

Navigating the global AI ethics landscape: Towards harmonised governance in the generative AI era

Introduction

The proliferation of generative artificial intelligence systems has fundamentally transformed the discourse surrounding AI ethics from theoretical frameworks to urgent governance imperatives. Correa et al.'s (2023) comprehensive meta-analysis of 200 AI governance documents provides critical insights into both the emerging global consensus on ethical principles and the substantial implementation challenges that persist. This analysis, when considered alongside Deckard's practical framework for AI ethics professionals, reveals the complexity of establishing effective governance mechanisms in an era of rapid technological advancement. This reflection examines these scholarly contributions to propose a coherent approach to addressing the multifaceted challenges posed by generative AI technologies.

The contemporary governance landscape: convergence and fragmentation

Correa et al.'s systematic analysis reveals significant convergence around core ethical principles, with transparency (86%), justice/equity (81%), non-maleficence (71%), accountability (71%), and privacy (56%) emerging as the most frequently cited concerns across global governance frameworks. However, this apparent consensus masks deeper structural challenges that have become increasingly pronounced with the advent of generative AI systems.

The geographical distribution of governance documents presents a concerning pattern of Western hegemony, with 77% of analysed documents originating from merely 13 countries, predominantly the United States, United Kingdom, and Germany. This concentration creates what may be characterised as a normative imperialism in AI ethics discourse, particularly problematic given that generative AI systems operate across diverse cultural, linguistic, and socioeconomic contexts whilst remaining governed by predominantly Anglo-European ethical frameworks.

Perhaps more critically, Correa et al. identify a substantial implementation deficit, with 98% of analysed documents proposing only legally non-binding guidelines whilst a mere 2% offer concrete implementation mechanisms. This phenomenon, which Mittelstadt (2019) terms "ethics washing," becomes particularly acute when addressing the immediate societal implications of generative AI, including deepfake proliferation, systematic misinformation campaigns, intellectual property violations, and widespread labour displacement.

Professional and technical imperatives

Deckard's emphasis on interdisciplinary collaboration proves particularly prescient given the multifaceted nature of generative AI challenges. The technology necessitates professionals capable of navigating complex technical architectures, evolving legal frameworks, diverse social implications, and significant economic disruptions simultaneously. However, Correa et al.'s demographic analysis reveals concerning patterns, with 66% of AI ethics documents lacking identifiable authorship and the remaining demonstrating significant gender imbalances, suggesting fundamental diversity deficits in governance development processes.

The "AI ethics boom" that Correa et al. documented in 2018 has evolved into what might be characterised as an implementation crisis with generative AI deployment. Whilst principles such as transparency remain paramount, their practical application to large language models with billions of parameters presents unprecedented technical challenges. Traditional explainability methodologies prove inadequate for understanding emergent behaviours in systems trained on internet-scale datasets, necessitating novel approaches to algorithmic accountability.

Legal and regulatory considerations

The regulatory fragmentation documented by Correa et al. has intensified with generative AI's rapid commercial deployment. The European Union's AI Act, the United Kingdom's principles-based regulatory approach, and China's algorithmic accountability measures represent fundamentally divergent regulatory philosophies that create substantial compliance complexities for globally deployed AI systems.

The implications for data protection frameworks, particularly the General Data Protection Regulation (GDPR), prove particularly severe. Generative AI systems' training methodologies, which utilise vast personal datasets without explicit consent, fundamentally violate core GDPR principles including purpose limitation and data minimisation. The practical implementation of rights such as data erasure becomes technically infeasible when personal information becomes embedded within model parameters across distributed computational infrastructure.

Societal and ethical implications

Correa et al.'s finding that merely 1.5% of analysed documents address long-term AI risks proves increasingly concerning given generative AI's rapid capability advancement. The predominant focus on short-term considerations (47%) inadequately addresses systemic risks including widespread misinformation campaigns, economic displacement, and potential pathways toward artificial general intelligence development.

The labour rights principle, mentioned in only a small fraction of documents, assumes critical importance as generative AI systems increasingly threaten knowledge-based occupations across creative industries, journalism, education, and professional services. Existing ethical frameworks, developed primarily to address blue-collar automation, prove insufficient for addressing cognitive labour displacement and its broader socioeconomic implications.

Proposed governance framework

Establishment of an International AI Ethics Coordination Mechanism

Building upon Correa et al.'s identification of principle convergence whilst addressing the documented geographical bias, I propose establishing an International AI Ethics Coordination Council operating under United Nations auspices (similar to the remit of the International Conference on Harmonisation for clinical trials). This mechanism would:

- Harmonise core ethical principles whilst accommodating legitimate cultural variations
- Develop binding implementation standards that transcend current soft law approaches

- Coordinate regulatory frameworks to prevent harmful fragmentation
- Ensure meaningful representation from Global South perspectives to address the documented Western bias

Implementation of mandatory algorithmic impact assessments

Responding to Deckard's emphasis on practical solutions and addressing the implementation deficit identified by Correa et al., generative AI deployments should require comprehensive Algorithmic Impact Assessments analogous to environmental impact evaluations (and Privacy Impact Assessments). These assessments would:

- Evaluate potential societal implications prior to system deployment
- Mandate comprehensive transparency reporting for high-risk applications
- Establish ongoing monitoring requirements for system behaviour evolution
- Create clear liability frameworks for AI-generated societal harms

Development of technical standards for responsible AI

Addressing the implementation gap documented by Correa et al. requires:

- Standardised evaluation metrics for bias detection, fairness assessment, and safety validation
- Technical specifications for explainability and auditability requirements
- Certification processes for high-risk AI applications
- Open-source compliance monitoring tools accessible to diverse stakeholders

Professional certification in AI ethics

Following Deckard's career development framework, establishing professional certification programmes would:

- Integrate technical expertise with ethical competency through interdisciplinary curricula
- Mandate continuing professional development given rapid technological evolution
- Ensure diverse representation in ethics professional development
- Create accountability mechanisms for certified practitioners

Impact assessment and implications

Legal implications

This framework would establish harmonised international standards whilst preserving national regulatory sovereignty. Mandatory impact assessments would create clear liability frameworks and establish due diligence requirements for AI developers and deployers, addressing current regulatory uncertainty.

Social implications

Global coordination mechanisms would ensure equitable representation in AI governance processes, directly addressing the geographical bias documented by Correa et al. Professional certification would enhance public trust through accountable expertise, whilst technical standards would provide measurable protection against AI-related societal harms.

Professional implications

The proposed framework would create substantive career pathways in AI ethics whilst establishing clear professional standards. Interdisciplinary requirements would elevate the field beyond purely technical considerations, incorporating social science, legal scholarship, and philosophical analysis as advocated by Deckard.

Conclusion

Correa et al.'s analysis demonstrates that whilst global consensus exists regarding fundamental AI ethics principles, the implementation crisis has intensified with generative AI's emergence. The path forward necessitates transcending aspirational principles toward enforceable standards, technical implementations, and effective global coordination mechanisms.

The generative AI revolution demands transforming AI ethics from an academic discipline into a practical governance framework. This transformation requires the interdisciplinary expertise championed by Deckard, the global perspective advocated by Correa et al., and the technical rigour demanded by emerging AI capabilities. Only through such comprehensive approaches can we ensure that AI development serves human flourishing rather than merely technological advancement.

The implications of contemporary governance decisions extend far beyond immediate technological considerations, fundamentally shaping the technological landscape for future generations. The urgency of establishing effective governance mechanisms cannot be overstated.

References:

Correa, N.K., Galvão, C., Santos, J.W., et al. (2023). Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. *Patterns*, 4(10), 100857.

Deckard, R. (2024). Building a career in AI ethics: A multidisciplinary approach. *BCS Professional Development Series*.

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1(11), 501-507.