



**Selected Papers of #AoIR2025:
The 26th Annual Conference of the
Association of Internet Researchers**
Niterói, Brazil / 15 – 18 Oct 2025

FROM PRINCIPLES TO PRACTICE: PARTICIPATORY INFRASTRUCTURES FOR INCLUSIVE AI

Dominique Carlon
Swinburne University of Technology

Anthony McCosker
Swinburne University of Technology

Introduction

Attacks on diversity, equity and inclusion initiatives in the United States mark a rupture in the trajectory of responsible artificial intelligence (AI) with implications across the global ecosystem (Mishra, 2025). These developments expose the limitations of prevailing approaches, predominately articulated through abstract principles (i.e. fairness, explainability, privacy) and operationalised through technical fixes embedded within models. While such initiatives have gained traction in policy and industry, they remain constrained by geopolitical and commercial imperatives, and insufficient attention to the social, cultural, and institutional contexts in which AI systems are implemented.

This paper introduces a practice-based framework for inclusive AI, developed through collaborations with community, humanitarian, health, legal, and learning organisations as part of a broader project investigating whether inclusive AI can be embedded in practice. We conceptualise inclusive AI not as a technological output but as a participatory process, situating inclusion at the centre of capability building, equitable benefit, and collective participation. Central to this reframing is the role of participatory processes and intermediaries—the individuals, organisations, and public or online spaces—who act as capability converters, translating technical features into community-relevant outcomes. We propose that achieving inclusive AI requires more than compliance with principles or technical fixes: it depends on intermediaries that can translate, adapt, and carefully situate AI technologies within community contexts.

Limitations of Responsible AI

Suggested Citation (APA): Carlon, D. & McCosker, A. (2025, October). *From principles to practice: Participatory infrastructures for Inclusive AI*. Paper presented at AoIR2025: The 26th Annual Conference of the Association of Internet Researchers. Niterói, Brazil: AoIR. Retrieved from <http://spir.aoir.org>.

Responsible AI has been shaped by two dominant paradigms: principlism and operationalism. Principlism grounds governance in normative values such as privacy, safety, and fairness, formalised through ethical guidelines and regulatory standards (Graves & Ratti, 2025; Ratti & Graves, 2025). Principles provide necessary orientation, but their presence alone does not guarantee adherence and they are often insufficiently actionable. As Ratti and Graves (2025, p.9) note, “principles are usually vague and not well specified, so that a technology company may understand respect for autonomy or explainability in one way, while external auditors in another.”

Operationalism attempts to embed responsible AI principles into systems through safeguards, metrics, and technical interventions (Zowghi & Bano, 2024; Shams et al., 2025). While this represents a shift toward practical application, it remains narrowly focused on the internal mechanics of technical systems, with limited attention to implementation factors. As Ibitoye, Nkwo, and Orj (2025, p.6218) argue, these initiatives often fracture under the messiness of “real-world deployment, sociopolitical contestation, and institutional complexity”. Moreover, the geopolitical dominance of a handful of AI companies means research scrutiny does not necessarily translate into impactful interventions.

Responsible AI principles and technical fixes are necessary but insufficient to ensure equitable outcomes. Both assumes universality and overlooks the situated contexts in which AI is deployed (Sartori & Theodorou 2022). Most critically, they fail to consider how diverse communities can convert technical systems into genuine capability gains for inclusive opportunities and benefits.

Evaluative framework for inclusive AI

As AI technologies are rapidly embedded into social and essential services, some community members find themselves excluded or disadvantaged, not only because of skills or resources, but because AI tools are often designed without attention to their needs or lived realities (Thomas et al., 2025). Our evaluative framework extends responsible AI by asking not only whether systems comply with responsible AI principles, but whether they enable communities to achieve tangible benefits within their contexts and allow for equitable participation from design, deployment, to governance.

Drawing on the human development capabilities approach (Sen, 1993; Nussbaum, 2009; UNDP, 2025), we propose that inclusive AI should be anchored in practices that ensure systems are widely, rather than narrowly, beneficial. In addition to regulation and technical design, this involves a ground-up approach to expanding participation into AI integration. Primarily, this is achieved by empowering diverse groups to critically engage with, contribute to, and understand the real-world application of AI (World Economic Forum, 2022). Accordingly, we define inclusive AI as securing equitable benefits across four dimensions in the AI process:

1. **Safety and accountability** – ensuring applications are reliable, explainable to diverse stakeholders, and free from unreasonable bias, error and harmful impact.

2. **Community and stakeholder participation** – assessing the depth and impact of community involvement in co-design, implementation, and governance.
3. **Community benefit** – evaluating empowerment, capability building, and agency derived from the AI system.
4. **Sustainability and maintenance** – assessing adaptability and feasibility of maintenance and governance.

This framework is premised on the sociotechnical proposition that norms and values embedded in AI systems confer power not only during design but also in deployment, user engagement, and maintenance (Noorman & Swierstra, 2023; Weidinger et al., 2023). Central to this evaluative lens is whether AI systems expand people’s real opportunities (their capabilities) and their agency to achieve what they value (Graves & Ratti, 2025). Accordingly, it acknowledges disengagement or refusal as valid responses (Zong & Matais, 2024), recognising that inclusivity must encompass agency in adoption, resistance, and learning.

Capability pathway to Inclusive AI in practice

Realising inclusive AI requires participatory infrastructures and processes that enable communities to navigate, adapt, and embed AI in line with their priorities. These infrastructures are facilitated by intermediaries who act as capability converters, bridging technical affordances with community needs. Capabilities encompass both technical components—such as datasets and models—and the knowledge, skills, and literacies of people affected by AI (Gupta & George, 2016). Technical developments however are advancing faster than our understanding of the human gains and conversion factors required to deliver benefit (Robeyns, 2006).

In our work, intermediaries have included community organisations, digital mentors, learning facilitators, local digital champions, and public spaces such as libraries and online forums. They function as bridges between technical and community capabilities, addressing the implementation factors (conditions needed for systems to operate effectively in specific settings) and conversion factors (personal, social, and environmental elements) that allow AI systems to generate real benefit.

Advancing inclusive AI depends on processes of translation, adaptation, and care via the social, cultural, and institutional arrangements through which communities access AI. *Translation* involves making technical concepts intelligible for diverse stakeholders and community members. *Adaptation* ensures technologies are reshaped to align with situated needs, whether tailoring outputs to local languages, embedding tools in analogue workflows, or contextualising knowledge for cultural or organisational settings. *Care* underscores the relational work needed to sustain inclusive participation and trust. Our work has found that intermediaries operate best when ‘double embedded,’ being both technically informed and culturally situated to enable agency and inclusive capability gains from the ground up. This ensures AI systems are not only technically sound but also socially meaningful, supporting communities in achieving what matters to them.

Conclusion

Inclusive AI will not emerge from principles or technical fixes alone. It will depend also on whether communities of users have the capabilities to shape, resist, and sustain AI systems on their own terms. Our evaluative framework extends responsible AI by asking whether systems enable communities to achieve tangible benefits and equitable participation across the process of AI. In doing so, it highlights the critical role of participatory processes and intermediaries in ensuring AI systems are not only technically informed but also socially and culturally grounded. The framework offers researchers and practitioners a conceptual and evaluative approach to examine inclusivity throughout the AI lifecycle and informs our ongoing research identifying what support intermediaries require to bridge technical change with community needs.

References

- Graves, M., & Ratti, E. (2025). A Capability Approach to Ethical Development and Internal Auditing of AI Technology. *Journal of Responsible Technology*, 100121, 1-7.
- Gupta, M., & George, J. F. (2016). Toward the development of a big data analytics capability. *Information & Management*, 53(8), 1049-1064.
- Ibitoye, A. O., Nkwo, M. S., & Orji, R. (2025). Rethinking responsible AI from ethical pillars to sociotechnical practice. *AI and Ethics*, 5(6), 6207-6223.
- Mishra, N. C. (2025) AI Under Donald Trump: What the New Executive Order Tells Us About America's Policy Approach, *Australian Institute of International Affairs*, 11 February.
<https://www.internationalaffairs.org.au/australianoutlook/ai-under-donald-trump-what-the-new-executive-order-tells-us-about-americas-policy-approach/>
- Noorman, M., & Swierstra, T. (2023). Democratizing AI from a sociotechnical perspective. *Minds and Machines*, 33(4), 563-586.
- OECD (2020). Addressing societal challenges using transdisciplinary research. In *OECD science, technology and industry policy papers 88*. OECD Publishing.
- Ratti, E., & Graves, M. (2025). A capability approach to AI Ethics. *American Philosophical Quarterly*, 62(1), 1-16.
- Sartori, L., & Theodorou, A. (2022). A sociotechnical perspective for the future of AI: narratives, inequalities, and human control. *Ethics and Information Technology*, 24(1), 4.
- Shams, R. A., Zowghi, D., & Bano, M. (2023). AI and the quest for diversity and inclusion: A systematic literature review. *AI and Ethics*, 1-28.

Thomas, J., McCosker, A., Parkinson, S., Hegarty, K., Featherstone, D., Kennedy, J., Ormond-Parker, L., Morrison, K., Rea, H., & Ganley, L. 'Measuring Australia's Digital Divide: 2025 Australian Digital Inclusion Index'. Melbourne: ARC Centre of Excellence for Automated Decision-Making and Society, RMIT University, Swinburne University of Technology, and Telstra.

Weidinger, L., Rauh, M., Marchal, N., Manzini, A., Hendricks, L. A., Mateos-Garcia, J., Mateos-Garcia, Bergman, S., Kay, J., Griffin, C., Bariach, B., Gabriel, I., Rieser, V & Isaac, W. (2023). Sociotechnical safety evaluation of generative ai systems. *arXiv preprint arXiv:2310.11986*.

World Economic Forum (2022). 'A Blueprint for Equity and Inclusion in AI'. <https://www.weforum.org/publications/a-blueprint-for-equity-and-inclusion-in-artificial-intelligence/>

Zong, J., & Matias, J. N. (2024). Data refusal from below: A framework for understanding, evaluating, and envisioning refusal as design. *ACM Journal on Responsible Computing*, 1(1), 1-23.

Zowghi, D., & Bano, M. (2024). AI for all: Diversity and Inclusion in AI. *AI and Ethics*, 1-4.