



Delft University of Technology

Document Version

Final published version

Citation (APA)

Gagua, A. (2025). Improvising Accountability: The Everyday Governance Work of Responsible AI in the Public Sector. In *Proceedings of the 2025 AAAI/ACM Conference on AI, Ethics, and Society (AIES 2025)*
<https://doi.org/10.1609/aies.v8i3.36775>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

In case the licence states "Dutch Copyright Act (Article 25fa)", this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

**Green Open Access added to [TU Delft Institutional Repository](#)
as part of the Taverne amendment.**

More information about this copyright law amendment
can be found at <https://www.openaccess.nl>.

Otherwise as indicated in the copyright section:
the publisher is the copyright holder of this work and the
author uses the Dutch legislation to make this work public.

Improvising Accountability: The Everyday Governance Work of Responsible AI in the Public Sector

Ana Gagua

Delft University of Technology, Department of Multi-Actor Systems
agagua@tudelft.nl

Abstract

Public sector organisations are increasingly expected to operationalise Responsible AI (RAI) principles such as transparency and accountability. Yet translating these principles into everyday practice remains challenging and underexplored, particularly in public sector settings. In this study, I explore how accountability practices around transparency requirements are navigated inside a Dutch regulatory agency developing AI systems for risk-based inspections, drawing on my 1.5 years of organisational ethnography. The preliminary findings highlight how public sector practitioners engage in everyday governance work – creating accountability processes through improvised practices when formal structures are absent. This work contributes to a practice-based understanding of RAI and challenges assumptions about organisational maturity as the solution to operationalisation challenges.

Introduction

In recent years, governments, companies, and international bodies have issued over 100 ethical frameworks for AI (Jobin et al. 2019). However, translating these high-level principles into actionable organisational practices remains a major challenge (Morley et al. 2020, 2023; Schiff et al. 2020). The issues include a lack of concrete methods (Mittelstadt 2019; Peters et al. 2020), blurred accountability across teams (Schiff et al. 2021), and organisational dynamics that discourage ethical reflection (Hagendorff 2020).

The practice of RAI is even more pressing in the public sector. Public organisations must uphold key public values such as fairness, justice, democratic participation, and accountability (Bozeman 2007; Symes 1999). While public values like accountability and transparency align with RAI principles, RAI brings more nuanced and detailed requirements that challenge existing bureaucratic traditions and processes. Public organisations face organisational constraints, legal obligations, and political pressure. These

intersect in complex ways with emerging RAI expectations. Existing organisational routines may support, delay, or distort RAI expectations, depending on how they intersect with evolving technologies and shifting responsibilities.

This creates institutional ambiguity - the absence of clear mandates, ownership structures, or procedural norms for taking responsibility in algorithmic decision-making (Wieringa 2020). For this study, Responsible AI governance is defined as encompassing formal and informal processes, informed by normative frameworks from digital ethics and regulation (Dignum 2019; Morley et al. 2020; Floridi and Cowsls 2022; Sigfrids et al. 2022). While many studies emphasise gaps and operationalisation failures, few examine empirically how public sector practitioners practice RAI.

This study examines accountability practices around transparency requirements in the context of institutional ambiguity. Transparency requirements refer to expectations that AI systems be communicated and understandable to relevant audiences, including explainability tools and broader disclosure practices (Ananny and Crawford 2018; Diakopoulos 2016). For the scope of this study, accountability refers to processes through which responsibility for transparency practices is established, following a relational understanding of accountability in algorithmic contexts (Wieringa 2020).

The analysis draws on ethical abduction (Tanweer, 2022), explaining how practitioners collectively reason through ethical problems when no predetermined solutions exist. This perspective aligns with broader movements in applied ethics, particularly empirical ethics (Widdershoven et al. 2009) and practice-based approaches (Verbeek 2020). During data analysis, I observed that governance activities accompanied this ethical reasoning - practitioners were not only making ethical decisions but taking responsibility for operationalising them. This led us to conceptualise everyday governance work - the situated, improvised efforts through

which practitioners navigate RAI expectations and create accountability without formal protocols.

Research Context

This study draws on original fieldwork I conducted in October 2023 - June 2025 in a Dutch regulatory organisation. Regulatory bodies, such as food safety, environmental protection, and financial oversight, increasingly employ AI technologies to enhance decision-making and resource allocation (Zardiashvili et al. 2019). Despite their growing reliance on AI, few empirical studies have examined how ethical and governance dilemmas emerge in such settings. The Dutch context is valuable as the Netherlands leads globally in RAI (Global Index on Responsible AI 2024) and is characterised by high structural capacity and citizen-orientation (Jugl 2025), complementing its global reputation for RAI leadership. The case organisation was chosen for its ongoing development of AI-informed regulatory tools and openness to research collaboration.

Research Design

The research employs a 1.5-year organisational ethnography to examine the social practices, meanings, and power dynamics within organisations (van Hulst, Ybema, and Yanow 2017). Rather than treating practitioners solely as research subjects, this study engaged them as collaborative partners in shaping the research focus and methods. This participatory approach included early-stage consultations with data scientists and AI developers to refine the research questions (Sanders and Stappers 2014; Creswell and Poth 2018), making the ethnography transdisciplinary in character. Data collection methods included:

1. 15 semi-structured interviews with key organisational roles, including data scientists, inspectors, domain experts, and managers.
2. Participation in five organisational meetings where AI systems were introduced and discussed.
3. Informal conversations and observational fieldnotes.
4. Document analysis of internal strategy papers, bias assessments, and policy templates.

The study focused on two AI systems at different lifecycle stages: an earlier inspection model (retrospective reflection) and a newer algorithm in development (active development). Through iterative coding and comparison across interviews, observations, and documents, the analysis traced how accountability practices emerged.

Preliminary Findings

Findings reflect the reality of everyday governance work - the challenging conditions practitioners face and how they respond when creating accountability around transparency requirements.

Practitioners operate within the reality of fragmented accountability. As one manager explained, "We thought legal was handling the risk documentation, but they assumed we had it.". Confusion over data ownership, model outputs, or documentation leads to "responsibility without mandate."

Faced with this fragmented reality, practitioners make situational decisions about who decides what to disclose, how to use explainability tools, and who bears responsibility. Some apply these tools to demonstrate compliance, while others adapt them pragmatically to context and audience. As one data scientist reflected, "We tried to build an explanation tool with SHAP values but then realised most users didn't understand what it meant." The variation in their use reflects the complexity of operationalising transparency with differing demands and expectations.

This navigation happens by learning by doing: Teams prototype their guidelines and iteratively adjust practices, cornerstones of ethical abduction. As one participant noted, "Sometimes we just had to go with what felt right. The formal process wasn't there yet, but we still had to make a call."

Conclusion

This study shows how practitioners improvise accountability around transparency work when formal governance is absent. Rather than signalling failure, this improvisation represents everyday governance that enables RAI operationalisation under institutional ambiguity.

These findings challenge assumptions about organisational readiness for RAI. What appears as organisational immaturity reflects institutional voids, where practitioners must invent governance processes. This improvisation demonstrates resilience and situated wisdom that deserve critical examination. If everyday governance work is how RAI actually happens, what does this mean for maturity models that assume formalisation is the answer?

While findings emerge from one Dutch regulatory agency, similar organisational challenges may exist in other public administration contexts. Future research should explore RAI across diverse bureaucratic settings to test this hypothesis. Practically, this underscores the urgency of supporting practitioners through ethical reasoning structures, not just compliance tools. For regulators, readiness may require rethinking where governance happens.

References

- Ananny, M.; and Crawford, K. 2018. Seeing Without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability. *New Media & Society* 20(3): 973–989. doi.org/10.1177/1461444816676645.
- Bozeman, B. 2007. *Public Values and Public Interest: Counterbalancing Economic Individualism*. Georgetown University Press.
- Creswell, J. W., & Poth, C. N. 2018. *Qualitative Inquiry and Research Design: Choosing Among Five Approaches* (4th ed.). SAGE Publications.
- Diakopoulos, N. 2016. Accountability in algorithmic decision making. *Communications of the ACM*, 59(2): 56–62. doi.org/10.1145/2844110.
- Dignum, V. 2019. *Responsible Artificial Intelligence*. Cham, Switzerland: Springer International Publishing. doi.org/10.1007/978-3-030-30371-6.
- Floridi, L.; and Cowls, J. 2022. A Unified Framework of Five Principles for AI in Society. In *Machine Learning and the City*, 535–545. Hoboken, NJ: Wiley. doi.org/10.1002/9781119815075.ch45.
- Global Index on Responsible AI. 2024. 2024 Global Index on Responsible AI. Accessed August 14, 2025. <https://www.global-index.ai/>
- Hagendorff, T. 2020. The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines* 30(1): 99–120. doi.org/10.1007/s11023-020-09517-8.
- Jobin, A.; Ienca, M.; and Vayena, E. 2019. The Global Landscape of AI Ethics Guidelines. *Nature Machine Intelligence* 1(9): 389–399. doi.org/10.1038/s42256-019-0088-2.
- Jugl, M. 2025. Patterns of bureaucracy: Conceptualizing administrative traditions. *Journal of Public Administration Research and Theory* 35(3): 277–291. doi.org/10.1093/jopart/maaf012.
- Mittelstadt, B. 2019. Principles Alone Cannot Guarantee Ethical AI. *Nature Machine Intelligence* 1(11): 501–507. doi.org/10.1038/s42256-019-0114-4.
- Morley, J.; Floridi, L.; Kinsey, L.; and Elhalal, A. 2020. From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods, and Research to Translate Principles into Practices. *Science and Engineering Ethics* 26(4): 2141–2168. doi.org/10.1007/s11948-019-00165-5.
- Morley, J.; Kinsey, L.; Elhalal, A.; Garcia, F.; Ziosi, M.; and Floridi, L. 2023. Operationalising AI Ethics: Barriers, Enablers, and Next Steps. *AI & Society* 38(2): 411–423. doi.org/10.1007/s00146-021-01308-8.
- Peters, D.; Vold, K.; Robinson, D.; and Calvo, R. A. 2020. Responsible AI—Two Frameworks for Ethical Design Practice. *IEEE Transactions on Technology and Society* 1(1): 34–47. doi.org/10.1109/TTS.2020.2974991.
- Sanders, E. B. N.; and Stappers, P. J. 2014. Probes, toolkits and prototypes: Three approaches to making in codesigning. *CoDesign*, 10(1): 5–14. <https://doi.org/10.1080/15710882.2014.888183>
- Schiff, D.; Biddle, J.; Borenstein, J.; and Laas, K. 2020. What's Next for AI Ethics, Policy, and Governance? A Global Overview. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 153–158. New York, NY: Association for Computing Machinery. doi.org/10.1145/3375627.3375804.
- Schiff, D.; Rakova, B.; Ayesh, A.; Fanti, A.; and Lennon, M. 2021. Explaining the Principles to Practices Gap in AI. *IEEE Technology and Society Magazine* 40(1): 81–94. doi.org/10.1109/MTS.2021.3056286.
- Sigfrids, A.; Nieminen, M.; Leikas, J.; and Pikkuaho, P. 2022. How should public administrations foster the ethical development and use of artificial intelligence? A review of proposals for developing governance of AI. *Frontiers in Human Dynamics*, 4: 858108. <https://doi.org/10.3389/fhumd.2022.858108>
- Symes, A. 1999. Creating public value: Strategic management in government. *International Public Management Journal*, 2(1): 158–167. [https://doi.org/10.1016/S1096-7494\(00\)87438-3](https://doi.org/10.1016/S1096-7494(00)87438-3).
- Tanweer, A. 2022. Tradeoffs all the way down: Ethical abduction as a decision-making process for data-intensive technology development. *Big Data & Society*, 9(1). <https://doi.org/10.1177/20539517221101351>.
- van Hulst, M.; Ybema, S. B.; and Yanow, D. 2017. Ethnography and organizational processes. In Tsoukas, H.; and Langley, A., eds., *The Sage Handbook of Process Organization Studies*, 223–236. London: Sage.
- Verbeek, P.-P. 2020. *Moralizing Technology: Understanding and Designing the Morality of Things*. University of Chicago Press.
- Widdershoven, G.; Abma, T.; and Molewijk, B. 2009. Empirical ethics as dialogical practice. *Bioethics*, 23(4): 236–248. <https://doi.org/10.1111/j.1467-8519.2009.01712.x>
- Wieringa, M. 2020. What to account for when accounting for algorithms: A systematic literature review on algorithmic accountability. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 1–18.
- Zardiashvili, L.; Bieger, J.; Dechesne, F.; and Dignum, V. 2019. AI ethics for law enforcement: A study into requirements for responsible use of AI at the Dutch police. *Delphi – Interdisciplinary Review of Emerging Technologies*, 2019(4): 77–93. <https://doi.org/10.21552/delphi/2019/4/7>.