The Equity Test:
A New Standard for AI in Public Services

Amanda Dahl
amandadahl.com

There is no shortage of frameworks for responsible AI. A 2020 review identified over 80 sets of ethical principles for artificial intelligence, produced by governments, companies, international organisations, and academic bodies. The principles are remarkably consistent: fairness, transparency, accountability, human oversight. The consensus is real. The implementation is not.

Across three decades building and governing technology inside public institutions, at the UK Government Digital Service, the White House US Digital Service, Sony PlayStation, the BBC, and now The Crown Estate, I have watched this gap between principle and practice grow wider. As Hudson et al. (2019) observe, a gap often exists between policy intent and implementation in public services, and AI governance is no exception. I have come to believe the problem is not a lack of principles. It is the wrong unit of analysis.

Most responsible AI frameworks ask: is this system fair? Is it transparent? Is there human oversight? These are important questions. But they are process questions. They evaluate the system. They do not evaluate the outcome.

I want to propose a different starting point. A single question, applied to any AI system deployed in a public service: does this system expand or narrow equity in access, treatment, and outcome for the people it affects?

I call this the equity test.

Three Questions

The equity test asks three things.

First: does this system improve access to the service for the people who need it most? Not access in general, but access for those who currently face the highest barriers. People with complex needs, limited (digital) literacy, language barriers, disabilities, caring responsibilities. If a system improves access for those already well-served while creating new barriers for those who were not, it has failed the equity test regardless of its aggregate performance.

Second: does it treat people with equivalent needs equivalently, regardless of background? Rather than identical treatment, the aim would be *equivalent* treatment. A triage system that routes complex claims to a slower manual track while accelerating straightforward ones may appear neutral. But if complexity correlates with disability, caring responsibilities, or socioeconomic disadvantage – as it almost always does – the system has introduced a new inequity while performing efficiency.

This is what Baert (1991) calls ‘counter-finality’, actions designed to achieve specific goals producing outcomes contrary to those goals. In my work on gender equity in government AI, I found that this dynamic is pervasive. Women are often underrepresented in available training data sets (Browne et al., 2024), and AI systems trained on historical data in which women’s employment patterns reflect interrupted careers due to caregiving can systematically disadvantage women when assessing eligibility for benefits or employment

services. When technology aimed at benefits eligibility is introduced, ‘women’s work becomes degraded and polarised from men’s work’ (Frennert, 2021, p. 12). The algorithm encodes the very inequality it was ostensibly deployed to overcome. The intention is equitable. The mechanism produces the opposite.

Third: does it produce more equitable outcomes than the process it replaces? If you cannot demonstrate this with evidence – not projections, evidence - then you do not have a governance case for deployment. You have a procurement decision dressed as a service improvement.

What AI Systems Disclose About Who Matters

Suchman (2006) argues that AI functions as a ‘disclosing agent for assumptions about the human’ (p. 226). Every AI system, through its design choices from what data it was trained on, to which outcomes it optimises for, and whose experience it models, reveals what its creators assumed about the people it would affect. As Kranzberg (1986) affirms, technology is not neutral: the choices embedded in design carry values whether or not they are acknowledged. In my research, I have developed Suchman’s concept into what I call the Disclosure-Scale-Trauma framework (Dahl, 2024a): the proposition that AI systems simultaneously disclose embedded assumptions about human nature, intervene across multiple scales from the individual to the institutional to the societal (following Agar, 2020), and modulate trauma responses, either amplifying or alleviating harm.

The equity test operationalises the disclosure dimension of this framework. When a government department deploys an AI system to triage benefit claims, the system’s design reveals what the department assumes about its claimants. If the system optimises for speed and cost reduction, it discloses an assumption that efficiency is the primary value. If it routes complexity to slower manual processes, it discloses an assumption that complex cases are exceptions rather than the core constituency. If it lacks mechanisms for challenge or appeal, it discloses an assumption that the people affected by its decisions do not need the power to contest them. Many AI systems use opaque algorithms to make decisions without transparency to humans (Grimmelikhuijsen, 2023), compounding the problem.

These disclosures operate across scales. At the individual level, a claimant experiences delay, confusion, or denial. At the institutional level, the system shapes how the department allocates resources and defines success. At the societal level, patterns of algorithmic decision-making compound into what Fricker (2009) calls epistemic injustice, the systematic marginalisation of certain people’s knowledge and experience. The equity test asks us to read these disclosures honestly, and to ask whether what the system reveals about our assumptions is something we are willing to defend.

The Problem of Thick Rules

Part of why the equity test matters is that the most consequential AI deployments in public services are replacing human decisions that were never simple rule-following. In my work on artificial moral agents in the civil service (Dahl, 2025a), I drew on Lorraine Daston’s distinction between ‘thick’ and ‘thin’ rules. Thin rules are straightforward, general, and easily applicable across many situations. Thick rules contain a high level of contextual understanding, in that they are those which foresee ‘the variability under which rules will be applied’ (Daston, as cited in Taylor, 2022).

Much of what frontline civil servants do operates in the realm of thick rules. A benefits assessor determining eligibility for a welfare assistance scheme is not simply applying a formula. They are exercising what Lipsky (1980) calls ‘street-level bureaucracy’:

discretionary judgement that accounts for context, complexity, and the particular circumstances of the person in front of them. In one council's Welfare Assistance Scheme, applicants must fit into certain 'categories of vulnerability' to receive a crisis grant, yet no published criteria are provided for how such vulnerability is assessed, allowing frontline workers to exercise discretion in evaluating each applicant's level of need (Perry et al., 2014). A good assessor recognises when someone's situation does not fit neatly into a category. They use practical wisdom, Aristotle's *phronesis*, to ensure administrative fairness.

AI systems, as currently deployed, operate overwhelmingly in the domain of thin rules. They classify, sort, and route based on pattern recognition. When these systems replace human discretion in thick-rule contexts such as welfare, healthcare, housing, and criminal justice, they flatten the complexity that fair decision-making requires. As Daston puts it, 'human beings can improvise. This is a very difficult challenge for programmes which expect the world to be steady as she goes, stable and predictable' (as cited in Mackereth & Drage, 2022). The equity test surfaces this problem. When you ask 'does this system treat people with equivalent needs equivalently?' you are asking whether the system can handle the thick-rule situations that most affect vulnerable people. In most cases, the honest answer is: not yet.

The Trauma Dimension

There is another reason the equity test requires more than a compliance framework. AI systems in welfare and social care contexts interact with populations that have experienced trauma: the trauma of poverty, of institutional hostility, of administrative burden, of being treated as a case number rather than a person. Trauma, understood as a deeply impactful experience that disrupts the sense of self, safety, and well-being (Herman, 1992), manifests at every scale, from individual psychological distress to collective harm resulting from systemic policy failures (Gray et al., 2004).

In my research on trauma-informed approaches to AI (Dahl, 2024a), I argued that designing algorithmic systems for care contexts without understanding trauma is designing systems that will cause harm. Chen et al. (2022) advocate for 'trauma-informed computing', safer technology experiences that account for the reality that many users carry histories of harm. The clinical literature on trauma-informed care identifies five core principles: safety, trustworthiness, choice, collaboration, and empowerment. These principles were developed for human practitioners. But they apply with equal force to the design and governance of algorithmic systems.

Consider the DWP's use of algorithmic messaging to prioritise welfare claimants for intervention. The system operates in a context where many claimants have experienced institutional hostility, administrative coercion, and the trauma of benefits sanctions. Barford and Gray (2022) document how the social safety net has frayed through technological and bureaucratic processes that treat citizens as cases to be managed rather than people to be served. A system designed without understanding this context may trigger threat responses, reduce engagement, and worsen outcomes – even if the system's stated intent is supportive. A trauma-informed equity test would ask: does this system provide safety and predictability? Does the claimant understand what is happening and why? Do they have genuine choice in how they engage? Is the system designed to empower, or to control?

These concerns are at the heart of whether AI in public services expands or narrows equity. A system that is technically accurate but retraumatising is not equitable. A system that improves efficiency while amplifying the administrative burden on the most vulnerable is not equitable. The equity test, informed by trauma-aware governance, catches these failures. A compliance checklist does not.

Framing, Power, and the Pro-Innovation Bias

The equity test also challenges a deeper structural problem: the way AI governance in public services is framed. In my analysis of big tech influence on UK AI regulation (Dahl, 2024b), I examined how corporate lobbying has produced what Ferreira et al. (2020) call a 'pro-innovation bias' - a framing in which regulation is positioned as inherently antagonistic to innovation, and governance is treated as friction to be minimised. This reflects what Pfotenhauer et al. (2018) term a 'deficit model of innovation', where systemic problems are blamed on over-regulation and risk-averse policymaking rather than on inadequate governance.

This framing has direct consequences for equity. When the default policy assumption is that the purpose of AI governance is to not impede innovation, equity becomes at best a secondary consideration, that is mentioned in white papers, absent from procurement specifications. The UK government's own policy paper on AI regulation (DSIT, 2023) exemplifies this: ethical and human rights risks are acknowledged but treated as manageable externalities rather than the central governance challenge. As Ochigame (2019) argues, much of the corporate-funded discussion surrounding AI ethics is strategically aligned with efforts to avoid legally enforceable restrictions on controversial technologies.
In my work on AI regulation and epistemic infrastructure (Dahl, 2025b), I argued that the real choice facing policymakers is not between innovation and regulation, but between innovation that serves extractive interests and innovation that democratises access and expands equity. Bradford (2024) describes this as a 'false choice' between regulation and innovation. The PSD2 directive in financial services provides evidence that well-designed regulatory intervention can stimulate rather than stifle genuine innovation, it transformed oligopolistic banking into a competitive ecosystem, enabling challenger entrants to emerge through mandated openness (Blind, 2012). As Stiglitz (2024) observes, a free market combined with democracy does not constitute a stable equilibrium without strong guardrails.

The equity test embodies this reframing. It does not ask 'does this governance slow AI deployment?' It asks 'does this deployment serve the people it is supposed to serve?' That shift in the governing question is itself a form of power redistribution — moving the frame from the interests of deployers to the experience of the deployed-upon.

Gender as a Case in Point

Gender equity in government AI ecosystems is a particularly revealing test case. In my research on pluralist approaches to human-centred government AI (Dahl, 2024c), I found that the absence of diverse representation in both training data and development teams produces systems that systematically disadvantage women — especially women from marginalised communities. As Connolly (2005) argues, the reduction of inequality requires the mobilisation of diverse perspectives, yet the number of women working in STEM research remains significantly smaller than men across most countries globally (Valls & Gibert, 2022). Without incorporating diverse expertise, as Forester (2018) warns, deliberative results risk being irrelevant or solving the wrong problems.
The consequence is counter-finality at scale: a system designed to streamline benefits access that disadvantages the people most likely to need benefits. This is not hypothetical. An AI system in Austria used labour market data to score jobseekers' employability, systematically giving lower scores to women with caring responsibilities and people with disabilities (Human Rights Watch, 2021), which are the populations the employment service was designed to support.

A pluralist approach to AI governance, one that mandates diverse representation in data, in development teams, and in governance structures, is a prerequisite for any system that

claims to be equitable. The equity test makes this visible: if you cannot demonstrate that a system treats women with caring responsibilities equivalently to other claimants, the system fails. No amount of aggregate accuracy compensates for distributional injustice.

What the Equity Test Is Not

The equity test is not a replacement for responsible AI frameworks. Fairness, transparency, and accountability remain essential. But they are means, not ends. The end is equity: that AI systems in public services expand access, fairness, and dignity for the people they are supposed to serve.

The equity test is also not a reason to avoid deploying AI. Technology can and does improve public services. I have seen this first-hand, at GDS and at USDS and at The Crown Estate. The question is not whether to deploy, but whether we have the governance to ensure that improvement is equitably distributed.

And the equity test is not a single assessment. It is a governing discipline. It should be applied before deployment, during operation, and in ongoing evaluation. It requires distributional analysis with accuracy disaggregated by the populations most affected. It requires procurement contracts with equity metrics and audit provisions. And it requires governance structures with meaningful representation from the people affected by algorithmic decisions — not consultation exercises, but standing mechanisms for challenge and accountability. As Ansell and Gash (2008) note, small wins may not be an appropriate strategy for trust-building where stakeholders have more ambitious goals that cannot easily be parsed into intermediate outcomes.

What Would Change

If the equity test were adopted as a governance standard, several things would shift. Algorithmic impact assessments would be required before deployment, not after. They would include distributional analysis across gender, disability, ethnicity, and socioeconomic status. Procurement contracts would include equity reporting and audit rights. Frontline staff would be trained on system limitations, not just capabilities. And critically, governance structures would include the voices of the people affected, like the welfare claimants, the patients, the jobseekers, the tenants. They would not just be consulted as stakeholders, but as participants with power.

As I argued in my work on human-centred AI policy interventions (Dahl, 2024c), one concrete step would be to integrate equity evaluation into the existing Service Assessment process that all UK government digital services must complete as part of the Service Standard (Central Digital and Data Office, 2024). This would mean no AI system could pass through spend control and go live without demonstrating that it had been assessed for distributional impact and that governance mechanisms were in place. The ‘indicators of perceived unfairness, inconsistency, or sloppy administration’ that Hood and Dixon (2015) identified as plaguing UK public services for decades would finally have a structured mechanism for prevention rather than post-hoc complaint.

None of this is technically difficult. All of it is institutionally hard. That is the nature of the implementation deficit, it’s the gap between our governance commitments and our institutional capacity to deliver them. Closing that deficit is the work.

The Governing Question

The equity test will not catch everything. But it changes the governing question from ‘is this system compliant?’ to ‘is this system just?’ From ‘does this system work?’ to ‘who does this system work for?’

In Kazuo Ishiguro’s novel Klara and the Sun, an AI companion reflects that despite her efforts to understand her human charge, ‘something would have remained beyond my reach’ (Ishiguro, 2021, p. 338). What remains beyond reach for AI in public services is not technical capability. It is the contextual understanding, the practical wisdom, the empathic discretion that equitable governance demands. The equity test does not solve this. But it names it, measures it, and insists that we govern for it.
That shift is overdue.

Amanda Dahl is Director of Product & Strategy at The Crown Estate and a Policy Fellow at the University of Cambridge Centre for Science and Policy. She is completing an MSt in AI Ethics and Society at Cambridge, where her research examines transparency, accountability, and equity in AI-enabled public services.

References

Agar, J. (2020). What is technology? Annals of Science, 77(3), 377–382. https://doi.org/10.1080/00033790.2019.1672788

Ansell, C., & Gash, A. (2008). Collaborative governance in theory and practice. Journal of Public Administration Research and Theory, 18(4), 543–571.

Baert, P. (1991). Unintended consequences: A typology and examples. International Sociology, 6(2), 201–210.

Barford, A., & Gray, M. (2022). The tattered state: Falling through the social safety net. Geoforum, 137, 115–125.

Blind, K. (2012). The influence of regulations on innovation: A quantitative assessment for OECD countries. Research Policy, 41(2), 391–400. https://doi.org/10.1016/j.respol.2011.08.008

Bradford, A. (2024). The false choice between digital regulation and innovation. Northwestern University Law Review, 19, 377–389.

Browne, J., Drage, E., & McInerney, K. (2024). Tech workers’ perspectives on ethical issues in AI development: Foregrounding feminist approaches. Big Data & Society, 11, Article 1.

Central Digital and Data Office, UK Cabinet Office. (2024). Service Standard. https://www.gov.uk/service-manual/service-standard

Chen, J. X., McDonald, A., Zou, Y., Tseng, E., Roundy, K. A., Tamersoy, A., Schaub, F., Ristenpart, T., & Dell, N. (2022). Trauma-informed computing: Towards safer technology experiences for all. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (pp. 1–20). https://doi.org/10.1145/3491102.3517475

Connolly, W. (2005). Pluralism. Duke University Press.

Dahl, A. (2024a). The intersection of disclosure and scale: A trauma-informed approach to AI. Leverhulme Centre for the Future of Intelligence, University of Cambridge.

Dahl, A. (2024b). The power of framing: Big tech influence on AI regulation and policy discourse. International School for Government, King's College London.

Dahl, A. (2024c). A pluralist approach to human-centred government AI with a focus on gender-based equity. International School for Government, King's College London.

Dahl, A. (2025a). Public trust in digital governance: The role of artificial moral agents in delivering ethical public services. Leverhulme Centre for the Future of Intelligence, University of Cambridge.

Dahl, A. (2025b). AI challengers: How AI regulation can stimulate innovation, democratise knowledge and break up digital monopolies. Leverhulme Centre for the Future of Intelligence, University of Cambridge.

Department for Science, Innovation, & Technology. (2023). A pro-innovation approach to AI regulation.
https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper

Ferreira, A., von Schönfeld, K. C., Tan, W., & Papa, E. (2020). Maladaptive planning and the pro-innovation bias: Considering the case of automated vehicles. Urban Science, 4(3), Article 41.

Forester, J. (2018). Deliberative planning practices — without smothering invention: A practical aesthetic view. In A. Bächtiger, J. S. Dryzek, J. Mansbridge, & M. Warren (Eds.), The Oxford handbook of deliberative democracy. Oxford University Press.

Frennert, S. (2021). Gender blindness: On health and welfare technology, AI and gender equality in community care. Nursing Inquiry, 28(4), Article e12433.

Fricker, M. (2009). Epistemic injustice: Power and the ethics of knowing. Oxford University Press.

Gray, M. J., Maguen, S., & Litz, B. T. (2004). Acute psychological impact of disaster and large-scale trauma: Limitations of traditional interventions and future practice recommendations. Prehospital and Disaster Medicine, 19(1), 64–72.

Grimmelikhuijsen, S. (2023). Explaining why the computer says no: Algorithmic transparency affects the perceived trustworthiness of automated decision-making. Public Administration Review, 83(2), 241–262.

Herman, J. L. (1992). Complex PTSD: A syndrome in survivors of prolonged and repeated trauma. Journal of Traumatic Stress, 5(3), 377–391. https://doi.org/10.1002/jts.2490050305

Hood, C., & Dixon, R. (2015). A government that worked better and cost less? Evaluating three decades of reform and change in UK central government. Oxford University Press.

Hudson, B., Hunter, D., & Peckham, S. (2019). Policy failure and the policy-implementation gap: Can policy support programs help? Policy Design and Practice, 2(1), 1–14. https://doi.org/10.1080/25741292.2018.1540378

Human Rights Watch. (2021). E.U. artificial intelligence regulation threatens social safety net.
https://www.proquest.com/wire-feeds/human-rights-watch-e-u-artificial-intelligence/docview/2595941831/se-2

Ishiguro, K. (2021). Klara and the Sun. Faber.

Kinsella, E. A., & Pitman, A. (2012). Phronesis as professional knowledge: Practical wisdom in the professions. Brill.

Kranzberg, M. (1986). Technology and history: “Kranzberg’s Laws.” Technology and Culture, 27(3), 544–560. https://doi.org/10.2307/3105385

Lipsky, M. (1980). Street-level bureaucracy: Dilemmas of the individual in public service. Russell Sage Foundation.

Mackereth, K., & Drage, E. (Hosts). (2022, September). The exorcism of emotion in rational science (and AI) with Lorraine Daston [Audio podcast episode]. The Good Robot Podcast. https://www.thegoodrobot.co.uk/

Ochigame, R. (2019). The invention of ‘ethical AI’: How big tech manipulates academia to avoid regulation. In Economies of virtue: The circulation of ‘ethics’ in AI (pp. 49–50).

Perry, J., Williams, M., Sefton, T., & Haddad, M. (2014). Emergency use only: Understanding and reducing the use of food banks in the UK. Child Poverty Action Group / Church of England / Oxfam / The Trussell Trust.

Pfotenhauer, S., Juhl, J., & Aarden, E. (2018). Challenging the ‘deficit model’ of innovation: Framing policy issues under the innovation imperative. Research Policy, 48(4), 895–904.

Stiglitz, J. (2024). The road to freedom: Economics and the good society. Penguin.

Suchman, L. (2006). Human-machine reconfigurations. Cambridge University Press.

Taylor, L. (Host). (2022, September 28). Rules and order [Audio podcast episode]. Thinking Allowed. BBC Radio 4.

Valls, A., & Gibert, K. (2022). Women in artificial intelligence (AI). Multidisciplinary Digital Publishing Institute.