frontiers | Frontiers in Political Science

TYPE Editorial
PUBLISHED 20 May 2025
DOI 10.3389/fpos.2025.1611563

EDITED AND REVIEWED BY
Leslie Paul Thiele,
University of Florida, United States

*CORRESPONDENCE
Ben Wagner
✉ b.wagner@tudelft.nl
Johanne Kuebler
✉ johanne.kubler@inholland.nl
Monika Zalnieriute
✉ monika.zalnieriute@tf.vu.lt

RECEIVED 14 April 2025
ACCEPTED 30 April 2025
PUBLISHED 20 May 2025

CITATION
Wagner B, Kuebler J and Zalnieriute M (2025) Editorial: Humans in the loop: exploring the challenges of human participation in automated decision-making systems. *Front. Polit. Sci.* 7:1611563. doi: 10.3389/fpos.2025.1611563

Editorial: Humans in the loop: exploring the challenges of human participation in automated decision-making systems

Ben Wagner[1,2]*, Johanne Kuebler[2,3]* and Monika Zalnieriute[4]*

[1]Interdisciplinary Transformation University Austria (IT:U), Faculty of Technology, Policy and Management, Delft University of Technology, Delft, Netherlands, [2]Department of Creative Business, Inholland University of Applied Sciences, Hoofddorp, Netherlands, [3]Sustainable Computing Lab, Vienna University of Economics and Business, Vienna, Austria, [4]Faculty of Law, Vilnius University, Vilnius, Lithuania

KEYWORDS

human-in-the-loop (HITL), human oversight, human-algorithm interaction, human factors in automation, human-centered AI, collaborative decision-making

Editorial on the Research Topic
Humans in the loop: exploring the challenges of human participation in automated decision-making systems

"Human in the loop" (HITL) refers to a process or system design where human oversight, intervention, and collaboration are integrated into Automated Decision Making Systems (ADMS) at strategic points (Mosqueira-Rey et al., 2023). The apparent paradox of inserting human oversight into systems which use algorithms, data analysis, and predefined rules to make decisions with minimal or no human intervention, stems from the realization that gains in efficiency are offset with a potential for serious harm in the instances when ADMS err. Errors can occur through bias amplification in predictive systems, a lack of contextual awareness in high-stakes scenarios and an automated system's inability to recognize outliers (Angwin and Larson, 2022). HITL is seen as a critical ethical safeguard against AI systems making consequential decisions without appropriate scrutiny. Many emerging AI regulations and frameworks (EU AI Act, NIST AI Risk Management Framework, for instance) explicitly require human oversight for high-risk applications, making HITL not just an ethical choice but a compliance necessity in many contexts.

Historically, there was a shift from fully automated systems to more collaborative approaches. Early forms of automated decision-making systems, such as expert systems MYCIN and DENDRAL developed in the 1960s and 1970s, operated under a design philosophy that sought to minimize human intervention, which was seen as inefficient or inconsistent. They attempted to capture human expertise in formal rules that could be executed without oversight. A series of high-profile failures catalyzed a significant change in ADMS design philosophy, for instance the discovery that the COMPAS Criminal Risk Assessment, used to predict recidivism rates in the American criminal justice system, produced racially biased outcomes, or IBM's Watson Health's "unsafe and incorrect" cancer treatment recommendations (Hao, 2019; Strickland, 2019). By the mid-2010s, researchers and developers increasingly recognized that the most effective systems would combine machine efficiency with human judgment rather than attempting to eliminate the human entirely.

Implementing HITL raises questions about the concrete frameworks in which humans interact with automated decisions. For instance, what kind of decision options are humans provided, what data are made available to inform their decisions, is the time they are allocated to make their decisions sufficient and what level of oversight, accountability and liability are attached to human-made decisions? Most importantly, effective human-machine collaboration requires that human input is meaningful, and not just rubber-stamping decisions from ADMS (Wagner, 2019).

The authors in this Research Topic of articles initiate a discussion on the socio-legal and socio-technical challenges associated with humans participation in ADMS, considering insights from law, social science, philosophy, computer science and engineering. Salvini et al., use case studies in social care, aviation, and vehicle driver monitoring systems to illustrate the challenges and tensions involved in the use of ADMS, and highlight that human oversight of ADMS is neither easily defined nor well implemented. Haitsma, in his analysis of a landmark judgment of the Court of Justice of the European Union in 2022 on discrimination and algorithmic profiling in a border security context, shows that courts dealing with legal challenges to ADMS struggle to assess risks and to prescribe clear safeguards and how to effectively implement them. Constantino and Wagner explore accountability principles that would effectively govern intelligence and security services in democratic societies to ensure responsible, answerable practices. These proposed principles of accountability include acting within duty, explainability, necessity, proportionality, reporting and record keeping, redress, and continuous independent oversight. Human, in his philosophical reflection on the loss of human agency and the threat to human rights in the digital age, argues for a paradigm shift from a predominantly "individual-centric" approaches to data protection and consenting toward human-compatible, collective approaches. He goes on to propose the establishment of novel sociotechnical mechanisms, such as the "Advanced Data Protection Control (ADPC)", within internet infrastructures to facilitate effective communication between users and stakeholders.

In sum, the articles in this Research Topic contribute to the debate how HITL should evolve beyond simplistic "human approval" models toward more sophisticated collaborative frameworks where humans and automated systems complement each other's strengths while mitigating respective weaknesses. Implementing effective human oversight remains challenging, and, most importantly, responsible development of automated systems means to go beyond merely implementing technical safeguards, and instead to thoughtfully design human-machine relationships that align with societal values and priorities.

Author contributions

BW: Writing – review & editing, Conceptualization. JK: Writing – original draft. MZ: Writing – review & editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Angwin, J., and Larson, J. (2022). "Bias in criminal risk scores is mathematically inevitable, researchers say," in *Ethics of Data and Analytics* (Boca Raton, FL: Auerbach Publications), 265–267.

Hao, K. (2019). "AI Is Sending People to Jail—And Getting it Wrong," in *MIT Technology Review*. Available online at: https://www.technologyreview.com/2019/01/21/137783/algorithms-criminal-justice-ai/ (accessed April 5, 2025).

Mosqueira-Rey, E., Hernández-Pereira, E., Alonso-Ríos, D., Bobes-Bascarán, J., and Fernández-Leal, Á. (2023). Human-in-the-loop machine learning: a state of the art. *Artif. Intell. Rev.* 56, 3005–3054. doi: 10.1007/s10462-022-10246-w

Strickland, E. (2019). IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care. *IEEE Spectrum* 56, 24–31. doi: 10.1109/MSPEC.2019.8678513

Wagner, B. (2019). Liable, but not in control? Ensuring meaningful human agency in automated decision-making systems. *Policy Intern.* 11, 104–122. doi: 10.1002/poi3.198