Dynamics of Automation Bias in Human-Algorithm Interaction based Public Policy Decision Making

Anshumaan Goel, Sudip Patra
PhD Scholar, Associate Professor
OP Jindal Global University
Sonipat, Haryana, India
{agoel, spatra}@jgu.edu.in

Abstract. *As public sector institutions increasingly transition from rule-based automation to advanced intelligent agents, artificial intelligence is reshaping strategic decision-making and administrative governance. However, embedding algorithmic decision support into bureaucratic workflows introduces a critical socio-technical paradox: the risk of automation bias, where human experts disproportionately defer to algorithmic outputs, potentially eroding institutional accountability and discretionary nuance. This study investigates the dynamics of human-algorithm interaction (HAI) within public policy decision-making, specifically examining how administrative officers interact with AI algorithms deployed in policing and social security organizations. Utilizing a mixed-methods approach grounded in Causal Loop Diagrams (CLDs), this research maps the complex feedback loops that govern trust, deskilling cognitive offloading, and systemic inertia in algorithmic public administration. Empirical data derived from a structured survey of administrative officers and use cases of predictive policing algorithms reveals that while automated risk profiling accelerates desk reviews, it simultaneously creates a psychological safety net that induces cognitive compliance. Preliminary systems-thinking models illustrate how a reliance on algorithmic parameters can inadvertently obscure contextual anomalies and distort frontline regulatory enforcement. Younger officers exhibited higher AI trust (3.42 vs. 3.00), indicating automation bias, while field officers preferred greater human oversight (3.71 vs. 3.13). The former system was prone to algorithmic bias while the latter system was prone to selective adherence bias, both resulting in amplified anti-minority outcomes in predictive policing HAI. However, as a novel tool, CLDs provide Balancing Loops to mitigate such HAI bias amplification in algorithmic decision making.*

By evaluating the transition of AI from a passive tool to an active cognitive partner, this paper critically addresses pivotal themes of meaningful human control, contextuality, cognitive sovereignty and institutional transparency. The findings contribute a concrete policy framework to act as an automated regulatory brake, counteracting the self-amplifying trajectory of HAI systems and to design intelligent systems that augment, rather than eclipse, human expertise. Ultimately, this research provides insights for policymakers and system architects seeking to align automated public infrastructure with robust accountability frameworks, ensuring that the next wave of intelligent systems enhances long-term institutional resilience and public trust.

Keywords. AI Bias, Public Policy, Decision making, Human-Algorithm Interaction

References

Laux, J., & Ruschemeier, H. (2025). Automation bias in the AI Act: On the legal implications of attempting to de-bias human oversight of AI. European Journal of Risk Regulation. Advance online publication. https://doi.org/10.1017/err.2025.10033

Busuioc, M. (2021). Accountable artificial intelligence in public administration: The challenge of machine learning. Public Administration Review, 81(5), 825–836.

Levy, K., Chasalow, K. E., & Riley, S. (2021). Algorithms and decision-making in the public sector. Annual Review of Law and Social Science, 17, 309–334. https://doi.org/10.1146/annurev-lawsocsci-041221-023808

Akbarighatar, P., Rinta-Kahila, T., & Soliman, W. (2025). Algorithmic decision-making in public services: Two empirical cases of algorithmic incompetence. Proceedings of the Hawaii International Conference on System Sciences (HICSS).

Salwei, M. E., & Carayon, P. (2022). A sociotechnical systems framework for the application of artificial intelligence in health care delivery. Journal of Cognitive Engineering and Decision Making, 16(4), 194–206. https://doi.org/10.1177/15553434221097357

Farivar, F. (2025). Industry 5.0: A socio-technical system perspective on human agency and institutional legitimacy. Journal of Management & Organization. Advance online publication.