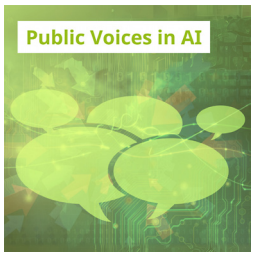
A Framework for Including Public Voices in AI

Reema Patel, Elgon Social &
Digital Good Network

elgon social research



ESRC Digital Good NETWORK

RAi

UKRI UK Research and Innovation

Table of Contents

Introduction and purpose

This report presents a conceptual framework and a practical workbook for including public voices in artificial intelligence (AI). It combines existing literature from the fields of AI ethics, the field of participatory and deliberative democracy, and the field of systems change. Drawing on these resources, it explores the critical role of public participation in shaping the development and deployment of AI.

AI comes with immense potential and opportunities, but also with significant risks and harms. Used in some contexts, AI's potential impacts are akin to those of nuclear power – and without a balanced approach to societal engagement, we risk creating the new 'Oppenheimer moment'.[1] It is important that governments and AI developers worldwide give as much consideration to potential hazards as they give to opportunities. At present, there is anxiety that commercial imperatives are being prioritised over societal concerns, resulting in the public and policymakers bearing the risks of AI development, while commercial interests gain the majority of the benefits and opportunities. This report demonstrates how participation in the field of AI can reset this balance so that opportunities are maximised for all stakeholders and risks are mitigated.

As previous debates about emerging science and technology have illustrated, we must engage in adequate consideration of what society expects from AI technologies before embarking on their use. We have learned these lessons before, with key examples including the debates around GM crops[2] and nuclear.[3] In both instances, a failure to engage directly with societal norms and values, combined with a top-down, communications-centric and patronising approach, resulted in backlash, increased risk and harm, and unrealised opportunities.

In the field of AI, we can already see numerous examples of such failures. In San Francisco, the use of facial recognition technology was banned due to inadequate public consultation and buy-in.[4] In Toronto, a smart city scheme was halted for similar reasons.[5] In the UK, efforts to roll out the Ofqual education algorithm met with considerable opposition, resulting in an embarrassing reversal for the UK government.[6] The digital contact tracing app cost the public taxpayer £35 million with limited uptake in the pandemic due to inadequate early engagement.[7] In response to a recent government consultation, the UK's national newspapers launched a campaign for a fair and just approach to the use of generative AI on copyright.[8]

---

1 Baig, M.A. and Khan, A.A. (2024) 'AI's Oppenheimer moment: Establishing regulations', *The Interpreter*, 16 January. Available at: https://www.lowyinstitute.org/the-interpreter/ai-s-oppenheimer-moment-establishing-regulations (Accessed: 13 March 2025).

2 Marchant, R. (2001) 'From the test tube to the table: Public perception of GM crops and how to overcome the public mistrust of biotechnology in food production', *EMBO Reports*, 2(5), pp. 354–357. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC1083893/ (Accessed: 13 March 2025).

3 American Nuclear Society (2017) 'Public engagement', *Nuclear Grand Challenges*. Available at: https://www.ans.org/challenges/engagement/ (Accessed: 13 March 2025).

4 Conger, K., Fausset, R. and Kovaleski, S.F. (2019) 'San Francisco bans facial recognition technology', *The New York Times*, 14 May. Available at: https://www.nytimes.com/2019/05/14/us/facial-recognition-ban-san-francisco.html (Accessed: 13 March 2025).

5 Jacobs, K. (2022) 'Toronto wants to kill the smart city forever', *MIT Technology Review*, 29 June. Available at: https://www.technologyreview.com/2022/06/29/1054005/toronto-kill-the-smart-city/ (Accessed: 13 March 2025).

6 University College London (2020) 'UCL welcomes government's reversal on A-level algorithm use', *UCL News*, 17 August. Available at: https://www.ucl.ac.uk/news/2020/aug/ucl-welcomes-governments-reversal-level-algorithm-use (Accessed: 13 March 2025).

7 National Audit Office (2021) 'Test and trace in England – progress update', *National Audit Office*. Available at: https://www.nao.org.uk/reports/test-and-trace-in-england-progress-update/?nab=1 (Accessed: 13 March 2025).

8 Landi, M. (2025) 'UK newspapers launch campaign against AI copyright plans', *The Independent*, 25 February. Available at: https://www.independent.co.uk/news/uk/politics/government-daily-express-the-sun-daily-star-daily-mail-b2704106.html (Accessed: 13 March 2025).

The costs of these unpopular schemes represent significant amounts of time, public funding, energy and resources, and these losses are borne by the public and policymakers. These examples show that the social contract for AI – worldwide and in the UK – is not stable or robust, and that it is further unravelling. If this trend continues, policymakers and regulators will carry on bearing the risk of AI decisions, the public will bear the harms and AI developers will bear the commercial costs, as trust and confidence diminish. This is bad for society and for the future of AI.

So, where next? Policymakers, innovators and regulators need to consider the question of how society can shape AI in its interest. We propose a framework to aid thinking about how best to include public voices in AI. By creating an ecology of public participation in AI, this powerful technology can be aligned with public norms and values so that it works for everyone. This means listening to participation on AI where it is already in use and enabling participation on AI on issues where it has not yet been taken up, and connecting participation insights with AI development, regulation and policymaking.

We also need to overcome some of the existing barriers to societal input and inclusion. Policymakers and regulators play a key role in encouraging responsible AI. They are in a position to push back against the 'move fast, break things' mantra that prioritises technology development at a fast rate of change, in contrast to more sustainable, durable ecosystems of development that serve society's needs.[9] Policymakers, regulators and AI developers must be supported by the right governmental, economic, regulatory and policy incentives – those that reward alignment with public and societal values, instead of short-term gains.

Getting inclusion and participation in AI right is an urgent priority, and not a 'nice to have'. This report calls for a transformative, power-literate approach to how we understand, design and implement participation in how AI is shaped and governed. It also engages with the challenges of fostering meaningful inclusion – such as resource constraints, competing incentives, and systemic inequalities, alongside the powerful instincts of existing commercial and public sector organisations to preserve the status quo ('the system immune response'). The report makes three broad recommendations:

1. Build and invest in broad coalitions for change and communities of practice
2. Scale and invest in disruptive participatory innovations as catalysts, using the framework for including public voices in AI
3. Tackle the social and commercial barriers to participatory and inclusive approaches to AI

The report also introduces a self-assessment workbook which can be completed in a stand-alone format. This is a practical tool which organisations, practitioners and technologists can use to critically evaluate and enhance their approaches to participation. The workbook supports practitioners, collectives, organisations and institutions to move beyond tokenistic efforts toward practices that

9 Gallo, C. (2019) 'Why "move fast and break things" doesn't work anymore', *Harvard Business Review*, 3 December. Available at: https://hbr.org/2019/12/why-move-fast-and-break-things-doesnt-work-anymore (Accessed: 13 March 2025).

are genuinely collaborative, equitable and empowering, and address the needs of those most impacted by AI systems. This approach sets the stage for a public culture of collaboration in AI design and governance, ensuring that these technologies reflect shared societal goals and promote a just and inclusive future.

What is a framework and why do we need one?

The framework set out within this report is for building AI systems that incorporate social values. Frameworks help us navigate a set of complex issues and problems. They seek to provide conceptual clarity, aid deep thinking, and provide practical assistance and guidance in navigating challenges.

This framework aims to make sense of how the practice of participation can shape AI. Informed by the thinking of US health policymaker and Vice President of the US National Health Council Sherry Arnstein and the International Association for Public Participation (IAP2), it brings together how we (currently) think about participation, how we think about AI, and our understanding of how systems change happens, into a coherent whole. Set out in Section 9 of this report, it seeks to encourage practitioners, policymakers, developers and the public to *design, evaluate, regulate, and operate AI systems in ways that benefit people, society and the nation*, through the lever of participatory practice.

A framework grows gradually less important as people master their work. This framework aims to prompt and to ask challenging questions. It is for developers, policymakers, regulators, participatory practitioners and the public to learn, respond and fill in the blanks.

Even when AI system builders, regulators and policymakers want to include the public, we know this can prove challenging for a variety of reasons. A critique of public participation in emerging technology policy is that it can be difficult, resource intensive, and time consuming to enable and to foster, given the complexity of the issues at hand. There is often limited political and commercial will to include those most excluded in key decisions. A further obstacle emerges due to the fact that AI is a complex field, moving at a fast rate of change.

Furthermore, at the time of writing, the so-called 'participatory turn' in AI is still a new and emerging field. There remains limited expertise, and that expertise is siloed.[10] This report argues that, despite the hype, the 'participatory turn' itself continues to remain far from widespread AI practice. Participation is still seen as a 'nice to have', rather than business as usual. Some researchers argue that it can be seen as more of a turn on a 'side road', rather than the highway.[11]

There already exist many resources, tools and workbooks for including the public in shaping complex and controversial technologies. These build on a rich tradition of public participation in science,

10 Bherer, L., Dufour, P. and Montambeault, F. (2016) 'The participatory democracy turn: an introduction', *Journal of Civil Society*, 12(3), pp. 225–230. Available at: https://www.tandfonline.com/doi/full/10.1080/17448689.2016.1216383 (Accessed: 13 March 2025).

11 Davies, T. and Tennison, J. (2025) 'The Participatory Turn: Side Road or Highway?', *TechPolicy.Press*, 26 February. Available at: https://www.techpolicy.press/the-participatory-turn-side-road-or-highway/ (Accessed: 13 March 2025).

economics, climate and technology. For example, in the UK, the Sciencewise programme has been fostering engagement between policymakers and the UK public on these questions since 2004, offering advice and support.[12] At the Ada Lovelace Institute,[13] there are the beginnings of a framework for data stewardship in 2021, Participatory Data Stewardship. This draws on Arnstein's ladder of citizen participation, which calls for approaches that increase people's agency, power and control over decision making.[14] Since the publication of that paper, the emergence of developments in the field of generative AI, and the creation of a new field of participatory AI governance, the field has advanced further, and it is time for some updated thinking.

Despite my own involvement in authoring this early report at the Ada Lovelace Institute, it is clear that much remains unsaid. At present, the biggest issue in the field is not the lack of resources and methodologies on how to design and run participatory processes. The biggest issue is the lack of an appropriately critical and power-literate lens to support the adoption, use and scaling of participatory methods in shaping AI in ways that truly shift power. Historically, there has been heavy reliance on embedding participatory processes within policy and industry organisations as a lever to scale participatory practice. Therefore, this report moves beyond the framework and the nuts and bolts of how to think about participatory practice. It also takes in the question of how to ensure that participatory and inclusive practice changes outcomes and makes systems change feasible.

The accepted wisdom in many participatory circles is that for a participatory process to have impact or influence, it must be commissioned or hosted by a policymaking or power-holding organisation. For instance, the Organisation for Economic Co-operation and Development (OECD) report on the deliberative wave focuses almost entirely on acceptance by policymakers as a key lever or vehicle for change.[15] There has also been limited attention to questions of equity and inclusion, and how these questions can be addressed in the context of good participatory practice. A new EU Horizon Europe programme, INSPIRE, is currently engaging extensively with these issues, making the case for more inclusive and intersectional spaces for participation. It recognises that change can be effected at the grassroots level and not just within technology development or policy circles.[16] In some instances, change may actually be more likely to take place through grassroots action. There is a risk that participatory processes that are closer to power will become co-opted back into existing power structures. We intuitively know this from historical examples of the many great social movements that have transformed political outcomes and led to major social shifts. The suffragette, civil rights, anti-apartheid, LGBT rights and colonial independence movements are all examples of how social movements have interacted with institutional and policy change. More recently, the work of Extinction Rebellion has been critical in creating a mainstream dialogue around citizen assemblies, ensuring that they have been adopted at scale across local government in the UK.[17] Extensive research suggests that the most effective social movements have successfully combined grassroots action with support from elites.[18]

As a consequence, many leading practitioners in the field argue that focusing exclusively on institutionalisation does not engage with how change through participation actually happens.[19] If including public voices in AI is to change the existing system, we cannot assume that embedding participatory processes within policy or technology development circles will do the job. A more power-literate theory of change will need to recognise that some institutions and developers may not respond to public voices in AI unless there is a change in narratives, incentives, regulation and pressure. It will also need to recognise that the public can influence change with or without institutional or developer support. Moreover, it must acknowledge that there is a significant gap between policy and implementation where policies do not always translate into effective action on the ground.[20] When considering how we include public voices in AI, there is a need to understand the full range of strategies for influence.

A framework for including public voices in AI

The below framework is based on Sherry Arnstein's 'ladder of citizen participation', developed in the late 1960s, which was, in turn, transformed into the IAP2's spectrum of participation.[21] The spectrum of citizen participation adapted from Arnstein and IAP2 was initially adopted in the Ada Lovelace Institute's *Participatory Data Stewardship* report (2021), which I authored.

In the context of AI, the levels of engagement can be understood as a progression from informing stakeholders about AI systems, to consulting them on their impacts, involving them in design, development, iteration and assurance processes, collaborating with them in decision-making, and ultimately empowering them to shape AI governance and development.

12 op cit

13 Patel, R. (2021) *Participatory Data Stewardship*. Ada Lovelace Institute. Available at: https://www.adalovelaceinstitute.org/wp-content/uploads/2021/11/ADA_Participatory-Data-Stewardship.pdf (Accessed: 13 March 2025).

14 Arnstein, S.R. (1969) 'A Ladder of Citizen Participation', *Journal of the American Institute of Planners*, 35(4), pp. 216–224. Available at: https://www.tandfonline.com/doi/abs/10.1080/01944366908977225 (Accessed: 13 March 2025).

15 Česnulaitytė, I. (2020) 'Innovative Citizen Participation and New Democratic Institutions: Catching the Deliberative Wave', *OECD*. Available at: https://www.oecd.org/en/publications/innovative-citizen-participation-and-new-democratic-institutions_339306da-en.html (Accessed: 13 March 2025).

16 Bussu, S. and Geissel, B. (2024) 'INSPIRE – Intersectional Spaces of Participation: Inclusive, Resilient, Embedded', *European Commission Community of Practice on Participatory and Deliberative Democracy*. Available at: https://cop-demos.jrc.ec.europa.eu/citizen-engagement-projects/inspire-intersectional-spaces-participation-inclusive-resilient (Accessed: 13 March 2025).

17 Smith, G. (2021) 'Citizens' assemblies are helping to fight the climate crisis', *University of Westminster*. Available at: https://www.westminster.ac.uk/research/impact/citizens-assemblies-are-helping-to-fight-the-climate-crisis (Accessed: 13 March 2025).

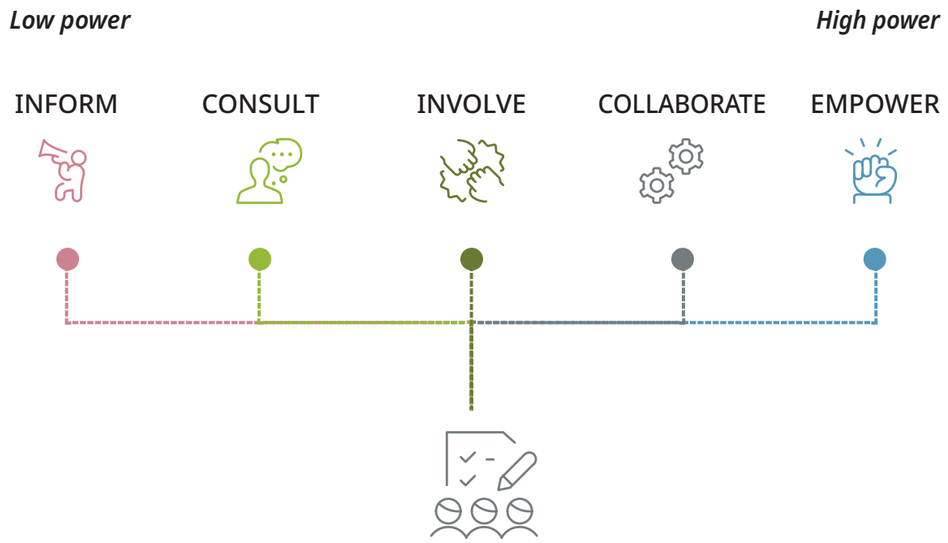
18 Institute of Medicine (2014) *Supporting a Movement for Health and Health Equity: Lessons from Social Movements: Workshop Summary*. National Academies Press. Available at: https://www.ncbi.nlm.nih.gov/books/NBK268722/ (Accessed: 13 March 2025).

19 Wilson, R. and Mellier, C. (2024) 'Getting Real About Citizens' Assemblies: A New Theory of Change for Citizens' Assemblies', *European Democracy Hub*. Available at: https://europeandemocracyhub.epd.eu/getting-real-about-citizens-assemblies-a-new-theory-of-change-for-citizens-assemblies/ (Accessed: 13 March 2025).

20 Goodwill, M., van der Bijl-Brouwer, M. and Bendor, R. (2021) 'Beyond Good Intentions: Towards a Power Literacy Framework for Service Designers', *International Journal of Design*, 15(3). Available at: https://www.ijdesign.org/index.php/IJDesign/article/view/4120 (Accessed: 13 March 2025).

21 International Association for Public Participation (IAP2), 'IAP2 Spectrum of Public Participation'. Available at: https://www.iap2.org/page/spectrum (Accessed: 13 March 2025).

Figure 1: Spectrum of citizen participation, originally drawn from IAP2 and Arnstein's ladder



Source: Spectrum adapted from Sherry Arnstein (1969) on 'A Ladder of Citizen Participation' and IAP2's spectrum – image source is the Ada Lovelace Institute (2021), and it was developed by Reema Patel, working with Octavia Field Reid.

While almost all organisations at the time of writing stop short of full empowerment, many advocate for increased public involvement and collaboration in shaping AI's societal role. Central to this framework is the Arnsteinian steer that a central goal of participatory policy making is that participation successfully increases levels of agency and power for those who participate. This is also a clear measure of its success.

This framework sets out the mutually complementary approaches to citizen participation – with examples that are feasible for all actors in society – that might support these increased levels of agency and power. Together, these approaches demonstrate the range of innovations that exist in the field of participatory and inclusive AI. They also illustrate that there are different ways to enable participation in shaping AI. These represent the possible outcomes and goals of informing, consulting, involving, collaborating and empowering people. They ascribe increased levels of agency and power to citizens in the process of shaping AI systems.

Figure 2: Framework for including public voices in AI:

Purpose	Description (shaped by Arnstein's ladder)	What people can expect from AI power holders – policymakers and developers (shaped by Arnstein's ladder)	Relevant mechanisms, methods that could suit (informed by Participatory Data Stewardship, Ada Lovelace Institute)	Examples of including public voices in AI
Informing: *Informing people about AI involves a one-way flow of information from those who develop, use and design AI, to 'beneficiaries'. This information flow can be direct or indirect.*	'A one-way flow of information'	*'We will keep you informed on how AI is being used'*	Communication and transparency Explainability – the process of enabling a data-driven system to be explained in human terms Mechanisms such as model cards and data sheets Rethinking and reframing how we talk about AI, so it is more accessible and inclusive	UK government, Department for Science, Innovation and Technology, AI algorithmic transparency record Global AI narratives project
Consulting: *Consultation can take place with individuals, groups, networks or communities, to enable people to voice their concerns, thoughts and perspectives. Consultation activities can take a range of forms but often involve the use of quantitative and qualitative methods including public-opinion 'attitude' surveys, neighbourhood meetings and public hearings.*	'Inviting people's opinions, through attitude surveys, neighbourhood meetings and public hearings'	*'We will listen to, acknowledge concerns and aspirations, and provide feedback on how public input influenced AI'*	Community networks User experience (UX) testing and co-design Surveys and public attitudes research Community engagements and consultations	Information Commissioner's Office public attitudes surveys Digital Identities public dialogue sandbox CityVerve community engagement and consultation, Manchester
Involving: *The process of 'involvement' positions beneficiaries in an advisory role, which helps inform early-stage decision-making by data stewards and fiduciaries. These initiatives convene non-specialist beneficiaries alongside specialists and stakeholders, with a view to informing key moments in the public policy landscape. They seek to advise and better inform government and regulatory bodies on the conditions for the acceptability of uses of AI.*	'Allow citizens to advise, but retain for power holder the continued right to decide'	*'We will work with you to ensure your concerns and aspirations are directly reflected in AI... We will provide eedback on how public input influenced these decisions'*	One-off and institutionalised public deliberation and deliberative democracy initiatives Lived experience panels Horizon scanning, design thinking and futures thinking	Institutionalised public deliberation, the Danish Board of Technology, 1986–2011 Ada Lovelace Institute deliberative body, the Citizen Biometrics Council Future of Flight Innovate UK dialogue

Purpose	Description (shaped by Arnstein's ladder)	What people can expect from AI power holders – policymakers and developers (shaped by Arnstein's ladder)	Relevant mechanisms, methods that could suit (informed by Participatory Data Stewardship, Ada Lovelace Institute)	Examples of including public voices in AI
Collaborating: *The process of collaboration in the context of public involvement in data and AI can be understood as enabling people to negotiate and engage in trade-offs with powerholders and those governing AI about specific aspects of decision-making. Collaboration also refers to co-creation and co-design with the public.*	'Enables people to negotiate and engage in trade-offs with powerholders'	*'We will look to you for advice and innovation in design of data and AI governance frameworks and incorporate your advice and recommendations to the maximum extent possible.'*	Bottom-up 'AI/data governance initiatives' managed by an independent fiduciary (e.g. trusts, and data-sharing contracts that build in collaboration)	Choral Data Trust Experiment with Serpentine Gallery and Alan Turing Institute People's Panel on AI, Public Voices in AI people's panel advisory group Liverpool Civic Data Cooperative Residents assembly on data and AI Sciencewise programme, NHS AI Lab and ScotGov juries (up to present day) NHS Waiting List prioritisation (Coventry and Warwickshire NHS Trust and Strategy Unit)
Empowering and / or empowerment: *Empowering people to shape AI enables them to exercise full managerial power and agency, and take responsibility for exercising and actively managing decisions about AI governance – specifically, how it is developed, designed, governed and used. In this model, the dynamic of power is shifted away from the AI developer and designer towards the public and communities impacted, who shape the decision or make the decision, advised where necessary by appropriate specialist expertise. In some instances, empowerment can simply refer to the realisation on the part of communities that they need to exercise their rights to self-determination of AI, in relation to an existing use case operating without their input.*	'Citizens obtain the majority of decision-making seats or full managerial power in shaping AI and/ or otherwise self-determine their own rights to shape AI issues in a bottom-up way'	*'We will provide advice and assistance as requested in line with your decisions for designing/ developing your own AI uses/applications' and/or 'We wish to retain for ourselves the right to deliberate about and self-determine our own views on a potential application of AI'*	AI governance rules shaped and routinely reviewed by beneficiaries Voting on governance boards of AI initiatives Ownership and/ or control of AI co-operatives Setting terms of AI licensing, use and access, with permissions overseen by citizens Assessing the impact of AI on diverse communities	A number of Public Voices in AI-funded projects are self-determined. These are: Workers' Observatory Migrants' Rights Network AI Under Watch project The People Speak, Knowing AI, Knowing u Unjust UK C.I.C. At the time of writing, and to my knowledge, no AI developer/ policymaker has handed over managerial power to citizens in shaping AI There is scope for further experimentation and innovation in this space

Why use the framework? The case for including public voices in AI

The public's concerns about AI are increasing

A recent UK survey undertaken by the Ada Lovelace Institute and the Alan Turing Institute on public attitudes to AI reveals a growing awareness of harms, with many individuals both self-reporting and experiencing negative impacts from emerging technologies.[22] On average, a third of the UK public (34%) report that they have encountered some form of AI-generated harm a few times, while a fifth (20%) report having encountered these harms many times. The most common harms reported are false information (61%), financial fraud (58%) and deepfakes (58%). Advances in AI raise concerns about its potential use by malicious actors for perpetrating these harms, especially as it enables these actors to generate malicious content more efficiently.

Given these risks, the public supports a multi-stakeholder approach to AI safety: 58% believe that both an independent regulator and AI companies should be responsible for ensuring AI is used safely. Trust in, and expectations of, companies versus regulators differs based on age. While younger people (18 – 44) predominantly say that AI companies should be responsible for the safe use of AI, those over 55 prefer regulators. The majority of the public (72%) indicate that laws and regulations would increase their comfort with AI technologies – an increase from 62% in our previous survey. This rise in demand for laws and regulation comes at a time when the UK does not have its own set of comprehensive regulations around AI.

This survey finds that public concern about AI is rising, yet perceptions of its benefits remain stable, suggesting a widening gap between risks and rewards. Notably, those who are most excluded from technological developments express the greatest concern, highlighting deepening inequalities in digital confidence and access. There is a strong public appetite for law and regulation to adapt more effectively to new technologies, providing protection while enabling innovation. However, attitudes vary significantly depending on specific use cases, underscoring the importance of context in shaping public opinion. A one-size-fits-all approach to governance is unlikely to succeed, as different technologies elicit different concerns, expectations and levels of trust across diverse communities.

There is also a strong theme of division, polarisation and ambivalence around the uses of AI. This is not specific just to the UK but is a global phenomenon. A recent Ipsos Global Survey on AI finds that globally people are both excited and nervous about AI. Fifty-three percent say they are excited for products and services that use AI, while 50% say AI makes them nervous. Excitement is highest in Asia while the Anglosphere and Europe are the most sceptical about AI.[23]

22 Modhvadia, R., Sippy, T., Field Reid, O., Margetts, H. (2025). How do people feel about AI? Wave 2 of a nationally representative survey of UK attitudes to AI designed through a lens of equity and inclusion. Ada Lovelace Institute and The Alan Turing Institute. Available at: https://attitudestoai.uk/

23 Carmichael, M. and Stinson, J. (2024) 'The Ipsos AI Monitor 2024: Changing attitudes and feelings about AI and the future it will bring', *Ipsos*. Available at: https://www.ipsos.com/en/ipsos-ai-monitor-2024-changing-attitudes-and-feelings-about-ai-and-future-it-will-bring (Accessed: 13 March 2025).

The rise of the 'AIcracy' or 'AI technocracy' in the place of democracy

The increasing use of AI and data-driven technology in almost all aspects of public life and public service demonstrates how AI is now influencing work, life, healthcare justice, social services and even arts and culture institutions.[24] Workplaces are rapidly developing large language models (LLMs) and companies are working on the diffusion of AI across industries, seeking to upskill employees to use AI across supply chains.

In day-to-day life, we use AI-driven technologies such as GPS, smartphones and real-time tracking, AI-based consumer products (through media and television) and AI-driven robotics in the home, often without critical awareness of the technologies involved. Within healthcare, AI technologies are being used to prioritise NHS waiting lists, triage patients, book appointments and allocate vaccines during health crises. Across social services, AI technologies have been embedded into decision-making about visas, and AI is also used in social care to provide rapid decision-making about cash-strapped public services.

Arts and culture institutions are using AI to enhance the viewer's experience at exhibitions, and yet are grappling with the ethics of using algorithms, rather than humans, to produce creative content. Educationalists, such as universities and schools, are similarly interrogating the appropriate use and government of AI in education, with stark implications for young students. Many of these decisions about the use and development of AI are made without appropriate societal input or consideration and are largely driven by the logic of consumerism and questions of productivity and efficiency. Every decision, process or experience is potentially now within the purview of analysis, quantification and potential influence by AI algorithms. This is often called datafication: the quantification of human life for value.[25]

Accelerating datafication serves to create the illusion of objectivity, which is used to justify the removal of public debate, public culture and societal considerations.[26] This loss is exacerbated by the fact that AI systems are often based on opaque algorithms and datasets, which are presented as objective and scientific, but which potentially reflect selection biases relating to human data sets, and narrow or biased perspectives. The result is that structural inequalities are reinforced.[27] This might not be easily understood or challenged by the public.

Furthermore, this situation presents particular challenges for inclusivity and public participation. As commercial developers and technical specialists present AI as a complex, technical domain, an 'AIcracy' emerges. AI is used to justify and enforce a particular worldview, while simultaneously excluding the public from understanding or challenging the underlying assumptions and biases of AI systems.[28] It sustains a power imbalance where those who control and understand AI wield significant influence, while the rest of society is left in the dark. The AIcracy effectively excludes the public from meaningful participation in decisions that affect their lives.[29] The public are often left feeling they don't 'own the language', imperatives and use of AI – the very opposite of the empowerment that Arnstein envisions.

Rethinking the 'deficit model' of technology development

Understanding that AI systems are developed by those who have **positionality and privilege** requires recognising the social and cultural contexts in which these systems are developed. AI creators, regulators and policymakers often come from privileged backgrounds with access to infrastructures, technologies and skill sets not available to the general public.[30] These can influence the priorities and perspectives embedded in both AI and its governance. This starting point shapes who is included or excluded in AI development from the very beginning. It is necessary to ensure that policy and technology development serves all walks of life and society, not simply the priorities of a smaller group, embedded within the technologies and their governance.

The *inclusion* of public voices and perspectives in AI thus requires us to take a multifaceted and collaborative approach. Creating responsible approaches to AI – that include the public – will also require us to take a 'socio-technical' approach, recognising that technologies themselves operate within complex environments, and create complex interactions between organisations, people, systems and activities, as part of wider society.[31]

Historically, scientists and technologists have adopted what can be described as a 'deficit model' of information flows, whereby technologists and scientists are positioned as 'on top', rather than as 'on tap', as in a 'dialogue model'.[32] In Figure 3, we can see that according to the deficit model, experts are positioned as filling the 'empty vessel' of the minds of the public with their technical knowledge and expertise. This assumes that society needs to understand the right technical information about AI to come to the 'right' conclusions about AI. This dynamic is explored and illustrated by a UCL survey of AI researchers, which finds that the majority perceive policymakers and the public as lacking understanding of AI.

24 Equality and Human Rights Commission (2025) 'Artificial intelligence in public services'. Available at: https://www.gov.uk/data-ethics-guidance/artificial-intelligence-in-public-services (Accessed: 13 March 2025).

25 Mikalef, P., Krogstie, J., Pappas, I.O. and Sidorova, A. (2023) 'Infosphere, Datafication, and Decision-Making Processes in the AI Era', *Journal of Technology and Innovation*, 23(4), pp. 345–360. Available at: https://link.springer.com/article/10.1007/s11245-023-09919-0 (Accessed: 13 March 2025).

26 Coyle, D. and Weller, A. (2020) '"Explaining" machine learning reveals policy challenges', *Science*, 368(6498), pp. 1433–1434. Available at: https://www.ijdesign.org/index.php/IJDesign/article/view/4120 (Accessed: 13 March 2025).

27 Patel, R., Field Reid, O. and Strait, A. (2021) 'How does structural racism impact on data and AI?', *Ada Lovelace Institute*. Available at: https://www.adalovelaceinstitute.org/blog/structural-racism-impact-data-ai/ (Accessed: 13 March 2025).

28 Dizikes, P. (2023) 'How an "AI-tocracy" emerges', *MIT News*, 13 July. Available at: https://news.mit.edu/2023/how-ai-tocracy-emerges-0713 (Accessed: 13 March 2025).

29 Sætra, H.S. (2020) 'A shallow defence of a technocracy of artificial intelligence: Examining the political harms of algorithmic governance in the domain of government', *Technology in Society*, 62, p. 101283. Available at: https://www.sciencedirect.com/science/article/pii/S0160791X19305925 (Accessed: 13 March 2025).

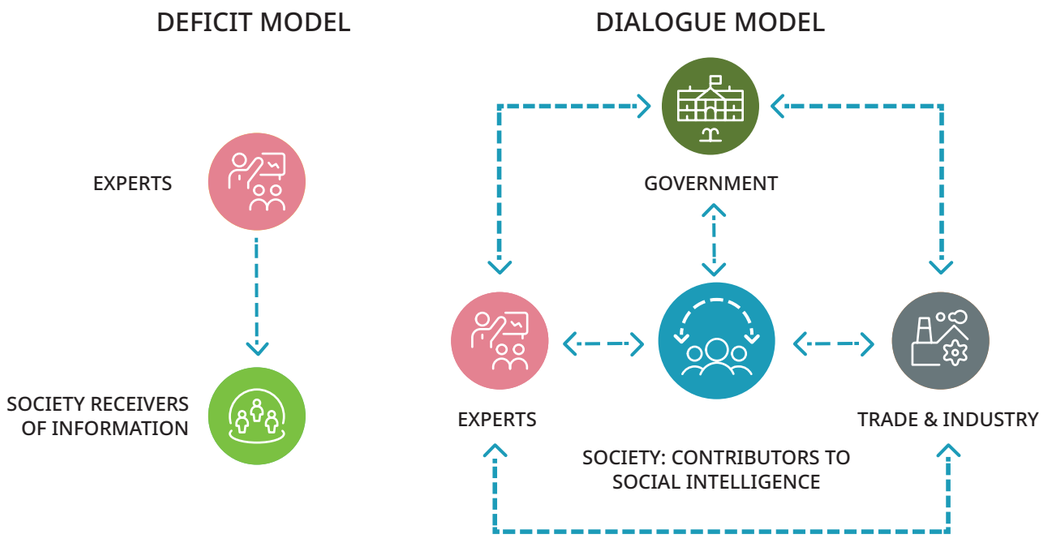
30 Crawford, K., Dobbe, R., Dryer, T., Fried, G., Green, B., Kaziunas, E., Kak, A., Mathur, V., McElroy, E., Nill Sánchez, A., Raji, D., Rankin, J.L., Richardson, R., Schultz, J., West, S.M. and Whittaker, M. (2019) *AI Now 2019 Report*. AI Now Institute, New York University. Available at: https://ainowinstitute.org/wp-content/uploads/2023/04/AI_Now_2019_Report.pdf (Accessed: 13 March 2025).

31 Salwei, M.E. and Carayon, P. (2022) 'A Sociotechnical Systems Framework for the Application of Artificial Intelligence in Health Care Delivery', *Journal of Cognitive Engineering and Decision Making*, 16(4), pp. 194–206. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC9873227/ (Accessed: 13 March 2025).

32 Reincke, C.M., Bredenoord, A.L. and van Mil, M.H.W. (2020) 'From deficit to dialogue in science communication: The dialogue communication model requires additional roles from scientists', *EMBO Reports*, 21(9), p. e51278. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC7506985/ (Accessed: 13 March 2025).

On the right-hand side, however, the dialogue model recognises that expertise resides with a broad range of actors, not just the technical experts. In the dialogue model, the need for dialogue between a wide range of actors stems from the fact that knowledge itself takes different forms. For example, knowledge can encompass an understanding of lived experience of AI, society, policy, regulation, markets and technology. In the dialogue model, we are better able to pool our knowledge, and this improves social and technological outcomes.

Figure 3: Moving from a deficit model to a dialogue model



Courchamp, F. et al., (2017) Invasion Biology: Specific Problems and Possible Solutions, Trends in Ecology & Evolution.

Not only has the deficit model approach failed the fields of technology and scientific development as a whole, it has also failed the communities that these fields seek to serve. Some of the challenges that come with the deficit model approach are well known and established. These include the likelihood that scientific or technological solutions will fail to command public legitimacy and confidence,[33] leading to the failure of high-profile, large-scale strategic and financial investments.[34] Less well articulated, however, is the fact that a deficit model approach leads to the marginalisation and exclusion of public voices, contributing towards ***marginalisation and exclusion*** from AI. Deficit model approaches to engagement perpetuate marginalisation in several ways:

33 Jansma, S.R., Gosselt, J.F., Kuipers, K. and de Jong, M.D.T. (2019) 'Technology legitimation in the public discourse: applying the pillars of legitimacy on GM food', *Technology Analysis & Strategic Management*, 32(2), pp. 1–13. Available at: https://www.tandfonline.com/doi/full/10.1080/09537325.2019.1648788 (Accessed: 13 March 2025).

34 Hemphill, T.A. (2019) "'Techlash," responsible innovation, and the self-regulatory organization', *Journal of Responsible Innovation*, 6(2), pp. 240–247. Available at: https://www.tandfonline.com/doi/full/10.1080/23299460.2019.1602817 (Accessed: 13 March 2025).

1. Deficit model approaches position the public as lacking: These approaches view communities who already experience structural and systemic barriers to participation and inclusion (such as racism, ableism, classism, sexism, homophobia, transphobia and colonialism) as lacking the skills, knowledge or resources needed to participate effectively. This ignores the barriers they might experience (which include financial, economic, linguistic, informational, cultural, social and geographical barriers, alongside the broader issues of stigmatisation).

Deficit model approaches focus on 'fixing' perceived deficits through training or education programs, neglecting the different lived experiences of technologies that these communities may bring. This model also neglects to explore how barriers intersect in excluding those who are disempowered. In doing so, it compounds the alienation and disempowerment from the policy and technology development process experienced by communities who are supposed to benefit from AI systems and tools.[35]

2. They reinforce the power imbalance: The deficit model positions technologists and scientists as the experts, and disempowers community members and diverse publics from having the space, time and contexts to articulate and present their own expertise, values, positions, needs and potential solutions. In her ladder of citizen participation, Arnstein understands this as a 'low power' approach to participation – an approach that she critiques extensively as 'an empty ritual'. Arnstein alludes to the potential for participation itself to simply be a performance rather than substantive – an empty box ticking exercise.[36]

Similarly, development theorists Bill Cooke and Uma Kothari highlight that 'acts and processes of participation… can both conceal and reinforce oppressions and injustices in their various manifestations…' (Cook and Kothari, 2001, p.13). Researcher Sophie Hope also describes participation as an invitation into something that is already partially formed, or as an enforced, coercive and potentially embarrassing encounter. Some highlight that participation simply mediates the existing and evolving power relations between people, things, technologies, environments and contexts. In a sense, participation and inclusion in AI serves to reinforce the existing power imbalances that already exist between people and AI.[37] In these contexts, participation can contribute to existing asymmetries of power.

3. They impose top-down and neocolonial technology solutions: Deficit models often lead to the imposition of pre-designed technology and AI solutions on communities by technologists and policymakers without considering the communities' specific needs, contexts or priorities. This approach overlooks the valuable lived experiences and cultural knowledge that underrepresented

35 Draper, N.A. and Turow, J. (2019) 'The corporate cultivation of digital resignation', *New Media & Society*, 21(8), pp. 1824–1839. Available at: https://journals.sagepub.com/doi/abs/10.1177/1461444819833331 (Accessed: 13 March 2025).

36 Arnstein, S.R. (1969) 'A Ladder of Citizen Participation', *Journal of the American Institute of Planners*, 35(4), pp. 216–224. Available at: https://lithgow-schmidt.dk/sherry-arnstein/ladder-of-citizen-participation.html (Accessed: 13 March 2025).

37 Hope, S. (2019) 'Participation as performance?', *Sophie Hope*. Available at: https://sophiehope.org.uk/blog/participation-as-performance/ (Accessed: 13 March 2025).

communities possess. In the most extreme cases, the application of the deficit model can be an extension of colonial practices[38] and part of a colonial matrix of power.[39]

For example, AI technologies developed with Global North investment and priorities focus on commercial benefit which returns back to venture capital, rather than focusing on safeguarding the ancestral values, needs and priorities of Global Majority perspectives, for whom value realised through technology could look different.[40] Even the Global North's terminology and framing of AI technologies, through terms such as 'frontier',[41] make uncritical use of the language associated with practices of colonisation and settlement. While the new colonisers do not seek to capture land, their efforts to capture and enclose data through AI bears a striking resemblance to old patterns of acquisition and colonialism.[42]

These practices can reinforce broader systems of oppression by prioritising the needs and interests of stakeholders who hold greater influence in technology development ecosystems. These might include the interests of venture capital, investing in technologies through what sociologist Shoshana Zuboff describes as 'surveillance capitalism',[43] and authoritarian governments or states. They might also lead to other more insidious efforts to exert influence and control over pre-identified 'problem communities'. However, participation by communities affected by such power dynamics has the potential to create anti-colonial dynamics, and new power structures that facilitate postcolonial visions for the use of AI.[44]

38 Hao, K. (2022) 'Artificial intelligence is creating a new colonial world order', *MIT Technology Review*, 19 April. Available at: https://www.technologyreview.com/2022/04/19/1049592/artificial-intelligence-colonialism/ (Accessed: 13 March 2025).

39 Muldoon, J. and Wu, B.A. (2023) 'Artificial Intelligence in the Colonial Matrix of Power', *Philosophy & Technology*, 36(4), pp. 1–24. Available at: https://link.springer.com/article/10.1007/s13347-023-00687-8 (Accessed: 13 March 2025).

40 Hale, S., Jones, R., Kennedy, H., Middlemass, R., Millings, A., Neff, G., Ong, J.C., Patel, R., Richards, D., Snooks, K., Wajid, S. and Williams, R. (2024) 'What Do We Mean When We Talk About a Good Digital Society?', *ESRC Digital Good Network*. Available at: https://www.thebritishacademy.ac.uk/publications/what-do-we-mean-when-we-talk-about-a-good-digital-society/ (Accessed: 13 March 2025).

41 Altenbernd, E. and Young, A.T. (2013) 'Introduction: The significance of the frontier in an age of transnational history', *Settler Colonial Studies*, 4(2), pp. 127–150. Available at: https://www.tandfonline.com/doi/abs/10.1080/2201473X.2013.846385 (Accessed: 13 March 2025).

42 Op cit

43 Chicago. Zuboff, Shoshana. 2019. The Age of Surveillance Capitalism. London, England: Profile Books.

44 A research paper is forthcoming, in which I argue that we need to look beyond simply 'decolonising', towards the creation of anti-colonial and post-colonial perspectives on AI.

Visions for including public voices in AI

Moving from the deficit to the dialogue model

How do we affect the shift from the deficit to the dialogue model? In theorist Paulo Freire's seminal text on critical engagement, *Pedagogy of the Oppressed*,[45] we can see the seeds of an approach that we could apply to AI – what he describes as the process of 'praxis'. Freire's focus is on the importance of building society's capacity for critical reflection. His concept of praxis has impacted the fields of educational, participatory and democratic theory, influencing a wide range of thinkers to this day. He argues that praxis – the cyclical process of reflection and action – is central to building society's capacity for critical engagement.

This lens can be applied directly to AI, suggesting a model by which people engage and work directly with AI tools and technologies, themselves shaping and governing the tools and technologies they use – in work, play, travel and everyday life. Thus, participation in shaping AI is not confined to an ad hoc, closed and invited, deliberative process, but can be part of an ongoing, dynamic, critical relationship between society and AI. We can foster open and inclusive conversations about AI's potential impacts, risks, and ethical implications across society as a whole, allowing diverse perspectives to be heard and considered. This process of creating internal or critical consciousness, alongside deliberation and dialogue, could help us move beyond simplistic narratives about AI to develop a more considered understanding of its complex social and ethical dimensions. Freire critiqued the 'banking model' of education – similar to the deficit model – in which knowledge is passively deposited into students. Using this comparison, it is clear that simply providing information about AI (a transparency-only approach) is insufficient.

Towards AI critical thinking – for everyone

Instead, we need transformative **'AI critical thinking'** initiatives that empower the public, policymakers, regulators and developers to actively engage with AI, question its underlying assumptions and critically evaluate its applications. Increasingly, data illustrates that the acceleration of AI itself is impacting our ability to think critically. This issue is compounded by the historical neglect of critical thinking within STEM disciplines – through which many AI developers may have forged an initial pathway. The need for critical thinking about AI has never been more urgent or important.

This involves demystifying or opening up AI, making it more accessible to non-technical audiences, and fostering critical thinking skills related to data, algorithms and AI's societal impact. This is just as important for AI developers – who may not have considered the implications of technological developments for broader society – as it is for members of the public.

45 Garavan, M. (2010) 'Paulo Freire's Pedagogy of the Oppressed', in Dukelow, F. and O'Donovan, O. (eds) *Mobilising Classics - Reading Radical Writing in Ireland*. Manchester: Manchester University Press, pp. 123–139. Available at: https://www.researchgate.net/publication/260297860_Paulo_Freire%27s_Pedagogy_of_the_Oppressed (Accessed: 13 March 2025).

Freire also stressed the importance of **solidarity and collaboration** among the oppressed to challenge oppressive systems. In the context of AI, this translates to building coalitions and networks of diverse stakeholders – including community groups, civil society organisations, and concerned citizens – to collectively advocate for responsible AI development and governance. By working together, these groups can amplify their voices and exert greater influence on technological development, policy and practice. Researchers describe this as an expansion of society's locus of control. The psychologist Julian Rotter, in particular, highlights the importance of cultivating an 'internal locus of control' in citizens – the belief that a person can influence events and their outcomes – rather than an 'external locus of control', where citizens may be inclined to place responsibility and blame on outside forces.[46]

By applying Freire's principles, we can cultivate a society which is capable of reflecting critically on the implications of AI's development, taking informed action, and ultimately shaping the use of AI to align with shared values. In a move away from deficit models, towards a more meaningful form of participation and the inclusion of a wide range of public voices, we can develop more effective AI tools and technologies, which command public and societal confidence. We can even develop AI technologies that help to address and tackle the growing data and AI divide between those who benefit from these technologies (often those with greater access to educational, social and economic capital) and those who do not.[47]

The sociologist and educator John Gaventa, who is a leading theorist on community power, suggests we think of agency and power in three different ways.[48] The first is **power to** – to do something, to exert some change in the world. The second is **power with** – building collective agency and the capabilities of networks and communities as a whole to find their agency and voice. And the last relates to **power within** – overcoming internal barriers or obstacles, a necessary but not sufficient condition for both power to and power with (we must first believe we have power before we can exert it either individually or collectively).[49]

Any assessment of the effectiveness of a citizen participation initiative in the field of AI should be able to articulate the extent to which it has met all three of these conditions, from the perspectives of the communities most affected. Moments such as protests against the Ofqual exam algorithm, an exam results prediction algorithm that privileged students from private, fee-paying schools in the UK, the 'techlash' against Cambridge Analytica's use of data for profiling in the run-up to the US elections, and the unionisation of creative workers in response to generative AI, are all ongoing illustrations of how communities have exercised their **power within**, found collective **power with** each other, and exerted their **power to** shape and influence AI technologies.

46 Rotter, J. (1966) Generalized expectancies for internal versus external control of reinforcements. Psychological Monographs, 80, pp. 609

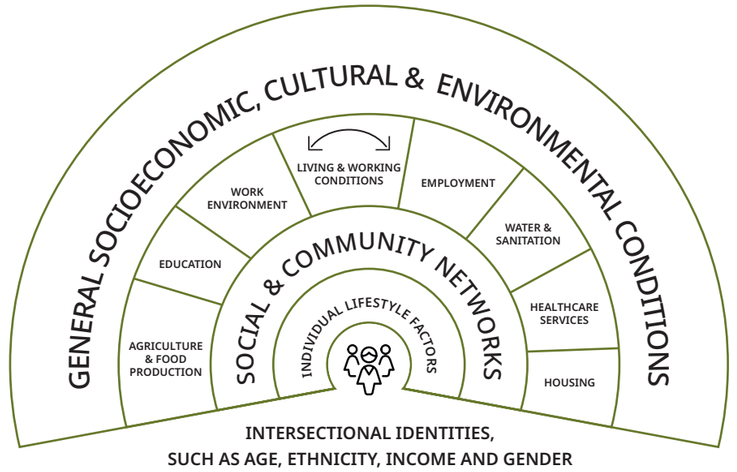
47 Patel, R. and Jones, E. (2021) 'The Data Divide: Public Attitudes to Tackling Social and Health Inequalities in the COVID-19 Pandemic and Beyond', *Ada Lovelace Institute*. Available at: https://www.adalovelaceinstitute.org/report/the-data-divide/ (Accessed: 13 March 2025)/

48 Gaventa, J, VeneKlasen, L. and Miller, V. (2002) *Power cube: A framework for analyzing power in spaces and levels*. Available at: https://www.powercube.net/other-forms-of-power/the-power-in-the-powercube/ (Accessed: 13 March 2025).

49 ibid

We also need to ensure that the use and development of AI tools and technologies considers the varied lived experience of diverse publics. Dahlgren-Whitehead's rainbow on the social determinants of health outcomes illustrates how lived experiences are shaped by a wide range of environments, cultures, societies and characteristics:[50]

Figure 4: Social and economic determinants of health outcomes (Dahlgren-Whitehead Rainbow Model)



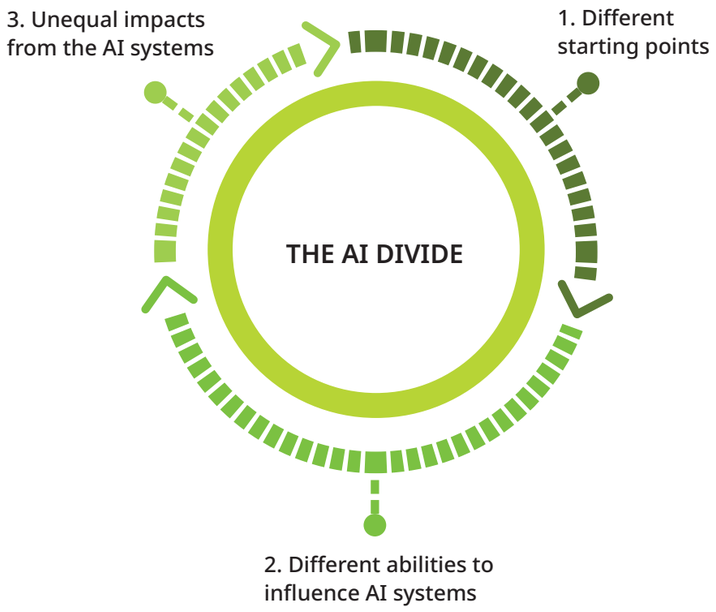
These broader factors at the individual, societal and environmental level have an enormous impact on how people experience, use and are affected by AI. In the present moment, a broad range of AI technologies is being developed, deployed and used, in a wide variety of settings – including medical, legal, policing, healthcare and labour-based contexts.

Cumulatively, these different social conditions and starting points contribute towards the AI divide – as represented below. Different and unequal starting points, such as collective disadvantage and privilege, result in stakeholders' differential abilities to design and influence AI systems. This then in turn informs who can benefit from AI systems, and in what ways they are able to do so. The unequally distributed outcomes from AI technologies, in various fields and contexts, result in real-world inequalities, such as unjust criminal justice decision making, finance decision making and law enforcement decision making, which, in turn, reinforces unequal starting points yet again.[51] This is exacerbated by the fact that AI systems themselves can shape and play an active role in structuring and shaping society, as uses of law enforcement technologies such as facial recognition, biometrics scanning and social credit scoring already illustrate. Here, real-world inequalities are baked into the way the AI itself is developed, designed and used, continuing to reinforce real-world inequalities in a vicious cycle:

50 Dahlgren, G. and Whitehead, M. (1991) *Policies and strategies to promote social equity in health*. Stockholm: Institute for Futures Studies. Available at: https://www.pslhub.org/learn/improving-patient-safety/health-inequalities/the-dahlgren-whitehead-rainbow-1991-r5870/ (Accessed: 13 March 2025).

51 https://www.adalovelaceinstitute.org/blog/structural-racism-impact-data-ai/

Figure 5: How AI can perpetuate inequality



Towards an intersectional approach to AI development to combat systems of oppression

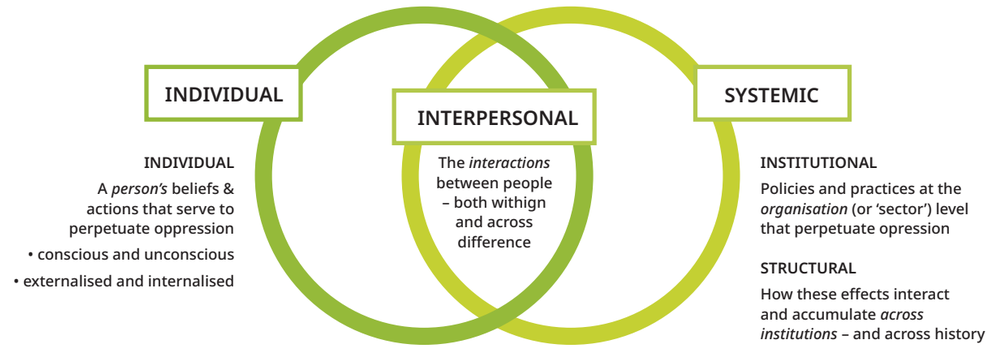
Intersectionality is defined by the National Conference of Community and Justice, NCCJ (a US-based thinktank) as 'the interconnected nature of social categorisations such as race, class and gender as they apply to a given individual or group, regarded as creating overlapping and interdependent systems of discrimination or disadvantage.' Scholars Kimberlé Crenshaw and Patricia Hill Collins' work highlights how Black women face distinct challenges that are not addressed when looking at race and gender separately. They also engage extensively with issues of class, income and geography as part of their analysis of what intersectionality is.[52] An intersectional analysis of technology's impacts and outcomes might consider how various forms of discrimination (race, gender, class, etc.) intersect to create unique experiences of exclusion and oppression.

Approaches to data collection, use and management struggle to take account of intersectionality, which has implications for the development of AI as a data-driven technology.[53] A forthcoming evidence review for the Public Voices in AI programme also finds that there is little to no engagement with intersectionality as a concept or a practice in AI engagement with publics. This demonstrates a risk that, within AI development at present, intersectionality, or exploring intersectional experiences of AI, is overlooked.

We know that a lens that classifies or segments people by their individual identities rather than taking a holistic view of their life experience is itself reductive. A meaningful dialogue with impacted communities can therefore help anticipate and address structural inequalities early on in AI development. An intersectionality-based approach to AI through the lever of inclusion and participation has the scope to reflect multiple dimensions of identity and their interactions with the use of AI. This approach could build upon the work of Crenshaw and Hill Collins, but also extend out to wider considerations such as socioeconomic status, disability, regional inequality and more, in an aim to create AI systems that are more inclusive and equitable.[54]

In an extension of this work, equity advocate Pharoah Bolding distinguishes between intersecting identities and intersectionality.[55] Drawing on the work of Crenshaw and Hill Collins, Bolding argues that this concept, rooted in an analysis of power structures that reinforce oppression for Black women, has been co-opted or conflated into a dialogue about intersecting identities. The main argument that Bolding posits is that the use of intersecting identities as a concept (defined as the idea that an individual's identity consists of multiple, intersecting factors) is often confused with intersectionality. Bolding suggests the conflation of these two key concepts results in the effect of diverting the narrative away from an acknowledgement of broader systems of oppression. See below a diagram that illustrates precisely how overlapping experiences of power, privilege and disadvantage at an individual level influences relational dynamics (interactions between people, and also with AI technologies), and filters into institutions, technologies and wider systems:

Figure 6: Systems of oppression



Source: National Equity Project

52 Collins, P.H. and da Silva, E.C.G. (2021) 'Intersectionality as Critical Social Theory', *Contemporary Political Theory*, 20(3), pp. 690–725. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC8127482/ (Accessed: 13 March 2025).

53 Bentley, C., Muyoya, C., Vannini, S., Oman, S. and Jimenez, A. (2023) 'Intersectional approaches to data: The importance of an articulation mindset for intersectional data science', *Big Data & Society*, 10(2). Available at: https://journals.sagepub.com/doi/10.1177/20539517231203667 (Accessed: 13 March 2025).

54 https://edisciplinas.usp.br/mod/resource/view.php?id=2211782&forceview=1

55 https://www.oregon.gov/deiconference/Documents/Pharoah%20Bolding%20-%20Intersectionality%20vs.%20Intersecting%20Identities.pdf

Who needs to be included in shaping AI?

As later sections in this report highlight, some bottom-up participatory processes emerge organically from the needs and interests of the communities themselves. For this reason, the question of 'who needs to be included' does not necessarily always need to be answered. This section therefore applies specifically to those taking a 'top-down' or 'messy middle' approach to engagement, participation and inclusion. This means seeking to integrate public voices into invited and closed spaces, as the ability to choose who is included and who is not is itself an exercise of power and positionality. However, it is clear that any broad move towards opening up or involving people who are not already involved is to be welcomed as part of more inclusive technology design and policymaking.[56]

When designing a participatory process to meet a specific objective in these settings, the choice about who to involve matters as much as which types of involvement or participation mechanisms are used. These choices will be dependent on context but the range of stakeholders who can be involved is broad and should move beyond those designing and deploying AI systems and governance frameworks to those affected by them.

When participatory mechanisms are introduced in policy or technical settings, there are some key questions that the developers and regulators of AI should answer about who to involve (who their beneficiaries are):

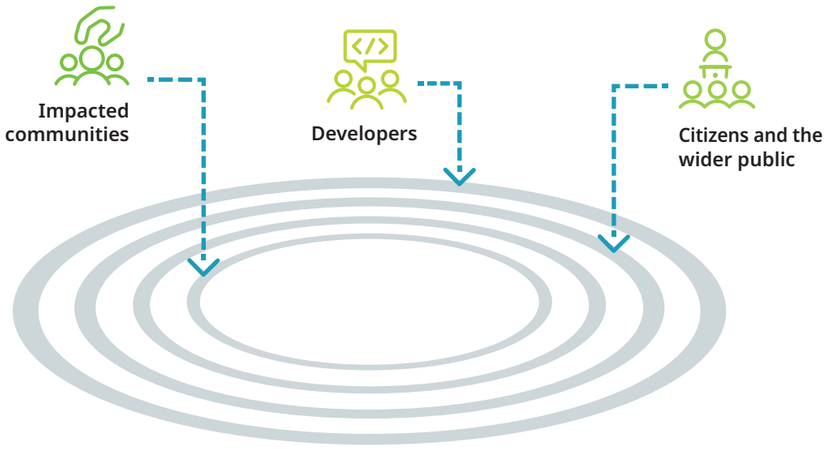
1. Who has a stake in the outcomes that emerge from the design, development, use and likely impact of this AI technology?
2. Who is included in the benefits and who is excluded from the benefits?
3. Who is likely to be impacted by any harms?
4. Who is less likely to be impacted by any harms?
5. Who is most likely to be directly affected and impacted, either benefiting or being adversely impacted?
6. Who is most likely to be overrepresented and/or underrepresented in the data training the AI system and technology?

In their responses to these questions, developers, regulators and policymakers are encouraged not to be reductive, but to consider how groups and people might be negatively affected by or underrepresented in AI. In particular, they might wish to consider the impact on those who experience multiple disadvantages. These groups might include racialised minorities; those experiencing regional inequalities; those experiencing digital exclusion; low-income groups; LGBTQ+ groups; women; people with disabilities; people who are dependent on public sector support; migrants, refugees and asylum seekers; people with experience of homelessness, clinical, financial or other vulnerabilities. Many people experience multiple disadvantages, and this is not an exhaustive list.

These key questions are informed by a recognition that the developers, designers and users of AI do not just have a responsibility to manage and design the tool responsibly, but also to recognise that it has impacts, either directly or indirectly, upon people (beneficiaries). For this reason, they must understand that when AI represents, omits or excludes people, there can be harmful consequences. This can happen, for instance, when AI discriminates against or underrepresents some people's interests and concerns. Participation can be as much about involving those who do not have a direct relationship with AI as including those who do.

The diagram below highlights the ripple effects of AI. It underscores the fact that those furthest from the system's effects often wield the greatest power and agency, while those closest to its impact may have the least. It illustrates the importance of working from the inside out, from the perspectives of those who are impacted by AI systems, rather than from the outside in.

Figure 7: Who is affected by AI systems?



A broad range of stakeholders are affected by AI, including not only those who directly use or interact with the system but also the broader public and society at large. Other beneficiaries include the professionals who use, create and regulate these technologies. At the centre of the diagram are the people who experience the impacts of the AI system directly, as the subjects of the system or AI subjects. These may be marginalised populations or those with lived experiences who are affected by the system's outcomes. Further outward, the ripple extends to citizens and the general public, representing society at large, who encounter the broader distributional and political implications of AI's use, experiencing the system less directly as indirect beneficiaries.

56 Patel, R., Brisley, R. and Gisborne, J. (2023) 'The Global Science Partnership: A toolkit for inclusive policymaking on climate change', *Ipsos*. Available at: https://www.ipsos.com/en-uk/global-science-partnership-toolkit-inclusive-policymaking-climate-change (Accessed: 13 March 2025).

Beyond this, domain experts – professionals in ethics, AI, law, medicine and related fields – inform the design and functionality of AI systems. Policymakers, industry, regulatory bodies and government authorities guide the ethical and legal framework within which AI operates, and are often the end users of the system. At the outermost layer are the developers and deployers, comprising engineers, data scientists, UX designers and industry professionals, who build, implement, use, update and quality assure these systems.

Since power often lies with those furthest removed from AI's immediate impacts (e.g. policymakers, developers and industry leaders), it is important to centre marginalised and less-visible voices. By intentionally prioritising equity and inclusion through the mechanism of participatory AI governance, we can work to counterbalance these power dynamics and ensure that those most directly affected by AI systems have a meaningful say in their development and governance.

How can we include the public in shaping AI?

How do we conceptualise or think about AI?

Defining the term Artificial Intelligence (AI) can be challenging. When policymakers, technologists and social scientists are asked to define AI, we struggle to identify a single defining characteristic common to all of the technologies that we use the term 'AI' to apply to. In the Philosophical Investigations, philosopher Ludwig Wittgenstein describes the concept of a game – and explores the ways in which there are overlapping and criss-crossing similarities, rather than a single characteristic that means everything is a game:[57]

> *'Consider for example the proceedings that we call "games". I mean board-games, card-games, ball-games, Olympic games, and so on. What is common to them all? – Don't say: "There must be something common, or they would not be called 'games'" – but look and see whether there is anything common to all. For if you look at them you will not see something that is common to all, but similarities, relationships, and a whole series of them at that. To repeat: don't think, but look!"*[58]

We can pick up Wittgenstein's exhortation to us to look at what we are ascribing the term 'AI' to, and observe how we use this term, to understand it better ('meaning as use').[59] Our use of the term AI seems to encompass a wide cluster of technologies, uses and applications, from predictive through to generative technologies. It covers a broad range of uses – from chess-playing algorithms to medical diagnostic tools – and applications. The same 'AI' technology (such as a large language model) can even be used to different ends and purposes. Some AI technologies are single use and developed for a bespoke purpose and context, others are dual use and many others are multiple use. The development process itself, including the datasets that the AI models are trained on, can vary enormously in terms of size, complexity and quality. The methods and approaches applied to train the models (machine learning, deep learning etc.) also vary enormously.

Together, these systems, processes, uses and approaches appear to form an identifiable and novel class of emerging technologies that we use the term 'artificial intelligence' to describe. They share overlapping capabilities that aim to resemble the capabilities of human intelligence, such as learning and processing information, but none are exactly alike. So, we understand AI as a rapidly evolving cluster of different technologies not through a fixed definition but through its resemblances – the shared traits that connect its various forms.

Why does this matter in considering how we include people in shaping AI?

When considering how we include people in shaping the development of AI, the implications of this understanding of AI are profound.

If AI is not a single, fixed technology but rather a cluster of overlapping, evolving systems with diverse applications, then involving people in its shaping must reflect this complexity and fluidity. This means that it is necessary to be specific about the use cases, purpose, application and impact of AI tools. It also highlights the need for a multifaceted, pluralistic and inclusive approach to participation – one that recognises the varied ways in which AI technologies affect different stakeholders and contexts. By including people who interact with or are affected by AI in different ways, we can uncover the specific challenges and opportunities unique to various contexts and applications. For example, the training datasets and design choices behind a predictive medical AI might require input from patients, healthcare workers and ethicists, while a generative AI used in the creative industries would benefit from the input of artists and cultural stakeholders.

If we acknowledge AI's evolving nature as an emergent technology, this means that participation needs to be an iterative and adaptable process. Stakeholders should be involved not only in the design and development phases of the AI but also throughout its deployment, evaluation and subsequent updates. This ongoing involvement allows for feedback loops that address new issues as they arise and ensures that AI technologies remain aligned with societal values over time. The varied and relational nature of AI calls for an equally diverse and relational approach to inclusion, one that sees participation not as a one-size-fits-all process but as an adaptive, contextual and equitable practice that evolves alongside the technologies themselves.

57 Grant, S. (2023) 'How playing Wittgensteinian language-games can set us free', *Aeon*. Available at: https://aeon.co/ideas/how-playing-wittgensteinian-language-games-can-set-us-free (Accessed: 13 March 2025).

58 ibid

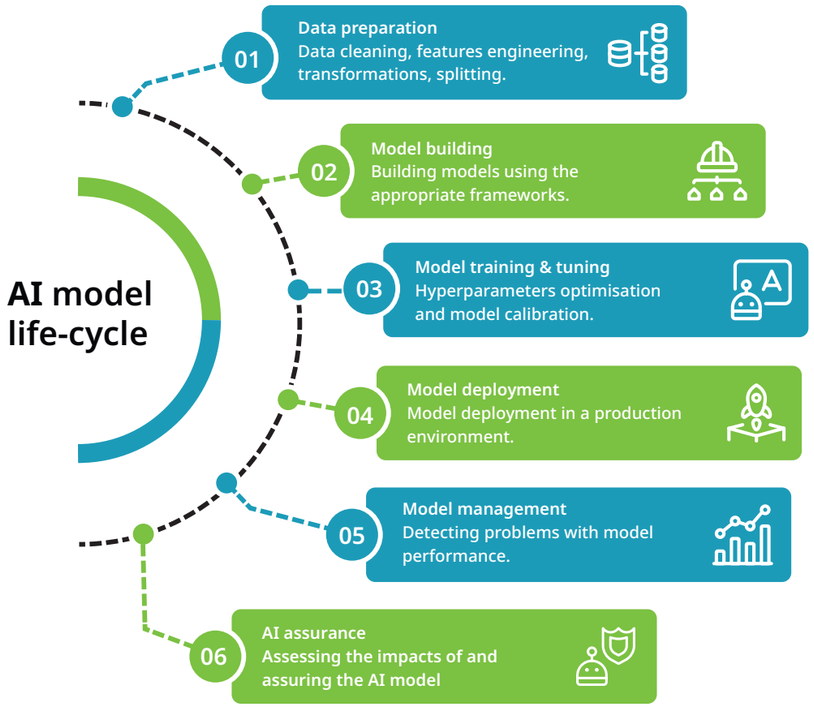
59 In the Philosophical Investigations, Ludwig Wittgenstein writes that the meaning of a word is understood by observing its use in the language, serving as a link between language and the real world.

When should the public be included in shaping AI?

People can be involved in shaping AI at many different stages of the eco-socio-technical and democratic systems. Their involvement need not be constrained to shaping AI's technical development ('midstream'), but can extend to questions about the role AI plays in society ('upstream'), as well as 'downstream' (thinking about AI's continued impact and influence on society).

Beyond the AI lifecycle: Many suggest using the AI lifecycle as a basis for including the public in AI – see the schematic of the AI lifecycle below. This diagram outlines the stages involved in creating and operationalising AI systems. It begins with data preparation, problem definition and data collection. This is followed by model building, selection and design, and then by training and tuning, where algorithms learn from the prepared data set. This is then used to shape and inform wider model development and management, which integrates the AI system into production environments and then tracks performance and updates the model as required.[60]

Figure 8: The AI development life cycle



60 De Silva, D. and Alahakoon, D. (2022) 'An artificial intelligence life cycle: From conception to production', *Patterns*, 3(6), p. 100489. Available at: https://www.sciencedirect.com/science/article/pii/S2666389922000745 (Accessed: 13 March 2025).

However, looking upstream and downstream of the AI lifecycle will be necessary if we are to meaningfully effect change in AI systems – due to some issues with the framing of this lifecycle. The traditional AI lifecycle framework is often criticised for its oversimplified portrayal of a process that is, in reality, much less linear, fixed and certain. Because of this, the model can be seen to veer into a form of technological determinism. The traditional AI lifecycle framework also fails to account for the complexities of AI development, such as the lack of a clear endpoint, as the reality of AI development is that advancements and iterations continue indefinitely. Moreover, the lifecycle does not create space for uncertain outcomes or indicate where there may be gaps – for instance, around mechanisms for removing or decommissioning AI systems. As such, it neglects the critical need for the responsible oversight and potential withdrawal of risky technologies. The seamless flow between the five stages overlooks the potential for accountability mechanisms, or socio-technical checks and balances that may need to be interleaved between the different stages.

The narrative around how AI is developed has largely been shaped by technology developers, which can mask the complexity and messiness of development. It can obscure accountability and diffuse responsibility across the multiple stages of development, leaving it unclear who is truly accountable for the decisions that are made. The lifecycle model also conceals the opacity inherent in AI development processes, relying on developers to voluntarily open their practices up for public participation – a condition that is often not met. As a result, the lifecycle model can reinforce exclusion in shaping AI, placing power in the hands of the developers to enable and foster participation in existing black-box models.

As a recent Ada Lovelace Institute paper (Groves, Strait, et al) has already found, the use of public participation in commercial contexts remains woefully constrained. We see public participation take place predominantly in the service of commercial goals and shareholder priorities, rather than in alignment with the broader societal ethos of stakeholder capitalism. Stakeholder capitalism proposes that companies should serve the interests of all stakeholders including employees, customers and broader society, and not just its shareholders.[61] The paper finds that participatory approaches encounter deep resistance in commercial settings and are at risk of being co-opted as a form of market research. Groves and Strait's report finds that public participation in the AI industry currently lacks a clear understanding of best practices, and pioneers within the industry report that they face obstacles such as resource intensity, misaligned incentives and atomisation.

A forthcoming survey of AI researchers undertaken by University College London and the Public Voices in AI project finds that these issues are worsened by the gap between AI researchers and the public. The two groups express differing levels of concern about AI, with the public more concerned about AI than those who are involved in developing and deploying AI. However, AI researchers do still have significant and wide-ranging worries, underscoring the importance of addressing these issues. The survey also finds that AI researchers form a highly unrepresentative group and do not currently prioritise questions of inclusion and exclusion. Despite these issues and concerns, they still

61 Schwab, K. and Vanham, P. (2021) 'What is stakeholder capitalism? Its history and relevance', *World Economic Forum*. Available at: https://www.weforum.org/stories/2021/01/klaus-schwab-on-what-is-stakeholder-capitalism-history-relevance/ (Accessed: 13 March 2025).

recognise a crucial role for the public in AI regulation and impact assessment and do not advocate for AI to develop at the fastest possible pace. While they perceive policymakers and the public as lacking understanding of AI (echoing the deficit model dynamic explored earlier), the survey also illustrates that they are, in the main, also largely unaware of the participatory and deliberative engagement methods that could help bridge these knowledge gaps.

Similarly, Sloane et al critique the ways in which the term participation is used in commercial settings to mean several substantively different things.[62] Sloane et al identify participation as work, 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. This is a common type of labour in AI development that is often undertaken by workers earning low wages in poor labour conditions. However, within commercial settings, it is framed as a form of participation. Such content moderation work is often traumatising and likely to cause lasting harm to workers.

Consultation can also be framed as a type of participation. This approach 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, using user experience (UX) methodologies. However, such approaches are critiqued for instrumentalising participatory approaches with a view to realising corporate targets (the incentive to sell or commodify, leaning into practices of consumption), rather than seeking to authentically transform the relationship between consumer and company to that of citizen and company.[63]

In this report, therefore, the lifecycle is not suggested as a basis for including public voices in AI. Instead, AI developers who are inclined to consider the meaningful use of public participation within commercial contexts might want to consider adopting participatory methodologies as part of a wider ethos of serving stakeholder goals and values. AI developers operating and working in alignment with the goals of wider stakeholder capitalism – rather than exclusively shareholder capitalism – could foster the emergence of glass box AI models, which are structured for direct transparency, interpretability and explainability.[64]

They may also wish to consider, as Groves and Strait suggest, developing standards of practice for public participation in a commercial setting. These may, in future, allow for increased participatory innovation in the context of commercial AI development. Such standards could encourage the trialling of public participation approaches in an open-source or transparent way with the goal of what Sloane describes as participation as justice. This framing focuses on longer-term partnerships and enquiries that contribute to agency over design and infrastructure that affect the lives of participants. Participation is less about delivering on technology-focused outcomes, but instead about inviting challenges to the existing power dynamics between the developers of AI systems and those who are impacted by them.

62 Sloane, M., Moss, E., Awomolo, O. and Forlano, L. (2022) 'Participation is not a design fix for machine learning', *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '22)*, pp. 1–6. Available at: https://dl.acm.org/doi/10.1145/3551624.3555285 (Accessed: 13 March 2025).

63 Alexander, J. (2022) The Citizens: *A new approach to solving society's biggest challenges*. London: Penguin Books.

64 InterpretML (2023) 'Glassbox Models', *InterpretML Documentation*. Available at: https://interpret.ml/docs/glassbox.html (Accessed: 13 March 2025).

Participation upstream of model development

In this framework, I do not engage with the broader questions of involving people in the governance, management and use of the data that AI models are often trained on, as these questions have been addressed in separate publications by the Ada Lovelace Institute in 2021 and 2024 (I was the lead author of the framework for participatory data stewardship in 2021). However, involvement in AI training data is another basis for including people in shaping AI systems upstream.

So, if the AI lifecycle does not work as a basis for understanding how to include people in shaping AI, what might work instead? We can conceptualise participation at different levels, rather than just at the level of the AI tool. We can include people in shaping the broader eco-socio-technical system itself, upstream of the development of specific tools or technologies, calling attention to questions of use, procurement, supply chain and labour force, data collection and management practices, environmental impacts and applications. We might wish to consider participation against the backdrop of the supply chains, organisational structures, cultural dynamics and hidden networks of people and environmental resources that underpin AI systems. This broadens the lens to account for some of the less visible, or 'invisible', aspects of AI systems: the influence of capital, hidden labour, and environmental costs and the unequal power dynamics around the use, development and application of AI between Global North and Global Majority contexts. By shedding light on these dynamics, we can develop a clearer understanding of how to redistribute power and create more equitable AI systems.

We also need to recognise and take account of wider trends in the way democratic and participatory systems are developing, and the political contexts within which AI systems are being developed worldwide. We can include people in broader questions of political choice, refusal, regulation and governance. Drawing on campaigner Simon Burall's deliberative systems model, this would look like active engagement with key decisions about AI governance and policy within the empowered institutionalised spaces of the United Nations, Parliament and civil service. It would be part of a widespread suite of 'bottom-up', 'messy middle' and 'top-down' participatory methodologies to open up dialogue and debate.

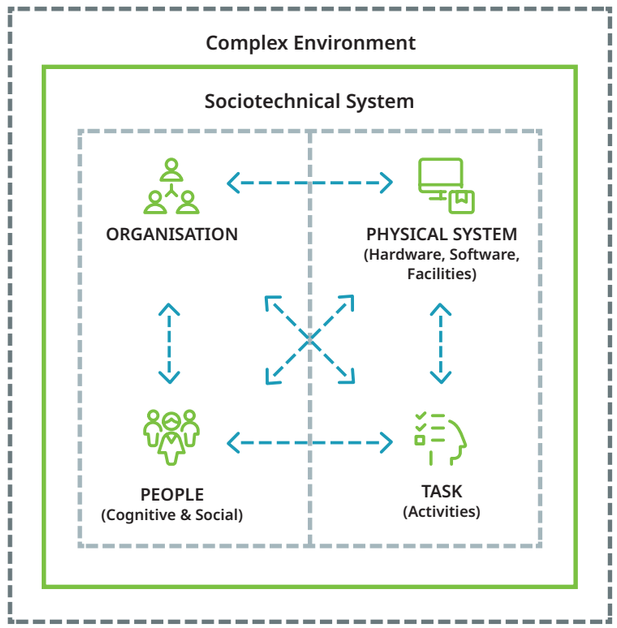
Participation downstream of AI model development

We can also include people in the shaping of AI technologies downstream of the development and design cycle. This would include looking at social and participatory assurance, and at the continued impacts and effects technologies are having on wider society. This might involve methodologies such as participatory impact assessments and transparency-based initiatives such as algorithmic risk assessment co-designed with members of the public impacted by the algorithms, alongside risk registers. A Responsible AI UK funded project at the University of Edinburgh, Participatory Harm Auditing Workbenches and Methodologies, is developing a new certification framework for participatory AI auditing, whilst a Nesta project is refining and developing a social assurance framework for public sector AI through the use of deliberative polling.[65]

Participation across the eco-socio-technical system

AI is embedded within wider socio-technical and democratic systems. Using the lens of the **eco-socio-technical system**, it is possible to recognise that AI operates within the context of the tasks, people, cultures, structures and physical hardware that make up society. It is both shaped by and affects these wider dynamics.

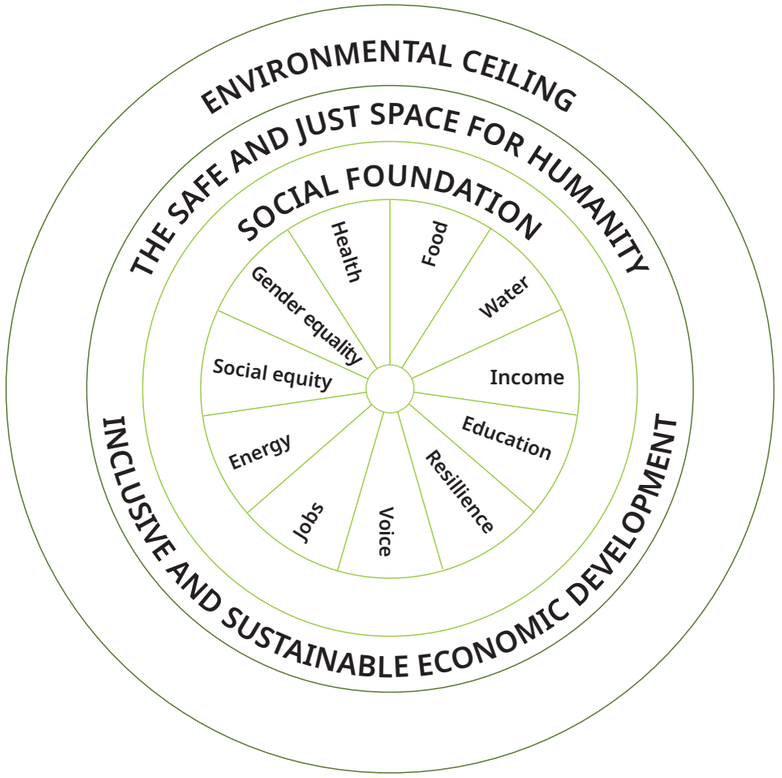
Figure 9: Eco-socio-technical systems



65 Digital Humanities Institute (no date) 'Participatory Harm Auditing Workbenches and Methodologies', *Digital Humanities Institute*. Available at: https://www.dhi.ac.uk/projects/phawm (Accessed: 13 March 2025).

Eco-socio-technical systems operate within planetary constraints and boundaries, or what economist Kate Raworth describes as the 'doughnut' model – see Figure 15 below. Just as AI technologies rely upon, are shaped by, and shape a social foundation, they also need to operate within existing planetary boundaries to avoid exacerbating the issues of climate change or overshooting the mark. The goal for participatory and inclusive AI is therefore to ensure that AI technologies are founded on and contribute towards a just social foundation, ***and*** operate within the parameters of an environmental ceiling, as part of a safe and just space for humanity:

Figure 10: The Doughnut model (Kate Raworth)

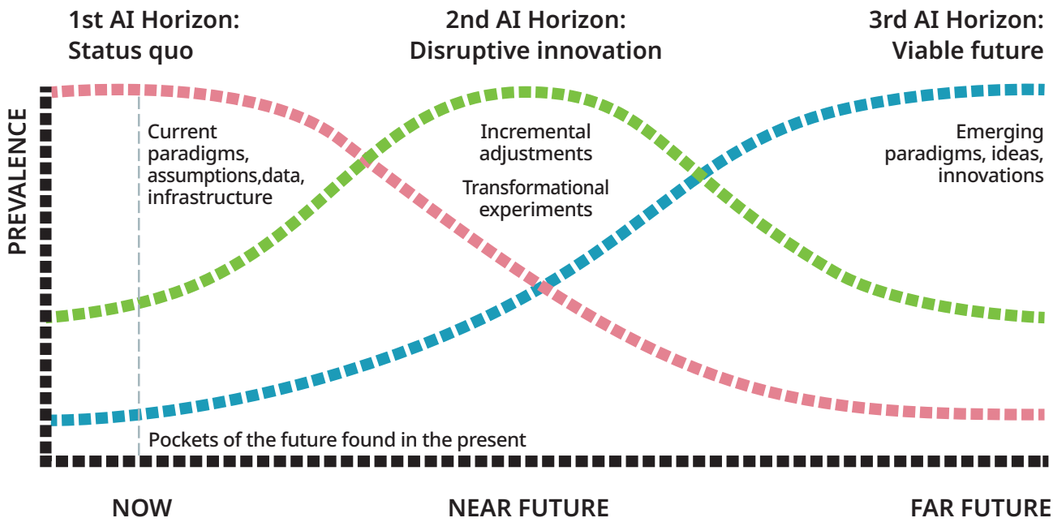


Participation and inclusion can be leveraged as a critical mechanism to secure the balance envisaged in the doughnut model. However, participatory efforts often encounter resistance due to entrenched power asymmetries, lack of transparency and cultural barriers. Participatory processes are also at risk of being co-opted into existing systems, losing their ability to make systems change happen. As the early part of this paper has made clear, the status quo is currently untenable, as it excludes the majority of the public from the ability to exercise agency or control over a powerful technology, AI. We can draw from the Three Horizons Framework to understand how to move towards justice, overcoming barriers and resistance.

The Three Horizons Framework for participation and inclusion in AI

The Three Horizons Framework helps us explore how things might evolve in future and how we can manage change over time. It's a key tool for thinking about how we can move from the present to a more sustainable, inclusive future, and identifies pathways, obstacles and barriers to systems change.

Figure 11: Three Horizons for Participation and Inclusion in AI



Horizon One (the status quo): This represents the present situation and the systems we have now. It includes existing issues, norms and structures. In the context of emerging AI technologies, Horizon One is about how these AI technologies are currently impacting us, showing both the benefits and the harms we're experiencing today. In the context of participation in AI, Horizon One is about the power asymmetries and exclusion of public voices and perspectives in AI.

At this horizon, sustaining participatory innovation keeps the lights on, but otherwise maintains the status quo. Horizon One represents the dominant AI systems today. These are often characterised by managerial stability, existing concentrations of technological and political power and exclusion. They represent black-box AI systems driven by commercial imperatives at the exclusion of societal and planetary benefit, and the perpetuation of existing power dynamics and processes of exclusion and marginalisation. At this horizon, there are some efforts to involve and include people in AI systems, but these are not widespread, they are largely tokenistic, and they struggle to meaningfully influence or exert change on established power structures within technology and AI development.

Horizon Two (disruptive innovation): This is the middle ground where things start to shift. Emerging trends or innovations, both in the field of AI and in the field of inclusion and participation in AI, may disrupt the status quo. In terms of including public voices in AI technologies, Horizon Two represents ongoing concerns and broad societal ambivalence about the growth of AI. There might be rising public worry about harms caused by AI, even while some of its benefits are still visible. Horizon Two represents a stage where many are working to actively respond to some of these worries, developing new participatory and democratic innovations as well as social movements. It's a time of experimentation, adaptation and transformation, but also resistance. The way ahead to Horizon Three is not linear and there will be barriers to systems change. Horizon Two encompasses widespread efforts to address these challenges, such as regulatory changes enabling increased participation across the ecosystem. There are also emerging participatory innovations such as citizen juries, data co-operatives, open-source data trusts, citizen assemblies, worker observatories and global participatory methods, including models aimed at fostering deeper systemic shifts.

Horizon Three (a viable future): This horizon is about the long-term future, in which new systems, ideas or ways of doing things may emerge. It's the realm of transformation and possibility. Looking ahead at Horizon Three, we see how society itself is transformed so that AI technologies adequately include public voices. We encounter visionary AI futures in which participation shifts power structures and business models. At this third horizon, emerging paradigms and innovations are adopted, so that they become the new status quo. Alternative ways of designing and governing AI prioritise equity, justice and participation.

Can participation really change anything?

We are seeing ever-increasing concentrations of technological power accelerated by the exponential growth of AI technologies, and, in particular, by the capabilities of generative AI. Generative AI introduces the ability to create new output and content that relies upon existing datasets, output and content. There are dynamics of inequity in who uses generative AI, and why. These dynamics persist in the relationship between those whose data trains the models, and those who use the model, and in the representation of societal norms within the content that emerges from the models.[66] Given these issues of equity, there are growing concerns from many established practitioners and policymakers that, as an approach, participation simply facilitates the existing status quo and direction of travel with regard to AI technologies, rather than fundamentally shaping or transforming outcomes.

Furthermore, large technology companies seek to misuse and appropriate the language of democracy and participation, and use their funding power to subvert existing standards of practice within participatory processes.[67] Beyond the field of AI and participation, academics such as Sonia Bussu and Adrian Bua have argued for the reclaiming of participatory governance away from the fields occupied by power holders (for instance, technologists and policymakers), towards the social movements within which there is greater urgency and appetite for change.[68] They argue that friction can be embedded in the relationship between social movements and state institutions, but that this can be a productive dynamic.[69]

One issue contributing towards these power dynamics is the unequal distribution of technological resources, capabilities and influence. In recent years, we have observed that a small group of powerful entities are increasingly controlling and enclosing key technologies, infrastructure, data and expertise, while others lack access or agency. These concentrations have emerged from phenomena such as market dominance and the monetisation of human behaviour through Zuboff's 'surveillance capitalism'.[70] These imbalances create asymmetries in people's ability to shape the benefits of AI or influence its development, exacerbating existing inequalities. Potentially, these imbalances lead to exploitation, dependence and the widening of social, economic and political divides.

How can citizen participation in the field of AI be anything but a sticking plaster given the scale of these challenges? As Arnstein argued in her 1969 work on citizen participation in the planning context, the purpose of *meaningful* participation is to shift power structures, creating increased levels of agency among communities, particularly those who are most marginalised.[71] She understood this 'high power' approach as distinct from *low power approaches*, described as tokenism, and as an *'empty ritual of participation'*. This is akin to the ritual that many large technology companies engage in when they claim to be involving the public in AI. Arnstein also critiqued efforts to 'participation-wash' and legitimise pre-existing power structures. She highlights the ways in which participatory processes can be appropriated or adapted to reinforce positionality and privilege:

> *'The idea of citizen participation is a little like eating spinach: no one is against it in principle because it is good for you... a revered idea that is vigorously applauded by virtually everyone.*
>
> *The applause is reduced to polite handclaps, however, when this principle is advocated by the have-not Blacks, Mexican-Americans, Puerto Ricans, Indians, Eskimos, and Whites.*
>
> *And when the have-nots define participation as redistribution of power, the American consensus on the fundamental principle explodes into many shades of outright racial, ethnic, ideological and political opposition.'*

In this quote Arnstein highlights one potential risk and challenge of participatory methods and practices, that they do not always serve all groups and communities equally. Participation expert Russ Dalton calls this the 'participation gap': some people, groups and communities have more power, agency and ability to ensure that their voices are heard.[72] In this context, participation alone, without an understanding of broader issues of power, equity and justice, will not result in the power shift Arnstein envisages. Nor will it change results and outcomes substantively for those who are already underrepresented. This issue with participation wash is illustrated to great effect in her original ladder of citizen participation, where she highlights the risks of manipulation and placation – see below:

66 Capraro, V., Lentsch, A., Acemoglu, D., Akgun, S., Akhmedova, A., Bilancini, E., Bonnefon, J.-F., Brañas-Garza, P., Butera, L., Douglas, K.M. and others (2024) 'The impact of generative artificial intelligence on socioeconomic inequalities and policy making', *PNAS Nexus*, 3(6), p. 191. Available at: https://doi.org/10.1093/pnasnexus/pgae191 (Accessed: 13 March 2025).

67 Tiku, N. (2025) 'OpenAI's push to democratize artificial intelligence raises concerns about bias and control', *Time*. Available at: https://time.com/6684266/openai-democracy-artificial-intelligence/ (Accessed: 13 March 2025).

68 Bua, A. and Bussu, S. (2023) *Reclaiming Participatory Governance: Social Movements and the Reinvention of Democratic Innovation*. London: Routledge.

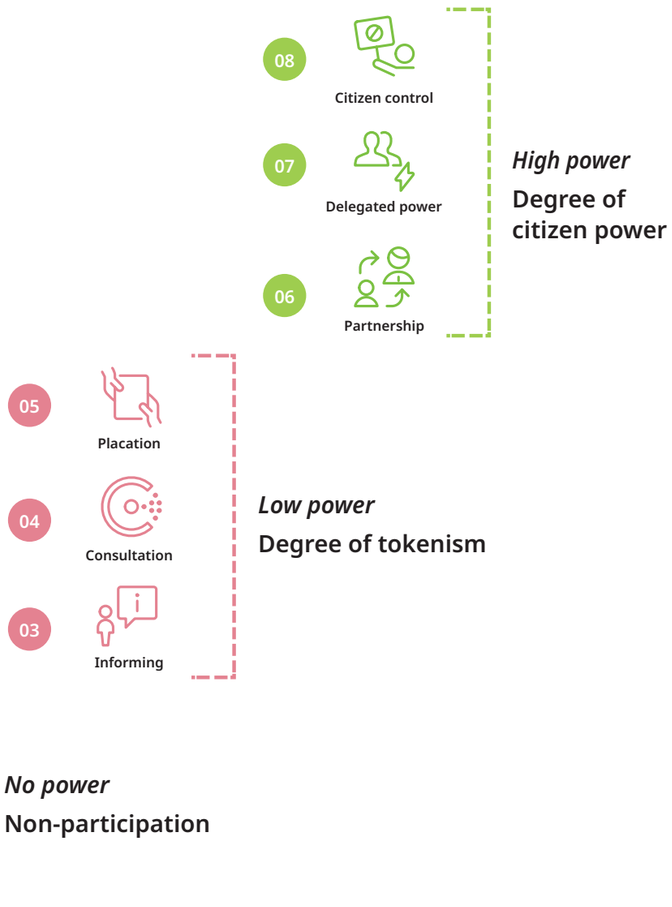
69 Schulbaum, O. (2024) 'Inspiring Democratic Innovation: A Conversation with Sonia Bussu on Reclaiming Participatory Governance', *Wilder Journal*, 2. Available at: https://journal.platoniq.net/en/wilder-journal-2/interviews/reclaiming-participatory-governance/ (Accessed: 13 March 2025).

70 Op cit

71 Op cit

72 Dalton, R. (2021) The Participation Gap. Oxford: Oxford University Press.

Figure 11: Arnstein's ladder of citizen participation



Arnstein herself acknowledged some of the critiques that have subsequently been levelled at the ladder of citizen participation. The IAP2 repositioned Arnstein's original ladder into the spectrum above to suggest that the activities of informing, consulting, involving, collaborating and empowering are not necessarily mutually exclusive. Some have critiqued the linearity of the model, suggesting that participation can be cyclical and multifaceted. Professor Tina Nabatchi highlights the elements of relationality, dialogue and wider sociopolitical contexts that the framing of the ladder overlooks in part, as it was necessarily a simplification. Researcher Hans Asenbaum argues that perspectives on participation are more kaleidoscopic rather than linear.[73] Some researchers now suggest that an alternative, less hierarchical framing might be one of a mosaic or of a star.[74] Few, however, seem to disagree with Arnstein's fundamental position that meaningful participation should be understood as a mechanism to enable the transfer of power. In centering power, rather than methods, approaches and processes, Arnstein encourages us to focus on the purpose of participation. In doing so, we are challenged to pay less attention to the methods and processes (the means of participation), and more attention to the outcomes that emerge from participation – the question of whether power shifts (the ends of participation).

Another critique of Arnstein's ladder emerges from the more recent work of Professor Rikki Dean. He argues that the ladder is a highly normative framing, and does not always clearly articulate the reality of the very different goals and purposes different actors have when embarking on citizen participation. He has created a typology assessing participation by its intended purpose and goals (named/stated/unnamed/unstated). He argues that the four goals and modes – labelled knowledge transfer, collective decision-making, choice and voice, and arbitration and oversight – are each linked to different traditions in democratic and public administration theory, with different normative assumptions and groundings:

We can apply this schema to understand that different actors and stakeholders in the AI ecosystem have different motives for participation. These may be, but are not necessarily, compatible with the idealised Arnsteinian goal of shifting power structures. Some actors, such as regulators and policymakers, are actively involving citizens in arbitration and oversight. Some, such as UK innovation agency Nesta,[75] and the NHS AI Lab,[76] are seeking to create collective intelligence and enable collective decision making. Some may be designing and developing methods to support participants to express their choice and voice, with initiatives such as the Workers' Observatory, funded by the Responsible AI UK Public Voices programme.[77] However, many are simply hoping to inform or effect knowledge transfer about AI, rather than to shift power structures at all. This process looks similar to the deficit model discussed earlier in this paper.[78]

With the exception of participation as knowledge transfer, the remaining three modes of implementing participation still map onto different ways of exercising or increasing citizen power, as described by citizen participation advocates VeneKlasen and Miller (2002).[79] Participation as arbitration and oversight maps onto **power to** – to do something, to exert some change in the world. The second is **power with** – building collective agency and the capabilities of networks and communities as a whole (participation as collective decision making). And the last example relates to **power within** or overcoming internal barriers or obstacles (participation as choice and voice).[80]

73 Asenbaum, H. (2021) 'Rethinking democratic innovations: A look through the kaleidoscope of democratic theory', *Political Studies Review*, 20(2), pp. 203–216. Available at: https://doi.org/10.1177/14789299211052890 (Accessed: 13 March 2025).

74 Collins, K. and Ison, R.L. (2009) 'Dare we jump off Arnstein's ladder? Social learning as a new policy paradigm', *Proceedings of the PATH (Participatory Approaches in Science & Technology) Conference*. Available at: https://oro.open.ac.uk/8589/1/Path_paper_Collins_Ison.pdf (Accessed: 13 March 2025).

75 Saunders, T. (2017) 'How to involve the public in the development of artificial intelligence', *Nesta*. Available at: https://www.nesta.org.uk/blog/how-to-involve-the-public-in-the-development-of-artificial-intelligence/ (Accessed: 13 March 2025).

76 Sciencewise (2022) 'NHS AI Lab Participatory Fund for Patient-Driven AI Ethics Research', *Sciencewise*. Available at: https://sciencewise.org.uk/projects/nhs-ai-lab-participatory-fund-for-patient-driven-ai-ethics-research/ (Accessed: 13 March 2025).

77 For more information see https://workersobservatory.org/

78 Könneker, C. (2024) 'How AI will change science communication', *Stanford Social Innovation Review*. Available at: https://ssir.org/articles/entry/science-communication-artificial-intelligence (Accessed: 13 March 2025).

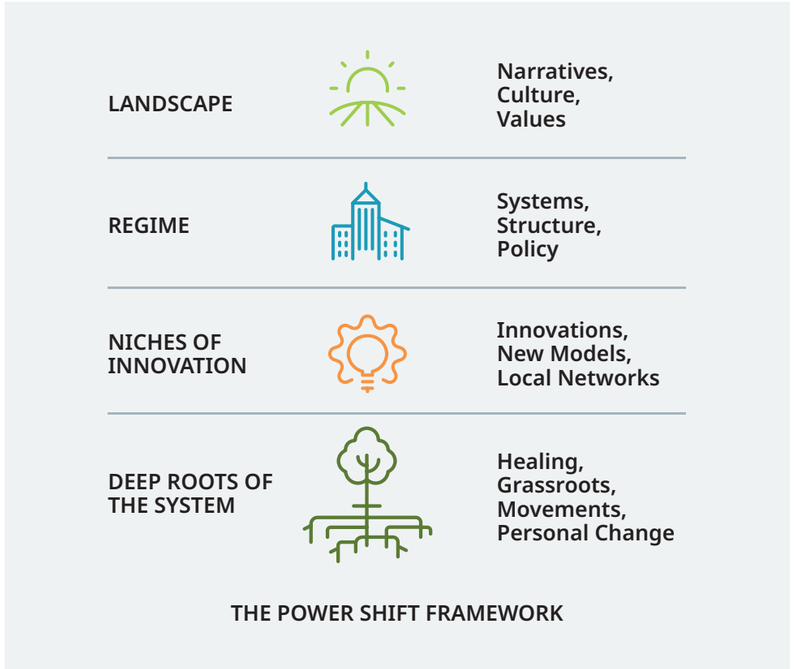
79 Just Associates (2007) *A New Weave of Power, People & Politics: The Action Guide for Advocacy and Citizen Participation*. Available at: https://justassociates.org/all-resources/a-new-weave-of-power-people-politics-the-action-guide-for-advocacy-and-citizen-participation/ (Accessed: 13 March 2025).

80 https://www.participatorymethods.org/method/power

Strategies to ensure power shifts through participation

Beyond exploring the development and use of the participatory process, we need to develop strategies to ensure that participation makes change happen. How do we understand how power might shift within the context of participation and AI – as power itself is such a broad concept. The Power Shift framework, developed by Tatiana Fraser and Juniper Glass, is useful here. It highlights the fact that power exists in different places and at different levels in a socio-technical system.[81] Drawing upon the metaphor of a tree, where roots run deep but are invisible, it illustrates how inward personal change, relationships and connections at the grassroots form the basis of the deep roots of the system. These deep roots foster the development of new models, innovations and connected networks, which, in turn, shape and influence the regime – our policies, systems and societal structures. Like the leaves on a tree, our narratives, culture, ways of working, practices and values draw from the deep roots, trunk and branches of the tree:

Figure 12: The Power Shift framework:



Source: Tatiana Fraser and Juniper Glass, adapted from Geels Transition Theory

81 https://www.innovationunit.org/thoughts/the-power-shift-framework/

Historically, power has been understood to reside within the hands of the traditional institutions, systems, structures, policymakers and regulators. It has also been seen to be held by those who shape and define narratives, culture and values across society (those with the power to make their voice heard). The philosopher Foucault describes this 'soft' exercise of power as 'biopower', which allows governmental and cultural institutions to produce norms, reinforce them, and (in some instances socially) discipline and punish those who fail to meet these societal expectations.[82]

As a consequence, some are critical of a focus on participation only operating at the level of traditional structures. Cruikshank explicitly applies this Foucauldian lens to thinking about participation itself. She explains how initiatives aimed at empowering marginalized groups – such as self-help movements, welfare programs and community activism – often function as a means of social control. Rather than reducing power imbalances, these initiatives can shape individuals into self-regulating subjects who internalise state norms.[83] She critiques the assumption that the goals of empowerment are inherently emancipatory, highlighting how these goals and processes can be co-opted to encourage compliance with existing power structures.

Power can also reside within emerging niches of participatory innovation. As Fraser and Glass's framework illustrates, however, power can reside invisibly at the deep roots of the system, where grassroots and bottom-up movements, drawing upon lived experience and collective experience of marginalisation, can represent strategies for systems change.[84]

These different types of power have been described by Jeremy Heimans and Henry Timms using the binary of new power (distributed, collective, solidaristic and organised), as opposed to old power (concentrated, hoarded, retained within institutions and existing power structures).[85] For example, in 2020, the death of George Floyd, a Black man who was murdered by a White police officer, sparked protests from Black Lives Matter activists and allies across the globe. The ascendance of this movement made mainstream a hitherto marginalised conversation about structural inequalities, injustice and asymmetries of power, particularly in the context of racial injustice. It also exposed the influence of 'new power' – the participation of many individuals in a distributed way – and its tension with 'old power', i.e. institutions that are exclusive, controlled and top-down, but whose authority and legitimacy is increasingly questioned in a networked society.[86] Old power is manifest, in this case, in the authority of the Minneapolis Police Department.

Old power can be understood as akin to a currency (hoarded), and new power as a current (relying on connections and networks to surge), drawing from the deep roots of the system. There are ongoing collective responses that illustrate new power in action in the field of AI, such as protests

82 Means, A.J. (2021) 'Foucault, biopolitics, and the critique of state reason', *Educational Philosophy and Theory*, 53(12), pp. 1161–1172. Available at: https://doi.org/10.1080/00131857.2021.1871895 (Accessed: 13 March 2025).

83 Cruikshank, B. (1999). *The Will to Empower: Democratic Citizens and Other Subjects*. Ithaca, NY: Cornell University Press.

84 Fraser, T. (2021) 'Towards a new holistic framework of systems change: Adapting Geels' transition theory', *Refuge for Systems Leaders*. Available at: https://medium.com/refuge-for-systems-leaders/towards-a-new-holistic-framework-of-systems-change-adapting-geels-transition-theory-8d589fb6de0a (Accessed: 13 March 2025).

85 Patel, R. (2020) 'Blending old and new power: Democratic innovation in 2020', *Involve*. Available at: https://www.involve.org.uk/news-opinion/opinion/blending-old-and-new-power-democratic-innovation-2020 (Accessed: 13 March 2025).

86 ibid

against the Ofqual exam algorithm in the UK, the 'techlash' against Cambridge Analytica's use of data for profiling in the run up to the US elections and the unionisation of creative workers in response to generative AI. Heimans and Timms argue that the dynamics of new power, often mediated through technology, are changing how people see themselves in relation to institutions, as well as in relation to each other.

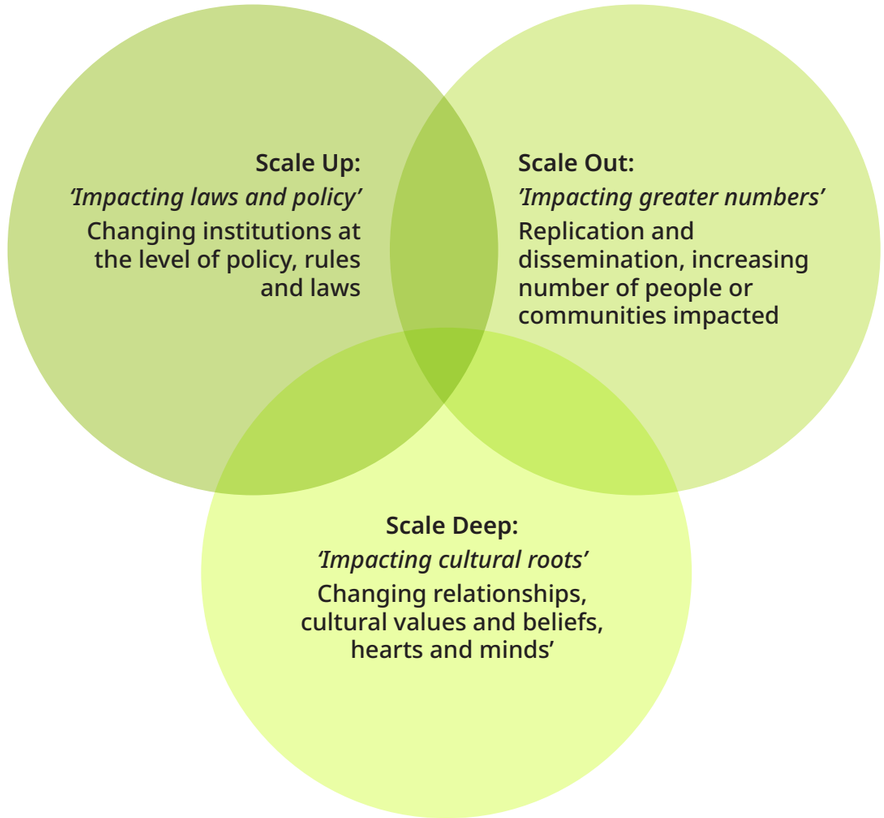
Using the Three Horizons Framework introduced earlier, the below grid presents how participatory and inclusive innovation in AI can affect a power shift at different levels of the system.

Figure 13: Mapping participatory and inclusive innovation at different horizons

Levels of the system (Power Shift Framework)	Horizon One (status quo)	Horizon Two (disruptive innovation)	Horizon Three (viable world)
Landscape *Refers to the broader environment, including macro-level trends like economic conditions, environmental challenges and societal expectations*	Dominance of established norms, systems and power structures within AI (e.g. commercial imperatives). Continued exclusion of publics and marginalised views and perspectives. At this level, Horizon One is stable and self perpetuating but unsustainable due to continued public and societal backlash.	Emerging pushback and pressures such as technology activism, class actions, pressure for addressing climate change and social justice issues, technological innovations that aim to engage the public, and experimental innovations that model AI inclusion. At this level, some innovations in Horizon Two point the way to the vision in Horizon Three, whilst others are at risk of being co-opted back into Horizon One to further reinforce the status quo.	A transformed environment in which AI considers the needs of all groups in society, especially those who are most excluded and marginalised. AI tools and technologies are primarily shaped by the goals of wellbeing, sustainability and equity (rather than commercial imperatives). AI works for people and society as a whole.
Regime *Represents dominant systems and structures, including industries, policies and cultural norms that resist or enable change.*	At Horizon One, existing industries, regulations, and cultural norms resist change in the form of inclusion, equity and participation to maintain stability and prioritisation of commercial imperative. They do little more than pay lip service to participatory and inclusive innovation, preferring instead to 'gatekeep' access to the development of and governance of AI technologies.	At Horizon Two, existing industries, regulations and cultural norms struggle between continuing to maintain the status quo and adopting participatory and inclusive innovations. Some organisations co-opt the participatory innovations back into their status quo-based models. Other organisations advance further, developing new and different AI models informed by participatory and inclusive approaches. However, these are siloed or unique approaches. In general there are fragmented adoption efforts with limited potential for scale or widespread adoption.	At Horizon Three, participatory and inclusive innovation in shaping, developing and using AI is usual practice. Horizon Three can become the new normal, or the new Horizon One. At this horizon, new AI technologies, uses, developments and applications are aligned with the goals of sustainability, equity and resilience
Deep roots of the system *The foundational beliefs, values, worldviews, and historical systems that underpin the status quo and influence all other levels.*	Historical and colonial structures, systems of oppression, cultural assumptions and vested interests perpetuate the current system, reinforcing the status quo. People developing, designing and governing AI systems lack awareness of their own positionality and privilege in doing so, or how this influences who AI systems work for and why. People excluded from AI systems lack the critical capacity, support and tools to overcome barriers and to shape the technologies themselves.	Historical and colonial structures, systems of oppression, cultural assumptions and vested interests become subject to increasing scrutiny and challenge, with growing awareness of the need for change, and willingness to effect it. More people excluded from AI systems develop some critical awareness of the technologies' impact. However, this remains limited/constrained to a small group of actors and individuals and is not widespread.	At Horizon Three, we see a complete transformation of foundational values and systems to reflect the need for AI systems to serve longer term, collective goals such as interdependence, global collaboration, and long-term well-being. All actors are able to shape and influence AI systems, with the requisite awareness of their own role in doing so.
Niches of innovation *Small-scale, experimental efforts that challenge the dominant regime and can seed transformative change.*	Small-scale, localised efforts to challenge dominant systems using inclusive participatory methods and approaches are developed and used, but these are marginalised, lack credibility and are widely discredited by powerful stakeholders. There is limited mainstream awareness of these innovations, and those who are aware do not feel these innovations are feasible.	There is growing awareness of the potential for participatory and inclusive innovation. Experimental projects and technologies begin gaining traction. Innovations are piloted, with some indication of emerging learnings, but there is a lack of a clear/systematic field or community of practice. There continues to be resistance from powerful stakeholders – the innovations remain niche/unusual.	Participatory and inclusive innovation is no longer considered to be 'niche'. Instead, the niche has become widespread and part of the regime, scaling up and out as well as deep. Participatory and inclusive innovation is fully integrated into processes of AI development, design, use and governance. Integrated and scaled innovations are also able to create meaningful and systems-level change.

To apply this lens effectively across all three horizons for systems change, we can recognise that effective approaches to participation and inclusion need to work at different levels of the system (deep roots, niches of innovation, regime and landscape). It's also important to blend a focus on old and new power. Effective approaches to participation and inclusion will need to aim to scale up (change and shape institutions, policies and laws), scale out (ensure participatory innovations have widespread use and uptake), and, crucially, to scale deep (transforming relationships and the culture of AI itself, investing in learning and field building, and providing support for lived experience and communities of practice). These elements are mapped out by Tatiana Fraser below.[87]

Figure 14: Three kinds of scaling – deep, up and out (Tatiana Fraser)



87 Moore, M.-L., Riddell, D.J. and Vocisano, D. (2015) 'Scaling out, scaling up, scaling deep: Strategies of non-profits in advancing systemic social innovation', *Journal of Corporate Citizenship*, 58, pp. 67–84. Available at: https://www.researchgate.net/figure/Three-types-of-scaling-and-their-main-strategies_tbl1_298971574 (Accessed: 13 March 2025).

Developing a public culture of AI and an ecology of participation

How, then, can participatory and inclusive practice in AI scale deep, scale up and scale out, effecting a real power shift and systems change? It will need to change and shape institutions, policies and laws, ensure participatory innovations have widespread use and uptake, and scale deep by transforming relationships and the culture of AI itself, investing in learning and field building, providing support for lived experience and communities of practice.[88] I call this the development of a public culture of AI, and an ecology of participation in relation to AI.

Meaningful change through participation in AI will, therefore, only be achieved if the cultural roots of how AI is designed and developed is transformed, through the creation of a public culture of AI design, use, deployment and development. What are the practical pathways towards a more participatory public culture of AI? This question can be answered by drawing from Burall's work on deliberative systems, Bussu's work on reclaiming participatory governance and assemblage theory, Chilvers et al's work on deliberative ecologies, and Gaventa's powercube in the context of participation and engagement.[89] These perspectives are all introduced and detailed below.

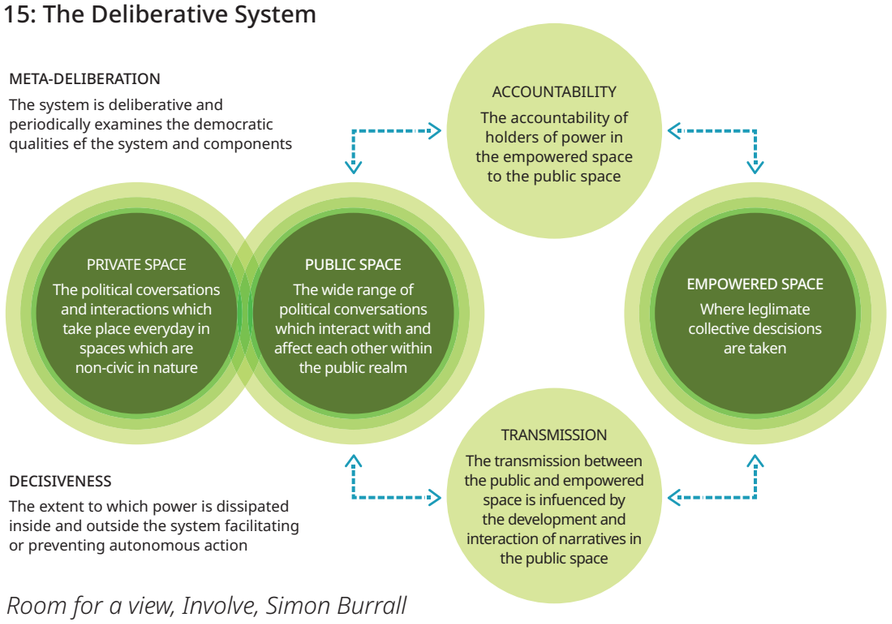
Deliberative systems: One approach is to consider the interaction of public and economic institutions as a system – what Burall refers to as a deliberative systems analysis. A critical failing in the democratic system is that it lacks that certain quality of dialogue and discussion that we have come to expect of properly democratic systems. Systems analysis identifies three spheres of political conversation, each of which is a deliberative space. The private space is made up of political conversations at home and in communal but bounded spaces such as the workplace or places of worship. The public space includes the media, civil society and citizens, and is where different narratives and arguments are tested and debated in open fora in which any citizen can participate (in theory). Critics of the deliberative systems analysis argue that the reality is very different, creating inequities in the scope to influence or to participate.[90] There is an empowered space where legitimate collective decisions are taken – for example, Parliament, or, more specifically to the focus of this document, within the commercial organisations in which AI is developed, designed and shaped.

88 ibid

89 Op cit

90 Holdo, M. and Öhrn Sagrelius, L. (2020) 'Why inequalities persist in public deliberation: Five mechanisms of marginalization', *Political Studies*, 68(3), pp. 634–652. Available at: https://doi.org/10.1177/0032321719868707 (Accessed: 13 March 2025).

Figure 15: The Deliberative System



Source: Room for a view, Involve, Simon Burrall

As Figure 10 indicates, and as thinkers like Dryzek and Burall argue, the health of the democratic system requires high levels of deliberative capacity in all its components. It can be judged first by the diversity and intensity of the exchange of views within each sphere, and second by the depth of interaction between the spheres. In particular, it is vital to a well functioning system that the empowered space (within which, for instance, technological decisions about AI are made) is accountable to the public sphere, and that there is effective transmission of narratives and strengthened accountability between the public and empowered spaces. Taking this systems approach allows us to evaluate the health of the overall system, not just the components and relationships within it. One of the qualities of the system is that it should, in some basic sense, be effective in delivering the outcomes expected by citizens. Outcomes emerging from AI and its use across society have fallen well short of standards considered acceptable to significant sections of the public. There are many examples of how data-driven AI technologies are failing society. These include the platform monopolies that have monetised human behaviour and data to influence the outcomes of elections,[91] an exam results prediction algorithm that privileged students from private, fee-paying schools in the UK,[92] and visa algorithms containing entrenched racism against immigrants[93].

It might be tempting to conclude that AI developers, policymakers and regulators simply need to up their game, and design and implement 'better' AI on behalf of citizens. Instead, what we need to focus on is the improvement of the functioning of the overall deliberative system, in order to improve the quality, diversity and interaction of deliberation over AI, as well as, relatedly, AI policy and regulation. This requires democratic and social change in several respects. First, it requires us to create a more understandable and empowering way of communicating about AI; one that feels more relevant, less intimidating and less distracting. It should both speak to citizens' lived experience and be able to explain it adequately. This is in contrast to much current media narrative, which veers between AI hype and AI scaremongering.[94] Second, it requires greater transparency about how AI works, to facilitate a meaningful dialogue between the public and the empowered space.[95] Third, it requires the creation of institutions and organisations developing AI that are more accountable to the public, and that can demonstrably act in more responsive ways to citizens. As Figure 10 highlights, accountability is a core requisite for the deliberative capacity of a system and yet it is notably absent in the relationship between those who design and develop AI, and those who use or are impacted by it.[96]

Assemblage theory: Recently, deliberative systems theorists have been critiqued for overlooking several important issues, due to the necessary oversimplification and idealisation of the deliberative system as set out above. Theorists argue that an approach that shifts the focus away from institutions towards practices, the messy realities of everyday design and practice, and the behaviour of actors within the ecosystems rather than the systems themselves, is preferable.[97] Others critique the way in which these approaches overlook how humans interact with society, technologies and democratic infrastructures and suggest an assemblage perspective. This implies investigating these changing human and nonhuman intersections as 'working arrangements' and through this identifying the emergent properties of assemblages as the 'grafting of new elements and reworking old ones; employing existing discourses to new ends'.[98]

Assemblage theory is a conceptual framework that emphasises the interconnectedness and dynamic nature of reality. It acknowledges the agency of various components, both human and non-human, and how these interact and coalesce to form temporary and ever-changing configurations known as 'assemblages'. As Bussu et al argue, 'the goal is to help overcome the linearity and oversimplification that sometimes characterise methodological approaches in the field, which tend to overlook the dynamism, complexity and messiness of participatory practices. Thus, it can help us to develop visions for participatory innovations, beyond just institutional characteristics and arrangements, as plural, experimental and multifaceted, highlighting interactions between different democratic practices, contexts and actors.'[99]

91 Rosenberg, M., Confessore, N. and Cadwalladr, C. (2018) 'Cambridge Analytica and Facebook: The scandal and the fallout so far', *The New York Times*, 4 April. Available at: https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html (Accessed: 13 March 2025).

92 Williamson, B. (2020) 'F**k the algorithm? What the world can learn from the UK's A-level grading fiasco', *LSE Impact Blog*, 26 August. Available at: https://blogs.lse.ac.uk/impactofsocialsciences/2020/08/26/fk-the-algorithm-what-the-world-can-learn-from-the-uks-a-level-grading-fiasco/ (Accessed: 13 March 2025).

93 Kelion, L. (2020) 'A-levels: Ofqual criticised over lack of algorithm details', *BBC News*, 4 August. Available at: https://www.bbc.com/news/technology-5365075 (Accessed: 13 March 2025).

94 Silverberg, D. (2023) 'Journalists are feeding the AI hype machine', *BBC News*. Available at: https://www.bbc.co.uk/news/business-68488924 (Accessed: 13 March 2025).

95 Bitzer, T., Wiener, M. and Cram, W.A. (2022) 'Algorithmic Transparency: Concepts, Antecedents, and Consequences – A Review and Research Framework', *Communications of the Association for Information Systems*, 52(1). Available at: https://aisel.aisnet.org/cais/vol52/iss1/16/ (Accessed: 13 March 2025).

96 Liu, X., Zeng, D. and Li, Y. (2023) 'Exploring the intersection of human-centered design and artificial intelligence: Opportunities and challenges', *AI & Society*. Available at: https://link.springer.com/article/10.1007/s00146-023-01635-y (Accessed: 13 March 2025).

97 Mouffe, C. (2023) 'Deliberative ecologies: A relational critique of deliberative systems', *European Political Science Review*. Available at: https://www.cambridge.org/core/journals/european-political-science-review/article/deliberative-ecologies-a-relational-critique-of-deliberative-systems/CDD5AA46BBF1222FCF60648E8C74331B (Accessed: 13 March 2025).

98 Bussu et al, INSPIRE: https://cop-demos.jrc.ec.europa.eu/citizen-engagement-projects/inspire-intersectional-spaces-participation-inclusive-resilient

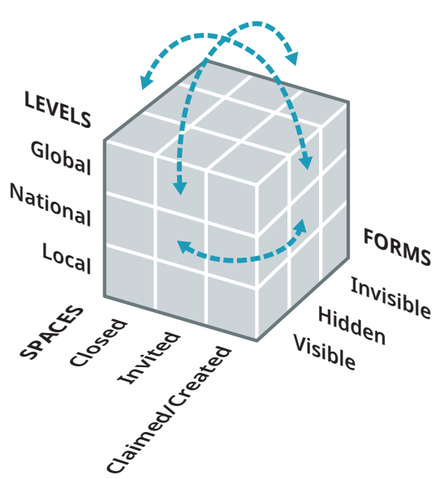
99 ibid

We turn, therefore, to deliberative ecologies and intersectional assemblages thinking. This complements deliberative systems thinking in answering how best to build a public culture of AI.

Different types of participatory processes

In designing participatory initiatives for inclusion in AI research, development and policy, the 'powercube' in Figure 11, developed by Gaventa, illustrates the multiple types of configurations of participatory design that are feasible or possible. It provides the scope for more creative approaches (rather than 'one-size-fits-all approaches') that serve communities most effectively. This matters, given the increasing discomfort many critical thinkers around participation have expressed over the ability of existing participatory approaches to account for the increasingly complex, diverse and interconnected roles of publics. The powercube suggests that there is scope for the 'unexpected' and the novel. The cube also points to the potential for participatory initiatives that can take place at a global, national and local level. It indicates how these processes and approaches can happen in spaces that are (historically) closed and invited, as well as claimed and created by the communities themselves. Last but not least, participatory processes can take forms that might be visible, delineated and defined, as with citizen juries, participatory budgeting, legislative theatre and citizen assemblies, but also hidden and invisible – as with the participation associated with many social movements.

Figure 16:
Gaventa's Power Cube for Community Power



Participatory ecologies: This range of configurations invites the possibility for what professors Jason Chilvers, Helen Pallett and Tom Hargreaves describe as a broad 'ecology of participation' in socio-technical change.[100] As important as the multiple possibilities and accessibilities of the *forms* of participation (minoritised communities cannot be included in spaces they cannot access), is the overall approach to *how we embark upon the activity of participation itself*. Examples of participatory spaces that have been claimed and created at the local level include The AI Under Watch project run by the Migrants' Rights Network, the Workers' Observatory mobilising platform gig workers, Reimagining AI and Digital Justice led by UNJUST C.I.C, and the Reclaim the Algorithm project led by the People`s Speak initiative.[101] There is also a growing tradition of arts-based methods, such as legislative and forum theatre, pioneered by political theorist Augustus Boal and undertaken by organisations such as Cardboard Citizens. There is also the Choral Data Trust experiment piloted between Serpentine Gallery and the Alan Turing Institute.[102]

Invited spaces at the local and the national level include deliberative processes and methods. Examples of these are citizen councils such as the Ada Lovelace Institute Citizens' Biometrics Council,[103] The Royal Academy of Engineering and People's AI Stewardship Summits,[104] the NHS Coventry and Warwickshire co-design of a Waiting List prioritisation AI tool[105] and citizen juries on AI algorithms, such as those run by the Scottish Government on QCovid®.[106]

Closed spaces have historically excluded lay perspectives. However, there is increasing practice of ensuring adequate lay perspective representation on boards and advisory decision-making bodies in the health data landscape, and the AI governance landscape in particular. At the global level, there is a live conversation about the creation and delivery of a Global Citizens Assembly on AI as an invited space.[107] This is supported by the bottom-up deliberations of community assemblies claimed and created worldwide in a range of different contexts by civil society, including international development and humanitarian projects such as the work undertaken by Nesta to localise AI as part of crisis response in Turkey, using a new participatory AI assurance methodology on working with impacted communities.[108]

These are just some of the examples of the initiatives that include public voices in AI, which, together, form an ecology of civic participation in AI. Open-source participatory technologies and tools such as Decidim and VTaiwan are also enabling digital participation at the local, national and global level, as pioneered by the policymakers of Brazil and Taiwan.[109] The Brazilian government has used open-source modular digital engagement platform Decidim to establish a digital participatory portal called Participativo (which has engaged 1.4 million people). VTaiwan is an open consultation process, consisting of online and in-person discussions, which brings together up to 200,000 experts, government officials, stakeholders and citizens to create consensus and recommendations for national legislation.

100 Chilvers, J., Pallett, H. and Hargreaves, T. (2018) 'Ecologies of participation in socio-technical change: The case of energy system transitions', *Environmental Science & Policy*, 88, pp. 33-44. Available at: https://www.sciencedirect.com/science/article/pii/S2214629618303025 (Accessed: 13 March 2025).

101 The Digital Good Network (2023) *Public Voices in AI: Announcement of Funded Projects*. Available at: https://digitalgood.net/public-voices-in-ai-announcement-of-funded-projects/ (Accessed: 13 March 2025).

102 Participedia page on forum theatre: https://participedia.net/method/forum-theatre

103 Peppin, A. and Patel, R. (2020) *Citizens' Biometrics Council*. Ada Lovelace Institute. Available at: https://www.adalovelaceinstitute.org/project/citizens-biometrics-council/ (Accessed: 13 March 2025).

104 McCarthy, N. and Lane, G. *People's AI Stewardship Summits*. Royal Academy of Engineering. Available at: https://raeng.org.uk/people-s-ai-stewardship-summits (Accessed: 13 March 2025).

105 The Strategy Unit (2022) *Coventry Waiting List Prioritisation Deliberative Report*. Available at: https://www.strategyunitwm.nhs.uk/sites/default/files/2022-09/Coventry%20waiting%20list%20prioritisation%20delib%20-Full%20Report.pdf (Accessed: 13 March 2025).

106 Scottish Government (2021) *Citizens' Jury on COVID-19: Report of the Jury's Conclusions and Key Findings*. Available at: https://www.gov.scot/publications/citizens-jury-qcovid-report-jurys-conclusions-key-findings/ (Accessed: 13 March 2025).

107 Davies, T. (2024) *Global Deliberation on AI*. Connected by Data. Available at: https://connectedbydata.org/resources/global-deliberation-ai (Accessed: 13 March 2025).

108 Berditchevskaia, A., Peach, K. and Stewart, I. (2020). *Localising AI: How to make artificial intelligence work for local communities*. Nesta. Available at: https://www.nesta.org.uk/report/localising-ai/ (Accessed: 13 March 2025).

109 *Decidim: A digital platform for democratic participation*. Available at: https://decidim.org/ (Accessed: 13 March 2025).

What makes for success across the range of participatory initiatives and how can we foster the ecology of participation in the field of AI?

Fostering this ecology of participation can be made possible by identifying the different types of participatory initiatives that have emerged, and ensuring that all of these categories of participatory approaches are supported and enabled. I have identified that participatory processes can be situated at the regime or institutional level, operating **top-down**, at the deep roots or niches of innovation level, operating **bottom-up**, and mediating between institutions and the grassroots or civil society, an approach I have called the **messy middle**.

Top-down: Top-down models include initiatives such as citizens' assemblies and juries, which aim to integrate citizen input into policy discussions at the local, national and international level. Deliberative polling similarly takes a predominantly top-down approach. Other examples in the UK include the Sciencewise programme, which funds and delivers a series of public dialogues on data and AI, driven primarily by UK policymakers, on issues that have been identified as being of particular interest. Top-down initiatives are usually commissioned, sponsored and championed by policymakers. They tend to leave limited space for agenda setting by the participants, since the topic of discussion has usually already been set by organisers.

Bottom-up/grassroots: In contrast, bottom-up initiatives often emerge from or as part of broader social movements. They can collectively identify and govern issues, ensuring that interventions are deeply rooted in local contexts and needs. Examples include the AI Under Watch project run by the Migrants' Rights Network, the Workers' Observatory mobilising platform gig workers and the Reclaim the Algorithm project led by the People's Speak initiative. In such initiatives, local communities actively engage in identifying barriers to technologies and AI, and co-creating solutions. The main challenge that bottom-up approaches experience and encounter is marginalisation – as they are already disconnected from regime-based power, they often struggle disproportionately to exert influence on policymakers and technologists. Lacking the ability to access the resources necessary to make their voices heard, they rely predominantly on methods such as civic activism and campaigning. Such initiatives are at risk of lacking access to the importance of particular forms of expertise (technical, scientific, policymaking, economic and otherwise) which are also required to meaningfully exert change.

Messy middle: The messy middle is represented by initiatives that seek to blend structured institutional input and engagement with grassroots participation, navigating the complexities and diversities of stakeholder engagement to foster inclusive and impactful outcomes. An example of this is the Ada Lovelace Institute deliberative inquiry in partnership with the Young Foundation, Liverpool Civic Data Cooperative and the People's Panel on AI, convened by Connected by Data, and the Nesta deliberations on AI working in the humanitarian context within Turkey.

The grid below sets out the respective benefits, as well as disadvantages, of these three approaches:

Figure 17: Benefits and drawbacks of different forms of participation

Approach	Benefits	Disadvantages
Top-down	– Efficiency, speed and coordination: Swift implementation of participation on a large scale, usually driven and mandated by significant resources. – Can enable coordination among different actors and parties. – Expertise and resources: Leverage significant expertise and resources, are often relatively well resourced, and have the potential to draw from international best practice. – Uniformity: Ensures the standardised/ co-ordinated implementation or international rollout crucial for issues requiring uniform or global approaches (e.g. on issues such as AI and climate change).	– Lack of local context: May overlook local nuances and cultural contexts, can be surface level, and lack the capacity to 'go deep'. – Findings and insights may be cherry picked or otherwise co-opted. – Limited stakeholder engagement: At risk of marginalising local voices and reducing community buy-in. – Issue identification: At risk of overlooking which are the right issues to hone in on, or discounting existing best practice where it exists. – Resistance: Imposing solutions from the top can lead to local resistance and non-compliance, and also might contribute to mistrust in contexts where power and authority is not already trusted.
Bottom-up/grassroots	– Local relevance: Tailors interventions to specific local needs and contexts, enhancing effectiveness and sustainability. – Empowerment: Involves local populations directly in decision-making, fostering ownership and responsibility. Can most effectively contribute towards building the capacity of those who are most marginalised and disempowered. – Innovation: Leads to innovative solutions grounded in real-world experiences, ensuring that innovation responds appropriately to need rather than being innovation for innovation's sake.	– Scalability: Struggles with applying successful local solutions to broader contexts. – Resource constraints: May lack the resources and expertise of larger organisations and centrally mandated structures – largely due to the limited visibility of the organisations working from the grassroots. – Coordination challenges: A multitude of local efforts and the coordination of numerous efforts can lead to inconsistencies and duplication. There is a risk that when similar issues emerge in different contexts this is overlooked and left uncoordinated.

Approach	Benefits	Disadvantages
Messy middle	– Balanced approach: Combines top-down expertise with bottom-up relevance for scientifically/technically sound and locally appropriate solutions. – Broader and more inclusive approaches to decision-making: Involves a range of stakeholders in dialogue, fostering more comprehensive and inclusive policymaking processes. – Adaptability: Flexible nature of this approach allows adjustments based on multi-level feedback, enhancing responsiveness to changing circumstances, while ensuring coordination challenges are addressed.	– Complex/challenging to balance: Integrating multiple perspectives can result in complex and cumbersome decision-making processes. Sometimes it can be unclear how best to strike the balance between top-down and bottom-up approaches. – Potential conflicts: Balancing diverse interests and priorities can lead to conflicts or diluted effectiveness as it is not always clear how best to negotiate particular trade-offs across the different stakeholder groups and interests. – Resource intensive: Requires significant time, effort and resources to manage interactions and ensure meaningful participation between grassroots and power holders, and requires significant investment in time and resources to ensure grassroots organisations can meaningfully influence power holders/change. – Gatekeepers represent the voices of the public, rather than participants in the processes representing themselves.

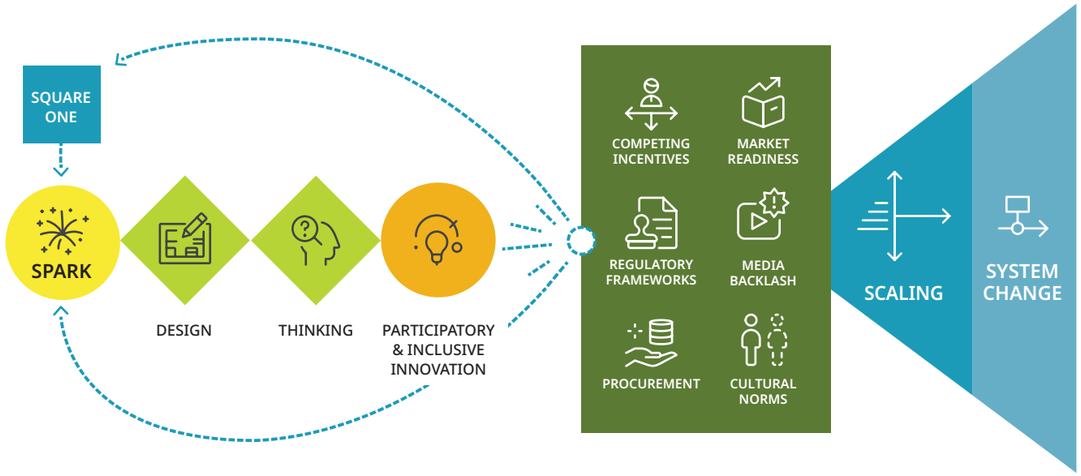
Overcoming barriers to participatory and inclusive innovation

The Three Horizons Framework also refers to some of the types of barriers that need to be overcome in order to make inclusion and participatory innovation a reality. These include, for instance, the instincts of the systems to preserve themselves, through entrenched power structures and cultural inertia, which can prevent participatory innovations from being embedded into or transforming technology processes. Conway indicates that we need to question the idea that innovation scales linearly.[110] It is necessary to recognise that ordinarily, there is a process through which innovation itself experiences rejection by what is described as the 'system immune response' – the tendency within the system to reinforce the existing status quo.

110 Conway, R. (2020). *Designing for an unpredictable future*. Medium. Available at: https://rowan-conway.medium.com/designing-for-an-unpredictable-future-c73d01940c00 (Accessed: 13 March 2025).

Figure 18 below shows that the key to scale and system change is to develop strategies that overcome the inevitable resistance that will be experienced. Some of the reasons for resistance include competing incentives (cultural and commercial), regulatory frameworks that do not reward participatory and inclusive innovations, an absence of market or social readiness, media backlash and existing cultural norms.

Figure 18: Why participatory and inclusive innovations may not scale and achieve systems change



While these may indeed be reasons for resistance, as the Conway article reveals, it can be observed that some are also key levers for change. For instance, through changing incentives (cultural and commercial) and procurement structures, creating regulatory frameworks that reward participatory and inclusive innovations, creating improved market conditions, and changing social, media and cultural readiness, we can ensure that participatory and inclusive innovation can overcome resistance and scale.

Indeed, leading systems change theorist Donella Meadows finds that the most influential ways to exert or leverage system change are as follows: through (5) enabling access to information flows, (4) changing the rules of the system (incentives, punishment and constraints), (3) shifting the distribution of power over the rules of the system, (2) changing the goals of the system, and (1) changing the mindset or paradigm out of which the system (goals, power structure, rules, culture) arises. She describes these mechanisms as 'leverage points' – places to intervene in a system. These are the leverage points that the field, and those seeking to involve the public, need to use to realise systems change.

Moving towards Horizon Three: Three key takeaways

Transitioning from Horizon Two (disruptive innovation) to Horizon Three (viable world) requires overcoming some of the significant barriers to system change outlined above. As already described, these barriers often stem from entrenched structures, resistance to change and the uncertainty associated with transformative shifts.

Some of the strategies I suggest include three broad routes to action that are mutually complementary and reinforcing – (1) building coalitions for change and communities of practice, (2) scaling and investing in participatory innovations at Horizon Two, and (3) addressing the system immune response/ cultural resistance through shifting and changing incentives for participatory AI.

1. **Build and invest in broad coalitions for change and communities of practice**

   It will be important for stakeholders to invest in communities of practice, and to support the development of accelerators in the field of participatory innovations for AI. In doing so, funders might consider fostering partnerships between publics, citizens, communities, governments, businesses, civil society and academia, to scale and leverage existing best practice. As part of this work, stakeholders might wish to consider actively investing in multi-sectoral collaborations and partnerships that benefit from participatory innovation.

2. **Scale and invest in disruptive participatory innovations as catalysts, using the framework for including public voices in AI**

   Starting from strategically investing into some of the innovations at Horizon Two, it will be important for stakeholders, funders and investors to invest in and work to scale existing good practice and successful niche innovations from Horizon Two that align with Horizon Three principles. Developers, policymakers, regulators and civil society can pioneer, develop and invest in participatory innovations that are situated across the public engagement spectrum, aiming to shift power in different ways. In doing so, existing innovations can be connected up with existing partners and technological innovators, creating room for the testing and refining of innovative approaches.

3. **Tackle the social and commercial barriers to participatory and inclusive approaches to AI**

   Stakeholders, funders and responsible AI pioneers can consider the importance of changing hearts and minds when it comes to the value that participatory innovations bring to shaping AI in responsible ways. They can invest in work to reframe and shift narratives and values within AI development. For instance, they can encourage responsible AI more broadly, and push back against the 'move fast, break things' mantra, which prioritises technology development at a fast rate of change, in contrast to more sustainable, durable ecosystems of technology development.[111] This will need to be supported by the right governmental, commercial, economic, regulatory and policy incentives – those that reward alignment with public and societal values over short-term gains.

Concluding thoughts

This report – and its associated framework – underscores the complexities and opportunities involved when we include public voices in AI development and deployment. It highlights the necessity of moving beyond surface-level engagement to foster meaningful, critical, collaborative and inclusive participation. By doing so, it aims to help stakeholders avoid the 'empty ritual of participation' described by Arnstein and move toward approaches that are more equitable, empowering and co-creative. By addressing the challenges inherent in this work (described as the 'system immune response') – such as misaligned incentives, limited resources, and systemic barriers – it emphasises the importance of rethinking current practices to better serve societal goals.

By scaling and investing in disruptive innovations that promote participatory practices, while building coalitions to safeguard against their co-option by entrenched systems, we can begin to catalyse a shift toward a more inclusive and justice-oriented AI ecosystem. This is envisaged in this paper as Horizon Three, moving beyond the waves of the status quo (Horizon One) and the emerging sparks of innovation (Horizon Two). This report sets out a pathway to support these efforts, offering both guidance and provocation to inspire action and refine strategies for embedding participation into the heart of AI design and governance.

Through the introduction of a self-assessment workbook below, I also provide organisations, practitioners and technologists with a structured tool to evaluate and enhance their approaches to public inclusion. As the framework above illustrates, however, a series of disparate one-off participatory initiatives do not add up to a whole. A deeper, more systemic shift is necessary to embed participation across all stages of AI development – upstream, midstream, and downstream – ensuring that AI systems draw from and connect with rather than undermine deep roots – the lived experiences, needs and expectations of diverse communities. This requires centering the voices of those who are most excluded by existing technologies, recognising that their unique and differential experiences contain valuable knowledge that can shape improved AI outcomes, and building a culture of public collaboration and openness in the design and governance of AI.

111 Zengler, T. (2019). *Why "move fast and break things" doesn't work anymore*. Harvard Business Review. Available at: https://hbr.org/2019/12/why-move-fast-and-break-things-doesnt-work-anymore (Accessed: 13 March 2025).

Self-assessment workbook

Ensuring that public voices are meaningfully included in AI development requires more than just theoretical commitments – it demands structured, actionable processes. This self-assessment workbook builds on the principles of critical thinking, inclusivity, equity, transparency and accountability, the goals of effecting a shift in power and transitioning towards a just and viable world, and the explorations of how participation in shaping and influencing AI might work. It provides a practical tool for organisations to evaluate and enhance their engagement strategies. It can support anyone who is seeking to initiate a participatory movement – from the grassroots, the messy middle or the institutional level – to plan their participatory process. It can also help assess how the process is developing as it evolves, and its ongoing impacts and outcomes. By systematically assessing key dimensions such as stakeholder involvement, responsiveness and impact measurement, practitioners developing and running participatory AI processes can identify strengths, address gaps, identify some potential risks and take concrete steps towards more equitable AI governance.

By introducing a self-assessment workbook, this paper moves beyond the theoretical towards the practical. It therefore offers a tool for organisations, practitioners and technologists to critically evaluate and enhance their approaches to public engagement in shaping AI.

A critique of public participation in emerging technology policy is that it can be difficult, resource intensive and time consuming to enable and to foster, given the complexity of the issues at hand. There are also numerous barriers, including a lack of political and commercial will. Furthermore, at the time of writing, the participatory turn in AI is still a new and emerging field, with new forms of knowledge emerging, where there still remains limited expertise and where that expertise is siloed. This is an effort to respond to some of these issues, providing support and guidance to those who wish to design, deliver and commission participatory practice in AI, where it is needed.

Why use this workbook?

Given the complexity of designing good and effective participation outlined above, it is helpful to develop tools that can prompt and orient towards good practice, acting as a compass that points in the right direction.

As a non-academic and practitioner, and drawing on over a decade's experience in participatory practice, I have therefore developed a workbook that I would personally find useful in approaching the design, delivery and evaluation of participatory approaches to AI. The questions below combine this practical experience with research about where and how participatory methods and approaches are situated in effecting wider systems change, and their role in relation to data and AI. The workbook aims to support organisations and institutions to move beyond tokenistic efforts toward practices that are genuinely collaborative, equitable and empowering, addressing the needs of those who are most impacted by AI systems and influencing change directly. Collectively, this can set the stage for a public culture of collaboration in AI design and governance, ensuring that these technologies reflect shared societal goals and promote a just and inclusive future.

Approach to using the workbook

The below section lays out a self-assessment workbook to provide institutions, organisations and practitioners with a way to review and enhance their participatory strategies and practices in the field of AI. I encourage you to think about public participation and inclusion as a culture, or a way of doing good technology, policy and regulatory design.

As such, inclusion and participation should be continuously embedded in the way that you make decisions, work and operate in cultures that have historically been technocratic and, by their very definition, exclusionary of the public. Because of this, you will have blind spots, and this document aims to support you to think and work in more reflexive ways about what those blind spots are likely to be, and, importantly, to take action to address them.

The workbook is therefore best used as a living document, supporting you to think reflexively about your approach to engagement at the beginning, middle and conclusion of an engagement project. It is appropriate both for people who are currently delivering a key participatory and inclusion project and/or are evaluating such projects, and for people who are at the earliest stages of project conceptualisation and design.

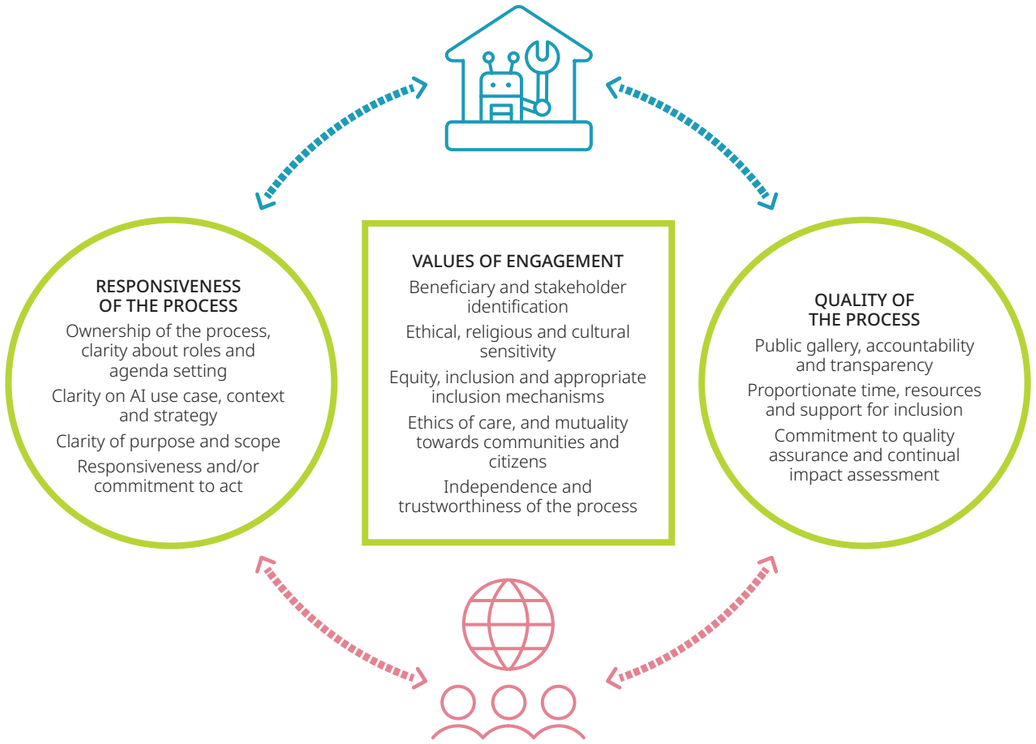
If your approach to engagement is likely to be continued, institutionalised and ongoing, you might want to think about the right touchpoints at which to review and assess your practice against these questions and develop processes to ensure that the questions asked within this framework are revisited.

How was the workbook developed?

This workbook draws on the thinking in this report, the framework it presents, and several co-design workshops, with the People's Panel on AI, the Public Voices in AI team and co-design workshops with practitioners and policymakers convened in Liverpool and in Barcelona. It also builds on the following diagram, which was initially developed in the Royal Society of Arts report Building a Public Culture of Economics.[112] The diagram articulates three strands that ensure citizen engagement is effective: (i) the quality, independence and design of the process (ii) the values at the heart of the participatory and engagement process and (iii) the way in which institutions respond to and engage with the process.

112 Patel, R., Gibbon, K., and Greenham, T. (2018). *Building a Public Culture of Economics*. Royal Society of Arts, Commerce and Manufactures

What makes for a good participatory and inclusion process in relation to AI?



Source: Adapted from Building a Public Culture of Economics, Patel, Gibbon, Greenham, Royal Society of Arts, Commerce and Manufactures (2018). This diagram was originally developed by Reema Patel, working with Kayshani Gibbon in 2018.

It builds on a research paper called Strategies for Including Public Voices in AI and emerges from a review of existing frameworks. It draws on prior RSA research on how institutions can most effectively assure the quality of their engagement approaches, the OECD's good practice principles for deliberative engagement, and Sciencewise's Quality in Public Dialogue framework.

The framework aims to provide a structured approach to understanding and improving how practitioners, technologists and institutions can engage with communities and the public. The core elements that emerge across these key areas include the importance of these twelve dimensions, which have been mapped against the three strands:

RESPONSIVENESS OF THE PROCESS

1. Ownership of the process, clarity about roles and agenda setting
2. Clarity on AI use case, context and strategy
3. Clarity of purpose and scope
4. Responsiveness and/or commitment to act

VALUES OF ENGAGEMENT

5. Beneficiary and stakeholder identification
6. Ethical, religious and cultural sensitivity
7. Equity, inclusion and appropriate inclusion mechanisms
8. Ethics of care, and mutuality towards communities and citizens
9. Independence and trustworthiness of the process

QUALITY OF THE PROCESS

10. Public gallery, accountability and transparency
11. Proportionate time, resources and support for inclusion
12. Commitment to quality assurance and continual impact assessment

Working through the workbook

Work through the workbook below to:

1. **Map your current status**: assess how you are doing on each of the criteria.
2. **Define specific actions**: identify some specific actions to improve each area or to learn more about it, alongside realistic target dates for achieving these improvements and clear allocation of responsibility to specific individuals or teams.

Responsiveness Of The Process

Dimension	Key Questions	Goal of the Dimension	Current Status (1)	Areas for improvement
1. Ownership of the process, clarity about roles, and agenda setting	*- Who owns (is paying for/resourcing and convening) the process? Why are they doing so?*	*- Clarity about who is driving the process and why*		
	- If the process is owned by a group or organisation that is not reflective of the group being included, is there scope for participants to shape and set the agenda? This can take the form of co-creation and co-design at an early stage.	*- Ensuring that all groups can benefit from the way the process is framed and the agenda is set*		
	- Is it clear what the roles for the process owners/ convenors are, and the limits to those roles?	*- Clear delineation of roles and responsibilities (facilitators, process designers, convenors, commissioners and participants)*		
2. Clarity on AI Use Case, Context and Strategy	*- Is it clear what the participation/inclusion is taking place in relation to (i.e. a use case, data set or other AI intervention)?*	*Clear sense of what participation will shape – consideration of historical and cultural factors*		
	- Is there a clear understanding of the broader environmental, societal, cultural and technological context shaping the AI use case?	*- Participation does not think about AI narrowly but within the broader eco-socio-technical context. Think about implications of the AI use case for environment and planetary boundaries, as well as its social and economic impacts*		
	- Is there a defined strategy for inclusion and participation tailored to this use case/context?	*- Clearly defined strategic approach to inclusion and participation*		
	If you have an existing strategy for your use case or application, is there any attention/time/ thought given to inclusion or equity within this overarching strategy?	*- Clearly defined approach to equity*		

Dimension	Key Questions	Goal of the Dimension	Current Status (1)	Areas for improvement
3. Clarity of Purpose and Scope	*- Are the goals of the participatory process, approach to inclusion and/ or public engagement clearly defined and communicated – in relation to your use case and the broader context?*	*Clear articulation of goals*		
	Is the scope of engagement realistic and well-defined?	*Well defined scope*		
	Is it clear to all actors what the limits of the engagement/scope are?	*Well defined boundaries*		
4. Responsiveness and Commitment to Act	*- How will you and/ or the practitioner/ organisation/technologist demonstrate a commitment to act on public input throughout the inclusion process and beyond?* *If you are not a technologist or power holder able to make change happen directly, have you considered strategies to influence change?* *These can include mobilising:* *– media and communications* *– legal mechanisms* *– coalitions and campaigns* *– advocacy strategies and briefing/engaging policymakers and technologists through the process*	*- Evidence of action based on input.*		
	- Are there mechanisms in place (eg through the design and development of the process, such as accountability and oversight groups) to ensure responsiveness?	*- Clear timelines and responsibilities*		
	- Who will hold the organisations and individuals who have the power to respond to account for its responsiveness to the inclusion process?	*- Mechanisms (eg governance or oversight bodies) in place to ensure that the inclusion/ engagement process effects real change.*		

Values Of The Participation And Inclusion Process

Dimension	Key Questions	Goal of the Dimension	Current Status (1)	Areas for improvement
5. Beneficiary and Stakeholder Identification	*Have all relevant stakeholder groups been identified? Consider a stakeholder mapping process with your team.*	*Stakeholders are mapped*		
	Of your listed stakeholders, is there a clear understanding of who might be adversely affected by the intervention/ use case, and why? Consider undertaking a power mapping process to identify who has the most and least power	*Stakeholder diversity*		
	Are marginalised or underrepresented communities identified and clearly included? *Consider the following questions:* *1. Who has a stake in the outcomes that emerge from the design, development, use and likely impact of this AI technology?* *2. Who is included in the benefits and who is excluded from the benefits?* *3. Who is likely to be impacted by any harms?* *4. Who is less likely to be impacted by any harms?* *5. Who is most likely to be directly affected and impacted, either benefiting or being adversely impacted?* *6. Who is most likely to be overrepresented and/ or underrepresented in the data training the AI system and technology?* *Consider that people impacted by the use of AI may experience multiple disadvantages. Consider what these may be and how they may intersect in relation to the AI use case*	*Mechanisms to engage marginalised communities*		
	Of your listed stakeholders, is there a clear understanding of who might benefit from the intervention and use case, and why? Is there also a clear understanding of who might be adversely impacted and why?	*Stakeholder diversity*		

Dimension	Key Questions	Goal of the Dimension	Current Status (1)	Areas for improvement
	Are beneficiaries identified and clearly included, or have they previously been engaged and included?	*Mechanisms to engage beneficiaries*		
6. Ethical, Religious and Cultural Sensitivity	*What form of governance or ethics review has been undertaken, to guide the engagement process, if any?*	*Review/Representation of governance and ethical concerns.*		
	Has the organisation/ practitioner taken account of cultural and religious sensitivities, e.g. considering timings and scheduling of key events and key holidays?	*Design, scheduling and practical matters accounts for religious and cultural contexts and differences. This may vary. For instance, it will be important to consider scheduling for gig workers, and timing of holidays for engagement with religious minorities.*		
	Does the facilitation and design team running the engagement process have a good understanding of the cultural and religious context?	*The facilitation and design team understands the real world cultural context in which stakeholders are working/operating in, and are able to navigate this.*		
	How are cultural values considered in the delivery and execution of the engagement?	*- The team are trained and have worked to develop a clear understanding of key ethics, religious and cultural considerations affecting the communities engaged.*		
	Are any of the topics or questions for discussion likely to be harmful, distressing, sensitive or triggering? If so, plans need to be in place to provide support and address risk and harm, drawing from trauma-informed and other approaches. *Consider using the National Trauma Transformation Programme's (Scotland) resources here.* *Consider the nature of the topic here, and whether discussion of any issues may make it hard for someone to participate.*	*Consideration of a trauma informed approach*		

Dimension	Key Questions	Goal of the Dimension	Current Status (1)	Areas for improvement
	Is there a plan to involve young people? If so, safeguarding risks and guidelines will need to be considered, in line with the appropriate legal governance of the country/context. *Please consult a qualified expert if you do not already understand your legal requirements.*	*Consideration of youth safeguarding if young people are involved*		
7. Equity, Inclusion, Participation and Appropriate Engagement Mechanisms	*What methods and mechanisms are being used to include public voices?*	*Variety and suitability of formats depending on the needs of the stakeholders and beneficiaries. These may vary depending on the needs of stakeholders.*		
	Are these mechanisms accessible and inclusive? *Is the engagement mechanism open, closed or invited? If not open, do you have strategies in place to include a diversity of voices?* *If open, how will you ensure that minority experiences and viewpoints are adequately considered?* *If invited, what mechanisms do you have to ensure that there is a broad/reflective spread or sample of the population in the group?*	*Accessibility measures are in place to support and encourage engagement* *The group being engaged is truly diverse.*		
	Has a needs assessment been undertaken to identify any potential barriers for stakeholders?	*A needs assessment for stakeholders is in place, and clear mapping of stakeholder needs*		
	Do your mechanisms for inclusion and participation clearly address these needs?	*Participation and inclusion approaches respond to the needs*		

Dimension	Key Questions	Goal of the Dimension	Current Status (1)	Areas for improvement
	Do all participants and stakeholders have the resources (time, incentives and accessible information) to be able to participate effectively? It will be helpful to focus on your most marginalised stakeholders to enable for inclusion here.	*Adequate time, knowledge and resources to participate*		
	How does your approach actively address power imbalances and contribute to increasing equity and inclusion, and combating systems of oppression? Refer to your power mapping, and identify strategies for inclusion.	*Clear plans for addressing power imbalances. Inequality is deeply embedded in our systems, structures, and institutions. We need to acknowledge, describe, and engage with those power asymmetries when we do public participation.*		
	If the proposed approach for engagement is digital a plan for digital inclusion should be in place.	*Plans for digital inclusion (if appropriate), working to ensure the accessibility of the digital process.*		
8. Ethics of care and mutuality towards communities and citizens	*How does your engagement practice ensure that we recognise and value the lived experiences of affected communities?*	*Approaches to data collection, reporting and management that engage participants*		
	How do you demonstrate reciprocity – ensuring that communities benefit from their participation and contributions?	*Focusing on the impacts on the participants, as well as on technology and policy design- ensuring that their capacity to exercise agency is built through the process.*		
	What measures are in place to ensure that the process prioritises the care and well-being of the participants over maximising the 'outputs' from the process?	*Understanding participant and community needs, and responding to them, even if it means that less is covered in the time available.*		
	Has thought being given to engaging with communities in an ongoing and iterative manner, rather than through one-off consultations?	*Plans for continued engagement and communication.*		

Dimension	Key Questions	Goal of the Dimension	Current Status (1)	Areas for improvement
9. Independence and trustworthiness of the engagement process	*How do you ensure that the participatory process is free from undue influence by private or political interests?*	*Arms length independence of the process through independent facilitation*		
	What governance safeguards do you have in place to maintain the integrity and impartiality of the participation and inclusion process? (These can be independent oversight bodies to review the process as it evolves)	*Mechanisms to scrutinise the process as it evolves*		
	How do we ensure that independent oversight bodies or mechanisms are in place to review AI-related decisions?	*Mechanisms to scrutinise the process as it evolves*		
	What mechanisms exist for communities and external stakeholders to hold you accountable if trust is compromised?	*Mechanisms to scrutinise the process as it evolves*		
	Independence of the information and the material that has been communicated - has the information about AI systems being shared with participants been independently verified by a trustworthy source? (A panel of specialists to sense check information can be established to review content and materials).	*Mechanisms to stress test and check quality of the information*		
	If there is a conflict of interest between what emerges in the process, is there a key organisation, institution or body that can help navigate or mediate that conflict of interest?	*Conflict of interests policy and approach*		

Quality Of The Process

Dimension	Key Questions	Goal of the Dimension	Current Status (1)	Areas for improvement
10. Public gallery, accountability, and transparency	*How are the goals, processes and outcomes of the engagement process communicated to the people participating/ being included, as well as to the wider public?*	*Communication strategy for the process*		
	Is the AI use case presented in a way that is accessible and understandable, without jargon or unnecessary complexity?	*Clear and accessible information and materials*		
	Are there mechanisms for addressing questions or concerns?			
	Are these engaged within a public domain and context? (An NDA or other commercially sensitive document is a red flag).	*Clear audit trail so anyone can understood how the process took place, and how it was incorporated into decision making*		
	Who analyses the data from, determines and shapes the key recommendations and findings from the participation process? Consider techniques and methods to ensure that the presentation of the key insights, advice and recommendations are co-created and/or created by the participants in the process	*Clear audit trail that shows how participatory process resulted in the findings/ key recommendations or advice from the process.*		
	Will the findings from the engagement process be made publicly available and transparent? If a researcher would like to understand more about how the process worked, would they be able to?	*Openness about the findings and consequences*		
	Will the way the engagement process shaped decision making be made available and transparent? ('You said, we did', as opposed to 'You said, we hid'.)	*An explanation from the power-holding organisation as to how the engagement process shaped decision making (if this is what was promised)*		

Dimension	Key Questions	Goal of the Dimension	Current Status (1)	Areas for improvement
11. Proportionate Time, Support and Resources	*Have there been adequate and proportionate resources allocated for meaningful and ongoing engagement? These will need to be proportionate to the complexity of the use case and issues at hand*	*Enough time for the engagement activities to take place meaningfully* *Please consider and balance the risk of participation fatigue (a longer engagement process is not necessarily better for everyone)*		
	Have there been adequate and proportionate resources allocated for meaningful and ongoing engagement? These will need to be proportionate to the complexity of the use case and issues at hand	*Proportionate resources allocated for engagement*		
		Clear, accessible materials supporting the engagement.		
	Are there support structures, such as resource libraries, training, networks, and communities of practice, in place to support ongoing inclusion/ participation?	*Support structures in place to support and enable ongoing inclusion and participation. In particular, staff and organisation has increased awareness and understanding about issues of marginalisation, exclusion and control, and their own potential role in reinforcing asymmetries of power.*		
12. Commitment to Quality Assurance and Continual Impact Assessment	*How is the effectiveness of the approaches to participation and inclusion assessed and quality assured throughout the process as well as in retrospect?*	*Clear approach to ensuring and securing the quality of the process with regard to key benchmarks, such as those outlined in the Sciencewise Quality in Public Dialogue Framework. This framework can also form a useful basis*		
	What evidence exists that public input shaped AI outcomes in this specific context, and continues to do so?	*Clear/Ongoing documentation of public contributions and impact.*		

Methodology

This report and framework has been developed and informed by extensive mixed-methods research and cross-sectoral consultation that encompasses:

- a rapid review and synthesis of existing conceptual frameworks for AI, participation and systems thinking
- a two-day co-design workshop with policymakers, regulators and participation practitioners
- a co-design session with the People's Panel on AI
- a co-design workshop with the Public Voices in AI team supported by the Responsible AI UK (RAIUK) project team
- a co-design session with the Public Voices in AI oversight and advisory group supported by the Responsible AI UK (RAIUK) project team
- dialogue and discussions with numerous Responsible AI UK project researchers and leads, including the leads of UK-based researchers considering and developing their use of public voice in AI
- three coaching and mentoring sessions with School of Systems Change and systems change specialist Jacqueline Lim
- insights generated by prior research undertaken by the lead author. The report has drawn substantially from Building a Public Culture of Economics, Royal Society of Arts (2018), Participatory Data Stewardship (2021), Ada Lovelace Institute, and the Data Divide (2020), Ada Lovelace Institute. The work also draws from the author's unpublished contributions to the EU INSPIRE project as well as the author's unpublished contributions to the Participatory and Inclusive Data Stewardship project hosted by the Digital Good Network and the Ada Lovelace Institute.

This framework has also been informed by the insights developed through the broader Public Voices in AI programme.

Bibliography

Alexander, J. (2022) The Citizens: A new approach to solving society's biggest challenges. London: Penguin Books.

Altenbernd, E. and Young, A.T. (2013) 'Introduction: The significance of the frontier in an age of transnational history', Settler Colonial Studies, 4(2), pp. 127–150. Available at: https://www.tandfonline.com/doi/abs/10.1080/2201473X.2013.846385

American Nuclear Society (2017) 'Public engagement', Nuclear Grand Challenges. Available at: https://www.ans.org/challenges/engagement/

Arnstein, S.R. (1969) 'A Ladder of Citizen Participation', Journal of the American Institute of Planners, 35(4), pp. 216–224. Available at: https://lithgow-schmidt.dk/sherry-arnstein/ladder-of-citizen-participation.html

Asenbaum, H. (2021) 'Rethinking democratic innovations: A look through the kaleidoscope of democratic theory', Political Studies Review, 20(2), pp. 203–216. Available at: https://doi.org/10.1177/14789299211052890

Baig, M.A. and Khan, A.A. (2024) 'AI's Oppenheimer moment: Establishing regulations', The Interpreter, 16 January. Available at: https://www.lowyinstitute.org/the-interpreter/ai-s-oppenheimer-moment-establishing-regulations

Bentley, C., Muyoya, C., Vannini, S., Oman, S. and Jimenez, A. (2023) 'Intersectional approaches to data: The importance of an articulation mindset for intersectional data science', Big Data & Society, 10(2). Available at: https://journals.sagepub.com/doi/10.1177/20539517231203667

Berditchevskaia, A., Peach, K. and Stewart, I. (2020). Localising AI: How to make artificial intelligence work for local communities. Nesta. Available at: https://www.nesta.org.uk/report/localising-ai/

Bherer, L., Dufour, P. and Montambeault, F. (2016) 'The participatory democracy turn: an introduction', Journal of Civil Society, 12(3), pp. 225–230. Available at: https://www.tandfonline.com/doi/full/10.1080/17448689.2016.1216383

Bitzer, T., Wiener, M. and Cram, W.A. (2022) 'Algorithmic Transparency: Concepts, Antecedents, and Consequences – A Review and Research Framework', Communications of the Association for Information Systems, 52(1). Available at: https://aisel.aisnet.org/cais/vol52/iss1/16/

Bua, A. and Bussu, S. (2023) Reclaiming Participatory Governance: Social Movements and the Reinvention of Democratic Innovation. London: Routledge.

Bussu, S. and Geissel, B. (2024) 'INSPIRE – Intersectional Spaces of Participation: Inclusive, Resilient, Embedded', European Commission Community of Practice on Participatory and Deliberative Democracy. Available at: https://cop-demos.jrc.ec.europa.eu/citizen-engagement-projects/inspire-intersectional-spaces-participation-inclusive-resilient

Capraro, V., Lentsch, A., Acemoglu, D., Akgun, S., Akhmedova, A., Bilancini, E., Bonnefon, J.-F., Brañas-Garza, P., Butera, L., Douglas, K.M. and others (2024) 'The impact of generative artificial intelligence on socioeconomic inequalities and policy making', PNAS Nexus, 3(6), p. 191. Available at: https://doi.org/10.1093/pnasnexus/pgae191

Carmichael, M. and Stinson, J. (2024) 'The Ipsos AI Monitor 2024: Changing attitudes and feelings about AI and the future it will bring', Ipsos. Available at: https://www.ipsos.com/en/ipsos-ai-monitor-2024-changing-attitudes-and-feelings-about-ai-and-future-it-will-bring

Česnulaitytė, I. (2020) 'Innovative Citizen Participation and New Democratic Institutions: Catching the Deliberative Wave', OECD. Available at: https://www.oecd.org/en/publications/innovative-citizen-participation-and-new-democratic-institutions_339306da-en.html

Chilvers, J., Pallett, H. and Hargreaves, T. (2018) 'Ecologies of participation in socio-technical change: The case of energy system transitions', Environmental Science & Policy, 88, pp. 33-44. Available at: https://www.sciencedirect.com/science/article/pii/S2214629618303025

Collins, K. and Ison, R.L. (2009) 'Dare we jump off Arnstein's ladder? Social learning as a new policy paradigm', Proceedings of the PATH (Participatory Approaches in Science & Technology) Conference. Available at: https://oro.open.ac.uk/8589/1/Path_paper_Collins_Ison.pdf

Collins, P.H. and da Silva, E.C.G. (2021) 'Intersectionality as Critical Social Theory', Contemporary Political Theory, 20(3), pp. 690–725. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC8127482/

Conger, K., Fausset, R. and Kovaleski, S.F. (2019) 'San Francisco bans facial recognition technology', The New York Times, 14 May. Available at: https://www.nytimes.com/2019/05/14/us/facial-recognition-ban-san-francisco.html

Conway, R. (2020). Designing for an unpredictable future. Medium. Available at: https://rowan-conway.medium.com/designing-for-an-unpredictable-future-c73d01940c00

Coyle, D. and Weller, A. (2020) "Explaining" machine learning reveals policy challenges', Science, 368(6498), pp. 1433–1434. Available at: https://www.ijdesign.org/index.php/IJDesign/article/view/4120

Crawford, K., Dobbe, R., Dryer, T., Fried, G., Green, B., Kaziunas, E., Kak, A., Mathur, V., McElroy, E., Nill Sánchez, A., Raji, D., Rankin, J.L., Richardson, R., Schultz, J., West, S.M. and Whittaker, M. (2019) AI Now 2019 Report. AI Now Institute, New York University. Available at: https://ainowinstitute.org/wp-content/uploads/2023/04/AI_Now_2019_Report.pdf

Cruikshank, B. (1999). The Will to Empower: Democratic Citizens and Other Subjects. Ithaca, NY: Cornell University Press.

Dahlgren, G. and Whitehead, M. (1991) Policies and strategies to promote social equity in health. Stockholm: Institute for Futures Studies. Available at: https://www.pslhub.org/learn/improving-patient-safety/health-inequalities/the-dahlgren-whitehead-rainbow-1991-r5870/

Dalton, R. (2021) The Participation Gap. Oxford: Oxford University Press.

Davies, T. (2024) Global Deliberation on AI. Connected by Data. Available at: https://connectedbydata.org/resources/global-deliberation-ai

Davies, T. and Tennison, J. (2025) 'The Participatory Turn: Side Road or Highway?', TechPolicy.Press, 26 February. Available at: https://www.techpolicy.press/the-participatory-turn-side-road-or-highway/

De Silva, D. and Alahakoon, D. (2022) 'An artificial intelligence life cycle: From conception to production', Patterns, 3(6), p. 100489. Available at: https://www.sciencedirect.com/science/article/pii/S2666389922000745

Digital Humanities Institute (no date) 'Participatory Harm Auditing Workbenches and Methodologies', Digital Humanities Institute. Available at: https://www.dhi.ac.uk/projects/phawm

Dizikes, P. (2023) 'How an "AI-tocracy" emerges', MIT News, 13 July. Available at: https://news.mit.edu/2023/how-ai-tocracy-emerges-0713

Draper, N.A. and Turow, J. (2019) 'The corporate cultivation of digital resignation', New Media & Society, 21(8), pp. 1824–1839. Available at: https://journals.sagepub.com/doi/abs/10.1177/1461444819833331

Equality and Human Rights Commission (2025) 'Artificial intelligence in public services'. Available at: https://www.gov.uk/data-ethics-guidance/artificial-intelligence-in-public-services

Fraser, T. (2021) 'Towards a new holistic framework of systems change: Adapting Geels' transition theory', Refuge for Systems Leaders. Available at: https://medium.com/refuge-for-systems-leaders/towards-a-new-holistic-framework-of-systems-change-adapting-geels-transition-theory-8d589fb6de0a

Gallo, C. (2019) 'Why "move fast and break things" doesn't work anymore', Harvard Business Review, 3 December. Available at: https://hbr.org/2019/12/why-move-fast-and-break-things-doesnt-work-anymore

Garavan, M. (2010) 'Paulo Freire's Pedagogy of the Oppressed', in Dukelow, F. and O'Donovan, O. (eds) Mobilising Classics - Reading Radical Writing in Ireland. Manchester: Manchester University Press, pp. 123–139. Available at: https://www.researchgate.net/publication/260297860_Paulo_Freire%27s_Pedagogy_of_the_Oppressed

Gaventa, J, VeneKlasen, L. and Miller, V. (2002) Power cube: A framework for analyzing power in spaces and levels. Available at: https://www.powercube.net/other-forms-of-power/the-power-in-the-powercube/

Goodwill, M., van der Bijl-Brouwer, M. and Bendor, R. (2021) 'Beyond Good Intentions: Towards a Power Literacy Framework for Service Designers', International Journal of Design, 15(3). Available at: https://www.ijdesign.org/index.php/IJDesign/article/view/4120

Grant, S. (2023) 'How playing Wittgensteinian language-games can set us free', Aeon. Available at: https://aeon.co/ideas/how-playing-wittgensteinian-language-games-can-set-us-free

Hale, S., Jones, R., Kennedy, H., Middlemass, R., Millings, A., Neff, G., Ong, J.C., Patel, R., Richards, D., Snooks, K., Wajid, S. and Williams, R. (2024) 'What Do We Mean When We Talk About a Good Digital Society?', ESRC Digital Good Network. Available at: https://www.thebritishacademy.ac.uk/publications/what-do-we-mean-when-we-talk-about-a-good-digital-society/

Hao, K. (2022) 'Artificial intelligence is creating a new colonial world order', MIT Technology Review, 19 April. Available at: https://www.technologyreview.com/2022/04/19/1049592/artificial-intelligence-colonialism/

Hemphill, T.A. (2019) '"Techlash," responsible innovation, and the self-regulatory organization', Journal of Responsible Innovation, 6(2), pp. 240–247. Available at: https://www.tandfonline.com/doi/full/10.1080/23299460.2019.1602817

Holdo, M. and Öhrn Sagrelius, L. (2020) 'Why inequalities persist in public deliberation: Five mechanisms of marginalization', Political Studies, 68(3), pp. 634–652. Available at: https://doi.org/10.1177/0032321719868707

Hope, S. (2019) 'Participation as performance?', Sophie Hope. Available at: https://sophiehope.org.uk/blog/participation-as-performance/

Institute of Medicine (2014) Supporting a Movement for Health and Health Equity: Lessons from Social Movements: Workshop Summary. National Academies Press. Available at: https://www.ncbi.nlm.nih.gov/books/NBK268722/

International Association for Public Participation (IAP2), 'IAP2 Spectrum of Public Participation'. Available at: https://www.iap2.org/page/spectrum

InterpretML (2023) 'Glassbox Models', InterpretML Documentation. Available at: https://interpret.ml/docs/glassbox.html

Jacobs, K. (2022) 'Toronto wants to kill the smart city forever', MIT Technology Review, 29 June. Available at: https://www.technologyreview.com/2022/06/29/1054005/toronto-kill-the-smart-city/

Jansma, S.R., Gosselt, J.F., Kuipers, K. and de Jong, M.D.T. (2019) 'Technology legitimation in the public discourse: applying the pillars of legitimacy on GM food', Technology Analysis & Strategic Management, 32(2), pp. 1–13. Available at: https://www.tandfonline.com/doi/full/10.1080/09537325.2019.1648788

Just Associates (2007) A New Weave of Power, People & Politics: The Action Guide for Advocacy and Citizen Participation. Available at: https://justassociates.org/all-resources/a-new-weave-of-power-people-politics-the-action-guide-for-advocacy-and-citizen-participation/

Kelion, L. (2020) 'A-levels: Ofqual criticised over lack of algorithm details', BBC News, 4 August. Available at: https://www.bbc.com/news/technology-5365075

Könneker, C. (2024) 'How AI will change science communication', Stanford Social Innovation Review. Available at: https://ssir.org/articles/entry/science-communication-artificial-intelligence

Landi, M. (2025) 'UK newspapers launch campaign against AI copyright plans', The Independent, 25 February. Available at: https://www.independent.co.uk/news/uk/politics/government-daily-express-the-sun-daily-star-daily-mail-b2704106.html

Liu, X., Zeng, D. and Li, Y. (2023) 'Exploring the intersection of human-centered design and artificial intelligence: Opportunities and challenges', AI & Society. Available at: https://link.springer.com/article/10.1007/s00146-023-01635-y

Marchant, R. (2001) 'From the test tube to the table: Public perception of GM crops and how to overcome the public mistrust of biotechnology in food production', EMBO Reports, 2(5), pp. 354–357. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC1083893/

McCarthy, N. and Lane, G. People's AI Stewardship Summits. Royal Academy of Engineering. Available at: https://raeng.org.uk/people-s-ai-stewardship-summits

Means, A.J. (2021) 'Foucault, biopolitics, and the critique of state reason', Educational Philosophy and Theory, 53(12), pp. 1161–1172. Available at: https://doi.org/10.1080/00131857.2021.1871895

Mikalef, P., Krogstie, J., Pappas, I.O. and Sidorova, A. (2023) 'Infosphere, Datafication, and Decision-Making Processes in the AI Era', Journal of Technology and Innovation, 23(4), pp. 345–360. Available at: https://link.springer.com/article/10.1007/s11245-023-09919-0

Modhvadia, R., Sippy, T., Field Reid, O., Margetts, H. (2025). How do people feel about AI? Wave 2 of a nationally representative survey of UK attitudes to AI designed through a lens of equity and inclusion. Ada Lovelace Institute and The Alan Turing Institute. Available at: https://attitudestoai.uk/

Moore, M.-L., Riddell, D.J. and Vocisano, D. (2015) 'Scaling out, scaling up, scaling deep: Strategies of non-profits in advancing systemic social innovation', Journal of Corporate Citizenship, 58, pp. 67–84. Available at: https://www.researchgate.net/figure/Three-types-of-scaling-and-their-main-strategies_tbl1_298971574

Mouffe, C. (2023) 'Deliberative ecologies: A relational critique of deliberative systems', European Political Science Review. Available at: https://www.cambridge.org/core/journals/european-political-science-review/article/deliberative-ecologies-a-relational-critique-of-deliberative-systems/CDD5AA46BBF1222FCF60648E8C74331B

Muldoon, J. and Wu, B.A. (2023) 'Artificial Intelligence in the Colonial Matrix of Power', Philosophy & Technology, 36(4), pp. 1–24. Available at: https://link.springer.com/article/10.1007/s13347-023-00687-8

National Audit Office (2021) 'Test and trace in England – progress update', National Audit Office. Available at: https://www.nao.org.uk/reports/test-and-trace-in-england-progress-update/?nab=1

Patel, R. (2020) 'Blending old and new power: Democratic innovation in 2020', Involve. Available at: https://www.involve.org.uk/news-opinion/opinion/blending-old-and-new-power-democratic-innovation-2020

Patel, R. (2021) Participatory Data Stewardship. Ada Lovelace Institute. Available at: https://www.adalovelaceinstitute.org/wp-content/uploads/2021/11/ADA_Participatory-Data-Stewardship.pdf

Patel, R. and Jones, E. (2021) 'The Data Divide: Public Attitudes to Tackling Social and Health Inequalities in the COVID-19 Pandemic and Beyond', Ada Lovelace Institute. Available at: https://www.adalovelaceinstitute.org/report/the-data-divide/

Patel, R., Brisley, R. and Gisborne, J. (2023) 'The Global Science Partnership: A toolkit for inclusive policymaking on climate change', Ipsos. Available at: https://www.ipsos.com/en-uk/global-science-partnership-toolkit-inclusive-policymaking-climate-change

Patel, R., Field Reid, O. and Strait, A. (2021) 'How does structural racism impact on data and AI?', Ada Lovelace Institute. Available at: https://www.adalovelaceinstitute.org/blog/structural-racism-impact-data-ai/

Peppin, A. and Patel, R. (2020) Citizens' Biometrics Council. Ada Lovelace Institute. Available at: https://www.adalovelaceinstitute.org/project/citizens-biometrics-council/

Reincke, C.M., Bredenoord, A.L. and van Mil, M.H.W. (2020) 'From deficit to dialogue in science communication: The dialogue communication model requires additional roles from scientists', EMBO Reports, 21(9), p. e51278. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC7506985/

Rosenberg, M., Confessore, N. and Cadwalladr, C. (2018) 'Cambridge Analytica and Facebook: The scandal and the fallout so far', The New York Times, 4 April. Available at: https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html

Rotter, J. (1966) Generalized expectancies for internal versus external control of reinforcements. Psychological Monographs, 80, pp. 609

Sætra, H.S. (2020) 'A shallow defence of a technocracy of artificial intelligence: Examining the political harms of algorithmic governance in the domain of government', Technology in Society, 62, p. 101283. Available at: https://www.sciencedirect.com/science/article/pii/S0160791X19305925

Salwei, M.E. and Carayon, P. (2022) 'A Sociotechnical Systems Framework for the Application of Artificial Intelligence in Health Care Delivery', Journal of Cognitive Engineering and Decision Making, 16(4), pp. 194–206. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC9873227/

Saunders, T. (2017) 'How to involve the public in the development of artificial intelligence', Nesta. Available at: https://www.nesta.org.uk/blog/how-to-involve-the-public-in-the-development-of-artificial-intelligence/

Schulbaum, O. (2024) 'Inspiring Democratic Innovation: A Conversation with Sonia Bussu on Reclaiming Participatory Governance', Wilder Journal, 2. Available at: https://journal.platoniq.net/en/wilder-journal-2/interviews/reclaiming-participatory-governance/

Schwab, K. and Vanham, P. (2021) 'What is stakeholder capitalism? Its history and relevance', World Economic Forum. Available at: https://www.weforum.org/stories/2021/01/klaus-schwab-on-what-is-stakeholder-capitalism-history-relevance/

Sciencewise (2022) 'NHS AI Lab Participatory Fund for Patient-Driven AI Ethics Research', Sciencewise. Available at: https://sciencewise.org.uk/projects/nhs-ai-lab-participatory-fund-for-patient-driven-ai-ethics-research/

Scottish Government (2021) Citizens' Jury on COVID-19: Report of the Jury's Conclusions and Key Findings. Available at: https://www.gov.scot/publications/citizens-jury-qcovid-report-jurys-conclusions-key-findings/

Silverberg, D. (2023) 'Journalists are feeding the AI hype machine', BBC News. Available at: https://www.bbc.co.uk/news/business-68488924

Sloane, M., Moss, E., Awomolo, O. and Forlano, L. (2022) 'Participation is not a design fix for machine learning', Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '22), pp. 1–6. Available at: https://dl.acm.org/doi/10.1145/3551624.3555285

Smith, G. (2021) 'Citizens' assemblies are helping to fight the climate crisis', University of Westminster. Available at: https://www.westminster.ac.uk/research/impact/citizens-assemblies-are-helping-to-fight-the-climate-crisis

The Digital Good Network (2023) Public Voices in AI: Announcement of Funded Projects. Available at: https://digitalgood.net/public-voices-in-ai-announcement-of-funded-projects/

The Strategy Unit (2022) Coventry Waiting List Prioritisation Deliberative Report. Available at: https://www.strategyunitwm.nhs.uk/sites/default/files/2022-09/Coventry%20waiting%20list%20prioritisation%20delib%20-Full%20Report.pdf

Tiku, N. (2025) 'OpenAI's push to democratize artificial intelligence raises concerns about bias and control', Time. Available at: https://time.com/6684266/openai-democracy-artificial-intelligence/

University College London (2020) 'UCL welcomes government's reversal on A-level algorithm use', UCL News, 17 August. Available at: https://www.ucl.ac.uk/news/2020/aug/ucl-welcomes-governments-reversal-level-algorithm-use

Williamson, B. (2020) 'F**k the algorithm? What the world can learn from the UK's A-level grading fiasco', LSE Impact Blog, 26 August. Available at: https://blogs.lse.ac.uk/impactofsocialsciences/2020/08/26/fk-the-algorithm-what-the-world-can-learn-from-the-uks-a-level-grading-fiasco/

Wilson, R. and Mellier, C. (2024) 'Getting Real About Citizens' Assemblies: A New Theory of Change for Citizens' Assemblies', European Democracy Hub. Available at: https://europeandemocracyhub.epd.eu/getting-real-about-citizens-assemblies-a-new-theory-of-change-for-citizens-assemblies/

Zengler, T. (2019). Why "move fast and break things" doesn't work anymore. Harvard Business Review. Available at: https://hbr.org/2019/12/why-move-fast-and-break-things-doesnt-work-anymore

Zuboff, Shoshana. 2019. The Age of Surveillance Capitalism. London, England: Profile Books.

Acknowledgements

This report is dedicated to my former colleague at Involve, mentor and advisor, **Simon Burall**, whom we sadly lost last year.

Without his contributions, continued support and commitment to developing the field of participation in data and AI, this report would not have been possible.

Acknowledgements for contributions to the report:

Teresa Cisneros, Wellcome Trust

Octavia Field Reid, Ada Lovelace Institute

Lara Groves, Ada Lovelace Institute

Helen Kennedy, University of Sheffield

Jacqueline Lim, School of Systems Change

Cian O'Donovan, UCL

Susan Oman, University of Sheffield

Mariana Silva, University of Cambridge

Lucy Peters, freelance copyeditor

Sakab Bashir, report designer

Collaboration informing the report:

I am also especially grateful for the support, collaboration and generosity shown by my colleagues on the RAIUK Public Voices in AI programme, hosted by the Digital Good Network. Their time and generosity has significantly advanced the thinking in this document. The full team is named below

Helen Kennedy Professor of Digital Society, University of Sheffield. Director of the Digital Good Network.

Ros Williams, Senior Lecturer in Digital Media and Society, University of Sheffield. Associate Director of the Digital Good Network.

Susan Oman, Lecturer in Data, AI and Society at The University of Sheffield. Policy Fellow within the UK Government's Department for Digital, Culture, Media and Sport. Senior Researcher at the Digital Good Network.

Helen Margetts, Director, Public Policy Programme, Alan Turing Institute for Data Science and AI.

Octavia Field Reid, Associate Director (Impact & research practice), Ada Lovelace Institute.

Jack Stilgoe, Professor of Science and Technology Policy, Dept of Science & Technology Studies, UCL.

Ellie O'Keefe, Public Participation & Research Practice Lead, Ada Lovelace Institute

Roshni Modhvadia, Researcher, Ada Lovelace Institute.

Mhairi Aitken, Ethics Fellow, Alan Turing Institute.

Cian O'Donovan, Senior Research Fellow, Department of Science and Technology Studies, UCL.

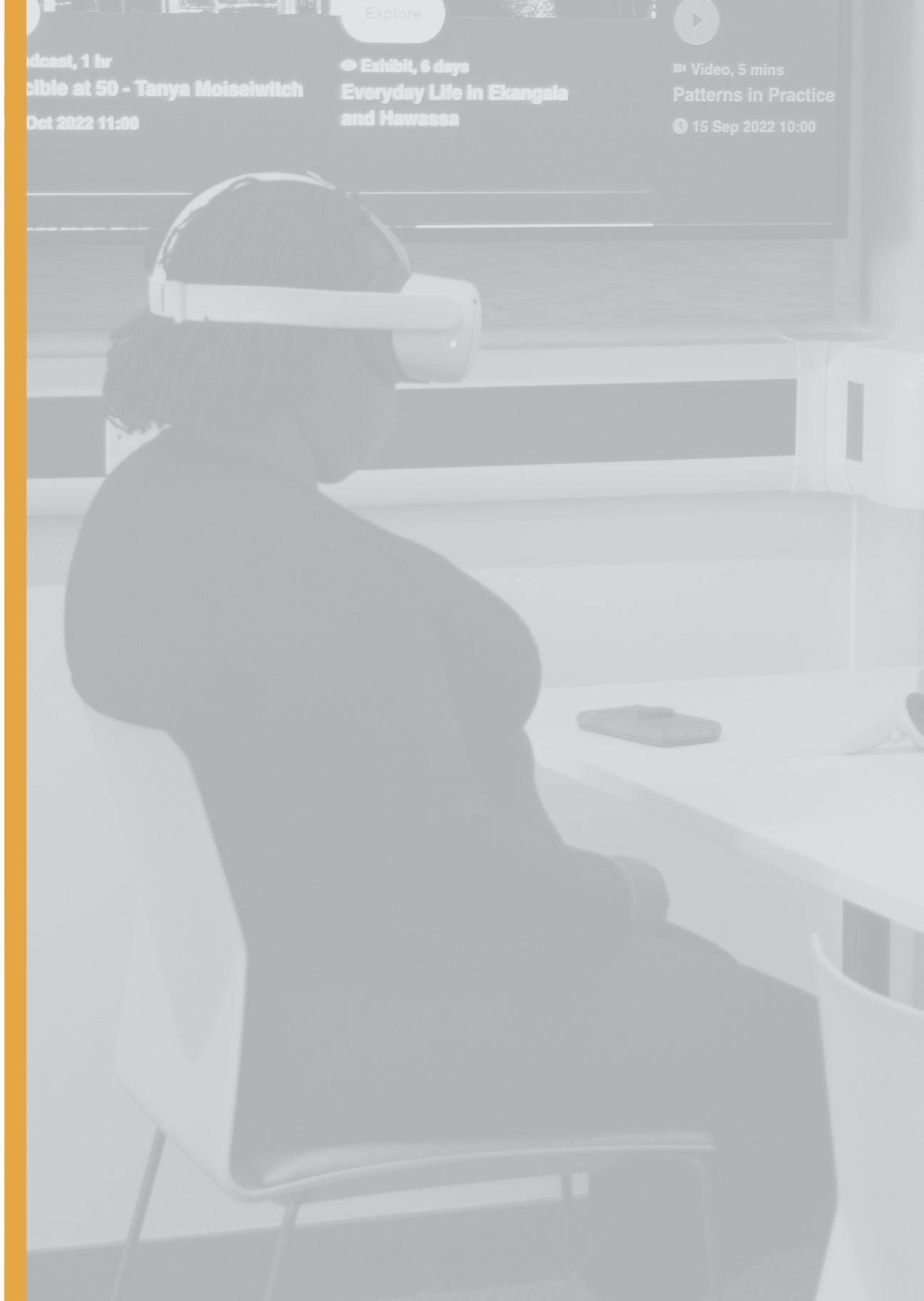
Tvesha Sippy, Researcher, Alan Turing Institute.

Sara Cannizzaro, Postdoctoral Researcher for the Public Voices in AI project.

Ruth Lauener (oversight of operations and logistics), University of Sheffield/Digital Good Network.

Sarah Givans, Research Support Project Administrator, University of Sheffield/Digital Good Network.

We are also grateful to the funders, Responsible AI UK, BRAID and the Economic and Social Research Council (ESRC) for their support



This framework and workbook, developed by Elgon Social Research, is a practical tool designed to empower organisations to shape AI systems that align with society's diverse values. Led by Reema Patel in collaboration with the Digital Good Network, Elgon has combined theoretical insights with real-world experience to create a self-assessment workbook for integrating public voices into AI decision-making. This workbook offers a structured approach to reflection, improvement, and action, ensuring AI systems truly serve the public good.

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

About Elgon Social:

Are you tired of top-down solutions? So are we.

We are forging a new path, where strategies, policies, research and innovations emerge from co-creation between organisations and society. We're doers and thinkers, developing inclusive research, design, and policy that puts people, planet, and purpose before profit. We champion and pioneer participatory and inclusive practices. We are passionate about amplifying underrepresented voices.

Partner with us, and let's build a just and viable world together.

www.elgon.social

reema@elgon.social