US public opinion of AI policy and risk

Jamie Elsey, David Moss

22 May 2023

Rethink Priorities is an independent, non-partisan, non-profit 501(c)3 think tank centered on policy analysis. As an extension of its work on policy analysis, Rethink Priorities regularly conducts polling and analyses of public attitudes. Rethink Priorities is not funded by any candidate or political party committee and does not poll on behalf of any political candidate or party.

Summary

On April 14th 2023, Rethink Priorities conducted an online poll to assess US public perceptions of, and opinions about, AI risk. The poll was intended to conceptually replicate and extend a recent AI-related poll from YouGov, as well as drawing inspiration from some other recent AI polls from Monmouth University and Harris-MITRE.

The poll covered opinions regarding:

1. A pause on certain kinds of AI research
2. Should AI be regulated (akin to the FDA)?
3. Worry about negative effects of AI
4. Extinction risk in 10 and 50 years
5. Likelihood of achieving greater than human level intelligence
6. Perceived most likely existential threats
7. Expected harm vs. good from AI

Our population estimates reflect the responses of 2444 US adults, poststratified to be representative of the US population. See the Methodology section of the Appendix for more information on sampling and estimation procedures.

Key findings

For each key finding below, more granular response categories are presented in the main text, along with demographic breakdowns of interest.

1. **Pause on AI Research.** Support for a pause on AI research outstrips opposition. We estimate that 51% of the population would support, 25% would oppose, 20% remain neutral, and 4% don't know (compared to 58-61% support and 19-23% opposition across different framings in YouGov's polls). Hence, support is robust across different framings and surveys. The slightly lower level of support in our survey may be explained by our somewhat more neutral framing.

2. **Should AI be regulated (akin to the FDA)?** Many more people think AI should be regulated than think it should not be. We estimate that 70% believe *Yes*, 21% believe *No*, and 9% don't know.
3. **Worry about the negative effects of AI.** Worry in everyday life about the negative effects of AI appears to be quite low. We estimate 72% of US adults worry little or not at all about AI, 21% report a fair amount of worry, and less than 10% worry a lot or more.
4. **Extinction risk in 10 and 50 years.** Expectation of extinction from AI is relatively low in the next 10 years but increases in the 50 year time horizon. We estimate 9% think AI-caused extinction to be moderately likely or more in the next 10 years, and 22% think this in the next 50 years.
5. **Likelihood of achieving greater than human level intelligence.** Most people think AI will ultimately become more intelligent than people. We estimate 67% think this moderately likely or more, 40% highly likely or more, and only 15% think it is not at all likely.
6. **Perceived most likely existential threats.** AI ranks low among other perceived existential threats to humanity. AI ranked below all 4 other specific existential threats we asked about, with an estimated 4% thinking it the most likely cause of human extinction. For reference, the most likely cause, nuclear war, is estimated to be selected by 42% of people. The other least likely cause - a pandemic - is expected to be picked by 8% of the population.
7. **Expected harm vs. good from AI.** Despite perceived risks, people tend to anticipate more benefits than harms from AI. We estimate that 48% expect more good than harm, 31% more harm than good, 19% expecting an even balance, and 2% reporting no opinion.

The estimates from this poll may inform policy making and advocacy efforts regarding AI risk mitigation. The findings suggest an attitude of caution from the public, with substantially greater support than opposition to measures that are intended to curb the evolution of certain types of AI, as well as for regulation of AI. However, concerns over AI do not yet appear to feature especially prominently in public perception of the existential

risk landscape: people report worrying about it only a little, and rarely picked it as a top existential threat.

Extrapolating from these findings, we might expect the US public to be broadly receptive to efforts aimed towards mitigating perceived risks of AI, for example through well-designed government regulation, or efforts to prevent risky arms-race type behavior from companies competing to develop AI.

We view these results as preliminary and our questions intentionally broadly replicated those asked in previous surveys in order to test the robustness of these earlier surveys to different framings. That said, as this topic is complicated and likely novel to most respondents, we think there is significant work to be done to further understand people's views and to ensure that our questions are eliciting meaningful attitudes, rather than pseudo-opinions. To this end, we have an ongoing project employing qualitative methodology to better understand how people think about these questions.

Report

Support vs. opposition for the pause on certain types of AI research

Our framing of this question largely mirrored that of a recent YouGov poll which found 58-61% (depending on framing) would support and 19-23% would oppose a pause on certain kinds of AI development. However, to try to reduce possible demand effects, we only noted that 'some' technology leaders signed an open letter (vs. ">1000" referenced in one YouGov framing), and provided a short piece of information about a countervailing perspective from 'other technology leaders'. One of the three YouGov framings also included a short statement of opposition, with 60% support and 21% opposition.

Our estimate is that 51% of US adults would be supportive of a pause, whereas 25% would oppose such a pause. The spread of separate response options for this question and the exact question framing are shown in Figure 1.

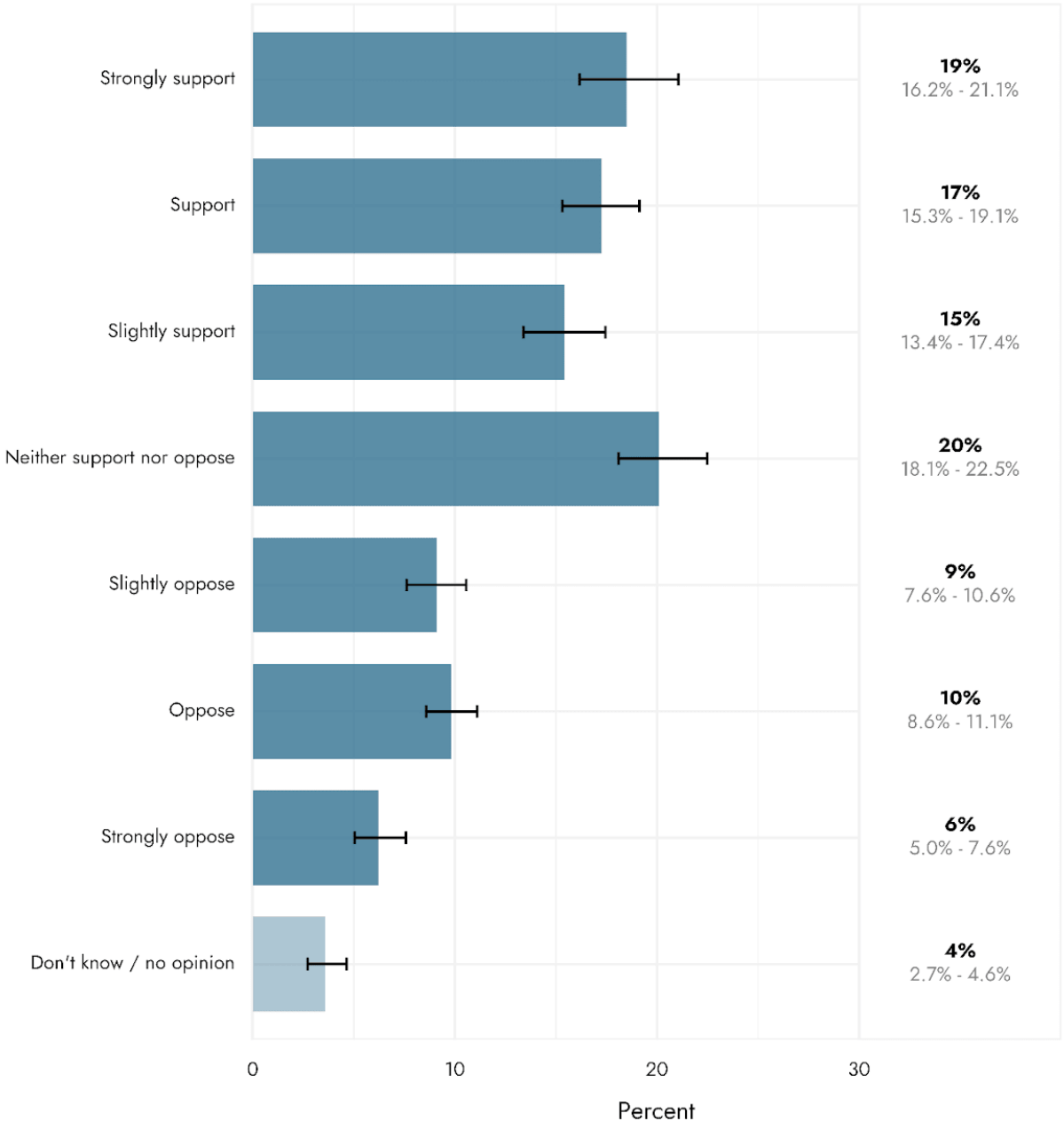
Though our more neutral framing found slightly lower levels of support than YouGov, this suggests that there is considerably more support than opposition for the AI open letter among the US population, and that this is robust to moderately different framings of the issue.

Population estimates of support / opposition for a pause on AI research

Some technology leaders recently signed an open letter calling on AI-labs to pause development of certain large scale AI systems for at least six months worldwide. They cited fears of the "profound risks to society and humanity". They argue we can use this time to better understand these AI systems and put safety measures in place.

Other technology leaders have argued that such concerns are overblown, and such a pause is unnecessary. They argue that such a pause would only hold us back from getting the benefits of developments in AI.

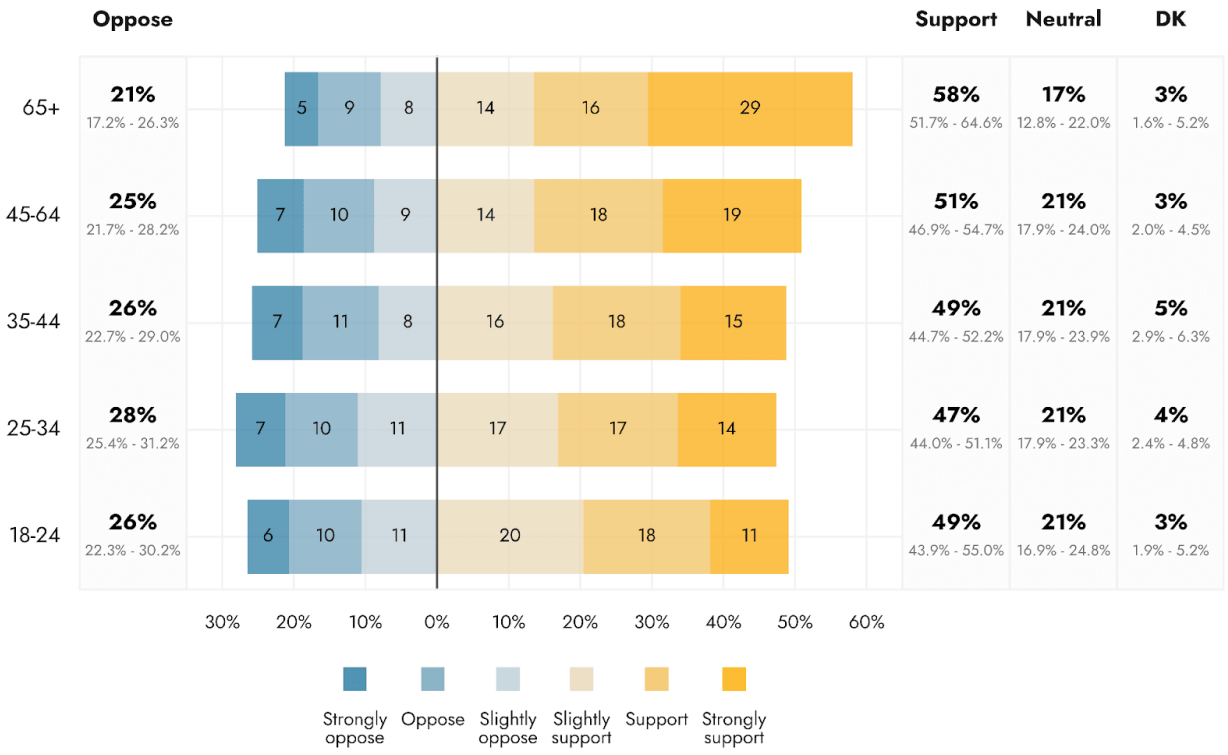
How much would you support/oppose pausing the development of large-scale AI systems for at least 6 months worldwide?

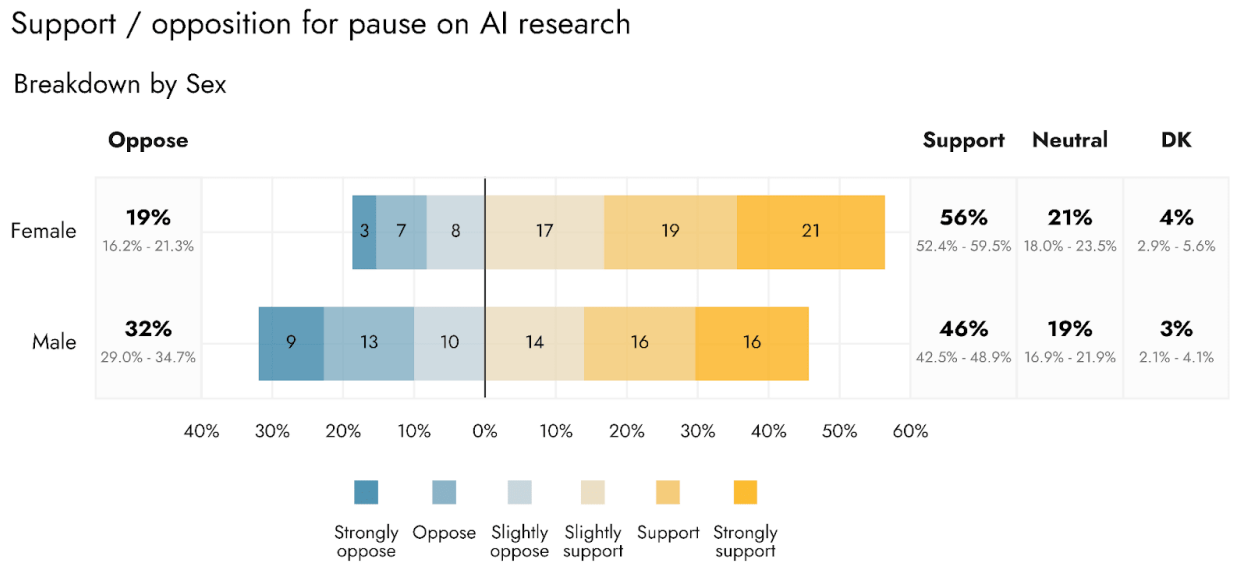


Looking at demographic breakdowns for this outcome, we note that respondents in the oldest age bracket appeared to be most supportive of such a pause, and also that men are expected to be less supportive than women. The difference between men and women is quite substantial, with about 1.5x more opposition among men than women. Nevertheless, across all subgroups we looked at we found that support outweighed opposition.

Support / opposition for pause on AI research

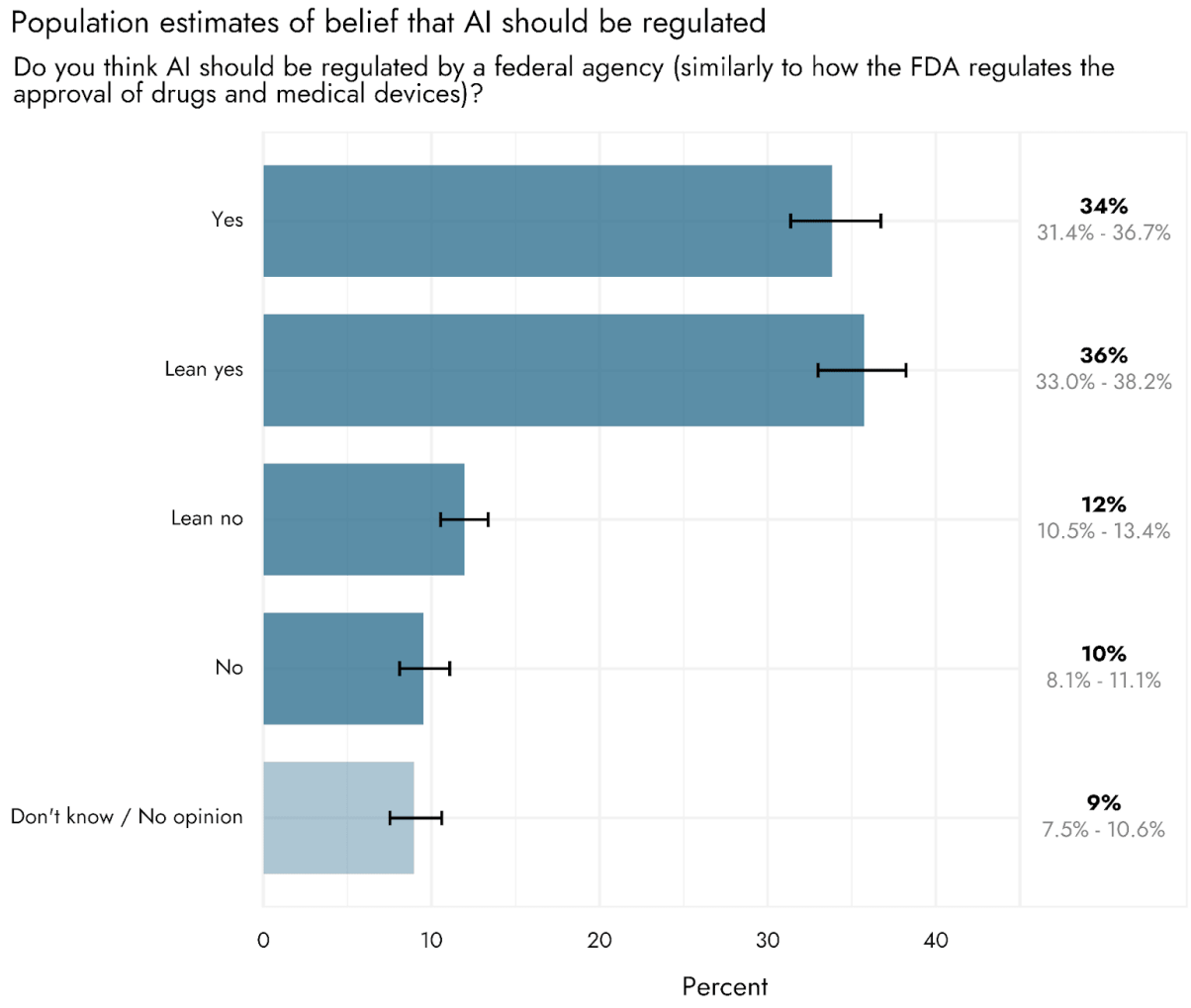
Breakdown by Age





Views on regulation of AI

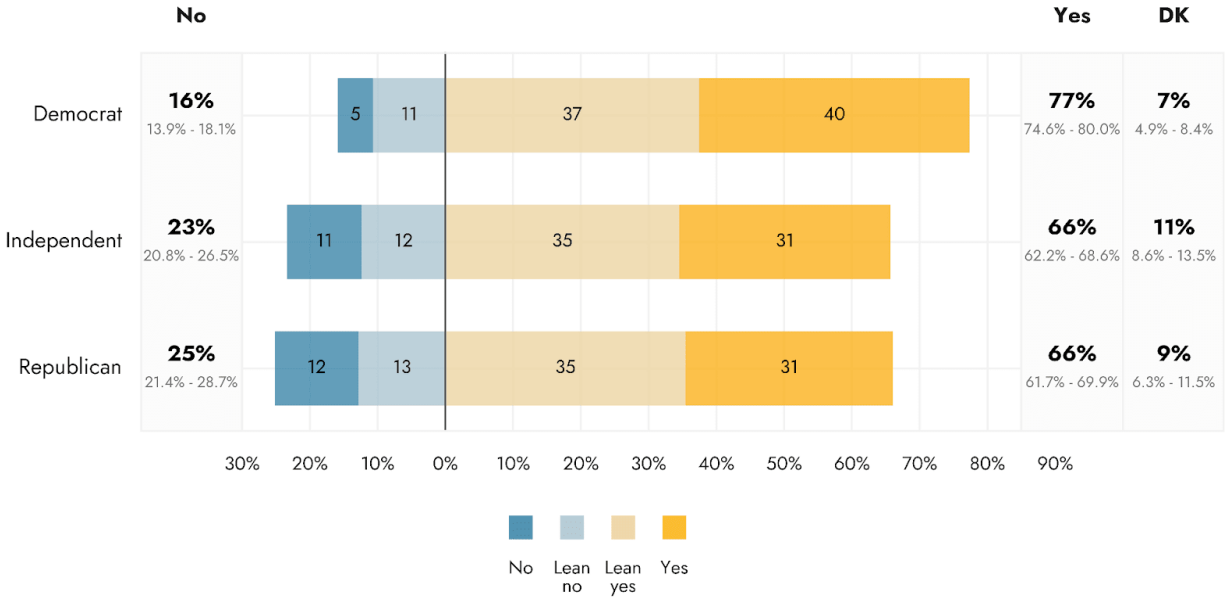
As well as views of a specific proposal for pausing AI development, we asked respondents whether or not they thought AI should be regulated by a federal agency, similarly to how the Food and Drug Administration (FDA) regulates the approval of drugs and medical devices. A Harris-MITRE poll of 2050 US adults, in November 2022, estimated that 82% of US adults would support government regulation of AI. A more recent Monmouth University poll from January 2023 estimated 55% favor, and 41% oppose, the idea of having 'a federal agency regulate the use of artificial intelligence similar to how the FDA regulates the approval of drugs and medical devices'. Using a very similar question framing, we estimate that a sizable majority of US adults would favor federal regulation of AI (70%), with 21% opposed.



Similarly to the question over pausing AI research, we found more females would support such federal regulation. As might be expected, we also estimate that those identifying with the Democratic party would be more favorable of regulation than those who identify as Republican or Independent/otherwise affiliated. Interestingly, we did not find shifts in support of a pause on AI research by political identity. This may be down to people considering factors such as a self-imposed pause (rather than government involvement), or also concerns over the specific form of 'FDA-like' regulatory approaches. In spite of shifts related to some such demographic features, we again found that support was sizable across demographic subgroups.

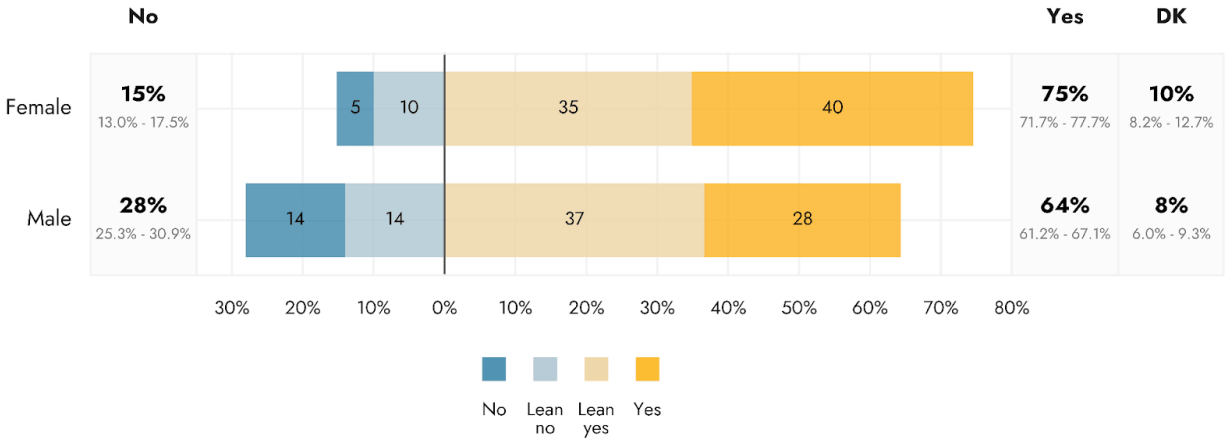
Favor AI being regulated, similarly to the FDA

Breakdown by Political Party Affiliation



Favor AI being regulated, similarly to the FDA

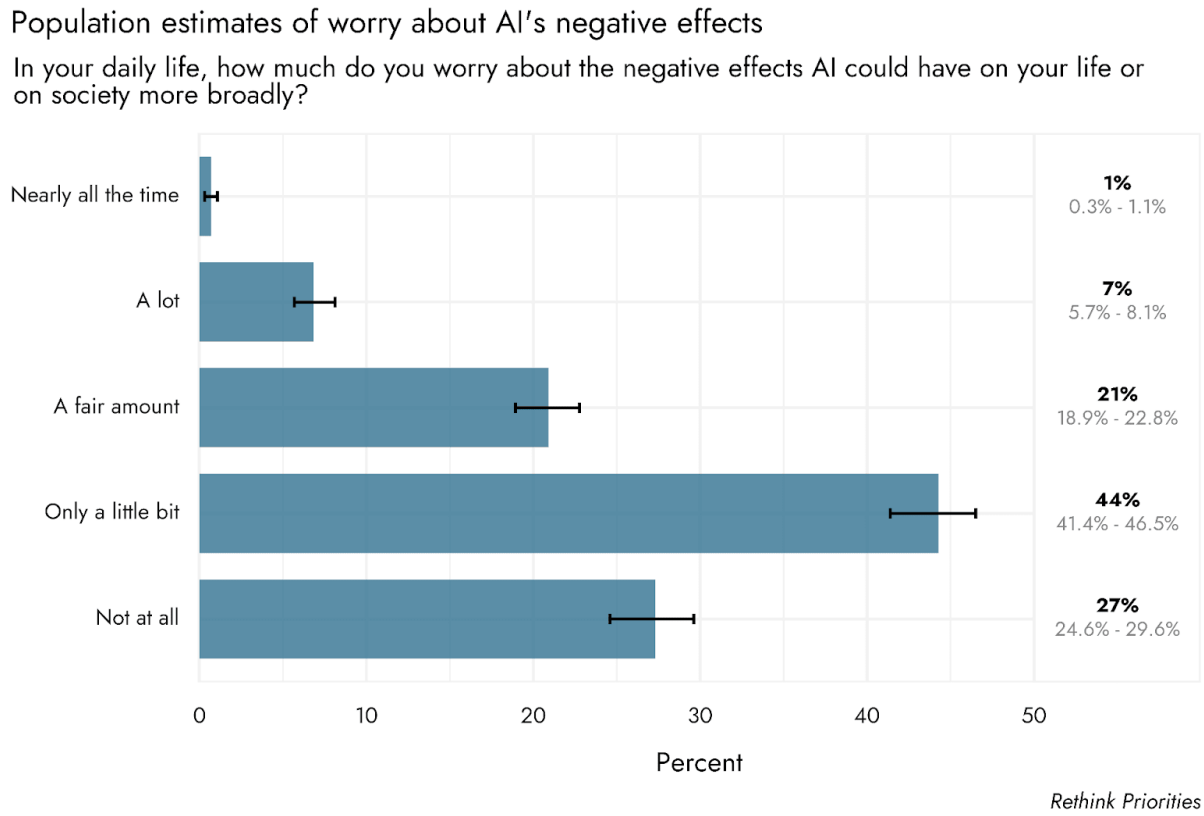
Breakdown by Sex



Worry about AI

The recent YouGov poll included a question regarding how concerned the respondents were about the possibility that AI might end the human race, finding 18% very concerned and 28% somewhat concerned. However, this question may have been interpreted by respondents in a number of different ways: how probable they thought the outcome was, how concerning they thought the outcome would be if it happened, literally how anxious they were about the possibility of the outcome, or some combination of these.

We wanted to instead get a relatively simple indication of how much people were actively worrying about AI, and a separate indication of their perceived likelihood of extinction risk. We asked respondents to indicate how much, in their daily lives, they worry about the negative effects of AI on their life and society more broadly. In a separate question, we asked about people's perceived likelihood of extinction caused by AI.



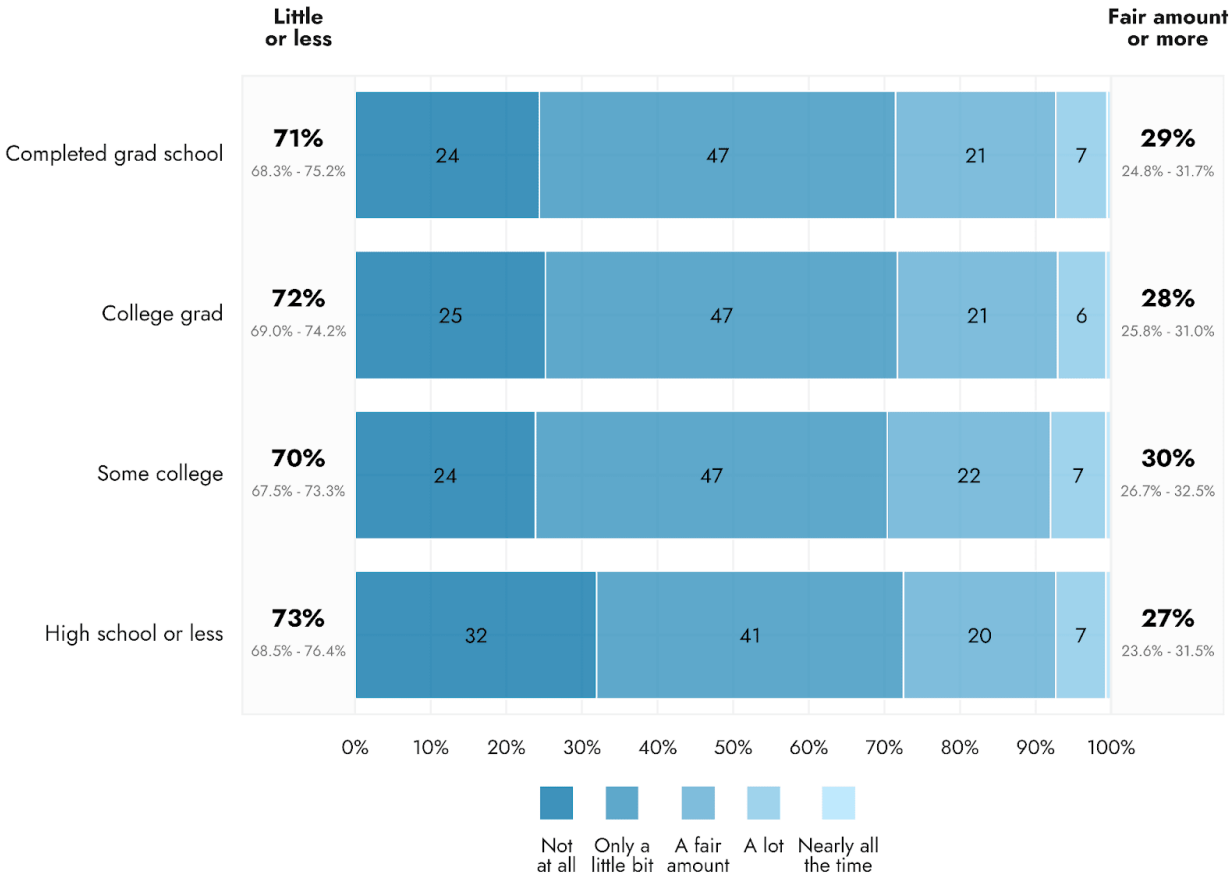
We estimate that the majority of US adults (72%) worry only a little or not at all, with 28% worrying a fair amount or more and less than 10% worrying 'a lot' or more. This might suggest that even if, when prompted, people express concern over certain aspects of AI and possible threats it poses to humanity broadly or to their jobs and the economy (as in the Monmouth poll), AI may not feature prominently among their daily worries. We think this can be important to consider, as when one reads that some sizable proportion of the population is very concerned about AI causing the end of humanity, one may imagine broad, active emotional engagement with this issue. Our findings suggest this might not be the case.

We found relatively little demographic variation in this outcome, although there was slightly greater endorsement for not worrying at all among those with at most high school education, as well as a slightly increasing share of people worrying a little vs. not at all with increasing income levels. People in these brackets may correspond to more highly educated and compensated individuals conducting 'knowledge work' and other white collar jobs that are expected to be most affected by near-term developments in AI.

Of course, even since our survey, there has been additional media coverage of AI risk. Our findings present a snapshot as of April 14th, but the landscape of concern may change substantially with current events or coverage of this issue.

How much people worry about AI in daily life

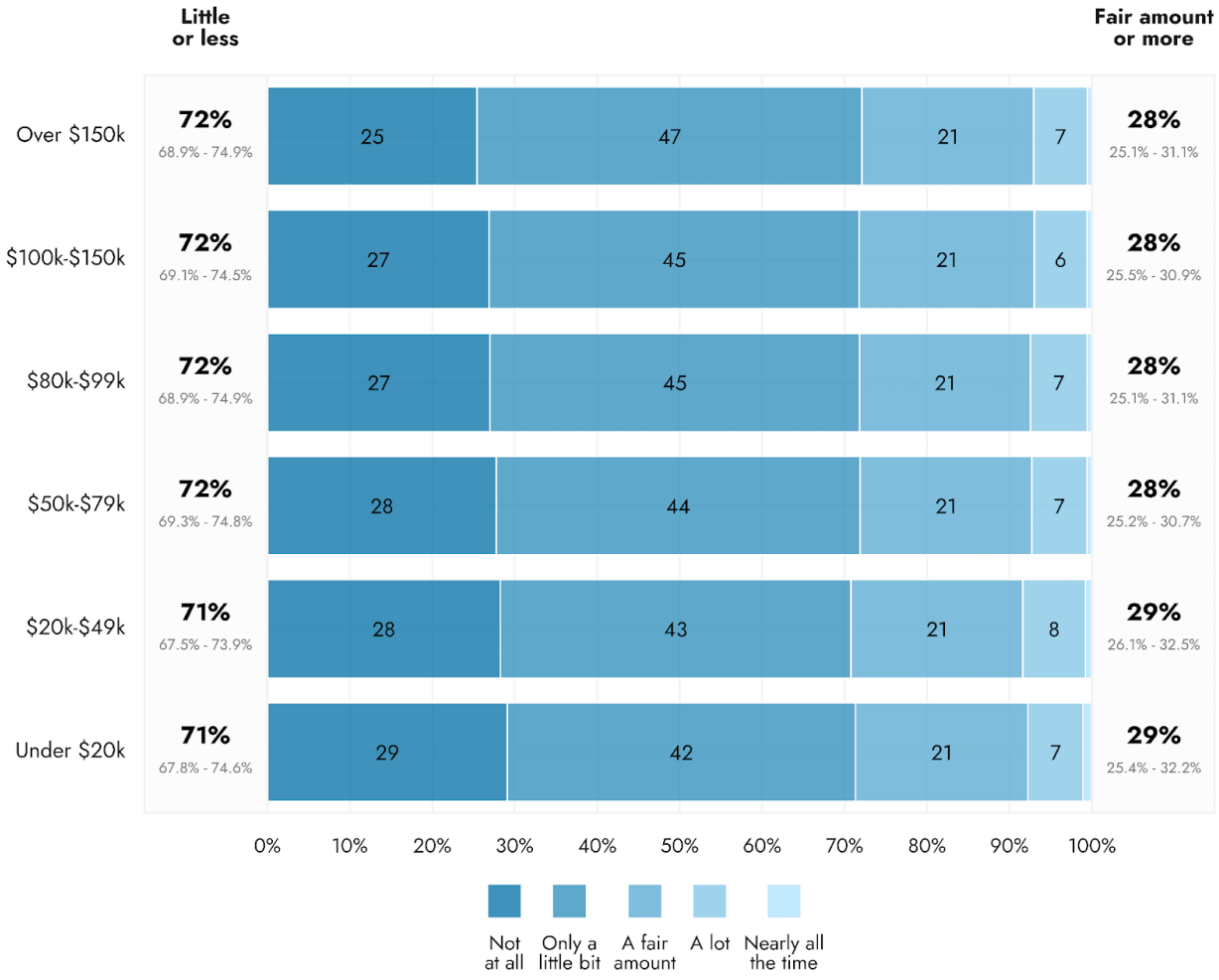
Breakdown by Education



Segments without a corresponding percentage represent at most 0.7% of responses in their subgroup

How much people worry about AI in daily life

Breakdown by Household Income

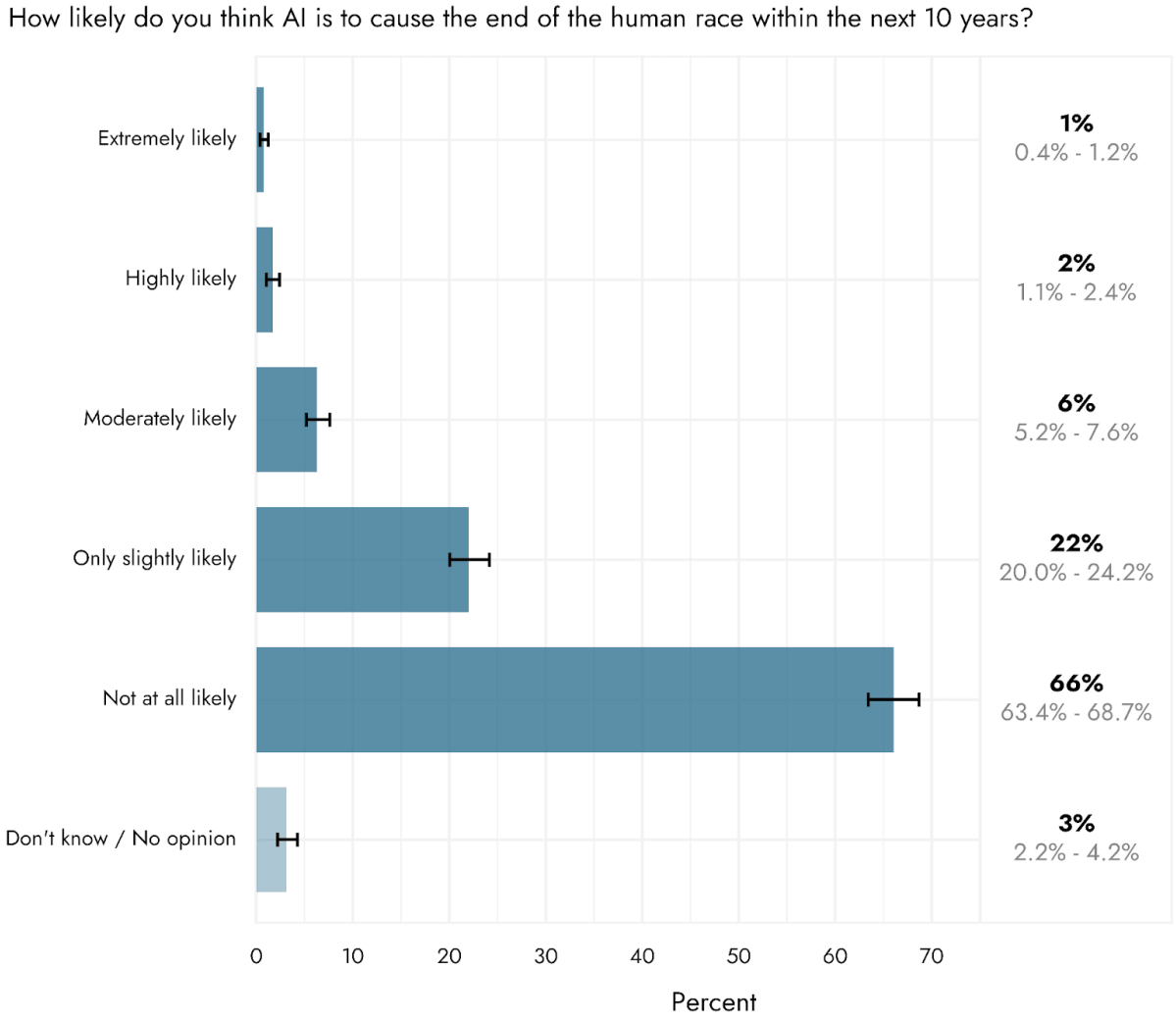


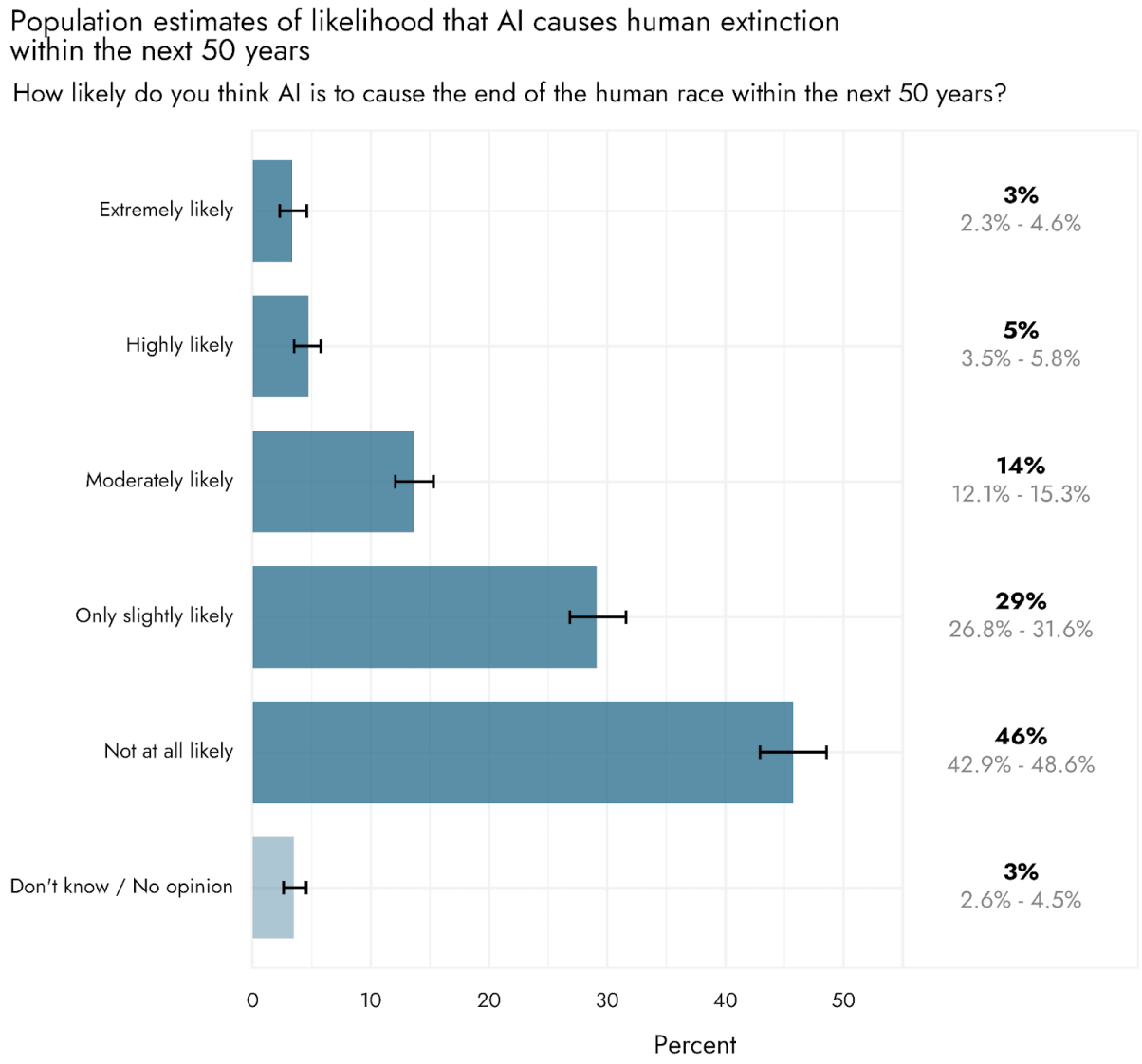
Segments without a corresponding percentage represent at most 1.1% of responses in their subgroup

Expectation that AI might lead to human extinction

We additionally asked respondents to indicate how probable they thought it was that AI would cause human extinction. Estimates from the recent YouGov poll suggested some perhaps surprisingly high estimates of the likelihood of extinction caused by AI: 17% reported it 'very likely' and a further 27% reported it 'somewhat likely'. One thing to note is that the question was not time bound, meaning that respondents may have been considering the possibility of AI representing a serious threat in the very distant future. We asked respondents two versions of this question: one time bound to within the next 10 years, and one time bound to within the next 50 years and, in a later question, simply about what was likely to cause extinction at all.

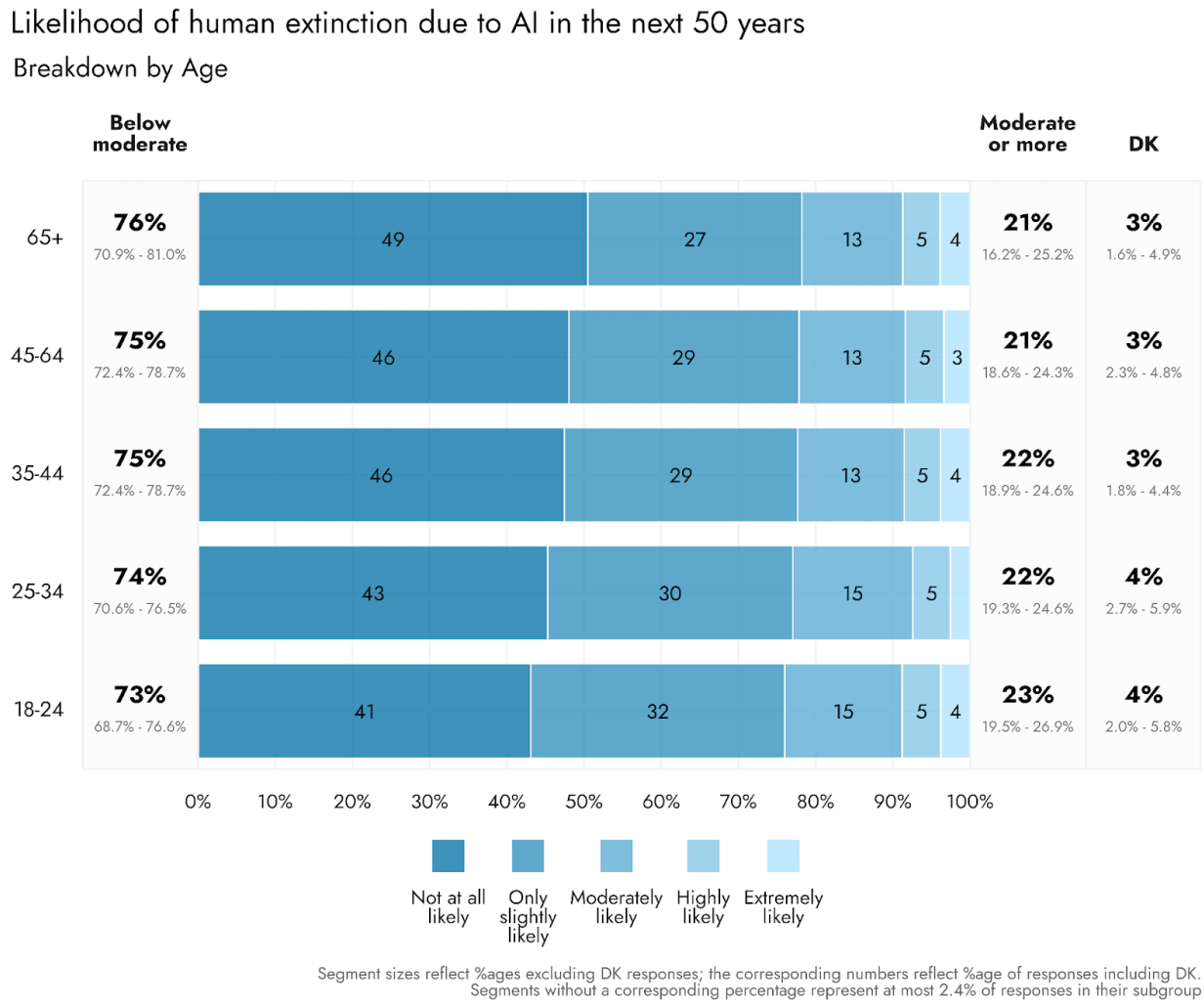
Population estimates of likelihood that AI causes human extinction within the next 10 years





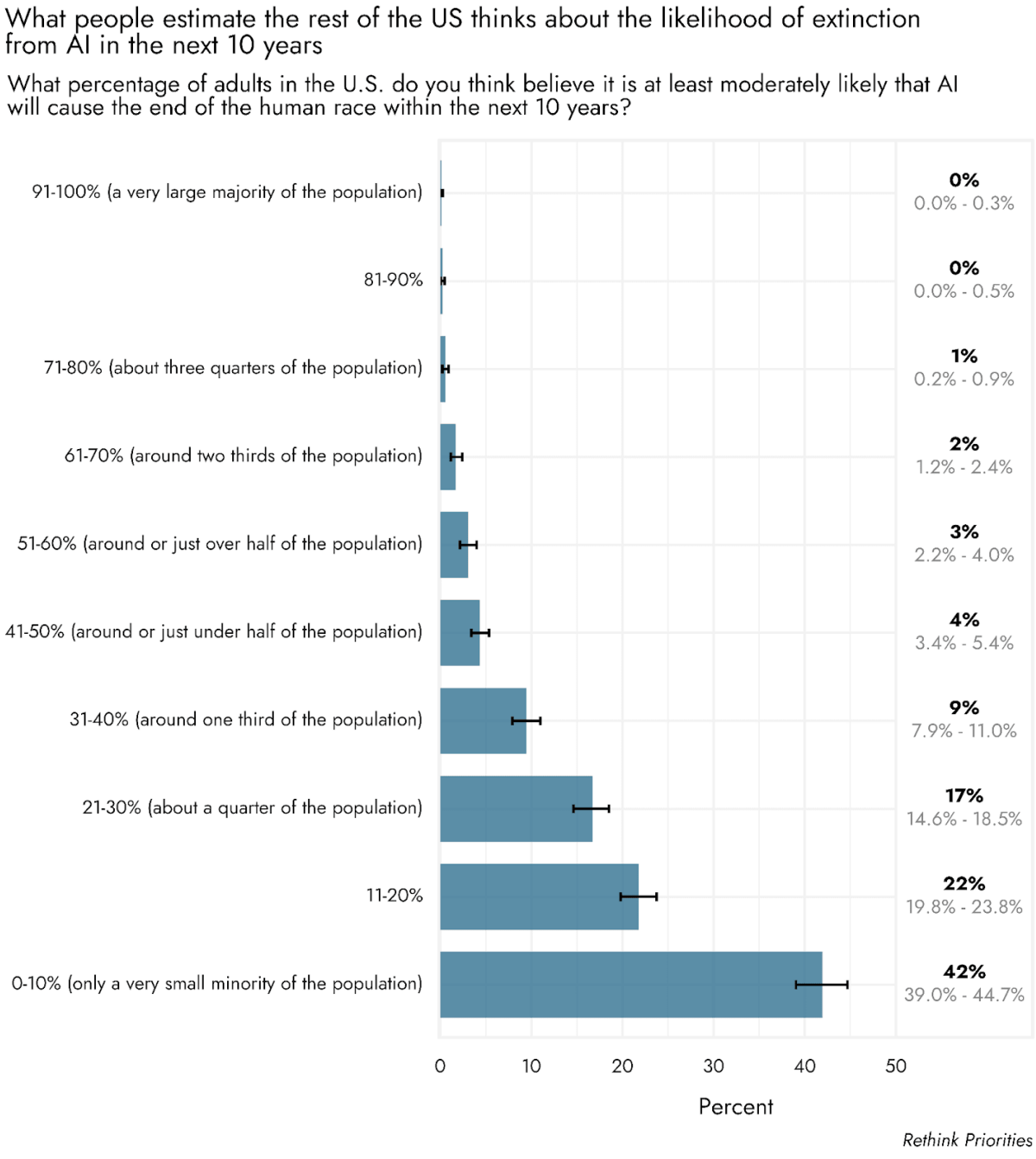
We estimated that the majority of US adults consider it either not at all likely or only slightly likely that AI would lead to human extinction within these timeframes, with the single most selected option being 'not at all likely'. However, the anticipated risk of extinction from AI does increase when moving from the next 10 to the next 50 years. We estimate that only 9% of the population think extinction from AI to be moderately likely or more over the next 10 years. This increases to 22% for the next 50 years.

The sense that extinction from AI was at all likely decreased with increasing age of respondents.



We additionally asked respondents to estimate what proportion of the US population they thought believed it at least moderately likely that AI would lead to human extinction. Assessing predictions about what other people think is one way in which it can be possible to gauge the extent to which people may be misrepresenting their own views. The idea is that even people who may not be willing to endorse an attitude for which they could incur criticism should be willing to honestly report what others' attitudes are. It seems possible

that social desirability concerns could lead to underreporting (e.g., if one fears appearing unhinged) *or* overreporting (e.g., if one fears appearing naive) of worries about AI.



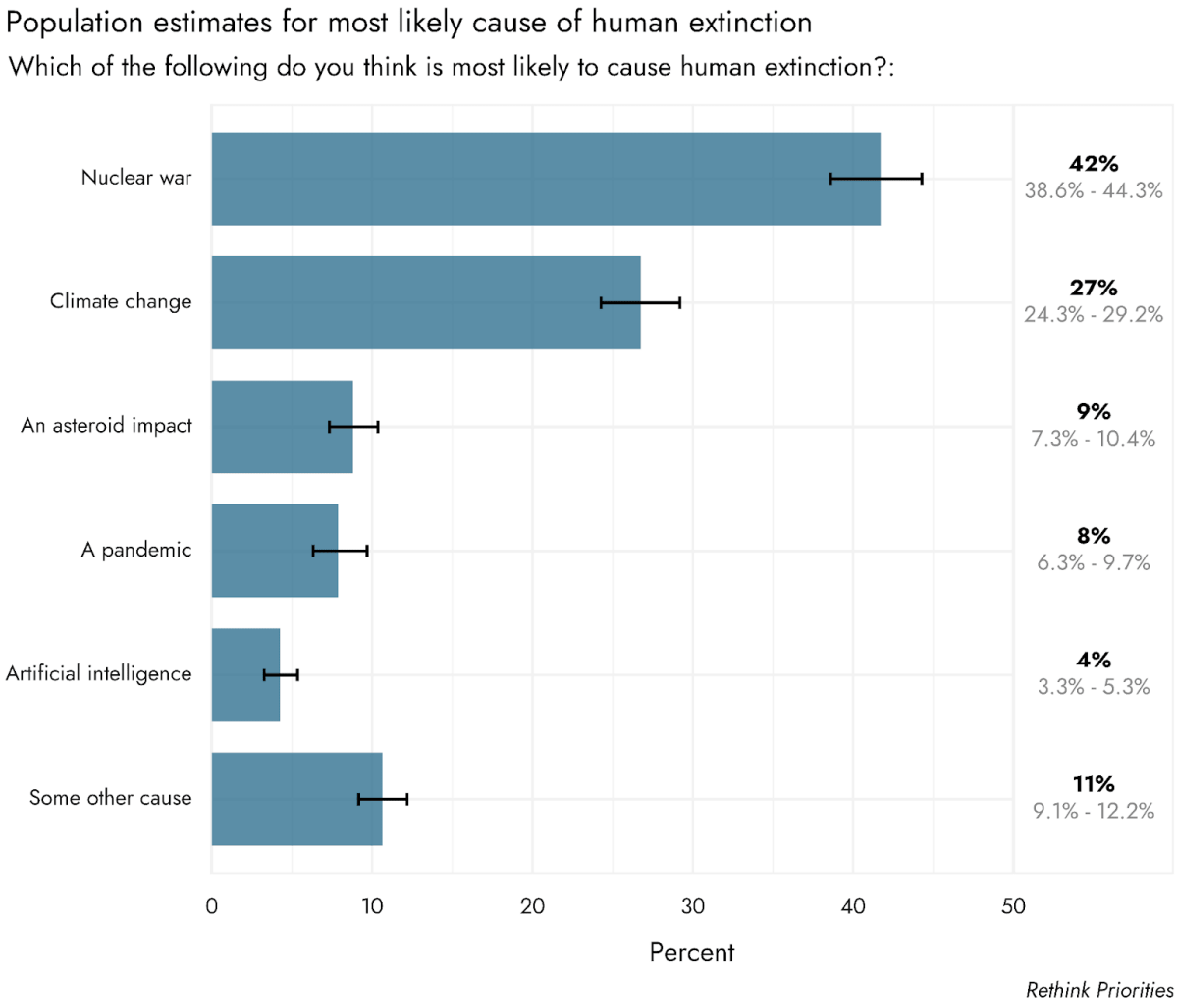
We estimate that most people (64%) expected somewhere between 0%-20% of the US population to grant at least a moderate likelihood of extinction from AI in the next 10 years, with the single-most endorsed option being 0-10% of the population.

In comparison, our results suggested that 9% of the population think it moderately likely or more that AI will cause extinction within the next 10 years (with an error margin just crossing over into the 10-20% bracket). Hence, there is an approximate correspondence between our estimated population level of perceived AI extinction risk based on responses in this survey, and how much people estimate the population to believe in extinction risks.

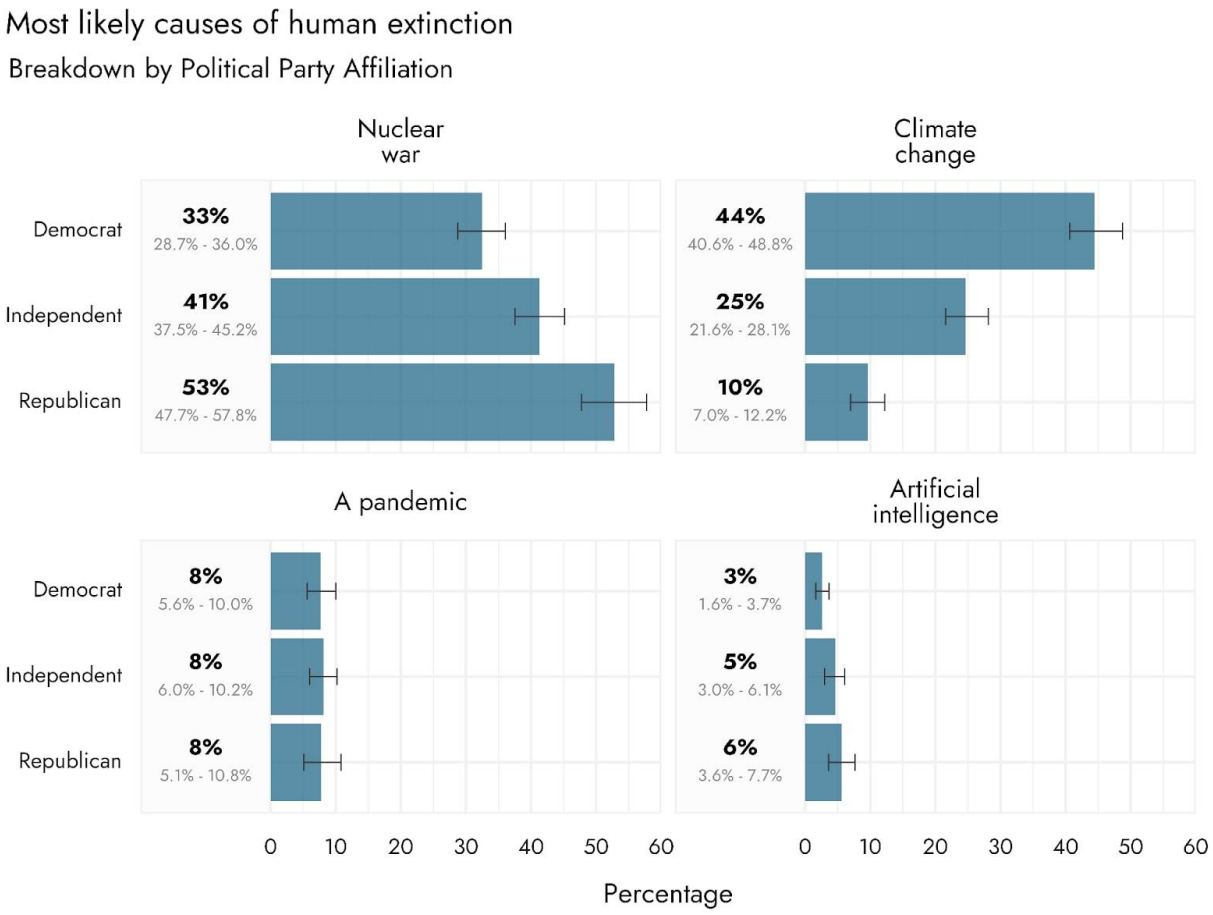
We can make estimates at a more granular level by fitting a beta distribution to the binned responses, thereby generating estimates of actual percentages. When doing this, we estimate that the average expectation among US adults for how much of the population believes AI extinction within 10 years to be 'moderately likely or more' is 18%, with a median expectation of 13%. If taken at face value as a 'wisdom of the crowds' estimate, with reduced risk of social desirability bias, then these numbers might indicate that respondents' reported level of belief in AI extinction was slightly suppressed in their direct responses to the possibility of extinction. However, discrepancies between the direct estimates of how many people anticipate AI extinction, and how much they believe others expect this, could reflect a whole range of factors, not only social desirability. For example, people may simply be overestimating the expectations of others, or there may be more general methodological issues such as the somewhat unfamiliar nature of this type of population prediction question. We are unable to disambiguate these different possibilities with the present data. Future work could examine these explanations further by looking at the association between measures of social desirability and first and third person judgements about AI risk.

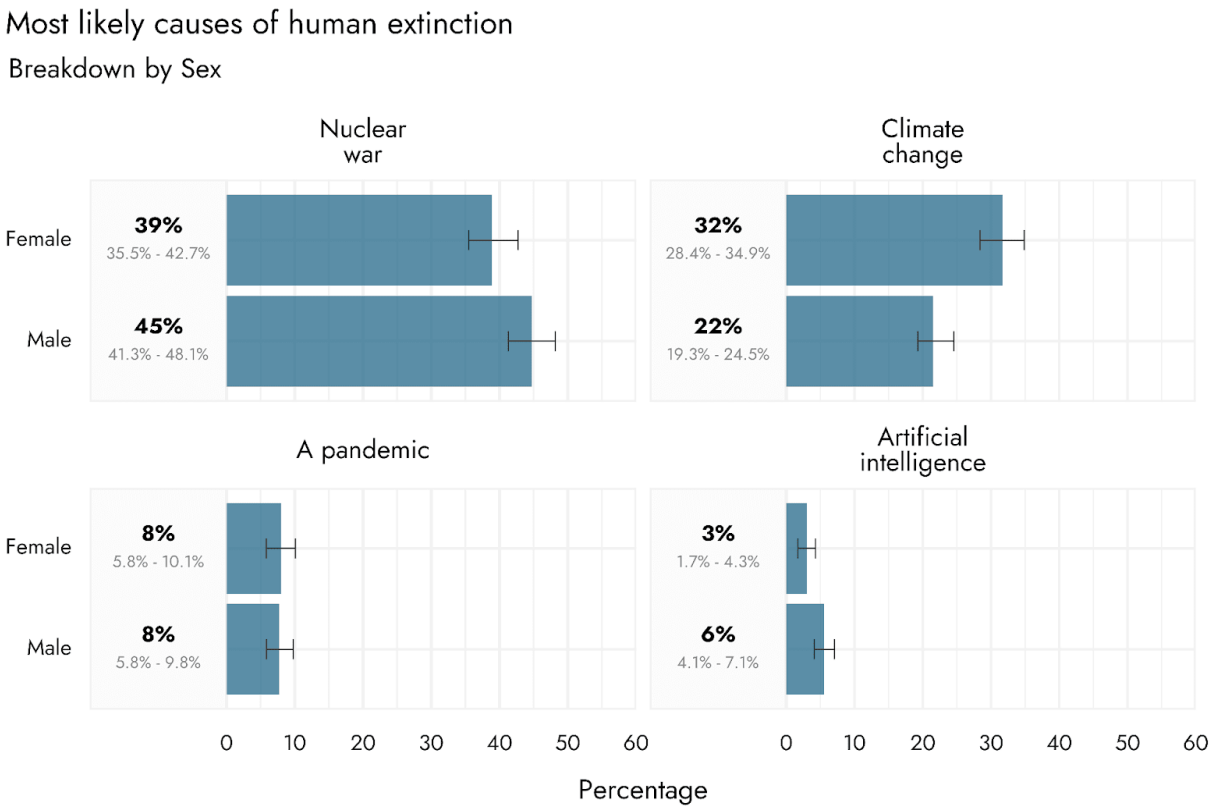
Most likely causes of human extinction

Considering extinction risk from AI relative to other possible causes may also be informative in terms of understanding public perception of AI risk, as well as other existential threats. In the recent YouGov poll, a range of causes were listed, and respondents had to rate how likely they thought each was to result in human extinction. Nuclear weapons were the top specific cause of concern and most likely cause of extinction. AI risk scored higher than alien invasions and infertility but lower than asteroid impacts. We simply had respondents pick the single most likely option among several possible causes of human extinction. Consistent with the YouGov findings, nuclear war ranked top among the choices, followed by climate change. AI risk was again outranked by Asteroid Impact.



The clearest demographic trends were for political differences, as well as differences between male and female respondents. Specifically, Democrats were far more likely than Republicans to endorse climate change as a possible cause of extinction, with Republicans more likely to endorse nuclear war. Independent voters were in between. Some of the Republicans not endorsing climate change also seemed to shift into AI risk, with an estimated 3% of Democrats vs. 6% of Republicans ranking AI risks as the most likely cause of human extinction. Male vs. Female respondents showed a similar pattern of responses as Republicans vs. Democrats.

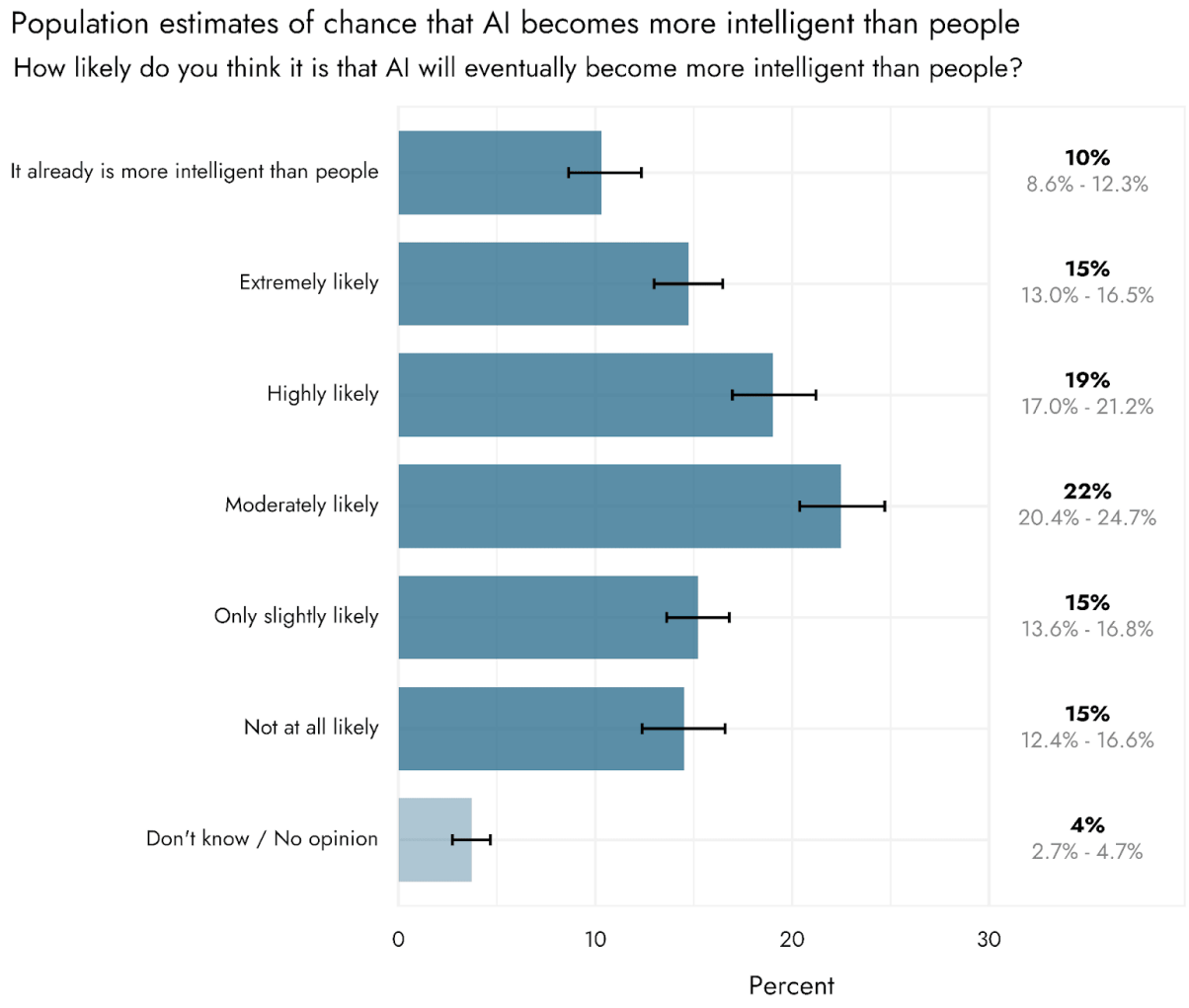

Most likely causes of human extinction
Breakdown by Political Party Affiliation
Nuclear war
Climate change
A pandemic
Artificial intelligence
Democrat
Independent
Republican
33%
28.7% - 36.0%
41%
37.5% - 45.2%
53%
47.7% - 57.8%
44%
40.6% - 48.8%
25%
21.6% - 28.1%
10%
7.0% - 12.2%
8%
5.6% - 10.0%
8%
6.0% - 10.2%
8%
5.1% - 10.8%
3%
1.6% - 3.7%
5%
3.0% - 6.1%
6%
3.6% - 7.7%
0
10
20
30
40
50
60
Percentage


Most likely causes of human extinction
Breakdown by Sex
Nuclear war
Climate change
A pandemic
Artificial intelligence
Female
Male
39%
35.5% - 42.7%
45%
41.3% - 48.1%
32%
28.4% - 34.9%
22%
19.3% - 24.5%
8%
5.8% - 10.1%
8%
5.8% - 9.8%
3%
1.7% - 4.3%
6%
4.1% - 7.1%
0
10
20
30
40
50
60
Percentage

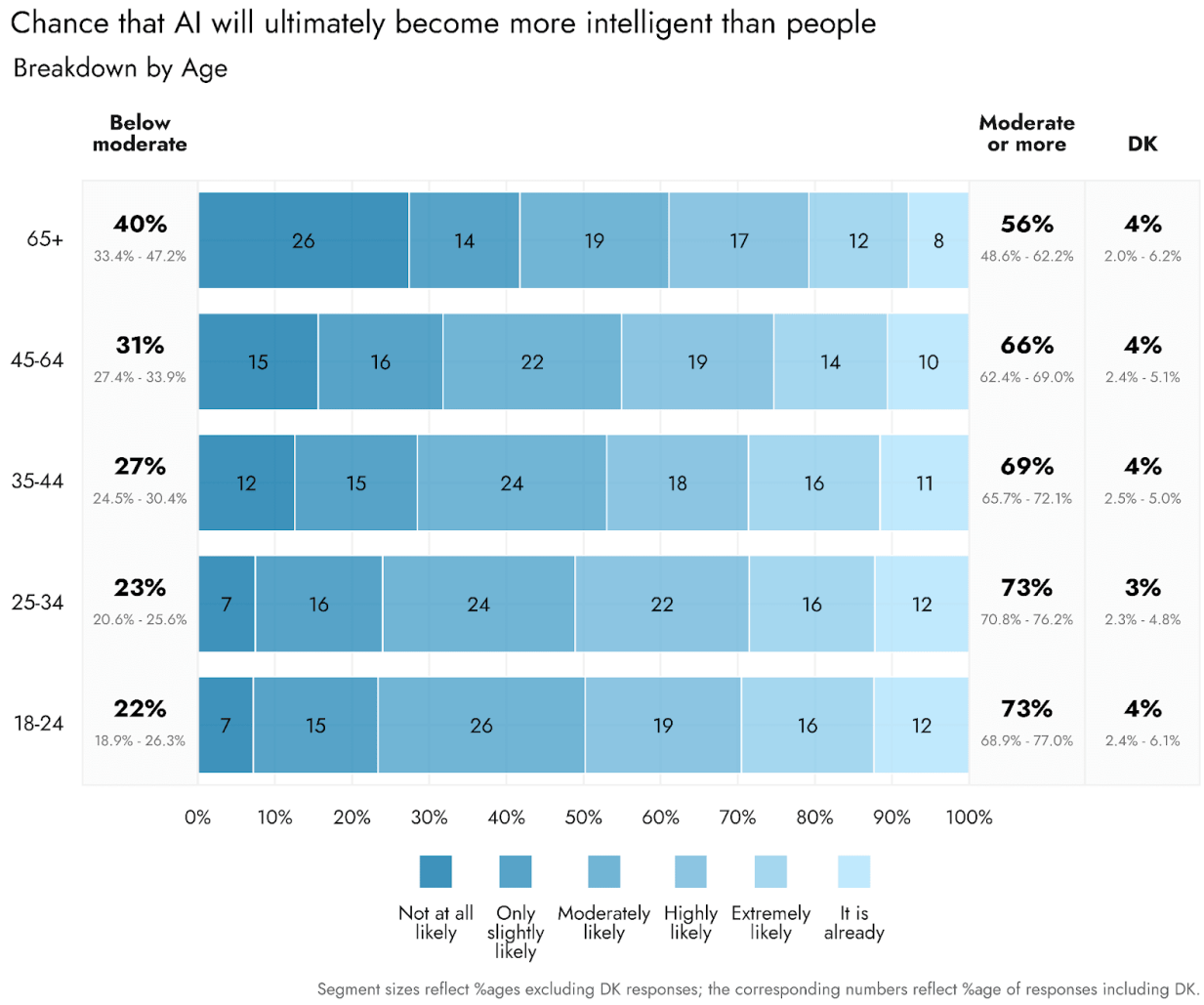
Greater than human intelligence

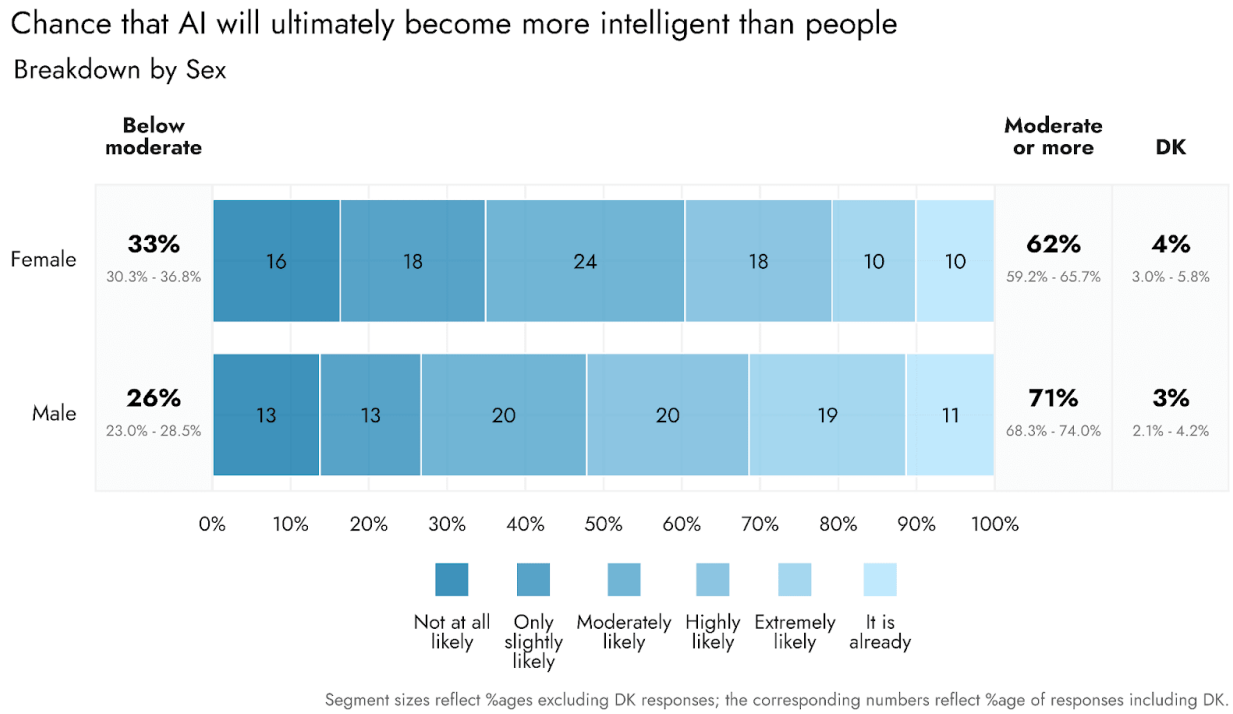
As many concerns over AI may depend on the extent to which people anticipate AI becoming competitive with human intelligence, we additionally asked respondents how likely they think it is that AI will ultimately become more intelligent than people.

We estimate that 67% of US adults think it moderately likely or more that AI will become more intelligent than people, with more than 40% of people thinking this outcome highly likely or more.



Older adults seemed more skeptical of this possibility than those in younger age brackets by a substantial margin. Additionally, female respondents were more skeptical than males. This is of interest given that females nevertheless favor a pause on AI development to a greater extent than males.



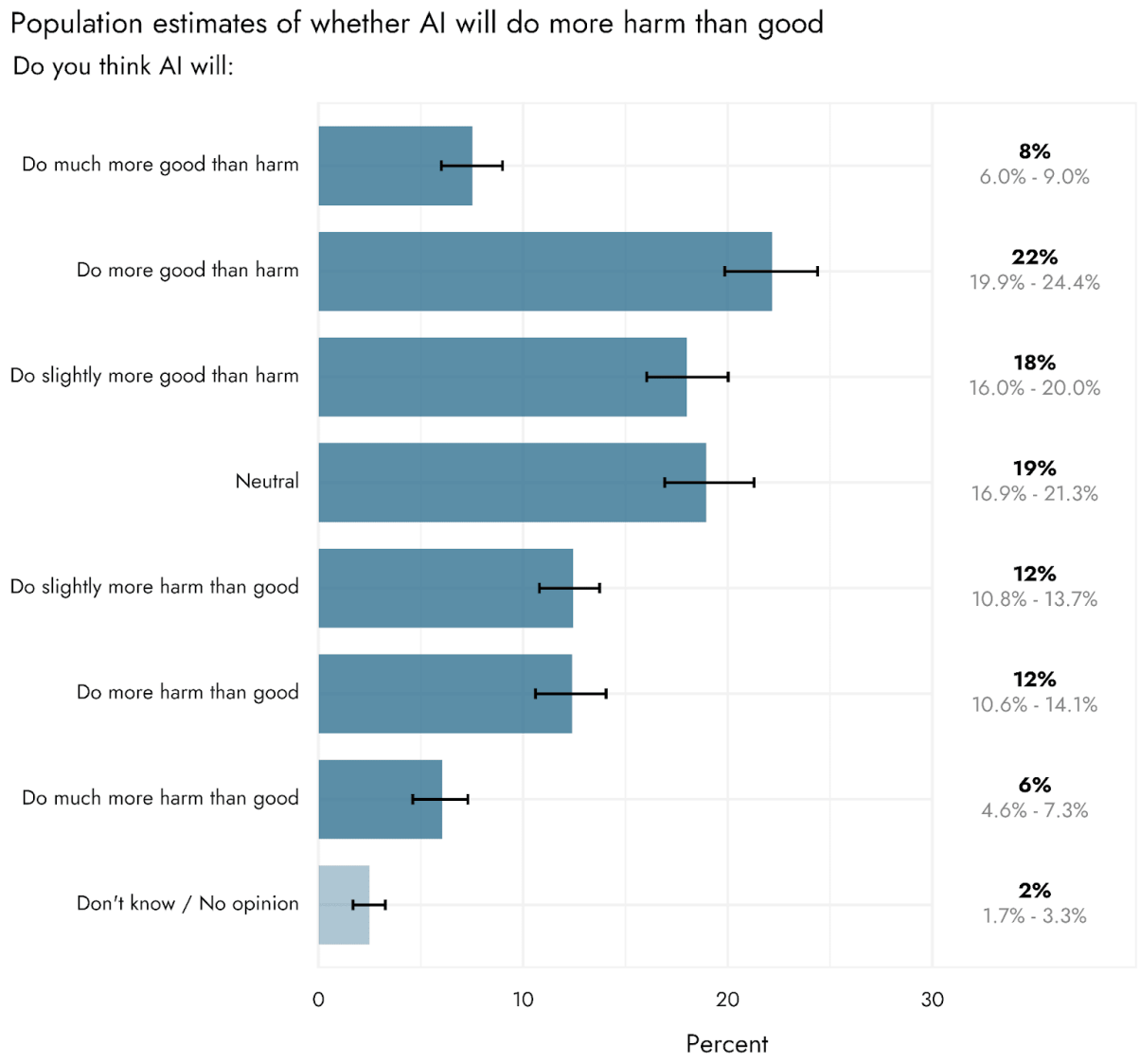


Good vs. Harm from AI

Finally, beyond catastrophic outcomes from AI, we were interested in public perceptions of the general good vs. harm that artificial intelligence might do. Our question was framed similarly to the recent Monmouth University poll, which estimated that just 9% of US adults expect more good than harm from developing artificial intelligence. An estimated 46% expected equal goods and harms, and 41% expected more harm than good. We expanded on the question's response options by allowing people to endorse more gradations of good and harm, which may have inflated estimates of equality.

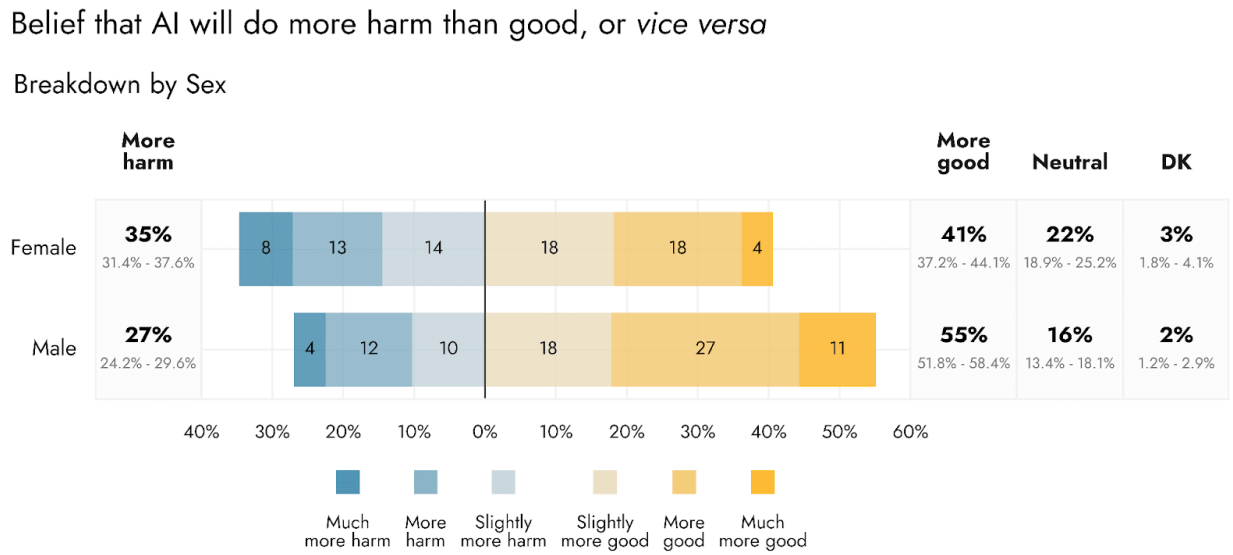
In contrast to the Monmouth poll, we estimate just 19% of the population are neutral on this issue, and that 48% lean in the direction of more good than harm, with 31% expecting more harm than good. This is the most substantial deviation from previous polls that we have observed in these AI-related questions. It is not immediately clear what might be the cause of this discrepancy, although the Monmouth poll was conducted 2 months ago and also included fewer respondents than our poll (805). The Monmouth poll also described AI as

the creation of 'computers that can think for themselves', which may be more conducive to imagining hostile or frightening agentic AI. If accurate, our findings indicate the US public is not as pessimistic about AI as some other polls might suggest. However, it is also plausible that this is dependent on exactly how AI is construed by the respondent.



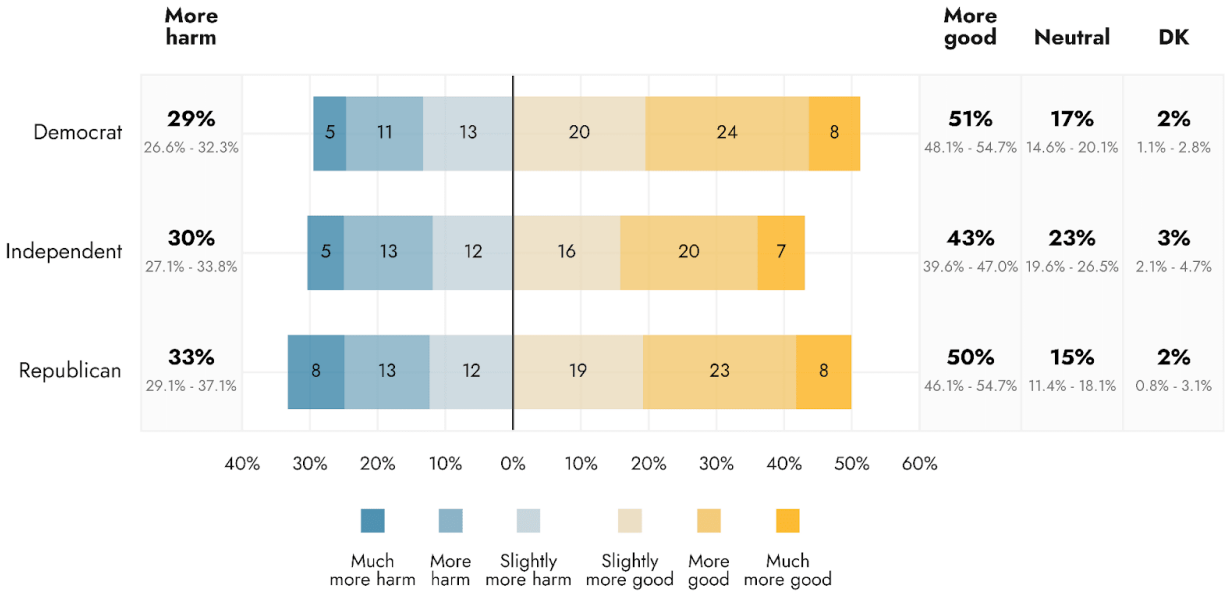
Male respondents were more likely to have a positive expectation for AI than female respondents, and both Democrats and Republicans had more positive expectations than

Independently affiliated respondents. Female respondents were clearly more negative and also more likely to endorse neutrality. For political affiliation, the difference between groups seemed largely due to Independents being more likely to pick the Neutral option than for them to have reliably negative expectations.



Belief that AI will do more harm than good, or *vice versa*

Breakdown by Political Party Affiliation

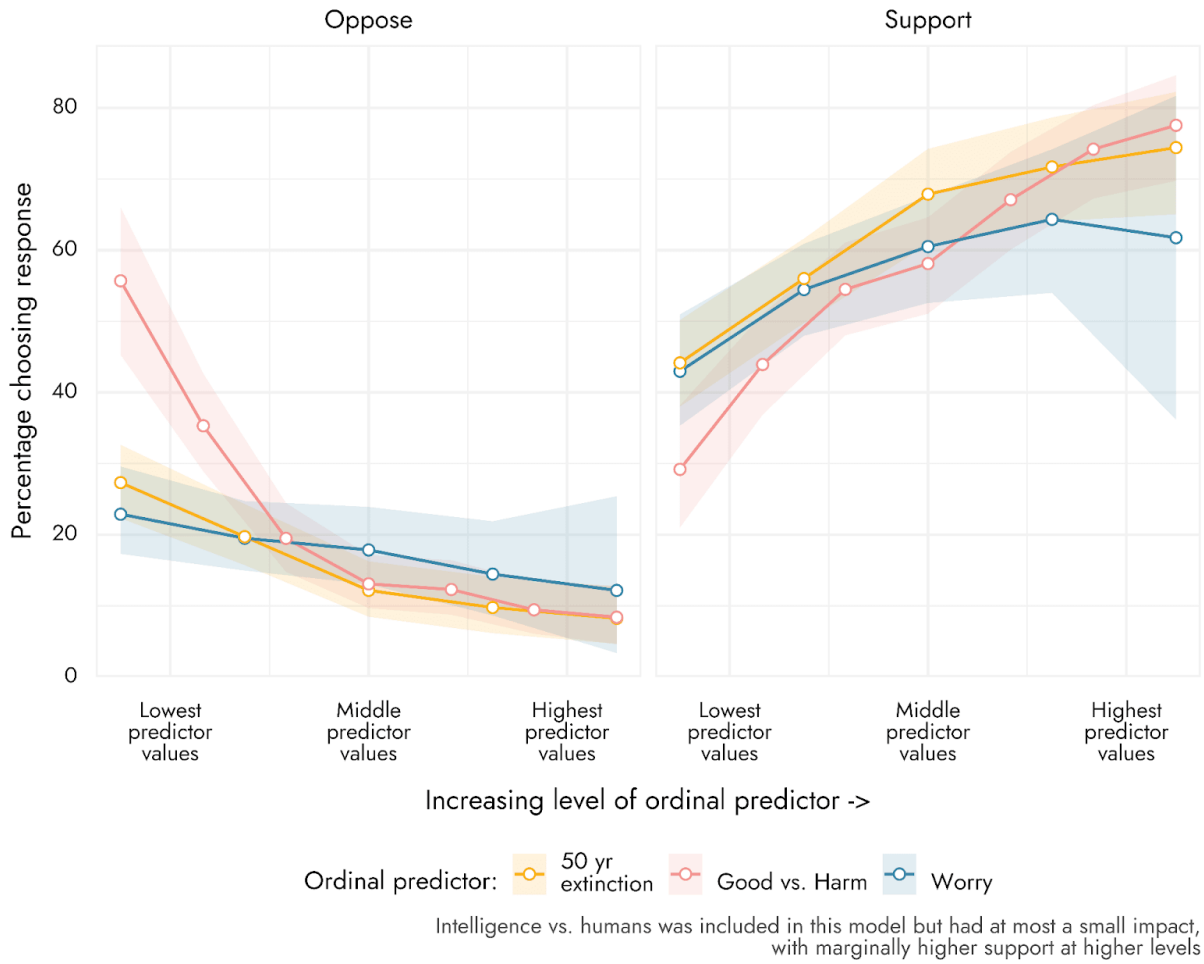


Associations among AI attitudes

In addition to estimating public opinion, we also assessed some relationships of potential interest between the AI-related outcomes (see the Methodology section for a description of these models). For support of a pause on certain kinds of AI research, we found that higher expectations of harm, more worry, and greater expectations of extinction in the next 50 years from AI were positively associated with support for a pause. Those who reported a clear expectation that AI would do more good or much more good than harm were especially likely to oppose a pause.

Support for pausing AI research depending on *worry*, *expectation of extinction*, and *expecation of good vs. harm*

Higher expectations of good from AI were associated with substantially greater opposition to pausing AI development. Higher levels of worry and expectation of extinction were associated with higher support.



Intelligence vs. humans was included in this model but had at most a small impact, with marginally higher support at higher levels

With respect to believing that AI should be regulated in a manner similar to how the FDA oversees food and drugs, the expectation of AI doing more good than harm was again associated with heightened disagreement with regulation. Worry and the belief in extinction from AI tended again towards being positively associated with support for regulation, but these associations were not robust.

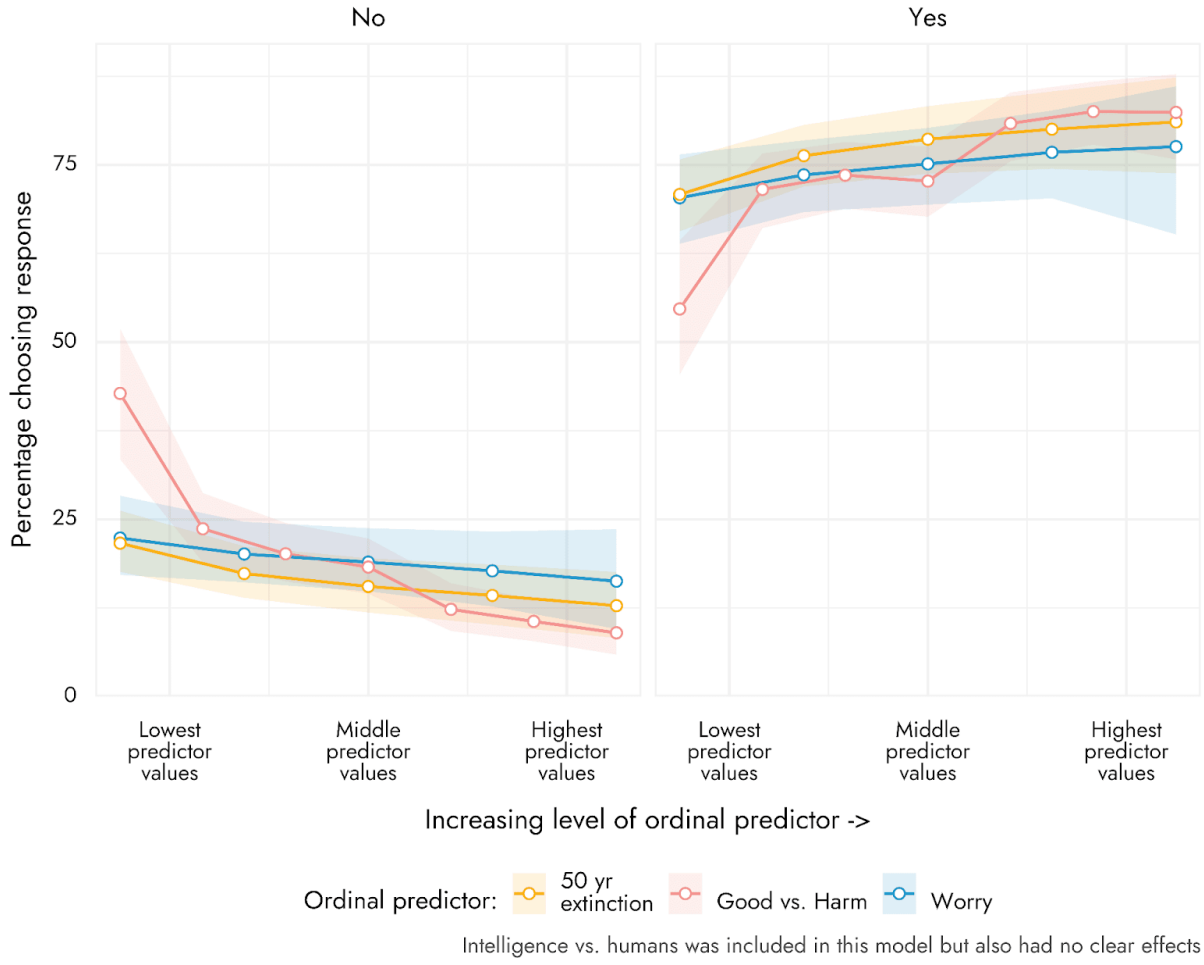
With respect to worry itself, we found that the anticipation of greater harm than good was strongly associated with worry - more so than both the expectation of extinction and the

belief that AI would achieve superhuman intelligence. The specific belief that AI might cause human extinction in the next 50 years was in turn more strongly associated with worry than the belief that it would achieve superhuman intelligence. It seems plausible that general expectations of harm, such as job loss or broader societal impact may be more concrete and imaginable than the potentially abstract conception of human extinction, even among those who really believe this might happen. Hence, general conceptions of harm may be more likely to provoke worry in one's daily life than the anticipation of extinction. People may also feel they have more agency with respect to more 'mundane' negative effects such as job loss, resulting in more rumination as people worry about how they might adapt. An alternative explanation might also be that general Good vs. Harm simply functions as a catch-all measure for the general goodness/badness of AI, and

therefore captures a much wider range of concerns than extinction expectations.

Agreement with regulation depending on *expectation of Good vs. Harm*, *worry*, and *expectation of extinction*

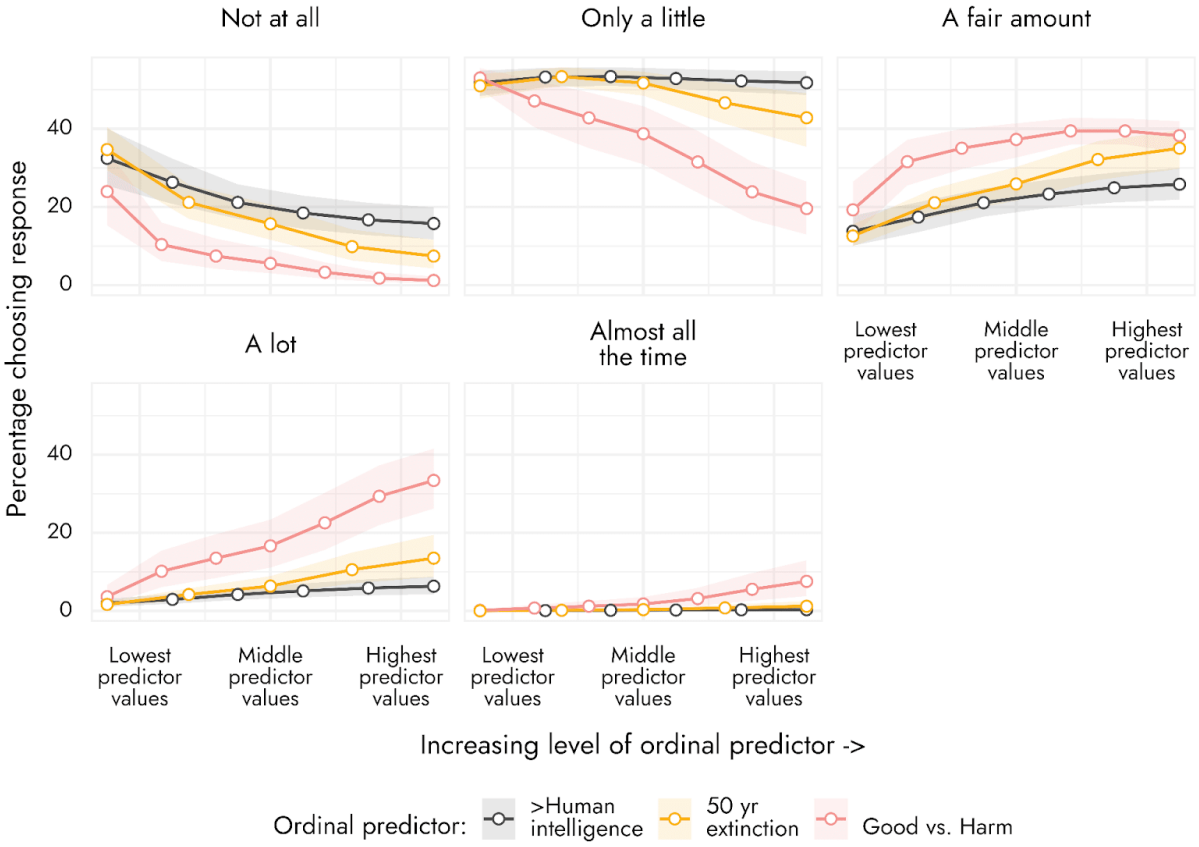
Higher expectations of good from AI were associated with less agreement with regulation. Worry and expectation of extinction did not show a reliable association.



Intelligence vs. humans was included in this model but also had no clear effects

Worry about AI depending on expectations of *extinction*, *good vs. harm*, and of *superhuman intelligence*

Specific expectation of extinction was more strongly associated with worry than was anticipation of superhuman intelligence. Worry increases substantially along with the expectation of more harm.



Conclusions

The estimates from this poll may inform policy making and advocacy efforts regarding AI risk mitigation. The findings broadly suggest an attitude of caution from the public, with substantially higher levels of support than opposition to measures that are intended to curb the evolution of certain types of AI (a possible pause of some kinds of AI development), as well as for regulation of AI.

However, concerns over AI do not yet appear to feature especially prominently in public perception of the existential risk landscape. Notably, extinction caused by AI was selected as the most substantial existential threat to humanity by only a small minority of people. In addition, it does not seem that risks from AI are something that most people are worrying about a lot in their daily lives (though note that we do not have a comparison for how much people report worrying about other issues). We are conducting additional qualitative research to better understand people's worries about, and their perceptions of risk from AI, which may further inform our understanding of AI risk perception.

US adults appear to appreciate that AI may well become more intelligent than people, and place non-negligible risk on the possibility that AI could cause extinction within the next 50 years. Nevertheless, people generally expect there to be more good than harm to come from AI.

Extrapolating from these findings, we might expect the US public to be broadly receptive to efforts aimed towards mitigating what are perceived as plausible and potentially highly concerning risks of AI, for example through well-designed government regulation, or efforts to prevent risky arms-race type behavior from companies competing to develop AI. However, there may be little mass appeal for what might be considered more extreme stances relative to where public perception and concern currently rests. This may be particularly the case given that people anticipate substantial good to come from AI, not just bad. Of course, this does not mean that such communication could not *shift* public opinion - we are describing where the US population appears to be at, and not suggesting where public opinion optimally should be, with respect to AI risk perceptions.

As AI risk represents a relatively new area of public discourse, we anticipate that current events and media discussion could still substantially shift public perception.

Appendix

Methodology

On April 14th 2023, Rethink Priorities conducted an online poll regarding public perceptions of AI risk, as well as attitudes towards regulation of AI development and support/opposition of the recently proposed pause on training/development of certain types of AI models.

The poll sampled 2523 US respondents aged 18 or above on the online sampling pool *Prolific,* of whom 2444 consented, answered questions, and passed requisite attention checks for the analyses presented below. We then used Multilevel Regression and Poststratification (MRP) to generate population-level estimates for US adult public opinion, accounting for Age, Sex, Race, Household Income, Education, Political party affiliation, as well as the US State/District and 2020 Republican vote share for the state.

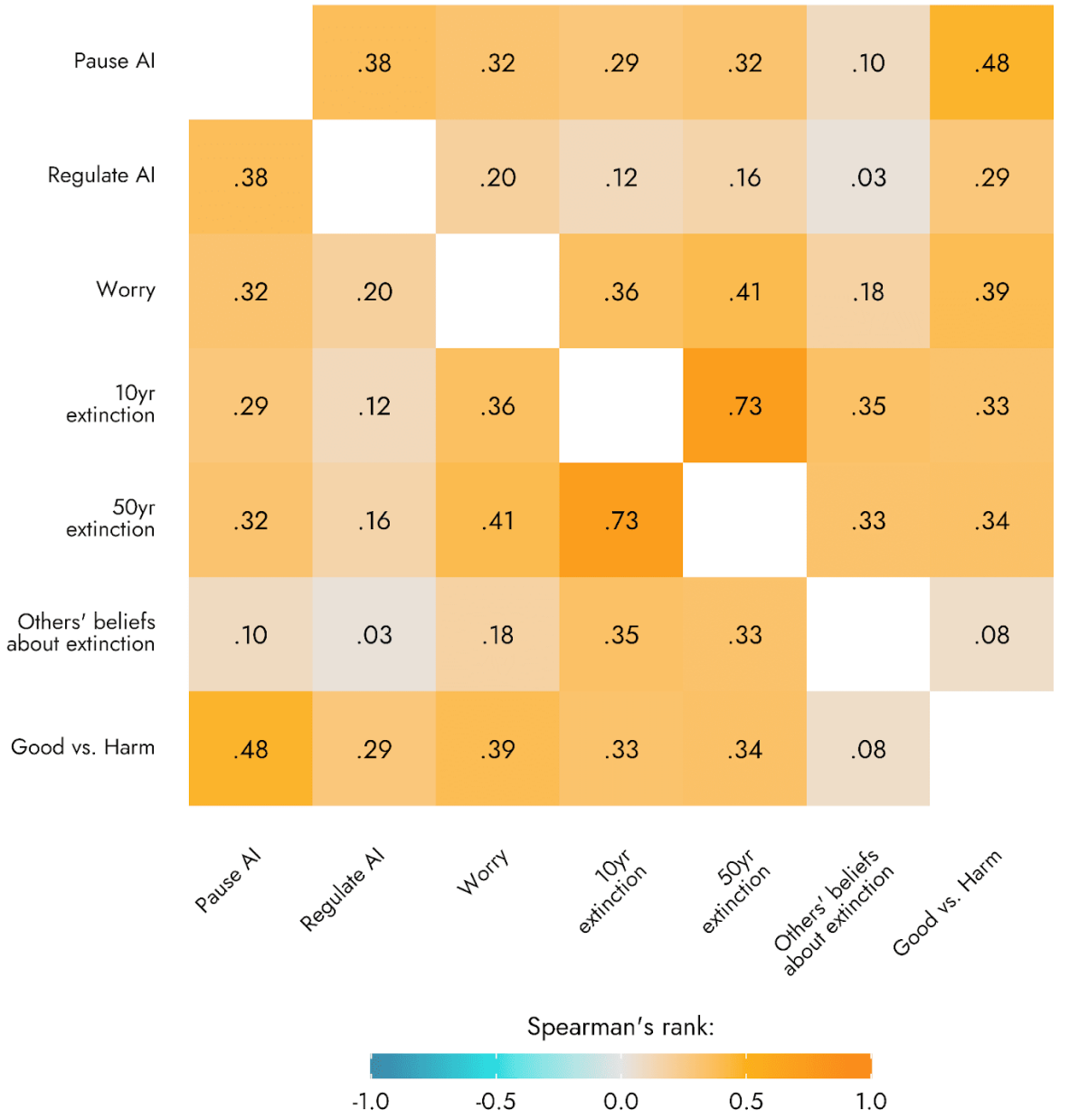
MRP is a technique that can be used to estimate outcomes in a specific target population based upon a potentially unrepresentative sample population. In brief, the technique involves generating estimates of how a range of features (e.g., education, income, age) are associated with the outcome of interest from the sampled population, using multilevel regression. Based on the known distribution of combinations of these features in the target population, the poststratification step then involves making predictions from the multilevel regression model for the target population. This approach is widely used to make accurate predictions of population level opinion and voting based upon unrepresentative samples (e.g., Wang, W., Rothschild, D., Goel, S., & Gelman, A. (2015). Forecasting elections with non-representative polls. *International Journal of Forecasting*, *31*(3), 980-991.), and also allows inferences to be made about specific subgroups within the population of interest.

Associations among AI measures

To assess possible associations among the different AI-related measures, we conducted Bayesian multiple regression with the respective AI-related predictor variables entered as ordinal predictors (i.e., monotonic effects). For each of these models, we additionally included Age, Race, Sex, Region, Education, Income, and Political Party Affiliation as control variables. When making predictions from the model, we varied the value of the ordinal predictor of interest while holding each of the other AI variables constant at their median in the sample data, and then averaged the predictions across all the demographic variables.

Simple pairwise associations (Spearman's rank) between all the AI-related variables presented above can be seen in the correlation matrix here:

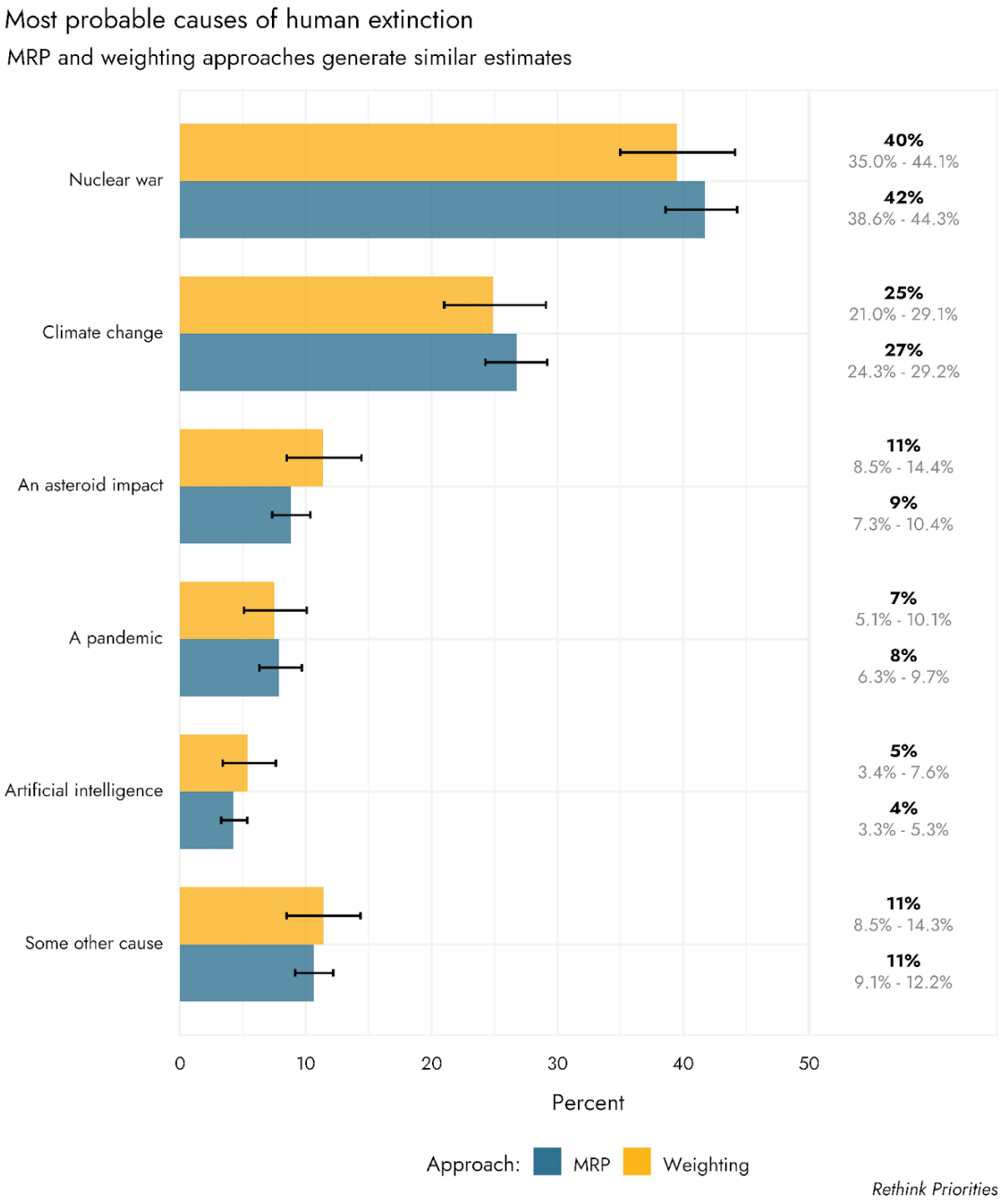
Pairwise associations between AI-related outcomes

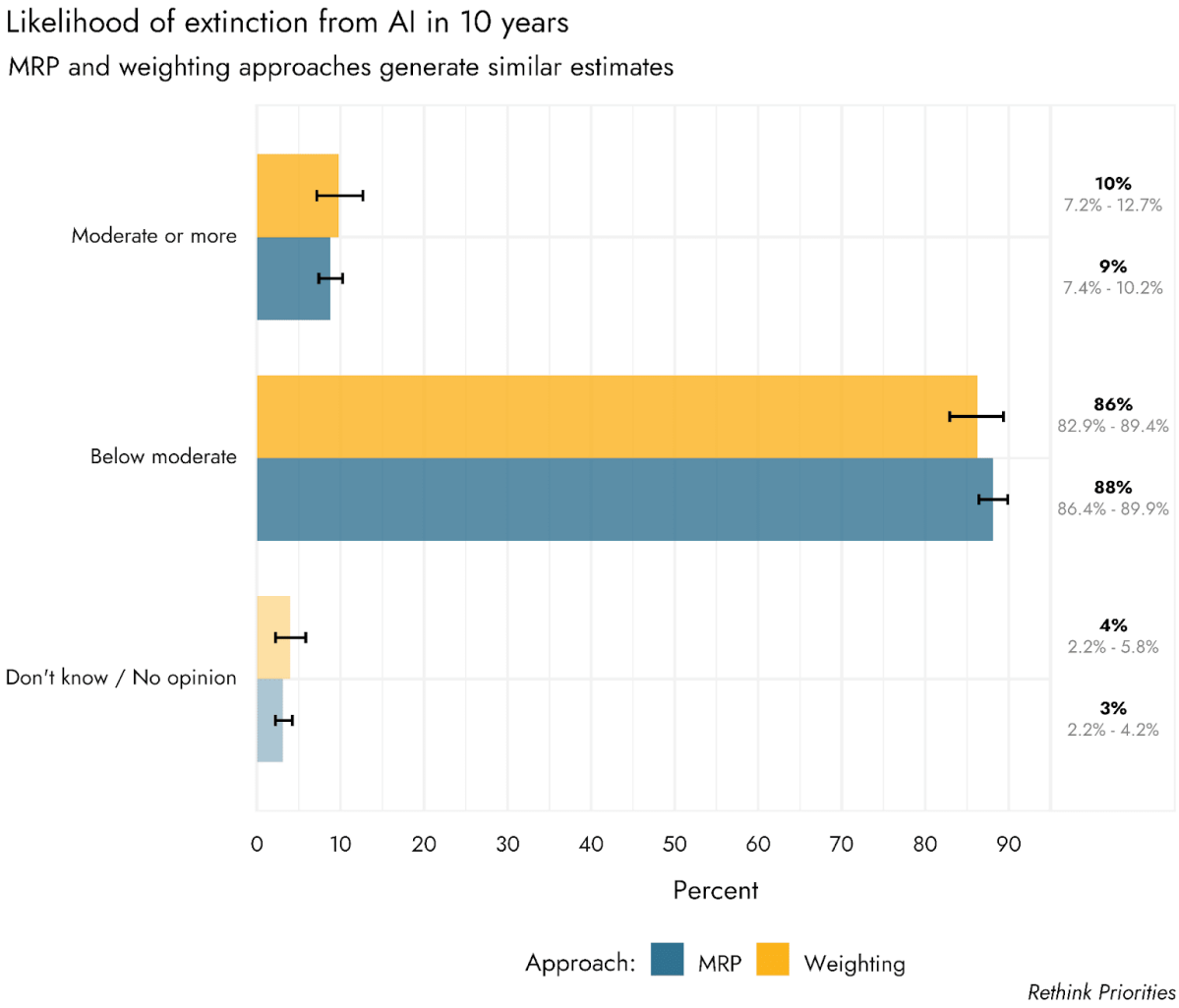


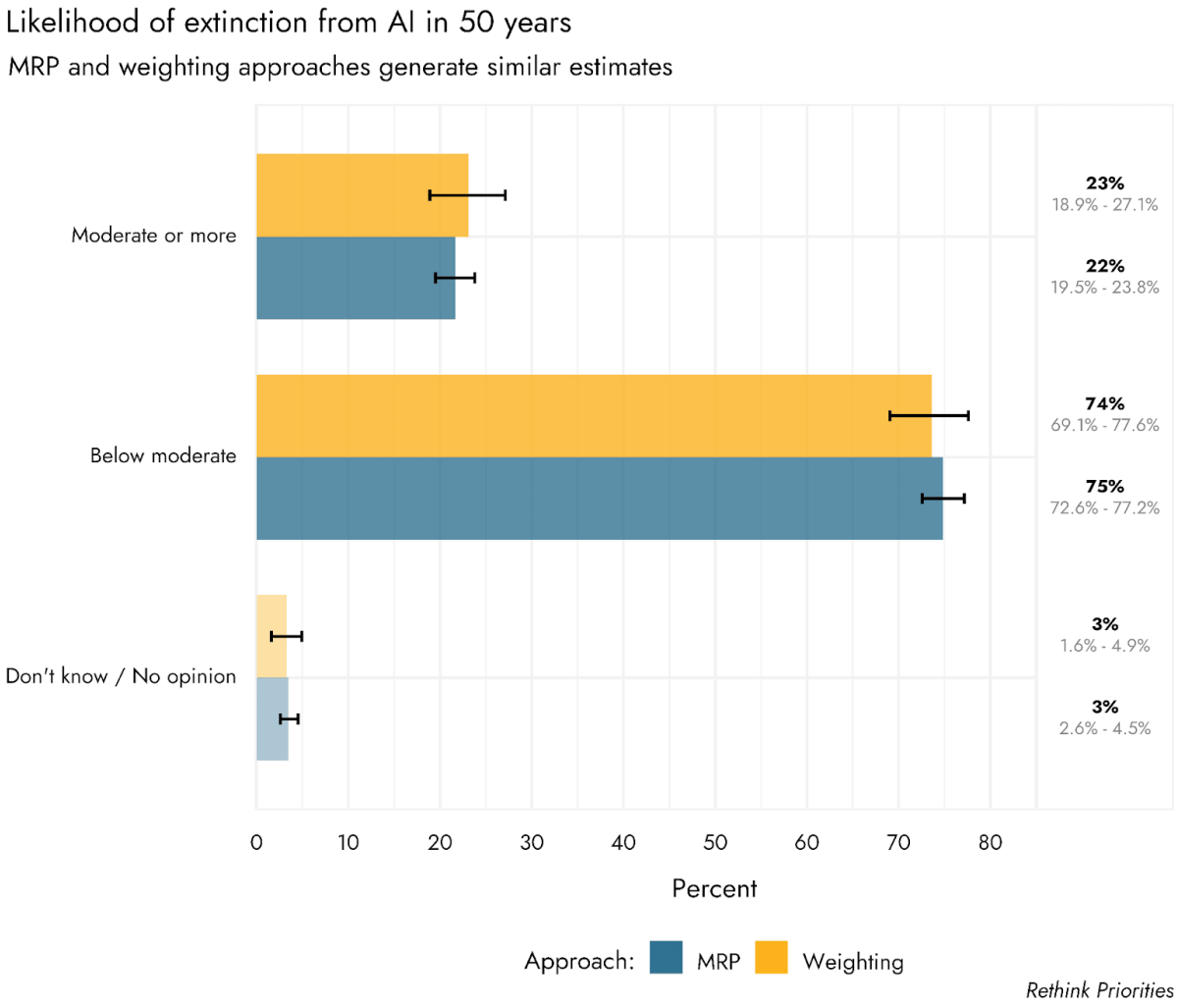
Only the *Others' beliefs* - *Regulate AI* association did not reach Bonferroni-corrected significance (p <.0024)

Sensitivity to poststratification/weighting approach

Given that our sample was generated from online respondents, there are some concerns that despite efforts to make the sample representative of the population, we cannot represent certain kinds of people who are simply not online. In data from 2021 from Pew Research, it was estimated that around 7% of US adults would report never using the internet. To try to correct for the possibility that an overly online sample might affect our results, we did include an assessment of internet frequency, and were then able to weight the sample according to answers to that question. Including this outcome variable is not possible in the MRP approach, but can be included in weights. As shown in the plots below, there is little if any difference between our MRP estimates and those generated by using an alternative weighting protocol that includes internet use frequency.

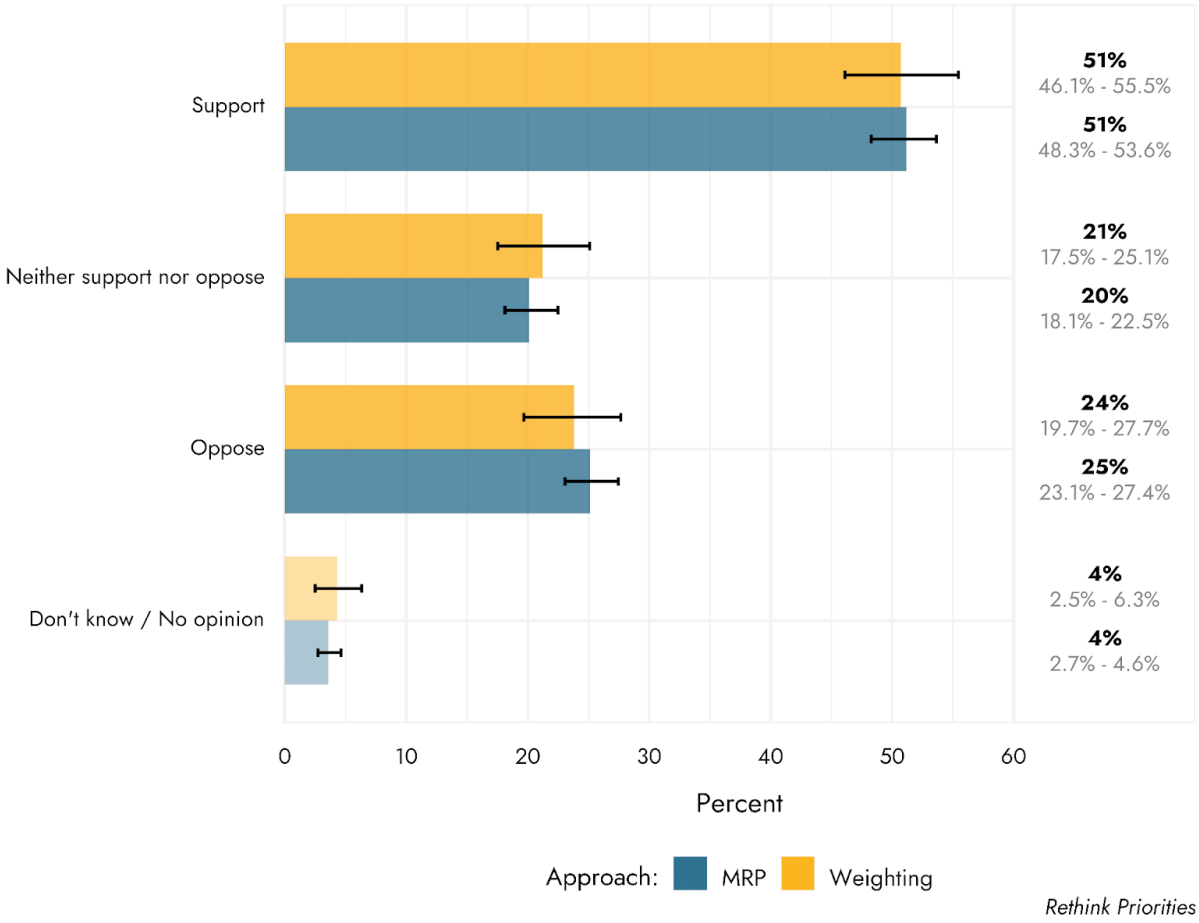

Most probable causes of human extinction
MRP and weighting approaches generate similar estimates
Nuclear war
Climate change
An asteroid impact
A pandemic
Artificial intelligence
Some other cause
40%
35.0% - 44.1%
42%
38.6% - 44.3%
25%
21.0% - 29.1%
27%
24.3% - 29.2%
11%
8.5% - 14.4%
9%
7.3% - 10.4%
7%
5.1% - 10.1%
8%
6.3% - 9.7%
5%
3.4% - 7.6%
4%
3.3% - 5.3%
11%
8.5% - 14.3%
11%
9.1% - 12.2%
0
10
20
30
40
50
Percent
Approach:
MRP
Weighting
Rethink Priorities


Likelihood of extinction from AI in 10 years
MRP and weighting approaches generate similar estimates
Moderate or more
Below moderate
Don't know / No opinion
10%
7.2% - 12.7%
9%
7.4% - 10.2%
86%
82.9% - 89.4%
88%
86.4% - 89.9%
4%
2.2% - 5.8%
3%
2.2% - 4.2%
0
10
20
30
40
50
60
70
80
90
Percent
Approach:
MRP
Weighting
Rethink Priorities


Likelihood of extinction from AI in 50 years
MRP and weighting approaches generate similar estimates
Moderate or more
Below moderate
Don't know / No opinion
23%
18.9% - 27.1%
22%
19.5% - 23.8%
74%
69.1% - 77.6%
75%
72.6% - 77.2%
3%
1.6% - 4.9%
3%
2.6% - 4.5%
0
10
20
30
40
50
60
70
80
Percent
Approach:
MRP
Weighting
Rethink Priorities

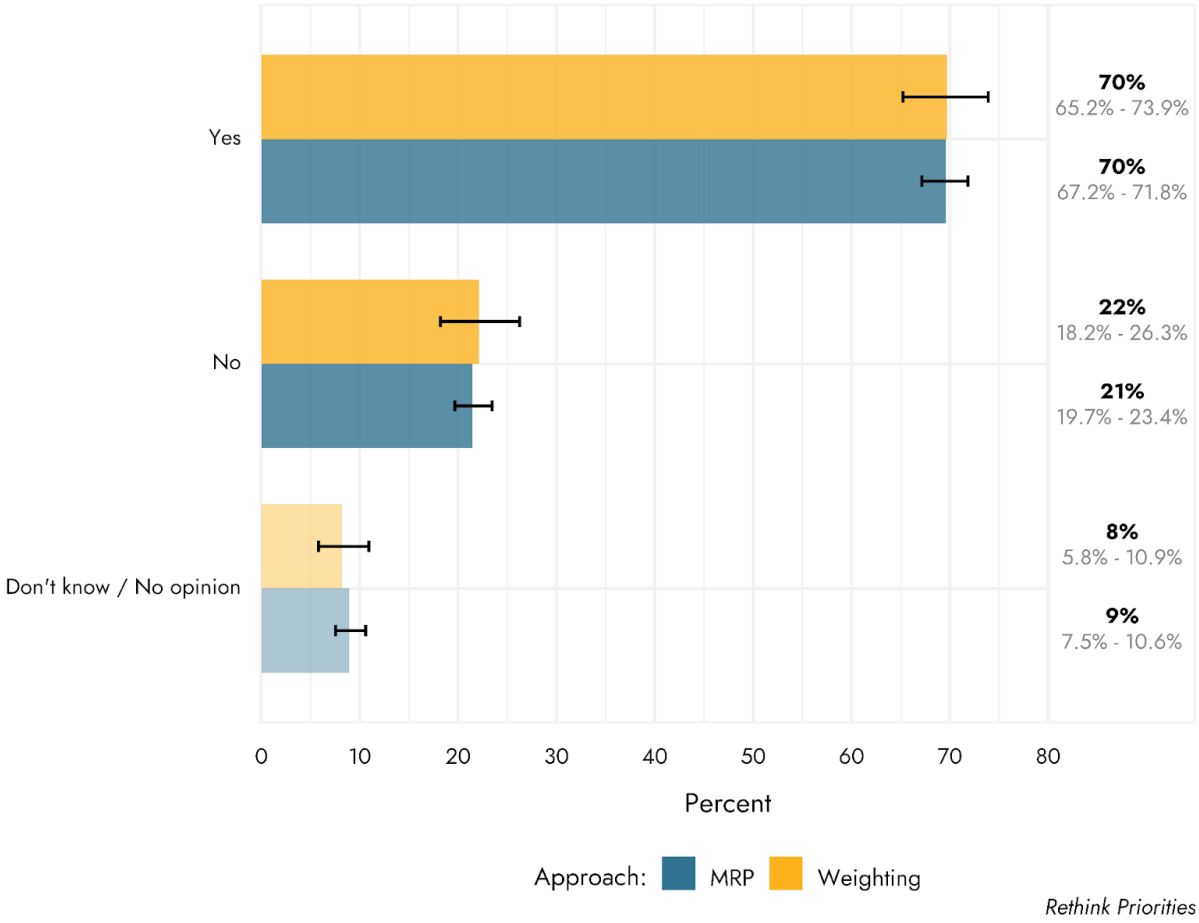
Pause on AI research

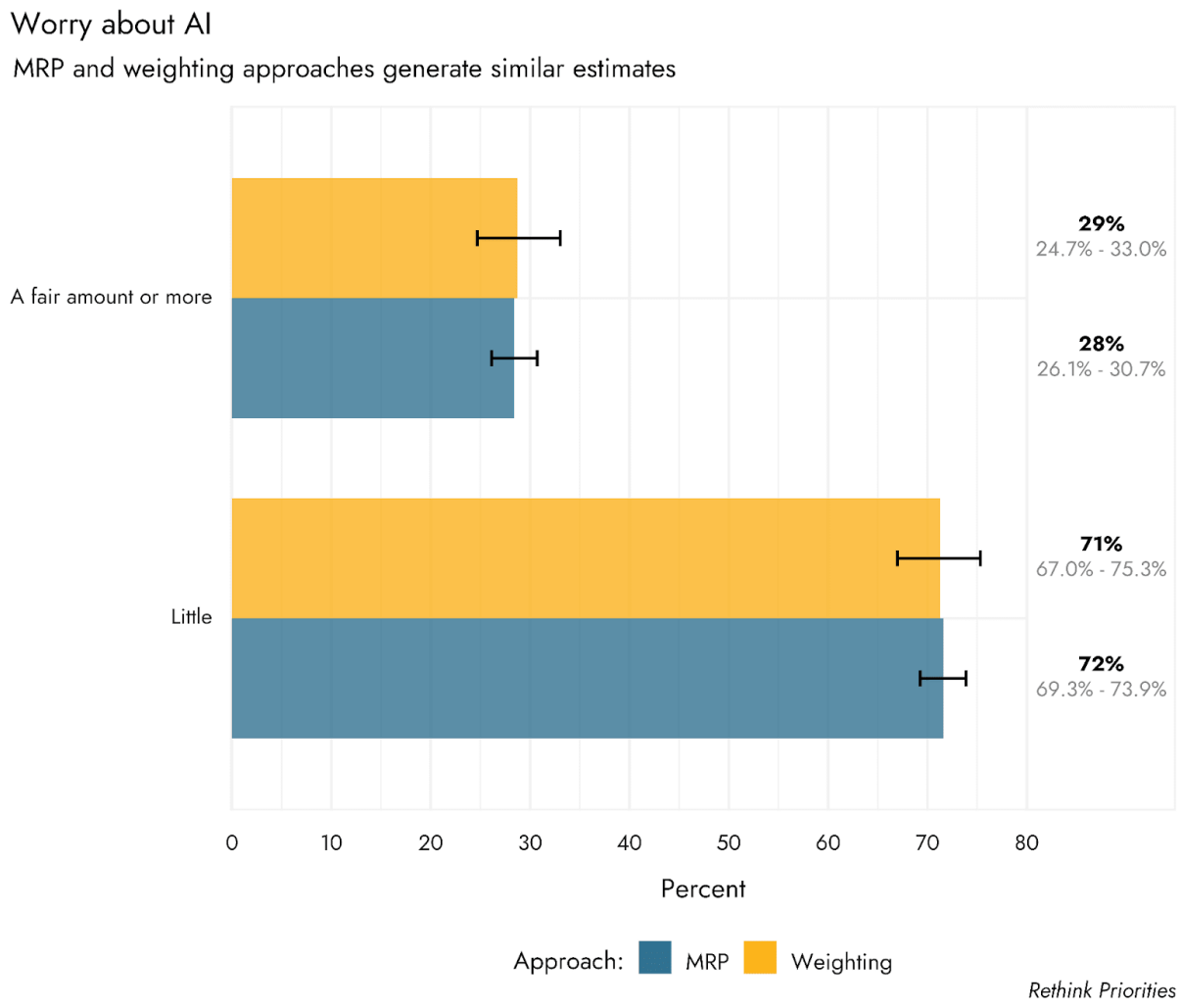
MRP and weighting approaches generate similar estimates



Regulation of AI

MRP and weighting approaches generate similar estimates

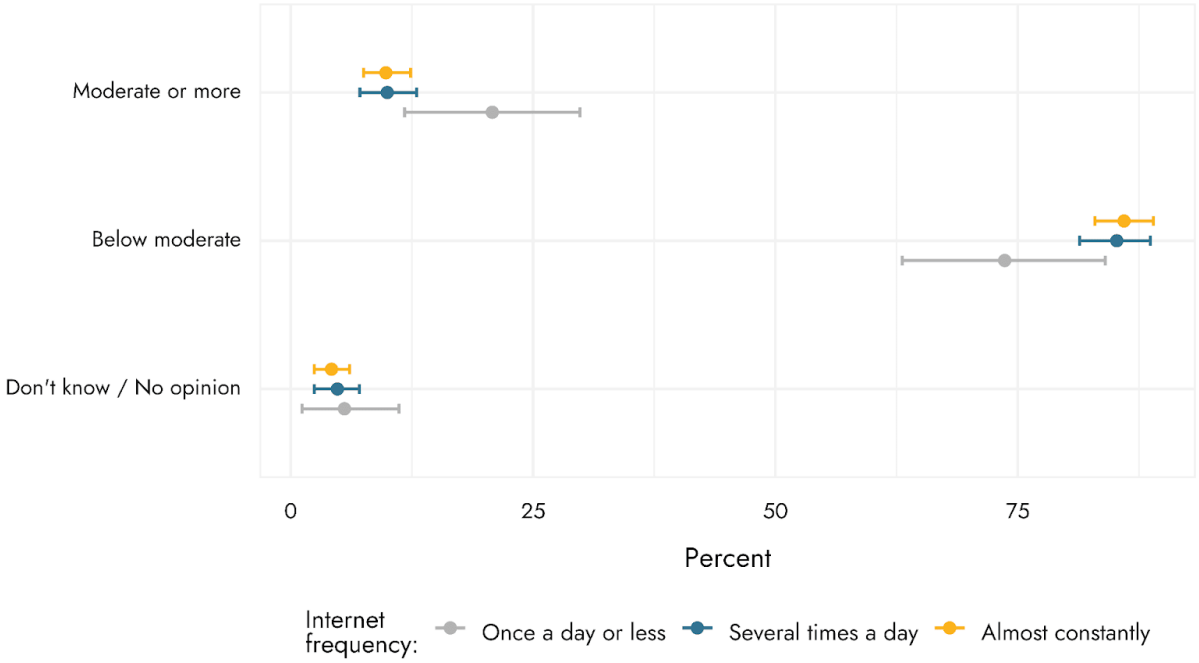




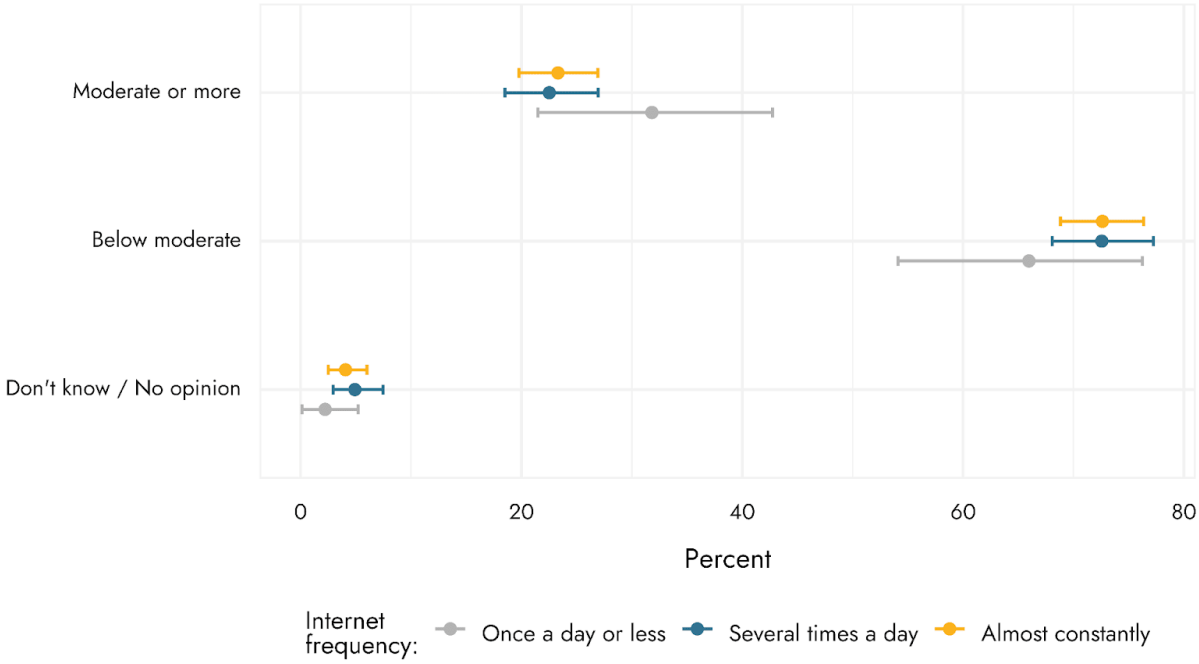
We additionally conducted some multiple regression analyses in which we included Age, Gender, Race, Education, Income, US Region, Political Party ID, and Internet Use Frequency as predictors of each of these main outcome variables. Averages over the other demographic variables and then assessing the specific differences among internet frequency answers, we see that for some variables, there appears to be a slight (or sometimes more substantial) effect of internet frequency on responses. However, the key consideration is that such respondents make up only a minority of the US population, and so even when such respondents' answers are weighted for inclusion, their responses would have to be very highly and consistently different in order to shift the overall population estimate.

Finally, where there do appear to be differences between those with higher and lower levels of internet use, infrequent internet users tend to be *more* rather than *less* concerned about AI. Hence, we do not think that our responses risk over-emphasising the concern people have over AI. However, we of course do not have information on people who actually never use the internet, and so if these people's responses would be really dramatically different from the lower frequency users we do have data on, then this could be a threat to the validity of the conclusions. Again, however, as these people represent just 7% of the population as of 2021, their responses would have to be different in the extreme to shift population estimates.

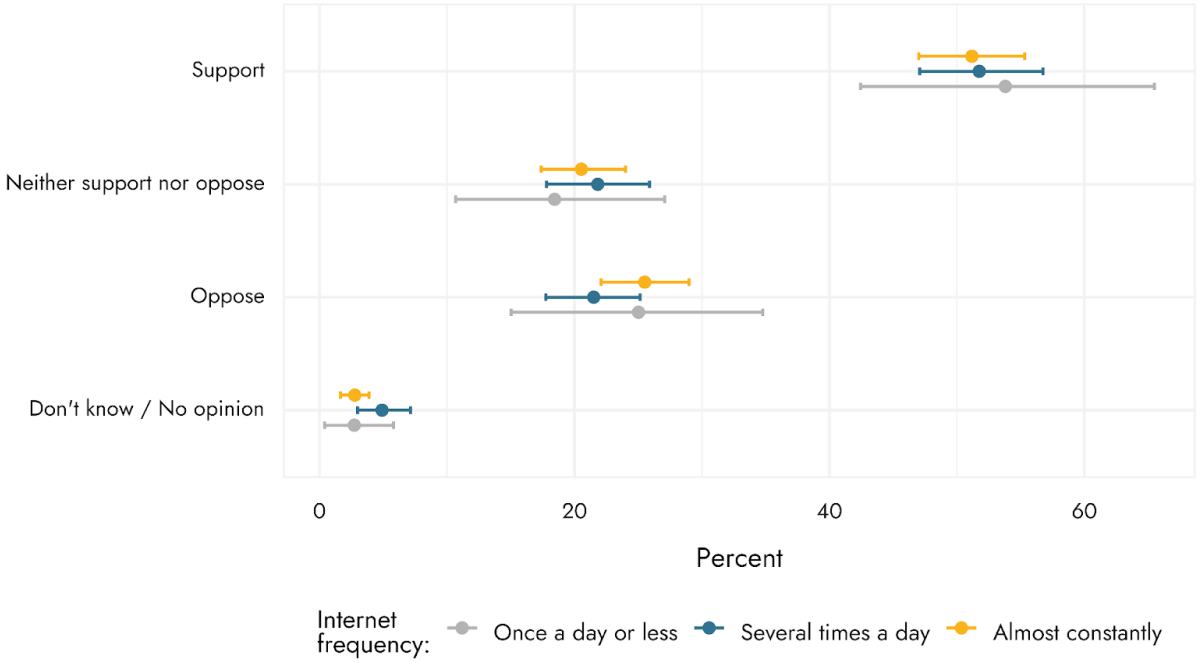
Extinction from AI within 10 years



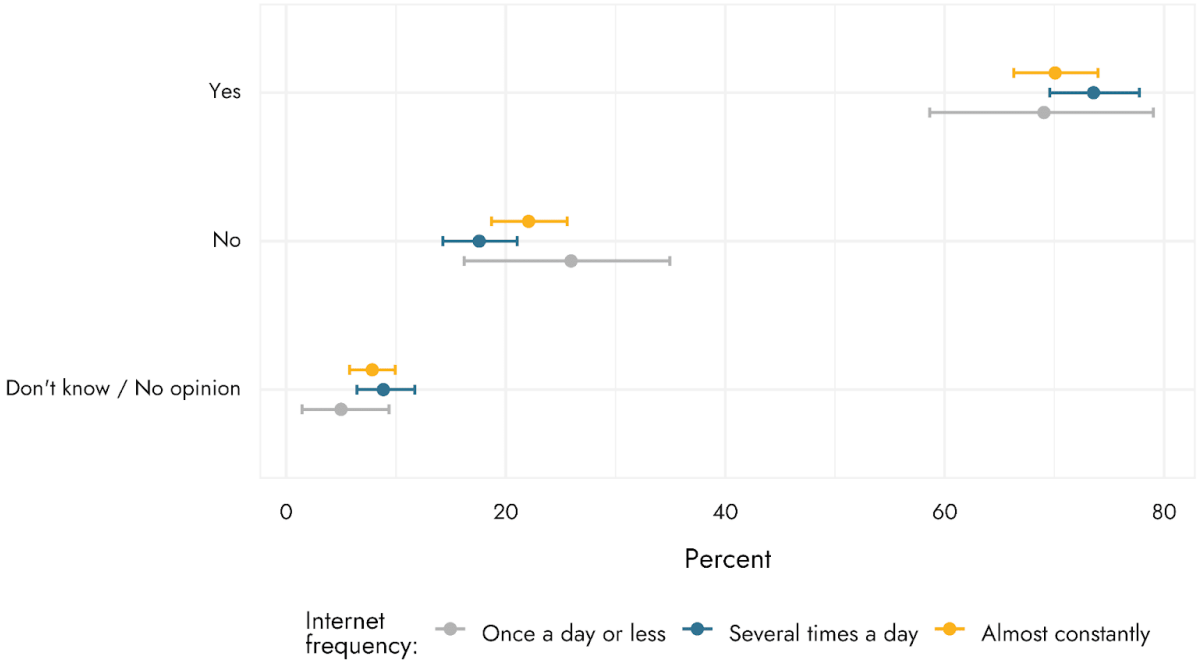
Extinction from AI within 50 years

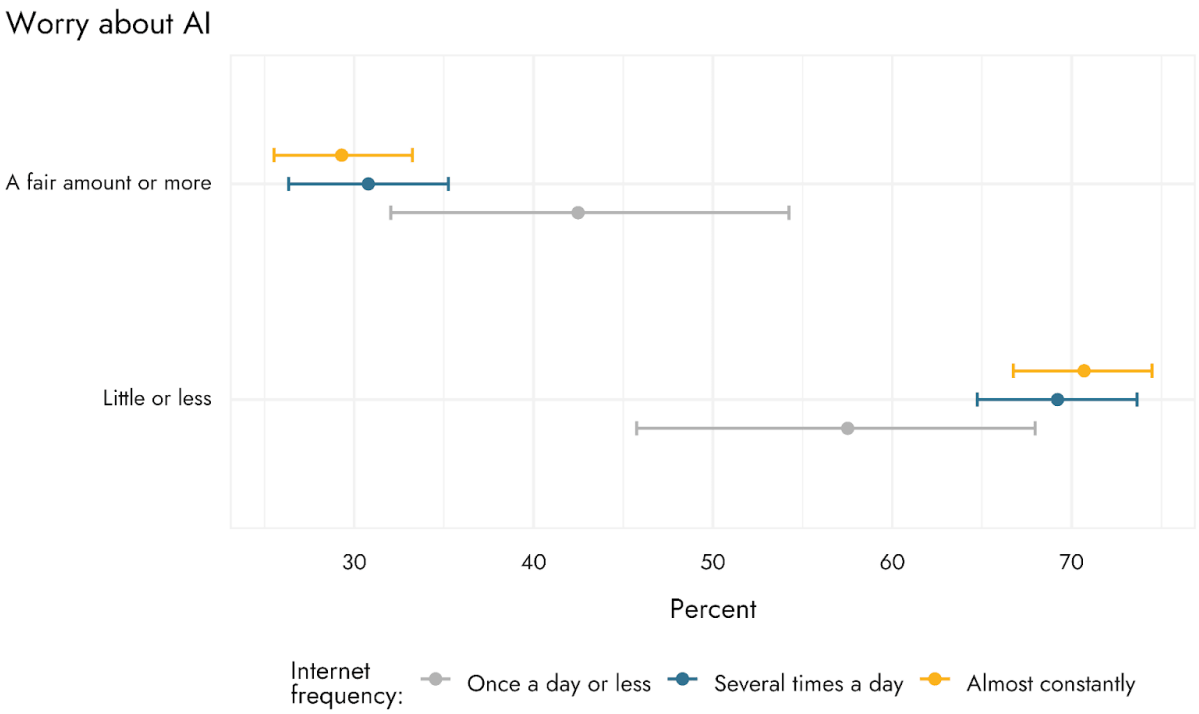


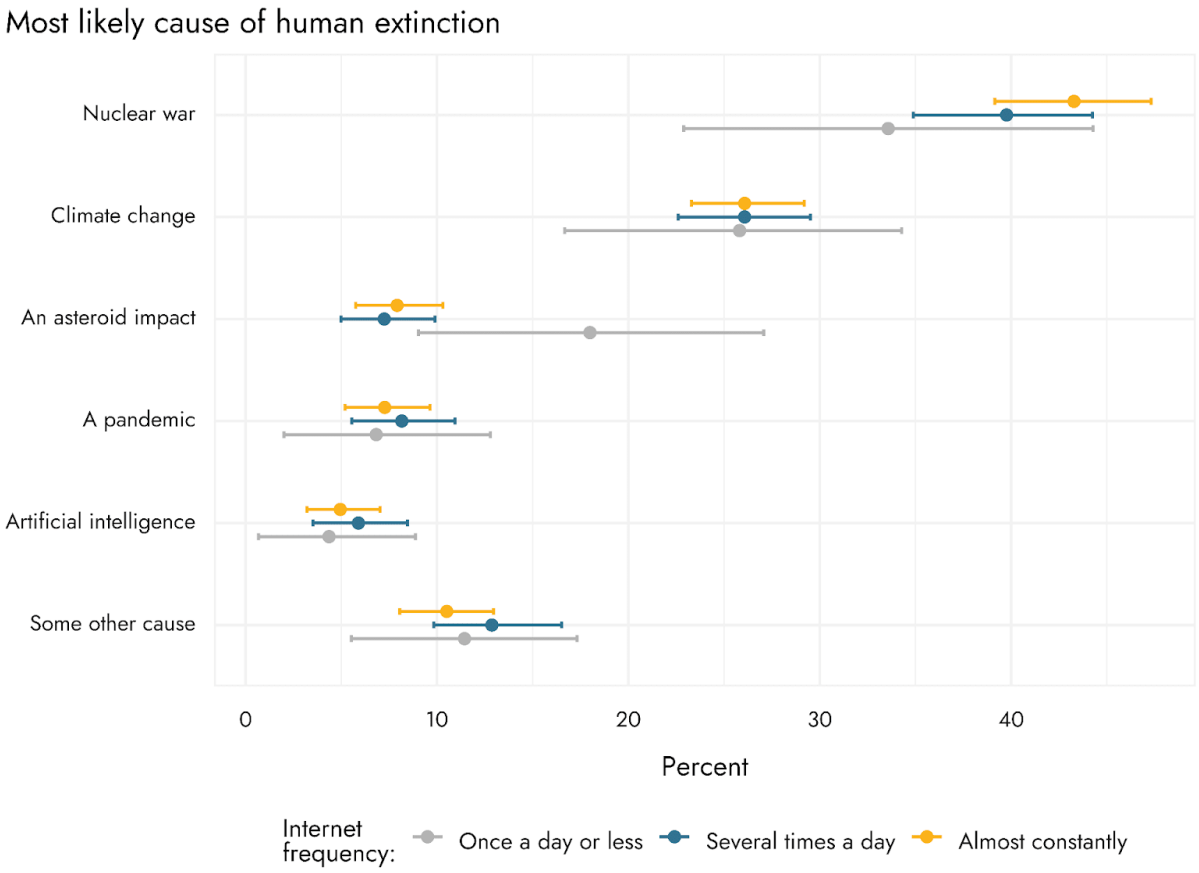
AI Pause



Regulate AI




Worry about AI
A fair amount or more
Little or less
30
40
50
60
70
Percent
Internet frequency:
Once a day or less
Several times a day
Almost constantly


Most likely cause of human extinction
Nuclear war
Climate change
An asteroid impact
A pandemic
Artificial intelligence
Some other cause
0
10
20
30
40
Percent
Internet frequency:
Once a day or less
Several times a day
Almost constantly

Acknowledgments

Jamie Elsey and David Moss wrote this report. Jamie Elsey, David Moss, and Will McAuliffe developed the survey. Will McAuliffe was also involved in discussion and interpretation of the results. Jamie Elsey conducted the analyses and data visualization. We would also like to thank Peter Wildeford and Renan Araujo for review of and suggestions to the final draft of this report.

If you are interested in RP's work, please visit our research database and subscribe to our newsletter.

Transparency Disclosures

This section contains information regarding details of the survey design, implementation, and analysis, as proposed by the American Association of Public Opinion Research (AAPOR).

1. **Data Collection Strategy:** This research is based on answers to a cross-sectional online survey.
2. **Who Sponsored the Research and Who Conducted It**. This research was funded and conducted by *Rethink Priorities* (RP), a registered 501(c)3 non-profit organization and think tank, registered in California, US. RP financial disclosures can be found here: https://rethinkpriorities.org/transparency

3. **Measurement Tools/Instruments.** The survey questions and response options are presented in the graphs depicting population-level estimates in the main document above.
4. **Population Under Study.** This research aimed to assess US Adult public opinion.
5. **Method Used to Generate and Recruit the Sample.** This research used a non-probability sampling approach, involving respondents opting-in to take part in the study via the *Prolific* participant pool platform. Respondents are not informed about the specific content of the survey until entering the survey, seeing only a description of *'A survey about current attitudes'*, thereby reducing bias from respondents who enter being specifically interested in artificial intelligence. Respondents were required to be registered on Prolific as adults (aged 18 and above), resident in the US. No quotas were used for sampling purposes. Respondents were informed that the survey was expected to take approximately 5 minutes, and they would be compensated £0.75 (~$0.93) for their completion of the survey.
6. **Method(s) and Mode(s) of Data Collection.** Respondents were recruited via the online platform *Prolific,* and completed the survey using the *Qualtrics* survey software platform. The survey was offered in English.
7. **Dates of Data Collection.** Data collection for this survey took place on April 14th, 2023.
8. **Sample Sizes and Precision of the Results.** The survey received 2523 respondents, of whom 2444 completed it, met inclusion criteria, and passed attention checks. The primary results are based upon these 2444 respondents. Margins of error in primary analyses represent the means and 95% highest density intervals (HDIs) of posterior distributions derived from Bayesian Multilevel Regression and Poststratification (MRP). Please see the plots and main text of the report for the uncertainty associated with each result, as there is no single +/- margin of error that applies to all estimates. The central estimates are presented as rounded percentages, while the margin of error is presented to 1 decimal place (this ensures that the width of the margin of error is not understated, as rounding the upper and lower bound of the error margin could artificially reduce its width). MRP is a technique that can be used to estimate outcomes in a specific target population based upon a potentially unrepresentative sample population. The technique involves generating estimates of

how a range of features (e.g., education, income, age) are associated with the outcome of interest from the sampled population, using multilevel regression. Based on the known distribution of combinations of these features in the target population, the poststratification step then involves making predictions from the multilevel regression model for the target population. This approach is widely used to make accurate predictions of population level opinion and voting based upon unrepresentative samples, and also allows inferences to be made about specific subgroups within the population of interest (e.g., Wang, W., Rothschild, D., Goel, S., & Gelman, A. (2015). Forecasting elections with non-representative polls. *International Journal of Forecasting*, *31*(3), 980-991. AND Park, D. K., Gelman, A., & Bafumi, J. (2004). Bayesian multilevel estimation with poststratification: State-level estimates from national polls. *Political Analysis*, *12*(4), 375-385.). Model specification for multilevel regression used respondent State/District, Region, Age bracket, Completed education, Household income bracket, Sex, Racial identification, Sex * Racial identification, Completed education * Age bracket, Political party identification, and the State/Disctrict's Republican vote share for the 2020 Presidential election. The following section describes the poststratification.

9. **How the Data Were Weighted**. Regression model predictions were poststratified according to the cross-tabulated proportions of the US population with the respective demographic features outlined in the previous section based upon public release of the *Census Bureau's 2020 American Community Survey* for US adults. This poststratification table was extended to include a posterior distribution of expected political party identification based upon multilevel regression using data from *Harvard University's 2020 Cooperative Election Study*.
10. **How the Data Were Processed and Procedures to Ensure Data Quality.** In addition to *Prolific's* in-house checks for participant quality and integrity, we included 2 attention checks - one at the beginning of the survey, and one towards the end - interspersed among typical demographic questions - to ensure participants were reading the questions and answering correctly. Respondents were required to pass both these checks in order to be included in analyses.
11. **Limitations of the Design and Data Collection**. Survey data and its analysis and interpretation can be prone to numerous issues. One particular concern for

non-probability samples (i.e., opt-in, online surveys) is the potential for biases in recruitment that are not or cannot be counteracted by weighting or poststratification approaches. In the appendix of our report, we detail an alternative weighting procedure geared towards incorporating possible biases from a largely online-savvy sample, which showed analogous results to our main analyses. However, it is not possible to correct for all possible sampling biases.