

Distributed Responsibility, Contested Accountability: Auditing and Governing AI Systems across the Supply Chain

Lelia Hanslik

Technische Universität Berlin, Berlin, Germany
hanslik@tu-berlin.de

Introduction

AI systems are increasingly produced through supply chains in which different actors, data workers, (foundation) model developers, cloud infrastructure providers, application developers, and deployers, each contribute components to systems which they collectively produce but over which none of the actors has full control or visibility (Cobbe, Veale, and Singh 2023; Widder and Nafus 2023). These supply chains are not democratic networks of equal participants. Their emergence has been driven by, among other cultural and economic shifts, a significant concentration of computational capacity, privatised funding, and technical innovation in the hands of a few powerful actors (Hopkins et al. 2025b; Rikap 2024).

They are characterised by asymmetric power dynamics (Valdivia 2025; Gray Widder, West, and Whittaker 2023), which become particularly apparent when central actors take unilateral actions that affect all downstream participants (Hopkins et al. 2025a), while the underlying innovation and knowledge from which AI systems are built is systematically captured and turned into rent-bearing assets by these central actors, termed intellectual monopolies by Rikap (2024). Valdivia (2025) further shows that harms and benefits are globally unequally distributed, i.e. AI supply chains reproduce social inequalities and environmental harms.

In what Muldoon, Valdivia, and Badger (2026) term the sphere of “AI deployment” (as opposed to “AI infrastructure” and “AI preparation”), the supply chain is non-modular (Hopkins et al. 2025a; Neumann and Singh 2026): components interact in ways that produce emergent effects no single actor designed, intended, or can observe (Cobbe, Veale, and Singh 2023). Harms surfaced at the application level, such as compounded representational or allocative harms (Shelby et al. 2023), may therefore originate in interactions far upstream.

When algorithmic harms are surfaced, questions arise about accountability, contestation, and redress. Accountability for such harms is often attributed to system components or actors through evaluations or audits, here broadly understood as post-market-entry assessments and analyses. When regulators determine whether an AI system is compliant,

when a company claims its system is safe, or when a third party audits a deployment, evaluations provide the evidentiary basis for those accountability claims.

The European Union’s recognition of evaluations as a cornerstone of algorithmic accountability is reflected in the evaluation ecosystem it is actively constructing (independent audits under Art. 37 Digital Services Act, audited descriptions of consumer profiling techniques under Art. 15 Digital Markets Act, and conformity assessments for high-risk AI systems under the AI Act), yet the conceptual foundations of what these evaluations should capture remain vague and heterogeneous across frameworks. How accountability is conceptualised within evaluations is therefore a socio-technical and politically relevant question, not merely a theoretical one, as flawed conceptualisations may produce systematically misleading accountability claims. Given power concentration, non-modularity, asymmetric control, and limited observability, evaluations designed around individual actors or components will measure compliance at the points they can see while remaining blind to the interactions between those points where cascade-emergent harms originate (Neumann and Singh 2026). The question should therefore not only be whether evaluations of AI systems are technically rigorous, but whether their underlying conceptualisation of accountability is adequate to the supply chain conditions under which algorithmic harms actually occur.

Existing approaches to improve AI evaluations advocate for enhanced traceability (Kroll 2021), participatory methods (Madaio et al. 2020) and strengthened external (Raji et al. 2022) and internal evaluation mechanisms (Raji et al. 2020). Yet the literature largely addresses accountability at the level of individual actors or components, without reaching the supply chain interactions where cascade-emergent harms originate and component-level harms are amplified. The risk that harmful systems receive clean audit reports therefore remains largely unaddressed.

Structure and Research Question

My overarching research question is: *How do the structural characteristics of AI supply chains, including power concentration, non-modularity, and opacity, shape the conditions under which accountability can be attributed, systems can be evaluated, and legal obligations can be enforced?* I address it in three mutually informing strands: a *conceptual*

strand asking whether accountability conceptualisations in the socio-technical AI evaluation literature account for supply chain interdependencies; an *empirical* strand on the role of AI-as-a-Service (AIaaS) and infrastructure providers in the deployment phase, combining quantitative measurement and qualitative methods; and an *institutional* strand on how staff of enforcement agencies, largely absent from the AI accountability literature, navigate supply chain dynamics and opacity.

Ongoing and Planned Work

My first conceptual dissertation paper, a PRISMA-guided systematic literature review (51 of 443 publications from Scopus, Web of Science, and the ACM Digital Library included), examines how algorithmic accountability is conceptualised in socio-technical AI evaluations and how supply chain interdependencies are reflected. Preliminary findings indicate that interdependencies are mostly acknowledged implicitly and decoupled from accountability claims. Two further papers from thesis supervisions compare software testing with AI auditing, and assess DSA audit reports against academic accountability standards.

In the empirical strand of my work, I plan to audit infrastructure opacity in Europe's most-used consumer LLM chatbots (via sub-processor lists and network traffic) and examine AIaaS provider power in EU public sector deployments (via procurement documents and FOI requests).

For the institutional strand of my work, I will interview staff of national (and possibly EU-level) enforcement agencies on what supply-chain-aware enforcement of, e.g., the AI Act requires and where current institutional capacity falls short.

Contributions

The dissertation links conceptual, empirical, and institutional dimensions largely treated in isolation, and aims to contribute (1) a mapping of how accountability is conceptualised in socio-technical AI evaluations and where supply chain interdependencies are missing, (2) replicable methods for auditing infrastructure opacity from outside, and (3) empirically grounded requirements for supply-chain-aware enforcement. Its working hypothesis is that power concentration along AI supply chains produces structural opacity that places actors responsible for cascade-emergent harms beyond the reach of existing accountability instruments.

References

Cobbe, J.; Veale, M.; and Singh, J. 2023. Understanding accountability in algorithmic supply chains. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 1186–1197.

Gray Widder, D.; West, S.; and Whittaker, M. 2023. Open (For Business): Big Tech, Concentrated Power, and the Political Economy of Open AI. *SSRN Electronic Journal*.

Hopkins, A.; Cen, S. H.; Struckman, I.; Ilyas, A.; Videgaray, L.; and Madry, A. 2025a. AI supply chains: An emerging

ecosystem of ai actors, products, and services. In *Proceedings of the 2025 AAAI/ACM Conference on AI, Ethics, and Society*, volume 8, 1266–1277.

Hopkins, A.; Struckman, I.; Klyman, K.; and Silbey, S. S. 2025b. Recourse, Repair, Reparation, & Prevention: A Stakeholder Analysis of AI Supply Chains. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, 209–227.

Kroll, J. A. 2021. Outlining Traceability: A Principle for Operationalizing Accountability in Computing Systems. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 758–771.

Madaio, M. A.; Stark, L.; Wortman Vaughan, J.; and Wallach, H. 2020. Co-Designing Checklists to Understand Organizational Challenges and Opportunities around Fairness in AI. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–14.

Muldoon, J.; Valdivia, A.; and Badger, A. 2026. The politics of artificial intelligence supply chains. *AI & SOCIETY*, 41(2): 1175–1187.

Neumann, A.; and Singh, J. 2026. AI Safety Evaluations Need To Consider Cascading Effects. In *Proceedings of 2026 IASEAI Conference*, volume 2, 460–487.

Raji, I. D.; Smart, A.; White, R. N.; Mitchell, M.; Gebru, T.; Hutchinson, B.; Smith-Loud, J.; Theron, D.; and Barnes, P. 2020. Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*, 33–44.

Raji, I. D.; Xu, P.; Honigsberg, C.; and Ho, D. 2022. Outsider Oversight: Designing a Third Party Audit Ecosystem for AI Governance. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 557–571.

Rikap, C. 2024. Varieties of corporate innovation systems and their interplay with global and national systems: Amazon, Facebook, Google and Microsoft's strategies to produce and appropriate artificial intelligence. *Review of International Political Economy*, 31(6): 1735–1763.

Shelby, R.; Rismani, S.; Henne, K.; Moon, Aj.; Ros-tamzadeh, N.; Nicholas, P.; Yilla-Akbari, N.; Gallegos, J.; Smart, A.; Garcia, E.; and Virk, G. 2023. Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 723–741.

Valdivia, A. 2025. The supply chain capitalism of AI: A call to (re) think algorithmic harms and resistance through environmental lens. *Information, Communication & Society*, 28(12): 2118–2134.

Widder, D. G.; and Nafus, D. 2023. Dislocated accountabilities in the “AI supply chain”: Modularity and developers' notions of responsibility. *Big Data & Society*, 10(1): 20539517231177620.