Ada Lovelace Institute

Discussion paper

Going public

Exploring public participation in commercial AI labs

December 2023

Emerging technology & industry practice

Contents

Executive summary

Major commercial AI companies are increasingly calling for more public participation in the development, deployment and governance of AI systems. According to their public statements, these companies believe the use of public participation methods can ensure these technologies are built according to the needs and values of people and society.[1] Some of these companies have gone so far as to hire dedicated roles to design and execute public participation projects.

Public participation approaches have a long history of use in public policymaking, citizen science projects and other fields. However, there is limited evidence around the kinds of public participation methods commercial AI labs are adopting, what objectives or goals they are using them for and what challenges practitioners in these labs face when using these methods.

Given the growing prominence of involving the public to make AI systems more accountable, this research project sets out to uncover how commercial AI labs are using public participation methods. This paper builds on Ada's existing work exploring AI accountability practices, which includes identifying and building the evidence base for tools, methods and approaches that enable scrutiny and oversight over AI technologies and the institutions developing and deploying them.[2]

It also builds on Ada's work in public participation methods,[3] which seeks to ensure the perspectives of people affected by data and AI are meaningfully embedded in shaping evidence, research, policy and practices related to data and AI.

1 OpenAI, 'How Should AI Systems Behave, and Who Should Decide?' (17 February 2023) https://openai.com/blog/how-should-ai-systems-behave accessed 20 April 2023.

2 Lara Groves and others, 'Algorithmic Impact Assessment: A Case Study in Healthcare' (Ada Lovelace Institute 2022) https://www.adalovelaceinstitute.org/report/algorithmic-impact-assessment-case-study-healthcare/ accessed 19 April 2022.

3 Ada Lovelace Institute, 'What do the public think about AI?' (2023) https://www.adalovelaceinstitute.org/evidence-review/what-do-the-public-think-about-ai//

We do not prescribe specific policy recommendations but this report highlights several areas that should be further investigated

We had three research questions we set out to uncover in this study:

1. How do commercial AI labs understand public participation in the development of their products and research?
2. What approaches to public participation do commercial AI labs adopt?
3. What obstacles and challenges do commercial AI labs face when implementing these approaches?

By conducting interviews with 12 industry practitioners and public engagement experts, our report surfaced five key findings:

1. Within commercial AI labs, researchers and teams using public participation methods view them as **a mechanism to ensure their technologies are beneficial for people and society**, and a way to support the mission and objectives of their organisation.

2. Our interviews with different practitioners revealed **a lack of consistent terminology** to describe public participation methods and **a lack of any consistent standards** for how to employ these methods.

3. Ultimately, **industry practitioners are not widely or consistently using public participation methods in their day-to-day work**. These methods tend to be deployed on an ad-hoc basis.

4. Industry practitioners **face multiple obstacles** to successfully employing public participation methods in commercial AI labs. These include resource intensity, misaligned incentives with management and teams, practitioners feeling siloed off from product or research teams, and commercial sensitivities constraining practitioner behaviour.

5. How public participation methods are used **in the foundation model (also known as 'general-purpose AI' or 'GPAI') supply chain requires further research**. It is challenging to adopt public participation in contexts that lack a clear use case, presenting implications for foundation models or generative AI systems and research.

This report discusses the opportunities and limitations for public participation in AI

Owing to the exploratory nature of this research project, and without definitive answers from our research participants about a course for further action, we do not prescribe specific policy recommendations for public participation in commercial AI companies. However, Ada's public participation research has set out frameworks and standards for meaningful public participation in data and AI, which we refer to in this report. For policymakers and members of the technology industry who wish to use participatory methods, this report highlights several areas that should be further investigated:

- Further trialling and testing of public participation approaches in industry 'in the open'.
- Collaborative development of standards of practice for public participation in commercial AI.
- Additional research into how public participation might complement other algorithm accountability methods or emerging regulation of AI.

This report offers an important 'link in the chain' to further understanding the opportunities and limitations for public participation in AI, and we hope the findings will shape and influence industry practice in this area.

This research will be of interest to industry practitioners working on issues of ethical AI, and to academic or civil society researchers with a background in public participation or community action interested in how participation might be adopted in commercial AI contexts. This research will also be of interest to policymakers or regulators, who might be curious about the role of participatory approaches as an accountability mechanism.

Glossary

Our report contains a brief discussion on the development and adoption of **foundation models** and **generative AI** systems. We reproduce the following definitions from the Ada Lovelace Institute explainer on foundation models:[4]

Foundation models

Also known as 'general-purpose AI' (or 'GPAI'), foundation models are AI models designed to produce a wide and general variety of outputs. They are capable of a range of possible tasks and applications, such as text, image or audio generation. They can be standalone systems or can be used as a 'base' for many other applications.[5]

Researchers have suggested the 'general' definition refers to foundation models' scope of ability, range of uses, breadth of tasks or types of output.[6]

Some foundation models are capable of taking inputs in a single 'modality' – such as text – while others are 'multimodal' and are capable taking multiple modalities of input at once (for example, text, image, video, etc.) and then generating multiple types of output (such as generating images, summarising text or answering questions) based on those inputs.

Generative AI

As suggested by the name, generative AI refers to AI systems that can generate content based on user inputs such as text prompts. The content types (also known as modalities) that can be generated include images, video, text and audio.

4 'Explainer: What Is a Foundation Model? | Ada Lovelace Institute' https://www.adalovelaceinstitute.org/resource/foundation-models-explainer/ accessed 17 July 2023.

5 Sabrina Küspert, Nicolas Moës and Connor Dunlop, 'The Value Chain of General-Purpose AI' (Ada Lovelace Institute Blog, 10 February 2023) https://www.adalovelaceinstitute.org/blog/value-chain-general-purpose-ai/ accessed 27 March 2023.

6 Philipp Hacker, Andreas Engel and Marco Mauer, 'Regulating ChatGPT and Other Large Generative AI Models' (arXiv, 12 May 2023) http://arxiv.org/abs/2302.02337 accessed 24 July 2023.

Like foundation models, forms of generative AI can be unimodal or multimodal. It is important to note that not all generative AI systems are foundation models. Unlike foundation models, generative AI can be narrowly designed for a specific purpose. Some generative AI applications have been built on top of foundation models, such as OpenAI's DALL·E or Midjourney, which use natural language text prompts to generate images.

Generative AI capabilities include text manipulation and analysis, as well as image, video and speech generation. Generative AI applications include chatbots, photo and video filters, and virtual assistants.

How to read this report

If you're an industry practitioner who is interested in or currently using participatory AI methods, and are interested in understanding how others in industry are experiencing these practices, you should read the findings on pages 35–51 and 'Areas for further input' on pages 52-61.

If you're an academic or civil society researcher interested in public participation methods, you should go to page 36 to understand how industry practitioners are conceptualising 'participatory AI' and the associated practices, and 'Areas for further input' on page 52, which presents considerations for whether and how civil society might engage with commercial AI labs on public participation projects.

If you're a policymaker or a regulator interested in what participatory approaches to AI might involve, go to page 36 of the findings and 'Areas for further input' on page 52.

Introduction

'[Our ambition is] to responsibly advance cutting-edge AI research and democratize AI as a new technology platform'[7]
Satya Nadella, CEO of Microsoft

'One way to avoid undue concentration of power is to give people who use or are affected by systems like ChatGPT the ability to influence those systems' rules'[8]
OpenAI

As AI systems become more accessible to everyday people and more impactful on their everyday lives, there is growing concern among some members of the public and global policymakers that these technologies are not designed for their benefit. In the wake of the release of powerful products like ChatGPT, some policymakers and technology companies have made calls to 'democratise AI' and use more 'participatory' methods to involve everyday people in the process of developing, deploying and governing AI. In May 2023, the commercial AI lab OpenAI launched an initiative seeking proposals for a democratic process to 'decide what rules AI systems should follow.'[9] That same month, the UK Government called for more public involvement in AI policy.[10]

What are we to make of these calls for 'democratisation' and 'participation' in AI? Both 'participation' and 'democracy' are capacious terms that may indicate a variety of different meanings. Broadly, they refer to the involvement of members of the public in decisions around how and where AI technologies are designed, governed, accessed and/

7 Microsoft Corporate Blogs, 'Microsoft and OpenAI Extend Partnership' (The Official Microsoft Blog, 23 January 2023) https://blogs.microsoft.com/blog/2023/01/23/microsoftandopenaiextendpartnership/ accessed 14 August 2023.
8 OpenAI (n 1).
9 'Democratic Inputs to AI' https://openai.com/blog/democratic-inputs-to-ai#fn-A accessed 31 May 2023.
10 Department for Science, Innovation & Technology and Office for Artificial Intelligence, 'A Pro-Innovation Approach to AI Regulation' (2023) https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper>accessed 16 June 2023.

or deployed.[11] They may allude to normative goals and have connotations of legitimacy, political agency or civic life, which may prove attractive to commercial companies looking to regain a positive public reputation, or to enable more corporate social responsibility.

Research into public participation has shown 'democratisation' and 'participation' are contested terms with different meanings that imply radically different goals.[12,13] In addition to the conceptual confusion around what 'participation' in AI means, there is a lack of empirical evidence exploring the practical adoption of public participation methods in the development of AI systems.

> There is a particularly notable lack of evidence of how public participation methods can be used in the sites driving major AI developments: commercial AI labs.

Commercial AI labs, as the centres driving AI innovation in commercial technology companies, wield significant influence in determining the trajectory of AI research and development.[14] This influence is also felt in conversations about AI oversight and ethics, where industry has begun to increase its visibility and influence in AI ethics journals and conferences[15] (many of which rely on industry funding to operate) or visibility at key AI policy convenings.[16] Industry's outsized influence in the field of 'ethical AI' has implications for the direction of 'participatory AI' discussions and

11 Elizabeth Seger and others, 'Democratising AI: Multiple Meanings, Goals, and Methods' (arXiv, 27 March 2023) http://arxiv.org/abs/2303.12642 accessed 26 April 2023.

12 Fernando Delgado and others, 'Stakeholder Participation in AI: Beyond "Add Diverse Stakeholders and Stir"' [2021] arXiv:2111.01122 [cs] http://arxiv.org/abs/2111.01122 accessed 28 April 2022.

13 Abeba Birhane and others, 'Power to the People? Opportunities and Challenges for Participatory AI' (15 September 2022) http://arxiv.org/abs/2209.07572 accessed 20 September 2022."plainCitation":"Abeba Birhane and others, 'Power to the People? Opportunities and Challenges for Participatory AI' (15 September 2022

14 Meredith Whittaker, 'The Steep Cost of Capture' (2021) https://papers.ssrn.com/abstract=4135581 accessed 23 January 2023."plainCitation":"Meredith Whittaker, 'The Steep Cost of Capture' (2021

15 'The 2022 AI Index: Industrialization of AI and Mounting Ethical Concerns' (Stanford HAI) https://hai.stanford.edu/news/2022-ai-index-industrialization-ai-and-mounting-ethical-concerns accessed 10 July 2023.

16 The White House, 'FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI' (The White House, 21 July 2023) https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/ accessed 25 July 2023.

It is important to explore whether the use of public participation methods can help ensure AI systems reflect the needs of people impacted by these technologies

may preclude civil society organisations, activists or members of the public from having a stake in the conversation.

Calls from major AI labs for increased public participation in AI have come at a time when the technology sector has faced a gloomy economic climate. Major players like Microsoft and Google have begun to speed up already fast-paced product development cycles as part of an 'arms race'[17] to deploy their technologies first.

> While calling for democratising AI, many of the biggest technology companies have also made mass redundancies to internal teams that are focused on 'responsible AI' and ethics initiatives – the precise teams that would presumably undertake public participation work.

In March 2023, Microsoft laid off its entire ethics and society team, the team with remit to ensure ethical principles are translated into product design.[18] Similarly, Meta dissolved their responsible innovation team in September 2022. Signalling from technology companies about their supposed ethical commitments also comes at a time when many of the largest and most powerful companies are engaging in widespread corporate lobbying efforts aimed at influencing emerging AI regulation in Europe, North America and other regions.[19]

These trends suggest it is reasonable to be wary of how commercial AI labs are using public participation methods. However, it is equally important to explore whether the use of these methods in commercial AI labs can help ensure AI systems are designed, deployed and governed in a way that reflects the needs of people impacted by these technologies. In this research project, we examine the aims and objectives commercial

17 Chris Stokel-Walker, 'TechScape: Google and Microsoft Are in an AI Arms Race – Who Wins Could Change How We Use the Internet' The Guardian (21 February 2023) https://www.theguardian.com/technology/2023/feb/21/techscape-google-bard-microsoft-big-ai-search accessed 10 July 2023.

18 Casey Newton, 'Microsoft Lays off Team That Taught Employees How to Make AI Tools Responsibly' (The Verge, 14 March 2023) https://www.theverge.com/2023/3/13/23638823/microsoft-ethics-society-team-responsible-ai-layoffs accessed 10 July 2023.

19 'Exclusive: OpenAI Lobbied E.U. to Water Down AI Regulation | Time' https://time.com/6288245/openai-eu-lobbying-ai-act/ accessed 10 July 2023.

This report builds on previous Ada work exploring the role of meaningful public participation approaches in data and AI policy and governance

AI labs have for experimenting with public participation approaches and explore what methods for public participation they may employ. This report seeks to answer three questions:

1. How do commercial AI labs understand public participation in the development of their products and research?
2. What approaches to public participation do commercial AI labs adopt?
3. What obstacles and challenges do commercial AI labs face when implementing these approaches?

As part of this research, we conducted nine interviews with practitioners working in commercial AI labs and involved in planning and delivering public participation experimentations or taking forward their findings. We also conducted three background interviews with public participation experts with knowledge of or experience working in the technology industry, with a view to empirically exploring on-the-ground practice.

This report was originally published as an academic paper at the 2023 Association of Computing Machinery (ACM)'s Fairness, Accountability and Transparency in Machine Learning (FAccT) conference,[20] and has been recreated here as a longer, policy-facing report.

This report builds on previous work the Ada Lovelace Institute has conducted to explore the role of meaningful public participation approaches in data and AI policy and governance. In our *Rethinking data* report, [21] we outline a vision for ensuring public participation in technology policymaking, which is that everyone who wishes to participate in decisions about data and data governance can do so. The report sets out possible approaches to get there, including democratic deliberative mechanisms and participatory co-design projects. Our 2023 evidence review, 'What do the public think about AI?', synthesises public attitudes toward AI from a range of studies, and finds that the public want to have a meaningful say in decisions related to data and AI and explores the value of different methods of engagement.[22] Finally, our *Participatory*

20 Groves L and others, 'Going Public: The Role of Public Participation Approaches in Commercial AI Labs', Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (Association for Computing Machinery 2023) https://doi.org/10.1145/3593013.3594071 accessed 6 December 2023.

21 Ada Lovelace Institute, Rethinking data and rebalancing digital power (2022) https://www.adalovelaceinstitute.org/project/rethinking-data/ accessed 3 February 2023.

22 Ada Lovelace institute (n 3).

This report aims to shape emerging policy and practice debates around public participation and help answer a burning question facing policymakers, industry partitioners, and civil society organisations: what role can public participation play in commercial AI labs?

data stewardship report sets out a framework for the different modes and practices for involving people in data.[23] In this report, we build on our work in *Participatory data stewardship*, adding a more grounded understanding of how industry organisations are integrating and using these methods in practice.

By exploring the role of public participation approaches in commercial AI contexts, this report seeks to provide richer insight into the opportunities for public participation approaches in the commercial AI space.

This report aims to shape emerging policy and practice debates around public participation and help answer a burning question facing policymakers, industry practitioners and civil society organisations: what role can public participation play in commercial AI labs?

23 Ada Lovelace Institute, Participatory data stewardship (2021) https://www.adalovelaceinstitute.org/report/participatory-data-stewardship/ accessed 10 January 2022.

Public participation in focus: theory and application

Public participation has a long history of use in democratic institutions and public policymaking, particularly for areas like healthcare[24] and the environment.[25] It also has a strong tradition in the history of design.[26] Existing literature demonstrates there may be a variety of different motivations for, and ambitions of, public participation:

- Participation might function as a procedural tool, generating 'wisdom of the crowd' to inform complex policy issues.[27]
- Participation may be leveraged by organisations as a means to raise social capital.[28]
- Participation might instrumentalise adjacent, but independent, goals such as increased inclusion.[29]

In this section, we examine the theory and conceptual underpinnings of public participation methods, as well as offering examples of participatory projects in other domains.

24 Josephine Ocloo and Rachel Matthews, 'From Tokenism to Empowerment: Progressing Patient and Public Involvement in Healthcare Improvement' (2016) 25 BMJ Quality & Safety 626.\\uc0\\u8216{}From Tokenism to Empowerment: Progressing Patient and Public Involvement in Healthcare Improvement\\uc0\\u8217{} (2016

25 Stephan Hügel and Anna R Davies, 'Public Participation, Engagement, and Climate Change Adaptation: A Review of the Research Literature' (2020) 11 WIREs Climate Change e645.

26 Tone Bratteteig and Ina Wagner, 'Unpacking the Notion of Participation in Participatory Design' (2016) 25 Computer Supported Cooperative Work (CSCW) 425.

27 'Designing Public Policies: Principles and Instruments - 2nd Edition -' https://www.routledge.com/Designing-Public-Policies-Principles-and-Instruments/Howlett/p/book/9781138293649 accessed 10 July 2023.

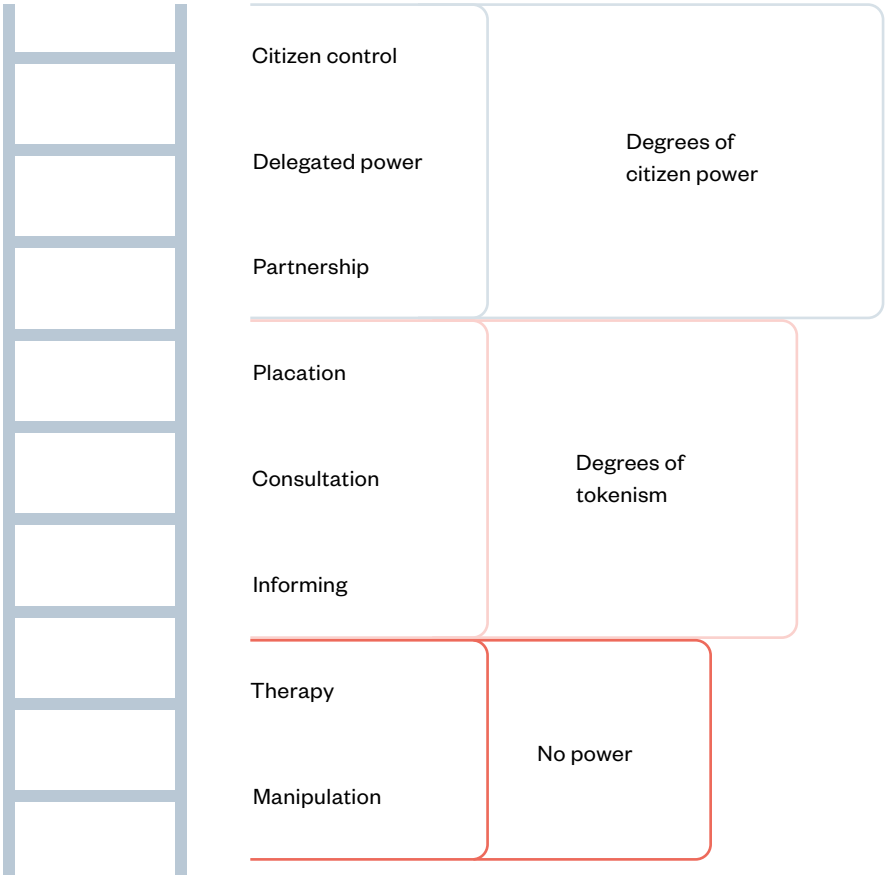
28 Robert D Putnam, Robert Leonardi and Raffaella Y Nonetti, Making Democracy Work: Civic Traditions in Modern Italy (Princeton University Press 1993) <https://www.jstor.org/stable/j.ctt7s8r7> accessed 10 July 2023.

29 Kathryn S Quick and Martha S Feldman, 'Distinguishing Participation and Inclusion' [2011] Journal of Planning Education and Research https://journals.sagepub.com/doi/epub/10.1177/0739456X11410979 accessed 10 October 2022.

Frameworks for public participation

One of the domains where public participation has been used is in public policymaking as a way to engage people in the creation of new laws and regulations. A widely cited framework is Sherry Arnstein's 'Ladder of Citizen Participation', which comprises different 'rungs' of public involvement in policy decision-making (see 'Figure 1: Arnstein's 'Ladder of Citizen Participation'). Designed to be provocative, Arnstein's Ladder characterises methods of participation predicated on one-way flows of information as 'manipulation' and 'tokenism', and suggests that truer forms of participation involve a greater share of decision-making and power to be shared with people (complete 'citizen control' sits as the final rung of the ladder).[30]

Figure 1: Arnstein's 'Ladder of Citizen Participation' [31, 32]



30 Sherry R Arnstein, 'A Ladder Of Citizen Participation' (1969) 35 Journal of the American Institute of Planners 216.
31 ibid.
32 Patel and others (n 22).

A more modern interpretation of Arnstein's Ladder comes from the International Association for Public Participation (IAP2), who offer five forms of public participation in decision-making: 'inform', 'consult', 'involve', 'collaborate' and 'empower'.[33]

The IAP2 spectrum matches the mode of participation to the goal of participation: for example, where the goal is to provide members of the public with 'balanced and objective information' about a particular project, proposal or intervention, then the mode of participation is 'inform'; whereas if the goal is to partner with members of the public across key moments of decision-making, the mode of participation is more aligned with 'collaborate'.[34]

The Ada Lovelace Institute's spectrum for participatory data stewardship builds both the Arnstein and IAP2 spectrum into a framework specifically designed to understand participation in data governance models (see page 20).

Public participation in action

We share three illustrative examples for how public participation is used in different domains:

- social infrastructure and urban planning
- participatory design
- citizen science.

Social infrastructure and urban planning

Arnstein's ladder was designed to conceptualise participation in urban planning contexts[35] which remains an area with well-established routes for members of the public to input on urban, regional and rural development. Many current international policy proposals around sustainable planning and infrastructure, such as the United Nation's

33 IAP2, 'Core Values, Ethics, Spectrum – The 3 Pillars of Public Participation - International Association for Public Participation' https://www.iap2.org/page/pillars accessed 4 May 2022.
34 ibid.
35 Arnstein (n 29).

Sustainable Development Goals, emphasise the importance of multi-stakeholder processes and citizen participation in the realisation of the goals.[36]

Example: We Can Make initiative in Bristol

Objective: Enabling a community to build affordable, sustainable housing designed in the community's interests.

The project: At a community event in Knowle West, Bristol in 2016, residents identified affordable local housing as becoming a serious issue. In a partnership between Knowle West Media Centre, a digital arts and social innovation centre, and White Design, an architects practice, residents collaborated with designers, academics and policymakers in the We Can Make initiative to set the agenda for housebuilding and advance a community-led vision for housing: locally-made, environmentally-friendly affordable homes, held in a trust for community benefit in perpetuity.[37] The project aims to build 300 affordable homes in Knowle West and will inform similar projects in other local authorities across the UK.

Level of participation: In this project, local residents worked closely with architects at the design phases to inform the design and ongoing project decision-making. This case study could be characterised as an example of the 'collaborate' level of the IAP2 Spectrum and Arnstein's ladder.

Participatory design

Additionally, the long history of participation in technology design offers useful learning for how public participation in AI could be structured in practice. Participatory design (PD) prioritises multi-stakeholder (end-users, designers, researchers, partners) collaboration into the design process for products, systems or services.[38] It emerged as a political movement – the Scandinavian workplace democracy movement in the 1970s[39] – and draws from fields such as sociology, political science,

36 Yasutaka Ozaki and Rajib Shaw, 'Citizens' Social Participation to Implement Sustainable Development Goals (SDGs): A Literature Review' (2022) 14 Sustainability 14471.
37 'Home' (WeCanMake) https://wecanmake.org/ accessed 11 July 2023.
38 'Participatory Design - an Overview | ScienceDirect Topics' https://www.sciencedirect.com/topics/computer-science/participatory-design accessed 30 May 2023.
39 Michael J Muller and Sarah Kuhn, 'Participatory Design' (1993) 36 Communications of the ACM 24.

public policy and communication studies.[40] User research, user testing and marketing-based approaches have drawn from the collaborative approach to PD, though with fewer explicitly political goals and more of a focus on the usability of the proposed product.[41]

University of the Arts London Creative Computing Institute, 'Syb'

Objective: Creating more queer representation in voice-based AI systems, promoting trans joy and connecting to queer and trans media.[42]

The project: University of the Arts London's Creative Computing Institute proposed a three-day workshop to imagine and prototype personal intelligent assistants, leading to the design Syb, a prototype voice interface

Developed through a participatory design process with a team of trans and non-binary people, Syb was designed to support and reflect the goals and values of this team of designers, according to a future where 'technology is developed by and for trans people, enabling them to imagine new and more liberating futures for themselves'.[43]

Level of participation: The process of co-defining goals, facilitating participation from start to finish, and placing total decision-making power in the hands of participants suggests this project might be considered an example of 'empower' on Arnstein's ladder or the IAP2 Spectrum.

Citizen science

Citizen science relies on the contributions of thousands of researchers and interested members of the public for co-production of knowledge. It facilitates public engagement and participation with the scientific community and science projects. Citizen science is a useful example of how to mediate participation at a substantial, even global scale.

40 Michael J Muller, 'Participatory Design: The Third Space in HCI' 32.
41 'Participatory Design: Bringing Users to the Design Process' (*Blog | Imaginary Cloud*, 17 June 2021) https://www.imaginarycloud.com/blog/participatory-design/ accessed 30 May 2023.
42 'Syb' https://www.feministinternet.com/syb accessed 24 July 2023.
43 'Syb: Queering Voice AI' (*The New New*) https://thenewnew.space/projects/syb-queering-voice-ai/ accessed 25 July 2023.

Zooniverse platform for citizen science projects

Objective: Facilitating partnerships with volunteers and professionals, benefiting from the 'wisdom of the crowd' to generate scientifically significant discoveries across a range of projects and disciplines.

The project: Zooniverse is the world's largest platform for 'people-powered research'.[44] Volunteers sign up to contribute to different research projects, usually for classification and pattern recognition tasks.

One of the most popular projects, Gravity Spy, involves participants helping astronomers label images produced by LIGO, an observatory detecting gravitational waves.[45] There are currently over 30,000 volunteers registered to take part, many of whom contributed to identifying novel 'glitches' in images to inform ongoing scientific inquiry.[46]

Level of participation: The exact shape of participation in citizen science project varies, but might be best characterised at the level of 'involve' or 'collaborate' on the IAP2 Spectrum.[47]

Theories of public participation in data and AI

Scholars in data and AI have built on the existing typologies of public participation in other domains to create frameworks and theories for public participation in the design and development of data and AI-driven technologies.

One of these frameworks is the Ada Lovelace Institute's *Participatory data stewardship* framework for involving people in the use of data. Drawn from the five levels of participation set out by Arnstein and IAP2, the framework details potential practical mechanisms that correlate with each of the levels and offers some real-world case studies of the use of these mechanisms.

44 'Zooniverse' https://www.zooniverse.org/about accessed 27 July 2023.

45 'Zooniverse' (nesta) https://www.nesta.org.uk/feature/ai-and-collective-intelligence-case-studies/zooniverse/ accessed 27 July 2023.

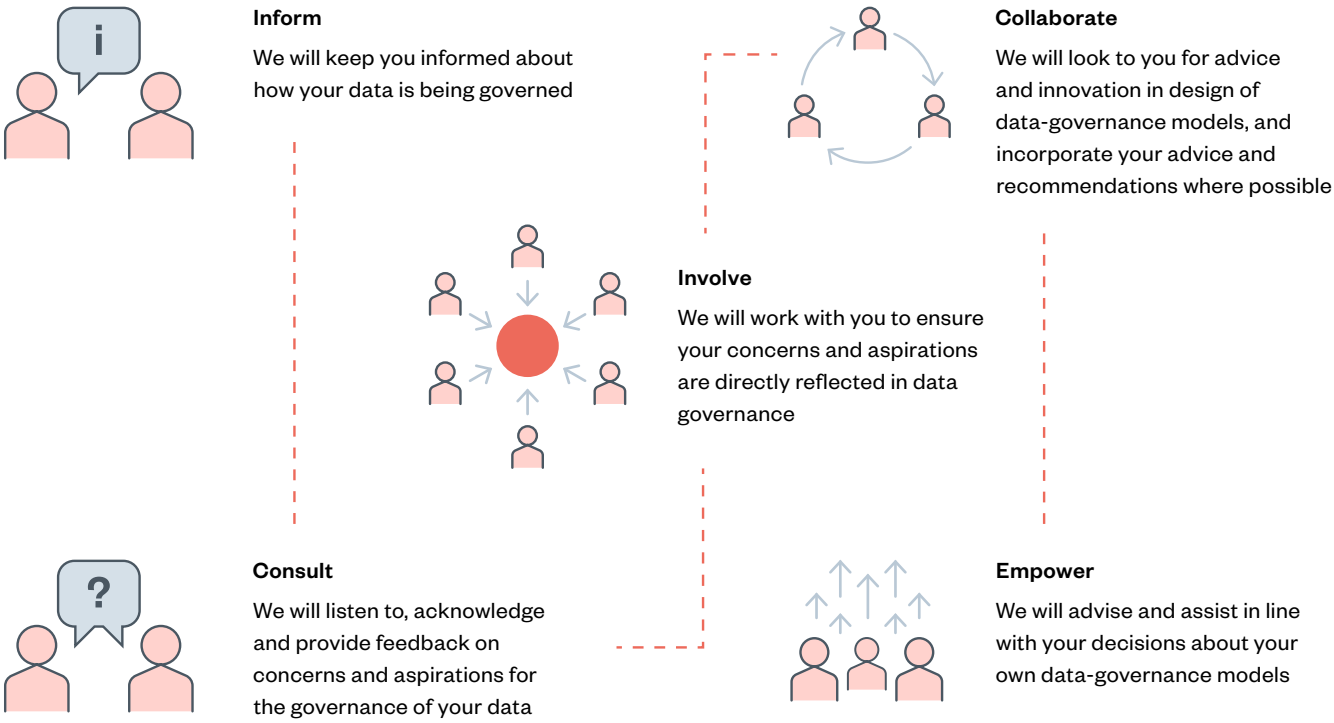
46 'Gravity Spy | Zooniverse - People-Powered Research' https://www.zooniverse.org/projects/zooniverse/gravity-spy accessed 27 July 2023.

47 IAP2 (n 32).

The five levels are:

1. **Informing** people about how data about them is used, such as through the publication of model cards.
2. **Consulting** people to understand their needs and concerns in relation to data use, such as through user experience research or consumer surveys.
3. **Involving** people in the governance of data, such as through public deliberation or lived experience panels.
4. **Collaborating** with people in the design of data governance structures and the technologies they relate to, such as through novel institutional structures like 'data trusts'.
5. **Empowering** people to make decisions about datasets and technologies built with them, such as through citizen-led governance boards.[48]

Figure 2: Ada's spectrum of participatory data stewardship[49]



48 Patel and others (n 22).

49 ibid.

Other frameworks and theories have sought to categorise different kinds of public participation in the machine learning and AI process. Sloane et al. (2020)'s framework identifies three distinct kinds of participation that occur throughout the AI lifecycle process:

- **Participation *as work*** recognises the often intensive labour that goes into the 'production or refinement' of AI systems. For example, in human content moderation to annotate and clean an AI system's training dataset, a common type of labour in AI development that is often undertaken by workers earning low wages and with poor labour conditions. Some of this work may even be outsourced to workers in the global majority who earn less than workers in the US or Europe.[50] Such content moderation work is often traumatising and likely to cause lasting harm to workers.[51]

- **Participation *as consultation*** involves seeking quick input or feedback on certain project decisions from certain stakeholder groups, for example, subject-matter experts or potential users of the proposed technology. Short-term participatory 'design sprints' often take this approach, as do urban planning projects.[52]

- **Participation *as justice*** centres on longer-term partnerships and enquiries that contribute to agency over design and infrastructure that affects the lives of participants. Participation is less about technology-focused outcomes, instead challenging existing power dynamics between developers of AI systems and those who are impacted by them. Sloane et al. suggest the work of organisations such as Data for Black Lives (a non-profit organisation with a mission to use data science to create concrete and measurable change in the lives of Black people[53]) might fall under the banner of 'participation as justice'.[54]

50 PAI Staff, 'Responsible Sourcing of Data Enrichment Services' (*Partnership on AI*, 16 June 2021) https://partnershiponai.org/responsible-sourcing-considerations/ accessed 13 July 2023.

51 Abeba Birhane, Vinay Uday Prabhu and Emmanuel Kahembwe, 'Multimodal Datasets: Misogyny, Pornography, and Malignant Stereotypes' (arXiv, 5 October 2021) http://arxiv.org/abs/2110.01963 accessed 13 July 2023.

52 Mona Sloane and others, 'Participation Is Not a Design Fix for Machine Learning' http://arxiv.org/abs/2007.02423 accessed 11 May 2023.

53 'Home' (*D4BL*) https://d4bl.org/ accessed 26 July 2023.

54 Sloane and others (n 51).

Sloane et al. (2020) also introduce the concept of 'participation washing',[55] which broadly refers to extractive and exploitative public participation practices.

> Participation washing can take many forms, including practitioners not respecting the needs and ideas of participants, not providing them with clear instructions about what they'll be expected to contribute or not compensating participants sufficiently.

Practitioners who engage in participation washing remove accountability for their actions and fail to delegate decision-making power to participants.

Following from Sloane et al (2020), Birhane et al (2022) introduce a framework for understanding participation as it appears in different stages of the machine-learning process. The authors identify three instrumental categories for participation:

- **Participation for *algorithmic performance improvement*:** participation in order to help refine or personalise the AI system or model. This dimension of participation might comprise a computer science 'hackathon'[56] to propose potential updates or collaboratively work through engineering challenges, or a 'red-teaming'[57] exercise where a team simulates an adversarial system attack to identity weak security points.

- **Participation for *process improvement*:** using participation to input on and inform the overall design or project process. Similar to Sloane et al. (2020)'s 'participation as consultation' dimension, participation for 'process improvement' might seek quick input or feedback on project objectives or activities. Birhane et al. consider citizen science as an example of 'participation for process improvement'.

55 ibid.
56 'Dreambooth-Hackathon (DreamBooth Hackathon)' (19 December 2022) https://huggingface.co/dreambooth-hackathon accessed 13 July 2023.
57 Deep Ganguli and others, 'Red Teaming Language Models to Reduce Harms: Methods, Scaling Behaviors, and Lessons Learned'.

- **Participation for *collective exploration*:** participants self-organising to facilitate discussion around shared goals for a particular community as opposed to purely technology- or project-driven goals. Participants might tackle questions around who participates, to what ends, and who stands to benefit.[58] Birhane et al. use the example of the *Te Hiku* NLP project, where the Māori community in New Zealand recorded and annotated 300 hours of audio data of the *Te Reo Māori* language to translate into tools such as speech-to-text technology.[59] In addition to this, the group crafted Māori Data Sovereignty Protocols to determine the shape of future contributions according to the needs, goals and values of the community.[60,61]

These frameworks address some of the stated goals of participation in the AI process, how these might be mediated through practical approaches and what the motivations might be for embedding public participation in the AI development process.

As we show above, while Arnstein is generally critical of approaches that appear at the bottom of the 'ladder', labelling them tokenistic, the IAP2 and our own *Participatory data stewardship* spectrum of participation aim to show the different degrees of participation without assigning a particular normative value to each approach. This is because different contexts will require a different level of engagement or involvement, which might accrue greater or fewer benefits to participants and the organising organisation. These frameworks also demonstrate there is no 'one size fits all' model for public participation, and that the right method should be chosen depending on the context, needs and objectives of the development team.

Where might public participation occur in the AI lifecycle?

In the context of AI development, opportunities for participation arise at different stages in the design and development lifecycle. For 'narrow AI' systems, these stages include:

58 Birhane and others (n 13).
59 'Māori Are Trying to Save Their Language from Big Tech | WIRED UK' https://www.wired.co.uk/article/maori-language-tech accessed 25 July 2023.
60 'Resources' (*Te Mana Raraunga*) https://www.temanararaunga.maori.nz/nga-rauemi accessed 25 July 2023.
61 Birhane and others (n 13).

Figure 3: Public participation across the AI lifecycle

1. **PROBLEM FORMULATION**
   What is the problem or challenge that we are trying to solve?

2. **IDEATION**
   What solution might help us address the problem, as formulated? What data or resources will I need?

3. **DATA COLLECTION AND ANALYSIS**
   Collecting, cleaning, annotating and analysing the data needed to address this tool

4. **MODEL OR TOOL DESIGN AND DEVELOPMENT**
   Design of the specific model architecture or wider AI-powered product that will address the problem

5. **TRAINING**
   Training of the model or system on the data collected

6. **TESTING AND EVALUATION**
   Evaluating the model or system's performance, safety, biases, accuracy, precision, recall and other metrics

7. **PILOTING**
   Testing the model or system in real-world settings or in a sandbox

8. **PROCUREMENT/DEPLOYMENT**
   Deploying the AI system or model in a real-world setting, or selling it to a third party who deploys it

9. **MONITORING AND EVALUATION**
   Auditing, testing and evaluating the system's actual performance over time. This can also include studying how the system integrates into complex real world environments, changing behaviours and relationships between different actors (e.g. nurses, doctors, and patients) or institutions (e.g. police departments and social workers)

For labs developing foundation models – a base model for different organisations to build applications on top of – opportunities for public participation would be limited to the data or model development layer (stage 3 or 4) and further stages might be undertaken by 'downstream' developers.

AI-based projects often require input and activity from many stakeholders in a range of capacities. The design and development of AI systems can often involve multiple teams and organisations, encompassing various stages from data analysis to deployment. These can include different teams within a company, such as product, engineering, legal and policy teams; but it can also include engagement and consultation with people outside of the organisation, such as potential buyers of the technology, regulators, data annotators who clean the data and even journalists. This has led some to argue that AI is intrinsically 'participatory'.[62]

Developers of AI systems can employ a range of participatory approaches in different stages of this process. Below are some examples:

Royal Society of Arts (RSA) and Google DeepMind Forum for Ethical AI citizens' jury

Single stage: Problem formulation

- In 2018, the RSA and Google DeepMind convened participants to deliberate on ethical issues surrounding AI and algorithmic decision-making systems, asking participants an open-ended question to facilitate rich discussion: 'under what conditions, if any, is it appropriate to use automated decision systems?'.[63]
- The motivation for conducting a citizens' jury was to fold public voice into broader discussion and debate around AI and use the results to inform future policies and design decisions. As such, we might characterise this participatory approach as appearing at the **problem formulation** phase, as it is intended to frame the problem and offer insight into potential actions.

- In this study, we don't have a clearly mapped process for how participants' contributions would inform ongoing research and development at DeepMind. So while early-stage public participation allows members of the public opportunity to contribute to agenda-setting, a potential limitation of this approach might be that contributions aren't taken forward.

62 A Feder Cooper and others, 'Accountability in an Algorithmic Society: Relationality, Responsibility, and Robustness in Machine Learning' [2022] arXiv:2202.05338 [cs] http://arxiv.org/abs/2202.05338 accessed 19 May 2022.

63 The RSA, 'Democratising Decisions about Technology: A Toolkit' (2019) https://www.thersa.org/reports/democratising-decisions-technology-toolkit accessed 3 February 2023.

Sepsis Watch clinical tool co-design project

Multiple stages: Problem formulation, data collection and analysis, model development, pilot

- Sendak et al.'s study to co-design a sepsis detection and management platform, Sepsis Watch, is a useful example of how participation can be leveraged at multiple points in the development process, particularly to meet non-technology-focused outcomes, including improved clinical decision-making.
- Frontline clinicians (such as nurses and doctors) were assembled at the **problem formulation** stage, so that Sepsis Watch was 'developed to meet a specific problem in a specific hospital, defined by clinicians working in that hospital'.[64] Clinicians were also invited to input on data curation (**data collection and analysis**) and two nurses informed **model development**.
- An interdisciplinary team was assembled at the **pilot phase**, including frontline nurses, clinical experts and innovation team staff, with clear lines of communication/feedback established.

Participation in commercial AI research and development

What can the commercial AI sector tell us about trends and debates in AI?

Since the early 2010s, major technology companies like Microsoft, Google, Amazon and Meta have invested heavily in creating dedicated in-house AI research labs that feed novel insights into their products. In recent years, some smaller AI labs like OpenAI and Google DeepMind have either signed landmark partnerships with major technology companies or have been fully acquired by these companies. As highlighted in the State of AI 2022 report, commercial AI labs retain significant resource and talent advantages over academic or government-funded labs. Most of the major developments in the capabilities and applications of AI technologies now take place in commercial AI labs or via partnerships between academic researchers

64 Mark Sendak and others, '"The Human Body Is a Black Box": Supporting Clinical Decision-Making with Deep Learning', Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (ACM 2020) https://dl.acm.org/doi/10.1145/3351095.3372827 accessed 25 July 2023.

and these labs.[65]

> As some researchers have noted, the concentration of compute and data resources by a small number of large AI companies has constituted a form of 'capture' over scientific advances and applications of AI.[66]

This concept of capture may also extend to what kinds of 'ethical' or 'responsible AI practices' companies put in place to mitigate harm to people and the environment.[67] In addition to funding major research and development projects into AI, the commercial AI industry exercises considerable influence in setting trends and tone for debate for AI more broadly, including around what constitutes ethical or responsible AI (RAI) practices. While RAI practices have included aspects like algorithmic auditing, the use of impact assessments and other kinds of data practices, some of these approaches have included 'participatory AI' practices.

Some AI labs, especially those in larger corporations, have in recent years hired dedicated teams to investigate ethics issues and manage internal ethics review processes for AI research. These teams tend to be called 'ethical AI' or 'responsible AI' teams, and examples include Microsoft's Office for Responsible AI, Google DeepMind's Ethics and Society team, and IBM's AI Ethics Board. The teams develop and use different methodologies, tools and approaches for implementing practices that seek to identify and mitigate potential ethical risks in the research process, including considerations for the broader societal impacts of how research may be used.

In some cases, these teams have publicly shared specific approaches and methods they've developed, such the Google Ethical AI team's model cards process for reporting transparent details about an AI

65 Nathan Benaich and Ian Hogarth, 'State of AI Report 2022' (2022) https://www.stateof.ai/> accessed 2 February 2023.
66 Whittaker (n 14).
67 'AI vs. Responsible AI: Why It Matters' (RAI Institute, 24 January 2023) https://www.responsible.ai/post/ai-vs-responsible-ai-why-is-it-important accessed 17 July 2023.

'Ethical AI' or 'responsible AI' teams seek to identify and mitigate potential ethical risks in the research process, including broader societal impacts

model's biases and intended uses[68] or Microsoft's Responsible AI Standard that establishes a series of practices for research and product teams to follow.[69] Similarly, Twitter's META (Machine Learning, Ethics, Transparency and Accountability) team ran the first algorithmic bias bounty challenge in 2021, which involved sharing access to some of Twitter's code and inviting the research community to identify potential ways to improve the performance of its image cropping algorithm.

This team was one of the first to be laid off in the restructuring of Twitter under Elon Musk. Further layoffs from ethics and responsible AI teams across the technology industry occurred in the following months, including at Meta, Microsoft and Google.[70] These layoffs and related 'hiring freezes' for these teams suggest that ethicists and other advocates for responsible innovation within technology companies occupy an institutionally insecure position when these companies face economic challenges.

There is a growing literature of social science research into the experiences of ethical and responsible AI teams, which highlights a number of challenges and considerations they face when implementing internal ethics processes:

Taking the spotlight: who 'owns' ethics?

In many companies, there is not a clear delineation of responsibilities for formulating and embedding ethical AI or responsible AI initiatives across the organisation. Moss and Metcalf coin the term 'ethics owners' to describe specific roles within companies that take on the burden of thinking about the ethical and societal risks of products or research. Companies may have different descriptions of these roles, but broadly these practitioners take on a unique set of skills and practices that include facilitating internal compliance with ethical frameworks, translating external public pressure into corporate practice and

68 Margaret Mitchell and others, 'Model Cards for Model Reporting' [2019] Proceedings of the Conference on Fairness, Accountability, and Transparency 220.

69 Microsoft, 'Microsoft Responsible AI Standard v2 General Requirements' [2022] Impact Assessment. https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2022/06/Microsoft-Responsible-AI-Standard-v2-General-Requirements-3.pdf.

70 Gerrit De Vynck and Will Oremus, 'As AI Booms, Tech Firms Are Laying off Their Ethicists' Washington Post (3 April 2023) https://www.washingtonpost.com/technology/2023/03/30/tech-companies-cut-ai-ethics/ accessed 30 May 2023.

preparing the company for future regulation.[71]

Technology companies may situate ethics roles in different parts of a company, such as in a legal, policy, research or communications team. A poorly situated and embedded team can create a risk that internal ethics initiatives are siloed away from the work of product and research teams. Additionally, there is a risk that ethics teams may only have a remit to critique and challenge individual researchers or product teams, but no remit to challenge institutional culture issues and leadership decisions that create the conditions for unethical behaviour.[72]

Striking the balance: ethics and AI development

Another challenge ethical and responsible AI teams face is that of doing slow, reflexive ethical deliberation about a research project or product while operating in a sector that incentivises fast publication and product launch timescales. Winecoff and Watkins (2021) demonstrate how technology entrepreneurs are incentivised to rapidly develop and launch technology products to the detriment of the slower, more meticulous practices that demonstrate scientific rigour.[73]

A similar challenge regularly arises for ethics practitioners: faced with metrics, targets and deadlines aimed at creating a quick launch schedule for a product, practitioners may struggle to enact the reflexivity and deliberation required to understand and mitigate the risks those products may pose for people and society.[74] The result can be a watered-down or 'tamer' ethics that does not seek to disrupt the institutional business model and pace of working.[75]

71 Emanuel Moss and Jacob Metcalf, 'Ethics Owners: A New Model of Organizational Responsibility in Data-Driven Technology Companies' 74.

72 Ben Green, 'The Contestation of Tech Ethics: A Sociotechnical Approach to Technology Ethics in Practice' (2021) 2 Journal of Social Computing 209.

73 Amy A Winecoff and Elizabeth Anne Watkins, 'Artificial Concepts of Artificial Intelligence: Institutional Compliance and Resistance in AI Startups' (14 June 2022) http://arxiv.org/abs/2203.01157 accessed 14 July 2022.

74 Bogdana Rakova and others, 'Where Responsible AI Meets Reality: Practitioner Perspectives on Enablers for Shifting Organizational Practices' (2021) 5 Proceedings of the ACM on Human-Computer Interaction 1.

75 Sanna J Ali and others, 'Walking the Walk of AI Ethics: Organizational Challenges and the Individualization of Risk among Ethics Entrepreneurs', Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (Association for Computing Machinery 2023) https://dl.acm.org/doi/10.1145/3593013.3593990 accessed 28 June 2023."plainCitation":"Sanna J Ali and others, 'Walking the Walk of AI Ethics: Organizational Challenges and the Individualization of Risk among Ethics Entrepreneurs', Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (Association for Computing Machinery 2023

Flying the flag: ethics washing

> A much-discussed challenge for ethics practitioners in the technology industry are claims of 'ethics washing', where internal ethics initiatives are prioritised as a form of self-regulation to delegitimise the need for external regulation.[76]

One example of a corporate ethics initiative that received accusation of ethics washing was Google's decision to stand up an ethics review board in 2019 that had no veto power over its decisions. The board garnered additional criticism over its membership, which included members who had expressed anti-LGBTQ views and spread misinformation relating to climate change.[77] Claims of ethics washing come when a technology company signals to the broader public that they are taking ethical concerns seriously, but often without substantive shifts to their status-quo practices. It is possible that public participation efforts might be vulnerable to this effect.

In summary, the conditions set out above present considerable challenges to teams or individuals who may wish to trial or experiment with embedding public participation approaches in industry. Participation projects also demand careful input from multiple stakeholders, and clearly scoped resourcing, which are likely to require at least some level of institutional buy-in.[78]

Public participation in commercial AI

The concept of 'participatory AI' has generated enthusiasm within the technology industry and is rapidly gaining traction. Perhaps in part due to the capaciousness of 'participatory AI' or 'democratising AI', we

76 essica Morley and others, 'Operationalising AI Ethics: Barriers, Enablers and next Steps' [2021] AI & SOCIETY https://link.springer.com/10.1007/s00146-021-01308-8 accessed 18 November 2021.

77 'Google Appoints an "AI Council" to Head off Controversy, but It Proves Controversial' (MIT Technology Review) https://www.technologyreview.com/2019/03/26/136376/google-appoints-an-ai-council-to-head-off-controversy-but-it-proves-controversial/ accessed 18 July 2023.

78 'How Do I Make the Case for Public Participation?' (involve.org.uk, 8 May 2018) https://involve.org.uk/resources/knowledge-base/how-do-i-make-case-public-participation accessed 18 July 2023.

have witnessed an uplift in technology industry actors adopting some of the language of public participation or democracy with regards to AI governance.

For example, in February 2023, OpenAI – creators of the popular LLM ChatGPT – published a blog post stating their intent to use public decision-making as a mechanism to align AI systems with societal values.[79] A few months later, in May 2023, OpenAI announced a fund for grants proposing processes to enable 'democratic input to AI', particularly around values and policies for use of products such as ChatGPT.[80]

Like almost all AI systems, OpenAI's ChatGPT relies on human annotation and labelling of the data used to train its system. ChatGPT also relies on human feedback on how the system behaves that is then used to retrain the system, a process known as reinforcement learning from human feedback (RLHF).[81] The people involved tend to be crowd-workers from websites like Mechanical Turk (MTurk) or providers of content moderation and data annotation services like Sama, a company that operates in Africa.[82] These workers are paid as part of a job, and the criteria for appropriate behaviour is usually set by the company developing the system.

> Recent press coverage has highlighted exploitative and harmful working conditions of these workers.[83, 84] Under the framework of public participation mentioned above, this process is not a meaningful form of consultation or collaboration with members of the public as it is not asking for their unfettered insights and perspectives on how the technology should be designed, deployed or governed.

79 OpenAI (n 1).
80 'Democratic Inputs to AI' (n 9).
81 'Introducing ChatGPT' https://openai.com/blog/chatgpt accessed 14 August 2023.
82 'OpenAI Used Kenyan Workers on Less Than $2 Per Hour: Exclusive | Time' https://time.com/6247678/openai-chatgpt-kenya-workers/ accessed 1 June 2023.
83 ibid.
84 'Introducing ChatGPT' (n 80).

Historically, there has been little public evidence of commercial AI labs trialling public participation methods in their research or product decisions

A commercial AI startup, Anthropic, has also adopted the language of political institutions and processes for an initiative they call 'Constitutional AI', a process for reducing the harmfulness and increasing the helpfulness of generated AI responses by creating a set of principles (what they call a 'constitution') to guide how the model is trained.[85] The goal of this process here is to minimise the requirement for widescale participation of human moderators in determining system behaviour by requiring the system to moderate itself. Anthropic has used the language of democratic institutions in this initiative, but the process does not adopt any components that enable people to input on the selection of these principles. This initiative cannot be categorised as 'involving' or even 'consulting' people, and therefore cannot be considered a meaningful form of 'public participation', despite its language.

Historically, there has been little public evidence of commercial AI labs trialling public participation methods in their research or product decisions. Where public participation projects have emerged, they have generally been driven by ethical AI teams within AI labs. A prominent example includes the Royal Society of Arts (RSA) and Google DeepMind's Forum for Ethical AI project in 2019, which ran a citizens' jury with members of the public, enabling deliberation on benefits and risks for algorithmic decision making.[86] In 2022, the Behavioural Insights Team (BIT) released a blogpost showcasing a recent partnership with Meta, also using citizens' assemblies for deliberation on climate misinformation. This blog was released with only minimal detail about the methodological components of this exercise.

More recently, IBM established a specific initiative tasked with developing participatory research called the Responsible and Inclusive Technology Participatory Initiative (RITPI). No publicly available information on RITPI exists, but according to Jing et al. (2023) RITPI has piloted a variety of approaches internally, including general participatory workshop methods, associated interface prompts and guidelines for community engagement, co-designed with people and communities affected by technologies with a view to creating documentation and

85 'Claude's Constitution' (*Anthropic*) https://www.anthropic.com/index/claudes-constitution accessed 22 May 2023.

86 RSA (n 62).

There are some examples of public participation methods being used in academic/industry research collaborations

artefacts reflecting their views.[87] In their study, the authors highlight some tensions around corporate incentives providing bias toward 'tangible, deliverable and solution oriented design artifacts'.[88]

There are also some examples of public participation methods being used in academic/industry research collaborations. A 2022 partnership between Meta and Stanford University's Deliberative Democracy Lab trialled a deliberative polling approach across several thousand participants from over 30 countries. Participants were asked to discuss policy proposals for countering bullying and harassment across Meta's services, rather than tweaks to specific AI services like a newsfeed algorithm. One of the advisers to the project suggested that the participants could have been afforded greater decision-making power, but the process represents a useful 'proof of concept' for further initiatives.[89] Meta has indicated they may adopt the approach to guide decision-making around their generative AI initiatives.[90]

This project, along with OpenAI's 'democratic input to AI' initiative, are interesting examples of public participation in commercial AI initiatives being pitched at a global level, with input from a large number of participants, as opposed to the RSA and Google DeepMind pilot which ran with a small number (25–29) of participants. For large technology companies, a large-scale exercise could potentially signal a desire to obtain broad social licence for AI and other technologies.

> The projects outlined above also point to another emerging trend in technology sector public participation: a tendency for projects in industry to involve external civil society organisations and partners.

87 Felicia S Jing, Sara E Berger and Juana Catalina Becerra Sandoval, 'Towards Labor Transparency in Situated Computational Systems Impact Research', *2023 ACM Conference on Fairness, Accountability, and Transparency* (ACM 2023) https://dl.acm.org/doi/10.1145/3593013.3594060 accessed 18 July 2023.

88 ibid.

89 'Meta Ran a Giant Experiment in Governance. Now It's Turning to AI | WIRED' https://www.wired.com/story/meta-ran-a-giant-experiment-in-governance-now-its-turning-to-ai/ accessed 19 July 2023.

90 Vandana Nair, 'Meta Needs You in Its Generative AI Gambit' (*Analytics India Magazine*, 17 July 2023) https://analyticsindiamag.com/meta-needs-you-for-their-generative-ai/ accessed 18 July 2023.

Multi-stakeholder collaboration may offer fruitful avenues to better quality public participation projects

A further example is the relationship between OpenAI, Anthropic and the Collective Intelligence Project, a non-profit organisation established in 2023 to facilitate research around alternate governance and participation models for emerging technologies. In May 2023, the three organisations announced an investigation into the potential for 'alignment assemblies' – opportunities for members of the public toarticulate their 'needs, preferences, hopes and fears' for AI in order to bring these technologies into alignment with societal need.[91]

Though we lack extensive evidence into the impact of these partnerships on participants and companies, when we consider that ethics initiatives in commercial AI are often encumbered by resourcing and capacity challenges, multi-stakeholder collaboration on participation (including with civil society actors) may offer fruitful avenues to better quality public participation projects.

The lack of public evidence on adoption of participatory approaches within commercial AI labs limits our ability to draw conclusions about the state of public participation in commercial AI. There is still limited understanding of what public participation practices AI labs may be experimenting with, what objectives they have for public participation and what challenges the teams running such projects face.

Without extensive evidence, there is little concrete understanding about what 'best practices' commercial AI labs should adopt and what kinds of problems these practices can help address. This gap is also to the detriment of policymakers and civil society organisations, who must determine their own role in relation to industry-led public participation projects.

91 'We Should All Get to Decide What to Do about AI' (*The Collective Intelligence Project*) https://cip.org/blog/alignment accessed 31 May 2023.

Interview findings from commercial AI lab practitioners and public participation experts

In order to better understand how commercial AI labs are using public participation methods, we conducted semi-structured research interviews with 12 participants: nine industry practitioners with a stake in 'participatory AI', and three public participation experts currently working in academia and civil society, but with knowledge of and/or experience working in technology companies. We used interviews to surface their experiences and understanding of public participation practices in commercial AI labs. These interviews were conducted in the period from Spring 2022 to Winter 2022/3. For a comprehensive overview of our research methods, analysis, and limitations, see 'Methodology' on page 64.

These interviews surfaced **five key findings**, summarised here and explored in detail below:

1. Within commercial AI labs, researchers and teams using public participation methods view them as a **mechanism to ensure their technologies are beneficial for people and society**, and a way to support the mission and objectives of their organisation.

2. Our interviews with different practitioners revealed **a lack of consistent terminology** to describe public participation methods and **a lack of any consistent standards** for how to employ these methods.

3. Ultimately, **industry practitioners are not widely or consistently using public participation methods in their day-to-day work**. These methods tend to be deployed on an ad-hoc basis.

4. Industry practitioners **face multiple obstacles** to successfully employing public participation methods in commercial AI labs.

These include resource intensity, misaligned incentives with management and teams, practitioners feeling siloed off from product or research teams, and commercial sensitivities constraining practitioner behaviour.

5. Public participation methods **are not well-suited for foundation models**. It is challenging to adopt public participation in contexts that lack a clear use case, presenting implications for foundation models or generative AI systems and research.

Public participation is viewed as beneficial for people and society, but may also support the mission and business direction of commercial AI labs

One of the research objectives of this study is to understand the objectives that commercial AI practitioners may have for using public participation methods. We asked interviewees to share with us their own understanding of participation as both a terminology and a methodology, and to describe the purpose of participation from their perspective. Responses largely coalesced around two main objectives or goals for participation in commercial AI labs:

- Participation might be useful or effective in producing societally 'good' outcomes and might advance social justice goals.
- Participation may support the business mission of an AI lab.

When probed on what societally 'good' outcomes might look like, interviewees put forward a number of ideas:

- Participation might help companies facilitate inclusion with marginalised and disadvantaged communities.
- Participation might help align AI with societal needs.
- Participation might foster greater transparency and accountability between technology companies and people affected by AI.

'It comes down to power and decision-making, and distributing power among stakeholders.' Industry practitioner A

'I'm trying to collectively imagine what a beneficial future might look like and use storytelling and designed objects as tools for discussion about kind of preferable futures and what it is that we're actually trying to design for.' Industry practitioner B

Interviewees also made the argument that public participation might be good for business, on two different grounds:

- In terms of creating products that are more profitable because they better fit the needs and desires of potential customers;
- In terms of improving the company's reputation and raising other forms of social capital like relationship-building and trustworthiness.

Most participants put forward a case for public participation primarily supporting or initiating outcomes that are beneficial for people and society, and contributing to societal-level goals or values, such as inclusion, power-sharing, collective decision-making and fairness. This reflects the findings from Sloane et al.'s (2020) study that participation can be viewed as a method to achieve just outcomes.92 Many participants described societal benefit as the primary goal for any sort of public-participation exercise in AI development. This was particularly true for public participation experts, who all supported this view.

A secondary objective for most participants was business-driven: to make better products or to improve corporate reputation. One practitioner suggested that incorporating user feedback (as a mode of participation) would generate 'better' products and contribute to bottom line:

'It should be for good business, right? Engaging with the public and engaging with people should help you build a product that addresses their wants and needs better, which in turn should probably make your company more profitable.' Industry practitioner C

Participation could, for example, be used to provide feedback or user testing on products and research, including the usability or robustness

92 Sloane and others (n 51).

of a product.[93] However, many participants noted that in practice, participatory projects were not always designed with these goals in mind (see page 44 below about the current obstacles for embedding public participation in commercial AI). Participants differed as to whether the two goals are inherently in opposition or whether there could be some alignment between them.

'We do a lot of AI for social good projects at [large company]. But I'm always wondering why we need the qualifier of AI for social good.'
Industry practitioner A

Several participants suggested that couching the benefits of public participation through the language of increased (social, economic) benefit for the company would be the most likely argument to carry weight for company shareholders to resource and justify such projects.

There is no clear shared terminology around public participation and no consistently used methods across commercial AI labs

Our interviews reveal a lack of consistently used terminology and public participation methods in commercial AI labs. When interviewees were asked what methods or approaches could potentially be considered 'participatory', responses reflected a wide variety of different terms. In total, 19 different methods and approaches were put forward as either:

- approaches that interviewees report using directly in their capacity as a commercial AI practitioners
- approaches that interviewees were familiar with/aware of being in use across the technology sector that could be potentially applicable to their work/organisation
- approaches that interviewees were familiar with/aware of being in use in other domains.

Most interviewees expressed familiarity with the idea that different participatory approaches might fulfil different needs and that different

93 Min Kyung Lee and others, 'WeBuildAI: Participatory Framework for Algorithmic Governance' (2019) 3 Proceedings of the ACM on Human-Computer Interaction 181:1."plainCitation":"Min Kyung Lee and others, 'WeBuildAI: Participatory Framework for Algorithmic Governance' (2019

modes of participation might be useful or desirable, with two explicitly using the word 'spectrum' to describe the range of approaches on offer (echoing frameworks including Arnstein's ladder). Approaches given ranged from classic methodologies associated with deliberative democracy, such as citizens' juries, to research methods for convening and crowdsourcing opinion, like workshops or focus groups. Two participants considered whether the release of an AI model via an open-source approach94 – in which the code and data are made publicly available via an online repository for other members of the research community – might be characterised as a form of public participation.

In Table 1 below, we reproduce the public participation approaches that participants cited and mapped them according to Arnstein's 'Ladder of Citizen Participation'.

94 Irene Solaiman, 'The Gradient of Generative AI Release: Methods and Considerations' (arXiv, 5 February 2023) http://arxiv.org/abs/2302.04844 accessed 14 February 2023.

Table 1: Plotting participatory methods against Arnstein's 'Ladder of Citizen Participation' and participants' understanding of the purpose of participation

Degrees of citizen power	Cooperatives	**Participation as a form of accountability**
	Citizen's jury	
	Community-based approaches/ participatory action research	**Embedding lived experience**
	Deliberative approaches	
	Participatory design	
	Speculative design/anticipatory futures	**Relationship building**
	Participation in governance mechanisms e.g. impact assessments	
Degrees of tokenism	Co-design	**Trust building**
	Community training in ML	
	Community-Based Systems Dynamics framework	
	Crowdsourcing	
	Fairness checklist	**Democratising AI**
	UX/user testing	
	Participatory dataset documentation	
	Value-Sensitive/Value-Centred Design	**Participation's intrinsic value**
	Diverse Voices method	
Non-participation	Workshops/convenings	
	Consultation	
	Surveys	**Soliciting input/ knowledge transfer**
	Request for comment	
		Appeasement

Most participants were able to name approaches they'd heard about being trialled or in active use across the sector, even where they had not personally applied them to their own work. Some of our interviewees expressed firm opinions about which might be best suited for application in commercial AI labs more generally, while others were unsure which would be the most effective or applicable, citing a lack of evidence across industry about best practice. Three interviewees reported feeling confused about the possible direction for certain practices in lieu of formally established standards or best practice guidelines for the sector:

'What does "responsible crowdsourcing" actually look like? How do we verify that [existing contractors] do their due diligence and also compensate and craft a representative group?' Industry practitioner D

From the interviews, the approach that emerged in most frequent use across the sector was a form of **consultation**.

The *Participatory data stewardship* framework characterises 'consult' as its own mode of participation, where organisers of participation should aim to 'listen to, acknowledge and provide feedback on concerns and aspirations' of the public.[95]

One public participation expert we spoke with described their understanding of how these consultations proceed in the technology sector:

'Technology companies, to the extent they do any consultation at all (which they tend not to), it's designed as window dressing, or as user experience to improve a product or get feedback on a specific product. It's not really about the broader question about the impacts of AI.' Public participation expert A

In a number of cases that practitioners referred to, consultations were not undertaken with members of the public or community groups at all, and instead the participants comprised domain experts from fields including law, education or health.

95 Patel and others (n 22).

One practitioner describes the process in the large technology company they work for:

'So engaging with subject matter experts, they're usually brought in under NDA. They don't become affiliated with [company], but we enter into [a] contractual relationship with them where the things that they're working on with us become our intellectual property. So there's sort of this protection bubble built around what's discussed, the issues that are raised. All organisations have sort of a reputational management thing that they want to do.' Industry practitioner E

Another interviewee drew attention to the ad hoc manner in which consultations and other methods get designed and implemented. Usually, these consultations occur to address a need for quick input or feedback on an existing product or research decision, generally later on in the development pipeline rather than at the 'problem formulation' or 'ideation' phases (see Figure 3).

As we demonstrate in the introduction, there is no 'right' approach to conducting public participation and for certain contexts, rapid input in the form of a consultation will be most appropriate for the problem to be solved.

However, the prevalence of narrowly scoped, rapid input consultations in labs may have concerning implications for the general trajectory of public participation in the commercial AI industry.

> In cases where the parameters for participation are narrowly scoped, and defined and determined by the company, there is a limit to the level of agency a participant can expect[96] and a risk that participation is co-opted or distorted by the company in order to become mere rubberstamping.

96 'Meta Ran a Giant Experiment in Governance. Now It's Turning to AI | WIRED' (n 88).

Second, consultative approaches with domain experts don't necessarily reflect the concerns or issues of members of the public, particularly those belonging to marginalised groups who are often excluded from the development and deployment of AI systems, and who might have lived experience of algorithmic harm.[97]

Commercial AI labs do not appear to prioritise public participation methods in the research and development of AI systems.

Based on the accounts of our interview subjects, commercial AI labs do not appear to use many public participation methods. Many of our interviewees admitted public participation is deprioritised within their companies and many cited numerous obstacles to embedding participation (discussed more below). Several were reluctant to share granular details about specific public participation projects in this study, likely due to apprehension about sharing commercially-sensitive company information. It is possible there is more public participation taking place in commercial AI labs than the researchers were able to surface, but the findings suggest it is more likely that these practices are not widely used or prioritised by commercial AI labs.

'The public doesn't meaningfully participate at [company name]. We don't even participate meaningfully. Like all we do is conduct this research and then we give vague recommendations, or consult with product teams.' Industry practitioner A

'I would have spent a lot more time on public engagement if I didn't already know that we had to make serious improvements on things like the truthfulness of our language model.' Industry practitioner F

The shortfall of public participation projects in commercial AI labs suggests that many of the programmes and initiatives spearheaded by these teams are facing enormous challenges when moving from ideation to practice.

97 Pratyusha Kalluri, 'Don't Ask If Artificial Intelligence Is Good or Fair, Ask How It Shifts Power' (2020) 583 Nature 169.

There are at present numerous obstacles to embedding public participation methods in commercial AI labs

We asked industry practitioner interview subjects what blockers or obstacles they face when seeking to conduct public participation initiatives. We also asked the public participation experts to speculate on the reasons why adoption of these methods in industry has been slow.

A range of different obstacles and concerns was cited by interviewees, with four broad types regularly cited as the most significant:

- Embedding public participation into projects requires considerable resources.
- Practitioners interested in public participation methods work in siloes and are removed from research and product teams.
- Practitioners are concerned about extractive or exploitative public participation practices.
- There is little incentive for commercial AI labs to share methods or learnings from experimentation with public participation.

Resource intensity

Interviewees reported a lack of necessary resources for running public engagement projects and struggling to fit these projects in tight project delivery timelines that disincentivised their use. Public participation projects of all shapes and sizes require significant resources and time to complete. The time and money involved in planning and delivery is characteristic of public participation projects regardless of sector or context.[98] This includes aspects of these projects like compensating participants.[99]

Practitioners working at AI labs reported that the speed of product development left little room for extra activities outside of the scoped product delivery:

98 'Costs of Public Participation' (*involve.org.uk*, 1 June 2018) https://involve.org.uk/resources/knowledge-base/what-impact-participation/costs-public-participation accessed 26 July 2023.

99 'Payment Guidance for Researchers and Professionals' https://www.nihr.ac.uk/documents/payment-guidance-for-researchers-and-professionals/27392 accessed 26 July 2023.

'In product teams, your goal is to ship product. And so in a lot of ways you're working against their incentives, but because at a really cut-and-dried level, their goal is to ship product.' Industry practitioner E

It was suggested by two interviewees, however, that the long history of user experience and user research practices in these teams might result in key actors and decision-makers being more receptive to 'participatory AI' activities, because activities such as gathering user feedback are analogous to some dimensions of public participation (specifically 'Consult' or 'Involve' dimensions on the *Participatory data stewardship* framework).[100]

Siloed practitioners and teams

Practitioners, particularly those from large companies, raised concerns that teams and individuals working on public participation methods might be siloed away from teams building products or research. Interviewees also noted that adopting public participation methods would require cooperation from multiple teams: for example, it might be policy/ethics/communications teams that identify a need for a public participation and contribute to the design and direction, but it might be research and development teams who would be expected to carry forward findings. This may also contribute to the resource intensity condition we outline above. A few practitioners suggested that it is often unclear who has the responsibility for setting up participatory projects in some companies:

'You might have a UX research team [conducting research with users], but you might have a team who wants to do research with policymakers. And so there becomes an organisational question of "whose job is it?".' Industry practitioner C

One participant, a practitioner in a large company, indicated that the extent to which practitioners can spearhead public participation projects in their company is dependent on their risk appetite:

100 Patel and others (n 22).

'If you're going to get dunked on for anything you do, [...] the game is just to stay out of the spotlight, because it doesn't really matter. You start to incentivise [...] basically doing zero risk-taking, becoming incredibly conservative in your approach.' Industry practitioner C

According to participants, siloing can create feelings of apathy or powerlessness among practitioners interested in 'participatory AI'.

Care and concern about extractive or exploitative practice

Another challenge cited by practitioners is avoiding practices that can attract accusations of participation washing.

'I think some [participatory AI practitioners] would say that some form of consultation is better than nothing and, you know, any feedback is valuable. I think that's true to an extent, but I'm really aware of [the] term 'participation washing'. I've been in lots of situations in the context of technology and AI where my presence in the meeting was taken by the powers that be as affirmation of what was going on.' Public participation expert A

Interviewees were aware of the concern that doing participation badly might further exacerbate harms caused to certain communities. One industry practitioner, when describing how the most common forms of engagement involve seeking quick input or feedback on a certain product, explains how this might complicate responsibility and accountability in a project:

'It's a really bad look that technology companies ship something that disproportionately hurts a particular community and then they rush over to the community to get help fixing it. And [from the community perspective] it's like, wait, you made this product, and now you need help fixing it! That's a terrible relationship to try to have with people.' Industry practitioner C

As evidenced in the literature, the risk of extractive or exploitative practice is live for public participation across a variety of contexts outside of commercial AI labs. [101,102,103]

However, two interviewees expressed that a commercial context, which prioritises profit, raises important considerations about power and decision-making that would have implications for the direction of any public participation activity:

'There's the concern about being exploitative in using the knowledge [from public participation projects] to do this sort of marketing veneer of responsible AI, and then we're still just going to make money on everything.' Industry practitioner A

'Given the position of power that we're in as a corporation, it's really important for us to acknowledge [that power] and bring folks in who don't have the same power, but who will be affected by how these technologies manifest in the world?' Industry practitioner B

As a result of these concerns, many of the practitioners we interviewed reported feeling apathetic toward the pursuit of adopting public participation approaches in industry, and some considered whether doing no engagement at all would be a preferable course of action to conducting projects that might cause risk of harm to communities.

A lack of incentives for conducting public participation research

Interviewees reported that commercial incentive structures often override incentives for conducting public participation work. One interviewee explained how corporate shareholder interests do not encourage or incentivise companies to experiment with public participation projects, particularly when this work might create more costs for a company:

101 Frances Cleaver, 'Paradoxes of Participation: Questioning Participatory Approaches to Development' (1999) 11 Journal of International Development 597.

102 Arnstein (n 29).

103 Ludo Glimmerveen, Sierk Ybema and Henk Nies, 'Who Participates in Public Participation? The Exclusionary Effects of Inclusionary Efforts' (2022) 54 Administration & Society 543.existing literature has shown that public participation often involves the co-optation of sympathetic citizens. In contrast, our study demonstrates that participatory advocates may discredit and marginalize critical voices despite their own inclusive, democratic ideals. We analyze the entangled legitimacy claims of participating citizens and "inviting" public-service actors, capturing (a

'[Company name] falls short of what we could consider public participation, but at the end of the day, they're really accountable to their shareholders and their customers, and so the fact that they're even considering these, you know, lukewarm public participation avenues is already a step.' Industry practitioner E

Commercial logics were described as underpinning many of the above obstacles: a desire for maximum profit means priorities more often lie with cost-effectiveness, efficiency and productivity, which might be antithetical to both a) the procedure of conducting public participation activities and b) the aims and objectives of members of the public and communities.

No incentive to share learnings

The competition dynamics at play within technology companies, and at the field level, were reported by interviewees as a contributing factor to why experimentations with public participation approaches are generally not made public.

One interviewee suggested they would be worried about public perception towards this work, which motivated a desire to not make experimentation public:

'It creates this really weird incentive structure where the people who engage [in public participation work] are signing up to very hard work that won't be supported and they may even be raked over the coals in the public sphere. So who's going to sign up to do that work?' Industry practitioner C

The lack of external transparency around what public participation approaches have been tested, trialled and iterated in commercial labs was put forward by one interviewee as potentially limiting the uptake of methods in industry for two related reasons: 1) a lack of evidence or norms around what 'good' practice might look like provides little direction for practitioners at the firm level, which in turn prevents 2) a phenomenon known as 'institutional isomorphism'[104] at the industry level – where

104 Paul J DiMaggio and Walter W Powell, 'The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields' (1983) 48 American Sociological Review 147.

companies observe and begin to mimic other practices from rivals in their industry.

The current AI climate presents challenges for embedding public participation

A final challenge our interviews identified is the difficulty of applying public participation methods to address current and emerging challenges in AI governance and development. In recent months, there has been considerable research attention towards foundation models and generative AI as the latest frontier in AI development (see 'Glossary: Foundation Models & Generative AI' on page 6). These models are designed to achieve generality of output and can be used as 'building blocks' for other applications.[105] At present, these models require considerable compute power to train and are characterised by their large size – number of parameters – and adaptability for a range of tasks.

The process of designing, developing and testing these models requires enormous amounts of labour. The data used to train them is annotated and labelled by human labellers: generally precarious workers, who are poorly paid and often subject to a lack of transparency about the nature of the work and of the technology company contracting it.[106] Notably, these systems are overwhelmingly developed in industry: for example, large multimodal model GPT-4 is developed by OpenAI, while Meta has developed LLaMa. Because of the ability for these models to give way to an extremely high number of different applications and be used in a wide variety of contexts, they could have numerous and wide-ranging potential impacts on people and society, many of which may be difficult to predict ahead of deployment.[107]

The scale, ubiquity and increasing popularity of foundation models with members of the public (ChatGPT, the chatbot built from GPT-3.5, is

105 'The Value Chain of General-Purpose AI' https://www.adalovelaceinstitute.org/blog/value-chain-general-purpose-ai/ accessed 17 February 2023.

106 Josh Dzieza, 'Inside the AI Factory' (Intelligencer, 20 June 2023) https://nymag.com/intelligencer/article/ai-artificial-intelligence-humans-technology-business-factory.html accessed 26 July 2023.

107 Rishi Bommasani and others, 'On the Opportunities and Risks of Foundation Models' (arXiv, 12 July 2022) http://arxiv.org/abs/2108.07258 accessed 26 July 2023."plainCitation":"Rishi Bommasani and others, 'On the Opportunities and Risks of Foundation Models' (arXiv, 12 July 2022

estimated to have over 100 million active users[108]), and the ever-growing supply chains for design and production, raise challenging questions about the role public participation can play in their governance. As noted above, OpenAI and other companies are experimenting with public participation methods for these models.

Our interviews reveal a concern that it is much more challenging to conduct public participation when the use case for a system or research is unclear, as it may be in the case of foundation models. The potential scale of use as well as both the ubiquity and the novelty of the technology were cited as concerns by one interviewee:

'What does it mean to engage people in systems like DALL-E and DALL-E mini [now Craiyon] as they start to go viral?' Industry practitioner D

We asked interviewees about their experiences implementing public participation in both product development contexts and either general AI research or foundation models/generative AI contexts. Practitioners considered that the AI product development context might offer fruitful avenues for feedback and engagement:

'Being in a product team context can be really focusing, right? Because we have these goals of the conversation. And you can be really clear with [participants] and get much clearer feedback from them.' Industry practitioner C

Best practice for public participation in other domains indicates that for deliberative public participation exercises, debating and deliberating on clearly defined policy problems, with extensive background information, facilitates decision-making and recommendations.[109] Supporting this view, one participant suggested that hinging participatory discussion off products that are already well-known or well-used offers more fruitful avenues for conversation around 'longer term societal issues'.

108 Krystal Hu and Krystal Hu, 'ChatGPT Sets Record for Fastest-Growing User Base - Analyst Note' Reuters (2 February 2023) https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/ accessed 30 June 2023."plainCitation":"Krystal Hu and Krystal Hu, 'ChatGPT Sets Record for Fastest-Growing User Base - Analyst Note' Reuters (2 February 2023

109 Julia Abelson and others, 'Deliberations about Deliberative Methods: Issues in the Design and Evaluation of Public Participation Processes' (2003) 57 Social Science & Medicine 239.

One interviewee raised concerns around participation in the context of foundation models, using a prior example of a consultation. They questioned the ability for any potential participants to effectively input or give feedback on a foundation model or generative AI system:

'For something like a medical device, you have this very specific intended use, where you don't for a general-purpose technology.' Industry practitioner A

With the pressure on AI labs to embed public participation already considerable, our interviewees were not confident about the ability for participation to be wrapped into foundation model development. This has implications for the future of public participation in the technology industry if trends continue toward extremely large systems with complicated use contexts. Public participation methods may not be appropriate for use in such contexts, and their use could lead to further claims of participation washing if the objectives of the exercise are not clear.

Areas for further input and remaining questions

This project surfaces novel empirical evidence of the use of public participation methods in commercial AI labs. Our findings do not point to clear recommendations for practitioners of public participation in commercial AI labs. However, our findings highlight several areas of research and further input that could help clarify how public participation in commercial AI labs can become a more meaningful and consistently used accountability practice.

In this chapter, we highlight three areas for further research to increase experimentation with participatory approaches and foster the conditions that might enable this experimentation to more readily take place:

1. Further trialling and testing of public participation approaches in industry 'in the open'.
2. Collaborative development of standards of practice for public participation in commercial AI labs.
3. Additional research into how public participation might complement other algorithm accountability methods or emerging regulation of AI.

1. Further trialling and testing of public participation approaches in industry 'in the open'

Problem statement

Practitioners interviewed in this study report a feeling of disconnection with other individuals and teams (both within their own labs and elsewhere in the industry). This finding suggests that more effort could be made to join up 'participation-interested' practitioners in a safe and comfortable forum to share learnings. As we outline above, there are at present strong disincentives for commercial AI labs to make public participation project trials public or to invest in resourcing and upskilling.

OpenAI's recent call for proposals for 'Democratic Inputs to AI'[110] in May 2023 is to date one of the most significant signals that industry is becoming more interested in experimenting with public participation methods in AI development. Initiatives like this may help to alleviate some of the fear of being the 'first mover' that causes risk aversion. However, further action and experimentation is needed.

What this might involve

Cultivating a robust ecosystem of public participation in the commercial industry requires the participation and cooperation of technology companies. Companies need to be incentivised to take part in these practices and both internal and external stakeholders, such as technology company practitioners and external researchers, should work together.

There are two potential ways this could be accomplished. First, policymakers and regulators could issue guidance for how tech companies can use public participation as an accountability practice. As outlined in the UK's white paper on AI regulation, the Government will set out a central set of principles for individual regulators to follow, which could include guidance on the use of public participation methods.[111] Governments could also make public participation a requirement in the public sector procurement process for AI technologies, and could set specific standards about technologies being developed with a minimum level of public participation. In future, transparency and open experimentation might be further incentivised through stricter policy guidance or statutory regulatory obligations.

> It is challenging at this juncture to make definitive recommendations for the shape of policymaker input, but initial policy guidance could be an interim measure to encourage the development of this norm.

110 'Democratic Inputs to AI' (n 9).

111 Department for Science, Innovation & Technology and Office for Artificial Intelligence (n 10).

Second, the establishment of a multi-stakeholder initiative around public participation methods could help create an industry norm around the use of these methods in the AI development and deployment process. A forum could create a space for industry practitioners to share their experiences of public participation (including failures) and build a community of practice around these methods. This initiative could be driven by an existing multi-stakeholder institution like the Partnership on AI, which has recently launched a Global Task Force on Inclusive AI – involving practitioners and researchers across academia, civil society, industry and policy – who are tasked with collectively establishing a possible framework to facilitate ethical and inclusive public engagement in AI.[112]

Tactically, this initiative could encourage industry practitioners to 'piggyback' on existing corporate practices in the AI development process and augment them in a way that involves more experimentation with public participation methods.[113] For example, red-teaming – a common practice in cybersecurity in which a team of internal or external experts attempts to break or challenge the security of an AI system – could be adapted to include more public participation. Similarly, bug bounties – in which external parties are encouraged to find vulnerabilities or problems in a system's functionality – could also include members of the public. Such initiatives often draw from a narrow group of participants that primarily include practitioners with technical skills, but as companies seek to diversify input into AI development,[114] there is room to expand participation to include other skills and lived experiences.

Potential research questions

- What are some potential levers to realign incentives between advocates of public participation and the motivations of the commercial technology industry?

112 'Global Task Force for Inclusive AI' (*Partnership on AI*) https://partnershiponai.org/global-task-force-for-inclusive-ai/ accessed 24 July 2023.

113 Wesley Hanwen Deng and others, 'Investigating Practices and Opportunities for Cross-Functional Collaboration around AI Fairness in Industry Practice', Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (Association for Computing Machinery 2023) https://dl.acm.org/doi/10.1145/3593013.3594037 accessed 28 June 2023.

114 OpenAI (n 1).

- How should commercial AI labs redefine and augment existing participatory activities in AI development so that they better serve people and society?

Who might be involved

- Existing practitioners of public participation in technology companies, including our interviewees and members of their teams, in tandem with other practitioners or executives in the technology industry
- The broader AI practitioner community: developers, designers, engineers and researchers, who might be motivated by questions of fairness or ethical practice in industry
- Technology companies that have already demonstrated interest in public participation in AI by publicly releasing details of participatory projects, including OpenAI, Meta and DeepMind
- Civil society, academic researchers, activists and members of the public to add external pressure and a supporting voice to practitioners interested in 'participatory AI'
- Policymakers and regulators who can add external pressure, and set guidance for companies to experiment with these methods

How this might help embed participatory approaches

- Further empirical evidence about effective approaches in different contexts will build a more comprehensive understanding of the limits and opportunities for public participation in commercial AI environments, and how the perspectives of people affected by AI are meaningfully embedded in their development.
- A culture of transparency will facilitate intuitive sharing of evidence on emerging good practice at the field-level, giving way to industry-wide norms of public participation practice.

2. Collaborative development of standards of practice

Problem statement

In this study, practitioners frequently reported the challenge of evaluating 'good' participation practice in these spaces, and a pragmatic awareness that commercial incentives might impede on the resources and time required to conduct a public participation project. If practitioners aren't able to meet a 'gold standard' of participation in the technology industry, what tools, tactics or approaches could be leveraged in the present conditions? Is there a 'minimum viable product' for public participation in industry? Without an extensive body of evidence in this emerging research area, these questions can be challenging to answer.

What this might involve

Addressing this challenge will require the development of public participation standards that outline what good practices look like. However, these standards must be developed through a multi-stakeholder process and cannot be led by industry. In this report, we present and discuss some of the risks of permitting the technology industry to exercise undue influence over ethical AI debates. To prevent corporate capture,[115] it is imperative that standards-setting for public participation in AI is driven by a broad coalition of civil society, policy and academic actors, as well as members of the public, particularly from marginalised groups, whose values and practices might reflect alternate priorities to those put forward by technology companies.

> It may also be fruitful for commercial AI labs to seek partnerships with civil society organisations or community groups when designing and executing participatory projects.

115 Whittaker (n 14).

This provides opportunity for direct collaboration around standards and practices for participation in industry, with a clearer line of influence and impact for external organisations.

Who might be involved

- To achieve broad and diverse input, bringing together a range of individuals, groups and organisations with expertise and background in public participation research, activism and community organising is critical. They might have prior engagement in applying this knowledge and experience toward questions of technology and society to offer in addition to experience convening and synthesising public perspectives. As part of the Ada Lovelace Institute's institutional commitment to ensuring decisions about AI are made with the views and experiences of members of the public, we are committed to driving forward this initiative.

- Organisations and groups with a history of collective bargaining tactics such as trade unions may also play a role in helping to establish standards around public participation with or on behalf of groups and individuals.

- Non-profit organisations in the field of data and AI have already made important contributions to the ongoing trajectory of 'participatory AI' and/or standards-setting around ethical and responsible AI. For example, the Collective Intelligence Project is currently conducting pilots around use of deliberation tools in technology companies and proposing models and frameworks of alternate governance for AI.[116] The Partnership on AI, the non-profit partnership of academic, civil society, industry and media organisations, has generated resources and policies geared toward guiding inclusive AI design.[117]

116 'Collective Intelligence Project - Whitepaper' (*The Collective Intelligence Project*) https://cip.org/whitepaper accessed 7 February 2023.

117 'Making AI Inclusive: 4 Guiding Principles for Ethical Engagement' (*Partnership on AI*, 20 July 2022) https://partnershiponai.org/paper/making-ai-inclusive-4-guiding-principles-for-ethical-engagement/ accessed 24 July 2023.

Potential research questions

- What is the role of standards bodies, civil society and other actors in driving the implementation of these methods in commercial AI?
- What lessons about meaningful practice can be gleaned from public participation practice in other domains?
- How might public participation in commercial AI be strengthened through industrial/civil society partnerships?

How this might embed participatory approaches

- It would ensure a coalition of actors can meaningfully contribute to and shape the trajectory of public participation in industry, instead of AI labs taking the lead.
- Standards might help to guide 'minimum viable' options for participation across a range of dimensions or modes (across the Participatory data stewardship spectrum, from 'Inform' to 'Empower'), further incentivising uptake, though they should not be considered an enforcement mechanism. Additional levers may be required to incentivise these standards.

3. Additional research into how public participation might complement other algorithm accountability methods and emerging regulation of AI

Problem statement

Recent AI policy and regulatory initiatives have called for external oversight and third-party evaluation of AI systems. For example, in 2023, the UK set out its pro-innovation approach to AI regulation, calling for diverse input into the execution of the regulatory framework and the use of mechanisms like algorithm audits.[118] Similarly, in the US, the

118 Department for Science, Innovation & Technology and Office for Artificial Intelligence (n 10).

National Institute for Science & Technology (NIST) AI Risk Management framework proposes risk assessment teams incorporating input from 'external collaborators',[119]and the White House voluntary commitments include a commitment to external red-teaming of systems to evaluate for safety.[120] EU's AI Act has explored a similar idea in the use of third-party conformity assessments for high-risk AI systems, as well as evaluation of foundation models by independent experts.[121] These proposals open an opportunity for participatory approaches to AI governance and policymaking.

Previous Ada research has explored whether public participation might function as an accountability mechanism, and facilitate oversight and scrutiny over AI systems.[122, 123]

> What remains underexplored is how public participation approaches can be embedded in other accountability mechanisms that are being proposed in AI regulatory approaches, such as algorithm audits and impact assessments.[124]

For example, Ada's NHS AI Lab algorithmic impact assessment includes a citizens' jury style methodology for patients and members of the public to deliberate on potential harms and benefits of proposed clinical AI systems. This approach might enable public participation methodologies to be designed for inclusion in existing proposals for AI oversight.

119 Elham Tabassi, 'AI Risk Management Framework: AI RMF (1.0)' (National Institute of Standards and Technology 2023) error: NIST AI 100-1 https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf accessed 24 July 2023.

120 House (n 16).

121 ConsolidatedCA_IMCOLIBE_AI_ACT_EN.pdf (europa.eu)

122 Groves and others (n 2).

123 Ada Lovelace Institute, AI Now Institute and Open Government Partnership, 'Algorithmic Accountability for the Public Sector' (Ada Lovelace Institute, AI Now Institute, Open Government Partnership 2021) https://www.opengovpartnership.org/documents/algorithmic-accountability-public-sector.

124 Inioluwa Deborah Raji and others, 'Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing' [2020] arXiv:2001.00973 [cs] http://arxiv.org/abs/2001.00973> accessed 12 March 2021.

What this might involve

Further research analysing the practical implementation of mechanisms such as impact assessments and audits, including their designed participatory components, to consider their potential for constructing accountability relationships between technology developers, deployers and procurers and people affected by the technologies.

Potential research questions

- If public participation were to be codified into law or policy for commercial actors developing AI, what would that look like?
- How should public participation methods interact with algorithm accountability mechanisms? What could a 'participatory' form of algorithm audit look like?
- How should public participation be implemented for foundation models, or models where the AI supply chain comprises multiple accountable actors?

Who would be involved

This initiative would require cooperation from policymakers, lawmakers and algorithm accountability researchers in academia, civil society, or industry, in order for mechanisms to be proposed, embedded or evaluated in step with both a) emerging public participation best practice and b) emerging AI accountability mechanisms.

How this might embed participatory approaches

Ada's work understands accountability as setting up an institutional mechanism between people affected by technologies and developers and procurers, to ensure AI systems are developed with proper oversight and scrutiny. It is essential that members of the public are invited to participate in this process, to pose questions and to pass judgement. At a moment where policy, industry and civil society actors are calling for both increased regulatory oversight of data-driven technologies and for wider public input into how these systems should

be designed, used and governed, there is opportunity for collective action into both policy asks simultaneously (which might otherwise become entrenched into distinct positions).

Our study highlights some participation methods and approaches, and we have set out how they could be applied to commercial AI, and where they might be used in the AI development pipeline

Conclusion

In this research, we have shed light on the state of 'participatory AI' in commercial AI labs and introduced the concept as one of importance to current conversations around ethical AI in AI development. Rather than attempt to set concrete normative ambitions for participation in industry or plot a path forward at this juncture, we have focused our efforts on plugging an evidence gap with some initial research.

Our study highlights some of the participation methods and approaches that are either currently in use across the sector or hold potential for use, and we have set out what some of what these approaches might accomplish for commercial AI, and where they might be used in the AI development pipeline. In addition, we have also set out some of the current challenges for public participation in commercial AI and current limitations of participatory research methods . In doing so, we provide clarity to researchers, practitioners and policymakers about what to expect in this emerging landscape. It is not our intention to dim the ambitions of motivated individuals and groups who wish to explore 'participatory AI'; rather, we hope these findings will generate further conversation and research into participation.

We have pointed to remaining questions and areas for further input in order to bring about 'more or better' participation,[125] including the involvement of civil society in standards setting for public participation in AI.

In all, through this research we hope to shape tone and terrain of ongoing debate around 'participatory AI' and influence the direction of public participation in commercial AI labs with evidence-led research.

125 Johannes Himmelreich, 'Against "Democratizing AI"' [2022] AI & SOCIETY https://link.springer.com/10.1007/s00146-021-01357-z accessed 3 August 2022.

> While participation is not a silver bullet, it is an important tool to ensuring data and AI debate and practice is in step with people's attitudes, opinions and concerns about technology.

By honing in on the sites driving major AI developments – commercial AI labs – we can contribute to explore public participation in a range of contexts and conditions, better driving understanding of how public participation in data and AI might generate benefits for people and society.

Methodology

To investigate our research questions, we adopted two research methods:

- A literature review
- Expert interviews

Our literature review conducted in summer 2022 explored the intersection of 'public participation' and 'commercial AI'. We analysed theories of participation from the fields of deliberative democracy, public policy and sociology, as well as conceptual theory around the role of participation in science, technology and design projects from the fields of science and technology studies (STS) and human computer interaction (HCI). We also reviewed institutional theory and scholarship centred around 'ethical AI', particularly within the commercial AI context.

We held 12 expert interviews in autumn 2022. We interviewed nine practitioners working in established and start-up commercial AI labs developing both AI products and AI research, who are interested in 'participatory AI', are involved in planning or implementation of public engagement/participation projects or who would be expected to carry forward findings of public participation projects into research and/or product development. For additional background, we also interviewed three subject-matter experts across participatory design, participatory AI and public engagement methods, and with knowledge of technology industry practice.

Our interview questions were split into three buckets to correspond with our project research questions:

- How do commercial AI labs understand public participation in the development of their products and research?
- What approaches to public participation do commercial AI labs adopt?
- What obstacles/challenges do labs face when implementing these approaches?

As an additional output for this project, the researchers authored a write-up of the project as an academic paper for submission in proceedings of the Association of Computing Machinery (ACM)'s Fairness, Accountability and Transparency in Machine Learning (FAccT) conference.

We call attention to two particular limitations of our research:

- Non-representative sample
- Barriers to participation

We regret that not every major commercial AI lab is represented in this study. We also spoke to people who have interest in or a stake in 'participatory AI', so our findings are not reflective of attitudes to participation from the broader practitioner population. We would have preferred to conduct more interviews to gain richer understanding of current practice: particularly for the largest organisations, speaking with individuals working in different teams across the organisation would have been useful in surfacing team-level versus organisational-level trends and initiatives.

We believe that a number of barriers prevented wider participation in this study. The first was an acute challenge in accessing the right people for interview: opaque technology company organisational structures create difficulties for even employees to make an assessment about who else within their organisation would have the requisite expertise. As a result, the researchers drew from mostly their existing industry networks. Many invited interviewees also turned down the request to participate, some explicitly citing burnout. Workers in large technology companies generally sign non-disclosure agreements, and even with researcher confidentiality we speculate that many practitioners felt too uncomfortable to divulge commercially sensitive practices to be able to participate.

Acknowledgements

This paper was lead authored by Lara Groves with substantive input from Andrew Strait, Aidan Peppin, Octavia Reeve, Jenny Brennan and Kira Allmann.

We would like to thank all our interview participants for their time and expert contributions to this study, both as individuals and representatives of organisations:

- Rachel Foley, Google DeepMind
- Lucia Komljen
- Tom Mason, Stability AI
- Meta
- Tina Park, Partnership on AI
- Antonia Paterson, Google DeepMind
- Irene Solaiman, HuggingFace
- Luke Stark, Western University
- And those who preferred not to be named.

About the Ada Lovelace Institute

The Ada Lovelace Institute was established by the Nuffield Foundation in early 2018, in collaboration with the Alan Turing Institute, the Royal Society, the British Academy, the Royal Statistical Society, the Wellcome Trust, Luminate, techUK and the Nuffield Council on Bioethics.

The mission of the Ada Lovelace Institute is to ensure that data and AI work for people and society. We believe that a world where data and AI work for people and society is a world in which the opportunities, benefits and privileges generated by data and AI are justly and equitably distributed and experienced.

We recognise the power asymmetries that exist in ethical and legal debates around the development of data-driven technologies, and will represent people in those conversations. We focus not on the types of technologies we want to build, but on the types of societies we want to build.

Through research, policy and practice, we aim to ensure that the transformative power of data and AI is used and harnessed in ways that maximise social wellbeing and put technology at the service of humanity.

We are funded by the Nuffield Foundation, an independent charitable trust with a mission to advance social well-being. The Foundation funds research that informs social policy, primarily in education, welfare and justice. It also provides opportunities for young people to develop skills and confidence in STEM and research. In addition to the Ada Lovelace Institute, the Foundation is also the founder and co-funder of the Nuffield Council on Bioethics and the Nuffield Family Justice Observatory.

Find out more:

Adalovelaceinstitute.org
@AdaLovelaceInst
hello@adalovelaceinstitute.org

Preferred citation: Ada Lovelace Institute. *Going public: Exploring public participation in commercial AI labs* (2023) https://www.adalovelaceinstitute.org/report/going-public-participation-ai/

ISBN: 978-1-7395236-0-2