AI

Procedure

We examine support for developing AI by the following demographic groups:

- Age group as defined by Pew Research Center: Millennial adults (born after 1980), Gen Xers (born 1965-1980), Baby Boomers (born 1946-1964), Silents/Greatest Generation (1945 and earlier)
- Gender: male, female
- Race: white, non-white
- Employment status: employed (full- or part-time), not employed
- Annual household income: less than \$30K annually, \$30-70K, \$70-100K, more than \$100K, prefer not to say
- Political party identification: Democrats, Republicans, Independents/Others
- Religion: Christian, follow other religions, non-religious
- Identifies as a born-again Christian: yes, no
- Completed a computer science undergraduate or graduate degree: yes, no
- Has computer science or programming experience: yes, no

We also use a multiple regression model to predict support for developing AI using the demographic variables listed above.

Results and discussion

The first figure below displays the average support for developing AI by demographic groups. Support for developing AI is negatively correlated with age. Males express greater support than females. Those who are employed express greater support than those who are not employed (i.e., retirees, students, or homemakers). Support is positively correlated with family income; support is the lowest among those who prefer not to report their income. Democrats express greater support for developing AI than either Republicans or Independents. Those who are Christian express less support compared to those who have other religious affiliation or are not religious; those who identify as “born-again” Christian exhibit less support than those who don’t identify as such. Those with computer science (CS) or engineering degrees or those with CS and programming experience are more likely support developing AI than those who do not.

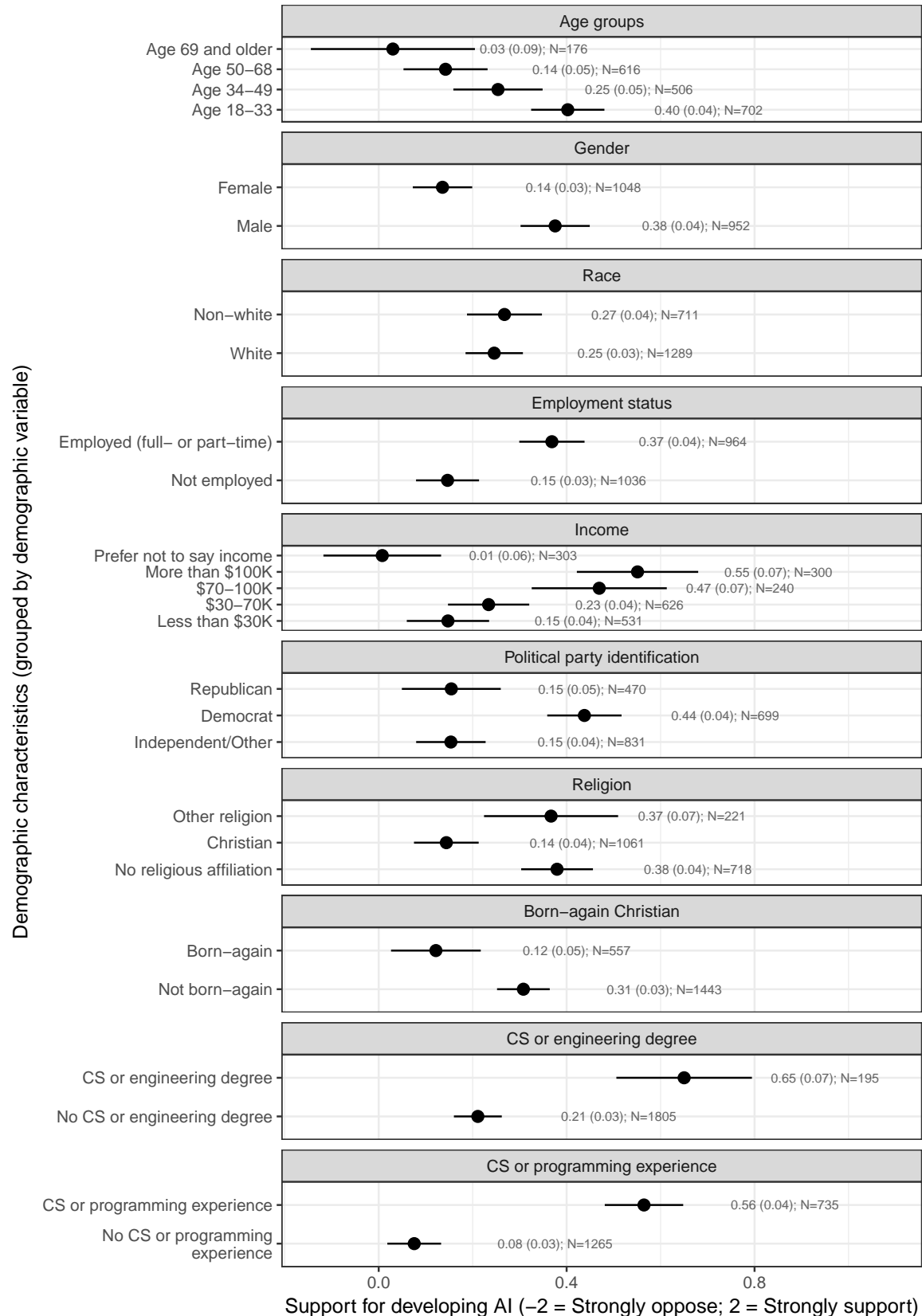
The second figure displays the coefficient plot from a multiple regression model that includes all of the demographic variables. The outcome variable has been standardized, so it has mean 0 and unit variance. Significant predictors correlated with support for developing AI include:

- identifying as a Democrat (versus identifying as an Independent)
- having a family income of \$70-100K or greater than \$100K annually (versus having an income of less than \$30K annually)
- having CS or programming experience (versus not having such experience).

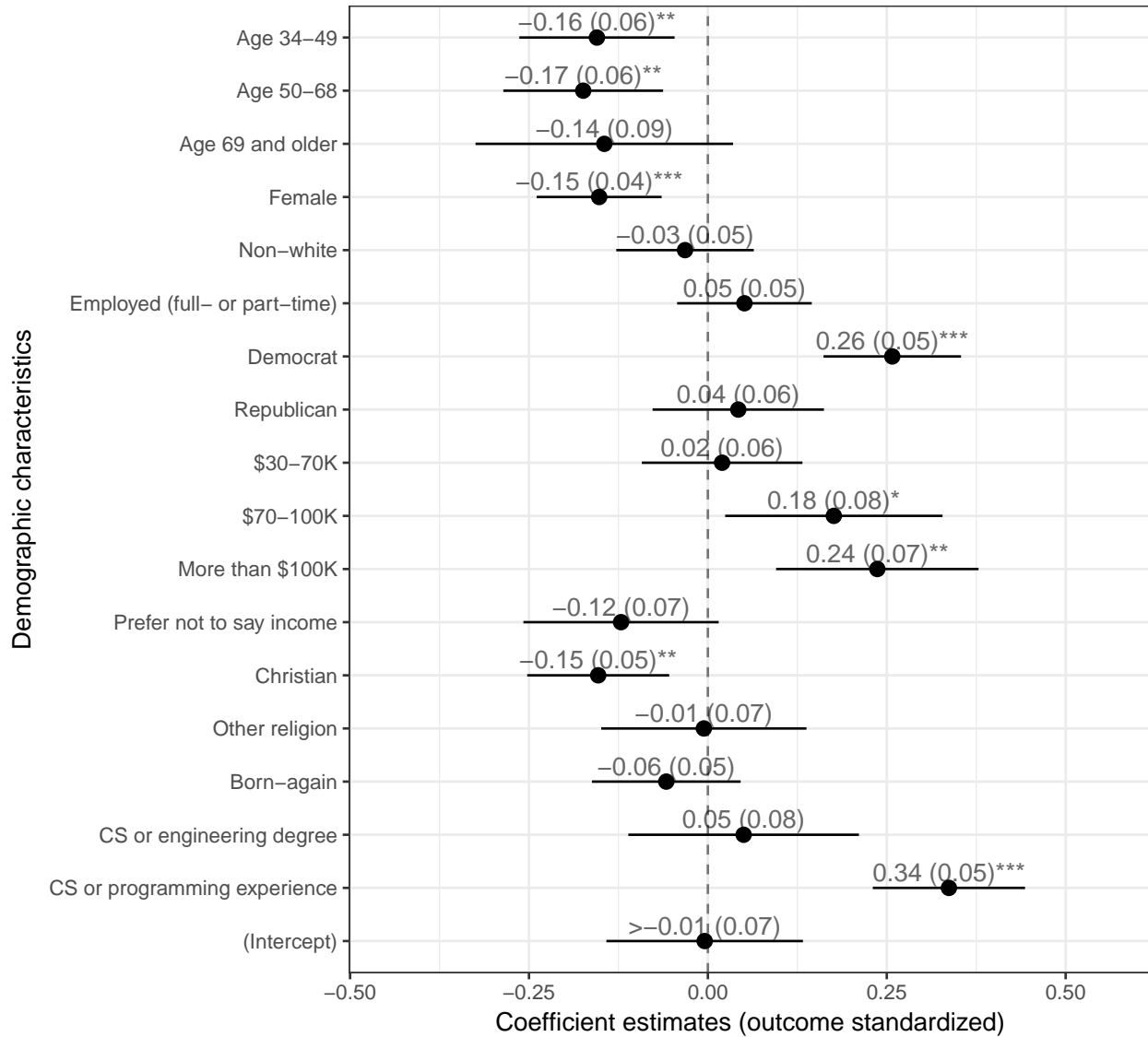
Significant predictors correlated with opposition to developing AI include:

- being a Gen Xer or a Baby Bloomer (versus being a Millennial)
- being a female (versus being a male)
- identifying as a Christian (versus identifying as non-religious)

Predicting support for developing AI using demographic characteristics: average support across groups



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



Survey experiment: AI and/or robots should be carefully managed

Procedure

We perform a survey experiment that seeks to replicate a question from the 2018 Special Eurobarometer #460. The original question asked respondents how much 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 are randomly assigned to consider one of these three statements:

1. AI and robots are technologies that require careful management.
2. AI is a technology that requires careful management.
3. Robots are technologies that require careful management.

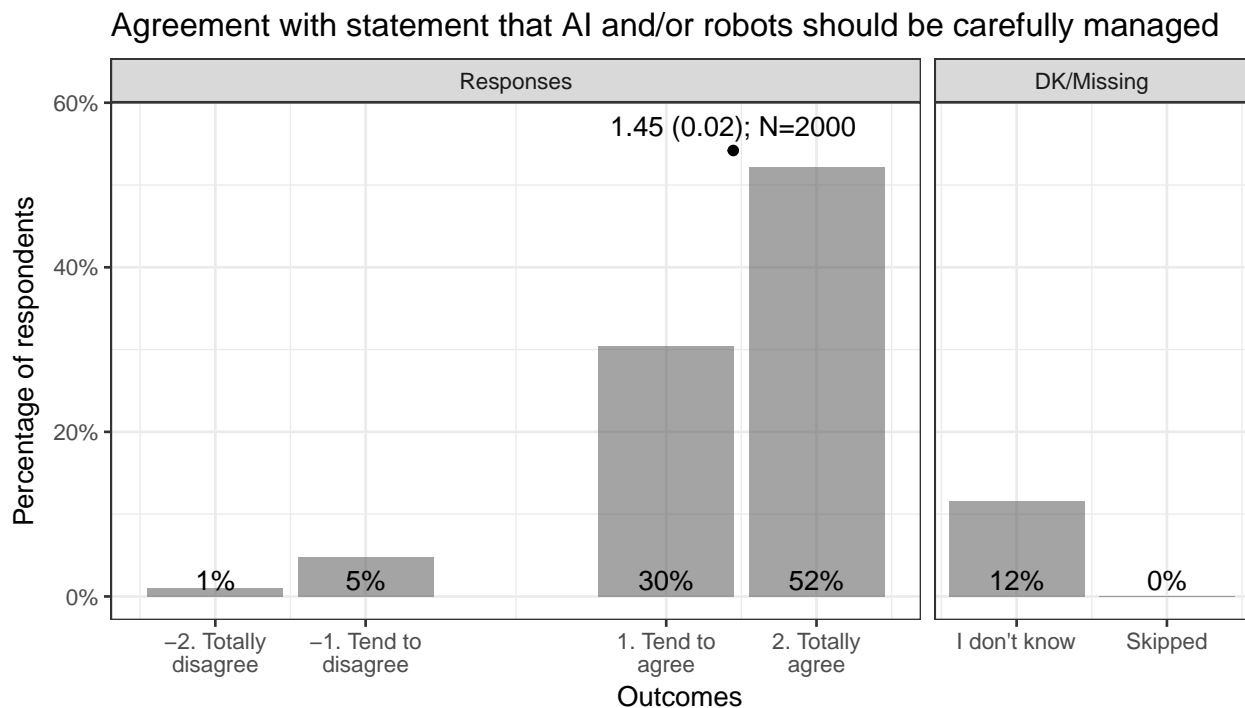
The responses are coded in the following way:

- -2 = Totally disagree
- -1 = Tend to disagree
- 1 = Tend to agree
- 2 = Totally agree

Results and discussion

The first figure below summarizes the results using all the data unconditional on treatment assignment. An overwhelming majority of Americans (82%) think that AI and/or robots should be carefully managed. Only 6% disagrees with the statement.

The second figure summarizes the results conditional on treatment assignment. The regression table presents the treatment effect estimates of the question-wording. We find that variations in the question-wording produce minor, non-significant, differences in responses.



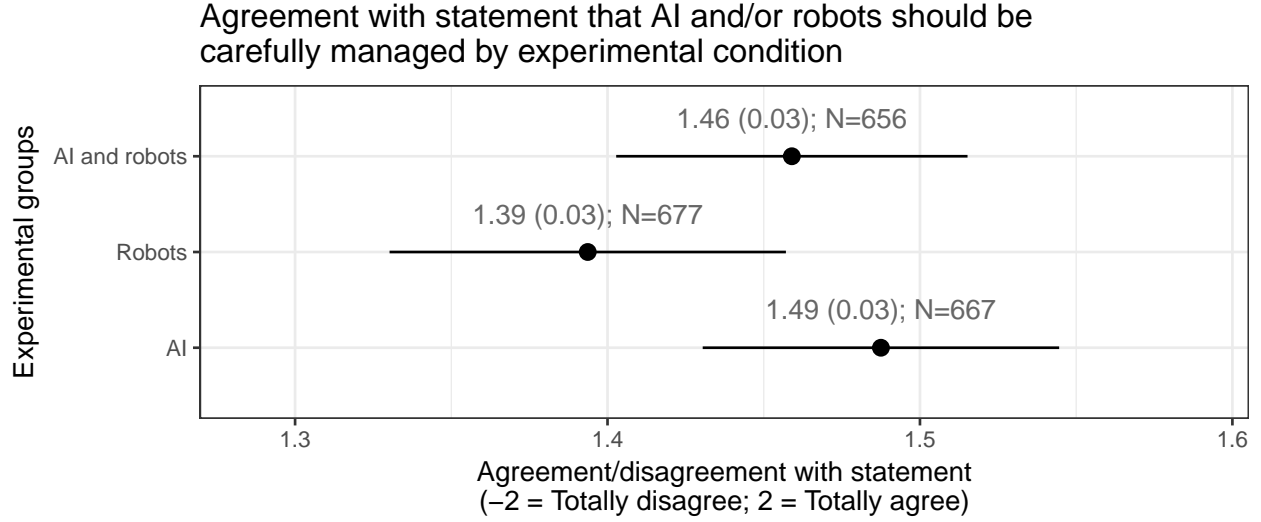


Table 1: Survey experiment attrition check: agreement with statement that AI and/or robots should be carefully managed

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

Variables	Coefficients
(Intercept)	0.11 (0.01)***
AI and robots	0.02 (0.02)
Robots	-0.01 (0.02)
N = 2000	F(2, 1997) = 1.21; p-value: 0.298

Table 3: Survey experiment results: agreement with statement that AI and/or robots should be carefully managed (controlling for DK/missing responses)

Variables	Coefficients
(Intercept)	1.46 (0.03)***
AI and robots	0.03 (0.04)
Robots	-0.07 (0.04)
N = 2000	F(5, 1994) = 0.98; p-value: 0.429

Comparison with Eurobarometer data

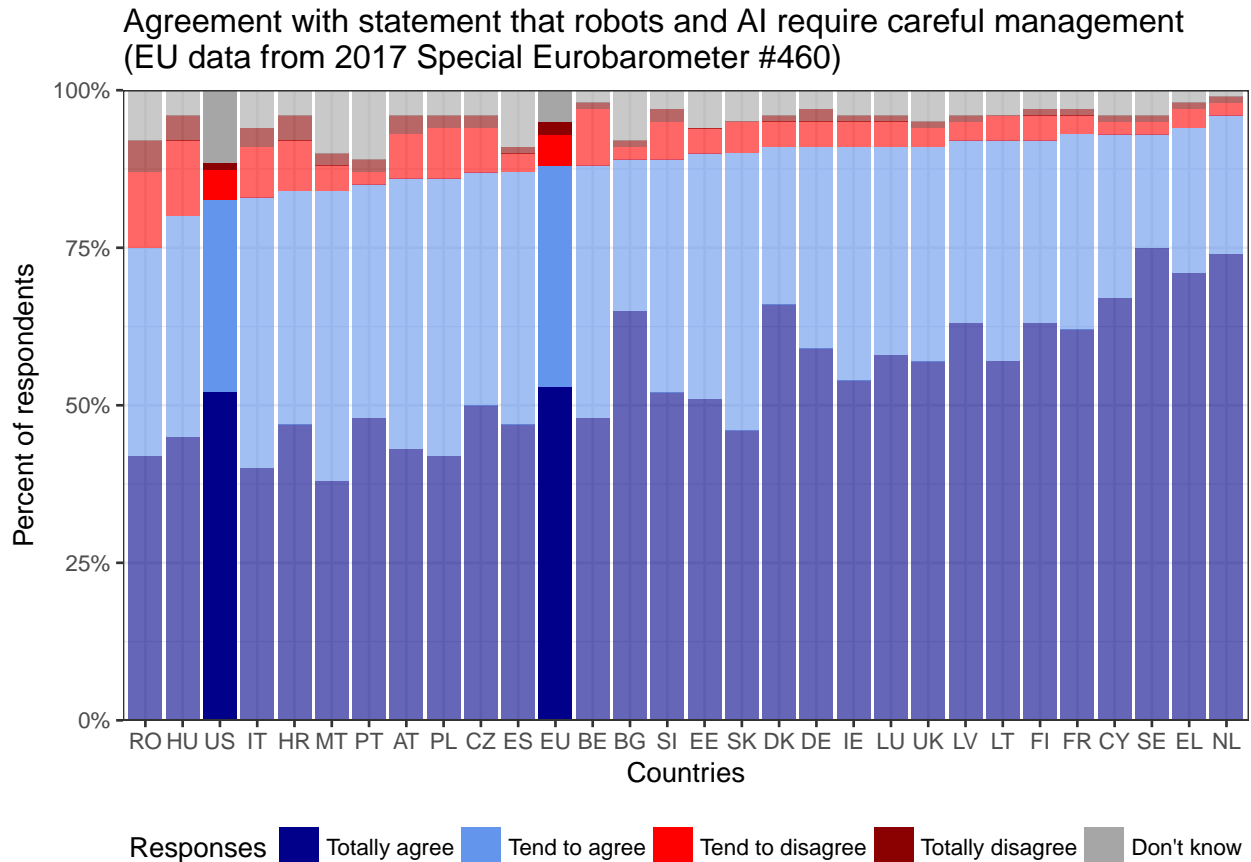
The figure below summarizes responses to the managing AI and robots question from the 2017 Special Eurobarometer #460 by country. The question asked whether respondents agreed or disagreed with the

following statement:

Robots and artificial intelligence are technologies that require careful management.

For the U.S., we use all the responses in our survey, unconditional on experimental condition given that the variation in question wording does not appear to affect responses.

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



AI governance challenges

Prioritizing governance challenges

Procedure

This set of questions seeks to understand how Americans prioritize policy issues associated with AI. Respondents are asked to consider five AI governance challenges, randomly selected from 13 potential ones. A short paragraph explains each AI governance challenge. The 13 challenges include:

- Bias in AI used in jobs hiring
- Bias in AI used in the criminal justice system
- Accuracy in diagnosing diseases
- Data privacy
- Digital manipulation

- AI-enhanced cyberattacks
- AI-enhanced surveillance
- U.S.-China arms race
- Value alignment
- Autonomous weapons
- Mass technical unemployment
- Failure of critical AI systems

After considering each governance challenge, respondents are asked how likely they think the challenge will affect large numbers of people 1) in the U.S. and 2) around the world within ten years. Respondents indicate the likelihood using the following multiple choices:

- Very unlikely: less than 5% chance
- Unlikely: 5-20% chance
- Somewhat unlikely: 20-40% chance
- Equally likely as unlikely: 40-60% chance
- Somewhat likely: 60-80% chance
- Likely: 80-95% chance
- Very likely: more than 95% chance

For the analysis, the multiple-choice answers are converted to the median value of each answer choice range. For instance, the “very unlikely” answer is converted to 2.5%.

Finally, respondents are asked how important it is for tech companies and governments to manage each governance challenge carefully. They are given the following answer options:

- 3 = Very important
- 2 = Somewhat important
- 1 = Not too important
- 0 = Not at all important

Results and discussion

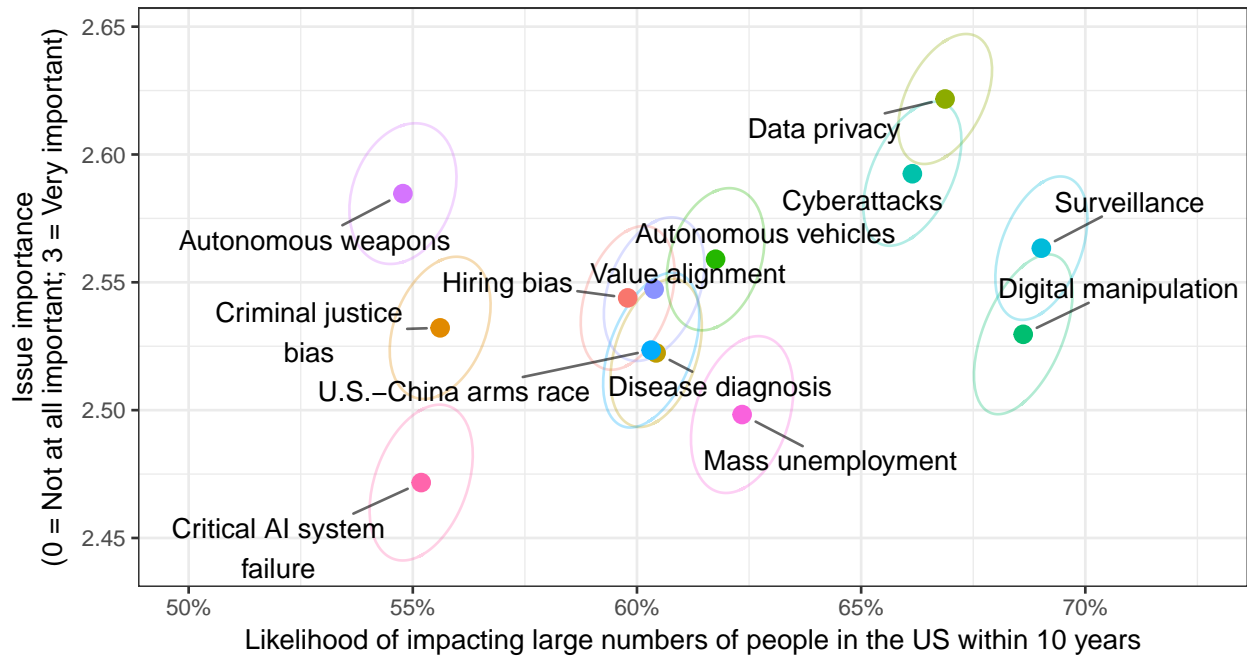
The results from the AI governance challenge questions are plotted below. In the first figure, the x-axis is the perceived likelihood of the problem happening to large numbers of people in the U.S. In the second figure, the x-axis is the perceived likelihood of the problem happening to large numbers of people around the world. The y-axes of the two plots are the same. A dot represents the mean perceived likelihood and issue importance, and the correspondent ellipse contains the 95% confidence region.

Americans consider all the AI governance challenges to be important: the mean perceived issues importance of each governance challenge is between “somewhat important” (2) and “very important” (3). While the variation in perceived issue importance is small, there is greater variation in Americans’ perceived likelihood of each governance challenge impacting a large number of people in the next 10 years – ranging from 55% likelihood for lethal autonomous weapons to 69% likelihood for AI-enhanced surveillance in the U.S.

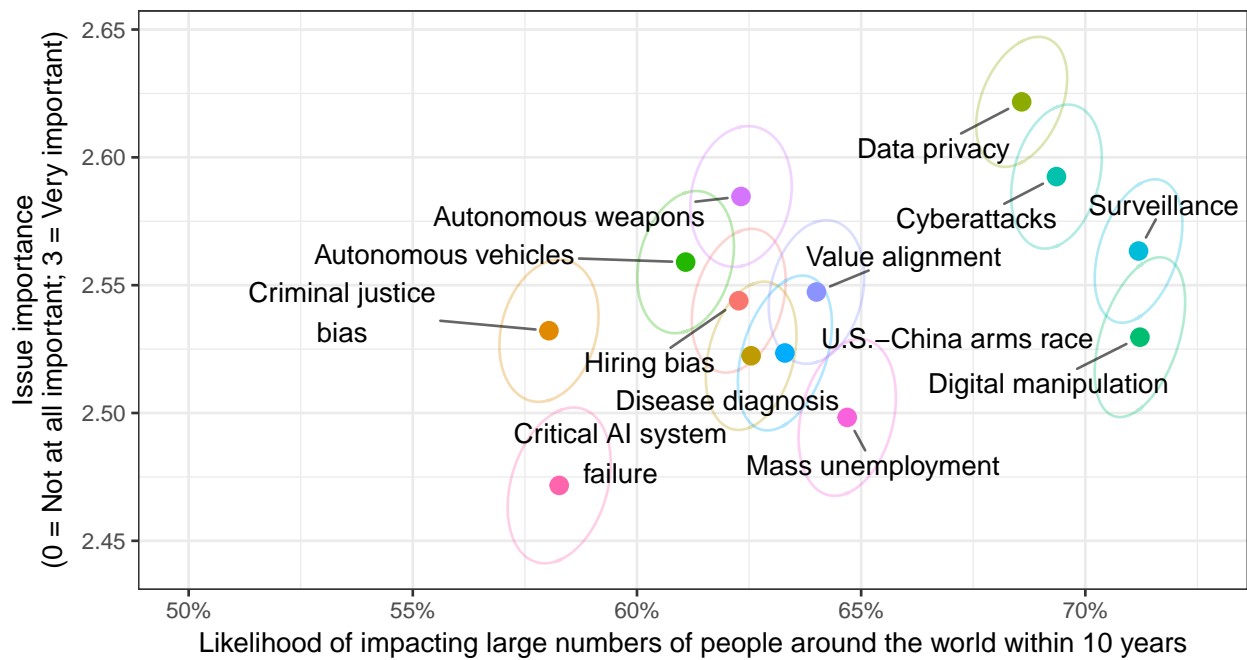
The AI governance challenges Americans think are most likely to impact large numbers of people and important for tech companies and governments to tackle are found in the upper-right quadrant of the two plots. These issues include protecting data privacy, AI-enhanced cyberattacks, AI-enhanced surveillance, and digital manipulation. These issues have been widely reported in the media in recent months. Interestingly, Americans also prioritize AI governance challenges that are not commonly discussed or may appear highly complex, such as the U.S.-China arms race or value alignment. Nevertheless, they do not consider the failure of critical AI systems to be likely or very important as a policy issue.

In the third figure, we compare respondents’ perceived likelihood of each governance challenge impacting people in the U.S. versus around the world. The results are largely similar for most governance challenges except for two. Americans think that value alignment and lethal autonomous weapons are more likely affect large numbers of people around the world than in the U.S. within the next decade.

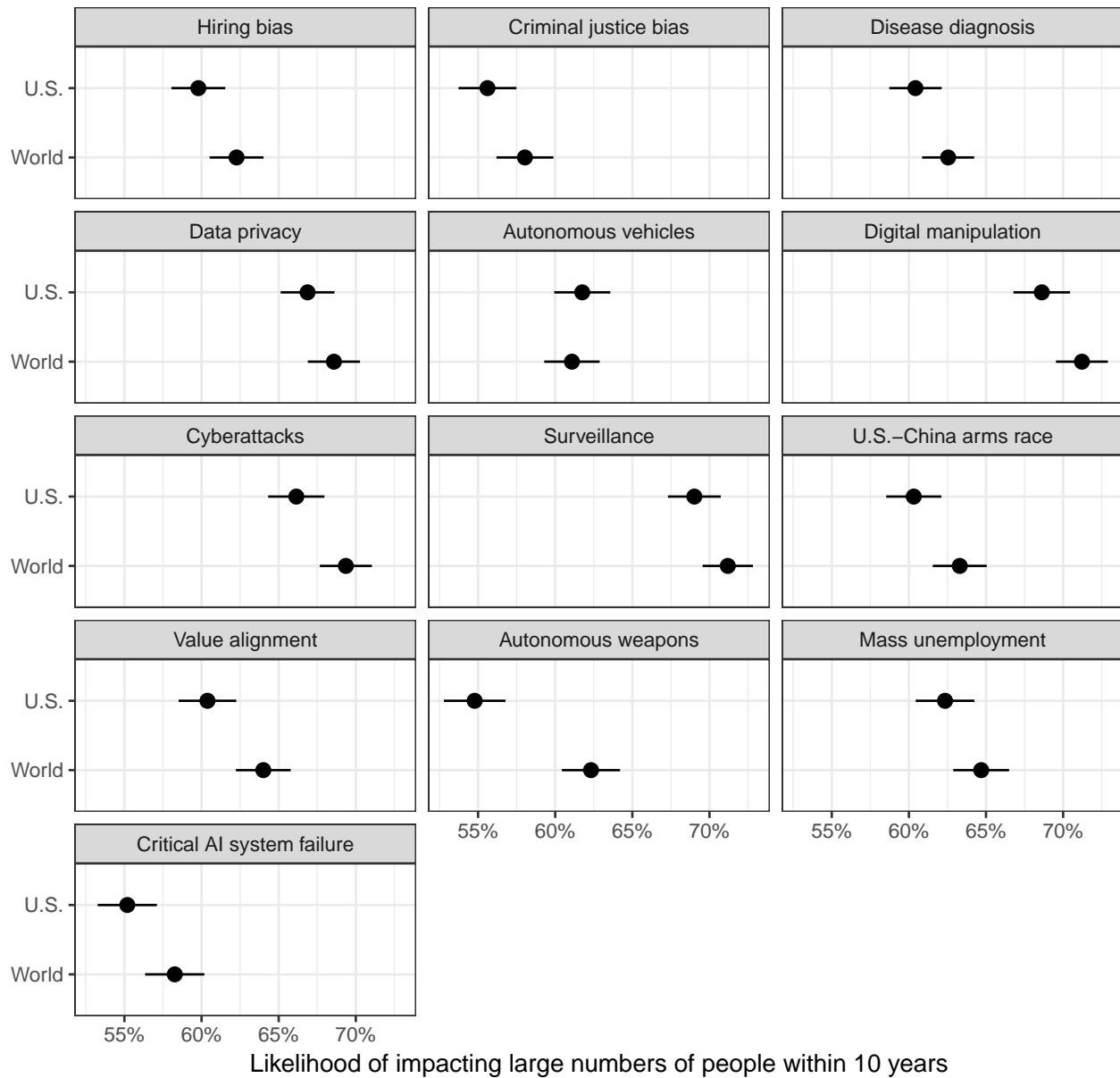
Perceptions of AI governance challenges in the U.S.



Perceptions of AI governance challenges around the world



Comparing perceived likelihood: in U.S. vs. around the world



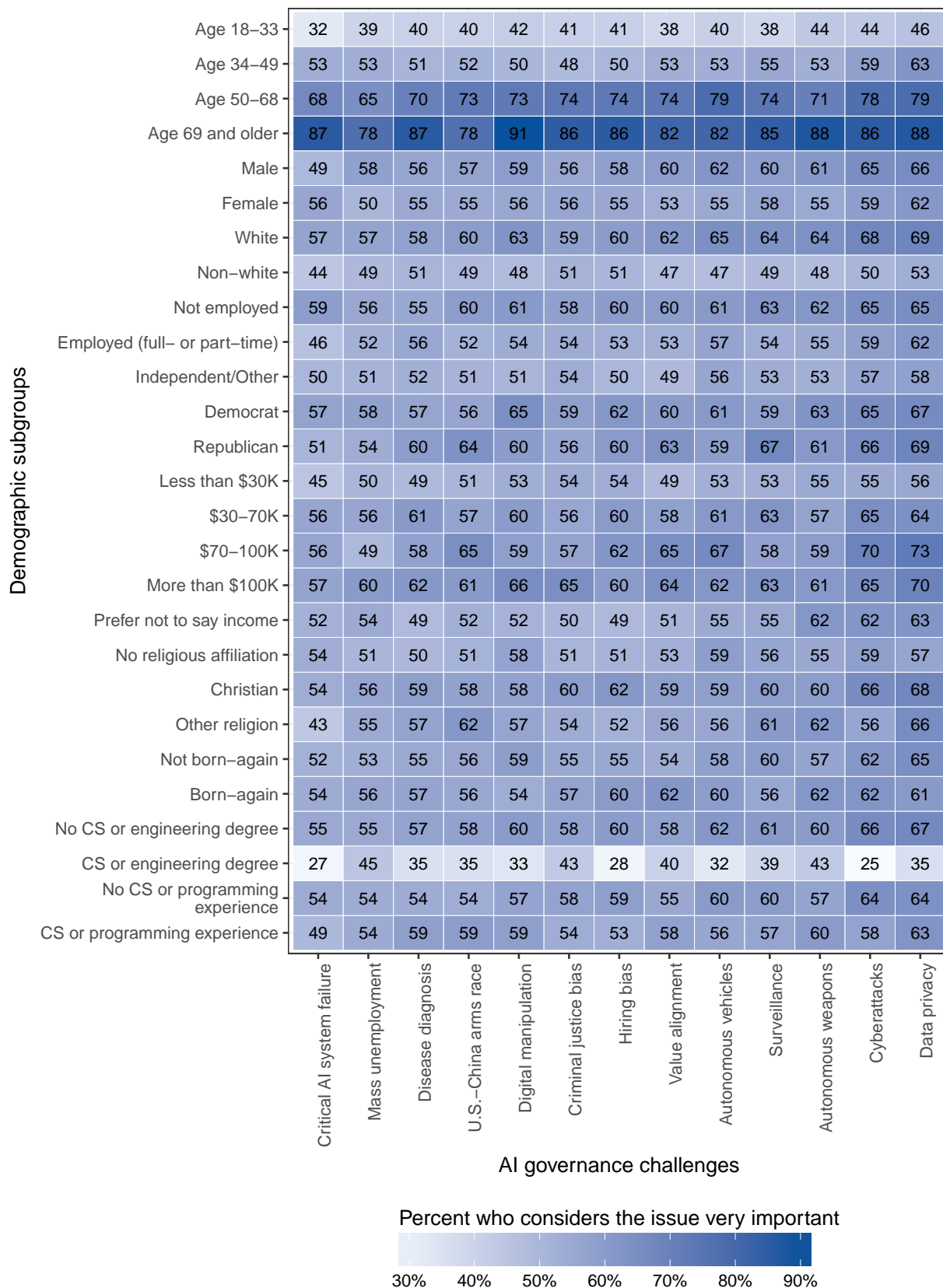
Demographic predictors of issue importance

In the figure below, we summarize perceptions of issue importance by demographic subgroup and AI governance challenge. In the first graph below, we show the percent of respondents within each demographic subgroup who considers each governance challenge a “very important” issue. For the second graph, to make the differences appear more visible, we subtract the overall mean across all responses from each subgroup-governance challenge mean.

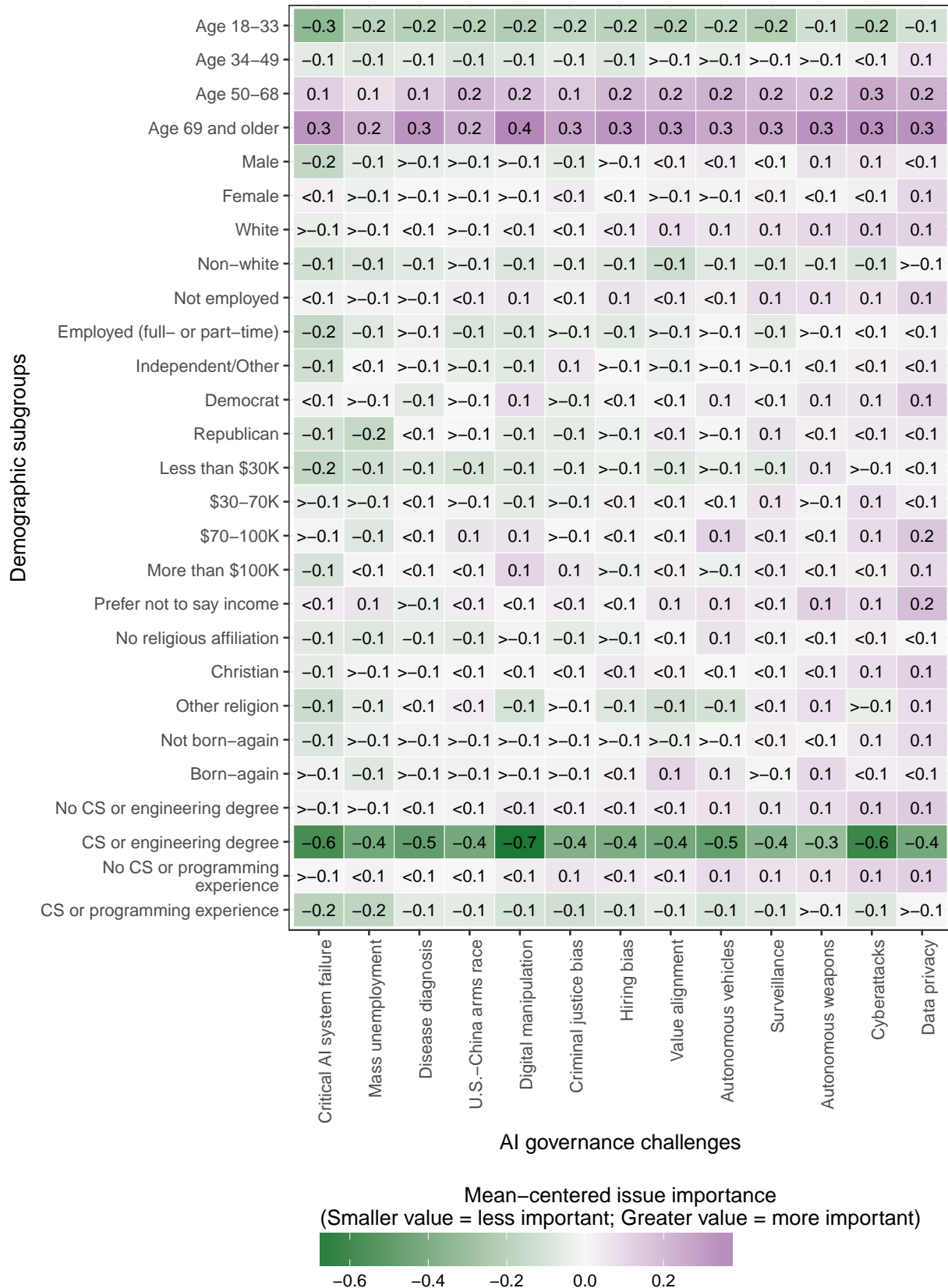
Two highly visible patterns emerge from the figures. First, age is positively correlated with perceived issue importance across all AI governance challenges. In a previous question, we find that older Americans, compared with younger Americans, are less enthusiastic about developing AI. Our nuanced questions about AI governance challenges might explain this lack of enthusiasm: older respondents express greater concern about each of the governance issues when compared with their younger counterparts.

Second, those with CS or engineering degrees – compare 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.

AI governance challenges: issue importance by U.S. demographic subgroups



AI governance challenges: issue importance by U.S. demographic subgroups



Trust of actors to develop and manage AI

Procedure

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

Confidence is measured using a four-point scale:

- 0 = No confidence at all
- 1 = Not too much confidence
- 2 = A fair amount of confidence
- 3 = A great deal of confidence

In addition, respondents are 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 are randomly assigned five out of 15 actors to evaluate. For actors that are not well-known to the public (e.g., AAAI, Partnership on AI, etc.), a short description is provided.

Confidence is measured using the same four-point scale used in the trust to develop AI question described above.

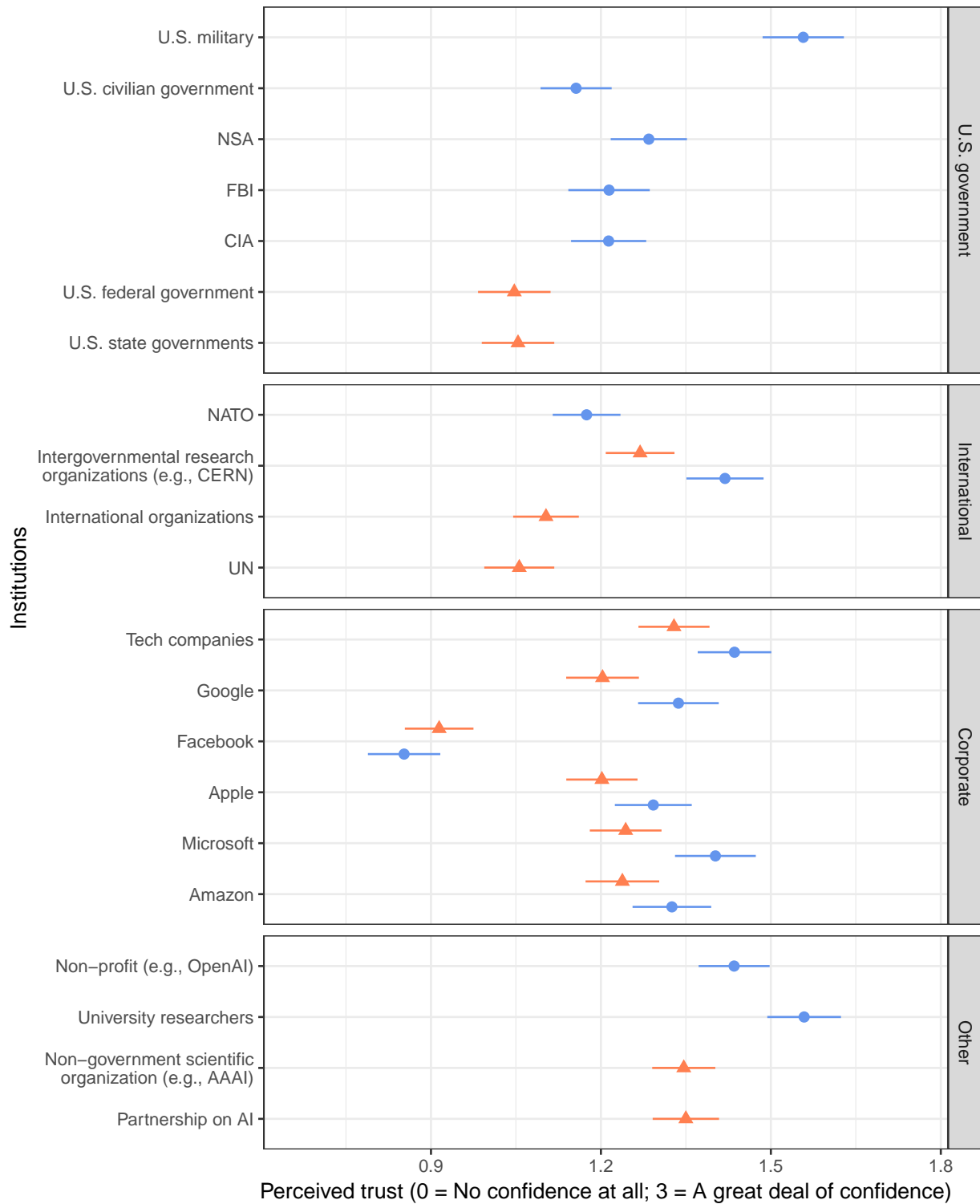
Results and discussion

The figure below compares respondents' evaluations of the actors. We sort the actors into four categories: U.S. government, international, corporate, and other.

In general, Americans do not have a great deal of confidence in any of the actors to develop AI in the best interest of the public. For all actors, the average rating is between "not too much confidence" (1) and "a fair amount of confidence" (2). The U.S. military and university researchers are the most trusted groups to develop AI. Intergovernmental research organizations and non-profit AI research organizations (e.g., OpenAI) have also received relatively high ratings. In general, respondents place greater confidence in tech companies than in civilian U.S. government agencies or NATO. But there is one notable exception: the public rates Facebook as the least trustworthy of all the actors. Our survey was conducted three months after the fallout of the Facebook/Cambridge Analytica scandal; therefore, respondents may have low levels of general trust in Facebook.

The results on the public's trust of various actors to manage the development and use of AI provided are similar to the results discussed above. Once again, Americans do not have a great deal of confidence in any of the actors. The public places the most trust in Partnership on AI and non-governmental scientific organizations (e.g., AAAI) to manage AI. Trust in the U.S. government to manage AI is somewhat lower than tech companies and international governmental organizations. Again, Facebook stood out as an outlier: respondents give it a much lower rating than any other group.

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



Outcome measure

- Trust in various actors to develop AI in the interest of the public
- ▲ Trust in various actors to manage AI in the interest of the public

Table 4: Trust in various actors to develop and manage AI in the interest of the public

Actors	Categories	Trust to develop AI	Trust to manage AI
U.S. military	U.S. government	1.56 (0.04); N=638	
U.S. civilian government	U.S. government	1.16 (0.03); N=671	
NSA	U.S. government	1.28 (0.03); N=710	
FBI	U.S. government	1.21 (0.04); N=656	
CIA	U.S. government	1.21 (0.03); N=730	
U.S. federal government	U.S. government		1.05 (0.03); N=743
U.S. state governments	U.S. government		1.05 (0.03); N=713
NATO	International	1.17 (0.03); N=695	
Intergovernmental research organizations (e.g., CERN)	International	1.42 (0.03); N=645	1.27 (0.03); N=747
International organizations	International		1.10 (0.03); N=827
UN	International		1.06 (0.03); N=802
Tech companies	Corporate	1.44 (0.03); N=674	1.33 (0.03); N=758
Google	Corporate	1.34 (0.04); N=645	1.20 (0.03); N=767
Facebook	Corporate	0.85 (0.03); N=632	0.91 (0.03); N=741
Apple	Corporate	1.29 (0.03); N=697	1.20 (0.03); N=775
Microsoft	Corporate	1.40 (0.04); N=597	1.24 (0.03); N=771
Amazon	Corporate	1.33 (0.04); N=685	1.24 (0.03); N=784
Non-profit (e.g., OpenAI)	Other	1.44 (0.03); N=659	
University researchers	Other	1.56 (0.03); N=666	
Non-government scientific organization (e.g., AAAI)	Other		1.35 (0.03); N=792
Partnership on AI	Other		1.35 (0.03); N=780

AI policy and China

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

Procedure

In this survey experiment, we ask respondents to consider either the U.S. or China’s status in AI research and development (R&D). Respondents are asked the following:

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

The responses are coded in the following way:

- 3 = Best in the world
- 2 = Above average
- 1 = Average
- 0 = Below average

Results and discussion

The figure below summarizes the responses by country. The regression table presents the estimated difference in respondents’ perceptions of the two countries’ status in AI research and development.

Americans do not perceive the U.S. to be better at AI R&D than China. According to our regression analysis, the respondents think China is better than the U.S. Ten percent of Americans think that the U.S. is the best in the world in terms of AI R&D while 7% think that about China. In contrast, 36% of Americans think that

the U.S.'s AI R&D is “above average” while 45% think China’s is “above average.” Given the technological nature of this question, about a quarter of the respondents in each condition indicated that they don’t know.

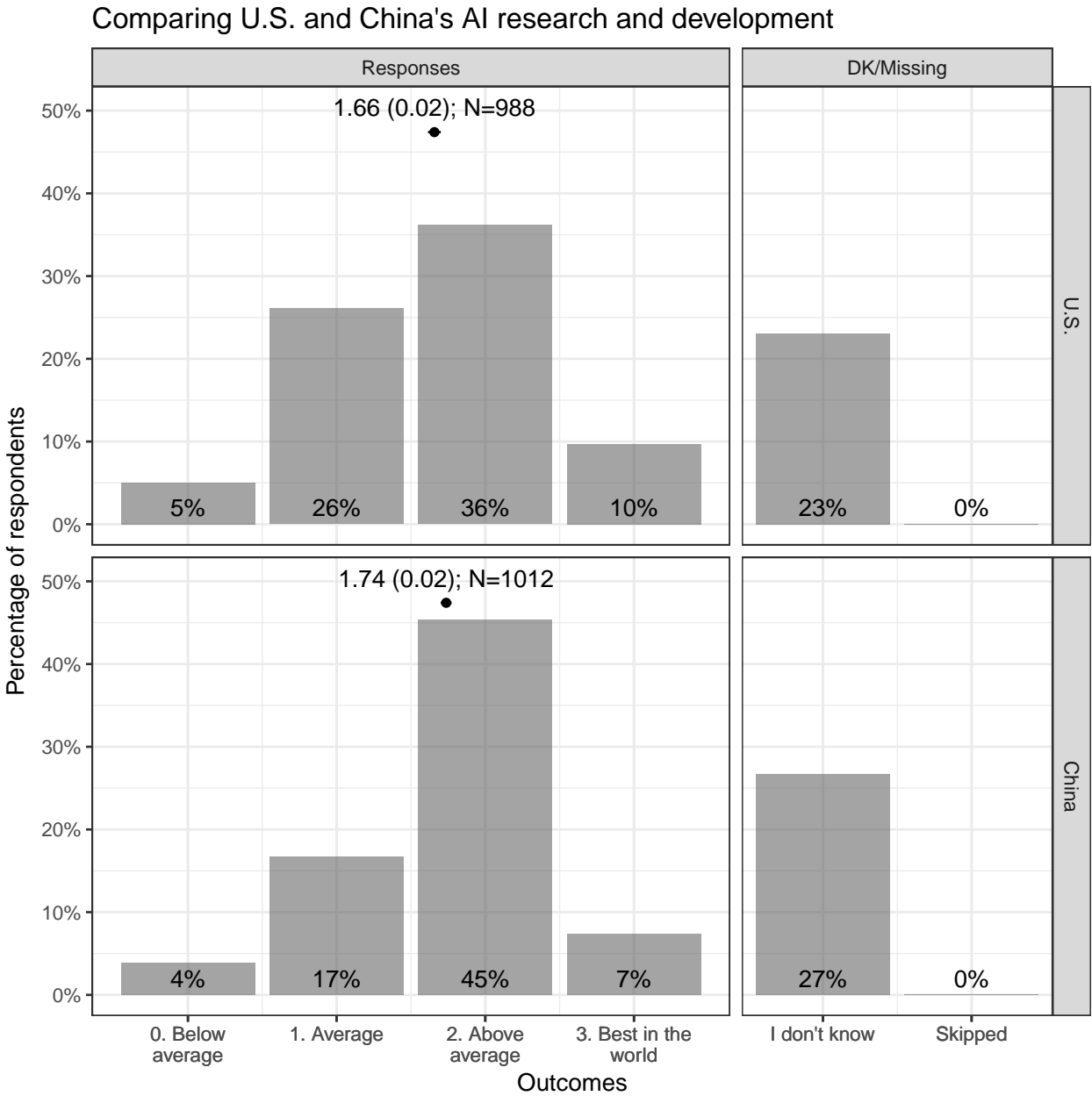


Table 5: Survey experiment attrition check: comparing U.S. and China’s AI research and development

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

Table 6: Survey experiment attrition check: comparing U.S. and China’s AI research and development

Variables	Coefficients
(Intercept)	0.27 (0.01)***
U.S.	-0.04 (0.02)
N = 2000	F(1, 1998) = 3.60; p-value: 0.058

Table 7: Survey experiment results: comparing U.S. and China’s AI research and development (controlling for DK/missing responses)

Variables	Coefficients
(Intercept)	1.74 (0.02)***
U.S.	-0.08 (0.03)**
N = 2000	F(3, 1996) = 5.74; p-value: <0.001

Survey experiment: U.S.-China arms race

Procedure

In this survey experiment, respondents are randomly assigned to consider different arguments about an U.S.-China arms race. All respondents are 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 are told they will read a short news article. The three treatments are:

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
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

The exact wording of these treatments are found in the Appendix.

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

- The U.S. should invest more in AI military capabilities to make sure it doesn’t fall behind China, 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.

Results and discussion

In the first figure below, the mean outcome, the standard error, the 95% confidence interval, and the number of responses are shown for each experimental group. The next figure shows the estimated treatment effect (difference in outcome between each treatment group and the control group).

The final figure summarizes the difference in responses to Statement 2 and Statement 1 by experimental group. Positive values mean greater agreement with Statement 2 than with Statement 1; negative values mean greater agreement with Statement 1 than with Statement 2.

Americans, in general, weakly agree that the U.S. should invest more in AI military capabilities *and* should cooperate with China to avoid the dangers of an AI arms race. Respondents assigned to Treatment 2, the risks of arms race treatment, indicate significantly greater agreement with Statement 2 (pro-cooperation) than Statement 1 (investing in AI military capabilities). 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, overall, do little to change the respondents' preferences, with one exception. Treatment 2 decreases respondents' agreement with the statement that the U.S. should invest more in AI military capabilities by 27%.

Responses from U.S.–China arms race survey experiment

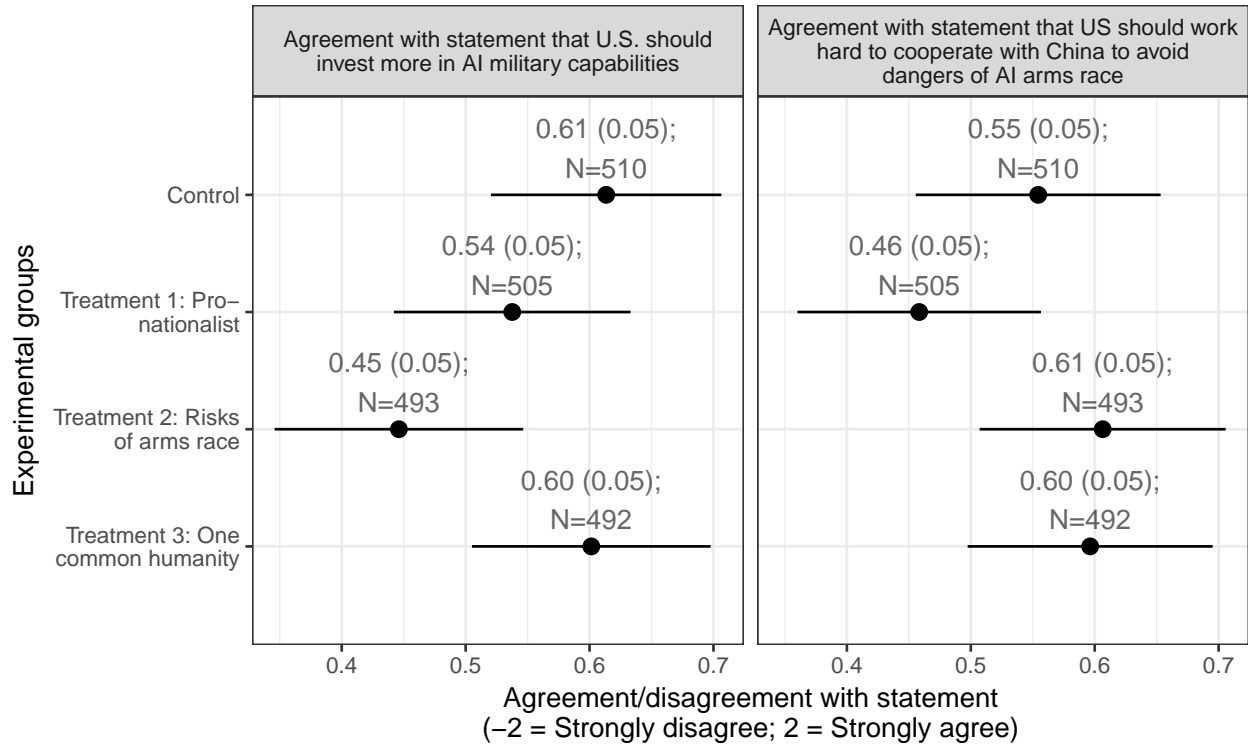


Table 8: Survey experiment attrition check: agreement with statement that U.S. should invest more in AI military capabilities

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

Variables	Coefficients
(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) = 0.09; p-value: 0.966

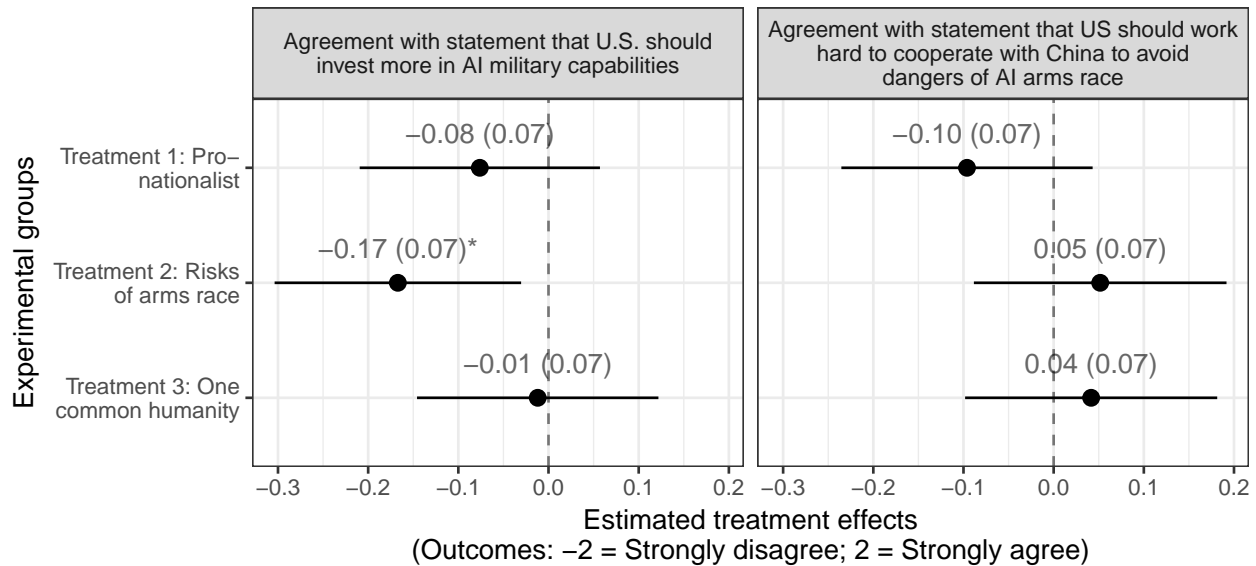
Table 10: Survey experiment attrition check: agreement with statement that US should work hard to cooperate with China to avoid dangers of AI arms race

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

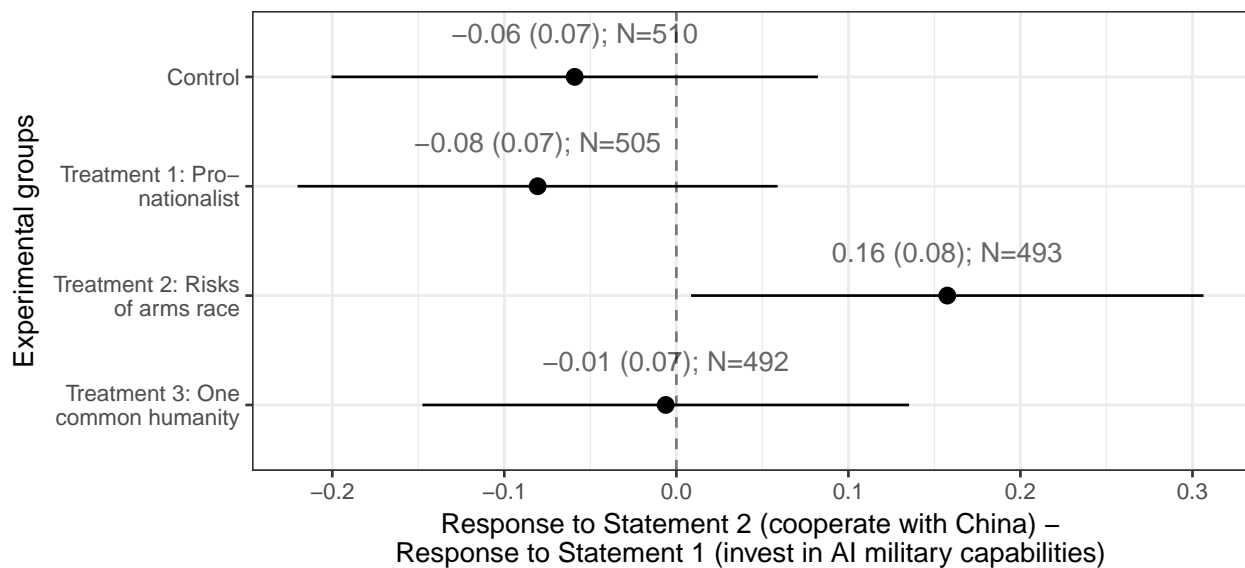
Table 11: Survey experiment attrition check: agreement with statement that US should work hard to cooperate with China to avoid dangers of AI arms race

Variables	Coefficients
(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) = 0.92; p-value: 0.429

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



Difference in response to the two statements by experimental group



Issue areas for possible U.S.-China cooperation

Procedure

This question investigates issue areas where Americans perceive likely U.S.-China cooperation. Each respondent is randomly assigned to consider three out of five AI governance challenges. For each challenge, he or she is asked, “For the following issues, how likely is it that the U.S. and China can cooperate?”.

The five AI governance challenges include:

- 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

Respondents indicate the likelihood of cooperation using the following multiple choices:

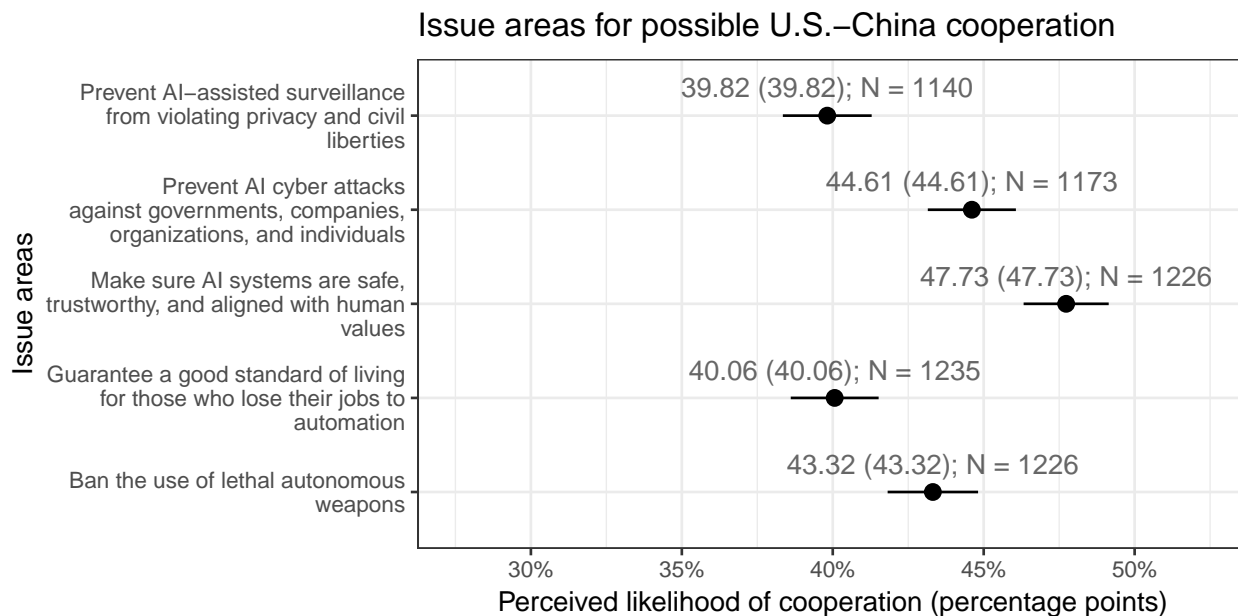
- Very unlikely: less than 5% chance
- Unlikely: 5-20% chance
- Somewhat unlikely: 20-40% chance
- Equally likely as unlikely: 40-60% chance
- Somewhat likely: 60-80% chance
- Likely: 80-95% chance
- Very likely: more than 95% chance

For the analysis, the multiple-choice answers are converted to the median value of each answer choice range. For instance, the “very unlikely” answer is converted to 2.5%.

Results and discussion

The figure below shows the mean perceived likelihood of cooperation on each of the issue areas.

On each of these issues, Americans – on average – think that cooperation is less likely than not. 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 perceived to be the least likely (40% mean likelihood).



Trend across time: job creation or job loss

This question seeks to understand how Americans’ attitudes towards automation changes over time. Between 1983 and 2003, U.S. National Science Foundation 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 continues this time trend study by posing a similar – but updated – question:

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.

The responses are coded in the following way:

- -2 = Strongly disagree
- -1 = Disagree
- 1 = Agree
- 2 = Strongly agree

Survey experiment: future time frame

Procedure

Our survey question also addresses the chief ambiguity of the original question: lack of a future time frame. We use a survey experiment to help resolve this ambiguity by randomly assigning respondents to one of four conditions. Besides the control condition in which we do not specify a future time frame, we also created three treatment conditions that defined the future time frame:

- In 10 years
- In 20 years
- In 50 years

Results and discussion

The figure below summarizes the responses by experimental group. The regression table presents the estimated differences in responses between the control and experimental groups.

On average, Americans disagree with the statement more than they agree with it, although about a quarter of respondents give a “don’t know” response in each experimental group. Those assigned to the 10 year time frame indicate greater disagreement than those assigned to the control condition of no specified time frame (two-sided p -value = 0.04). Agreement with the statement appears to increase with the time frame, but there exist no significant differences between the responses of differing time frame.

This is puzzling from the perspective that AI and robotics will increasingly automate human performed tasks. Such a perspective would expect more agreement as one looks further into the future. One hypothesis to explain this pattern is that respondents believe the disruption from automation is greater in the coming ten years, then it will be over the longer run once institutions adapt. Another hypothesis is that asking about the next ten years makes this question more concrete, whereas in the more abstract framing different reasoning processes may operate. The effects are sufficiently uncertain that we cannot conclude with confidence that agreement goes up with time frame; however, we can conclude with confidence that agreement does not go strongly down with time frame (e.g., we can reject that average agreement decreases by one point going from 10 to 50 years frame).

Agreement with the statement that automation will create more jobs than it will eliminate

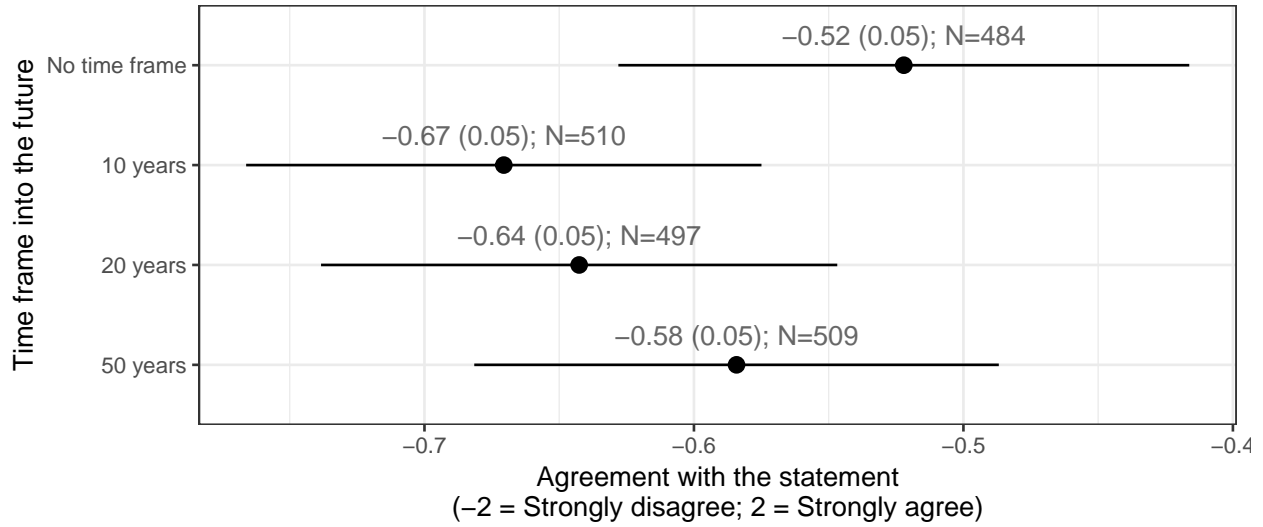


Table 12: Survey experiment attrition check: future time frame

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 13: Survey experiment attrition check: future time frame

Variables	Coefficients
(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) = 0.38; p-value: 0.767

Table 14: Survey experiment results: future time frame (controlling for DK/missing responses)

Variables	Coefficients
(Intercept)	-0.52 (0.05)***
10 years	-0.15 (0.07)*
20 years	-0.12 (0.07)
50 years	-0.06 (0.07)
N = 2000	F(7, 1992) = 0.84; p-value: 0.553

Table 15: Testing coefficients for time frames are equivalent

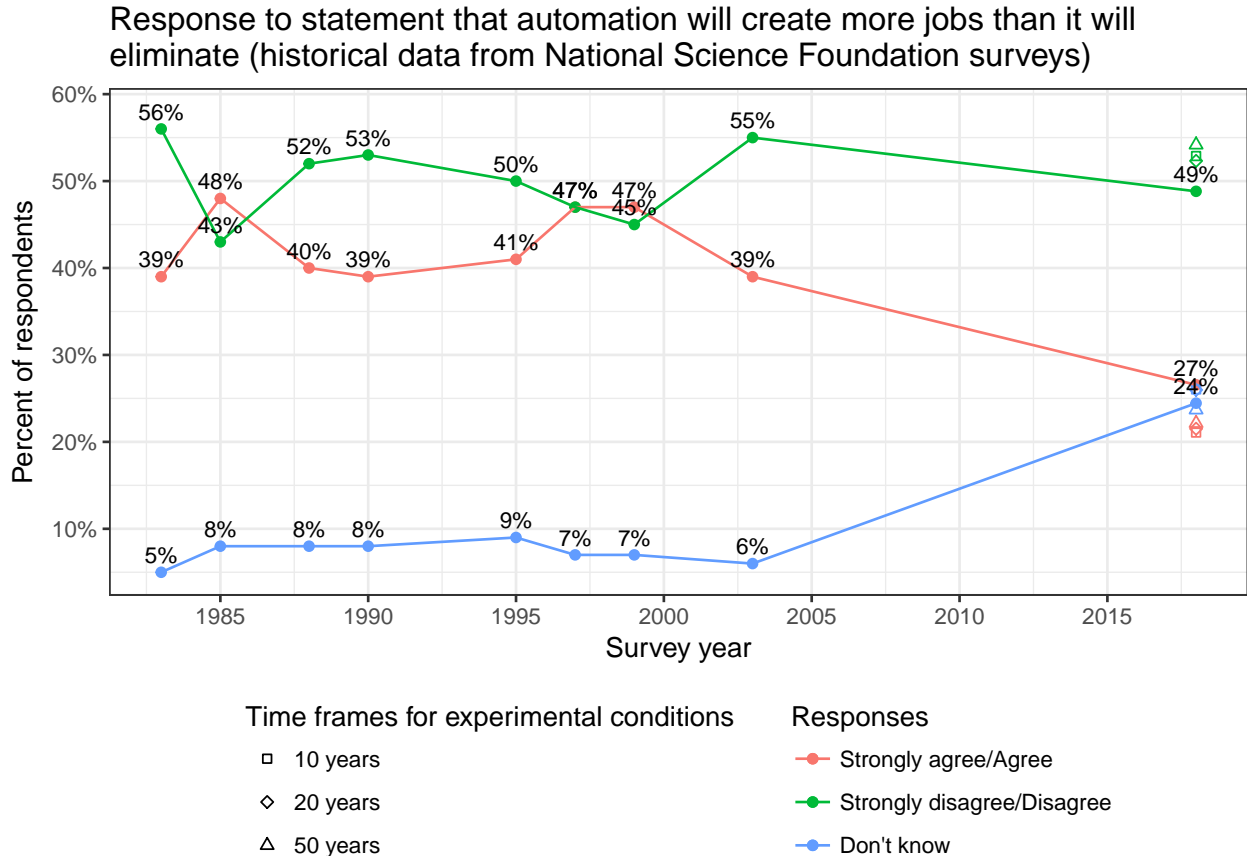
Tests	F-statistic	p-value
10 years = 20 years	$F(1, 1992) = 0.15$	0.70
10 years = 50 years	$F(1, 1992) = 1.41$	0.24
20 years = 50 years	$F(1, 1992) = 0.63$	0.43

Extending the time trend

The figure below summarizes the historical polling data while adding in data points from our survey. The “strongly agree” and “agree” responses are grouped together; likewise, the “strongly disagree” and “disagree” responses are grouped together.

The percent of Americans today who disagree with the statement is on par with historical levels. Nevertheless, the percent who agree with the statement has decreased by 12 percentage points since 2003 while the percent who responded “don’t know” has increased by 18 percentage points since 2003.

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 their 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); however, other studies have shown that online surveys cause no such effect (Shin, Johnson, and Rao 2012; Bech and Kristensen 2009). Third, the changes in response could have arisen due to the actual changes in respondents’ perceptions over time.



High-level machine intelligence

The next set of questions concerns respondents’ perceptions of high-level machine intelligence.

Forecasting timeline

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

Say we have high-level machine intelligence when machines will be able to perform almost all tasks that are economically relevant today better than the median human (today) at each task. Note that this would include subtle common sense queries such as a travel agent would provide. Ignore tasks which are legally or culturally restricted to humans, such as being a jury member.

Note that our definition of high-level machine intelligence is equivalent to what many would consider human-level machine intelligence.

Respondents predict the probability that high-level machine intelligence will be built in 10, 20, and 50 years using the following multiple choices:

- Very unlikely: less than 5% chance
- Unlikely: 5-20% chance
- Somewhat unlikely: 20-40% chance
- Equally likely as unlikely: 40-60% chance
- Somewhat likely: 60-80% chance
- Likely: 80-95% chance
- Very likely: more than 95% chance

For the analysis, the multiple-choice answers are converted to the median value of each answer choice range. For instance, the “very unlikely” answer is converted to 2.5%.

We present our 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 (see Grace et al. (2018)), we aggregated the respondents’ forecasts using the same method. Note that Grace et al. (2018) asked about a much higher definition of high-level machine intelligence involving machines being better than all humans at all tasks.

Each respondent provides three data points for their forecast, and these are fitted to the Gamma CDF by least squares to produce the grey CDFs. The “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. Please see the code and (Grace et al. 2018) for details. Note that survey weights are not used in this analysis due to problems incorporating survey weights into the bootstrap.

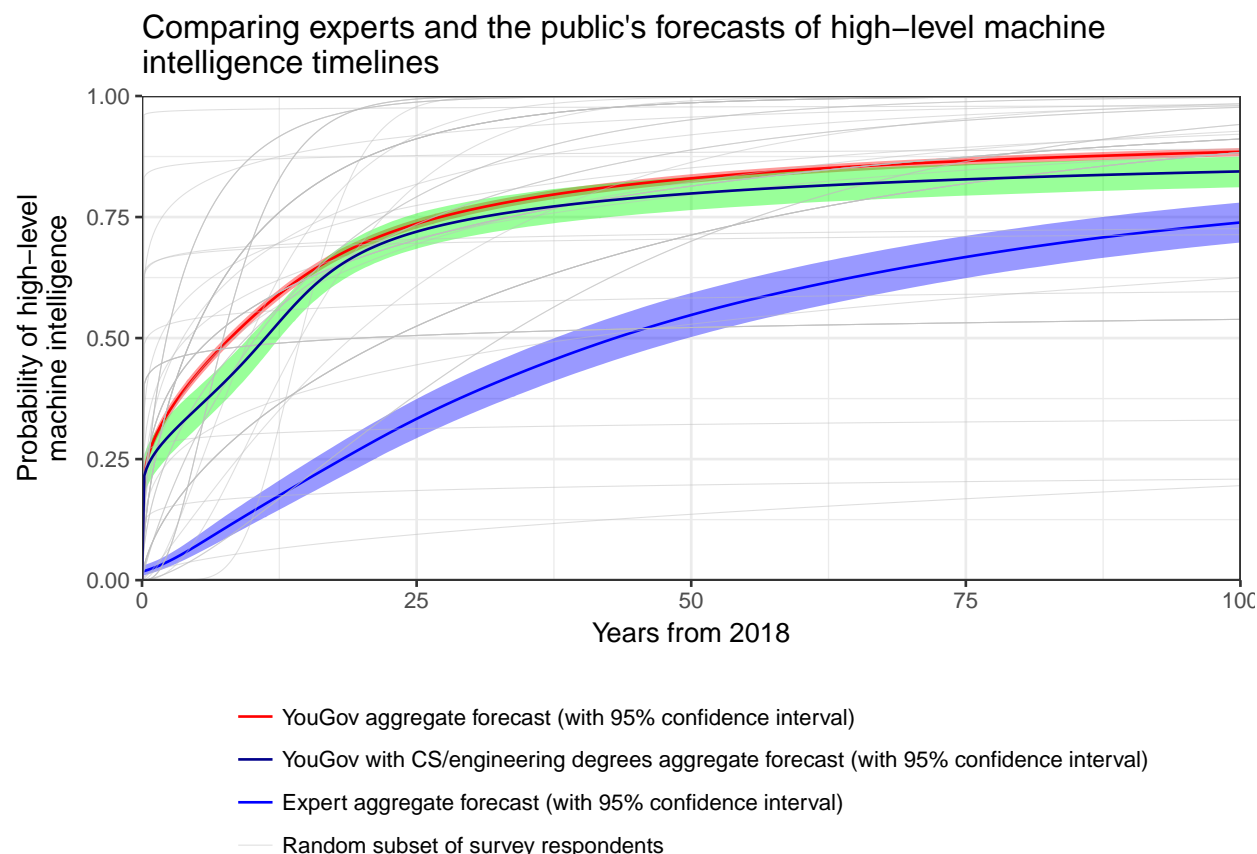
Results and discussion

Respondents predict that high-level machine intelligence will arrive fairly quickly. The median respondent predicts there is 54% probability of high-level machine intelligence by 2028, 70% probability by 2038, and 88% probability by 2068.

These predictions are considerably more optimistic than predictions by experts in two previous surveys. In Müller and Bostrom (2014), expert respondents (the “All” prediction) predict a 50% probability of high-level human intelligence by 2040 and 90% by 2075. In Grace et al. (2018), experts predict that there is a 50% chance that high-level machine intelligence will be developed by 2061. Plotting the public’s forecast with the expert forecast from Grace et al. (2018), we see that the public perceives high-level machine intelligence arriving much sooner than experts predict.

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. While nearly 80% of

non-experts think that high-level machine intelligence will be built by 2050, only 40% of experts predict that. In our survey, respondents with CS or engineering degrees give a somewhat longer timeline for the arrival of high-level machine intelligence. Nevertheless, these respondents' forecasts are more optimistic than those made by experts from Grace et al. (2018) and show considerable overlap with the overall YouGov forecast.



Support for developing high-level machine intelligence

Procedure

Respondents are asked how much they support or oppose the development of high-level machine intelligence. The responses are coded in the following way:

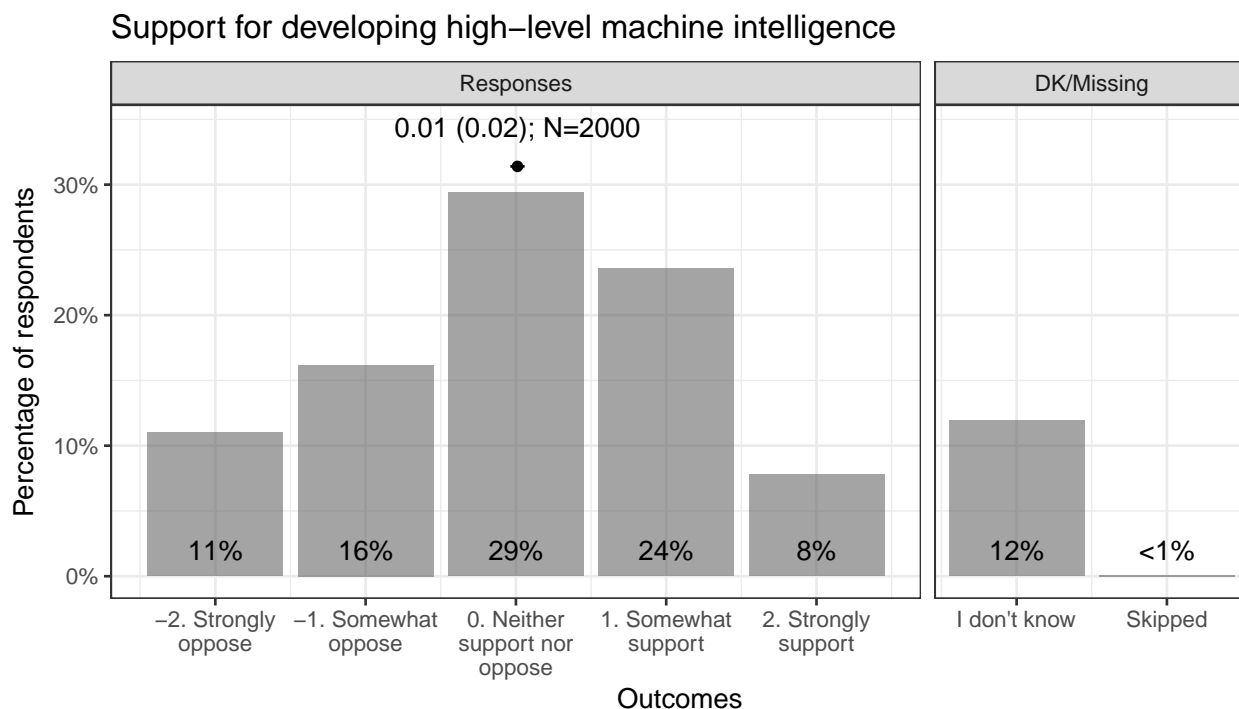
- -2 = Strongly oppose
- -1 = Somewhat oppose
- 0 = Neither support nor oppose
- 1 = Somewhat support
- 2 = Strongly support

Results and discussion

The distribution of responses is displayed in the figure below. The mean outcome and the corresponding standard error are also presented.

Similar to the results about support for developing AI, Americans express mixed support for the development of high-level machine intelligence. About one-third of Americans (32%) somewhat or strongly support the development of AI while 27% somewhat or strongly oppose the development of AI. 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 0.62. The mean level of support for developing AI, compared with the mean level of support for developing high-level machine intelligence, is 0.24 points (SE = 0.03) higher on a five-point scale (two-sided p -value < 0.001).



Demographic predictors of support for developing high-level machine intelligence

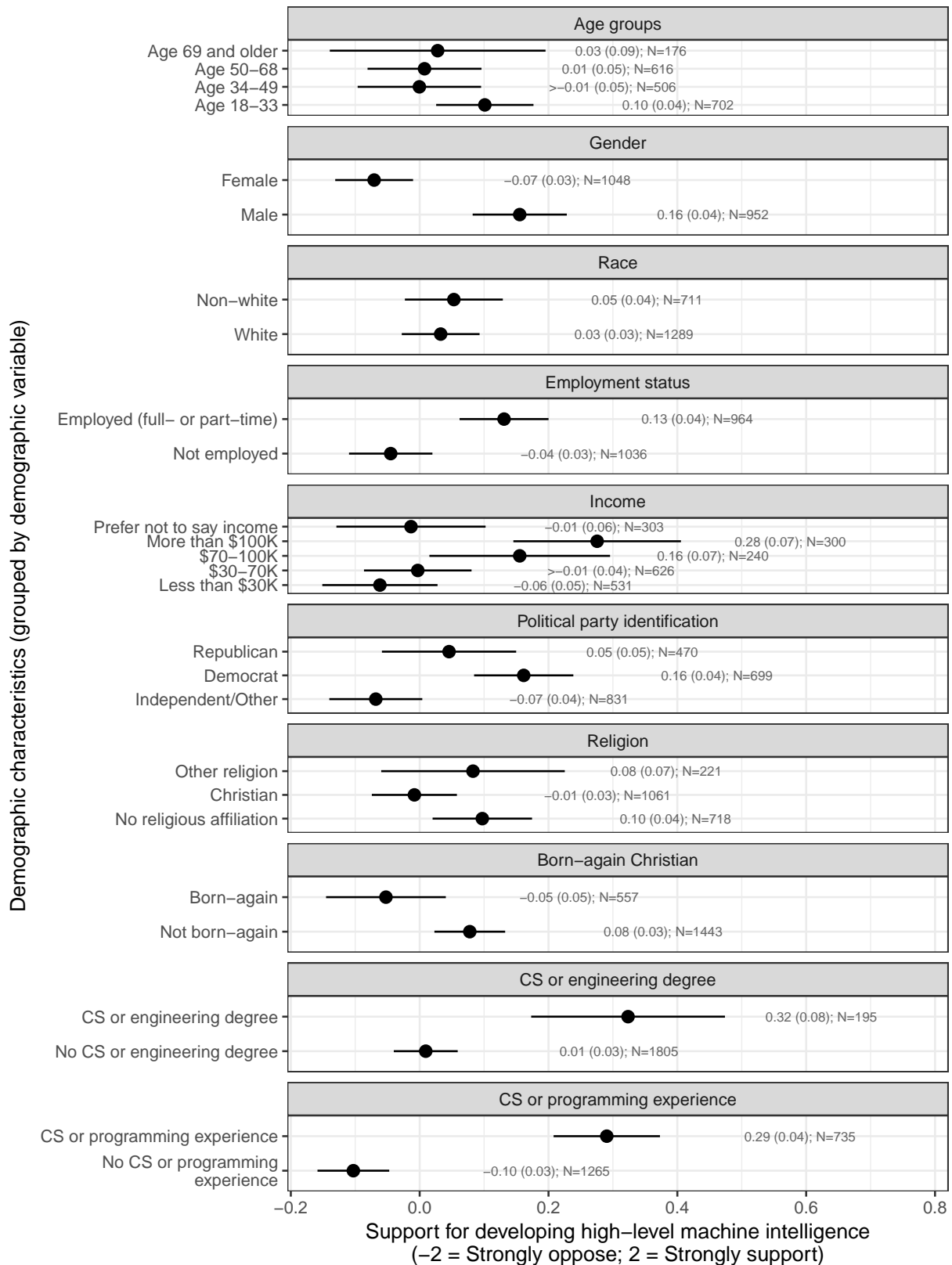
The first figure below displays the average support for developing high-level machine intelligence by demographic groups. Males express greater support than females. Those who are employed express greater support than those who are not employed (i.e., retirees, students, or homemakers). Support is positively correlated with family income. Democrats express greater support for developing high-level machine intelligence than Independents. Those with computer science (CS) or engineering degrees or those with CS/programming experience are more likely support developing the technology than those who do not.

The second figure displays the coefficient plot from a multiple regression model that includes all of the demographic variables. The outcome variable has been standardized, so it has mean 0 and unit variance. Significant predictors correlated with support for developing high-level machine intelligence include:

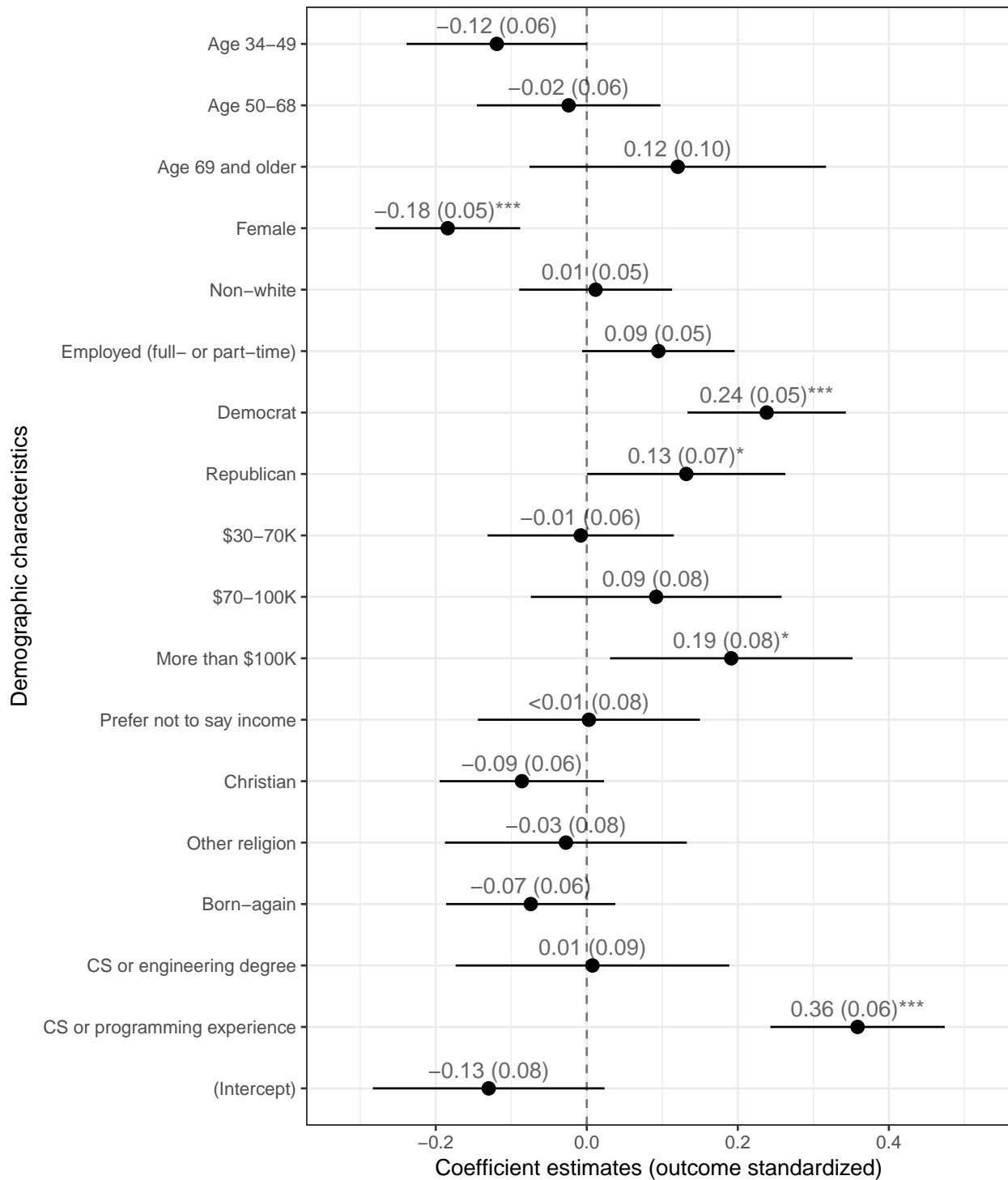
- identifying as a Democrat or a Republican (versus identifying as an Independent)
- having a family income of more than \$100K annually (versus having a family income of less than \$30K annually)
- having CS or programming experience (versus not having such experience).

Being a female (versus being a male) is a significant predictor of opposition to developing high-level machine intelligence.

Predicting support for developing high-level machine intelligence using demographic characteristics: average support across groups



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



Expected outcome of high-level machine intelligence

Procedure

This question seeks to quantify respondents' expected outcome of high-level machine intelligence. Respondents are 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?

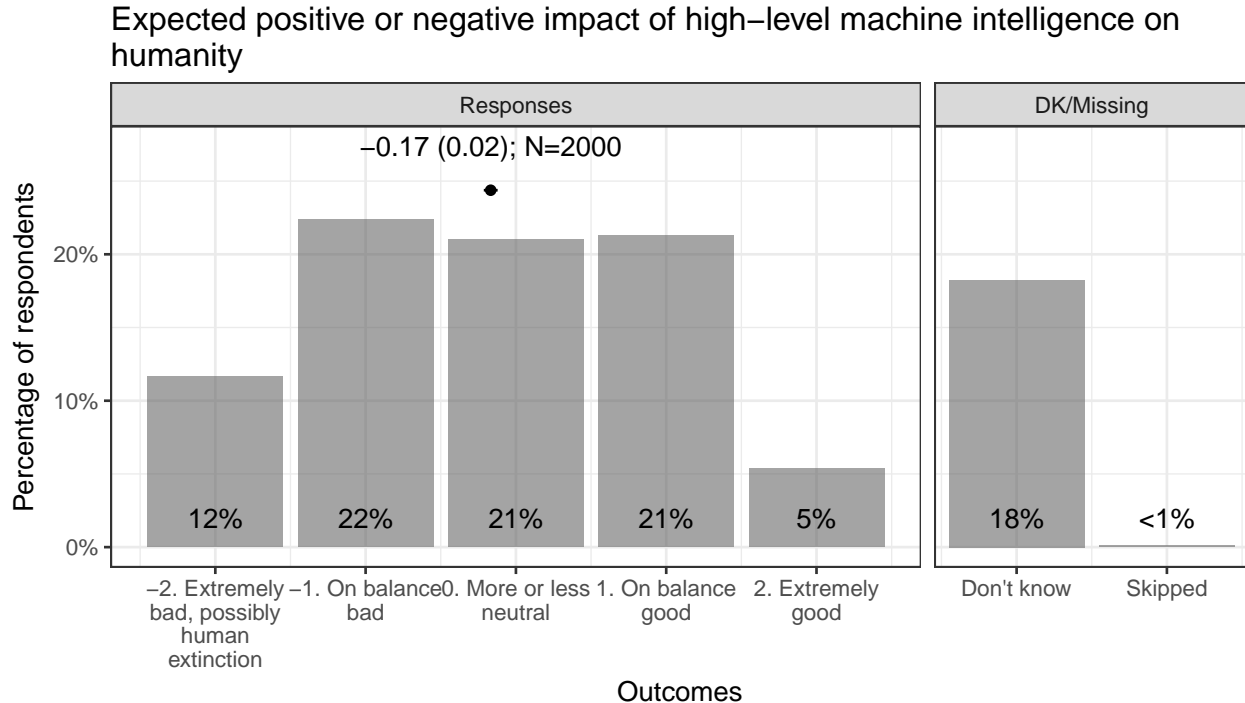
The responses are coded in the following way:

- 2 = Extremely good
- 1 = On balance good
- 0 = More or less neutral
- -1 = On balance bad
- -2 = Extremely bad, possibly human extinction

Results and discussion

The distribution of responses is displayed in the figure below. The mean outcome and the corresponding standard error are also presented.

Americans, on average, expect the high-level machine intelligence will have a harmful impact on balance. Of the 34% who think the technology will have a harmful impact, 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.”



Comparing perceptions of global risks

Procedure

At the beginning of the survey, respondents are asked to consider five out of 15 potential global risks. The purpose of this task is to compare respondents' perception of AI as a catastrophic risk compared with other potential ones. The global risks are selected from the World Economic Forum's "The Global Risk Report 2018". The language used to describe each risk has been simplified so that the average reader can understand it; nevertheless, the substantive content remains the same. We give 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.

This definition borrows 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."

The 15 global risks respondents are asked to consider include:

- Failure of climate change mitigation and adaptation
- Failure of regional/global governance
- Interstate conflict
- Weapons of mass destruction
- Large-scale involuntary migration
- Spread of infectious diseases
- Water crises
- Food crises
- Harmful consequences of AI
- Harmful consequences of synthetic biology
- Cyberattacks
- Terrorist attacks
- Global recession
- Extreme weather events
- Natural disasters

After considering each potential global risk, respondents are asked to evaluate the likelihood of it happening globally within ten years and its impact on several countries or industries. For assessing the likelihood, respondents choose from a multiple-choice list (like in the AI governance challenge question). For the analysis, the multiple-choice answers are converted to the median value of each answer choice range. For instance, the "very unlikely" answer is converted to 2.5%.

For evaluating the potential negative impact of each global risk, respondents are given the following multiple-choice scale:

- 0 = Minimal
- 1 = Minor
- 2 = Moderate
- 3 = Severe
- 4 = Catastrophic

Results and discussion

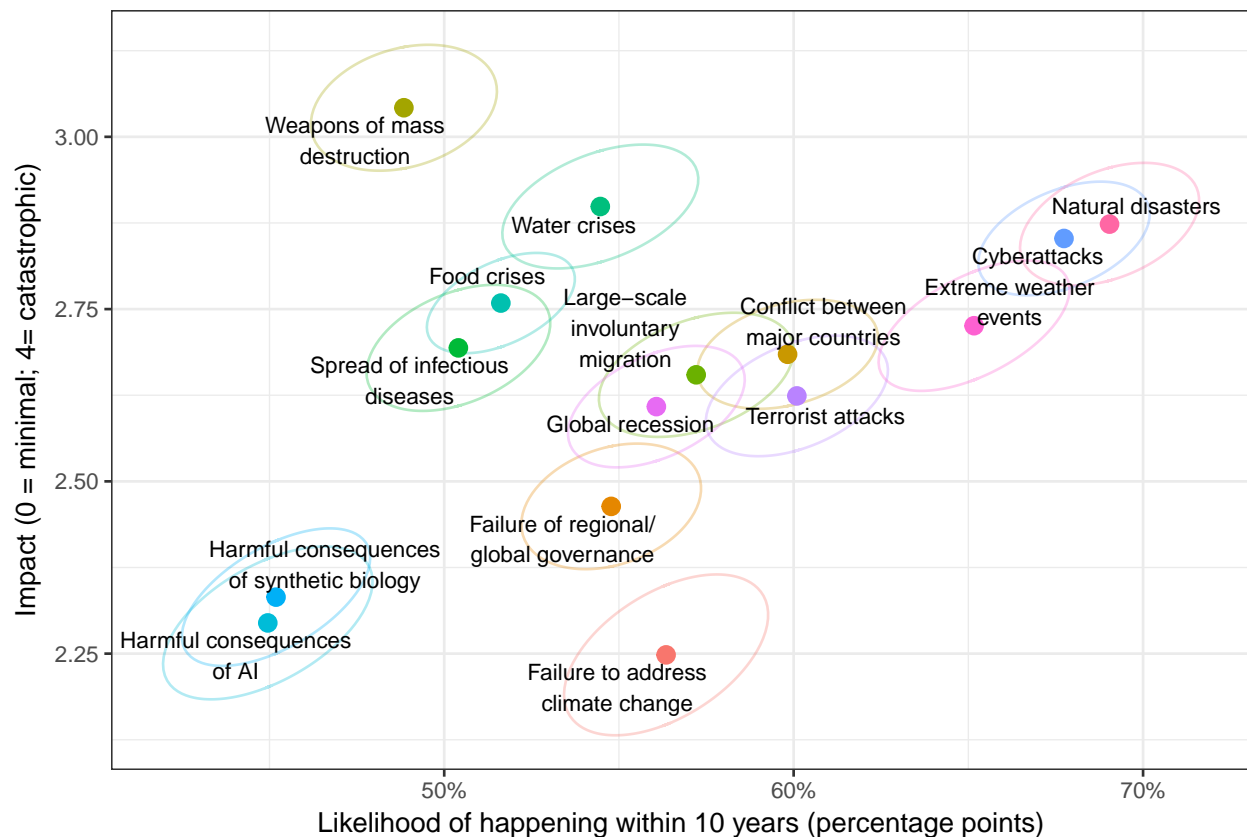
The figure below presents the results from respondents' evaluations of global risks. The x-axis is the perceived likelihood of the risk happening globally within ten years. The y-axis is the perceived impact of the risk. The mean perceived likelihood and impact is represented by a dot. The corresponding ellipse contains the 95% confidence region.

In general, Americans perceive all these risks to be impactful: 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 48% probability of occurring within 10 years. Risks in the upper-right quadrant are perceived to be the most likely as well as the most impactful. These include natural disasters, cyberattacks, and extreme weather events.

The likelihood of these risks appear excessive. These mean responses suggest (assuming independence) that about eight (out of 15) of these global risks, which will 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 actually is (Pinker 2018; H. Rosling, Rönnlund, and Rosling 2018). Another explanation is that it arises as 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.”

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 corner, which contains what are perceived to be lower-probability, lower-impact risks. These risks are perceived to be as impactful (within the next 10 years) as the failure to address climate change. One interpretation of this is that while Americans’ have substantial concerns about the long run impacts of advanced AI, they do not expect these risks to manifest within the next 10 years. We can contrast this interpretation with the above aggregate expectation of a 54% chance of human-level AI within 10 years. One potential interpretation of these two statistics emerges if we assume respondents believe global risks only emerge from human-level AI; in that case, conditional on human-level AI (within 10 years), respondents believe there is about an 80% chance of global risks from AI.

The public's perceptions of 15 potential global risks



Appendix: survey test questions

The text of some important survey questions is included in this Appendix. The Codebook contains the full text of every question in the survey.

AI governance challenges

- **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 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.
- **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 algorithm 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.
- **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.
- **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.
- **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.
- **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.
- **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.
- **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.
- **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.
- **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.

- **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 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.
- **Guarantee a good standard of living for those whose jobs are lost 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.
- **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.

Description of actors

In the support for developing and managing AI questions, respondents are asked about their trust in various actors. We provide descriptions of actors that are not well-known to the public.

- **NATO:** NATO is a military alliance that includes 28 countries including most of Europe, as well as the U.S. and Canada.
- **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.
- **OpenAI:** Open AI is an AI non-profit organization with backing from tech investors that seeks to develop safe AI.
- **AAAI:** Association for the Advancement of Artificial Intelligence (AAAI) is a non-government scientific organization that promotes research in, and responsible use of AI.

Furthermore, Partnership on AI is described as “an association of tech companies, academics, and civil society groups.”

U.S.-China arms race survey experiment treatments

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 AI 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.”

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.

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.”

Global risks

- **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.

Appendix: methodology

The analysis of the real survey data will closely mimic the analysis of the simulated data. Unless otherwise specified, we perform the following procedure:

- Survey weights provided by YouGov are used in our main analysis. For transparency, in the final report, we will perform the analysis without using survey weights.
- For estimates of summary statistics or coefficients, “don’t know” or missing responses are re-coded to the weighted overall mean, unconditional on treatment conditions. Almost all questions have a “don’t know” option. If more than 10% of the variable’s values are “don’t know” or missing, we will include a (standardized) dummy variable for “don’t know”/missing in the analysis. For survey experiment questions, we will compare “don’t know”/missing rates across experimental conditions. In our final report, we will also check whether “don’t know”/missing is correlated with treatment, pre-treatment covariates, and the interactions between treatment and covariates. Our decision is informed by recommendations in the Standard Operating Procedures for Don Green’s Lab at Columbia University.
- Heteroscedasticity-consistent standard errors are reported.
- 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.

References

- Bech, Mickael, and Morten Bo Kristensen. 2009. “Differential Response Rates in Postal and Web-Based Surveys in Older Respondents.” In *Survey Research Methods*, 3:1–6. 1.
- 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). SAGE Publications Sage UK: London, England: 167–90.
- Grace, Katja, John Salvatier, Allan Dafoe, Baobao Zhang, and Owain Evans. 2018. “When Will Ai Exceed Human Performance? Evidence from Ai Experts.” *Journal of Artificial Intelligence Research*. Forthcoming. <https://arxiv.org/pdf/1705.08807.pdf>.
- Müller, Vincent C, and Nick Bostrom. 2014. “Future Progress in Artificial Intelligence: A Poll Among Experts.” *AI Matters* 1 (1). ACM: 9–11. <http://www.nickbostrom.com/papers/survey.pdf>.
- 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). BioMed Central: 351.
- Pinker, Steven. 2018. *Enlightenment Now: The Case for Reason, Science, Humanism, and Progress*. New York: Penguin.
- 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.
- 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). Sage Publications Sage CA: Los Angeles, CA: 212–28.
- Walsh, Toby. 2018. “Expert and Non-Expert Opinion About Technological Unemployment.” *International Journal of Automation and Computing*. Forthcoming. Springer. <https://doi.org/10.1007/s11633-018-1127-x>.