The Indian Journal for Research in Law and Management
Open Access Law Journal – Copyright © 2026
Editor-in-Chief – Dr. Muktai Deb Chavan; Publisher – Alden Vas; ISSN: 2583-9896

This is an Open Access article distributed under the terms of the Creative Commons Attribution- Non-Commercial-Share Alike 4.0 International (CC-BY-NC-SA 4.0) License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium provided the original work is properly cited.

RESPONSIBLE AI - A LITERATURE REVIEW

AKANKSHA

ABSTRACT

This article reviews the principal academic and institutional contributions to artificial intelligence governance scholarship, incorporating regulatory developments through April 2026. Drawing on peer-reviewed literature in law, political science, ethics, and information systems - and on international instruments from the United Nations, the OECD, and the European Union - the article maps five themes: the ethical principle landscape and its limits; accountability and transparency mechanisms; the governance of algorithmic bias; the comparative regulatory landscape, updated to reflect post-2025 amendments and new national laws; and the governance of generative AI. The article concludes that despite nominal convergence on principle -level commitments, the governance architecture remains fragmented, enforcement-thin, and increasingly pulled in opposite directions by a deregulatory United States and an enforcement-oriented European Union.

KEYWORDS - RESPONSIBLE AI , AI GOVERNANCE FRAMEWORKS ,GENERATIVE AI , EU ARTIFICIAL INTELLIGENCE ACT , INDIA AI SUMMIT , DIGITAL GOVERNANCE.

I. INTRODUCTION

Artificial intelligence now makes decisions - about creditworthiness, parole eligibility, and the content that reaches billions of screens - once reserved for human institutions. The governance question is not abstract. It concerns who controls consequential automated systems, under what rules, and with what remedies for those harmed. The scholarly response has been substantial but uneven, and the regulatory landscape has shifted faster than any prior review has been able to track. This article reviews that literature and updates it to reflect developments through April 2026, a period in which several foundational assumptions of the governance field have been tested by political reversals, legislative proliferation, and the maturation of generative AI as a distinct regulatory object.

The article proceeds in six parts. Part II surveys the foundational debate on ethics principles and their limits. Part III addresses accountability and transparency. Part IV examines algorithmic bias. Part V surveys the comparative regulatory landscape - the EU, the United States, China, and international frameworks - incorporating amendments and new laws enacted since August 2025. Part VI addresses generative AI governance. Part VII offers a synthesis.

II. FROM PRINCIPLES TO GOVERNANCE: THE FOUNDATIONAL DEBATE

The modern AI governance literature begins with the proliferation of ethics guidelines in the second half of the 2010s. The most widely cited study of this phenomenon was conducted by Anna Jobin, Marcello Ienca, and Effy Vayena, who mapped eighty-four AI ethics guidelines issued by companies, research institutions, and public bodies across the globe.[1] Their central finding - that a global convergence is emerging around five core principles (transparency, justice and fairness, non-maleficence, responsibility, and privacy) - anchored subsequent debate.[1] What they also found was that the apparent consensus concealed deep divergence: the five principles were interpreted differently, applied to different actors, and accompanied by radically different implementation strategies.

Brent Mittelstadt's influential essay argued that principles alone cannot guarantee ethical AI, comparing the field unfavourably to medical ethics, which

achieved normative force only after decades of institutional investment and legally binding instruments.[2] Thilo Hagendorff's evaluation of thirty-two ethics guidelines confirmed the pattern: the most technically tractable issues received concrete guidance, while the ethically weightiest - fairness, accountability, societal impact - were least likely to be operationalised.[3] These critiques set the agenda for the governance turn that followed, driving scholars toward regulatory theory, comparative law, and the study of international organisations.

III. ACCOUNTABILITY, TRANSPARENCY, AND THE LIMITS OF EXPLAINABILITY

Transparency and accountability are the two principles most consistently cited across AI governance frameworks, and among the most contested in implementation. Sandra Wachter, Brent Mittelstadt, and Luciano Floridi argued that the explanations produced by post-hoc interpretability tools frequently fail to provide affected individuals with information adequate to understand or contest consequential decisions.[4] Joy Buolamwini and Timnit Gebru's empirical study of commercial facial recognition systems demonstrated how algorithmic auditing could expose systematic performance disparities across gender and skin type - disparities that developers had not disclosed and regulators had not detected.[5]

The systematic literature review published in *AI and Ethics* in 2025 synthesised accountability scholarship across twenty-eight primary studies, organising the field around four questions: who is accountable, what elements are subject to governance, when governance occurs in the AI lifecycle, and how it is implemented.[6] Papagiannidis, Mikalef, and Conboy's research framework extended this analysis, arguing that governance must be understood as a sociotechnical enterprise requiring changes not only to AI system design but to the organisational structures and incentive systems of deploying institutions.[7]

IV. ALGORITHMIC BIAS, FAIRNESS, AND THE GOVERNANCE OF DISCRIMINATION

The mathematical literature on fairness has established that several intuitive definitions - calibration, equal false positive rates, equal false negative rates - are mutually incompatible in all but trivial cases.[8] This impossibility result carries significant governance implications: the choice of a fairness criterion is a political

decision, reflecting judgments about whose errors are tolerable and which conception of equality should prevail. Solon Barocas and Andrew Selbst's analysis of how disparate impact doctrine applies to machine learning systems has been particularly influential, demonstrating that existing anti-discrimination law is at least theoretically equipped to address many instances of algorithmic discrimination without requiring new legislation.[9]

The global governance literature has added a further complication: Mona Sloane and colleagues have argued that AI bias research and governance have been dominated by concerns arising from the specific social context of the United States, and that frameworks developed there may not travel well to jurisdictions with different social cleavages, legal traditions, and configurations of state and market power.[10] This observation carries particular force as AI deployment accelerates across the global South, where governance frameworks are thinnest and affected populations least represented in the scholarly literature.

V. THE COMPARATIVE REGULATORY LANDSCAPE

A. European Union: A Framework Under Amendment

The EU Artificial Intelligence Act - Regulation (EU) 2024/1689 - entered into force on August 1, 2024, and remains the most ambitious binding AI regulatory instrument in the world.[11] Its risk-based architecture classifies AI systems into four tiers: prohibited, high-risk, limited-risk, and minimal-risk, with conformity assessment and post-market monitoring obligations attaching to the most consequential applications.[11] The GPAI provisions, governing general-purpose AI models, became applicable in August 2025, requiring providers of the most capable models to conduct systematic risk assessments and publish summaries of training data.

The article's earlier characterisation of the Act as a settled framework must now be qualified. In November 2025, the European Commission published a Digital Omnibus legislative proposal seeking to amend the AI Act and the GDPR simultaneously - deferring the date of applicability of high-risk AI rules, allowing GPAI providers additional time to update documentation, and narrowing the definition of personal data to ease AI training.[12] These proposals, which must be approved by the

European Parliament, reflect a significant recalibration: the EU is now attempting to pair enforcement maturity with economic competitiveness, acknowledging that its regulatory ambition may be constraining the innovation it also seeks to lead. Italy enacted a national AI law on October 10, 2025, closely aligned with the Act but containing additional protections for minors, illustrating the member-state variation that will complicate uniform enforcement.[13]

B. United States: Federal Retreat and State-Level Proliferation

Executive Order No. 14,110, which established a comprehensive federal AI governance framework in October 2023, was revoked by Executive Order No. 14,179 in January 2025.[14] In December 2025, a further Executive Order established a national policy framework explicitly designed to preempt state AI laws that conflict with a minimally burdensome federal approach, directing the Secretary of Commerce to evaluate state laws that require AI models to alter their outputs or compel disclosures that could raise First Amendment concerns.[15] The constitutional and political viability of that preemption claim remains contested.

Despite federal retreat, state-level activity has accelerated sharply. California enacted training-data transparency requirements for generative AI developers under AB 2013, mandating disclosure of dataset composition, intellectual property content, and licensing arrangements.[16] New York's RAISE Act, signed in December 2025, and Colorado's AI Act, amended in August 2025 to delay implementation to June 2026, have extended high-risk AI obligations to developers and deployers operating in those states.[16] The NIST AI Risk Management Framework (NIST AI 100-1) retains its role as the principal voluntary federal standard, but the widening gap between voluntary federal guidance and binding state law creates compliance complexity that the governance literature has not yet fully analysed.[17]

C. China and Asia-Pacific: Layered Enforcement Deepens

China's AI regulatory architecture has continued to deepen. The Measures for Labelling AI-Generated and Synthetic Content came into effect in September 2025, requiring platforms to implement detection mechanisms including encrypted metadata and watermarking systems.[18] An amended Cybersecurity Law, explicitly

referencing AI, became enforceable on January 1, 2026, adding requirements for security reviews and data localisation of AI systems.[18] Japan's Act on the Promotion of Research and Development and Utilisation of AI-Related Technologies, enacted in May 2025 and effective in June 2025, establishes a non-binding coordination framework - consistent with Japan's established preference for voluntary self-regulation over prescriptive mandates.[19] South Korea's Basic AI Act entered into force in January 2026, adopting a risk-based classification approach similar to the EU framework and extending obligations to employment, education, and essential services.[20]

D. International Frameworks: The Governance Gap Widens

The international AI governance architecture rests on the OECD Recommendation on Artificial Intelligence - updated in 2024 and now endorsed by forty-seven jurisdictions[21] - the UNESCO Recommendation on the Ethics of Artificial Intelligence, endorsed by all 194 member states[22], and the Council of Europe Framework Convention on Artificial Intelligence, the first legally binding international AI treaty, opened for signature in September 2024.[23] Matthijs Maas's comprehensive literature review of advanced AI governance documented the full range of institutional proposals - from expanding existing international bodies to creating a dedicated intergovernmental AI agency - and assessed the political obstacles facing each.[24]

The governance gap identified by the UN High-Level Advisory Body - that 118 countries were not party to any significant international AI governance initiative as of 2024[25] - has not closed. By early 2026, over seventy-two countries have launched more than one thousand AI policy initiatives[26], but the distribution remains skewed: Brazil, Korea, Kazakhstan, and Vietnam have adopted risk-based AI laws[26], while most of the global South remains without binding AI legislation and without meaningful participation in the frameworks that shape global AI norms. The ITU's AI Governance Day Report observed that most countries still lack the technical capacity to enforce the obligations they have nominally adopted.[27]

VI. THE GOVERNANCE OF GENERATIVE ARTIFICIAL INTELLIGENCE

Generative AI presents governance challenges that earlier frameworks, designed with narrower applications in mind, are poorly equipped to address. Taeihagh and colleagues identified four categories of concern requiring dedicated regulatory attention: synthetic media and information integrity; copyright and intellectual property; amplification of bias and discrimination; and systemic risk arising from the concentration of generative AI development in a small number of large corporations.[28] Bender and colleagues' foundational analysis of large language models established the core technical basis for these concerns, demonstrating how models trained on unrepresentative corpora reproduce and amplify the biases embedded in their training data.[29]

The regulatory response to generative AI has been rapid but uneven. The EU AI Act's GPAI provisions - now in force - represent the most demanding binding framework yet applied to foundation models, requiring transparency, copyright compliance, and safety evaluation. China's layered approach, adding watermarking and content labelling obligations in September 2025, reflects a distinct priority: controlling the information environment rather than protecting individual rights. The United States, at the federal level, has largely declined to impose binding obligations on generative AI developers, relying instead on voluntary commitments and the emerging state-level patchwork. The governance literature has not yet developed an adequate account of how these divergent national approaches will interact - whether they will produce regulatory arbitrage, convergence around the most demanding standard, or a fragmented global ecosystem in which the terms of AI access are determined by regulatory geography.

VII. SYNTHESIS

Several conclusions emerge clearly from this review. First, the gap between ethics principles and enforceable governance obligations remains structural and persistent - it reflects the absence of enforcement mechanisms, asymmetric resources between developers and regulators, and the incentive structures of a competitive global industry. Second, the post-2025 period has produced the most rapid proliferation of AI laws in history, but also the sharpest divergence in regulatory philosophy: the EU is tightening implementation while simultaneously seeking to

lighten its compliance burden; the United States is actively retreating from federal governance while states rush to fill the vacuum; and China is deepening technical enforcement without the rights-protective orientation that animates Western frameworks.

Third, the enforcement question remains critically under-theorised. The EU AI Act's national market surveillance authorities are widely regarded as under-resourced for the task ahead. No binding international enforcement mechanism exists. Fourth, and most urgently, the geographic distribution of governance remains profoundly unequal. The populations most exposed to AI-related harm - in healthcare, in criminal justice, in public administration across the global South - are least represented in the governance frameworks designed to protect them. Birkstedt and colleagues identified this geographic bias as among the field's most significant knowledge gaps[30], and nothing in the developments of the past year has narrowed it. Effective AI governance will require not merely technical standards and legal instruments, but a long-overdue reckoning with who bears the costs of AI-related harm and who has the power to define the terms on which that question is answered.

REFERENCES

1. Anna Jobin, Marcello Ienca & Effy Vayena, *The Global Landscape of AI Ethics Guidelines,* 1 *Nature Mach. Intelligence* 389 (2019), https://doi.org/10.1038/s42256-019-0088-2.
2. Brent Mittelstadt, *Principles Alone Cannot Guarantee Ethical AI,* 1 *Nature Mach. Intelligence* 501 (2019), https://doi.org/10.1038/s42256-019-0174-5.
3. Thilo Hagendorff, *The Ethics of AI Ethics: An Evaluation of Guidelines,* 30 *Minds & Machines* 99 (2020), https://doi.org/10.1007/s11023-020-09517-8.
4. Sandra Wachter, Brent Mittelstadt & Luciano Floridi, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation,* 7 *Int'l Data Privacy L.* 76 (2017), https://doi.org/10.1093/idpl/ipx005.
5. Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification,* 81 *Proc. Mach. Learning Rsch.* 77 (2018), http://proceedings.mlr.press/v81/buolamwini18a.html.
6. Amna Batool et al., *AI Governance: A Systematic Literature Review, AI & Ethics*

(Jan. 14, 2025), https://link.springer.com/article/10.1007/s43681-024-00653-w.

7. Emmanouil Papagiannidis, Patrick Mikalef & Kieran Conboy, *Responsible Artificial Intelligence Governance: A Review and Research Framework,* 34 *J. Strategic Info. Sys.* 101885 (2025), https://doi.org/10.1016/j.jsis.2024.101885.

8. Alexandra Chouldechova, *Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments,* 5 *Big Data* 153 (2017).

9. Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact,* 104 *Cal. L. Rev.* 671 (2016), https://doi.org/10.15779/Z38BG31.

10. Mona Sloane et al., *Participation Is Not a Design Fix for Machine Learning, Equity & Access in Algorithms, Mechanisms, & Optimization* (2022), https://doi.org/10.1145/3551624.3555290.

11. 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), 2024 O.J. (L 1689) 1.

12. European Commission, Digital Omnibus Legislative Proposals (Nov. 19, 2025); *see* Morgan Lewis, *The New Rules of AI: A Global Legal Overview* (Dec. 22, 2025), https://www.morganlewis.com/pubs/2025/12/the-new-rules-of-ai-a-global-legal-overview.

13. Wilson Sonsini, *2026 Year in Preview: AI Regulatory Developments for Companies to Watch Out For* (Jan. 13, 2026), https://www.wsgr.com/en/insights/2026-year-in-preview-ai-regulatory-developments-for-companies-to-watch-out-for.html.

14. Exec. Order No. 14,110, 88 Fed. Reg. 75,191 (Oct. 30, 2023) (revoked by Exec. Order No. 14,179, 90 Fed. Reg. 8,741 (Jan. 23, 2025)).

15. Exec. Order on Ensuring a National Policy Framework for Artificial Intelligence (Dec. 11, 2025), https://www.whitehouse.gov/presidential-actions/2025/12/eliminating-state-law-obstruction-of-national-artificial-intelligence-policy/.

16. Cal. AB 2013 (2024); *see* Wilson Sonsini, *supra* note 13.

17. Nat'l Inst. of Standards & Tech., U.S. Dep't of Commerce, *AI Risk Management Framework,* NIST AI 100-1 (Jan. 2023), https://doi.org/10.6028/NIST.AI.100-1.

18. Cyberspace Admin. of China, Measures for Labelling AI-Generated and Synthetic Content (eff. Sept. 2025); Amended Cybersecurity Law (eff. Jan. 1, 2026); *see* GDPR Local, *AI Regulations Around the World: Everything You Need to Know in 2026* (Jan. 28, 2026), https://gdprlocal.com/ai-regulations-around-the-world/.

19. Act on the Promotion of Research and Development and Utilisation of AI-Related Technologies (Japan, eff. June 2025).

20. Basic AI Act (South Korea, eff. Jan. 2026); *see* OneTrust, *Where AI Regulation Is Heading in 2026: A Global Outlook* (2026),

https://www.onetrust.com/blog/where-ai-regulation-is-heading-in-2026-a-global-outlook/.

[21.] Organisation for Economic Co-operation and Development, *OECD Principles on Artificial Intelligence,* OECD/LEGAL/0449 (1st ed. 2019, updated 2024), https://oecd.ai/en/dashboards/policy-initiatives/oecd-ai-principles-9705.

[22.] U.N. Educ., Sci. & Cultural Org., *Recommendation on the Ethics of Artificial Intelligence,* SHS/BIO/PI/2021/1 (Nov. 23, 2021), https://www.unesco.org/en/artificial-intelligence/recommendation-ethics.

[23.] Council of Europe, Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, C.E.T.S. No. 225 (opened for signature Sept. 5, 2024).

[24.] Matthijs M. Maas, *Advanced AI Governance: A Literature Review of Problems, Options, and Proposals,* AI Foundations Rep. 4 (Nov. 10, 2023), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4629460.

[25.] U.N. High-Level Advisory Body on Artificial Intelligence, *Governing AI for Humanity: Final Report* 12 (Sept. 2024), https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf.

[26.] GDPR Local, *supra* note 18; Holistic AI, *AI Regulation in 2026: Navigating an Uncertain Landscape* (Jan. 19, 2026), https://www.holisticai.com/blog/ai-regulation-in-2026-navigating-an-uncertain-landscape.

[27.] Int'l Telecomm. Union, *Key Findings on the State of Global AI Governance* (July 2024), https://www.itu.int/hub/2024/07/key-findings-on-the-state-of-global-ai-governance/.

[28.] Araz Taeihagh et al., *Governance of Generative AI,* 44 *Pol'y & Soc'y* 1 (2025), https://academic.oup.com/policyandsociety/article/44/1/1/7997395.

[29.] Emily M. Bender et al., *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?,* FAccT '21: Proc. 2021 ACM Conf. on Fairness, Accountability & Transparency 610 (2021), https://doi.org/10.1145/3442188.3445922.

[30.] Tiina Birkstedt et al., *AI Governance: Themes, Knowledge Gaps and Future Agendas,* 33 *Internet Rsch.* 133 (2023), https://doi.org/10.1108/INTR-01-2022-0042.