Visions, values, voices: A Survey of Artificial Intelligence Researchers

UCL
Centre for Responsible Innovation

Cian O'Donovan, Sarp Gurakan,
Xiaomeng Wu, Jack Stilgoe.
With Nicolas Bert, Ekaterina Dmitrichenko,
Embla Gjørva, Stella Liu,
Teddy Zamborsky IV, Tianyu Zhao.

March 2025

Visions, values, voices: A Survey of Artificial Intelligence Researchers

Executive Summary

As excitement and investment in artificial intelligence grow, a number of surveys have sought to understand public views. There have been very few attempts to understand the attitudes of AI researchers. Given the uncertainties around the opportunities and threats of AI technologies, the views of those closest to the technology are crucial.

In summer 2024, a research team from University College London's Department of Science and Technology Studies fielded a survey of AI researchers designed to understand their values, their visions for the future of AI, and what they though about the role of public voices in AI. Our survey included questions that had been asked in representative UK public surveys, to map overlaps and gaps between public and AI researchers' views. We analysed the responses from 4,260 AI researchers, making it the largest survey of AI researchers to date.

Our insights include the following:

- Researchers do not speak with one voice: they report diverse and divergent views about innovation and responsibilities in AI
- Researchers are more positive than members of the public about the benefits of AI
- Researchers and the public share concerns about disinformation, data use and cybercrime
- There is a sense of technological inevitability in AI research
- 'Optimist' and 'pessimist' researchers report different views on AI
- Researchers tend to have a 'deficit model' of the public
- Researchers want the public involved downstream, not upstream
- Researchers want AI to reflect human values but do not pay attention to social science research
- Researchers think it is more important for society to debate risks than benefits
- Researchers and the public disagree about who should be responsible for the safe use of AI
- Researchers want greater care for training data
- Researchers are less concerned than the public are when it comes to explaining AI outputs
- Researchers are concerned about who sets research agendas for AI

This project is part of the *Public Voices in AI* project under the UKRI Responsible AI programme.

UCL
Centre for Responsible Innovation

Table of contents

1. Introduction

As excitement and investment in artificial intelligence grow, a number of surveys have sought to understand public views about these emerging technologies. Yet with only a few exceptions, the views the attitudes of AI researchers have not been systematically surveyed at scale. Given the uncertainties surrounding the opportunities and threats of AI technologies, it is important to understand the views of those closest to the technology.

In summer 2024, a research team from UCL's Department of Science and Technology Studies ran a survey of AI researchers. It was designed to understand their hopes and fears about AI, their values and their impressions of public concerns. The work was part of the Public Voices in AI project, funded by the Responsible AI UK programme.[1] Our survey included some questions that had been asked in representative UK public surveys, to map overlaps and gaps between public views and AI researchers' views.

We wanted to know:

1. What do AI researchers think of AI risks, benefits, responsibilities and other issues?
2. Where are the gaps and overlaps between public and AI researchers' attitudes?
3. How do AI researchers imagine public hopes and fears?
4. What do AI researchers think about involving the public in AI?

Why focus on AI researchers?

Scientists, innovators and policymakers often worry about public trust in new technologies. In the past, the assumption has often been that an apparent lack of trust is due to public ignorance about science. We now have substantial evidence to challenge this 'deficit model'.[2] Members of the public often have legitimate questions about the benefits and risks of technology, and technologies are surrounded by uncertainties. For these reasons, there is growing interest in the idea of public dialogue - a genuine conversation between the public and the people making decisions about new technologies. Understanding the views of people on both sides of this conversation is vital, but while we have seen plenty of attempts to understand what the public think, we have seen relatively little curiosity about the other side.[3]

Over the last two years, AI has entered the public consciousness rapidly and in a way that feels confusing. Interfaces like ChatGPT have quickly become tools or playthings. And prototype technologies like self-driving cars have become test cases for AI on some cities' streets. These technologies depend on massive amounts of data, so it is

---

[1] See https://digitalgood.net/dg-research/public-voices-in-ai/ and www.rai.ac.uk

[2] Wynne, B. (2006). Public engagement as a means of restoring public trust in science–hitting the notes, but missing the music?. Public Health Genomics, 9(3), 211-220.

[3] Many public surveys have been framed narrowly. The Public Voices in AI project is seeking to improve this area of research.

unsurprising that the companies developing them are, or are supported by, the biggest technology firms. The people in charge of AI companies have sought to persuade us that the technology is revolutionary and will be hugely beneficial. But they have also expressed concerns that its risks could be catastrophic. It is hard, if not impossible, for even those closest to the technology to fully understand how AI systems do what they do. And because AI is already in the pipework of so many digital technologies, we may not know when we are interacting with AI systems.

With any new technology, there is a range of issues about which people - AI researchers, members of the public or anyone else - might be concerned. Experts in a particular technology may focus their excitement or concern on different issues from the public. They might also claim that their priorities, by virtue of their expertise, are the 'correct' ones.[4] This can have the effect of closing down democratic debate. Our study tried to avoid presumptions about which risks and benefits were 'correct'.

We know from other technologies that the people doing the R&D tend to be more aware of uncertainties than those at one stage removed who have bought into the hype.[5] With AI, we have heard a lot about both the world-changing benefits and the existential risks of technology from company leaders, self-appointed gurus and enthusiastic politicians. We don't tend to hear from the people who are trying to improve the technologies and understand their inner workings.

There have to date been few serious attempts to survey AI researchers' views at scale. Many previous surveys of AI insiders have asked for predictions. The approach has been to ask AI scientists to estimate when AI will achieve some level of superhuman performance or the probability of catastrophic risks from AI. Some of these have been widely cited and reported, despite having poor survey design and small numbers of responses.[6]

A notable exception has been a survey from Luye Bao, Dominique Brossard and colleagues.[7] Their survey received 2,352 responses from scientists working on AI. The 15-minute online survey asked these scientists for opinions on risks, benefits, regulations and who should be included in making decisions about AI. The research

---

[4] This characterisation of concerns in terms of science, and the effects this has on democratic debate, has been labelled 'scientism' (see Welsh, I., & Wynne, B. (2013). Science, scientism and imaginaries of publics in the UK: Passive objects, incipient threats. Science as Culture, 22(4), 540-566.)

[5] See Collins, H. M. (1997). Expertise: Between the Scylla of certainty and the new age Charybdis. Accountability in Research, 5(1-3), 127-135. And MacKenzie, D (1998). "The Certainty Trough," Palgrave Macmillan Books, in: Robin Williams & Wendy Faulkner & James Fleck (ed.), Exploring Expertise, chapter 15, pages 325-329, Palgrave Macmillan.

[6] See, for example, Müller, V. C., & Bostrom, N. (2014). Future progress in artificial intelligence: A poll among experts. AI Matters, 1(1), 9-11. https://nickbostrom.com/papers/survey.pdf; Grace, K., Salvatier, J., Dafoe, A., Zhang, B., & Evans, O. (2018). When will AI exceed human performance? Evidence from AI experts. Journal of Artificial Intelligence Research, 62, 729-754.; Grace, K., Stewart, H., Sandkühler, J. F., Thomas, S., Weinstein-Raun, B., & Brauner, J. (2024). Thousands of AI authors on the future of AI. arXiv preprint arXiv:2401.02843. Clarke, S, Schuett, J, 2021, Survey on AI existential risk scenarios, AI Alignment Forum

[7] Bao, L., Calice, M. N., Brossard, D., Li, N., Newman, T. P., Scheufele, D. A., & Xenos, M. A. (2023). AI scientists' perspectives on AI. University of Wisconsin-Madison. Madison, WI: Department of Life Sciences Communication. Available from https://scimep.wisc.edu/projects/reports/

found a high degree of concern about AI risks and ethical dilemmas but a conviction that science and scientists should lead the debate. Less than 15% of their respondents agreed with the statement that "public opinion is more important than the scientists' opinions when making decisions about the ethical implications of scientific research."

Some research on small samples of AI scientists has sought to understand the values that they bring to their research. Maurice Jakesch and colleagues found from comparative surveys of the US public and AI researchers (n=175) that researchers have different ethical priorities when it comes to things like fairness, safety, accountability and performance of AI systems.[8] (We know that scientists often avoid explicit discussion of public values in their research). Abeba Birhane and colleagues analysed 100 of the most influential scientific papers from machine learning conferences and found that science and engineering values like efficiency and novelty are prioritised over discussions of what societies need or what the risks of AI might be.[9]

Our survey is, as far as we know, the largest social science survey of AI researchers to date. Our findings, based on survey responses from more than 4,000 AI researchers point to some opportunities and challenges for enriching a dialogue between scientists and society on this high-stakes issue. We have summarised our findings as a set of headline insights based on our interpretation of our data.

[8] Jakesch, M., Buçinca, Z., Amershi, S., & Olteanu, A. (2022, June). How different groups prioritize ethical values for responsible AI. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (pp. 310-323).

[9] Birhane, A., Kalluri, P., Card, D., Agnew, W., Dotan, R., & Bao, M. (2022, June). The values encoded in machine learning research. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (pp. 173-184).

2. Study design and analysis

The survey questions were designed to address the study's research aims and overarching research questions: AI researchers' views of the developments of the field, the gap between researchers and the public, how researchers conceptualise expertise in their field, and if, and how, they include public voices in their work.

We selected some questions based on existing surveys of public opinion, to enable comparisons to be made, notably by Ada Lovelace and Alan Turing institutes[10] and by the Office for National Statistics, both in 2023.[11] Both surveyed the UK public, so some of the questions are taken verbatim to compare these sentiments with the sentiments of AI researchers (see Box 1 for further information). Direct comparison is not always appropriate between these surveys which were conducted at different times, and amongst different samples. In this report we highlight differences in *response profiles* - significant differences in strength and prioritisation of respective answers to questions. These profiles reveal important areas for follow-up studies that might investigate misalignments between AI researchers and members of the public identified here, as well as internal heterogeneity and diversity within AI researcher sentiment.

Our survey benefitted from two rounds of piloting. The first was with UCL postgraduate students in Science and Technology Studies. The second was with members of UCL's Centre for Artificial Intelligence. Their feedback helped us to reword and reorganise the survey.

2.1 Participant selection and recruitment

Our goal was to recruit participants who were recently active in AI research and related fields. Because our interest was in gaining a deep understanding of a set of AI researchers rather than general perspectives, we took a purposive, convenience sampling approach. We identified candidate participants on the online academic article repository, ArXiv. We identified preprint papers in the following AI computer science subcategories: cs.AI, cs.LG, cs.CV and cs.CL. This includes researchers in different branches of AI, researchers from other science and engineering fields that applied AI tools and methods, and researchers from social science who study AI and its applications. The protocol subsection in the Appendix explains specific choices made. We built a corpus of approximately 140,000 papers. Using the Qualtrics survey platform we sent an initial invitation email and two follow-up emails to a sample of 99,516 authors of articles in our corpus, randomly selected. No incentives were provided to survey respondents. The study was reviewed and awarded research ethics clearance by the UCL Department of Science and Technology Studies Research Review Panel (reference STSEth369).

---

[10] See: The Ada Lovelace Institute and The Alan Turing Institute (2023). How do people feel about AI? A nationally representative survey of public attitudes to artificial intelligence in Britain. AI Survey. https://attitudestoai.uk/

[11] See: Office for National Statistics. (2023). Public awareness, opinions and expectations about artificial intelligence. Office for National Statistics. https://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/articles/publicawarenessopinionsandexpectationsaboutartificialintelligence/julytooctober2023

2.2 Data collection and analysis

The survey was open for a month, between June 20th and July 22nd 2024. At the end of this period, 7,595 surveys had been started and 5,318 of those completed. We considered for analysis only completed surveys and those that had clicked 'yes' on the pre-survey consent form. The final dataset for analysis contained 4,260 responses.

Before the analysis, all data collected that included identifiable personal information such as email addresses were removed. Where respondents requested further information about the survey results, contact details were stored separately. The original dataset was stored in UCL's secure data repository.

Descriptive statistics for each survey question were prepared using a set of Python scripts. In the external comparison where other public survey results are referred to, the statistics are retrieved from the original dataset provided. We also manually read some of the free text results. Quotes from the free text responses used in the report are original quotes without processing. Survey results were presented at an internal UCL seminar and feedback was also solicited from colleagues on the Public Voices in AI project team.

2.3 What we know about the survey respondents

AI researchers who took part in the survey are not a homogenous group. This report shows significant variation in how researchers expect technology to develop, and who should be responsible for the positive and negative impacts of AI. Throughout the report we highlight multiple differences, first in terms of underlying demographics of respondents, the sectors they work in as well as their roles. Second, we highlight instances where there are strong associations between sub-groups we have identified in the analysis. This includes groups we have labelled as 'optimist'/'pessimist' researchers, groups who express opposing views on the inevitability of technological change, and groups of researchers whose views on social and technological issues differ significantly based on the country or region they live in.

77% of our respondents work in university or research organisations, while 23% are in industry or 'government / public sector'. 71% of respondents have PhDs, and another 22% have some graduate degree.

Almost all of our respondents' jobs involve AI. 75% replied 'mostly' to the question 'to what extent does your job involve AI?' A further 20% replied 'somewhat', 4.7% replying 'slightly' and 32 responding 'not at all'.

Soliciting open text responses, we asked 'How would you describe your job or role?' Responses (4,163) included a mix of job titles (for instance 'chief data scientist' or 'I am working as a machine learning engineer'), indicators of academic status (for instance postdoc, Academic Faculty and professor) and descriptive accounts of common tasks (e.g. medical image analysis, neural analysis, large language modelling). 611 respondents told us they were assistant, associate or full professors and 412 respondents told us they were students, mostly at graduate or PhD level.

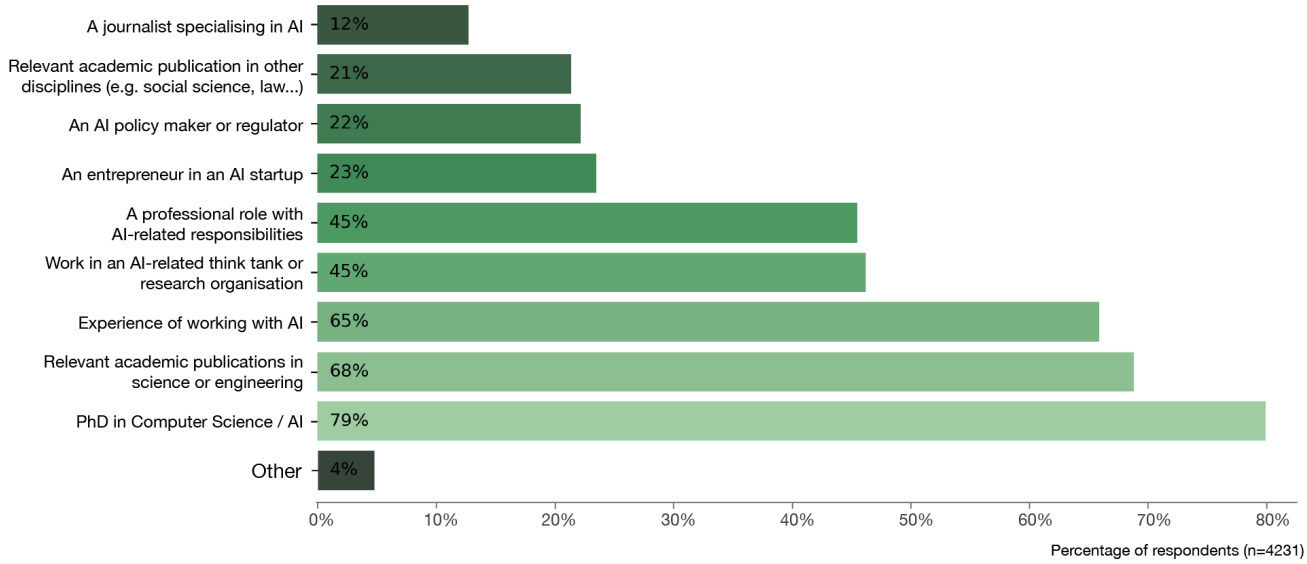
Where indicated, the majority of tasks and roles were in science, technology, engineering and mathematics disciplines. Checking for non-technical tasks and roles, only 12 responses included the word 'social', for instance 'social research',

'computational social scientist', 'social researchers' and 'social impact'. 19 responses included 'safety' and one response included either 'policy', 'politics' or 'governance' ('AI policy/governance'). Using topic modelling processes we produced 16 coherent topics which give insight into jobs and roles across all responses. The top six topics, which account for more than half the responses, are: postdoctoral research and teaching (15% of responses), graduate students (9%), AI model developers (9%), data software engineers (7%), machine learning engineers (7%), computer science image processing (7%). 42% of our respondents are under 35 years old. We have a relatively young pool of respondents with a high ratio of advanced degrees to primary degrees.

81% of respondents described their gender as male and 18% female. 40 respondents identified as non-binary, 10 self-described their gender and 117 chose not to identify with a given gender or add a response of their own.

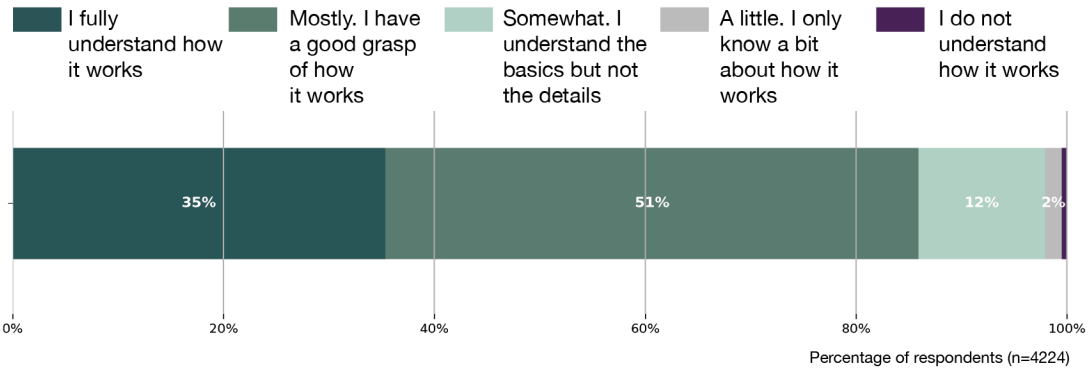
When it comes to views on what counts as AI expertise (figure 1), most respondents valued hands-on technical experience such as working with AI (65%) and academic signifiers such as PhDs in technical subjects (79%) and publications (68%). Far fewer respondents told us that relevant journalism, academic experience from non-technical domains, policy making or experience in an AI start-up qualified someone as an expert in AI (each less than 25%).

Figure 1 'What do you think qualifies someone as an expert in artificial intelligence?' (click all that apply)



Most respondents told us they understand how AI works (figure 2). This is important because it runs counter to some commentary about AI innovation, which emphasises the mysterious nature of technologies like large language models.

Figure 2 'Think about an AI system that you are most familiar with. To what extent do you understand how it works?'



We asked respondents where they live. Respondents (n=3,765) listed 92 countries. The top countries are the USA (25%), the UK (11%) and China (8%). 29% of respondents live in European Union countries. 714 respondents, (17%), are from one of 47 low and middle-income countries entered. Of this number 60% are from China and India. It is important to note that our survey was available only in English, and that the population of researchers we sampled had also published their papers in English. Statistical analysis shows that the country respondents live in has strong associations with the answers to most questions. For instance, when compared with all aggregate responses, significantly more respondents living in China (8%) agree that AI should be developed as quickly as possible, and that artificial general intelligence is inevitable (for graphs and analysis see Insight 3 on technological inevitability). It is also noteworthy that responses from people living in the USA do not tend to deviate strongly from aggregate responses across the survey.

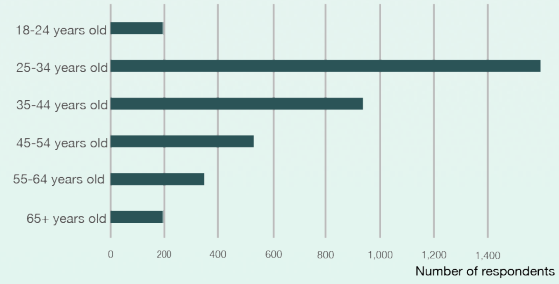
Box 1 **Researchers and the public in the United Kingdom**

Our analysis shows that where AI researchers live matters (see Figure 9), despite the global and interconnected nature of AI R&D. We looked at responses from researchers who live in the UK to observe differences and similarities with the global sample, and because we were comparing with data from nationally-representative UK public attitude surveys. UK was strategically selected - the project team has expert knowledge of UK institutions and public attitudes to science and technology. Similar to the USA, UK responses are close to all aggregate responses. However there are some subtle but significant differences:
Slightly more respondents in the UK agree that the people who create AI systems should be held responsible for their use (70% versus 66%). Respondents living in the UK are slightly more pessimistic about AI benefits and more sceptical about positive claims (for instance fewer respondents compared with global analysis think there are more benefits than risks). UK respondents tend to be more likely to agree to involving the public in deploying AI, but importantly, not developing AI. Significantly fewer respondents living in the UK agree that if the public understood AI they would trust it more. More respondents in the UK compared with the USA or global pool are concerned that negative impacts of AI will be felt by some groups in society more than others. Furthermore, UK respondents more than any other top 10 country agree that regulators need to consider different groups in society when regulating AI.

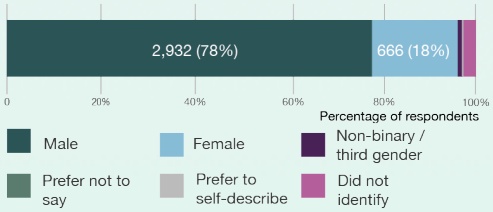
Demographic and geographic highlights are presented on the next page.

Box 2 **Demographic information about survey respondents**

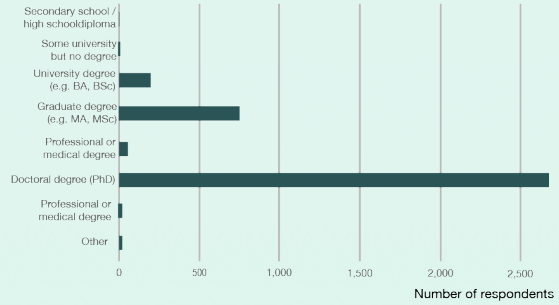
Age of respondents



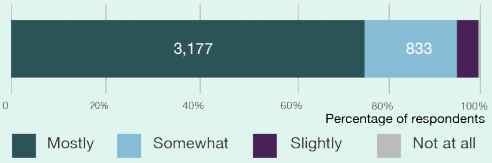
Reported gender n=3,765



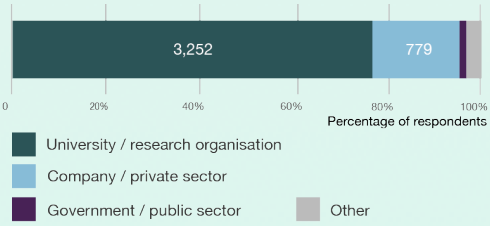
Educational attainment



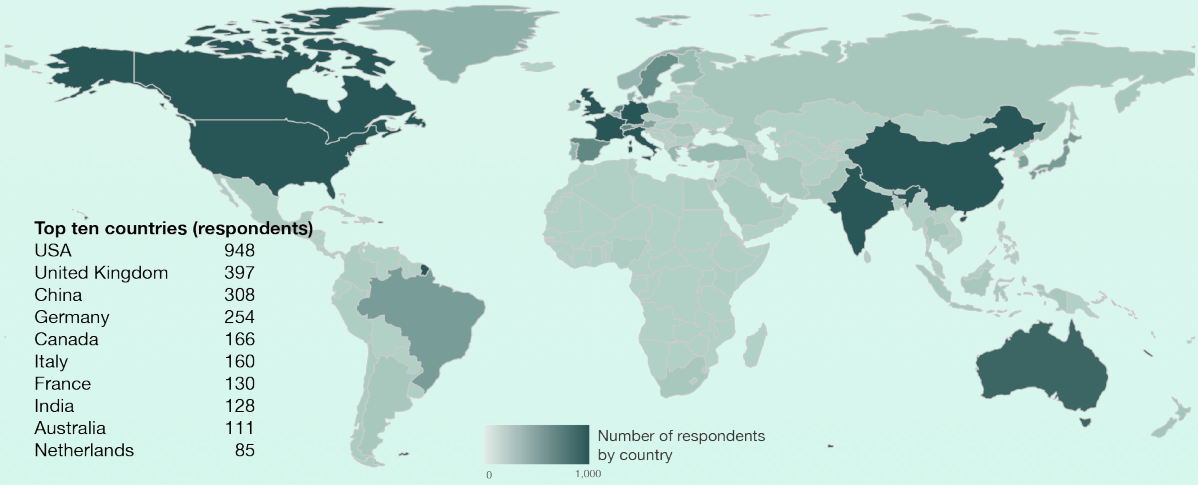
Extent to which respondents' jobs involve artificial intelligence



Sector in which respondents work



Country in which respondents live n=3,765



Top ten countries (respondents)

USA	948
United Kingdom	397
China	308
Germany	254
Canada	166
Italy	160
France	130
India	128
Australia	111
Netherlands	85

Note: The map uses a normalised colouring scale. This helps identify the top countries by respondents. However, the gradient does not represent in a linear fashion the considerable differences between these countries. The USA is a significant outlier.

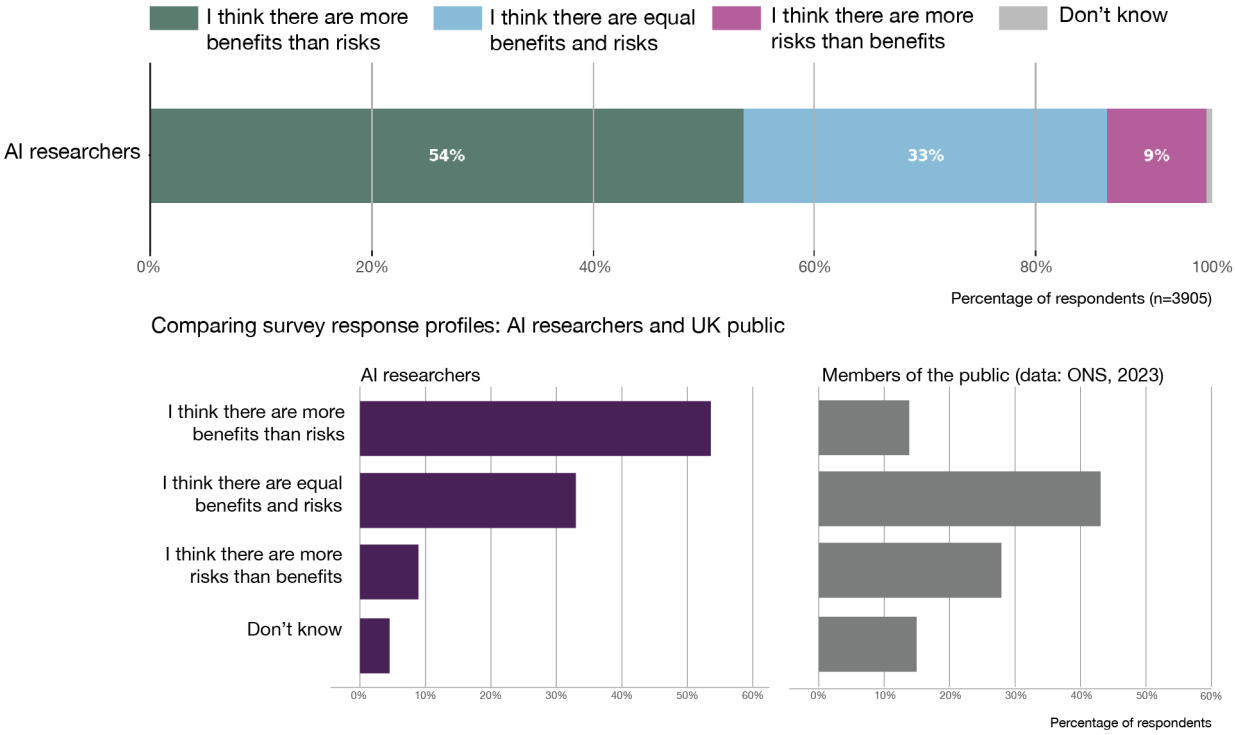
3. Survey results

3.1 AI researchers' views on hopes, fears, benefits and risks

Insight 1

Researchers are more positive than members of the public about the benefits of AI

Figure 3 'Which statement best reflects your view about the benefits of AI?'

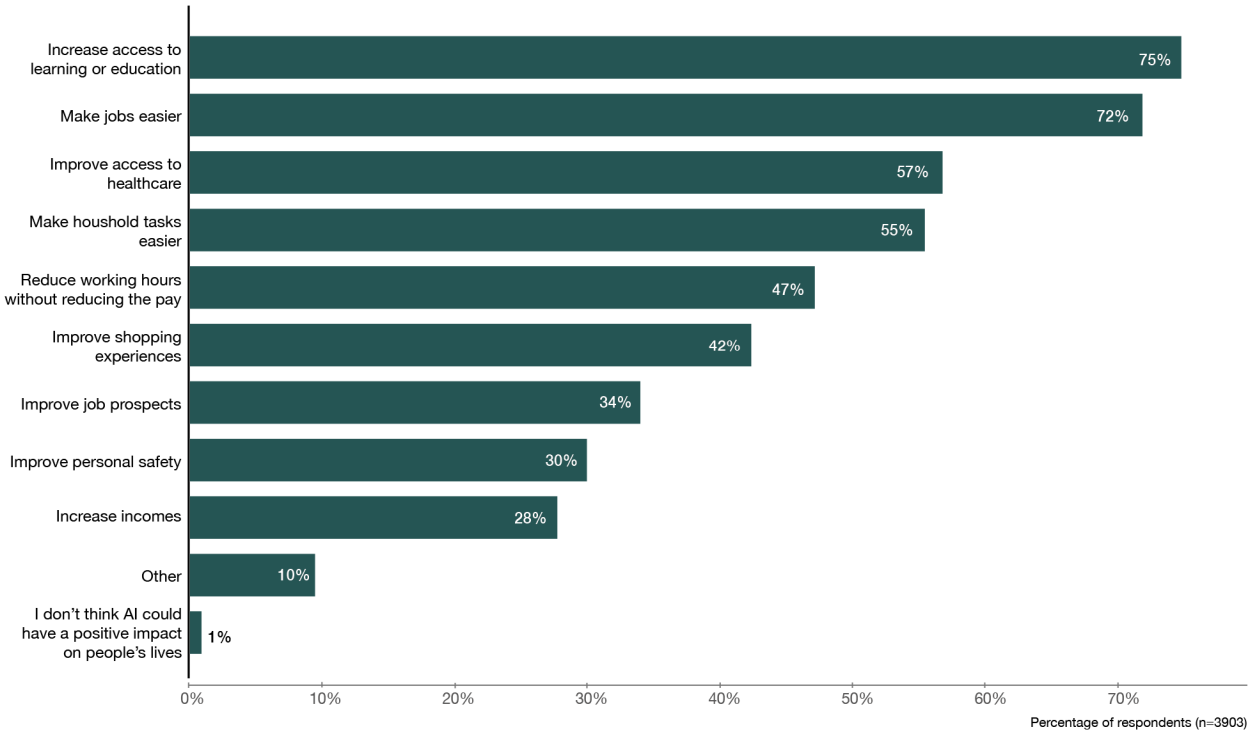


When asked about the impact of artificial intelligence, researchers have a more positive outlook than the public. A large majority of researchers (87%) believe the benefits either outweigh or balance the risks. Only 9% believe the risks outweigh the benefits. In comparison, the response profile from an ONS study that surveyed members of the UK public shows a slight majority of the public (57%) believe the benefits either outweigh or balance risks. 28% of the UK public (28%) believe the risks outweigh the benefits.[12]

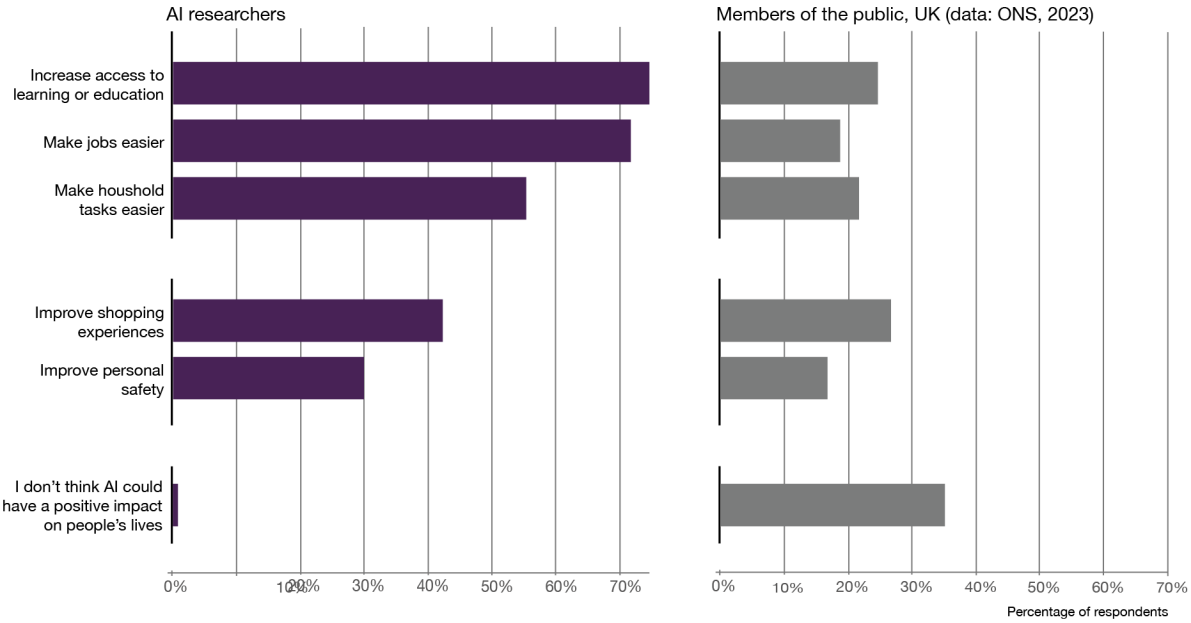
UK-based researchers in our study are less positive about AI benefits, nevertheless their responses tend much closer to the aggregate views of AI researchers than the UK public, responding 45%, 36% and 13% respectively to the response options stated above.

[12] For this question we have used data on the views of the UK public obtained from the Office for National Statistics' 2023 survey, 'Public awareness, opinions and expectations about artificial intelligence' (ONS 2023). 'Don't know' in the ONS data is represented here as 15% adjusting for a rounding difference of 1% in the original ONS visualisation.

Figure 4 'In which ways, if any, do you think artificial intelligence could have a positive impact on people's lives?'



Comparing survey response profiles: AI researchers and UK public



AI researchers and the public vary significantly in how they think AI can have a positive impact on people's lives, but not across all categories we asked about.[13] When asked to consider the positive impacts AI could have on their lives, researchers and the public agreed on two of the top three

[13] For this question we again compared AI researcher views from our own survey with responses to the 2023 ONS survey of members of the public in the UK.

answers: Increased access to education and healthcare. However, responses from AI researchers are much more positive when it comes to issues such as *increasing access to education* and *making jobs easier* and *making household tasks easier (see the response profile comparison between AI researchers and UK public survey responses)*.

Responses from researchers and the public are more similar when it comes to issues such as *improving shopping experiences* and *personal safety*. Nevertheless, AI researchers report more positive views on the impacts of AI across every issue we asked about.

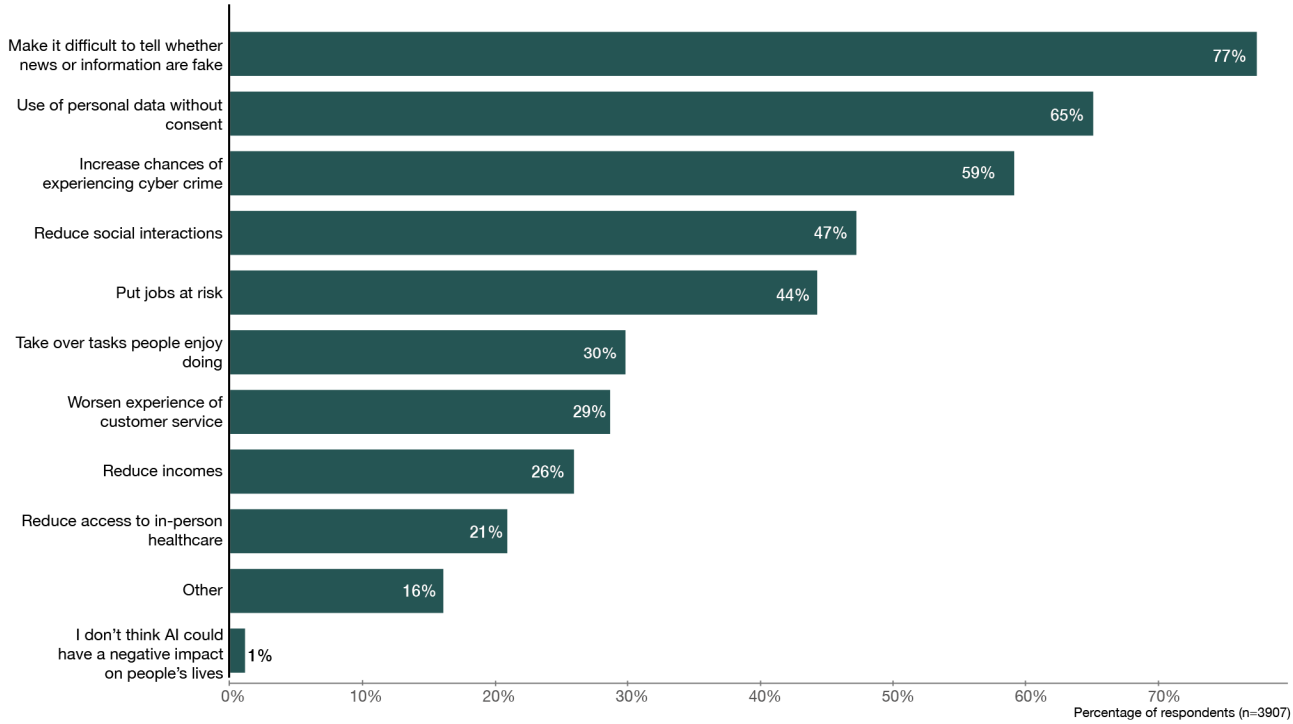
Less than 1% of AI researchers agreed with the statement 'I don't think AI could have a positive impact on people's lives'. In comparison, 36% of respondents to the 2023 ONS survey of members of the public in the UK do not believe AI will have a positive impact on people's lives.

It is interesting to consider the gaps between researchers and the public in perceptions about the future of work. The public do not share AI researchers' optimism about the technology improving job prospects but, as we can see in the next section, are more relaxed about AI putting jobs at risk.

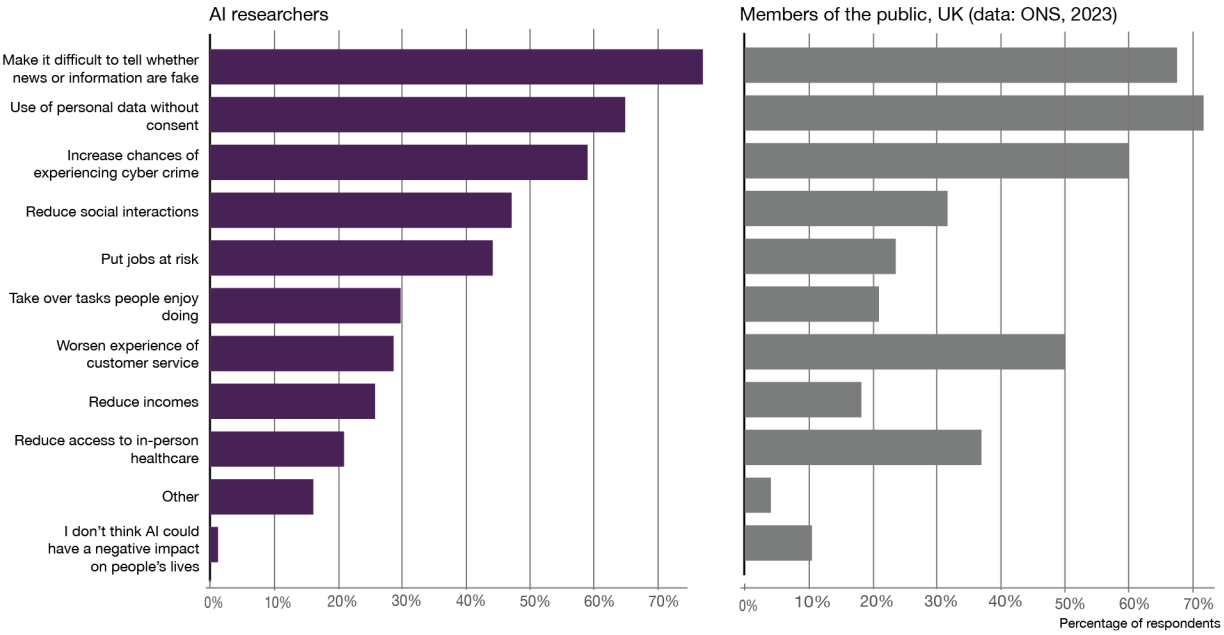
Insight 2

Researchers and the public share concerns about disinformation, data use and cybercrime

Figure 5 'In which ways, if any, do you think artificial intelligence could have a negative impact on people's lives?'



Comparing survey response profiles: AI researchers and UK public



When asked to select negative impacts AI could have on their people's lives, researchers and the public agreed on the top three: disinformation, use of data without consent, and cybercrime. Perspectives diverge after

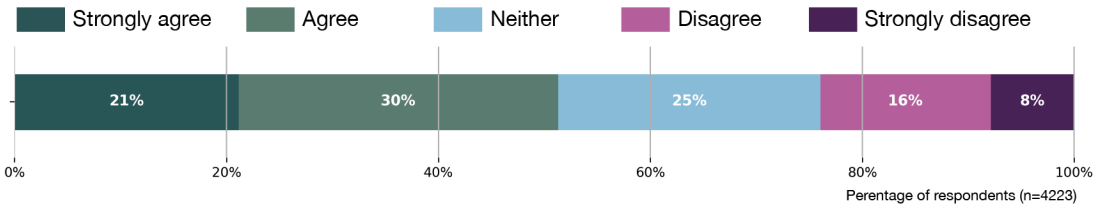
that.[14] For the public, the fourth agreed concern is worsening experiences of customer service. For AI researchers, it is reduced social interactions.

What comparison of the response profiles shows us here however is the relative consensus between researchers and members of the UK public when it comes to negative impacts of AI, compared with an 'optimism gap' when it comes to benefits.

Insight 3

There is sense of technological inevitability in AI research

Figure 6 'To what extent do you agree or disagree with the following statement?: Artificial general intelligence (AGI) is inevitable'



The idea of artificial general intelligence (AGI) is not well-defined, but it has become a prominent term in debates about AI. It has a number of alternative technical definitions (see Box 3), but it has also served as a shorthand for a scientific end goal for AI research. We wanted to know not just whether AI researchers bought into this term and its implied status as a goal for AI research, but also if they saw some sense of inevitability. Just over half of respondents (51%) agreed that artificial general intelligence (AGI) is inevitable.

Box 3 Perspectives on artificial general intelligence

OpenAI: "highly autonomous systems that outperform humans at most economically valuable work"

Pennachin and Goertzel: "a software program that can solve a variety of complex problems in a variety of different domains, and that controls itself autonomously, with its own thoughts, worries, feelings, strengths, weaknesses and predispositions".

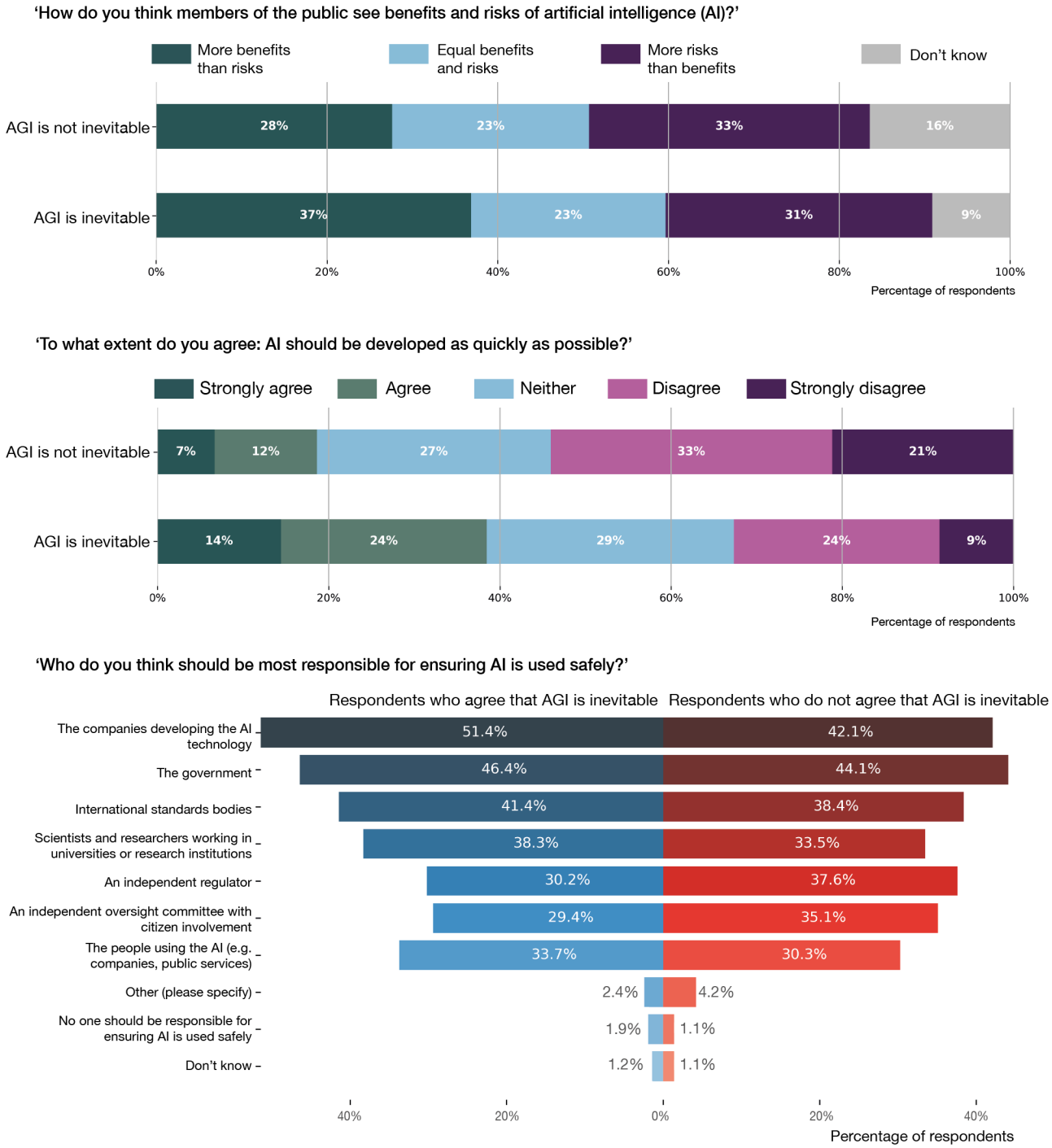
Peter Voss: "a computer system that matches or exceeds the real time cognitive (not physical) abilities of a smart, well-educated human" [3].

Stuart Russell and Peter Norvig: "a universal algorithm for learning and acting in any environment"

From Gebru & Torres (2024). The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence. First Monday. https://doi.org/10.5210/fm.v29i4.13636

[14] For this question we again compared AI researcher views from our own survey with responses to the 2023 ONS survey of members of the public in the UK. See: Office for National Statistics. (2023). Public awareness, opinions and expectations about artificial intelligence. Office for National Statistics. https://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/articles/publicawarenessopinionsandexpectationsaboutartificialintelligence/julytooctober2023

Figure 7 The visions and values of researchers who think that AGI is inevitable

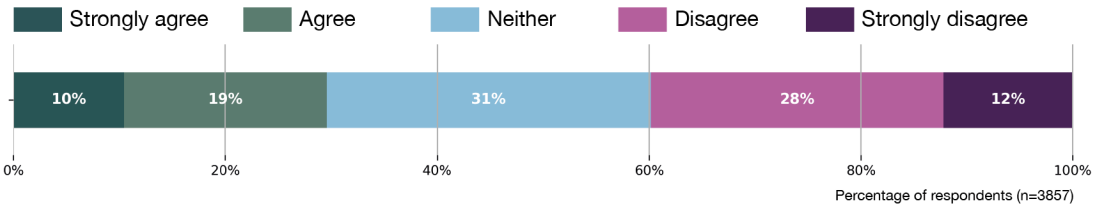


There is some disagreement about the prospect of AGI among respondents there is also disagreement about whether developing AI as quickly as possible would be the best approach.

Researchers who agree that AGI is inevitable are more likely than other respondents to think that AI should be developed as quickly as possible and that there are more benefits than risks to be gained from AI. This group of respondents are also more strongly associated with views that experts such as companies and scientists should be responsible for ensuring AI is

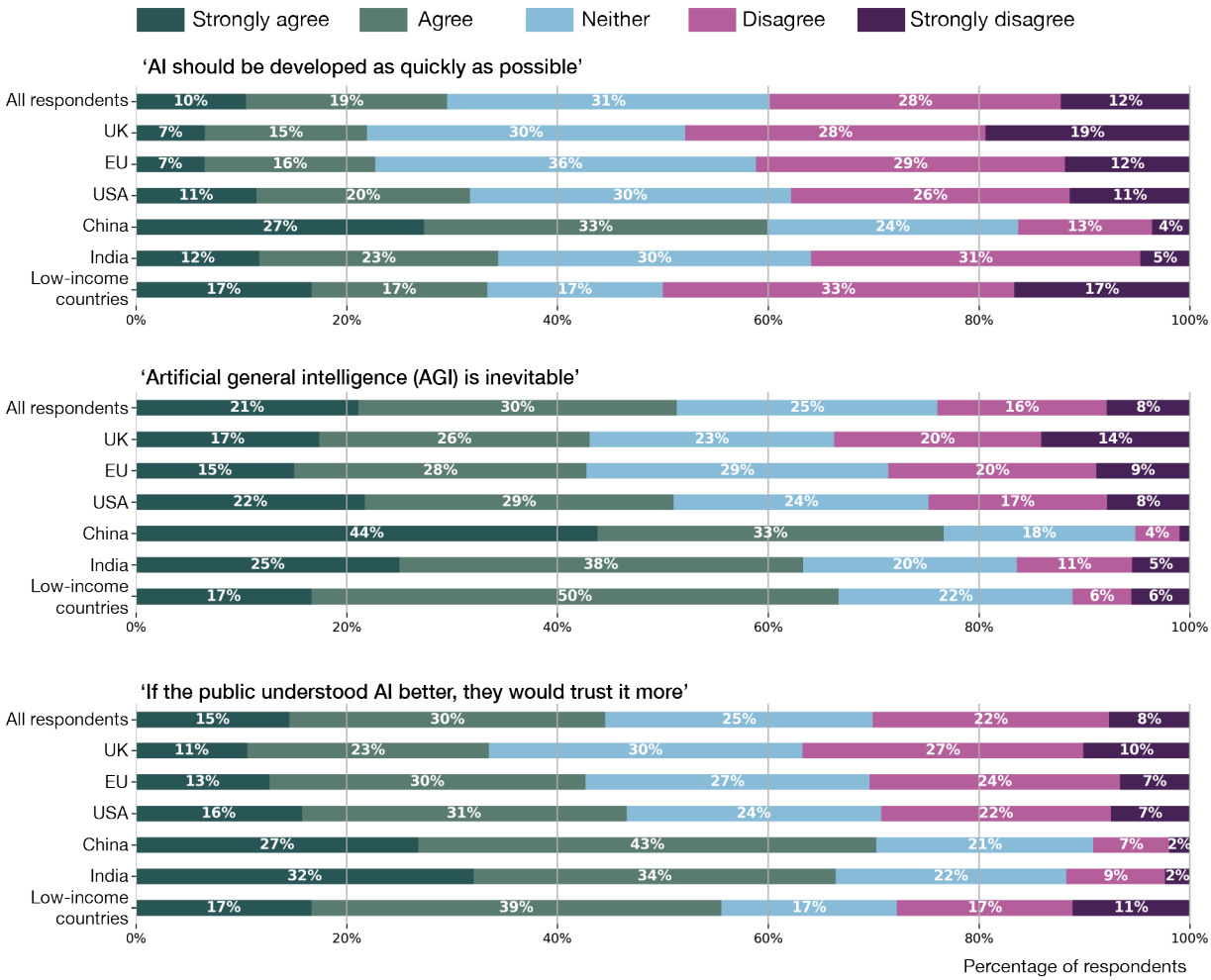
used safely, and less likely to think that independent regulators or citizens should be responsible. They also report significantly less experience of using methods for understanding public views, values and voices which we cover in Figures 25 and 26.

Figure 8 'To what extent do you agree or disagree with the following statement?: AI should be developed as quickly as possible'



Only 29% of all respondents believe that AI should be developed as quickly as possible, although as we see in the next section this idea is more attractive for AI researchers who take an optimistic stance on artificial intelligence.

Figure 9 A sense of technological inevitability in China and India.
Extent to which respondents agree with the following statements:



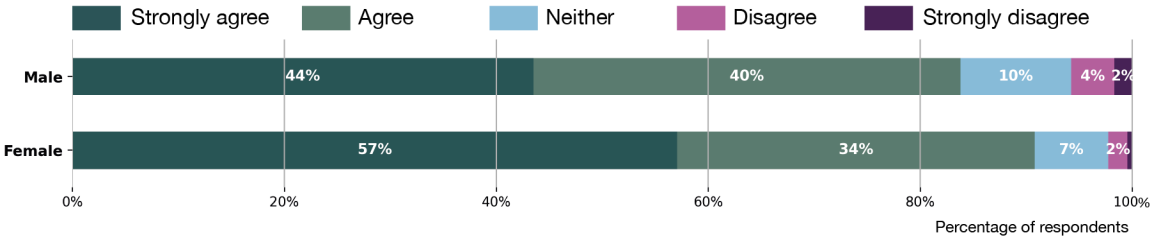
We found a strong association between reported technocratic views of technology and the country where those respondents live. Significantly more respondents living in China (8.2% of respondents who told us where they live) agree that AI should be developed as quickly as possible, that AGI is inevitable and that if the public understood AI better, they would trust it more. Likewise, respondents living in India were more likely than not to think AGI is inevitable and that public trust would increase with public understanding of AI.

Also, respondents living in China are less worried that some members of the public will be more negatively impacted by AI than others (71% vs 82% for all respondents) and fewer of them are worried that some members of the public will benefit from AI more than others (57% versus 70%).

Insight 4

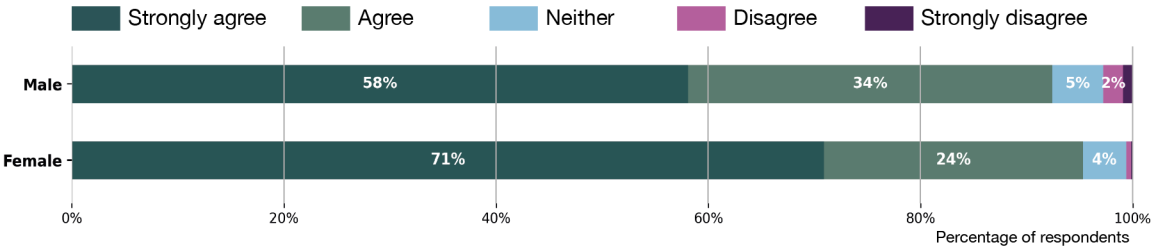
Different views on AI associated with gender and the country where researchers live

Figure 10 'To what extent do you agree it is important that developers consider how AI impacts different groups?'



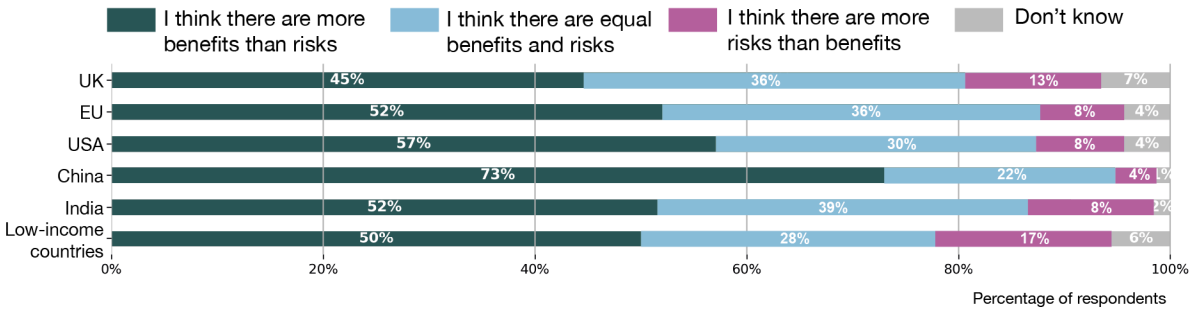
Analysis reveals a strong association between reported gender and questions concerning the impact of AI, with significantly more female than male respondents agreeing with the following statements: '*It is important that developers consider how AI impacts different groups'* and '*It is important that regulators consider how AI impacts different groups'*. (Note: we have graphed only responses selected as male or female given the very small numbers of respondents who told us they were non-binary/third gender or who self-reported their gender).

Figure 11 'To what extent do you agree it is important that regulators consider how AI impacts different groups?'



We also see strong association between reported gender and questions concerning decisions about AI research funding in education and policing (but not in healthcare and military) and in questions about public understanding of AI. Conversely there is not a strong association between gender and questions of who should be responsible for AI safety or the real-world impacts of AI systems.

Figure 12 What AI researchers think about benefits and risks is strongly associated with the country they work in: 'Which statement best reflects your view about the benefits and risks of AI?'



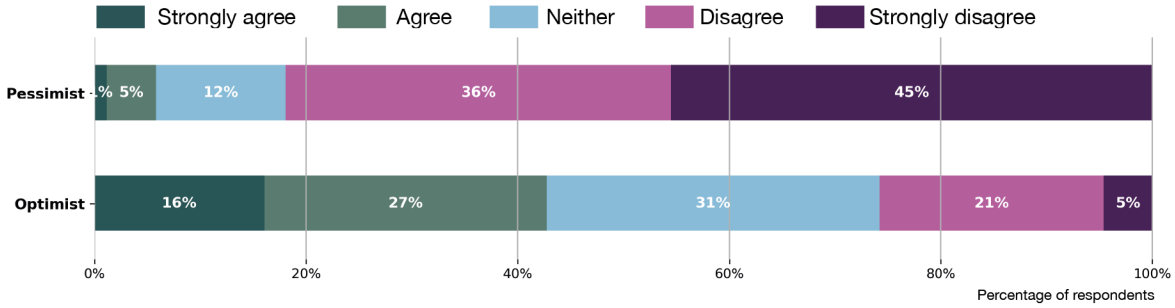
AI researchers report significantly different views on the benefits and risks of artificial intelligence based on the country they live. Researchers in China (8% of respondents) are much more likely than not to think there are more benefits than risks. Results elsewhere are more mixed.

Insight 5 'Optimist' and 'pessimist' researchers report different views on AI

Whether or not a researcher believes AI has more risks or benefits correlates with their outlook on a number of questions we surveyed. Using a subset of questions to appraise outlook,[15] we identified two groups of respondents. We have labelled these optimists and a minority group we have labelled pessimists.

'Optimists' (54%) outnumber 'Pessimists' (9%), and strong discrepancies are visible between the two groups.

Figure 13 'To what extent do you agree or disagree with the following statement: AI should be developed as quickly as possible.'

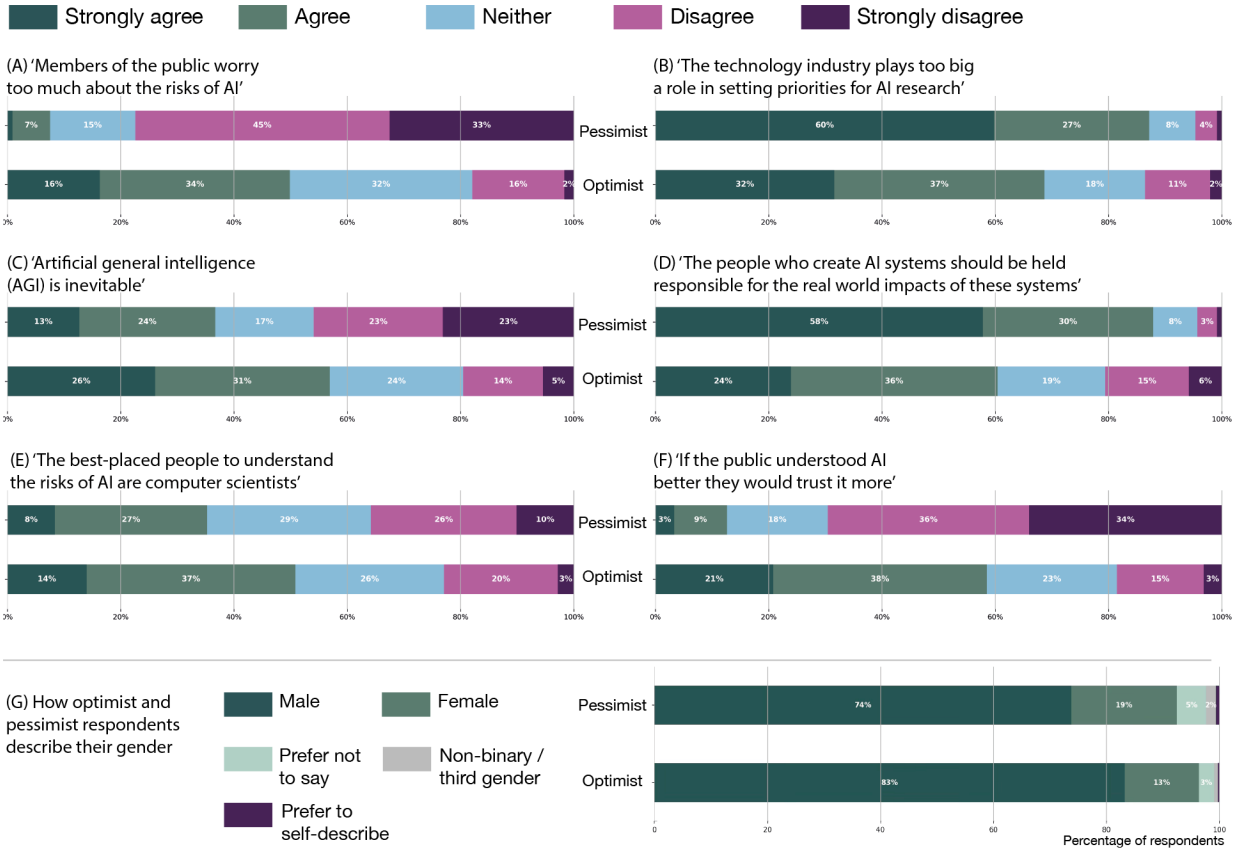


'Optimist' and 'pessimist' respondents differ significantly on the question of how fast AI should be developed. The strongest matter of difference was over how fast AI should be developed. 42.8% of Optimist researchers agreed that AI should be developed as quickly as possible, only 8% of pessimist researchers shared the same sentiment. We explore these differences further in Box 4.

[15] The baseline question was Q2.3. 'Which statement best reflects your view about the benefits and risks of AI?' See appendix for full list of survey questions.

Box 4 **Comparing optimist and pessimist outlooks**

Extent to which pessimist and optimist respondents agree with the statements



Pessimists are much more likely to agree that 'members of the public worry too much about AI' (fig. A). Pessimists are also more likely to agree that 'The technology industry plays too big a role in setting priorities for AI research (fig. B).

Optimists are more likely to agree that 'AGI is inevitable' (fig. C). Optimists are more likely to disagree that 'the people who create AI systems should be held responsible for the real world impacts of these systems' (fig. D).

Optimists are also more likely to agree that 'the best-placed people to understand the risks of AI are computer scientists' (fig. E). Optimists are much more likely to have a 'deficit model' of trust (see below). Less than one in five disagree with the statement 'if the public understood AI better, they would trust it more' (fig. F). Pessimists are also much more likely to agree that it's important that members of the public know about the risks of AI and that the public should be involved in regulating AI.

Finally, the pessimist group of respondents is more diverse when it comes to gender with significantly more researchers reporting as female and non-binary/third gender (fig. G).

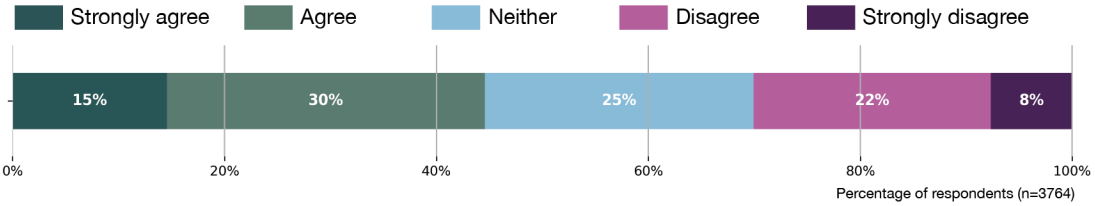
3.2 Public involvement

Insight 6

Many researchers have a 'deficit model' of the public

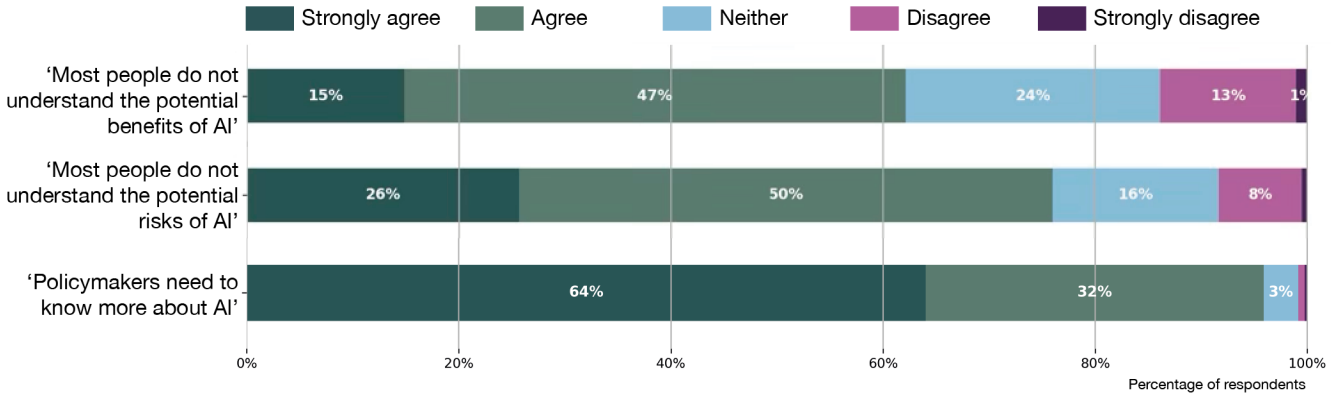
The previous section shows that researchers and the public differ in their responses when asked about AI risks and benefits. And UK public surveys such as those by by The Ada Lovelace Institute/Alan Turing Institute and the Office for National Statistics reveal substantial scepticism about other aspects of AI. With other technologies, it has been common for scientists and policymakers to put public mistrust down to ignorance or misunderstanding of science, although we know from social research that this is explanation is flawed.[16] However, our survey data provide strong evidence that this 'deficit model' persists among AI researchers.

Figure 14 'To what extent do you agree or disagree with the following statements?: If the public understood AI better they would trust it more'



Researchers believe that the public would trust AI more if they understood it better (45% agree vs 30% disagree). Positive responses to this statement are more strongly associated with male researchers than female researchers.

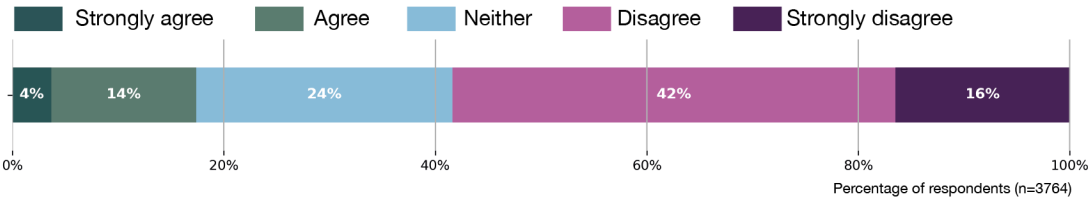
Figure 15 'To what extent do you agree or disagree with the following statements?'



The public are seen as ignorant. Respondents think the public understand neither the benefits of AI (62% agree vs 14% disagree) nor the risks (76% vs 24%). 96% of respondents think that policymakers need to know more about AI.

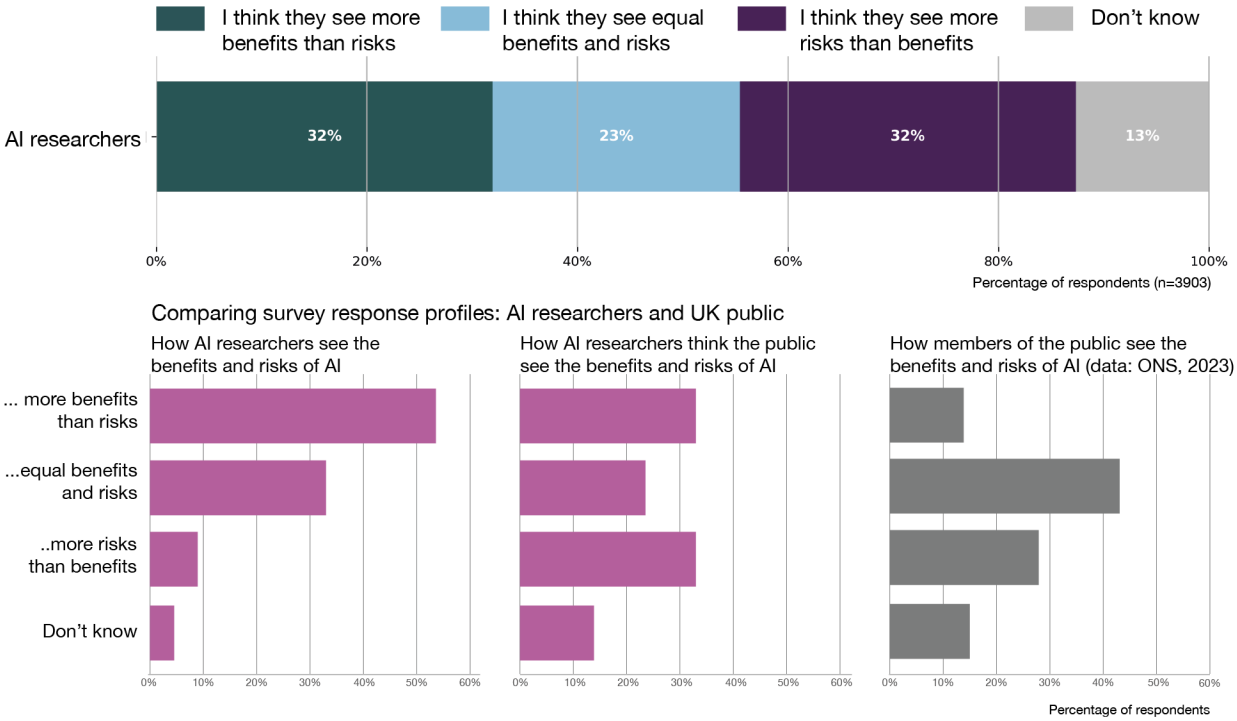
[16] Wynne, B. (2006). Public engagement as a means of restoring public trust in science–hitting the notes, but missing the music?. Public Health Genomics, 9(3), 211-220.

Figure 16 'To what extent do you agree or disagree with the following statement? - The public are able to weigh up the risks and benefits of AI'



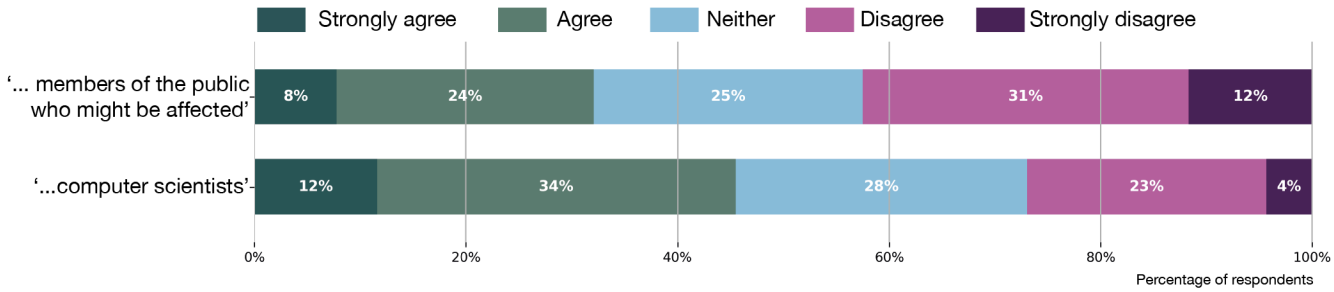
This deficit logic follows through to how respondents think about the ability of the public to evaluate risk and benefit. Only 18% of respondents agreed that public are able to weigh up the risks and benefits of AI.

Figure 17 'How do you think members of the public see the benefits and risks of artificial intelligence (AI)?'



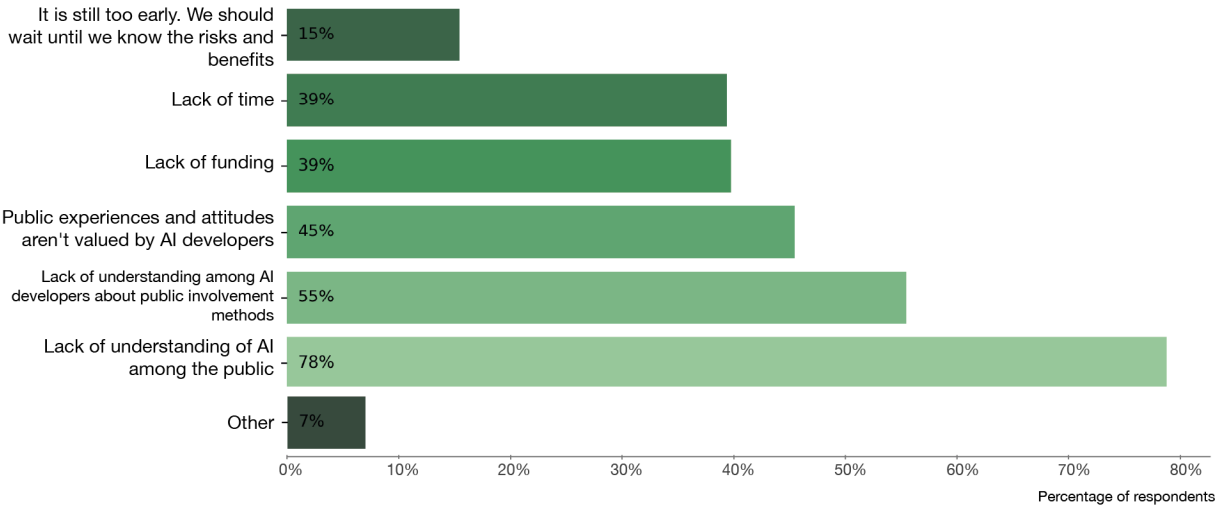
Respondents are split on how they understand the public's appraisal of risks and benefits. When it comes to understanding how the public distribute the benefits and risks of AI, respondents were split: just under one in three replied that they think the public see more benefits than risks, while a similar number replied that they think the public see more risks than benefits while 23% replied that they thought the public see equal benefits to risks.

Figure 18 'The best-placed people to understand the risks of AI are…'



Computer scientists trust themselves more than the public to understand the risks of AI. 46% of researchers believe that computer scientists are best placed to understand the risks of AI. 32% think that the people who might be affected are best-placed. There is a tension within current debates about 'AI Safety' between technical understandings of AI risk which rely on the knowledge of experts such as AI researchers, and 'sociotechnical' understandings which incorporate different kinds of knowledge and expertise.[17]

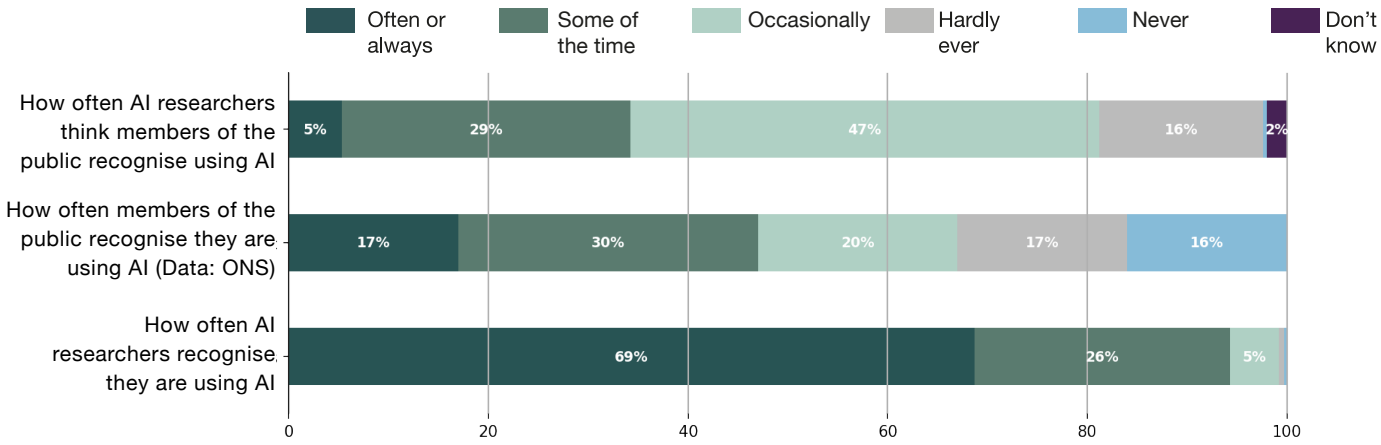
Figure 19 'What do you think are the barriers to involving the public in AI? (Click all that apply)'



Researchers think the biggest barrier to involving the public in AI is "lack of understanding of AI among the public".

[17] Lazar, S., & Nelson, A. (2023). AI safety on whose terms?. Science, 381(6654), 138-138.

Figure 20 'How often do you think you can recognise when you are using artificial intelligence?'

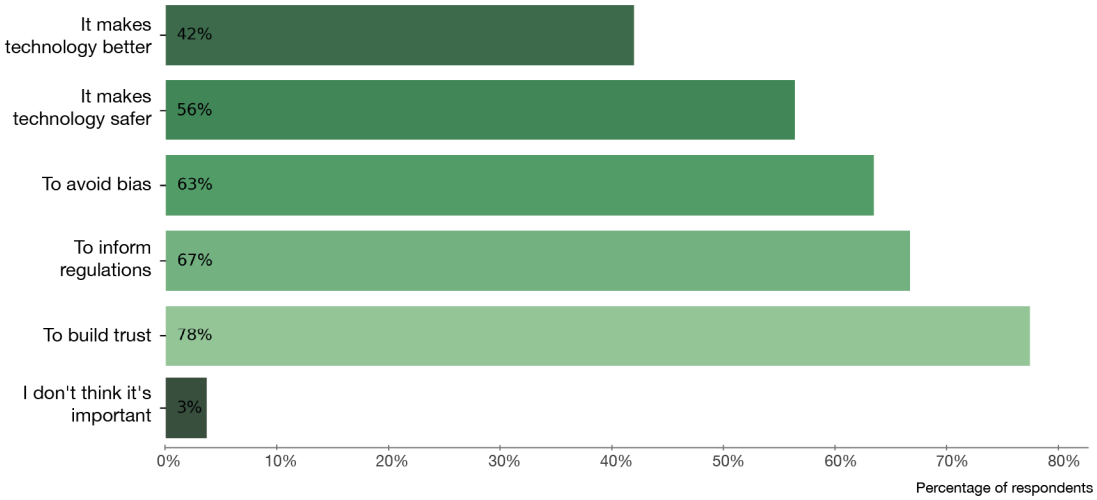


Researchers think they can recognise AI and they think the public can't recognise AI. Researchers are confident that they know when they are using AI. ONS data shows that the UK public believe they have a higher level of understanding, recognising AI often or "some of the time". But researchers think the public are unlikely to know.

Insight 7

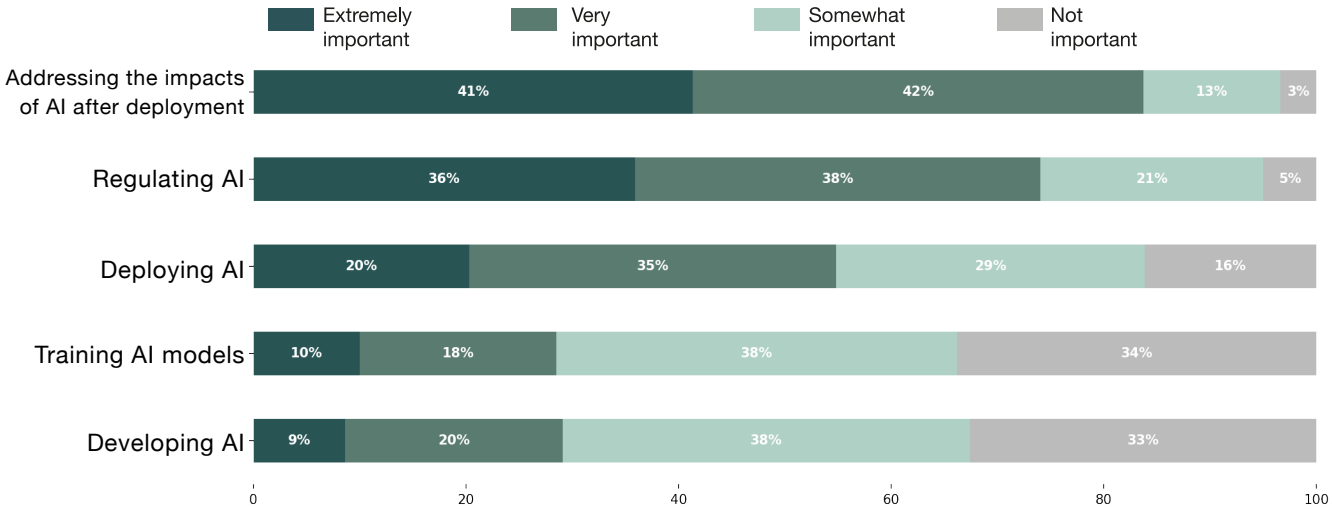
Researchers want the public involved downstream, not upstream

Figure 21 'Why do you think it might be important to involve the public?' (click all that apply)



Almost all respondents agreed with reasons for involving the public in AI. Again, a deficit logic is evident here. Motivations for involving the public in downstream processes such as regulation and building trust were chosen by many more respondents than reasons for involving the public in upstream processes such as safety and making technology better.

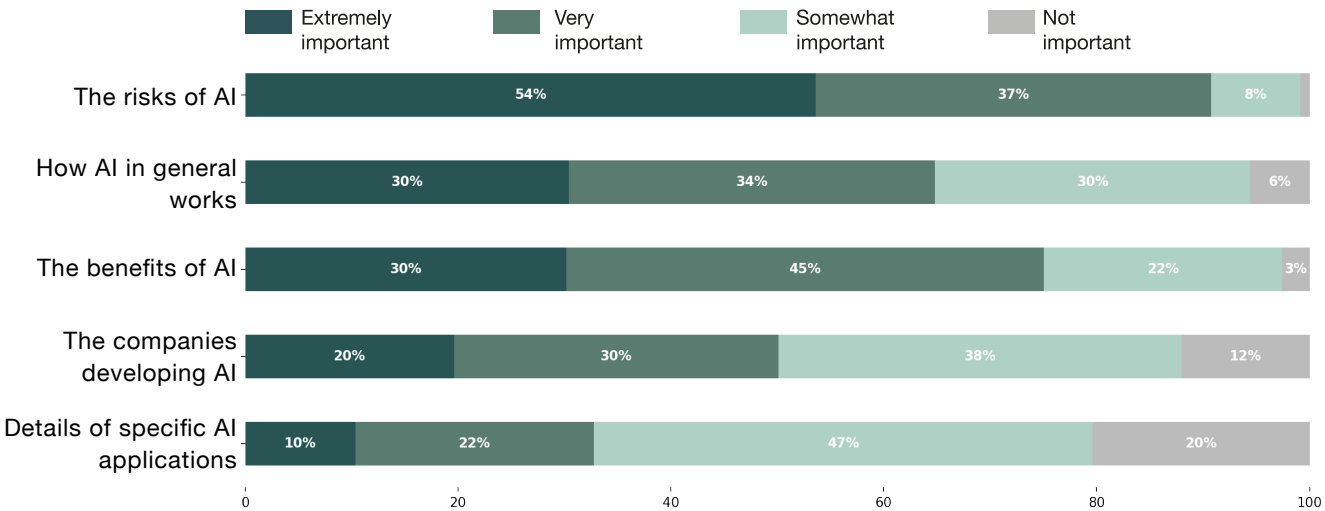
Figure 22 'How important do you think it is to involve the public in decisions about the following tasks?'



Most respondents reported that they believe it is important to involve the public in assessing the impact of AI and regulating AI. Only 3% and 5% of respondents respectively said these were not important issues.

Researchers believe the public should be involved in downstream decisions, 84% said this was at least somewhat important. However, fewer researchers want members of the public involved in upstream decisions, about developing AI or in training models.

Figure 23 'How important do you think it is for members of the public to know about the following:'



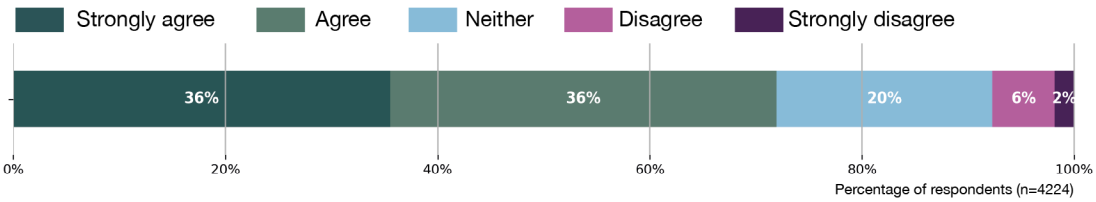
Researchers think it is very important for the public to know about the risks of AI, but fewer think it is important for the public to know about the companies developing AI or details of specific AI applications.

Insight 8

Researchers want AI to reflect human values but they do not pay attention to social science research

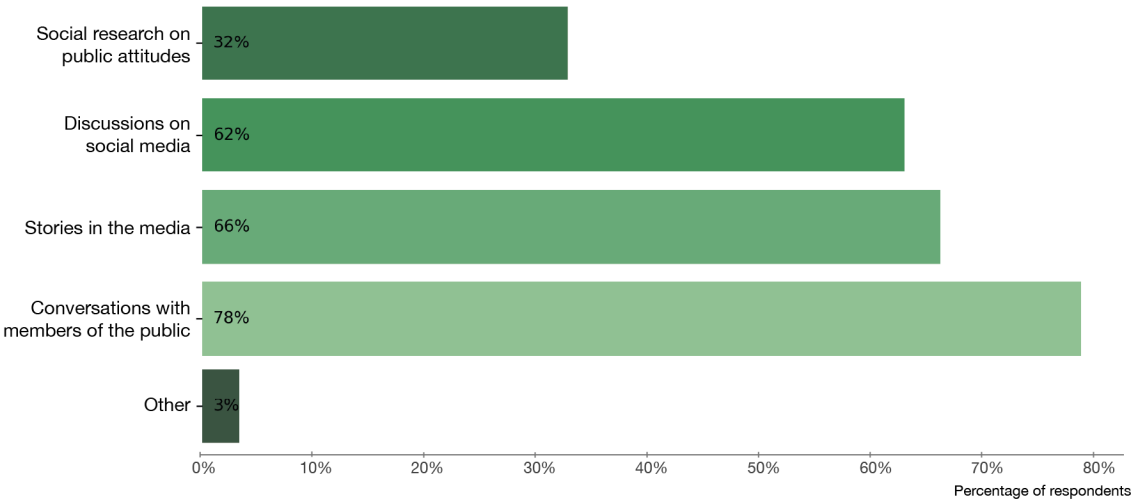
Insight 6 showed us that many researchers have a 'deficit' model of the public. We wanted to know what they knew about the public and how they found out what members of the public think?

Figure 24 'To what extent do you agree or disagree with the following statements? AI systems should be designed to reflect human values'



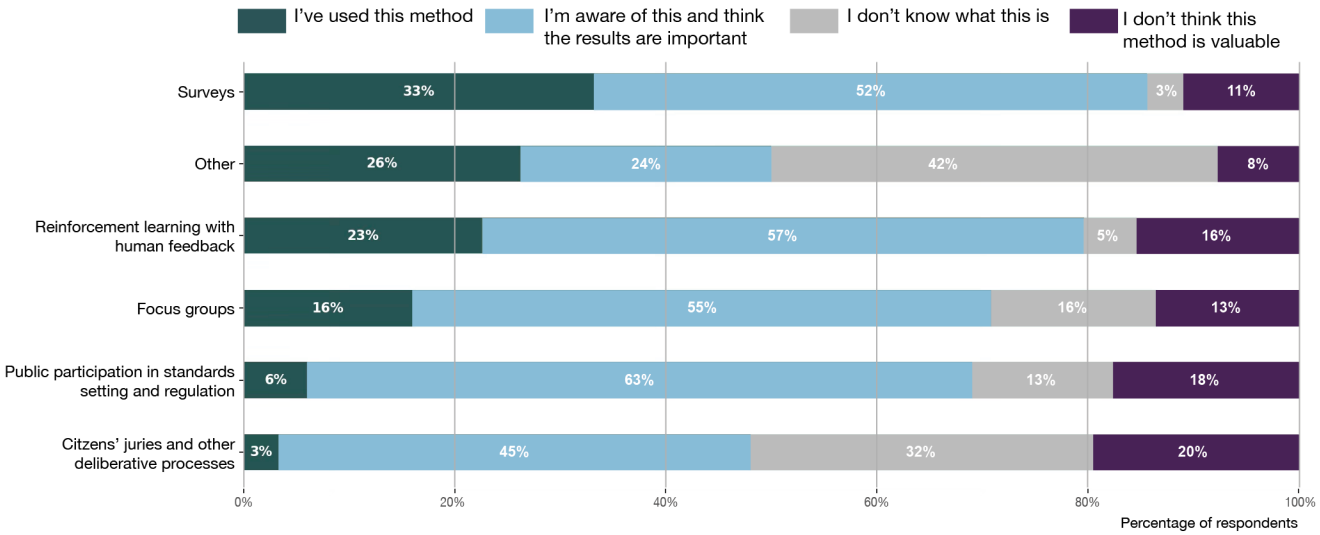
Researchers want AI to reflect 'human values'. But our evidence suggests a lack of demand for the social science that might reveal how diverse these values are.

Figure 25 'How do you normally find out what the public think about AI? (Click all that apply)'



Researchers report learning about public opinion from personal conversations the most, followed by the media and social media. Few AI researchers (32%) report learning from social research.

Figure 26 'Do you have any knowledge or experience of the following methods for public involvement?'



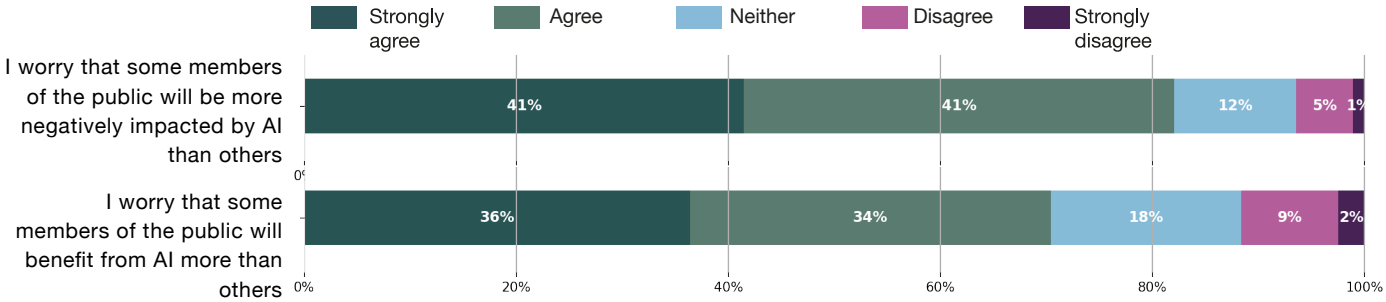
Most respondents have not used methods for involving the public we asked about. The majority of respondents have used or are aware of surveys. Most know about reinforcement learning with human feedback, a now-common technical procedure for the fine-tuning of AI systems. Very few have used citizens' juries and other deliberative processes.

3.3 Responsibilities of AI researchers

Insight 9

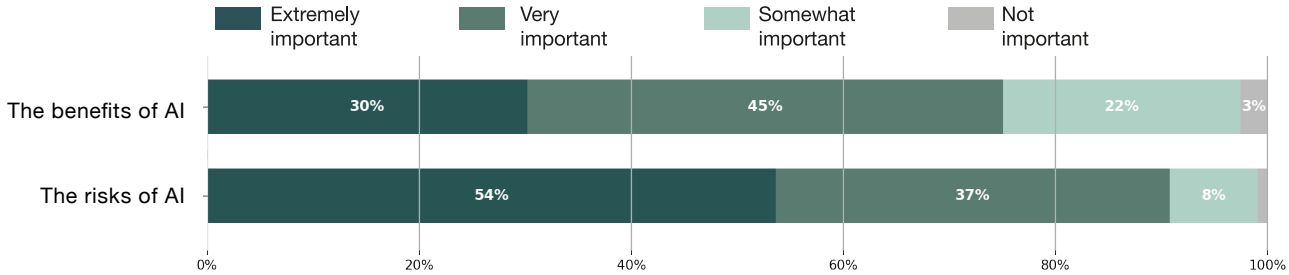
Researchers think debates about risk matter more than those about benefits

Figure 27 ‘To what extent do you agree with the following statements?’



Worries about negative impacts trump hopes about benefits. More researchers are worried about the unequal distribution of negative impacts than positive ones.

Figure 28 ‘How important do you think it is for members of the public to know about the following?’



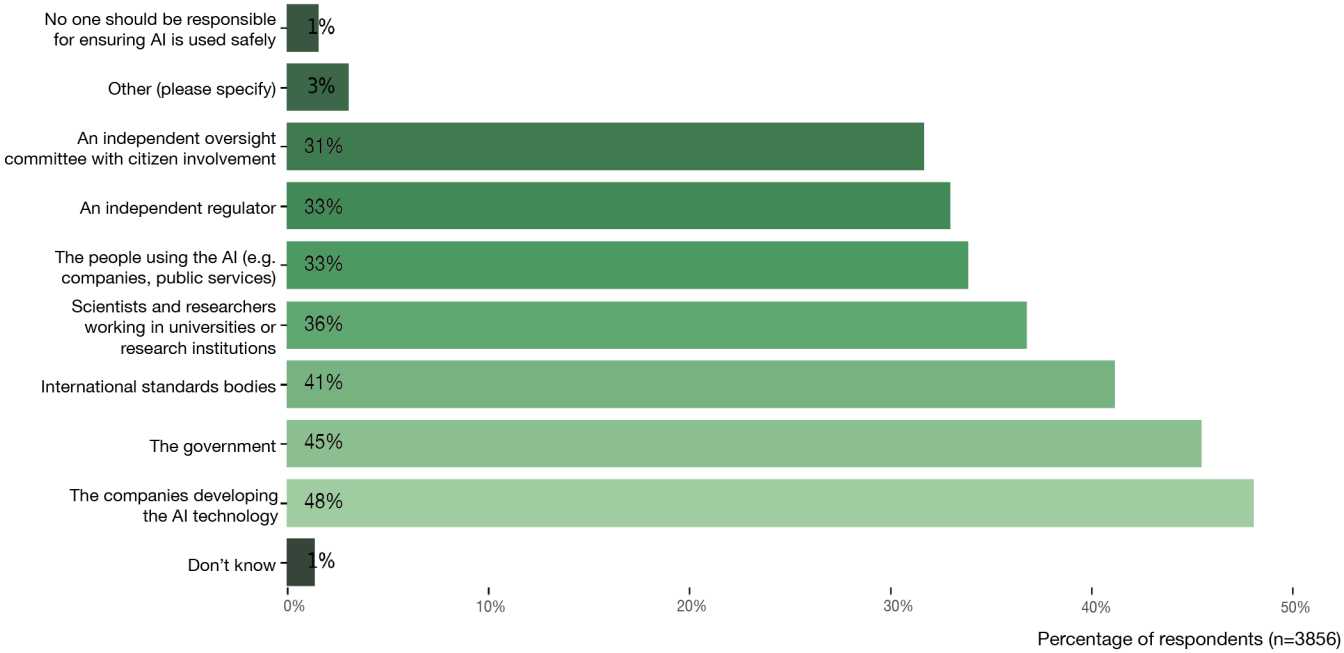
AI researchers think it is more important for the public to know about the risks of AI than the benefits of AI (75% vs 91% saying it is extremely or very important).

This asymmetry reflects a dominant view of innovation and democracy, in technological benefits are taken-for-granted or seen as an expert matter while questions of risk are seen as more appropriate for public debate.

Insight 10

Researchers and the public disagree about who should be responsible for the safe use of AI

Figure 29 'Who do you think should be most responsible for ensuring AI is used safely?' (multiple choices permitted)



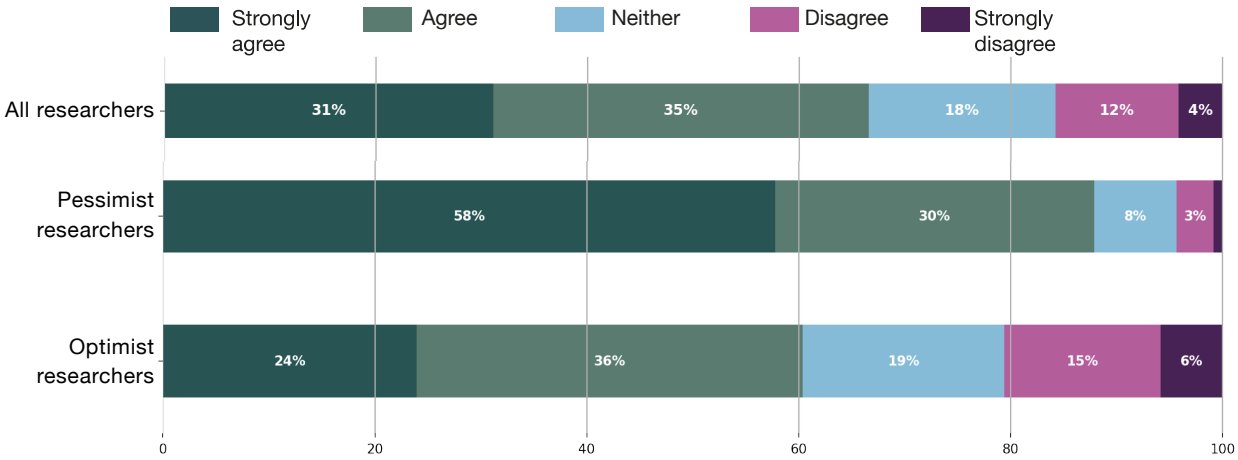
For researchers, the top three responses when asked who should be most responsible for safe use of AI are (1) Companies developing AI, (2) the government, and (3) international standards bodies.

Meanwhile, when asked by a survey, a representative sample of the UK public, listed different priorities.[18] They said: (1) an independent regulator, (2) the companies developing AI and (3) independent oversight committee with citizen involvement.

These findings need to be interpreted in the light of low public trust in politicians. The public have a strong desire for an independent regulator, while researchers think they and companies should be more responsible.

[18] See: The Ada Lovelace Institute and The Alan Turing Institute (2023). How do people feel about AI? A nationally representative survey of public attitudes to artificial intelligence in Britain. AI Survey. https://attitudestoai.uk/

Figure 30 'To what extent do you agree or disagree: The people who create AI systems should be held responsible for the real-world impacts of those systems'



There is disagreement within researchers on the responsibilities of AI creators. Researchers that are pessimistic about AI are more likely to believe that the people who create AI systems should be held responsible for their real-world impacts. (58% strongly agree vs 24% for optimists).

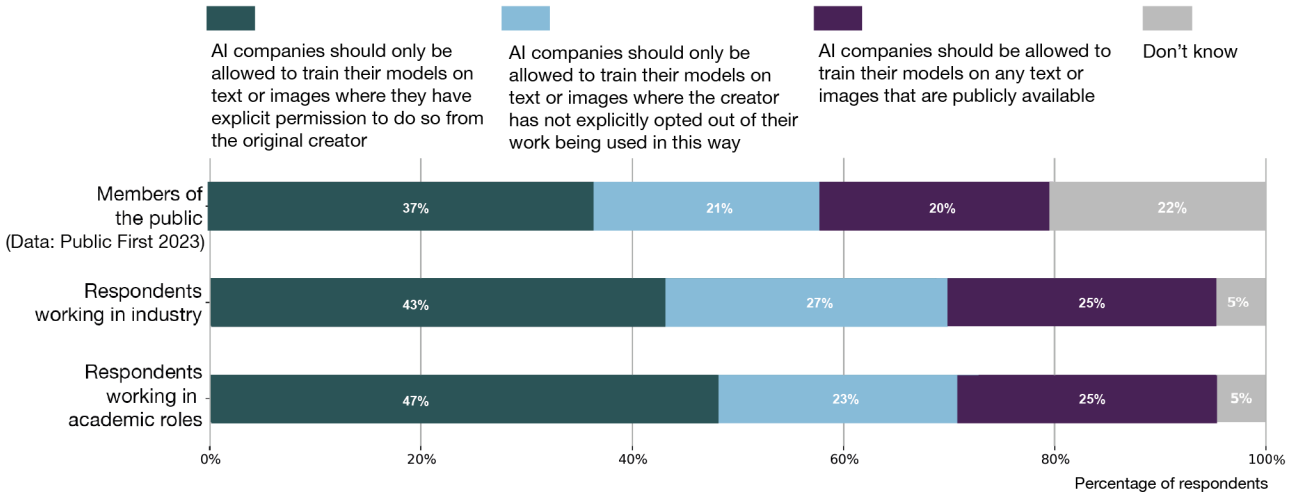
In total, 66% of researchers believe that people who create AI systems should be held responsible for the impacts of those systems. Meanwhile, 12% disagree, and 4% strongly disagree.

Given the conventional division of moral labour in innovation, in which society, users, companies or others are often blamed for the negative downstream consequences of scientific research, this finding might be seen as surprising. But we do not know how many respondents interpreted responsibility in terms of credit rather than blame here.

Insight 11

Researchers want greater care about training data

Figure 31 'Which statement best reflects your views about training data?'



Most respondents think there should be constraints on how training data is acquired. More than 65% of respondents agreed that there should be some sort of constraints, either explicit permission, or opt out systems. Only one in four of the survey's respondents think that AI companies should be allowed to train their models on any publicly available text or images.

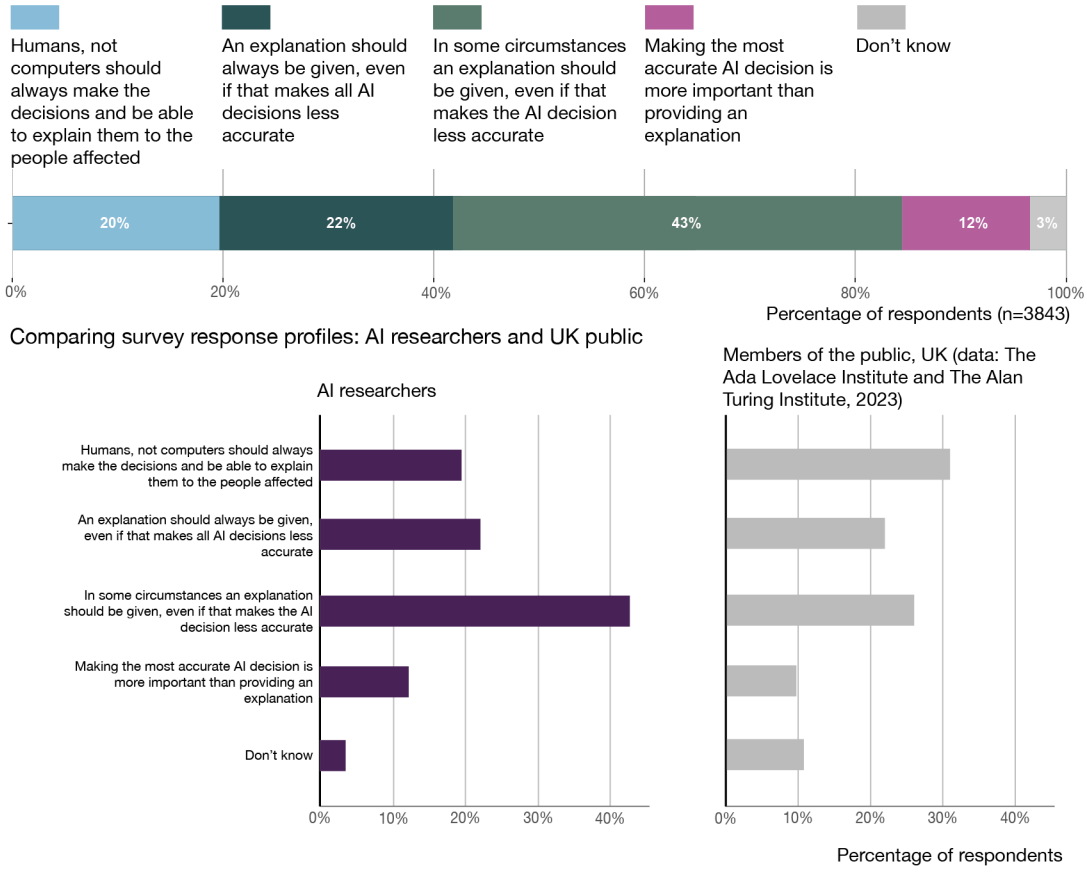
Responses did not vary significantly across industry and academic settings and did not differ markedly from a survey of members of the public in the UK published by Public First in 2023.[19]

[19] Dupont, J., Wride, S., & Ali, V. (2023). What does the public think about AI? Public First. https://publicfirst.co.uk/ai/

Insight 12

Researchers are less concerned than the public are when it comes to explaining AI outputs

Figure 32 'Below are four statements that reflect different opinions toward explaining how AI systems make decisions. Overall, which statement do you feel best reflects your personal opinion?'



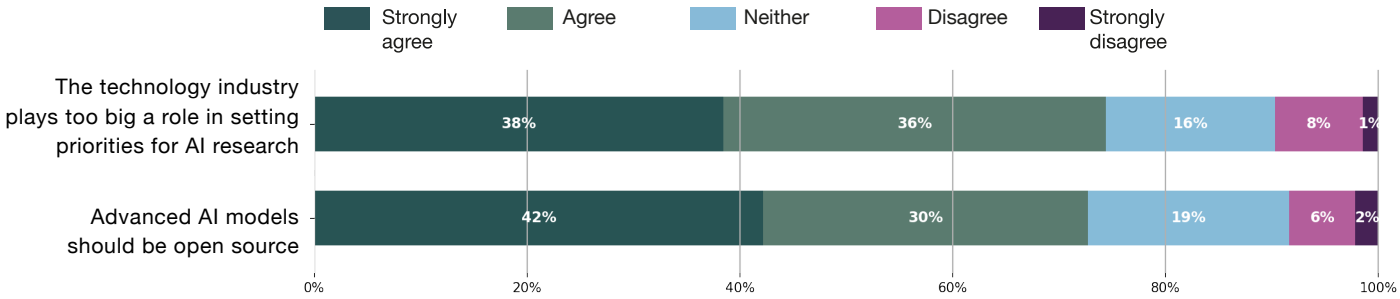
When asked about a tradeoff between explanation and accuracy, 43% of AI researchers agree that "in some circumstances an explanation should be given, even if that makes the AI less accurate". 20% think "humans, not computers, should always make the decisions and be able to explain". 12% think "making the most accurate AI decision is more important than providing an explanation".

AI researchers are less concerned than members of the public about explaining AI decisions: the most popular response for public respondents is that "humans, not computers, should always make the decisions and be able to explain". 10% of public respondents to the 2023 survey by The Ada Lovelace Institute and Alan Turing Institute believe that "making the most accurate AI decision is more important than providing an explanation".

Insight 13

Researchers are concerned about research agendas for AI

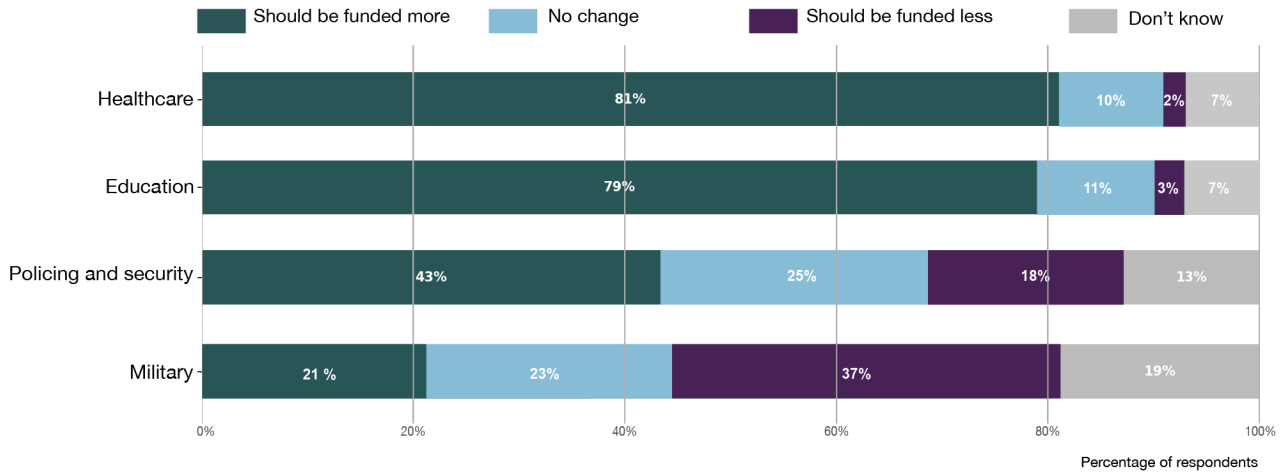
Figure 33 'To what extent do you agree or disagree with the following statements?'



Most researchers agree that the industry plays too big a role in setting priorities for AI research. This effect is less for industry researchers.

Most agree that AI models should be open source, although the proportion is lower for industry researchers.[20]

Figure 34 'How should government funding for AI be changed in these areas?'



A majority of researchers (60%) think that military funding in AI should not increase. Most researchers think education and healthcare should be funded more.

[20] The label 'open source' is contested. See Widder, D. G., Whittaker, M., & West, S. M. (2024). Why 'open'AI systems are actually closed, and why this matters. Nature, 635(8040), 827-833.

4. Final thoughts, future agendas

It is clear that policymakers, journalists and others interested in the public debate on AI should seek more diverse views on AI. There is a wide range of views among those researching AI, but these are currently drowned out by the loud voices of a few powerful people. Some of these people lead AI companies, some are evangelists for the technology and some are self-proclaimed 'experts'. Many of these people have direct financial interests in accelerating current AI trajectories and they are likely to downplay uncertainties about the technology. Our survey results support the hypothesis that 'distance lends enchantment' in AI. There is more uncertainty among the researchers who are closest to the technology. Our survey reveals that, beneath the surface, AI researchers have a range of hopes and fears about the technology that are broader and more complex than the public debate suggests.

While there are gaps between researchers' and public views on AI, it would be a mistake to see these just as gaps in understanding. Public views on the risks and benefits of AI should be taken seriously, rather than viewed as misconceptions that will improve with education or the further development of technologies.

Many AI researchers see the need for public input, but they largely think that this should take place downstream, around the risks, uses and regulations of AI. They do however recognise a range of upstream concerns, such as those to do with the data that feeds AI models, industry control of research agendas and the need for steering of AI research.

We should be careful not to read too much into these results. Further qualitative research is needed to explore AI researchers' ideas of risk, opportunity and responsibility. Even the most carefully-designed survey questions will be interpreted differently by each respondent. Respondents will have understood terms in different ways and, where we have made comparisons between researchers' and public views, it should be noted that researchers may have more technical understandings of issues. We have therefore tried to observe patterns of responses rather than zooming into particular risks, for example.

We can assume some self-selection bias in the survey. People who are interested in responding to questions such as ours are unlikely to be perfectly representative of the community of AI researchers. We are also conscious that professional groups' responses on issues that seem to be about professional virtue may over-emphasise concerns that are seen as in the public interest. Past surveys on scientific fraud, for example, have tried to get around this by asking respondents to talk about their colleagues' views rather than their own.[21] Follow-up surveys could frame questions in this way.

'Responsibility' might be seen by some people in terms of blame and others as taking credit or developing social obligations. AI researchers' sense of what AI is will also vary. Our respondents here include researchers whose work is highly theoretical and others who focus on uses of particular AI systems. There is some evidence here that researchers see a division of moral labour, in which scientists are responsible for

[21] See, for example, Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. PloS one, 4(5), e5738.

research and society is responsible for the uses of technology. But there is also evidence here that researchers do not see AI technologies as morally neutral, nor do they see data as merely a raw material.

We began this project with interests in the possibilities of democratising AI research. Our survey suggests there is a need for further dialogue, an opportunity to engage AI researchers in a debate that both they and the public find important, and some challenges in overcoming AI researchers' sense of public attitudes.

About this report

Lead authors

Cian O'Donovan, Sarp Gurakan, Xiaomeng Wu, Jack Stilgoe

Contributing authors

Nicolas Bert, Ekaterina Dmitrichenko, Embla Gjørva, Stella Liu, Teddy Zamborsky, Tianyu Zhao

Preferred citation

O'Donovan, Gurakan, Wu, Stilgoe, Bert, Dmitrichenko, Gorba, Liu, Zamborsky and Zhao. 2025. Visions, values, voices: A Survey of Artificial Intelligence Researchers. UCL Centre for Responsible Innovation Working Paper 2025/01. London, United Kingdom. https://doi.org/10.5281/zenodo.15080287

Revisions

Version 1.01 - Updated appendix to include all response text options. Revised author name text 'Gjørva'; Revised public data in Figure 31; Revised Box 2 text 'Country in which respondents live'; Revised text in Section 2.3, Box 1, Figure 9 to reflect where respondents 'live' rather than 'work'. Amended Figure 15 to include 'policymakers' responses.

To find out more, email the corresponding authors

Cian O'Donovan cian.o@ucl.ac.uk; Jack Stilgoe j.stilgoe@ucl.ac.uk

Funding

The research reported here was undertaken as part of **Public Voices in AI**, a satellite project funded by Responsible AI UK and EPSRC (Grant number: EP/Y009800/1). Public Voices in AI was a collaboration between: the ESRC Digital Good Network at the University of Sheffield (Grant number: ES/X502352/1), Elgon Social Research Limited, Ada Lovelace Institute, The Alan Turing Institute and University College London.

The project also received support from UCL's Department of Science and Technology Studies, The UCL Centre for Responsible Innovation, the Economic and Social Research Council's Metascience Programme, and UCL's UKRI Impact Acceleration Accounts.

About Public Voices in AI

Public Voices in AI was a year-long (2024-25) research project that aimed to ensure that public voices are attended to in artificial intelligence (AI) research, development, deployment and policy ('AI RDD&P'). It synthesised, reviewed, built and shared knowledge about public views on AI and engaging diverse publics in AI RDD&P, with and in consultation with target beneficiaries working in (responsible) AI and members of the public, especially from groups most negatively affected by and underrepresented in AI.

Public Voices in AI aimed to achieve the following:

1. Increase understanding of the value of meaningful inclusion of public voice amongst beneficiary groups and increased methodological capacity to do so.
2. Increase understanding of public views and experiences of AI amongst beneficiaries, especially of how underrepresented groups are differentially impacted by AI and the subsequent need for equity-driven approaches.
3. Demonstrate good practice in how to engage underrepresented communities in developing, co-designing and producing AI public voice research.
4. Ensure that people from underrepresented groups participate in AI RDD&P and shape this project.
5. Better enable public voices to inform AI RDD&P.

A central aspect of responsible AI is ensuring that it takes account of public hopes, concerns and experiences. As concern about the societal impacts of AI grows and pressure for its effective regulation mounts, understanding and anticipating societal needs and values can inform responsible AI developments and deployments. Yet public voice is frequently missing from conversations about AI, an absence which inhibits progress in responsible AI. Addressing this gap is essential to enable 'good AI' – that is, AI which maximises benefits, prevent harms and works for everyone.

However, 'public voice' is not easy to define, as there is no one 'public'. Different groups benefit from and are affected by AI differently, and their hopes, concerns and experiences also vary. Centuries of structural inequities and overlapping systems of oppression (eg racism, sexism, ableism, colonialism, transphobia, classism) mean that some groups have more resources and access to power to shape AI technologies than others. There is a related participation gap between those with the social capital to participate in shaping AI and those without. AI public voice and public participation initiatives therefore need to centre those most impacted and underrepresented. Public Voices in AI did that.

Public Voices in AI consisted of:

- A review of existing evidence about public thoughts and feelings about AI, which asked who is included and who is excluded from existing evidence, and how this impacts knowledge;
- A number of participatory projects with people from disadvantaged, marginalised, minoritised communities;
- A survey of public attitudes to specific uses of AI, with boosts of some disadvantages/minoritised groups;
- A survey of AI researchers' attitudes to AI benefits and harms, to including publics in their work, and to public attitudes;
- The production of resources to support the inclusion of public voices in AI RD&P;
- The production of a framework which seeks to encourage practitioners, policymakers, developers and the public to design, evaluate, regulate, and operate AI systems in ways that benefit people, society and the nation;
- The production of a self-assessment workbook which practitioners and technologists can use to critically evaluate and enhance their approaches to including public voices in AI.

Public Voices in AI team

- **Helen Kennedy**, Professor of Digital Society, University of Sheffield. Director, Digital Good Network
- **Ros Williams**, Senior Lecturer in Digital Media and Society, University of Sheffield. Associate Director, Digital Good Network
- **Susan Oman**, AI and In/equalities Lead, Centre for Machine Learning & Senior Researcher at the Digital Good Network, University of Sheffield
- **Helen Margetts**, Director, Public Policy Programme, Alan Turing Institute for Data Science & AI
- **Octavia Field Reid**, Associate Director (Impact & research practice), Ada Lovelace Institute.
- **Jack Stilgoe**, Professor of Science and Technology Policy, Dept. of Science & Technology Studies, University College London.
- **Eleanor O'Keeffe**, Public Participation & Research Practice Lead, Ada Lovelace Institute.
- **Roshni Modhvadia**, Researcher, Ada Lovelace Institute
- **Mhairi Aitken**, Ethics Fellow, Alan Turing Institute.
- **Cian O'Donovan**, Director, UCL Centre for Responsible Innovation. Dept. of Science and Technology Studies, University College London.
- **Tvesha Sippy**, Researcher, Alan Turing Institute.
- **Sara Cannizzaro**, Postdoctoral Researcher, Public Voices in AI project
- **Ruth Lauener**, Manager, Digital Good Network, University of Sheffield
- **Sarah Givans**, Research Support Project Administrator, Digital Good Network, University of Sheffield

Appendix: Survey questions

1.1. To what extent does your job involve AI?
- Mostly
- Somewhat
- Slightly
- Not at all

1.2. How would you describe your job or role?
- [open text]

1.3. Which of the following best describes where you work?
- Company or private sector
- University / research organisation
- Government or public sector
- Other [open text]

1.4. What do you think qualifies someone as an expert in artificial intelligence? (click all that apply)
- PhD in Computer Science / AI
- Relevant academic publications in science or engineering
- Relevant academic publications in other disciplines (e.g. social science, philosophy, economics, law)
- A professional role with AI-related responsibilities
- A journalist specialising in AI
- An entrepreneur in an AI startup
- An AI policy maker or regulator
- Experience of working with AI
- Work in an AI-related think tank or research organisation
- Other (please provide an example) [open text]

1.5. How often do you think you can recognise when you are using artificial intelligence?
- Often or always
- Some of the time
- Occasionally
- Hardly ever
- Never

1.6. How often do you think members of the public can recognise when they are using artificial intelligence?
- Often or always
- Some of the time
- Occasionally
- Hardly ever
- Never

1.7. Think about an AI system that you are most familiar with. To what extent do you understand how it works?
- I fully understand how it works
- Mostly. I have a good grasp of how it works
- Somewhat. I understand the basics but not the details
- A little. I only know a bit about how it works
- I do not understand how it works

1.8. To what extent do you agree or disagree with the following statements? *Response options: Strongly agree; agree; neither agree nor disagree; disagree; strongly disagree.*
- AI systems should be designed to reflect human values
- The people who create AI systems should be held responsible for the real-world impacts of those systems
- The best-placed people to understand the risks of AI are computer scientists
- The best-placed people to understand the risks of AI are members of the public who might be affected
- Artificial general intelligence (AGI) is inevitable

2.1. What one thing most excites you about AI?
- [open text]

2.2. What one thing most worries you about AI?
- [open text]

2.3. Which statement best reflects your view about the benefits and risks of AI?
- I think there are more benefits than risks
- I think there are equal benefits and risks
- I think there are more risks than benefits

2.4. In which ways, if any, do you think artificial intelligence could have a negative impact on people's lives? (click on all that apply)
- Use of personal data without consent
- Make it difficult to tell whether news or information are fake
- Increase chances of experiencing cyber crime
- Worsen experience of customer service
- Reduce social interactions
- Put jobs at risk
- Take over tasks people enjoy doing
- Reduce incomes
- Other [open text]
- I don't think AI could have a negative impact on people's lives

2.5. In which ways, if any, do you think artificial intelligence could have a positive impact on people's lives? (click on all that apply)
- Improve access to healthcare
- Improve shopping experiences
- Increase access to learning or education

- Make household tasks easier
- Make jobs easier
- Improve personal safety
- Reduce working hours without reducing the pay
- Improve job prospects
- Increase incomes
- Other [open text]
- I don't think AI could have a positive impact on people's lives

2.6. How do you think members of the public see the benefits and risks of artificial intelligence (AI)?
- I think they see more benefits than risks
- I think they see equal benefits and risks
- I think they see more risks than benefits
- Don't know

3.1. To what extent do you agree or disagree with the following statements? *Response options: Strongly agree; agree; neither agree nor disagree; disagree; strongly disagree.*
- Advanced AI models should be open-source
- AI should be developed as quickly as possible
- The technology industry plays too big a role in setting priorities for AI research

3.2. Below are four statements that reflect different opinions toward explaining how AI systems make decisions. Overall, which statement do you feel best reflects your personal opinion?
- Making the most accurate AI decision is more important than providing an explanation
- In some circumstances an explanation should be given, even if that makes the AI decision less accurate
- An explanation should always be given, even if that makes all AI decisions less accurate.
- Humans not computers should always make the decisions and be able to explain them to the people affected
- Don't know

3.3. Who do you think should be most responsible for ensuring AI is used safely? (Choose up to two options)
- The companies developing the AI technology
- International standards bodies
- The people using the AI (e.g. companies, public services)
- Scientists and researchers working in universities or research institutions
- The Government
- No one should be responsible for ensuring AI is used safely
- An independent oversight committee with citizen involvement
- An independent regulator
- Don't know
- Other [open text]

3.4. Which statement best reflects your views about training data?
- AI companies should only be allowed to train their models on text or images where they have explicit permission to do so from the original creator
- AI companies should only be allowed to train their models on text or images where the creator has not explicitly opted out of their work being used in this
- AI companies should be allowed to train their models on any text or images that are publicly available
- Don't know

3.5. How should government funding for AI be changed in these areas? *Response options: Should be funded more; no change; should be funded less; don't know.*
- Military
- Healthcare
- Policing and security
- Education

4.1. How important do you think it is for members of the public to know about the following? *Response options: Extremely important; very important; somewhat important; not important.*
- How AI in general works
- Details of specific AI applications
- The benefits of AI
- The risks of AI
- The companies developing AI

4.2. To what extent do you agree or disagree with the following statements? *Response options: Strongly agree; agree; neither agree nor disagree; disagree; strongly disagree.*
- If the public understood AI better, they would trust it more
- Most people do not understand the potential benefits of AI
- Most people do not understand the potential risks of AI
- Policymakers need to know more about AI
- Members of the public worry too much about the risks of AI
- The public are able to weigh up the risks and benefits of AI

4.3. How do you normally find out what the public think about AI? (click all that apply)
- Conversations with members of the public
- Social research on public attitudes
- Discussions on social media

- Stories in the media
- Other [text]

4.4. How important do you think it is to involve the public in decisions about the following tasks? *Response options: Extremely important; very important; somewhat important; not important.*
- Training AI models
- Developing AI
- Deploying AI
- Assessing the impacts of AI after deployment
- Regulating AI

4.5. Why do you think it might be important to involve the public? (click all that apply)
- To build trust
- To avoid bias
- It makes technology better
- It makes technology safer
- To inform regulations
- I don't think it's important

4.6. Do you have any knowledge or experience of the following methods for public involvement? *Response options: I've used this method; I'm aware of this method and think the results are important; I don't think this method is valuable; I don't know what this is.*
- Surveys
- Focus groups
- Citizens' juries and other deliberative processes
- Public participation in standards setting and regulation
- Reinforcement learning with human feedback
- Other [open text]

4.7. What do you think are the barriers to involving the public in AI? (click all that apply)
- Lack of understanding of AI among the public
- Lack of understanding among AI developers about public involvement methods
- Lack of time
- Lack of funding
- It is still too early. We should wait until we know the risks and benefits
- Public experiences and attitudes aren't valued by AI developers
- Other [open text]

4.8. To what extent do you agree with the following statements? *Response options: Strongly agree; agree; neither agree nor disagree; disagree; strongly disagree.*
- I worry that some members of the public will benefit from AI more than others
- I worry that some members of the public will be more negatively impacted by AI than others
- It is important that developers consider how AI impacts different groups
- It is important that regulators consider how AI impacts different groups

5.1. How old are you?
- 18-24 years old
- 25-34 years old
- 35-44 years old
- 45-54 years old
- 55-64 years old
- 65+ years old

5.2. How do you describe yourself?
- Male
- Female
- Non-binary / third gender
- Prefer to self-describe [open text]
- Prefer not to say

5.3. In which country do you currently live?
- [Dropdown list of countries]

5.4. What is your highest education qualification?
- Secondary school / high school diploma
- Some university but no degree
- University degree (e.g. BA, BS, BSc)
- Graduate degree (e.g. MA, MS, MSc, MPhil)
- Professional or medical degree (e.g. MD, MBA, JD)
- Doctoral degree (e.g. PhD)
- Professional qualification
- Other [open text]

UCL Centre for Responsible Innovation
Working paper 2025/01

This is the first publication from the UCL Centre for Responsible Innovation.
Version 1.01 March 2025.

Centre for
Responsible
Innovation