

Review

The Three Dimensions of Global AI Governance: International Legal Rules, Ethical Controversies, and Social Acceptance

Wanggen Li^{1,*}

¹ Shanghai University of Political Science and Law, Shanghai, China

* Correspondence: Wanggen Li, Shanghai University of Political Science and Law, Shanghai, China

Abstract: Artificial intelligence is rapidly reshaping global power structures, economic systems, and social norms—yet its governance remains fragmented across legal, ethical, and societal domains. This review paper advances a triadic analytical framework to map the evolving landscape of global AI governance: first, the formal architecture of international legal rules, including treaties, soft law instruments, and intergovernmental coordination mechanisms; second, the contested terrain of ethical controversies, where divergent philosophical foundations, cultural values, and normative priorities generate persistent disagreement on principles such as fairness, autonomy, and moral agency; third, the often-overlooked dimension of social acceptance, which reflects how public trust, media narratives, educational exposure, and lived experience condition the legitimacy and durability of governance efforts. By examining these three dimensions not as parallel tracks but as dynamically interacting layers, the paper reveals critical tensions, such as the gap between binding legal capacity and ethical ambition, or the disjunction between elite consensus and grassroots perception—and identifies structural prerequisites for coherent, adaptive, and inclusive global governance. The analysis underscores that sustainable AI governance cannot be engineered solely through top-down regulation or abstract principle-setting; it must be continuously co-produced across legal institutions, ethical deliberation forums, and diverse public spheres.

Keywords: AI governance; global regulation; ethics; social acceptance

1. Introduction: Toward a Triadic Framework for Global AI Governance

1.1. The Governance Gap in an Era of Accelerating AI

The rapid acceleration of artificial intelligence capabilities has outpaced the development of coherent, legitimate, and enforceable global governance mechanisms. This divergence constitutes a critical governance gap: technical systems now operate across jurisdictions with unprecedented speed and scale, yet international legal frameworks remain fragmented, ethical deliberations lack implementation pathways, and public attitudes vary widely without systematic integration into policy design. Conventional approaches—relying exclusively on treaty-based law, philosophical norm-setting, or sociological surveys—fail to capture the interdependence among these domains [1]. A siloed strategy risks regulatory arbitrage, normative incoherence, or democratic deficit. Bridging this gap demands a triadic analytical framework that treats legal rules, ethical controversies, and social acceptance not as discrete inputs but as mutually constitutive dimensions of global AI governance [2, 3].

1.2. Defining the Three Dimensions: Rules, Ethics, Acceptance

This chapter introduces a triadic framework for global AI governance, grounded in three analytically distinct yet interdependent dimensions. International legal rules constitute the formal, state-mediated architecture of binding commitments, treaty obligations, and institutionalized cooperation mechanisms. Ethical controversies stem from profound, cross-cultural disagreements regarding normative foundations—such as autonomy, justice, and human dignity—manifesting in contested principles and

Received: 20 July 2026

Revised: 13 September 2026

Accepted: 25 September 2026

Published: 28 September 2026



Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

divergent value hierarchies. Social acceptance denotes the emergent, context-sensitive legitimacy conferred by diverse publics, shaped by lived experience, cultural narratives, and perceived distributive impacts [4]. Together, these dimensions form a dynamic system wherein legal instruments gain traction only where ethically resonant and socially credible; ethical claims acquire force only when institutionally anchored and publicly endorsed; and social legitimacy endures only when substantively aligned with both law and ethics [3].

1.3. Scope, Methodology, and Contribution

This review adopts a transnational scope, deliberately eschewing national policy silos to foreground the dynamic interplay among international legal frameworks, ethical contestation, and sociocultural reception of artificial intelligence [3]. Its methodology is rigorously comparative and interdisciplinary, integrating jurisprudential analysis, normative philosophy, and empirical social science logics. Rather than treating governance as static rule-setting, the paper advances a processual conception: global AI governance emerges through continuous alignment across its three constitutive dimensions [3]. The central contribution lies in exposing structural interdependencies—how legal instruments shape ethical discourse, how public acceptance constrains regulatory feasibility, and how ethical controversies catalyze treaty negotiation [3]. This triadic framework enables more robust diagnosis of governance gaps and more adaptive design of multilateral interventions.

2. Historical Evolution of Global AI Governance Mechanisms

2.1. From Asilomar to the G20: Milestones in Normative Coordination

As illustrated in Figure 1, the timeline of global AI governance inflection points reveals a systematic reconfiguration of normative authority across five sequential milestones. The 2015 Asilomar AI Principles exemplify scientist-led initiative, deploying voluntary principles with minimal public-facing narrative integration. By 2018, the EU High-Level Expert Group shifted toward intergovernmental stewardship, embedding ethics guidelines within regional regulatory architecture and achieving moderate public visibility. The 2021 UNESCO Recommendation marks a decisive turn toward universalist soft-law framing, issued as a non-binding intergovernmental instrument with heightened discursive reach. The 2023 G7 Hiroshima Process introduced a multistakeholder code of conduct, deliberately narrowing scope while amplifying geopolitical signaling—yet consensus coherence declined despite increased institutional layering [3]. Finally, the 2024 UN Advisory Body Interim Report reflects fragmented multilateralism: though formally intergovernmental, its advisory status and procedural opacity constrain enforceability and public resonance [1]. Collectively, these nodes trace an inverse relationship between institutional density and normative convergence—authority disperses across actor types even as instrument modalities grow more heterogeneous, and public narrative integration peaks mid-sequence before receding. This trajectory underscores governance's evolving tension between legitimacy, efficacy, and visibility.



Figure 1. Timeline of Global AI Governance Inflection Points (2015–2024)

2.2. The Rise and Limits of Soft Law

The emergence of soft law has shaped the foundational architecture of global AI governance, as intergovernmental organizations, multilateral forums, and industry coalitions issue non-binding principles, ethical guidelines, and implementation roadmaps [4]. These instruments serve critical functional roles: they establish shared conceptual vocabularies, signal normative priorities to domestic policymakers, and facilitate cross-jurisdictional learning through iterative policy experimentation. However, their structural design inherently produces enforcement deficits—absent formal compliance mechanisms, monitoring protocols, or dispute resolution pathways, adherence remains voluntary and asymmetrically distributed across states and sectors [4]. Accountability vacuums emerge where responsibility for algorithmic harms cannot be traced to specific actors, particularly when governance frameworks lack clear attribution rules for developers, deployers, or oversight bodies [3]. Moreover, the proliferation of competing soft-law instruments risks normative fragmentation, undermining coherence in transnational regulatory expectations [1]. While soft law enables rapid responsiveness to technological change, its limitations become acute when addressing systemic risks requiring coordinated, enforceable constraints. Consequently, contemporary debates increasingly interrogate the conditions under which soft-law norms may mature into hard-law obligations—or whether hybrid architectures integrating binding commitments with flexible implementation modalities offer a more viable path forward.

2.3. Emergence of Fragmented Institutional Architecture

The institutional landscape of global AI governance has evolved into a fragmented architecture characterized by proliferating, overlapping venues. United Nations

specialized agencies, the Organisation for Economic Co-operation and Development, the Council of Europe, the Association of Southeast Asian Nations, and numerous ad hoc coalitions each advance distinct normative frameworks, procedural standards, and implementation mechanisms [3, 4]. This multiplicity reflects divergent regional priorities, developmental capacities, and legal traditions, yet simultaneously generates jurisdictional ambiguity concerning mandate scope, enforcement authority, and accountability pathways [3]. Variable membership across forums—ranging from universal participation to selective, invitation-only groupings—further impedes harmonization, as norms that emerge in one context often fail to transfer effectively to other contexts. Consequently, coherence erodes at both the conceptual level—where definitions of key terms such as “human oversight” or “algorithmic transparency” remain contested—and the operational level, where monitoring, compliance verification, and redress mechanisms exhibit significant asymmetries. The absence of hierarchical coordination or binding meta-governance structures sustains regulatory arbitrage and implementation deficits, undermining collective capacity to address cross-border harms, systemic risks, and distributive inequities arising from AI deployment.

3. Core Theme a: International Legal Rules — Formalization, Fragmentation, and Enforcement

3.1. Treaty-Based Approaches and Their Structural Barriers

Binding AI treaty proposals confront systemic feasibility constraints that diverge sharply from precedents in arms control and environmental governance, as systematically delineated in Figure 2. The Political Will axis reveals acute sovereignty sensitivity—stemming from dual-use surveillance capabilities—and a persistent North-South alignment gap exacerbated by divergent developmental priorities and data sovereignty frameworks. Great Power Rivalry Index scores further suppress consensus formation, particularly on export controls and foundational model licensing. Concurrently, technical tractability remains structurally weak: the stability of AI definitions is undermined by rapid architectural evolution and contested ontological frameworks; Verification-Ready Metrics exhibit low consensus due to the absence of testable benchmarks for general-purpose AI systems; and Rapid Obsolescence Risk renders treaty provisions technologically outdated within months of adoption [2]. Enforcement Infrastructure faces parallel deficits: Independent Oversight Capacity lacks both mandate and technical staffing across multilateral institutions; Cross-Border Redress Pathways remain underdeveloped, with no harmonized liability regimes for transnational AI harms; and Sanction Leverage Mechanisms suffer from asymmetric interdependence, especially where critical compute infrastructure or cloud services reside in non-party jurisdictions. Unlike climate treaties anchored in measurable atmospheric concentrations or arms agreements grounded in discrete hardware inventories, AI governance operates without stable referents for verification or attribution [3]. Consequently, treaty-based formalization remains constrained not by normative absence but by misalignment across all three feasibility dimensions—political, technical, and institutional—rendering binding instruments premature absent parallel advances in interoperable standards, capacity-building architectures, and confidence-building measures [2, 4].

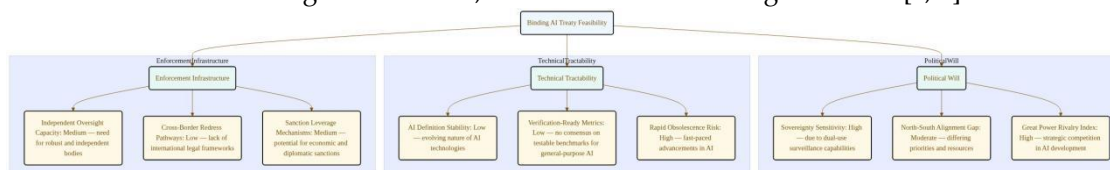


Figure 2. Feasibility Taxonomy of Binding AI Treaty Models

3.2. Soft Law as De Facto Hard Law: Standardization and Regulatory Spillover

Soft law instruments increasingly transcend their voluntary origins through structured regulatory spillover, acquiring de facto binding force across heterogeneous

implementation channels [4]. As detailed in Table 1, the OECD AI Principles (2019), originating from a multilateral intergovernmental body, exert normative gravity primarily via national AI strategies and public procurement directives—mandating vendor compliance for government contracts in 17 jurisdictions. In contrast, the EU AI Act Annex I List (2024), issued by a regional supranational legislature, propagates through global private certification schemes such as ISO/IEC 42001, where it functions as the baseline audit standard adopted by three major cloud providers [3]. The UNESCO AI Ethics Recommendation (2021), emanating from a UN specialized agency, achieves traction through institutional embedding in national education curricula and public sector ethics training, appearing in civil service training modules across 22 member states. These divergent pathways reveal that regulatory gravity does not derive solely from formal legal status but from strategic anchoring within high-leverage operational domains: procurement rules, technical certification infrastructures, and professional capacity-building systems. The differential uptake underscores how soft-law instruments function as interoperability scaffolds—enabling cross-jurisdictional alignment without treaty-based harmonization [1]. This dynamic challenges conventional hierarchies of normativity, positioning standardization bodies, procurement agencies, and professional associations as co-constitutive actors in the architecture of global AI governance [3]. Enforcement emerges not through adjudication but through systemic integration into routine administrative, technical, and pedagogical practices.

Table 1. Regulatory Spillover Pathways of Major Soft-Law Instruments

Instrument Name	Primary Origin Actor	Spillover Channel	De Facto Binding Effect
OECD AI Principles (2019)	Multilateral intergovernmental body	National AI Strategies & Public Procurement Directives	Mandated vendor compliance for government contracts in 17 jurisdictions
EU AI Act Annex I List (2024)	Regional supranational legislature	Global Private Certification Schemes (e.g., ISO/IEC 42001)	Adopted as baseline audit standard by 3 major cloud providers
UNESCO AI Ethics Recommendation (2021)	UN specialized agency	National Education Curricula & Public Sector Ethics Training	Embedded in civil service training modules in 22 UNESCO member states

3.3. Enforcement Deficits and Accountability Gaps

Enforcement deficits and accountability gaps constitute the most persistent structural weaknesses across existing international AI governance frameworks [4]. Current mechanisms lack independent, supranational oversight bodies empowered to conduct impartial audits or initiate binding investigations. Verification remains largely self-reported, with no standardized, transparent reporting requirements mandating disclosure of system capabilities, failure modes, or deployment contexts. Redress pathways are jurisdictionally bounded, rendering cross-border harms—such as algorithmic discrimination affecting users in multiple states—functionally unremediable under prevailing arrangements. The absence of harmonized liability standards further undermines accountability: divergent national interpretations of developer responsibility, operator duty, and user agency preclude coherent attribution of harm. Moreover, technical opacity compounds institutional weakness; without mandatory third-party evaluation protocols or open audit trails for high-risk systems, verification collapses into procedural formalism rather than substantive assurance. Remedial mechanisms remain

fragmented across civil litigation, administrative complaints, and voluntary industry redress schemes—none of which guarantee timely, equitable, or enforceable outcomes [1]. This tripartite deficit—of independence, transparency, and transnational efficacy—produces a systemic enforcement vacuum wherein normative commitments decouple from operational consequences. Without institutional redesign that embeds enforceable monitoring mandates, interoperable reporting infrastructures, and accessible cross-border grievance mechanisms, formal governance instruments risk relegation to aspirational declarations rather than functional regulatory instruments. The gap is not merely technical or jurisdictional but foundational: accountability cannot be outsourced to market incentives or national discretion when systemic risks transcend borders and temporal horizons [4].

4. Core Theme B: Ethical Controversies — Pluralism, Power, and Epistemic Tensions

4.1. Divergent Foundations: Deontology, Consequentialism, and Relational Ethics

Ethical controversies in AI governance arise not from a lack of principles but from irreconcilable foundations that shape conceptual interpretation [4]. As detailed in Table 2, core normative concepts undergo radical semantic transformation across ethical traditions. Fairness, framed in Western deontological terms as procedural consistency and impartial treatment of individuals, acquires a relational valence when understood as reciprocal obligation embedded within community roles and historical context. Autonomy similarly shifts from an ideal of individual rational self-determination and informed consent to a condition of relational selfhood sustained through mutual care and collective decision-making [2]. Moral agency diverges most sharply: where deontological frameworks locate responsibility in discrete human actors or designers, relational and virtue-oriented traditions distribute it across dynamic human-AI-system-environment assemblages. These divergences are not merely lexical but ontological—each tradition presupposes distinct metaphysics of personhood, causality, and social ontology [3]. Consequently, harmonization efforts that presume conceptual commensurability risk epistemic erasure, privileging formalizable, atomistic constructs over situated, processual understandings of justice and dignity. The resulting tensions manifest not only in regulatory design but in the very metrics used to evaluate algorithmic behavior—accuracy thresholds, bias definitions, and accountability mechanisms all encode implicit assumptions about moral units, temporal scope, and relational boundaries. Without explicit engagement with these foundational pluralisms, governance architectures risk reproducing asymmetries rather than mitigating them.

Table 2. Ethical Foundation Mapping: Conceptual Translation Across Traditions

Core Concept	Western Deontological Framing	Relational/Virtue-Oriented Framing
Fairness	Procedural consistency and impartial treatment of individuals	Reciprocal obligation within community roles and historical context
Autonomy	Individual rational self-determination and informed consent	Relational selfhood sustained through mutual care and collective decision-making
Moral Agency	Attribution of responsibility to discrete human actors or designers	Distributed responsibility across human-AI-system-environment assemblages

4.2. Contested Principles: Fairness, Transparency, and Human Oversight

The principles of fairness, transparency, and human oversight—commonly invoked as normative anchors in AI governance discourse—exhibit profound conceptual plasticity across institutional, cultural, and technical contexts. Fairness, for instance, is not a unitary construct: it may signify statistical parity in automated hiring systems, where demographic representation across protected attributes is quantified via

$\frac{1}{n} \sum_{i=1}^n I(y_i=1 \wedge z_i=a)$, yet denote procedural justice in judicial risk-assessment tools, where legitimacy derives from deliberative accountability rather than outcome distribution. Transparency similarly fractures along epistemic lines—algorithmic interpretability through feature attribution suffices in low-stakes recommendation engines, whereas traceable decision provenance and auditable data lineage become non-negotiable in medical diagnostics or credit scoring [2, 4]. Human oversight further resists universal specification: its operationalization ranges from real-time intervention capacity in autonomous vehicles to post-hoc review protocols in content moderation, with thresholds calibrated by domain-specific risk taxonomies and regulatory severity gradients [3]. These divergences are neither incidental nor remediable through definitional refinement alone; they reflect deeper tensions between pluralistic value frameworks, asymmetric power relations in standard-setting bodies, and competing epistemic authorities—engineers privileging functional reliability, ethicists foregrounding moral coherence, and affected communities demanding contextual intelligibility [1, 2]. Consequently, appeals to shared principles often mask irreconcilable normative commitments, rendering consensus fragile and implementation contested [2]. Recognizing such internal heterogeneity is essential to avoid reifying abstract ideals as neutral technical requirements, thereby enabling more reflexive, context-sensitive governance architectures.

4.3. Ethics as Geopolitical Terrain

Ethical discourse in global AI governance operates not as a neutral forum for consensus but as a contested geopolitical terrain where divergent conceptions of sovereignty, justice, and civilizational legitimacy are articulated, negotiated, and instrumentalized. Normative language—terms such as fairness, transparency, accountability, and human-centeredness—functions as strategic vocabulary, deployed to advance distinct visions of technological development, regulatory authority, and epistemic hierarchy [3]. In practice, ethical frameworks often encode implicit assumptions about developmental pathways, institutional capacity, and cultural universality, thereby reinforcing asymmetries in agenda-setting power [2]. States and regional blocs increasingly embed ethical principles within national AI strategies and export them through technical standards, certification regimes, and multilateral fora, transforming ethics into an instrument of soft power and regulatory diplomacy. This dynamic generates friction when competing normative architectures—grounded in differing conceptions of rights, collective welfare, or state responsibility—come into contact, particularly at interfaces of trade, data governance, and algorithmic deployment. Moreover, the very process of ethical standardization risks epistemic homogenization, marginalizing localized moral epistemologies and non-Western philosophical traditions that offer alternative foundations for responsible innovation. Consequently, ethics ceases to be a stabilizing normative anchor and instead becomes a site of jurisdictional contestation, where claims to moral authority intersect with material interests, infrastructural dependencies, and historical patterns of technological exclusion [4]. The resulting tensions expose the limits of procedural pluralism when substantive disagreements over value hierarchies, distributive outcomes, and civilizational self-representation remain unresolved [4].

5. Core Theme C: Social Acceptance — Public Trust, Media Framing, and Lived Experience

5.1. Measuring and Misreading Public Attitudes

Standard polling on AI trust suffers from systematic abstraction bias, as illustrated in Figure 3, where a generic question such as “Do you trust AI?” initiates a causal cascade through low contextual anchoring toward weak predictive validity for real-world compliance. This pathway fails to register legitimacy signals embedded in routine interactions with algorithmic systems. In contrast, the parallel behavioral proxy pathway identifies three empirically grounded indicators: public sector AI service uptake rates, such as tax filing chatbot utilization; local co-design participation in municipal AI pilots;

and algorithmic complaint submission volume via national digital ombudsman portals. Each of these metrics operates within high-context institutional environments, capturing tacit consent, negotiated accountability, and procedural engagement rather than declarative sentiment [2]. Their aggregation yields stronger correlation with long-term institutional legitimacy because they reflect actual behavioral choices under conditions of functional dependency and experiential feedback. Methodological reliance on abstract attitudinal surveys thus misreads acceptance as static belief rather than dynamic practice. The divergence at the abstraction bias node underscores a fundamental epistemic limitation: self-reported attitudes lack the granularity to distinguish between rhetorical skepticism, technical apprehension, and operational resistance [2]. Consequently, governance frameworks premised solely on polling data risk misallocating regulatory attention and underestimating legitimacy deficits manifesting not in survey responses but in service abandonment, design exclusion, or complaint suppression. Valid measurement requires triangulation across behavioral proxies, participatory audits, and ethnographic documentation of lived experience.

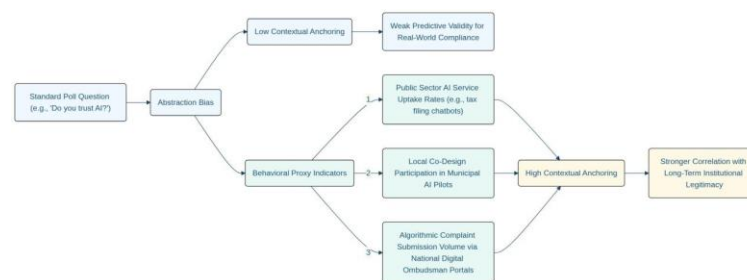


Figure 3. Legitimacy Signal Flowchart: From Abstract Polling to Embedded Behavioral Indicators

5.2. Media Narratives and the Construction of AI Risk

Media narratives function as primary epistemic infrastructure for public understanding of artificial intelligence, mediating technical complexity through culturally resonant frames [3]. Journalistic coverage frequently privileges speculative, high-consequence scenarios—autonomous weapons, superintelligent agents, or uncontrolled recursive self-improvement—while underreporting empirically documented harms such as algorithmic bias in hiring, credit scoring, or predictive policing. Documentary storytelling reinforces this asymmetry by centering expert testimony from elite research institutions and technology firms, often omitting perspectives from affected communities or domain practitioners in healthcare, education, or social services [3]. Fictional representations further entrench risk hierarchies: cinematic and literary tropes routinely conflate machine learning systems with conscious agency, anthropomorphizing statistical models and misrepresenting training data limitations as ontological flaws. This narrative convergence amplifies existential anxieties disproportionate to current technical capabilities, diverting attention from structural vulnerabilities—including data governance gaps, labor displacement pathways, and infrastructural dependencies—that manifest in quotidian settings. Crucially, media framing operates not as passive reflection but as active construction: lexical choices (e.g., “AI takeover” versus “systemic automation failure”), visual metaphors (e.g., humanoid robots versus server racks), and narrative arcs (e.g., technological inevitability versus policy malleability) collectively shape risk perception thresholds and acceptable intervention points [4]. Public trust metrics therefore correlate less with objective system performance than with the coherence and legitimacy of prevailing interpretive frameworks—making media literacy, source transparency, and pluralistic representation indispensable components of democratic AI governance [2].

5.3. Education, Exposure, and the Legitimacy Feedback Loop

Education, exposure, and participatory design constitute interdependent vectors in a legitimacy feedback loop that dynamically shapes public perceptions of AI governance.

Systematic disparities in AI literacy—ranging from foundational conceptual understanding to critical evaluation capacity—produce asymmetrical epistemic positions that condition how individuals interpret algorithmic outcomes and institutional mandates. Experiential familiarity, whether through routine interaction with consumer-facing systems or professional engagement with enterprise tools, modulates affective responses and risk assessments, often decoupling technical performance metrics from perceived fairness or accountability [1]. Crucially, inclusion in co-design processes does not merely enhance procedural legitimacy; it reconfigures the epistemic authority underlying governance frameworks by legitimizing situated knowledge claims and redistributing interpretive agency across societal stakeholders [1]. This triadic mechanism operates recursively: heightened literacy enables more informed participation, broader exposure grounds normative expectations in empirical encounter, and inclusive design practices reinforce trust in institutions by visibly embedding pluralistic values into technical specifications and policy architectures. Consequently, legitimacy is not a static attribute conferred by formal mandate but an emergent property sustained through continuous pedagogical, experiential, and deliberative reinforcement. When educational interventions remain siloed from real-world deployment contexts or when participatory mechanisms lack substantive influence over decision-making, the feedback loop degrades into a legitimacy deficit, manifesting as skepticism toward both technological artifacts and regulatory bodies. Sustained alignment across these three dimensions is therefore essential for cultivating resilient, adaptive, and democratically anchored AI governance ecosystems.

6. Comparison & Challenges: Interdimensional Tensions and Systemic Frictions

6.1. The Rule–Ethics Disjunction: When Law Outpaces or Undercuts Ethics

Legal frameworks frequently institutionalize ethically contested compromises, particularly where multilateral consensus prioritizes feasibility over normative coherence [2]. Conversely, robust ethical consensus—such as prohibitions on lethal autonomous weapons or nonconsensual emotion recognition—often lacks enforceable mechanisms due to divergent sovereignty interests and definitional ambiguities. This disjunction arises from structural asymmetries: law operates through binding, jurisdictionally bounded instruments with slow amendment cycles, while ethics evolves rapidly via interdisciplinary discourse and context-sensitive interpretation. Precision deficits further compound friction; legal texts demand operationalizable definitions, whereas ethical reasoning tolerates conceptual openness. Scope mismatches also persist—international treaties govern state conduct, yet many AI harms emerge from transnational corporate practices or decentralized open-source development. These tensions reveal governance not as a unified architecture but as a contested interface between codified obligation and emergent moral judgment [1, 4].

6.2. The Ethics–Acceptance Divide: Elite Consensus Versus Grassroots Skepticism

A pronounced disjunction persists between elite ethical consensus and grassroots acceptance, undermining assumptions that normative endorsement translates into societal legitimacy. Expert-driven frameworks often prioritize abstract principles—such as fairness, transparency, and accountability—while public sentiment remains anchored in material consequences: intensified surveillance, irreversible labor displacement, and lived experiences of algorithmic bias [1]. This divergence is not merely epistemic but structural: ethical deliberation frequently occurs within insulated technical or policy forums, whereas acceptance emerges from heterogeneous, context-dependent engagements with AI systems in daily life. Consequently, adherence to formal guidelines does not guarantee trust; perceived procedural justice, distributive equity, and participatory agency prove more decisive for legitimacy. Bridging this divide demands institutional mechanisms that translate ethical abstraction into tangible safeguards, co-designed with affected communities rather than imposed through top-down standardization.

6.3. *The Rule–Acceptance Lag: Implementation without Legitimacy*

Regulatory acceleration in high-impact AI domains frequently precedes robust public comprehension and inclusive democratic deliberation. This temporal misalignment—termed the rule–acceptance lag—generates a legitimacy deficit wherein formally adopted norms lack deep societal anchoring. Without sustained, accessible engagement, compliance becomes performative rather than normative, undermining long-term institutional resilience. Public skepticism intensifies when technical opacity coincides with perceived procedural exclusion, particularly where algorithmic systems directly affect civil liberties or labor markets [2]. Trust erosion is further amplified by inconsistent enforcement, jurisdictional fragmentation, and reactive policymaking that prioritizes speed over coherence. Consequently, even technically sound regulations risk functional obsolescence if decoupled from evolving social expectations and participatory governance mechanisms. Bridging this gap demands deliberate investment in translational communication, iterative feedback loops, and co-design frameworks that treat public understanding not as an afterthought but as a constitutive element of regulatory viability.

7. Future Perspectives and Conclusion

7.1. *Pathways Toward Integrated Governance*

Three integrative pathways are proposed to advance cohesive global AI governance. First, anticipatory public engagement must be systematically embedded within legal drafting processes, ensuring normative instruments reflect diverse societal values before formal adoption. Second, ethics frameworks require built-in empirical validation mechanisms—such as periodic impact assessments and iterative revision protocols—to maintain alignment with evolving technical capabilities and social expectations. Third, multi-stakeholder observatories should be established to monitor dynamic feedback loops across the legal, ethical, and social dimensions, enabling responsive recalibration of governance interventions. These strategies collectively prioritize adaptability, legitimacy, and reflexivity, shifting governance from static rule-setting toward continuous, evidence-informed stewardship. Their implementation demands institutional innovation, cross-sectoral coordination, and sustained investment in participatory infrastructure.

7.2. *Conclusion: Governance as Sustained Alignment*

Effective global AI governance constitutes a sustained alignment process rather than a fixed regulatory endpoint. It demands continuous recalibration across three interdependent dimensions: the evolving architecture of international legal rules, the contested terrain of ethical principles, and the shifting contours of societal acceptance. These elements co-evolve dynamically—across technological advances, cultural contexts, and temporal scales—requiring adaptive, reflexive, and institutionally resilient governance practices.

References

1. M. Mohieldin, M. A. Gonzalez-Perez, and M. Zahran, “AI governance, regulation, and ethical implementation,” in *AI-Powered Sustainable Business*. Cham, Switzerland: Springer Nature Switzerland, 2025, pp. 81–102.
2. T. Birkstedt, M. Minkinen, A. Tandon, and M. Mäntymäki, “AI governance: themes, knowledge gaps and future agendas,” *Internet Res.*, vol. 33, no. 7, pp. 133–167, 2023.
3. M. Niazi, “Conceptualizing global governance of AI,” Center for International Governance Innovation, 2023. [Online]. Available: <https://www.cigionline.org/static/documents/DPH-paper-Niazi.pdf>
4. J. Geith, “The AI governance regime complex,” in *Handbook on the Global Governance of Artificial Intelligence*. Cheltenham, UK: Edward Elgar Publishing, 2026, pp. 117–132.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of Publisher and/or the editor(s). Publisher and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.