



Cian O'Donovan

June 19th, 2025

Most AI researchers reject free use of public data to train AI models

3 comments | 7 shares

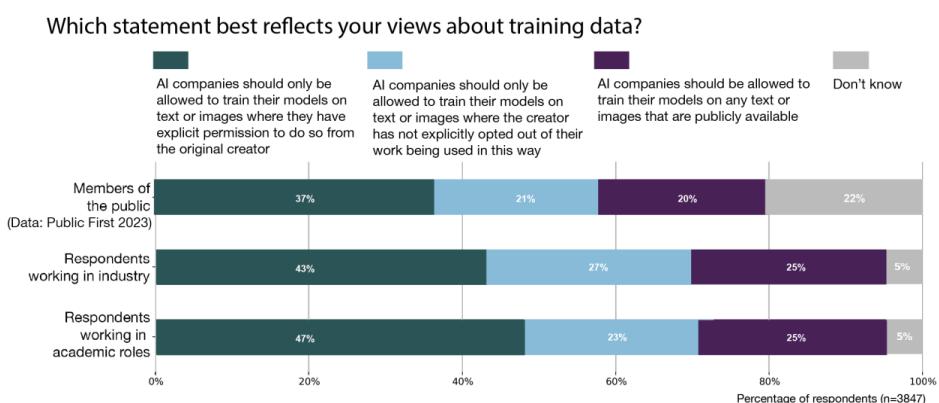
Estimated reading time: 10 minutes

A new survey finds only a quarter of AI researchers support unrestricted AI model training using publicly available data. As creatives protest unauthorised use of their work, tensions mount between government policy, industry ambitions and public expectations. Cian O'Donovan explores what researchers think, where they align with the public and where critical gaps remain.

There are growing concerns among writers and academics about the

use of their work in AI development **without consent**. The **Atlantic magazine** has covered that engineers at Meta had downloaded more than 7 million books and 80 million research papers from vast pirated libraries online. A recent petition by the **Society of Authors**, representing over 11,000 members, highlights the call for transparency and fair use practices in AI training.

New **research from UCL** reveals that many AI researchers support stronger ethical standards for training data. A global survey found that only one in four respondents believe AI companies should be allowed to train their models on any publicly available text or images. The majority favour more restrictive approaches, including the requirement for explicit permission from content creators.



This result is important because it shows that most researchers disagree with the UK government's **current proposals** for training data. The legislation would place the responsibility governing the use of creative data back on creatives to explicitly "opt out" of their work being used in this way. This runs contrary to decades of copyright norms in the industry.

This debate over copyright shows us that despite intense hype about the opportunities AI offers, AI systems and the firms that run them have not yet achieved a broad licence to operate in society. Any such social

licence will depend on satisfying the needs and norms of people in diverse sectors such as our creative industry.

What AI researchers think about training data

In June 2024, I was part of a team that fielded the largest international survey of AI researchers to date to try to understand more about what AI researchers themselves thought about a range of issues on this topic.

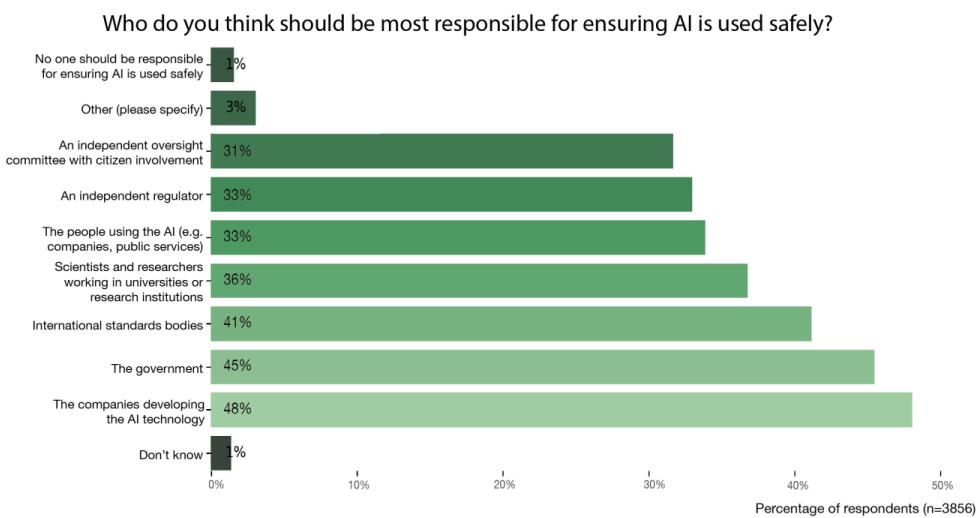
The conversation about the benefits of AI is dominated by a small number of industry voices who tend to emphasise their own sectoral interests. The UK's *AI Opportunities Action Plan*, for instance, written by industry insider Matt Clifford, is ebullient on the benefits of AI. But according to **Gaia Marcus**, Director of the Ada Lovelace Institute, the plan so far neglects "a credible vehicle and roadmap for addressing broader AI harms".

We wanted to know how those closest to developing AI systems understand public hopes and fears and where the gaps and overlaps between public and AI researchers' attitudes emerge.

Analysis of 4,260 responses from published AI researchers shows that respondents have diverse and divergent views about innovation and responsibilities in AI. These researchers and the public agree on AI's risks but disagree about who should be responsible for its safe use. And while researchers told us they want AI to reflect people's values, we found that most of our respondents do not pay attention to social science research that might tell them more about those values.

Given these disagreement between government, tech leaders and

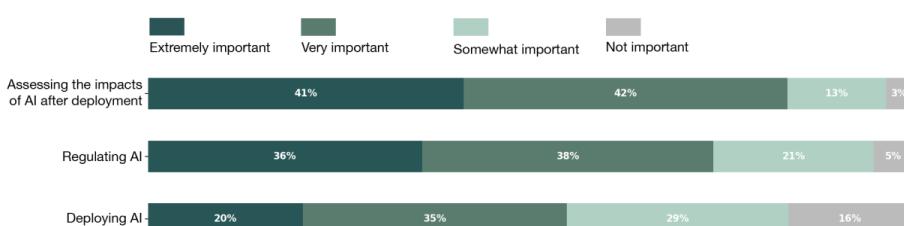
members of the public and over how to govern AI, the question of who should be responsible for ensuring that AI is used safely gains importance. For researchers, the top three responses when asked who should be most responsible for safe use of AI were companies developing AI, the government and international standards bodies.



Meanwhile, when surveyed in a **prior study** by the Alan Turing Institute and Ada Lovelace Institute, a representative sample of the UK public listed different priorities. They said that those most responsible are: (1) an independent regulator, (2) the companies developing AI and (3) an independent oversight committee with citizen involvement.

We also asked about involving the public in AI research processes. Researchers believe the public should be involved in decisions as, or after, AI is deployed. In the survey, 84% said this was at least somewhat important. But fewer researchers want members of the public involved in decisions before deployment, training models or in developing AI.

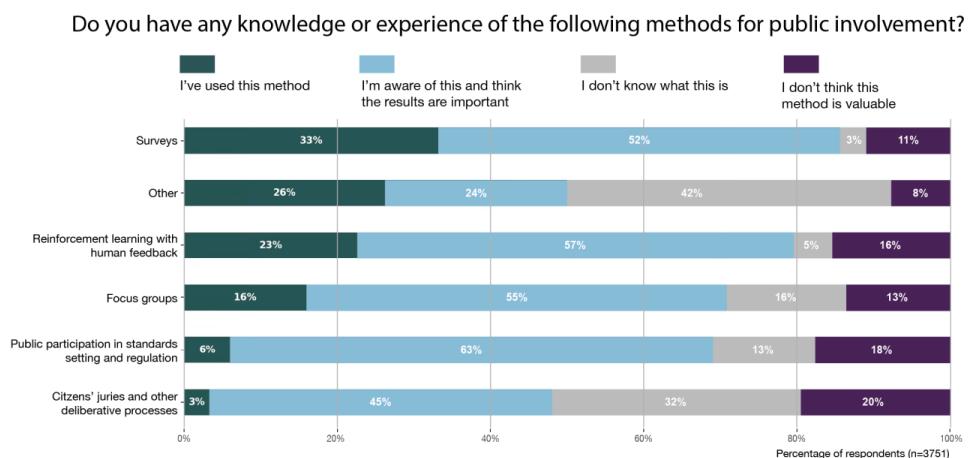
How important do you think it is to involve the public in decisions about the following tasks?





Many AI researchers see the need for public input, but more respondents than not think that this should take place when it is deployed in the world, around the risks, uses and regulations of AI.

Analysis also shows that relatively few researchers have used methods that involve the public in AI innovation. Where respondents do have experience, it is mostly with “low participation” methods such as surveys. The use of techniques that empower the public to help design standards or open up research to deliberative processes are reported by far fewer respondents.



Researchers do, however, recognise a range of upstream concerns about the data that feeds AI models, or about industry control of research agendas and the need for steering AI research.

An agenda for AI research policy

While there are gaps between researchers' and the views of authors, it would be a mistake to see these only as gaps in understanding. Song

writer and surviving Beatle Paul McCartney's comments to the BBC are a case in point: "I think AI is great, and it can do lots of great things," **McCartney told Laura Kuensberg**, "but it shouldn't rip creative people off."

It's clear that McCartney gets the opportunities AI offers. For instance, he used AI to help bring to life the voice of former bandmate John Lennon in a recent single. But like the writers protesting outside of Meta's office, he has a clear take on what AI is doing wrong and who should be responsible. These views and the views of over members of the public should be taken seriously, rather than viewed as misconceptions that will improve with education or the further development of technologies.

Increasing dialogue and deliberation between diverse public groups and developers across the AI landscape can help. However, almost 40% of researchers agreed that barriers to involving the public included a lack of time and a lack of funding. The good news is that these kinds of barriers can be addressed with finance, time and resource commitments from bodies that fund and train AI researchers. **Research coming out of Public Voices in AI**, the project that funded our survey, show how that can be achieved.

We should of course be careful not to read too much into these results. Even the most carefully designed survey questions will be interpreted differently by each respondent. Further qualitative research is needed to explore how AI researchers think about risk, opportunity, responsibility and AI itself.

Our survey was designed to test researcher interests in the possibilities of democratising AI research. The results suggest 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.

For more on the report mentioned above see: <https://doi.org/10.5281/zenodo.15080287>.

The research reported here was undertaken as part of Public Voices in AI, a 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 content generated on this blog is for information purposes only. This Article gives the views and opinions of the authors and does not reflect the views and opinions of the Impact of Social Science blog (the blog), nor of the London School of Economics and Political Science. Please review our [comments policy](#) if you have any concerns on posting a comment below.

Image Credit: [PJ McDonnell](#) on Shutterstock.

 Print Friendly

About the author

**Cian O'Donovan**

Dr Cian O'Donovan is Director of the UCL Centre for Responsible Innovation. Cian's work uses social science methods to tell stories about who benefits from technologies like AI, robotics and digital systems. In 2014, he co-founded Uplift, Ireland's largest people-powered digital advocacy organisation for progressive social change.

Posted In: AI Data and Society

3 Comments

Related Posts

[AI Data and Society](#)

AI is transforming education into a collaborative interaction between humans and machines

MAY 7TH, 2025

5 

AI Data and Society

What happens when you let ChatGPT assess impact case studies?

JANUARY 15TH, 2025

2 

AI Data and Society

Does AI have a copyright problem?

MAY 6TH, 2025

6 

AI Data and Society

Can AI make research more open?

APRIL 4TH, 2025

2 



© LSE 2026