

AI ethics is an epistemic bubble that must burst

AI ethics promises fairness, accountability, and transparency, yet cannot deliver them. Contemporary AI ethics is an epistemic bubble in which the conditions required to realise these goals do not exist within its own structure. Its goals are admirable, but those it claims to protect have little meaningful role in defining the systems and criteria by which ethical compliance is determined (Birhane et al., 2022). Similar self-reinforcing tendencies have been identified in technological discourse and innovation-driven systems more broadly (Vinsel and Russell, 2020; Gertz, 2024). AI ethics is therefore asking the wrong question. Rather than pursuing impossible neutrality, it must interrogate whose knowledge builds systems and whose is excluded.

This paper introduces Equity Bias to explain how institutional AI ethics frameworks reproduce exclusion. It occurs through the selective incorporation of knowledge compatible with existing power structures, whilst marginalising forms of knowledge that challenge them. It provides a philosophical and practical framework for identifying how epistemic exclusion is embedded across the AI development lifecycle. It also shows how ethical governance processes can appear corrective whilst remaining structurally reinforcing.

The consequences are measurable. Hundreds of AI ethics guidelines now exist globally (Corrêa et al., 2023), yet automated decision-making in areas such as housing continues to produce discriminatory outcomes (Cheng et al., 2024). Those affected often have little meaningful recourse (Alon-Barkat et al., 2025). Harm persists not despite the ethics apparatus, but beneath its cover. The bubble not only fails but blocks the conditions under which genuine protection is possible.

Responses to ethical failures often follow a pattern which includes more guidelines and continued consultation with the same expert communities (Maclure and Morin-Martel, 2025). Equity Bias, when applied to AI ethics, reveals this pattern as epistemic replication: the reproduction of the same epistemic commitments regardless of outcome or harm. This invites comparison with bioethics, a field that has similarly grappled with questions of epistemic authority and representation (Hofmann, 2023).

Applied recursively, Equity Bias makes visible what the epistemic bubble conceals. Internal reform cannot resolve exclusions produced by the field's own structures. Bursting the bubble requires creating conditions in which multiple knowledge systems can contest and inform both AI development and the ethics field itself. The question is not whether AI ethics must change, but who gets to determine what it changes into.

Bibliography

- Alon-Barkat, S. *et al.* (2025) "Algorithmic discrimination in public service provision: Understanding citizens' attribution of responsibility for human versus algorithmic discriminatory outcomes," *Journal of Public Administration Research and Theory*, 35(4), pp. 469–488. Available at: <https://doi.org/10.1093/jopart/maaf024>.
- Birhane, A. *et al.* (2022) "Power to the People? Opportunities and Challenges for Participatory AI," *ACM International Conference Proceeding Series*. Association for Computing Machinery. Available at: <https://doi.org/10.1145/3551624.3555290>.
- Cheng, L. *et al.* (2024) "Algorithm-Assisted Decision Making and Racial Disparities in Housing: A Study of the Allegheny Housing Assessment Tool," *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. Association for the Advancement of Artificial Intelligence (AAAI), pp. 281–292. Available at: <https://doi.org/10.1609/aies.v7i1.31636>.
- Corrêa, N.K. *et al.* (2023) "Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance," *Patterns*, 4(10). Available at: <https://doi.org/10.1016/j.patter.2023.100857>.
- Gertz, N. (2024) *Nihilism and technology*. Second edition. London: Rowman & Littlefield.
- Hofmann, B. (2023) "Biases in bioethics: a narrative review," *BMC Medical Ethics*. BioMed Central Ltd. Available at: <https://doi.org/10.1186/s12910-023-00894-0>.
- Maclure, J. and Morin-Martel, A. (2025) "AI Ethics' Institutional Turn," *Digital Society*, 4(1), p. 18. Available at: <https://doi.org/10.1007/s44206-025-00174-x>.
- Vinsel, L. and Russell, A.L. (2020) *The innovation delusion: how our obsession with the new has disrupted the work that matters most*. New York: Currency.