The Case for Enforceable AI Governance

A Critical Reflection on Generative AI Ethics and Global Regulatory Challenges

Tobias Zeier, 12696372, the 27th of October 2025
University of Essex Online

Since late 2022, generative artificial intelligence has transformed the technological landscape, raising significant ethical concerns about privacy, algorithmic discrimination and transparency (Corrêa *et al.*, 2023). Whilst substantial work defines guiding values, establishing global agreement remains challenging given diverse stakeholder perspectives. However, a more fundamental concern emerges: AI ethics risks remaining "toothless" without enforcement mechanisms (Resseguier and Rodrigues, 2020). This reflection examines current international approaches, recommends a hybrid governance framework emphasising enforcement, and explores the implications of this position.

Global Approaches and Critical Limitations

The international landscape reveals different approaches. The EU AI Act represents comprehensive, risk-based legislation categorising AI systems by potential harm, requiring strict conformity assessments for high-risk applications (European Commission, 2024; World Economic Forum, 2024). The UK adopts principles-based, sector-specific regulation empowering existing regulators (Takhar, 2023), whilst the US pursues decentralised regulation through federal agencies (Lorica, 2025; World Economic Forum, 2024).

Critically, Resseguier and Rodrigues (2020) argue that contemporary AI ethics is "largely ineffective, trapped in an ethical principles approach" that industry actors manipulate as substitutes for regulation, enabling "ethics washing". Stahl, Rodrigues, Santiago and Macnish (2022) similarly identify that ethical guidelines lack "democratic legitimacy and binding nature". Without enforcement mechanisms such as financial penalties, criminal liability or regulatory sanctions, ethical frameworks remain simply aspirational.

Kieslich, Keller and Starke (2022) show through empirical research involving 1,099 German citizens that public perception demands accountability as the most important ethical principle, followed by fairness, security, privacy and accuracy. However, 24.3% remained "Indifferent" to ethical design, "not affected by compliance with ethical principles" (Kieslich, Keller and Starke, 2022). Without public demand for ethical standards, market forces alone fail to drive compliance.

Recommended Course of Action

This author proposes a hybrid governance model combining regulatory certainty with innovation flexibility, crucially incorporating mandatory enforcement mechanisms. The framework should establish internationally harmonised core principles through multilateral agreement whilst permitting regional implementation flexibility (World Economic Forum, 2024; Stahl, Rodrigues, Santiago and Macnish, 2022).

Firstly, mandatory algorithmic impact assessments must happen before high-risk deployment, with results publicly disclosed and subject to independent audit (European Commission, 2024; World Economic Forum, 2024). Organisations failing to conduct proper assessments should face substantial financial penalties. Kieslich, Keller and Starke (2022) note that "achieving fairness in AI systems is very costly" and "the market does not reward putting a massive amount of money into collecting data of marginalised groups". Without regulatory mandates, profit-focused firms lack motivation for ethical investment.

Secondly, human oversight mechanisms must be embedded within important AI decisions, with clear liability frameworks (World Economic Forum, 2024; Kieslich, Keller and Starke, 2022). Current ambiguity enables organisations to deflect responsibility. Stahl, Rodrigues, Santiago and Macnish (2022) recommend establishing "benchmarks for enforcement" through regulatory opinions or intervention advisories.

Thirdly, regular independent audits by accredited external bodies must assess compliance, with findings submitted to authorities possessing enforcement powers (European Commission, 2024; Stahl, Rodrigues, Santiago and Macnish, 2022). Fourthly, strong whistleblowing protections must extend beyond existing frameworks (Whistleblowing Network, 2024). Finn and Shilton (2023) document how researchers resigned when controversial research emerged, highlighting practitioner vulnerability.

Finally, an international AI governance coordinating body must facilitate information sharing and harmonise standards with binding authority to impose sanctions for non-compliance (World Economic Forum, 2024; Stahl, Rodrigues, Santiago and Macnish, 2022). Stahl, Rodrigues, Santiago and Macnish (2022) propose a European Agency for AI with powers to "make recommendations for legislative amendments", "identify red lines for AI development", and "maintain an AI risk alert system".

Professional bodies like the ACM and BCS must go beyond current limitations. Resseguier and Rodrigues (2020) argue that ethics must recover its "teeth". Professional standards should establish minimum competence thresholds with membership revocation for violations (Association for Computing Machinery, 2018; British Computer Society, 2022).

Critical Impact Analysis

The proposed framework carries significant legal implications. Extraterritorial jurisdiction claims create compliance complexity (European Commission, 2024; Stahl, Rodrigues, Santiago and Macnish, 2022). Liability frameworks must clarify responsibility allocation when AI causes harm (World Economic Forum, 2024). Stahl, Rodrigues, Santiago and Macnish (2022) emphasise that many ethical issues "either have human rights implications or can be described as human rights infringements, which would suggest that legal remedies are appropriate".

From a social perspective, algorithmic bias reinforces societal inequalities (Leslie, 2020; Kieslich, Keller and Starke, 2022). Kieslich, Keller and Starke (2022) document how the COMPAS algorithm "systematically disadvantaged black defendants" whilst Amazon's hiring algorithm "discriminated against female candidates". The digital divide threatens to widen as wealthy populations access AI benefits whilst marginalised communities bear greater exposure to harms (World Economic Forum, 2024). Stahl, Rodrigues, Santiago and Macnish (2022) note concerns about "worker surveillance and generalised exploitation". Public trust depends critically on transparency. Kieslich, Keller and Starke (2022) identify the "black box" nature undermining accountability, whilst noting that "too much transparency can impair user experience".

Professional implications demand enhanced ethical competence beyond technical skills (Dilmegani, 2023; Stahl, Rodrigues, Santiago and Macnish, 2022). Professional accountability extends beyond contractual obligations (Association for Computing

Machinery, 2018). However, the current system fails computing professionals. Finn and Shilton (2023) document how researchers faced ethical uncertainties with inadequate institutional support. Kieslich, Keller and Starke (2022) note that "ethical considerations are often left out of software development". Whistleblowing protections remain inadequate, computing professionals who prioritise public good risk employment security without robust legal safeguards.

Conclusion

The governance challenge requires practical solutions balancing competing interests. However, as Resseguier and Rodrigues (2020) argue, AI ethics must recover its enforcement capacity. Kieslich, Keller and Starke (2022) show empirically that whilst public accountability demands exist, substantial populations remain "Indifferent". Stahl, Rodrigues, Santiago and Macnish (2022) conclude that new regulatory bodies will "form part of the regulatory AI ecosystem" by providing "flexible approach" and enabling "exchange of knowledge". The proposed hybrid framework combines regulatory certainty, innovation flexibility and international coordination, but critically mandates enforceable accountability mechanisms. However, geopolitical tensions nearly make it impossible to establish an international AI ethics body since competitive pressures between states may lead to underinvestment in safe, ethical AI development, with every country wanting to win the AI race (ÓhÉigeartaigh *et al.*, 2020). Without enforceable accountability, generative AI governance risks remaining merely aspirational, allowing powerful actors to pursue commercial interests whilst placing harms onto vulnerable populations, an ethically unacceptable and socially unsustainable outcome.

Word count: 1,025

References

Association for Computing Machinery (2018) *ACM Code of Ethics and Professional Conduct*. Available at: https://www.acm.org/code-of-ethics (Accessed: 26 October 2025).

British Computer Society (2022) *Code of Conduct for BCS Members*. Available at: https://www.bcs.org/membership/become-a-member/bcs-code-of-conduct/ (Accessed: 26 October 2025).

Corrêa, N.K., Galvão, C., Santos, J.W., Del Pino, C., Pinto, E.P., Barbosa, C., Massmann, D., Mambrini, R., Galvão, L., Terem, E. and Oliveira, N. (2023) 'Worldwide AI Ethics: A Review of 200 Guidelines and Recommendations for AI Governance', *Patterns*, 4(10), p. 100857. Available at: https://doi.org/10.1016/j.patter.2023.100857

Dilmegani, C. (2023) *Generative AI Ethics: Concerns and How to Manage Them*. Available at: https://research.aimultiple.com/generative-ai-ethics/ (Accessed: 26 October 2025).

European Commission (2024) *AI Act*. Available at: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai (Accessed: 26 October 2025).

Finn, M. and Shilton, K. (2023) 'Ethics Governance Development: The Case of the Menlo Report', *Social Studies of Science*, 53(3), pp. 315-340. Available at: https://doi.org/10.1177/03063127231151708

Kieslich, K., Keller, B. and Starke, C. (2022) 'Artificial Intelligence Ethics by Design: Evaluating Public Perception on the Importance of Ethical Design Principles of Artificial Intelligence', *Big Data & Society*, 9(1), pp. 1-15. Available at: https://doi.org/10.1177/20539517221092956

Leslie, D. (2020) 'Understanding Bias in Facial Recognition Technologies', *The Alan Turing Institute*, 1, pp. 1-50. Available at: https://doi.org/10.5281/zenodo.4050457

Lorica, B. (2025) *AI Governance Cheat Sheet: Comparing Regulatory Frameworks Across the EU, US, UK and China*. Available at: https://gradientflow.com/ai-governance-global-cheat-sheet/ (Accessed: 26 October 2025).

ÓhÉigeartaigh, S.S., Whittlestone, J., Liu, Y., Zeng, Y. and Liu, Z. (2020) 'Overcoming Barriers to Cross-cultural Cooperation in AI Ethics and Governance', *Philosophy & Technology*, 33, pp. 571-593. Available at: https://doi.org/10.1007/s13347-020-00402-x

Resseguier, A. and Rodrigues, R. (2020) 'AI Ethics Should Not Remain Toothless! A Call to Bring Back the Teeth of Ethics', *Big Data & Society*, 7(2), pp. 1-5. Available at: https://doi.org/10.1177/2053951720942541

Stahl, B.C., Rodrigues, R., Santiago, N. and Macnish, K. (2022) 'A European Agency for Artificial Intelligence: Protecting Fundamental Rights and Ethical Values', *Computer Law & Security Review*, 45, pp. 1-25. Available at: https://doi.org/10.1016/j.clsr.2022.105661

Takhar, J. (2023) *UK vs EU Approach to Regulating AI: From One Extreme to Another*. Available at: https://healthcarelifesciences.bakermckenzie.com/2023/04/05/uk-vs-eu-approach-to-regulating-ai-from-one-extreme-to-another/ (Accessed: 26 October 2025).

Whistleblowing Network (2024) *A Tech Workers Guide to Whistleblowing: UK Edition*. Available at: https://whistleblowingnetwork.org (Accessed: 26 October 2025).

World Economic Forum (2024) *AI Governance Alliance Briefing Paper Series*. Available at: https://www.weforum.org (Accessed: 26 October 2025).