EUROPEAN LAW & GOVERNANCE UNIVERSITY
An International University by
EPLO European Public Law Organization

ELGU Students & Alumni Policy Briefs

When Humans Oversee Machines: Accountability and the Rule of Law in AI-Assisted Public Decision-Making

Milica Jevtić

Abstract

The EU Artificial Intelligence Act relies on human oversight as the main safeguard for high-risk AI systems used in public administration. This article examines if oversight, as regulated by Articles 13, 14 and 26, guarantees accountability. It argues that, although the Act recognises automation bias and provides for human intervention, it does not sufficiently specify how effective oversight is to be achieved or what happens when it fails. Oversight thus remains a formal requirement rather than a real guarantee. The article recommends clearer rules on overseers' disagreement with the system and on the explanation of decisions.

Keywords: human oversight; accountability; EU AI Act; automation bias; public administration.

Introduction

The increased use of artificial intelligence (AI) systems in public administration raises the question of accountability for decisions these systems help shape and make. Cases such as Robodebt [1] and SyRI [2] illustrate the risks of automated systems operating within public administration without precisely defined human oversight, without transparency and without adequate legal frameworks. In this context, the EU Artificial Intelligence Act (hereafter EU AI Act) [3] establishes a regulatory framework for high-risk AI systems. Human oversight is seen as the main factor of control: natural persons with appropriate competence, training and authority should oversee AI systems and seek to ensure accountability and alignment with the rule of law (Art. 14). However, human oversight alone does not guarantee accountability, and research identifies automation bias and limitations in human oversight as significant

[1] Robodebt was an Australian automated debt-recovery scheme that raised welfare debts by comparing income data, largely without individual verification by officials. See *Report of the Royal Commission into the Robodebt Scheme*, Commonwealth of Australia 2023.

[2] SyRI (System Risk Indication) was a Dutch government system that linked administrative data to identify individuals at heightened risk of benefit and tax fraud. In 2020 the Hague District Court held that the legislation governing it violated Article 8 of the European Convention on Human Rights (ECHR), inter alia for insufficient transparency and verifiability. Rechtbank Den Haag, 5 February 2020, ECLI:NL:RBDHA:2020:1878.

[3] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), OJ L, 2024/1689, 12.7.2024. Unless otherwise indicated, articles cited in the text are articles of this Regulation.

challenges [4]. This article argues that human oversight under the EU AI Act is not sufficient enough on its own to ensure accountability in AI-assisted public decision-making. It examines the legal framework, the relationship between oversight and the rule of law, and three challenges: automation bias, the black box problem and the moral crumple zone.

Accountability, Rule of Law and AI-Assisted Public Decision-Making

Human oversight and the EU AI Act

Article 14 of the EU AI Act provides that high-risk AI systems must be designed and developed in such a way that natural persons can oversee them during their use (Art. 14(1)). To prevent or minimise the risks to health, safety or fundamental rights (Art. 14(2)), oversight measures must be either built into the system by the provider before its placement on the market or identified by the provider and implemented by the deployer (Art. 14(3)). The system must enable the persons assigned to oversight to understand its capacities and limitations and to monitor its operation, remain aware of the risk of automation bias, correctly interpret the system's output, and, if needed, disregard, override or reverse that output, or interrupt the system through a stop procedure (Art. 14(4)). Article 13 further elaborates this with transparency obligations and Article 26(2) requires that deployers assign human oversight to persons with the necessary competence, training, authority and support. It appears that important elements of oversight, such as competence, awareness of the risk of automation bias and the possibility of intervention, are recognised, but insufficiently developed [5]. Additionally, Green challenges the assumption that oversight is an effective safeguard and argues that the formal presence of a human decision maker might create a false sense of security, as public institutions would rely on systems that are still flawed [6]. Similarly, Fink examines the requirements of Article 14, highlighting the importance of introducing meaningful oversight over the formal assignment of such responsibilities [7]. As a result, the question arises of what constitutes meaningful oversight that enables accountability and who bears the consequences if oversight fails.

Human oversight, accountability and the rule of law

To determine whether human oversight is sufficient to guarantee the accountability of public administration, it must be assessed against the requirements of accountability and the rule of law. The Venice Commission's Updated Rule of Law Checklist identifies, among others, legality, legal certainty, the prevention of abuse of power, equality and access to justice as core elements of the rule of law[8]. However, the use of automated systems

[4] J. LAUX, Institutionalised distrust and human oversight of artificial intelligence: towards a democratic design of AI governance under the European Union AI Act, *AI & Society* 39 (2024), pp. 2853-2866, https://doi.org/10.1007/s00146-023-01777-z.

[5] LAUX, note 4 above.

[6] B. GREEN, The flaws of policies requiring human oversight of government algorithms, *Computer Law & Security Review* 45 (2022), 105681. https://www.sciencedirect.com/science/article/pii/S0267364922000292

[7] M. FINK, Human oversight under Article 14 of the EU AI Act, SSRN, 14 February 2025, http://dx.doi.org/10.2139/ssrn.5147196.

[8] Venice Commission, *Updated Rule of Law Checklist*, CDL-AD(2025)002.

may complicate these requirements. Busuioc defines public accountability as a process involving information, explanation and the possibility of consequences. She argues that algorithmic decision-making creates challenges at each of these stages, particularly by making it more difficult for citizens to understand, question and challenge decisions[9], which calls into question both the rule of law and the accountability of public administration. In this article, accountability of public administration means that it must always be clear which authority or person is responsible for a decision, that they can explain it, and that they bear the consequences if it is wrong. It is not a separate principle alongside the principles of the rule of law, but a condition for their functioning. Overall, the EU AI Act seeks to ensure accountability through human oversight of high-risk AI systems, in order to ensure legality and prevent arbitrariness.

Is human oversight enough?

Considering the legal solutions outlined above, the key question is whether human oversight, with its problems, is a sufficient guarantee of accountability.

The first limitation is automation bias. Research suggests that people may overestimate the accuracy of AI[10] and that this tendency is weaker among people with expertise[11]. Laux therefore suggests that better training can reduce the risk of automation bias [12]. However, expertise is not an ideal solution either, because expert overseers may rely too much on their own judgement[13] so training does not remove the error, but may shift it in another direction. The problem of automation bias is also addressed by Laux and Ruschemeier. They examine the legal implications of the EU AI Act's requirement that human overseers should be aware of the risk of over-relaying on AI outputs, which raises questions about how can this requirement be translated into effective oversight in practice [14]. Article 14(4)(b) requires awareness of automation bias, but awareness alone cannot prevent it. The provision also requires the ability to disregard or override outputs, but exercising that ability in practice may be difficult where an overseer lacks sufficient information, time or institutional support. Article 26(2) likewise does not specify what happens when oversight fails, which suggests that the Act establishes conditions for oversight without specifying how institutions should manage the practical limitations.

Closely connected is the black box problem: the difficulty of understanding how AI systems produce particular outputs. Burrell identifies this opacity as a significant challenge, particularly where machine-learning processes are difficult for humans to interpret [15] . Competence helps

[9] M. BUSUIOC, Accountable artificial intelligence: holding algorithms to account, *Public Administration Review* 81 (2021), pp. 825-836.
[10] S. M. JONES-JANG / Y. J. PARK, How do people react to AI failure? Automation bias, algorithmic aversion, and perceived controllability, *Journal of Computer-Mediated Communication* 28 (2022), zmac029.
[11] J. M. LOGG / J. A. MINSON / D. A. MOORE, Algorithm appreciation: people prefer algorithmic to human judgment, *Organizational Behavior and Human Decision Processes* 151 (2019), pp. 90-103.
[12] LAUX, note 4 above.
[13] LAUX, note 4 above.
[14] J. LAUX / H. RUSCHEMEIER, Automation bias in the AI Act: on the legal implications of attempting to de-bias human oversight of AI, *European Journal of Risk Regulation* 16 (2025), pp. 1519-1534.
[15] J. BURRELL, How the machine 'thinks': Understanding opacity in machine learning

only partly, because some systems are not sufficiently transparent. Busuioc argues that transparency alone is insufficient for accountability, since affected individuals must also be able to understand, question and challenge algorithmic outcomes [16]. The Act addresses this through Article 13, which ties sufficient transparency to two elements: a system design that enables the deployer to interpret the output, and instructions for use (Art. 13(1)-(2)). However, understanding an output is not necessarily equivalent to being able to explain the legally relevant reasons for an individual administrative decision. Article 86(1) gives the affected person the right to a clear and meaningful explanation of the role of the AI system and of the main elements of the decision, but this right is limited by the conditions and exceptions set out in that provision and does not itself guarantee a complete legal justification for every AI-assisted decision.

Finally, Elish introduces the concept of the moral crumple zone, which illustrates a situation in which human operators are held responsible for the actions of complex automated systems despite having limited control over them [17]. This concern is relevant to Green's analysis of government algorithms [18] and the risk that responsibility for systemic failures may be shifted to individual human operators. It shows that formally regulated oversight does not, by itself, guarantee meaningful control over the system, or accountability.

Conclusion

It can be concluded from the above that human oversight, as regulated by the EU AI Act, is not sufficient to guarantee accountability when decisions are made with the help of a high-risk AI system. The Act recognises automation bias and requires public deployers of high-risk AI systems to assess the impact on fundamental rights (Art. 27), but does not sufficiently specify how effective oversight and accountability could be ensured when oversight fails in practice. Thus, Article 14(4)(b) requires awareness of automation bias, while Article 13 enables interpretation of the output, but not explanation of the decision. Therefore, the Act should specify what the overseer should do when they disagree with the system or suspect that their judgement is affected by automation bias. It should also ensure that a decision can be explained by legally relevant reasons, not just by interpreting the system's output. Without this, human oversight remains a formal requirement rather than a real guarantee that public administration decisions will remain clear, explicable and open to challenge, which is what the rule of law requires.

---

algorithms, *Big Data & Society 3(1) (2016),* pp. 1–12. https://doi.org/10.1177/2053951715622512

[16] BUSUIOC, note 9 above.

[17] M. C. ELISH, Moral crumple zones: cautionary tales in human-robot interaction, *Engaging Science, Technology, and Society* 5 (2019), pp. 40-60.

[18] GREEN, note 6 above.