

# Co-Constructing Sociotechnical AI Governance: Participatory System Mapping Using Algorithm Registers (Extended Abstract)

Íñigo de Troya\*, Maurus Enbergs, Neelke Doorn, Roel Dobbe

TU Delft

\*i.m.d.r.detroya@tudelft.nl

## Abstract

Algorithm registers have been championed as a means of providing transparency on the use of algorithms in public services. Yet potential publics differ in their expectations of what should be made transparent and how, as well as in their interest in and ability to parse the information currently published in the registers. Moreover, it remains unclear how these instruments can represent the sociotechnical systems in which these algorithms are embedded, and how system-level transparency can facilitate accountability. In this paper, we ask, *what do algorithm registers reveal (and occlude) about the sociotechnical systems governing algorithmic systems, and how can diverse stakeholder perspectives inform a more pluralistic system-theoretic safety analysis?* To do this, we probe the municipal algorithm register of a Dutch city through a case study of a decision-support tool for caseworkers' assessment of citizens' welfare benefits eligibility based on legal automation through a business rule engine. Through interviews, surveys, and participatory system mapping workshops (with staff from a municipality, civil society organisations, and municipal Ombudsmen, N=8), we seek to understand to what extent the register allows stakeholders to map the algorithmic system in question. These maps inform a System-Theoretic Process Analysis (STPA) that situates the register within a wider sociotechnical governance structure. Participants' contributions allow us to identify potential safety hazards which would not have been possible to see using the algorithm register alone, including benefits eligibility denial, system performance deterioration, and inability to contest wrongful decisions. By engaging both direct and indirect stakeholders, we reflect on the normative dimensions of algorithm governance efforts and how politics shape the practice of system safety analysis.

**Extended version** — <https://arxiv.org/abs/2608.12166>

## Ethics Statement

This study was approved by a Human Research Ethics Council review at TU Delft, which included providing a risk management and mitigation plan, and a data management plan. Participants were informed about the scope of the study and gave their informed consent prior to their participation.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

## Acknowledgements

We thank the participants of the study for their generosity in sharing their time, perspectives, and contributions to this study, and in particular the municipality, for their openness and willingness to reflect on their governance efforts. We also want to thank the AIES reviewers for their thoughtful suggestions and critiques. This research was conducted as part of the Gravitation research program Hybrid Intelligence, funded by the Dutch Research Council (Nederlandse Organisatie voor Wetenschappelijk Onderzoek) under file number 024.004.022.

## References

- Binns, R. 2018. Algorithmic accountability and public reason. *Philosophy & Technology*, 31(4): 543–556.
- Blacktrace Mergers & Acquisitions. 2024. Overname Mendix partner Bizzomate versterkt hyperautomation propositie Ciphix. <https://www.blacktrace.nl/nl/transacties/transactie/overname-van-mendix-platinum-partner-bizzomate-versterkt-hyperautomation-propositie-van-ciphix.html>.
- Burrell, J. 2016. How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1): 2053951715622512.
- Cath, C.; and Jansen, F. 2021. Dutch Comfort: The limits of AI governance through municipal registers. *arXiv preprint arXiv:2109.02944*.
- Cellard, L. 2020. *Theatres of algorithmic transparency: a post-digital ethnography*. Ph.D. thesis, University of Warwick.
- Chalke, J. 2023. Implications and Challenges of AI for Parliamentary Ombuds Work in Canada. *Canadian Parliamentary Review*, 46(2).
- Chen, B. J.; and Metcalf, J. 2024. Explainer: A sociotechnical approach to AI policy. *Data & Society*.
- de Troya, Í.; Kernahan, J.; Doorn, N.; Dignum, V.; and Dobbe, R. 2025. Misabstraction in Sociotechnical Systems. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, 1829–1842.
- Dobbe, R. 2022. System Safety and Artificial Intelligence. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, 1584–1584.

- Dobbe, R. 2025. AI Safety is Stuck in Technical Terms—A System Safety Response to the International AI Safety Report. *arXiv preprint arXiv:2503.04743*.
- Escher, N.; Bilik, J.; Banovic, N.; and Green, B. 2024. Code-ifying the Law: How Disciplinary Divides Afflict the Development of Legal Software. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW2): 1–37.
- Felzmann, H.; Villaronga, E. F.; Lutz, C.; and Tamò-Larrieux, A. 2019. Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data & Society*, 6(1): 2053951719860542.
- Fenster, M. 2005. The opacity of transparency. *Iowa L. Rev.*, 91: 885.
- Frederik, J. 2021. De tragedie achter de toeslagenaffaire. *De Correspondent*. Published: 2021-01-15.
- Greenbaum, J.; and Kyng, M. 1991. Introduction: Situated Design. In *Design at Work: Cooperative Design of Computer Systems*, 1–24. CRC Press.
- Haataja, M.; van de Fliert, L.; and Rautio, P. 2020. Public AI Registers: Realising AI transparency and civic participation in government use of AI. <https://openresearch.amsterdam/en/page/73074/public-ai-registers>.
- Haraway, D. 1988. Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies*, 14(3).
- Iwanska, K.; Skoric, V.; Fanucci, F.; Keskindemir, B.; and Kokkula, S. 2024. Towards an AI Act that serves people and society: strategic actions or civil society and funders on the enforcement of the EU AI Act.
- Kaeseberg, T. 2019. The Code-ification of law and its potential effects. *Computer Law Review International*, 20(4): 107–110.
- Kallina, E.; Bohné, T.; and Singh, J. 2025. Stakeholder Participation for Responsible AI Development: Disconnects Between Guidance and Current Practice. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, 1060–1079.
- Kroll, J. A.; Huey, J.; Barocas, S.; Felten, E. W.; Reidenberg, J. R.; Robinson, D. G.; and Yu, H. 2017. Accountable Algorithms. *University of Pennsylvania Law Review*, 165(3): 633.
- Leveson, N. 2004. A new accident model for engineering safer systems. *Safety Science*, 42(4): 237–270.
- Leveson, N. G. 2011. *Engineering a safer world: Systems thinking applied to safety*. The MIT Press.
- Leveson, N. G.; and Thomas, J. P. 2018. STPA Handbook. *MIT Partnership for Systems Approaches to Safety and Security (PSASS)*.
- Madise, Ü.; and Pilvik, K. 2024. The role of Ombuds institutions in a rush of digitalization. In *La Ley del Ararteko. Construyendo el futuro: una reflexión sobre las defensorías del pueblo: 2023ko ekainaren 12an eta 13an Vitoria-Gasteizen egindako mintegia= Seminario celebrado en Vitoria-Gasteiz los días 12 y 13 de junio de 2023*, 239–264. Eusko Legebiltzarra= Parlamento Vasco.
- Ministry of Internal Affairs, U. I., (Ministerie van Binnenlandse Zaken en Koninkrijksrelaties). 2023. Doelgroepenanalyse Algoritmeregister: Kwalitatief onderzoek naar hoe (potentiële) gebruikers het Igitmeregister ervaren. Accessed on 30-12-2025.
- Murray-Rust, D.; Alfrink, K.; and Zaga, C. 2025. Towards Meaningful Transparency in Civic AI Systems. *arXiv preprint arXiv:2510.07889*.
- Nieuwenhuizen, E. 2025. Algorithm Registers: A Box-Ticking Exercise or Meaningful Tool for Transparency? *Information Polity*, 29(4).
- Norval, C.; Cornelius, K.; Cobbe, J.; and Singh, J. 2022. Disclosure by Design: Designing information disclosures to support meaningful transparency and accountability. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 679–690.
- Nouws, S.; Martinez De Rituerto De Troya, Í.; Dobbe, R.; and Janssen, M. 2023. Diagnosing and addressing emergent harms in the design process of public AI and algorithmic systems. In *Proceedings of the 24th Annual International Conference on Digital Government Research*, 679–681.
- Paudyal, P.; and William Wong, B. 2018. Algorithmic opacity: making algorithmic processes transparent through abstraction hierarchy. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 62(1): 192–196.
- Raji, I. D.; and Dobbe, R. I. J. 2020. Concrete problems in AI safety, revisited. In *Workshop on Machine Learning In Real Life at the International Conference on Learning Representations. Addis Abeba*.
- Rismani, S.; Shelby, R.; Smart, A.; Jatho, E.; Kroll, J.; Moon, A.; and Rostamzadeh, N. 2023. From plane crashes to algorithmic harm: applicability of safety engineering frameworks for responsible ML. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–18.
- Rotterdam Court of Auditors. 2024. Kleur bekennen: vervolgonderzoek naar algoritmes. <https://rekenkamer.rotterdam.nl/onderzoeken/kleur-bekennen/>. Published: 2024-05-13.
- Sloane, M.; Moss, E.; Awomolo, O.; and Forlano, L. 2022. Participation is not a design fix for machine learning. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, 1–6.
- van Vliet, M.; Schuitemaker, N.; Espana, S.; van de Weerd, I.; and Brinkkemper, S. 2024. Defining and implementing algorithm registers: an organizational perspective. In *ECIS 2024 Proceedings*. 5.
- Wieringa, M. 2020. What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 1–18.
- Zejnilović, L.; Lavado, S.; Martínez de Rituerto de Troya, Í.; Sim, S.; and Bell, A. 2020. Algorithmic long-term unemployment risk assessment in use: counselors’ perceptions and use practices. *Global Perspectives*, 1(1).