

Three legal forms of human oversight of AI and ADM: How human oversight manages agency and accountability in organisational context

Jake Goldenfein

As automation and AI reorganise social, political, and economic life, law's reflex is often towards the human. By making human oversight a central feature of AI governance, policy makers hope to promote the safe realisation of automation's economic and social promises. What exactly the human overseer is implemented to achieve, however, is less clear. Over time, the literature has described multiple goals such as system performance, imbuing systems with the normative value of human agency, and locating responsibility within automated workflows. But the efficacy and value of those rationales is difficult to demonstrate. In light of empirical studies consistently indicating that human oversight does not improve system performance, that human oversight cannot subordinate automation to human agency, and that human oversight improperly responsabilises humans, how it remains an entrenched regulatory orthodoxy is difficult to understand.

This paper suggests that the growing scholarly and policy literatures on human oversight makes understanding how human oversight interacts with those goals largely impossible. Empirical literatures attempt to draw generalisable findings about system performance from highly stylised experiments and simulations. Normative scholarship describes the value of human agency without any connection to real examples of automation. Studies that interface regulatory thinking with disciplines like human factors engineering and ergonomics become context agnostic, treat all examples of automation and human oversight as effectively equivalent, and propose that policy makers pursue *sui generis* strategies for whatever instance of automation is being governed. All three approaches treat human oversight as a generic and unitary concept, in ways that obscure its operational value and rationales.

To better understand the ways that human oversight interacts with the goals of efficacy, agency, and accountability, this paper rejects the genericisation and describes three different legal forms of human oversight, taxonomized by organisational and labour arrangements that elucidate the functions of human oversight in operation. These legal regimes alternatively demand a human overseer as a 'final decision-maker', as a 'system operator', and as a 'supply chain input', with each: producing different relationships between human overseers and the organisation that controls the means and purposes of automation; configuring human agency in different ways with different accountability consequences; and enabling different organisational forms, automation workflows, and labour relations for the strategic management of efficiency, controllability, and liability.

In defining these legal forms and offering case studies of how they are operationalised, the paper provides conceptual clarification of a complex and confused body of research that often treats human oversight as a generic and unitary regulatory strategy. And more importantly, it elucidates how regulation requiring human oversight functions, in different ways, as a managerial instrument for strategically coordinating competing organisational objectives, such as regulatory compliance, system efficiency, and the distribution of accountability and liability, though enabling discrete organisational and labour arrangements.

Bibliography

Sarah Sterz, Kevin Baum, Sebastian Biewer, Holger Hermanns, Anne Lauber-Rönsberg, Philip Meinel, and Markus Langer, 'On the Quest for Effectiveness in Human Oversight: Interdisciplinary Perspectives' (2024) *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*.

Jake Goldenfein, 'Lost in the loop – Who is the “human” of the Human in the Loop' in Gavin Sullivan, Fleur Johns, Dimitri Van Den Meerssche (eds) *Global Governance by Data* (Cambridge University Press, 2026).

Rebecca Crootof, Margot E Kaminski and W Nicholson Price II, 'Humans in the Loop' (2023) 76(2) *Vanderbilt Law Review* 429.

Matteo Pasquinelli, Christina Alaimo and Alessandro Gandini, 'AI at Work: Automation, Distributed Cognition, and Cultural Embeddedness' (2024) 15(1) *Tecnoscienza* 99.

Jennifer Raso, 'Displacement as Regulation: New Regulatory Technologies and Front-Line Decision-Making in Ontario Works' (2017) 32(1) *Canadian Journal of Law and Society* 75.