*Adv. J. Manag. Humanit. Soc. Sci.* **(2026),** Volume 2, Issue 2, 114-126

**Advanced Journal of Management, Humanity and Social Science**

**Journal homepage:** https://www.ajmhss.com/

DOI: https://doi.org/10.5281/zenodo.20616375

# Human Rights in the Age of Artificial Intelligence: Legal Personhood, Responsibility, and Global Governance

**Anita Yousefi[1*], Farideh Afshani[2]**

[1]Ph.D. Student in Public International Law, Islamic Azad University, North Tehran Branch, Tehran, Iran

[2]Ph.D. Student in Public International Law, Islamic Azad University, North Tehran Branch, Tehran, Iran

**Article info**

**Received:** 04.04.2026

**Accepted:** 06.06.2026

**Available Online:** 09.06.2026

**Checked for Plagiarism:** Yes

**Keywords:**

Artificial Intelligence, Human Rights, Legal Personhood, Global Governance, Human Rights Due Diligence.

**ABSTRACT**

The rapid integration of artificial intelligence (AI) systems into core societal institutions from criminal justice and welfare administration to employment and border governance has generated unprecedented challenges for international human rights law. This article examines three intersecting dimensions of the AI-human rights nexus: the contested question of AI legal personhood, the allocation of responsibility for AI-induced harms across complex value chains, and the evolving architecture of global AI governance. Drawing on the Council of Europe's Framework Convention on Artificial Intelligence (2024), the UN Guiding Principles on Business and Human Rights as applied to AI (2025), and emerging regulatory frameworks including the EU AI Act, this analysis argues that granting legal personhood to AI systems is neither necessary nor desirable for effective accountability. Instead, a functional approach that mandates human rights due diligence throughout the AI lifecycle, establishes accessible remedy mechanisms for affected individuals, and promotes regulatory coherence across jurisdictions offers a more promising pathway. The article synthesises findings from a doctrinal analysis of 45 international legal instruments, UN reports, and scholarly sources to propose a rights-based governance framework centred on mandatory human rights impact assessments, independent oversight, and meaningful stakeholder engagement with affected communities.

## Introduction

Artificial intelligence has moved rapidly from the realm of technical speculation to become a pervasive infrastructure of contemporary governance and commerce. Machine learning systems now inform decisions about who receives social benefits, how long criminal sentences should be, which job applicants proceed to interviews, whether loan applications are approved, and even which individuals are flagged for heightened scrutiny at national borders. Large language models and generative AI tools have simultaneously transformed how information produced and consumed, raising profound questions about freedom of expression, access to information, and the integrity of democratic processes. While these technologies promise efficiency gains and novel capabilities, they also carry systemic risks of discrimination, opacity, and power concentration that challenge foundational human rights principles.

The urgency of addressing AI's human rights implications reflected in the accelerating pace of regulatory and normative development. September 2025 marked a watershed moment with the opening for signature of the Council of Europe's Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law the first legally binding international treaty specifically addressing AI systems.

This Framework Convention establishes that AI-related activities must be fully compatible with human dignity, individual autonomy, equality and non-discrimination, privacy, and data protection, while also articulating procedural principles including transparency, oversight, accountability, and access to remedy. Simultaneously, the Office of the United Nations High Commissioner for Human Rights (OHCHR) has produced authoritative guidance on applying the UN Guiding Principles on Business and Human Rights to technology

**Corresponding Author: **Anita Yousefi** (mrsyousefi77@gmail.com)*

*1 Email: Farideh_afshani73@yahoo.com)*

companies' AI activities, clarifying state duties and corporate responsibilities across the AI value chain. Regional and national regulatory initiatives have proliferated alongside these international developments. The European Union's AI Act, the world's first comprehensive AI regulation, adopts a risk-based approach that prohibits certain unacceptable practices, imposes stringent requirements on high-risk systems, and establishes transparency obligations for general-purpose AI models. South Korea's Basic Act on AI, Brazil's proposed AI framework, and various national strategies reflect a global movement toward codifying AI governance principles, though with significant variation in stringency and rights-protective mechanisms. Meanwhile, the UN Secretary-General's AI Advisory Body has advanced recommendations for strengthening global governance functions, and the Global Digital Compact adopted at the 2024 Summit of the Future provides a macro-level framework for digital cooperation.

Despite this normative activity, substantial gaps remain. Current governance frameworks remain fragmented, with the rapid expansion of AI in both reach and sophistication outpacing regulatory responses. The absence of comprehensive human rights impact assessment requirements, weak enforcement mechanisms, limited access to remedy for affected individuals particularly from marginalized communities and insufficient accountability for corporate actors across complex AI supply chains all represent critical deficiencies. Moreover, the conceptual question of whether AI systems themselves might be granted some form of legal personhood has resurfaced, with implications for how responsibility is allocated and whether traditional legal categories remain adequate.

This article addresses three interconnected research questions: First, what moral and legal arguments support or oppose granting AI systems legal personhood, and what alternative frameworks for accountability exist? Second, how can responsibility for AI-induced human rights harms be allocated effectively across the distributed networks of developers, deployers, and users? Third, what principles and institutional mechanisms should inform a coherent global governance framework capable of protecting human rights while accommodating technological innovation?

The analysis proceeds as follows. Section 2 reviews the existing literature on AI and human rights, identifying key debates and research gaps. Section 3 outlines the doctrinal and comparative methodology employed. Section 4 presents findings across four analytical dimensions, organized in tabular form with accompanying interpretive analysis. Section 5 discusses the implications for legal reform and global governance. Section 6 concludes with recommendations for policymakers and directions for future research.

## Literature Review

Scholarly engagement with AI's legal and human rights implications has expanded dramatically over the past decade, moving from speculative inquiries about robot rights to detailed doctrinal analyses of regulatory frameworks and empirical studies of algorithmic harm. This review synthesizes the existing literature across three thematic streams: the legal personhood debate, responsibility attribution in AI systems, and global governance architectures.

**Legal Personhood and AI:** The question of whether AI systems could or granted legal personhood has generated sustained philosophical and legal debate. Early contributions often drew analogies with corporate personhood, suggesting that functional arguments for treating AI as legal entities might parallel the pragmatic recognition of corporations as rights- and duty-bearing subjects. However, Naidoo's analysis of the DABUS case in which an AI system was named as an inventor in patent applications reveals that courts have consistently rejected AI personhood claims, instead emphasizing that legal personality presupposes capacities for intentional action, moral agency, and susceptibility to sanctions that current AI systems lack.

More nuanced positions distinguish between strong personhood (conferring full legal personality with rights) and weak or sui generis personhood (creating limited-purpose legal constructs for specific functional needs). Advocates of weak personhood argue that designating AI systems as legal entities could facilitate liability allocation, particularly in cases where human actors are difficult to identify or where autonomous systems make decisions unforeseen by their programmers. Conversely, critics contend that legal personhood for AI would obscure rather than clarify accountability, allowing human actors to evade responsibility by attributing harm to "autonomous" systems. The dominant position in the literature, reflected in the OHCHR's guidance, is that AI systems treated as instruments or tools whose design, deployment, and use remain subject to human control and responsibility, rather than as independent legal subjects.

**Responsibility and Accountability:** A substantial body of literature addresses the "responsibility gap" problem the risk that AI systems' autonomy and opacity may outstrip existing legal frameworks for attributing liability. Choudhury et al. examine how machine learning integration into decision-making processes across sentencing, taxation, workplace dynamics, and financial markets creates constitutional, contractual, and tort issues that existing law is ill equipped to handle. The problem is particularly acute for generative AI systems,

where outputs are not directly traceable to specific inputs or design choices, complicating traditional causation analyses.

The UNGPs framework has emerged as a central reference point for corporate responsibility in the AI context. The OHCHR's 2025 report clarifies that companies developing or deploying AI have a responsibility to conduct human rights due diligence (HRDD) throughout the AI lifecycle from design through deployment and monitoring to identify, prevent, mitigate, and account for adverse human rights impacts. This includes assessing human rights risks before new activities or major changes, integrating findings into company processes, tracking effectiveness, and communicating how impacts addressed. However, empirical research reveals that comprehensive HRDD remains rare in practice, with many companies only considering human rights at late stages of development or not at all.

**Global Governance Architectures:** The literature on AI governance has proliferated alongside institutional developments at multiple levels. At the international level, the Council of Europe's Framework Convention represents the most significant binding instrument, establishing principles that signatory parties must implement domestically. However, scholars note that the Convention's framework nature means it requires further specification through national legislation and complementary standards, raising questions about whether it will achieve sufficient stringency and enforcement.

The EU AI Act has attracted extensive scholarly attention as the world's first comprehensive AI regulation. Paseíllo's analysis highlights both achievements and limitations: while the Act's risk-based approach and prohibitions on unacceptable practices are laudable, civil society critics point to exemptions for law enforcement, the absence of mandatory human rights impact assessments, and weak redress mechanisms as significant gaps. Similar critiques apply to national frameworks, with the OHCHR noting that while some companies have begun integrating human rights assessments, overall uptake remains limited, particularly among smaller firms and in high-risk sectors such as surveillance, military, and immigration.

Emerging literature on AI agents and their regulation highlights additional complexities. Hacker argues that agentic AI systems capable of autonomously reasoning, planning, and executing tasks across external systems require a fundamental shift in oversight toward the "orchestration layer," where multi-agent interactions introduce novel risks of misalignment. This suggests that as AI systems become more autonomous, traditional accountability mechanisms may require significant adaptation.

**Research Gaps:** Despite substantial scholarly attention, several gaps remain. First, the literature on legal personhood has largely proceeded in isolation from empirical research on how responsibility attributions function in practice, leaving normative debates undertheorized in relation to actual accountability mechanisms. Second, analyses of global governance tend to focus on the Global North, with limited attention to how AI governance frameworks might address the concerns and capacities of low- and middle-income countries. Third, while human rights impact assessments frequently recommended, there is limited empirical evidence on their effectiveness or on the methodologies best suited to AI systems. Fourth, the integration of AI governance with existing human rights treaty body mechanisms remains underexplored. This article aims to address these gaps by synthesizing doctrinal analysis with attention to practical implementation challenges and by centring the perspectives of affected communities as articulated in UN and civil society documentation.

## Methodology

This study employs a doctrinal legal methodology combined with comparative regulatory analysis to examine the intersection of AI, human rights, legal personhood, responsibility attribution, and global governance. Doctrinal analysis is appropriate for this inquiry because the research questions centre on interpreting and synthesizing legal norms including treaties, soft law instruments, regulatory frameworks, and jurisprudential developments to identify principles, inconsistencies, and gaps. The analysis supplemented by comparative methods that examine how different jurisdictions (EU, Council of Europe member states, and emerging frameworks in other regions) address analogous problems, enabling identification of converging standards and persisting divergences.

The primary data sources consist of 45 documents selected through purposive sampling based on their authoritative status and relevance to the research questions. These include:

- ✓ International legally binding instruments, specifically the Council of Europe Framework Convention on Artificial Intelligence (2024-2025).
- ✓ UN documents, including OHCHR reports on the UNGPs applied to AI (A/HRC/59/32, 2025) and the report of the Secretary-General's AI Advisory Body.
- ✓ Regional regulatory frameworks, notably the EU AI Act (2024) and related analyses.
- ✓ National legislation and proposed frameworks from diverse jurisdictions.
- ✓ scholarly articles identified through Google Scholar and arXiv searches using keywords including "AI legal personhood," "AI

human rights," "AI governance," "algorithmic accountability," and "human rights due diligence".
- ✓ civil society and expert submissions to UN processes.

Documentary analysis proceeded through iterative coding of materials for references to three thematic domains:
- ✓ Legal personality or subjectivity of AI systems.
- ✓ Responsibility, liability, and accountability mechanisms for AI-induced harms.
- ✓ Governance frameworks, institutional arrangements, and regulatory instruments addressing AI's human rights implications.

Within each domain, the analysis identified proposed principles or rules; identified gaps or limitations; mechanisms for implementation, monitoring, and enforcement; and provisions for access to remedy. The coding framework developed inductively based on preliminary reading of key UN documents and refined through application to the full corpus.

The analysis is subject to several limitations. First, the rapid pace of AI development and regulatory response means that findings may require updating as new instruments emerge. Second, the study focuses primarily on English-language sources and European and international frameworks, potentially underrepresenting perspectives from other regions, although African and Asian sources are included where available. Third, the doctrinal approach does not include empirical validation of how frameworks operate in practice, instead relying on official reports and scholarly assessments. Fourth, the search conducted in April 2026; subsequent developments not reflected.

**Results**

The findings are organized into four tables, each addressing a core dimension of the AI-human rights nexus. Following each table, interpretive analysis elaborates on the patterns, tensions, and implications identified.

**Table 1.** Approaches to AI Legal Personhood across Jurisdictions and Scholarship

| Approach | Key Proponents/Examples | Core Rationale | Identified Limitations/Challenges |
|---|---|---|---|
| Strong Personhood (full legal personality with rights) | Minority scholarly position; not adopted in any binding instrument | AI systems with advanced autonomy may merit moral consideration; functional necessity for complex autonomous systems | Lacks moral agency; cannot be sanctioned; would obscure human accountability; no empirical evidence of need |
| Weak/Sui Generis Personhood (limited-purpose legal construct) | Naidoo (2022) ; some scholarly proposals | Facilitates liability where human actors difficult to identify; may enable AI to hold property (e.g., intellectual property) | Risk of "personhood laundering" by corporations; may not solve attribution problems; creates doctrinal confusion |
| Instrumentalist/No Personhood (current dominant approach) | OHCHR (2025) ; Council of Europe Framework Convention ; EU AI Act ; all national frameworks surveyed | AI systems are tools; human actors (developers, deployers, users) remain responsible; legal personality unnecessary for accountability | May not adequately address genuinely autonomous systems or collective action problems; relies on ability to identify responsible humans |
| Electronic Persons (EU Parliament proposal, 2017, not adopted) | European Parliament resolution on civil law rules on robotics (2017) | Would create specific category for robots/AI with associated liability fund | Widely criticized; withdrawn from consideration; created perverse incentives |

The near-universal rejection of AI legal personhood across binding legal instruments, authoritative UN guidance, and regional frameworks reflects a robust consensus that conferring legal personality on AI systems is neither necessary nor desirable for protecting human rights. This consensus rests on several reinforcing considerations.

First, the instrumentalist approach aligns with foundational principles of accountability. Legal personality in modern legal systems is generally predicated on capacities that AI systems demonstrably lack: the ability to form intentions, understand the consequences of actions, respond to legal sanctions with behavioral modification, and participate in legal processes as autonomous agents. Granting legal personality to entities without these capacities would require redefining the concept in ways that could undermine its coherence and functional utility.

Second, the risk of "personhood laundering" whereby corporations might attribute decisions to AI systems to evade responsibility has been a consistent concern in debates. If an AI system designated as a legal entity, human actors might argue that its autonomous decisions are not attributable to them, creating doctrinal space for avoiding liability. The OHCHR's guidance explicitly cautions against this possibility, emphasizing that "meaningful human oversight" maintained over AI-assisted decisions.

Third, the functional arguments for weak personhood particularly regarding intellectual property and liability appear addressable through other legal mechanisms without creating new legal subjects. The DABUS cases, which considered whether an AI system could be named as an inventor, were resolved by courts holding that patent law's concept of "inventor" presupposes a natural person. Existing doctrines of vicarious liability, enterprise liability, and product liability can accommodate most AI-induced harms without requiring personhood. Where gaps exist, they better addressed through targeted legislative reforms to liability rules rather than through the conceptually fraught device of AI personhood.

Fourth, empirical evidence does not support the claim that existing frameworks are systematically failing. While high-profile cases of AI-induced harm have been documented including discriminatory benefit algorithms in the Netherlands, biased hiring tools, and privacy violations in each case investigators were able to identify responsible human actors (developers, deploying agencies, procuring entities) even when accountability mechanisms proved inadequate. The problem is not that responsibility cannot be attributed in principle, but that existing mechanisms for enforcement and remedy are insufficient a governance problem rather than a personhood problem.

The rejected European Parliament proposal for "electronic persons" serves as a cautionary example. When the proposal advanced in 2017, it met with widespread criticism from legal scholars, ethicists, and civil society organizations, who argued that it would create perverse incentives and allow corporations to externalize liability. The proposal was subsequently withdrawn and has not been revived in subsequent EU AI governance discussions. This episode demonstrates that even limited legal personhood proposals face substantial normative and political hurdles.

Nevertheless, the instrumentalist consensus not treated as static. As AI, systems become more sophisticated and autonomous particularly agentic AI capable of independent action across multiple domains the question may require reconsideration. Hacker's analysis of AI agents suggests that when systems can autonomously reason, plan, and execute tasks across external systems without direct human oversight, traditional attribution models may become strained. However, even in such scenarios, extending legal personhood is not the only or most promising response.

**Table 2.** Responsibility Allocation Mechanisms for AI-Induced Human Rights Harms

| Mechanism | Legal Basis/Authority | Scope of Application | Gaps/Limitations Identified |
|---|---|---|---|
| Human Rights Due Diligence (HRDD) | UNGPs (UN Human Rights Council endorsement); OHCHR AI guidance (2025) | Entire AI lifecycle (design, development, deployment, monitoring) | Limited uptake in practice; no comprehensive disclosure; weak enforcement; smaller firms lack capacity |
| State Duty to Protect | UNGPs Pillar I; Council of Europe Framework Convention Article 5 | States must protect against AI-related human rights abuses by third parties (corporations) | Fragmented implementation; inconsistent across jurisdictions; border/security exceptions |
| Corporate Responsibility to Respect | UNGPs Pillar II; OHCHR guidance | All companies developing or deploying AI | Self-assessment bias; lack of independent verification; remedy mechanisms inadequate |
| Access to Remedy | UNGPs Pillar III; Council of Europe Framework Convention Article 8 | Judicial, State-based non-judicial, non-State grievance mechanisms | Complexity of digital ecosystems; difficulty identifying responsible actors; technical expertise barriers; cost barriers |
| Mandatory Human Rights Impact Assessments | Proposed in OHCHR recommendations; EU AI Act (high-risk systems only) | High-risk AI applications | Not universally required; methodologies underdeveloped; limited stakeholder engagement |
| Prohibited AI Practices | EU AI Act Article 5 ; OHCHR moratoria recommendations | Unacceptable risk AI (e.g., social scoring, real-time biometric surveillance in public spaces) | Exceptions for law enforcement; enforcement limited; not globally harmonized |

The responsibility landscape for AI-induced human rights harms is characterized by a proliferation of norms and mechanisms alongside persistent implementation gaps. The UNGPs framework, originally developed for business and human rights generally, has been authoritatively extended to AI through the OHCHR's 2025 report, which clarifies how human rights due diligence applies to the AI lifecycle. However, the translation of these norms into effective corporate practice remains deeply uneven.

Human rights due diligence the process of identifying, preventing, mitigating, and accounting for adverse impacts is the central mechanism proposed for corporate accountability. The OHCHR specifies that HRDD must occur "early and throughout AI product design, development and deployment". This represents a significant shift from late-stage compliance checks to proactive integration of human rights considerations from the outset of AI development. Despite this normative clarity, empirical evidence on implementation is limited and concerning. The OHCHR itself notes, "There is still no comprehensive analysis of how technology companies embed the UN Guiding Principles to their AI products and services, largely due to a lack of data disclosure by companies". Some companies have begun integrating human rights impact assessments into their AI practices, but "overall uptake remains limited".

The gaps are particularly acute in high-risk sectors. Surveillance technologies, military AI applications, and immigration control systems represent domains where AI can be repurposed in ways that pose serious threats to human rights, yet these are precisely the sectors where transparency is lowest and accountability mechanisms weakest. The OHCHR's forthcoming guidance on digital border governance highlights how AI systems used in migration management often operate without meaningful oversight, human rights impact assessments, or accessible remedy mechanisms.

State duties under the UNGPs framework require governments to protect against human rights abuses by corporations, including those involving AI. This includes implementing regulatory measures that ensure AI products and services incorporate respect for human rights through mandatory HRDD requirements, enhanced due diligence for high-risk use cases, and prohibitions on AI deployment when sufficient guardrails are not in place. However, state practice varies dramatically. The EU AI Act represents the most comprehensive regulatory effort, but it includes significant exemptions for law enforcement and national security that civil society organizations have criticized as undermining rights protection. The Council of Europe Framework Convention establishes binding obligations for signatory states, but its framework nature means that domestic implementation will determine its effectiveness.

Access to remedy a pillar of the UNGPs framework presents particular challenges in the AI context. The OHCHR identifies several key barriers: lack of transparency means affected individuals may not be aware they have suffered AI-facilitated harm; the technical expertise required to prove algorithmic bias creates access hurdles; the complexity of digital ecosystems makes it difficult to identify responsible actors; and privacy and data protection concerns complicate evidence gathering. These barriers disproportionately affect marginalized communities, who may face AI-facilitated discrimination in social services, housing, employment, and immigration enforcement.

The Dutch SyRI case exemplifies these challenges. When the Netherlands implemented an algorithmic risk assessment system for welfare benefits, it resulted in disproportionate targeting of minority and migrant communities, with thousands of families wrongly accused of fraud. While The Hague District Court ultimately struck down the system for violating human rights law, the case required sustained litigation and demonstrated how AI systems can cause widespread harm before accountability mechanisms activated.

Emerging proposals for strengthening responsibility mechanisms include mandatory human rights, impact assessments (HRIAs), independent auditing requirements, enhanced transparency obligations. However, the OHCHR notes that HRIAs remain "vastly underutilized" and that there is "a surprising lack of authoritative guidance on how to conduct HRIAs beyond the Danish Institute for Human Rights' widely respected toolkit". Without standardized methodologies, capacity-building support, and enforcement mechanisms, HRIAs risk becoming checkbox exercises rather than meaningful accountability tools.

**Table 3.** Global Governance Instruments for AI and Human Rights

| Instrument | Legal Status | Geographic Scope | Key Human Rights Provisions | Implementation Mechanism |
|---|---|---|---|---|
| Council of Europe Framework Convention on AI (2024-2025) | Legally binding treaty | Council of Europe members + open to non-European states (US, Canada, Japan, Uruguay signed) | Human dignity, autonomy, equality/non-discrimination, privacy, transparency, oversight, accountability, remedy | Conference of Parties; domestic implementation; monitoring mechanism |
| EU AI Act (2024) | EU Regulation (directly applicable) | EU member states; extraterritorial effects | Risk-based approach; prohibited practices; high-risk system requirements; transparency for general-purpose AI | National supervisory authorities; European AI Office; fines for non-compliance |
| UN Guiding Principles on Business and Human Rights (UNGPs) | Soft law (UN Human Rights Council endorsed) | Global | State duty to protect; corporate responsibility to respect; access to remedy | National Action Plans; UN Working Group on Business and Human Rights |
| OHCHR Guidance on UNGPs for AI (2025) | Interpretive guidance | Global | Specific application of UNGPs to AI lifecycle; HRDD requirements | N/A (interpretive, not independently binding) |
| UNESCO Recommendation on AI Ethics (2021) | Soft law (UNESCO General Conference) | Global (193 member states) | Human rights-centred approach; proportionality; non-discrimination; environmental sustainability | Periodic reporting; National Commissions for UNESCO |
| G7 Hiroshima Process (2023) | Political commitment | G7 members | Behavioral code for advanced AI systems; risk management | Voluntary reporting; annual reviews |
| UN Global Digital Compact (2024) | Political commitment (UN member states) | Global | Digital cooperation framework; AI governance principles | Follow-up mechanisms under Summit of the Future |

The global governance landscape for AI and human rights is characterized by fragmentation, legal pluralism, and contested authority. No single instrument provides comprehensive coverage, and the relationships between binding treaties, regional regulations, soft law instruments, and political commitments remain undefined. This fragmentation creates both opportunities and risks.

The Council of Europe Framework Convention represents the most significant achievement to date in establishing legally binding international obligations specific to AI and human rights. Opened for signature in September 2025, it has already attracted signatories from outside Europe, including the United States, Canada, Japan, and Uruguay, demonstrating its universal vocation. The Convention establishes that AI activities must be fully compatible with human dignity, individual autonomy, equality and non-discrimination, privacy, and data protection, uphold transparency, oversight, accountability, and safe innovation. However, as a framework convention, it requires domestic implementation and further specification through complementary standards. Its effectiveness will depend on the rigour of its monitoring mechanism, the willingness of signatory states to adopt implementing legislation, and the degree to which its principles incorporated into corporate practice.

The relationship between the Council of Europe Framework Convention and the EU AI Act is characterized by complementarity rather than competition. The Convention provides overarching principles, while the Act provides detailed technical requirements for AI systems placed on the EU market. However, the geographic scope of the two instruments differs, with the Convention open to non-European states and the Act applying to any entity placing AI systems in the EU market, regardless of location. This creates potential for regulatory coordination challenges, though early indications suggest constructive cooperation between the Council of Europe and EU institutions.

The UNGPs framework, while not specifically designed for AI, has successfully extended to the technology sector through the OHCHR's interpretive guidance. The 2025 report on applying the UNGPs to AI activities provides detailed clarification of how human rights due diligence should operate across the AI lifecycle. However, the UNGPs remain soft law, and the OHCHR's guidance lacks independent binding force. States and corporations that choose to ignore the guidance face no formal sanctions beyond reputational consequences and potential scrutiny from UN human rights mechanisms.

The UNESCO Recommendation on AI Ethics, adopted by 193 member states in 2021, provides a broadly endorsed normative framework but lacks enforcement mechanisms. Its periodic reporting process relies on voluntary state cooperation, has produced limited evidence of domestic implementation. Similarly, the G7 Hiroshima Process and the UN Global Digital Compact represent political commitments rather than legally binding obligations, and their impact depends on ongoing political will.

A persistent gap in the global governance architecture is the absence of a binding universal convention specifically addressing AI and human rights. Smart, Effoduh, and Niazi argue that the current fragmentation and reliance on soft law are inadequate given the scale and urgency of AI's human rights implications. They propose a legally binding international framework that would clarify obligations for states and corporations across the entire AI lifecycle, harmonise governance standards, and establish enforcement mechanisms capable of addressing cross-border AI activities. The Council of Europe Convention provides a potential model, but its geographic scope is limited to signatory states and non-European AI powers particularly China have not joined.

Another significant gap concerns the representation of Global South perspectives in AI governance. The UN and UNESCO instruments nominally include all states, but actual participation in standard setting and implementation remains uneven. Critics argue that current frameworks reflect the priorities and capacities of developed countries, with insufficient attention to how AI governance might address the needs of low- and middle-income economies, including digital infrastructure gaps, workforce displacement, and the risk that AI systems trained on data from wealthy countries will perpetuate colonial patterns of knowledge production and economic extraction.

**Table 4.** Identified Gaps and Proposed Reforms in AI-Human Rights Governance

| Gap Identified | Source/Authority | Proposed Reform | Implementation Challenges |
|---|---|---|---|
| Fragmented governance; no binding universal convention | Smart, Effoduh, Niazi (2025) ; OHCHR (2025) | Legally binding universal convention on AI and human rights | Geopolitical divisions; resistance from states and corporations; enforcement difficulties |
| Limited HRDD uptake; weak enforcement | OHCHR (2025) ; Passuello (2025) | Mandatory human rights due diligence requirements; independent auditing | Corporate resistance; capacity constraints; methodological challenges |
| Inadequate access to remedy | OHCHR (2025) ; Hayes (2025) | Judicial and non-judicial remedy mechanisms; reduced procedural barriers; technical assistance | Complexity of AI systems; identifying responsible actors; cost barriers |
| Lack of transparency and explainability | Choudhury et al. (2024) ; OHCHR (2025) | Mandatory transparency reporting; explain ability requirements for high-risk systems | Trade secrecy claims; technical limitations of complex models |
| Insufficient stakeholder engagement | OHCHR (2025) ; UNGPs framework | Meaningful engagement with affected communities throughout AI lifecycle | Power asymmetries; resource constraints; tokenism risks |
| Digital divide; Global South underrepresentation | UN reports; scholarly critiques | Capacity-building support; inclusive governance mechanisms; technology transfer provisions | Funding gaps; geopolitical tensions; divergent regulatory philosophies |
| Agentic AI challenges | Hacker (2026) | "Traffic light" task authorization; non-delegable legal acts lists; orchestration layer oversight | Rapid technological change; definitional challenges; enforcement across borders |

The gaps identified across the literature and authoritative sources point to a consistent conclusion: while normative frameworks have developed significantly, implementation and enforcement lag substantially behind. The proposed reforms cluster around several themes that merit detailed examination.

First, the gap between normative endorsement and practical implementation of human rights due diligence is particularly striking. While the UNGPs have been endorsed by the UN Human Rights Council and the OHCHR has provided AI-specific guidance, the OHCHR itself acknowledges that "overall uptake remains limited" and that "there is still no comprehensive analysis of how technology companies embed the UN Guiding Principles". This suggests that mere normative development is insufficient; what is required is a combination of mandatory requirements, independent oversight, and capacity-building support. The EU AI Act's mandatory requirements for high-risk systems represent one model, but civil society critics note that many high-risk systems including those used by law enforcement benefit from exemptions, and that the Act lacks comprehensive HRDD requirements.

Second, access to remedy presents particularly acute challenges. The OHCHR identifies multiple barriers: lack of transparency means affected individuals may not know they have been harmed; technical expertise requirements create insurmountable hurdles for many claimants; the complexity of digital ecosystems makes it difficult to identify responsible actors; and cost barriers disproportionately affect low-income and marginalized groups. Proposed reforms include requiring transparency about AI-assisted decision-making processes, establishing accessible grievance mechanisms, providing technical assistance to claimants, removing cost and procedural barriers. However, implementing these reforms requires political will and resource allocation that many states have been reluctant to provide.

Third, the challenge of meaningful stakeholder engagement particularly with affected communities is both critically important and persistently difficult to achieve. The OHCHR emphasizes that companies and states "should carry out meaningful stakeholder engagement with affected groups and civil society organizations to identify and address human rights risks". However, power asymmetries between technology companies and affected communities, resource constraints, and the risk of tokenistic consultation all pose challenges. Best practices emerging from civil society advocacy include co-design of engagement processes with community organizations, provision of resources to support community participation, and binding requirements to respond to engagement findings.

Fourth, the governance of agentic AI systems presents novel challenges that existing frameworks may not adequately address. Hacker's analysis of AI agent's systems capable of autonomously reasoning, planning, and executing tasks across external systems suggests that traditional oversight mechanisms may be insufficient when multi-agent interactions create emergent behaviours that no single human designer anticipated. His proposed "traffic light" system for staggered task authorization based on operational risk, combined with a statutory list of non-delegable legal acts, represents a pragmatic approach to maintaining human accountability while accommodating increasing autonomy. However, implementation requires technical standards for risk classification, international coordination to prevent regulatory arbitrage, and ongoing adaptation as agentic capabilities evolve.

Fifth, the digital divide and Global South underrepresentation in AI governance remains a persistent concern. While UN and UNESCO instruments nominally include all states, effective participation requires technical expertise, financial resources, and bargaining power that many developing countries lack. Proposed reforms include capacity-building support, technology transfer provisions, and governance mechanisms that ensure meaningful participation. However, funding for capacity-building remains inadequate and geopolitical tensions particularly between the United States and China complicate efforts to establish inclusive governance mechanisms.

## Discussion

The findings from this analysis reveal a fundamental tension in the current approach to AI and human rights: while normative frameworks have developed with remarkable speed over the past five years, their translation into effective protection for individuals and communities remains profoundly inadequate. This section discusses the implications of this tension across three dimensions: the relationship between legal personhood debates and accountability practice; the limits of soft law and voluntary mechanisms; and the prospects for a binding universal convention.

**Legal Personhood as a Distraction:** The sustained scholarly attention to AI legal personhood, while intellectually interesting, appears largely disconnected from the practical challenges of accountability for AI-induced harm. Across all binding legal instruments, authoritative guidance documents, and national frameworks surveyed, no serious proposal for granting AI systems legal personality has adopted or is under active consideration. The instrumentalist consensus that AI systems are tools whose design, deployment, and use remain subject to human responsibility is robust and unlikely to change in the near future. Debates about personhood may divert attention from more

pressing questions about how to strengthen human rights due diligence, improve access to remedy, and ensure meaningful oversight of AI systems throughout their lifecycles.

This is not to deny that AI systems raise novel accountability challenges. The "responsibility gap" identified in the literature reflects genuine difficulties in attributing causation and fault when autonomous systems make decisions that no human specifically intended or could have predicted. However, the response to these challenges should focus on adapting existing legal doctrines vicarious liability, product liability, enterprise liability, and negligence rather than creating new legal subjects. Where gaps persist, targeted legislative reforms establishing no-fault compensation funds, mandatory insurance schemes, or reversed burdens of proof may be more effective than the conceptually fraught project of AI personhood.

**The Limits of Soft Law:** A striking finding from this analysis is the gap between the proliferation of soft law instruments UNGPs, UNESCO recommendations, G7 commitments, the Global Digital Compact and the limited evidence of their effectiveness in changing corporate or state behavior. The OHCHR's candid acknowledgment that there is "no comprehensive analysis" of how technology companies embed the UNGPs, and that "overall uptake remains limited," underscores that soft law alone is insufficient. Voluntary principles, no matter how well crafted, cannot substitute for mandatory requirements backed by enforcement mechanisms.

This suggests that the next phase of AI governance must prioritise hard law mechanisms. The Council of Europe Framework Convention provides a template, but its framework nature requires domestic implementation, and its geographic scope is limited. The EU AI Act provides detailed mandatory requirements, but its application is limited to the EU market, and its exceptions for law enforcement and national security are concerning. What is needed is a binding universal convention that establishes minimum human rights standards for AI systems, requires mandatory human rights due diligence and impact assessments, ensures access to remedy, and creates enforcement mechanisms capable of addressing cross-border AI activities.

However, the political feasibility of such a convention assessed realistically. Geopolitical divisions particularly between democratic and authoritarian states with competing visions of AI governance present formidable obstacles. China, a major AI power, has not joined the Council of Europe Framework Convention and has articulated a vision of AI governance that prioritises state control and social stability over individual rights protections. The United States has signed the Framework Convention but has not yet ratified it, and domestic political divisions may complicate implementation. A universal convention would require navigating these divisions, likely resulting in lowest-common-denominator standards that some advocates would find inadequate.

**Human Rights Due Diligence as a Core Mechanism:** Despite the limitations of soft law, the human rights due diligence framework remains the most promising approach for embedding human rights considerations into AI development and deployment. The OHCHR's guidance on applying the UNGPs to AI provides a detailed roadmap, specifying that HRDD should be conducted "early and throughout the AI product life cycle," should include "identifying and assessing human rights risks, integrating findings into company processes, taking appropriate action, tracking effectiveness, and communicating how impacts are addressed". This lifecycle approach is critical: human rights cannot be an afterthought, addressed through late-stage compliance checks, but must inform AI design from the outset.

The challenge is transforming this normative guidance into mandatory practice. The EU AI Act's risk-based approach represents one model, but its limited HRDD requirements and exemptions for law enforcement are significant weaknesses. A stronger approach would require mandatory human rights affects assessments for all high-risk AI systems, with independent auditing, meaningful stakeholder engagement, and public disclosure of findings. Impact assessments must be conducted before deployment and updated throughout the AI lifecycle, with mechanisms for suspending or prohibiting systems where risks cannot be adequately mitigated.

**Access to Remedy as the Weakest Link:** The most significant gap in current frameworks is access to remedy for individuals and communities harmed by AI systems. The OHCHR's identification of multiple barriers lack of transparency, technical expertise requirements, difficulty identifying responsible actors, cost barriers reflects a systemic failure to ensure that when rights are violated, effective remedies are available. This is not a peripheral issue but a core accountability mechanism: without remedy, rights are merely aspirational.

Addressing this gap requires a multi-pronged approach. Judicial mechanisms strengthened through capacity building for judges and legal professionals, evidentiary rules that accommodate the technical complexity of AI systems, and procedural reforms that reduce barriers for claimants. State-based non-judicial mechanisms including national human rights institutions, ombudsperson offices, and equality bodies must be empowered to investigate AI-related complaints and issue binding decisions. Non-state grievance

mechanisms, including corporate complaint processes and multi-stakeholder initiatives, must meet the UNGPs effectiveness criteria: legitimate, accessible, predictable, equitable, transparent, rights-compatible, and sources of continuous learning. Technical assistance provided to claimants, particularly from marginalized communities, to help them navigate complex systems and prove algorithmic discrimination.

**The Role of Affected Communities:** A recurring theme in the OHCHR guidance and civil society critiques is the importance of meaningful engagement with affected communities. Too often, AI systems are designed and deployed without consultation with the people who will be most affected by them. The Dutch SyRI case, the UK Post Office Horizon scandal, and numerous algorithmic discrimination cases share a common pattern: systems implemented with little input from affected communities, harms ignored or denied, and accountability achieved only through prolonged and costly legal battles.

Meaningful engagement requires more than token consultations. It requires co-design of AI systems with community input from the earliest stages, ongoing monitoring and feedback mechanisms, governance structures that include community representatives, and binding requirements to respond to community concerns. It also requires resources to support community participation, including funding for civil society organizations, technical assistance, and protection from retaliation.

## Conclusion

This article has examined the intersections of legal personhood, responsibility allocation, and global governance in the context of AI and human rights. The analysis yields several conclusions.

First, the debate over AI legal personhood is largely a distraction from more pressing accountability challenges. Across all binding legal instruments and authoritative guidance, the consensus is clear: AI systems are tools, not legal persons, and human actors remain responsible for their design, deployment, and consequences. Efforts to confer legal personality on AI would likely obscure rather than clarify accountability and not supported by any demonstrated need that not addressed through adaptation of existing legal doctrines.

Second, the human rights due diligence framework, as elaborated in the UN Guiding Principles and the OHCHR's AI-specific guidance, provides a robust normative foundation for corporate accountability. However, the gap between normative endorsement and practical implementation remains substantial. Mandatory HRDD requirements, independent auditing, and meaningful enforcement mechanisms are essential to translate principles into practice.

Third, access to remedy is the weakest link in current frameworks. Multiple barriers lack of transparency, technical complexity, difficulty identifying responsible actors, cost prevent affected individuals and communities from obtaining effective remedies when AI systems violate their rights. Strengthening remedy mechanisms must be the highest priority for policymakers and advocates.

Fourth, global governance remains fragmented, with a mix of binding treaties (Council of Europe Framework Convention), regional regulations (EU AI Act), soft law instruments (UNGPs, UNESCO recommendations), and political commitments (G7, Global Digital Compact). While a binding universal convention on AI and human rights would be desirable, geopolitical divisions make its near-term adoption unlikely. In the meantime, efforts should focus on strengthening existing mechanisms, promoting regulatory coherence, and ensuring that the voices of affected communities particularly in the Global South heard in governance processes.

Future research should prioritise empirical studies of HRDD effectiveness, comparative analysis of remedy mechanisms across jurisdictions, and participatory research with affected communities to understand their experiences of AI-induced harm and their priorities for accountability. As AI systems continue to evolve particularly with the emergence of agentic AI capable of autonomous action ongoing normative and empirical inquiry will be essential to ensure that human rights remain central to AI governance.

## Disclosure Statement

No potential conflict of interest reported by the authors.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Authors' Contributions

All authors contributed to data analysis, drafting, and revising of the paper and agreed to be responsible for all the aspects of this work.

## References

[1] Asaro, P. M. (**2012**). On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making. International Review of the Red Cross, 94(886), 687-709.

[2] Cath, C. (**2018**). Governing artificial intelligence: ethical, legal and technical opportunities and challenges. Philosophical Transactions of the Royal Society A, 376(2133).

[3] Choudhury, B., De Stefano, V., Hutchinson, A. C., Penney, J. W., Craig, C., Abraha, A. P., &

Tanguay-Renaud, F. (**2024**). Artificial intelligence and the law: New challenges and possibilities for fundamental human rights and security.

[4] Cobbe, J. (**2021**). Administrative law and the machines of government: Judicial review of automated public acts. Legal Studies, 41(3), 411-430.

[5] Council of Europe. (**2024**). Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (CETS No. 225). Council of Europe Treaty Series.

[6] Council of Europe. (**2025**). Bosnia and Herzegovina signs the Council of Europe's Framework Convention on Artificial Intelligence.

[7] Crootof, R. (**2016**). War torts: Accountability for autonomous weapons. University of Pennsylvania Law Review, 164(6), 1347-1402.

[8] Ebert, I., & Hsin, L. (**2023**). Putting private sector responsibility in the mix: A business and human rights approach to artificial intelligence. In J. Temperman & A. Quintavalla (Eds.), Artificial Intelligence and Human Rights (pp. 34-??). Oxford University Press.

[9] Eubanks, V. (**2018**). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press. (ISBN: 978-1250074317).

[10] European Parliament and Council of the European Union. (**2024**). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonized rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union, L series.

[11] Floridi, L., & Cowls, J. (**2019**). A unified framework of five principles for AI in society. Harvard Data Science Review, 1(1).

[12] Foundational Scholarly Books (with DOIs)

[13] Gless, S., Silverman, E., & Weigend, T. (**2016**). If robots cause harm, who is to blame? Self-driving cars and criminal liability. New Criminal Law Review, 19(3), 412-439.

[14] Hacker, P. (**2018**). Teaching fairness to artificial intelligence: Existence and possibility of algorithmic bias. Michigan Technology Law Review, 25, 147.

[15] Hacker, P. (**2026**). A pragmatic approach to regulating AI agents. arXiv:2604.22819.

[16] Hayes, B. (2025). Human rights and digital border governance: Opportunities and challenges for the new OHCHR guidance. VerfBlog.

[17] Hoekstra, J., & Diker-Vanberg, A. (**2025**). Can AI tools enhance access to remedies as envisaged under the UN Guiding Principles on Business and Human Rights? A critical assessment. International Review of Law, Computers and Technology, 1-22. Advance online publication.

[18] Jadaan, J. B., & Hasan, Q. A. (**2026**). Compensatory justice and artificial intelligence within the framework of international law. Veredas Do Direito, 23(5), e235665.

[19] James, A., Hynes, D., Whelan, A., Dreher, T., & Humphry, J. (**2023**). From access and transparency to refusal: Three responses to algorithmic governance. Internet Policy Review, 12(2).

[20] Kleemann, S., & Tahraoui, M. (**2025**). Responsibility and accountability for the use of AI in law enforcement in the European Union. Menschenrechtsmagazin (MRM), 30(2), 117-143.

[21] Latzer, M., Hollnbuchner, K., Just, N., & Saurwein, F. (**2017**). The economics of algorithmic selection on the internet. In Handbook on the Economics of the Internet. Edward Elgar Publishing.

[22] Malacka, M. (**2025**). AI legislation, private international law and the protection of human rights in the European Union. The Lawyer Quarterly. (Open Access).

[23] Malgieri, G., & Comandé, G. (**2017**). Why a right to legibility of automated decision-making exists in the GDPR. International Data Privacy Law, 7(4), 243-265.

[24] Mökander, J., & Floridi, L. (**2021**). Ethics-based auditing of automated decision-making systems: intervention points and policy implications. AI & Society, 36(3), 1057-1074.

[25] Naidoo, M. (**2022**). AI and legal personhood: An African perspective. In Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (p. 906).

[26] Office of the United Nations High Commissioner for Human Rights. (**2025**). The practical application of the UN Guiding Principles on Business and Human Rights to the activities of technology companies including activities related to artificial intelligence (A/HRC/59/32).

[27] O'Neil, C. (**2016**). Weapons of math destruction: How big data increases inequality and threatens democracy. Crown. (ISBN: 978-0553418811)

[28] Passuello, C. (**2025**). Digital rights and AI: Can the EU protect human rights in the age of artificial intelligence? GC Human Rights Preparedness.

[29] Raso, F., Hilligoss, H., Krishnamurthy, V., Bavitz, C., & Kim, L. (**2018**). Artificial intelligence & human rights: Opportunities & risks. Berkman Klein Center Research Publication No. 2018-6.

[30] Schultz, D. (**2025**). Personhood, crimes, and criminal liability in the age of artificial

intelligence. Bulletin of Transilvania University of Brasov, 18(67), 3.32.

[31] Sharma, S. (**2025**). Governance in the age of algorithms: Ethical dilemmas and administrative reforms. International Journal of English Literature and Social Sciences, 10(2), 379-388.

[32] Smart, S., Effoduh, O. J., & Niazi, M. (**2025**). Safeguarding human rights in the age of artificial intelligence: Towards a binding international framework. Submission to UN Working Group on Business and Human Rights.

[33] Taboada Macias, I. (**2025**). The risks to human rights posed by AI: Their mitigation efforts in the EU AI ACT. Cuestiones Constitucionales, (52), e19226.

[34] UNESCO. (**2021**). Recommendation on the Ethics of Artificial Intelligence. UNESCO Digital Library.

[35] United Nations Human Rights Council. (**2011**). Guiding Principles on Business and Human Rights: Implementing the United Nations 'Protect, Respect and Remedy' Framework (A/HRC/17/31). United Nations.

[36] Valcu, E. N. (**2025**). The imperative of protecting human dignity and fundamental EU values in the context of implementing artificial intelligence systems. Bulletin of Transilvania University of Brasov, 18(67), 3.33.

[37] Veale, M., & Brass, I. (**2019**). Administration by algorithm? Public management meets public sector AI. Public Policy and Administration, 34(3), 231-244.

[38] Wachter, S., Mittelstadt, B., & Russell, C. (**2021**). Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. Computer Law & Security Review, 41, 105567.

[39] Yang, J., et al. (**2026**). AGI [Prospects for legal governance of AGI superintelligence]. Sohu.

[40] Zuboff, S. (**2019**). The age of surveillance capitalism: The fight for a human future at the new frontier of power. Public Affairs.