REGULATING AI ACROSS BORDERS: THE CASE FOR A BINDING INTERNATIONAL FRAMEWORK

Sibghat Ullah and Simra Sohail[*]

Abstract

This article argues that fragmented governance approaches to artificial intelligence are insufficient and that a binding international treaty is urgently required. Unlike existing soft-law initiatives, the proposed framework is rooted in enforceability, adaptability, and multilateral legitimacy. Drawing from lessons in arms control, climate accords, and data protection, the paper advances original design pathways, including tiered obligations, sector-specific annexes, and the creation of a Global AI Observatory. The analysis demonstrates that without such a treaty, regulatory arbitrage, inequities between the Global North and South, and risks to fundamental rights will intensify. The contribution lies not only in diagnosing governance failures but in proposing a practical architecture for binding global regulation. In doing so, the article extends scholarly debates beyond descriptive accounts toward normative institutional design. The findings are relevant to policymakers, international lawyers, and scholars concerned with embedding human rights and equity in the governance of emerging technologies.

Keywords: Artificial Intelligence, International Law, Global Governance, Treaty, Human Rights, Regulation.

Introduction

Over the last decade, Artificial Intelligence (AI) has become omnipresent within all areas of life, including governance, law, education, health, etc. It is evolving quickly, raising urgent questions about how it should be governed.

[*]Sibghat Ullah is a recent Law Graduate from Lahore University of Management Sciences (LUMS), Lahore currently serving as a Legal Officer at Almoiz Industries. Passionate about Constitutional Law, International Law, Governance, and Rights – based reform in Pakistan, he aims to contribute to legal and policy advancements. Simra Sohail did LLB Hons from the University of London with an upper second – class hons. She is currently a law faculty member at LGS international degree programme and a research fellow at the Global Institute of Law. She has keen interest in International Law, Constitutional Law and Human Rights Law.

As its reach expands, so do the risks—misinformation spreads with precision, and surveillance increasingly operates across borders without oversight. Yet regulation remains weak. Most countries rely on voluntary codes and industry-led standards, leaving major gaps in accountability. Binding laws are rare, and enforcement is inconsistent. With private actors shaping the future of AI, the absence of strong legal frameworks threatens to undermine public trust. Without clear rules, the technology's promise risks being overshadowed by its potential for harm. Making such a scenario even worse are considerations related to the nature of AI itself.

The lack of clear decision-making in AI and hidden values in these systems make the current crisis worse. As AI is used more in important sectors, it can significantly impact public discourse and increase inequality. If it is not controlled, AI may create a new type of digital colonialism and cause global disparities.[1] The European Union's AI Act tries to set high standards for responsible AI governance, but it may push Western values on non-Western regions.[2] Critics highlight that AI systems are not neutral but are political tools replicating the ideas and power of those who make them.[3] Dominant technological companies have made their own charters and ethics guidelines, but these are not enforceable and just retain power in the hands of a few.[4] As of now, there is no binding international treaty to control AI technologies. Fragmented initiatives and global rivalries prevent a cohesive and comprehensive worldwide response.[5] The gap between fast-growing AI capabilities and the weak regulatory frameworks showcases the need for a binding, universal legal structure.

This article makes the case that the current trajectory, characterised by fragmentation, techno-nationalism, and regulatory capture, is unsustainable. It argues that only a binding international treaty on AI, rooted in inclusive, rights-based, and enforceable principles, can close the widening governance gap. Drawing on insights from global treaty-making processes, recent legal models such as the Council of Europe's Framework Convention on AI, and emerging scholarship from the Global South, this paper outlines both the necessity and feasibility of such a treaty. This article advances the debate by moving from descriptive mapping to a prescriptive treaty design that operationalises enforceability, adaptability, and multilateral legitimacy within a binding framework.

The article proceeds in seven parts. Section 2 identifies the systemic regulatory vacuum in AI governance and outlines the key global harms that justify a treaty-based approach. Section 3 critiques the current patchwork of national and industry-led regulation, emphasising the risks of fragmentation. Section 4 evaluates the role and limitations of soft law instruments and voluntary norms. Section 5 highlights the structural and political barriers to treaty-making. Section 6 proposes a way forward by drawing analogies from environmental and disarmament regimes and analysing the design of the Council of Europe AI Convention. Section 7 concludes with a call for inclusive, enforceable, and equitable global governance of AI.

The Case for International Regulation of AI

AI is now everywhere in the world. For example, facial recognition works in airports, train stations, and streets in many different countries. Language models use information from all over the globe. AI systems are able to traverse easily across national borders or different laws. But the regulations and laws that try to control AI are still different in each country. This disconnect is a significant problem. It makes it difficult to protect people's rights, fragments regulations, and causes global disparities.[6] Now that AI is more common and powerful, it is clear that there is an urgent need for international regulations.

AI technologies operate globally but remain shielded from legal scrutiny. Algorithms can recommend curated videos, track people via biometrics, or predict what someone will do next - no matter where the user lives. However, if something goes wrong structurally, there will be an international fallout. Each country tries to manage problems on its own, but this is not enough because AI problems can easily cross borders.[7] For example, fake news made by AI can spread quickly online and can cause confusion or harm in any country.[8] Also, when AI is used in policing, military, or border control, it brings up many questions around legal redress, responsibility, and fairness.[9]

The risks from AI are not just abstract ideas - they are happening in daily life in the form of algorithmic bias, black box decision-making, and surveillance. Facial recognition technology sometimes makes errors, especially with people of colour, at a disproportionate rate. This has led to people facing discrimination and being wrongfully arrested.[10]

Large language models are trained on internet datasets that are highly unmoderated. This means they perpetuate wrong or unfair stereotypes about gender, race, or culture on a wider scale.[11] The damage caused by this is significant within poor countries where vulnerable populations, particularly in the Global South, do not have strong courts or rights to protect themselves.

This reality has profound implications for human rights, democratic legitimacy, and public trust. AI systems, if unchecked, risk becoming tools of oppression rather than empowerment. Authoritarian regimes have already begun deploying AI to surveil citizens, censor dissent, and manipulate public discourse.[12] At the same time, democracies have failed to develop adequate safeguards, often prioritising innovation over inclusion.[13] In either case, the lack of binding multilateral norms has enabled a digital arms race, where the incentives to regulate lag behind the incentives to dominate.

AI also exacerbates existing global inequalities. The current model of AI development is resource-intensive, driven by tech conglomerates based in the Global North, and reliant on vast computing power and data infrastructure.[14] This has created a regulatory asymmetry: a handful of companies and states are effectively shaping the digital future for the rest of the world.[15] Countries in the Global South, already marginalised in international decision-making processes, are further disempowered by their lack of access to computational resources, open datasets, and AI talent.[16] The result is not only technological dependency but a new form of digital colonialism, where the Global South becomes a passive consumer of systems it had no role in designing.[17] This imbalance extends beyond technological dependency to normative exclusion, where countries with the least say in shaping AI governance must live under the rules made elsewhere.[18] As Monasterio Astobiza emphasises, these inequalities are compounded by extractive practices that offload the environmental and labour costs of AI onto the Global South.[19] Okolo similarly warns that infrastructural deficits and dependency on foreign corporations constrain equitable AI adoption in developing regions.[20]

In this context, legal certainty, equitable state responsibility, and interoperability are vital to ensuring that AI systems are developed and deployed in ways that respect fundamental rights and promote shared prosperity.

Legal certainty is essential to prevent a race to the bottom in regulatory standards. It offers clarity to developers, reassurance to users, and an everyday basis for redress. Equitable state responsibility ensures that all countries, not just powerful ones, have a role in setting the normative frameworks that govern AI. Interoperability, meanwhile, allows national legal regimes to align on minimum thresholds for transparency, safety, and rights protection without sacrificing sovereignty.[21]

Some efforts to address these concerns already exist. The Council of Europe's 2024 Framework Convention on Artificial Intelligence is a promising step, as it anchors AI regulation in human rights law and introduces binding obligations on signatories.[22] However, its limited geographical scope and opt-out clauses for national security threaten to weaken its global impact. Private companies also try to manage risks through their own voluntary measures, like creating internal ethics boards or carrying out technical audits. These can support formal regulation, but they cannot replace it.[23] Past experience with digital technologies shows that when powerful firms set their own regulations, the outcome usually harms the public. With weak oversight and strong commercial interests, the systems usually grow in a way that puts profit above privacy, safety, and truth.[24] The rise of AI is in danger of repeating the same mistakes. Voluntary ethical codes and industry-led actions, even if meant well, do not have the legal power or public trust needed to protect society. If the vacuum is left for companies to decide regulations on their own, it can result in opaque systems that affect society without any accountability.[25] This makes the need for binding legal rules stronger, as clearly defined obligations are far more effective than self-restraint. Soft law, which is based on non-binding rules, has already proven too weak to control a technology as big and fast-moving as AI.[26]

The international treaty would have a number of benefits compared to the current regime. To begin with, it would establish a level playing field, since it would require states to use shared standards in risk assessment, transparency, and redress. Second, it would help prevent regulatory arbitrage, where companies relocate to jurisdictions with weaker safeguards. Third, it would embed AI governance within existing international human rights frameworks, strengthening global mechanisms for accountability.

Fourth, it would formalise the inclusion of the Global South, civil society actors, and marginalised groups, ensuring that AI development reflects the interests of all communities—not merely those with power and resources.

While drafting a treaty on AI governance presents real technical and political hurdles, the cost of inaction is far greater. Without a global framework, AI will continue to amplify inequality, distort knowledge systems, and weaken democratic institutions. The creation of an international regulatory structure is no longer a distant ideal—it is a necessary foundation for ensuring that this technology develops under the rule of law, with justice and democratic oversight at its core.

The Governance Gap: Soft Global Norms, Regional Fragmentation, and Corporate Power

Today, the global approach to handling AI governance is fragmented. Different regions follow their own rules, many of which are soft law principles, while private actors also play a strong role. Many global initiatives have been created, but most are only aspirational and not legally binding. The 2021 UNESCO Recommendation on the Ethics of Artificial Intelligence was the first global normative framework accepted by 193 states.[27] It focused on human rights, transparency, and environmental sustainability, but it had no enforcement mechanisms.[28] The OECD's AI Principles 2019 also influenced discussions, but they are voluntary and dependent upon domestic/local action.[29] In 2023, the G7 started the Hiroshima Process on Generative AI in order to make common principles, but it did not give clear regulatory plans.[30] These efforts highlight normative convergence on values, but they cannot be enforced strongly, hence causing them to be inadequate. Because of the ineffectiveness of global frameworks, regions are now making their own AI regulation models with extraterritorial impacts.

This global normative softness contrasts with the emergence of regionally assertive, and often conflicting, legal frameworks. The European Union's AI Act, adopted in 2024, has been hailed as the most comprehensive attempt to regulate AI through a risk-based approach. It imposes strict obligations on high-risk applications and prohibits specific unacceptable uses of AI outright.

However, its extraterritorial scope—applying to any provider whose AI systems affect people within the EU—has stirred controversy, especially among developers in the Global South who lack the resources to ensure compliance.[31]

Meanwhile, China's approach to AI regulation is marked by state-centric control, political conformity, and real-time oversight. China's *Generative AI Measures* (2023) emphasise ideological alignment, content control, and cybersecurity, imposing strict pre-release review requirements and real-name user authentication.[32] In contrast, the United States continues to rely largely on sectoral guidance, executive orders, and voluntary frameworks. The Biden Administration's 2023 *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* encourages federal agencies to adopt AI risk management standards, but stops short of introducing binding legislation.[33] The lack of a comprehensive federal AI law in the U.S. reflects a broader hesitancy toward centralised regulatory intervention, in part due to fears of stifling innovation and private enterprise.

This patchwork of rules has enabled larger tech companies (private actors) to act as de facto regulators of AI. Companies like Google, OpenAI, and Microsoft set global technical and ethical standards via their design choices, deployment norms, and voluntary safety commitments. These companies also engage with standard-setting bodies like IEEE and ISO so as to shape the rules internally.[34] However, this dominance can lead to regulatory capture. Mira Lane and Stacey King state that corporate-led governance without democratic accountability allows profit goals to override public interest, particularly within areas like surveillance, content moderation, and biometric data use.[35] Voluntary AI governance plans like the OpenAI Charter and Google's AI Principles showcase ethical goals but do not have meaningful accountability, transparency, or independent checks. This leads to norm-making power being in company boardrooms, not in parliaments. These corporate-led systems do not work in isolation; instead, they support broader geopolitical agendas of powerful states.

The way the United States and China manage artificial intelligence also affects how global politics work. In the U.S., most of the work and progress in AI comes from private companies. The government steps in lightly, giving more space to businesses to grow. This shows that the U.S. supports open systems, new ideas, and competition in the market.

China does things differently, whereby it focuses more on strong government control and national stability. The Chinese government plays a bigger role in making decisions, and the system supports collective goals. Both countries want other parts of the world to follow their styles. The U.S. tries to spread its model through corporate growth and signing trade deals. China, however, uses regulatory tools like the Digital Silk Road to promote its style.[36] This struggle between their two systems is also seen in global forums. In forums like the International Telecommunication Union and the United Nations, both sides push their own views on how AI should be handled. But they often disagree - mainly on topics like privacy, human rights, and algorithmic transparency. These differences make it hard to reach an agreement, and this is why AI regulations remain fragmented in many regions.[37]

States in the Global South face problems because of this fragmented system. They do not have enough regulatory capacity or technological sovereignty, so they are pushed to pick between models that are incompatible. Many of them end up taking slightly modified versions of Western or Chinese models, or they just use imported AI technologies without updating their own regulations. This situation negatively impacts global inequality and inclusive governance. As Kathleen Siminyu explains, if actors in Africa, Latin America, and Asia are not part of making global AI rules, then the current governance will continue the colonial patterns of exclusion.[38]

In summary, the present state of AI governance is marked by a dangerous combination: soft law at the global level, hard borders through regional legislation, and unaccountable power concentrated in the private sector. These trends collectively weaken the prospects of a unified, inclusive, and enforceable international regime. Without stronger coordination mechanisms and a legally binding framework, the global AI landscape will remain fragmented, vulnerable to arbitrage, inequality, and geopolitical tension. A binding international framework is therefore not only desirable, but it is necessary to bridge these divides. The need for a binding international framework becomes more urgent when the legal, political, and structural consequences of fragmentation are made visible. These divisions are not theoretical. They shape how AI is regulated, contested, and experienced across different jurisdictions.

Consequences of Fragmentation: Legal, Political, and Structural

AI governance fragmentation is causing real harm and not just administrative inefficiencies. There is legal uncertainty about jurisdiction. Without shared rules, it is hard to pinpoint which laws apply to AI systems operating across countries, and this, resultingly, creates enforcement gaps. States may not want or may not be able to hold foreign AI developers responsible for harm caused within their region, especially if these developers are protected via favourable regulations or no treaties. These legal uncertainties lead to unequal protection for people depending on the jurisdictional strength.

This uncertainty leads to a fractured ecosystem of rights protection and accountability. In the EU, individuals harmed by AI may have recourse under GDPR or the AI Act's risk-based provisions. In contrast, residents of jurisdictions lacking strong legal protections, particularly in the Global South, often face significant barriers to redress, even when affected by the same technologies.[39] Algorithmic bias, surveillance, and disinformation become borderless harms, yet the protection against them is intensely local and unequally distributed.[40]

This gap allows regulatory arbitrage wherein companies utilising or developing AI can choose host countries with weak oversight to avoid human rights safeguards or ethical review. This creates an unfair system where legal loopholes guide innovation instead of ethics or substantial compliance.[41] Major technological firms utilise this loophole to concentrate power and wealth in a few countries while externalising social and ethical risks to countries with weaker institutions and laws.[42]

Nowhere are the political consequences more apparent than in the Global South. Without a binding international framework, these states are left to import foreign technologies and standards while remaining excluded from the processes that shape them.[43] As Gwagwa notes, this dynamic represents a form of digital neo-colonialism, where governance decisions made in Washington, Brussels, or Beijing disproportionately shape the digital futures of Nairobi, Dhaka, or La Paz.[44] This exclusion of the Global South from governance forums further entrenches power disparities.

Moreover, the absence of a dedicated global enforcement forum for AI—comparable to the WTO in trade or the ICC in international criminal law—means there is no institutional mechanism to adjudicate disputes, resolve normative clashes, or ensure accountability across borders.[45]

Finally, this fragmented governance structure suffers from a democratic deficit. Multilateral negotiations are often dominated by technologically advanced powers and private actors, while civil society, underrepresented nations, and vulnerable communities are marginalised. Without democratic legitimacy and global representation, the emerging AI governance order risks deepening geopolitical inequities and undermining the very legitimacy it seeks to create. These consequences expose the entrenched barriers that any attempt at building a binding global treaty must confront.

Barriers to a Binding Treaty: Power, Norms, and Institutional Gaps

A binding global treaty on AI is still complicated to achieve. The main reason is that countries have different structures, values, and geopolitical goals. The biggest divide comes from how major AI powers think about governance. Liberal democracies, such as the European Union and Canada, focus on human rights, transparency, and security when they create AI regulations. On the other hand, authoritarian states like China give more importance to techno-sovereignty, centralised control, and ideological alignment.[46] Because of these very different political and normative views, it becomes difficult to incorporate regulations and even harder to form a multilateral framework.[47]

Strategic competition in AI is getting stronger as differences become hardened. Countries now see AI not as a collective project for humanity, but more as a tool for power and advantage in global politics. The United States has placed export restrictions on advanced semiconductors to slow down China's rise in technology. In reply, China has restricted its export of rare earth minerals, which shows how this struggle is two-sided.[48] This rivalry has also reached international platforms. Initiatives like the G7 Hiroshima Process on Generative AI and the Global Partnership on AI (GPAI) have either excluded China or created divisions between countries. Because of this, their fairness and openness are being questioned.[49]

Private sector pushback, mainly from large tech companies in the Global North, makes it harder to set binding international regulations. These companies promote "agile governance," which focuses on flexibility and innovation instead of strict rules. Some flexibility is useful because technology changes quickly, but in many cases, this flexibility is used as an excuse to avoid legal obligations and to slow down the implementation of accountability systems.[50] Corporate lobbying has shaped governance models where companies take part in creating ethics guidelines while avoiding independent checks or control.[51] Scholars such as Bernd Carsten Stahl warn that this kind of self-regulation is not enough.[52]

Institutional fragmentation compounds the problem. As Huw Roberts and colleagues observe, the current international order is characterised by a "weak regime complex," a web of overlapping, non-binding initiatives lacking substantive alignment.[53] This enables forum-shopping, where both states and corporations choose regulatory frameworks that impose the least constraints. For example, a company may align with the OECD's voluntary AI principles while avoiding the legally binding provisions of the EU AI Act. Meanwhile, institutions such as the UN, OECD, and ITU operate in silos, often duplicating efforts without progressing toward harmonisation or enforcement.[54]

Another problem is that there is no clear definition of "artificial intelligence," i.e., is it a set of tools, a socio-technical system, or general-purpose infrastructure?[55] This ambiguity regarding foundation models makes it hard to draft future-proof legal obligations.[56] Although the EU and the Council of Europe have begun efforts to try to define AI formally, there is no global agreement yet. Making laws prematurely without clarity would risk hardcoding outdated or narrow interpretations into international law.[57]

Sovereignty concerns also inhibit consensus. Many states, especially in the Global South, are reluctant to cede regulatory authority to supranational frameworks.[58] They fear that binding treaty obligations may conflict with domestic priorities or be used to reinforce asymmetrical power structures. Past withdrawals from global accords, such as the U.S. exit from the Paris Agreement, have deepened scepticism about the reliability of international legal commitments.[59] For states with nascent AI ecosystems, binding global obligations may appear misaligned with their technical capacities and development agendas.

Soft law's limitations are already well established; the more pressing challenge is how these limitations complicate treaty negotiations. States that rely heavily on voluntary codes often resist binding commitments, while powerful private actors invoke the flexibility of soft law to avoid stricter obligations.[60] In this way, soft law does not merely fail to regulate; it actively entrenches barriers to consensus on a binding treaty.[61]

Overcoming these entrenched obstacles demands a recalibration of ambition. The objective should not be to draft a flawless treaty, but to pursue procedural realism: a framework that embraces normative pluralism, supports iterative revision, and foregrounds global legitimacy. As Seán Ó hÉigeartaigh and others have emphasised, international cooperation does not require total value alignment, only governance processes that are inclusive, dynamic, and attentive to underrepresented communities, particularly from the Global South.[62] Without such foundations, any treaty regime for AI will remain institutionally fragile, politically skewed, and ethically incomplete.

Lessons from Other Regimes: Arms Control, Climate, and Data Protection

Global efforts to regulate AI can draw inspiration from past treaty-based systems in arms control, climate governance, and data protection. Although these domains are vastly different in terms of technicalities, they also face the same issues, such as cross-border risks, scientific uncertainty, and uneven stakeholder power. The examples can be helpful to create tools and models required to formulate an AI treaty that will be reliable and binding.

The Paris Agreement showcases how tiered obligations and differentiated responsibilities can work well by enabling flexibility while also preserving global coherence. It allows the states to determine their own nationally determined contributions (NDCs), which are reviewed regularly and follow transparency guidelines. In AI governance, a similar treaty can also use graduated commitments based on each state's capacity and technological level. This keeps inclusiveness intact while providing strong regulation.[63] Kathleen Siminyu supports contextual AI governance and highlights the need for differentiated standards that match regional capacities, especially for the Global South.[64]

Another significant insight can come from the Biological Weapons Convention (BWC), which is a regime for high-risk dual-use technologies. Even though it does not have a strong enforcement body, it leverages a normative ban, peer pressure, and voluntary compliance to stop misuse.[65] In the context of AI, similar bans and transparency obligations can help prevent harm. António Guterres highlighted that it is extremely urgent to have enforceable limits on military AI usage in order to prevent humanitarian disasters.[66]

Mechanisms like peer review and compliance monitoring also offer vital procedural lessons. The BWC and Paris Agreement deploy reporting cycles, expert panels, and review conferences to foster accountability without overstepping sovereignty. A similar model could be adapted for AI through an independent AI Treaty Oversight Body that reviews implementation progress, assesses technical updates, and recommends corrective actions.[67] Scholars like Roberts and others advocate for strengthening the existing regime complex rather than erecting an entirely new global body, emphasising the role of polycentric institutions and expert coordination.[68]

Besides regimes focused on dual-use technologies, data protection laws help to understand how to incorporate individual rights into AI governance. The General Data Protection Regulation (GDPR) helps with enforcement and individual rights protection.[69] GDPR's focus on consent, purpose limitation, and data minimisation showcases how law can incorporate safeguards into technical systems.[70] Floridi and others argue that AI regulation should also go beyond ethics and consider functional oversight of specific risks in socio-technical systems.[71] GDPR's extraterritorial reach, applying to any actor utilising EU citizens' data, is a viable example for building global compliance and interoperability pressure into AI regulation.[72]

AI governance may also draw on the Martens Clause, an interpretive norm from humanitarian law. It affirms that human dignity and ethical constraints apply even in areas not explicitly regulated by treaty. Embedding such principles into an AI treaty would ensure ethical anchoring even amid rapidly evolving technological frontiers.[73] This echoes Arthur Gwagwa's call to embed the values of historically oppressed communities into AI governance frameworks, ensuring that justice transcends technical compliance.[74]

Adaptive oversight is also essential in these systems. Climate governance uses Intergovernmental Panel on Climate Change (IPCC) reports to adjust global policy.[75] GDPR utilises periodic revisions to incorporate changes when new threats appear. AI is evolving quickly; therefore, it needs a similar system that allows the governance frameworks to update standards, fix regulatory gaps, and respond to failures. An AI treaty should have a strong institutional mechanism to revise obligations as technology evolves. Lee, Floridi, and Taddeo suggest a stage-based model for AI regulation. It starts with pre-market risk assessments and moves to performance reviews, hence closely matching the adaptive trajectory.[76]

Ultimately, these regimes show that legal resilience, procedural adaptability, and normative clarity can coexist. For AI governance, the lesson is that enforceability must be balanced with adaptability, ensuring regulation remains effective without becoming obsolete.

Toward a Binding Global Treaty: Design Pathways and Proposals

A binding international treaty on AI should be clear, practical, and inclusive. One overarching body to control AI is not possible. The treaty can take inspiration from polycentric models like climate agreements and arms control. Polycentric governance refers to a decentralised model wherein different groups are working together towards shared rules and accountability.[77] For this treaty to work well, it should use modularity, focus on different risks, and make regulations that are sector-based. This helps the treaty deal with all the different types of technology and geopolitics. The treaty also needs compliance mechanisms and procedures that are adaptive. This makes sure the treaty stays effective as AI changes.

A risk-based framework, like the Paris Agreement, would let countries join according to what they can manage in terms of capacity, regulation, and risk. This would follow the principle of common but differentiated responsibilities and capabilities (CBDR+RC). In this way, developed and developing countries could take part without carrying the same load.[78] For specific sectors, annexes could be added, similar to the model of the Biological Weapons Convention. These annexes would provide regulations for sensitive areas such as biometric surveillance, lethal autonomous weapons, and algorithmic decisions in criminal justice.[79]

General obligations would apply to all states, but the sector-specific rules could be updated over time. This would make the system flexible while also keeping it enforceable.

The regulation of General-Purpose AI (GPAI) must be central to any binding framework.[80] These systems, such as large language models and multimodal transformers, are increasingly foundational to critical infrastructure, consumer platforms, and security applications, making their governance an urgent priority. These systems, capable of functioning across multiple domains, pose unique challenges due to their potential for rapid diffusion, dual-use capabilities, and unpredictability.[81] A binding treaty should require GPAI developers to undergo pre-deployment risk assessments, post-deployment audits, and ongoing impact disclosures.[82] This aligns with the layered, function-first regulatory logic proposed by Lee, Floridi, and Taddeo, who argue that risk should be determined by how AI functions in socio-technical systems, not merely by technological classification.[83] Given GPAI's geopolitical sensitivity, transparency obligations, and usage disclosures must also apply to states and militaries deploying such models.

A treaty negotiation process should be anchored in multilateral institutions with normative legitimacy and technical capacity. The United Nations High Commissioner for Human Rights (OHCHR) is well-positioned to foreground rights-based obligations, while UNESCO's experience with normative instruments can guide ethical harmonisation.[84] The International Telecommunication Union (ITU) can offer technical standardisation pathways, and the UN High-Level Advisory Body on AI can serve as a coordinating mechanism across sectors.[85] Given the failure of prior efforts to achieve consensus through exclusive clubs such as the G7 or OECD, inclusivity will be paramount. The exclusion of China from key forums like GPAI has already undermined its legitimacy, reinforcing the need for a truly global process.[86]

To support implementation and compliance, the treaty should establish an independent Global AI Observatory tasked with monitoring risks, evaluating compliance, and supporting capacity-building, particularly in the Global South. This proposal builds on the model of the Intergovernmental Panel on Climate Change (IPCC) and the Aarhus Convention's Compliance Committee.[87] The Observatory would aggregate data on AI harms, conduct public risk assessments, and issue annual compliance reviews.

It could also issue interpretative opinions, non-binding but norm-shaping, similar to how the Human Rights Committee shapes state conduct under the ICCPR. In turn, peer review mechanisms, modelled on those in the Paris Agreement, could create pressure for normative convergence without coercive enforcement.[88]

Legal harmonisation is one of the treaty's primary benefits. The current patchwork of soft norms and regional legislation results in regulatory arbitrage and uneven protection of fundamental rights. A binding treaty would consolidate a baseline of enforceable norms, reducing jurisdictional uncertainty and creating a more stable legal environment for AI development and deployment.[89] As noted by Vardiashvili, such harmonisation also empowers weaker jurisdictions by insulating them from external technological imposition and setting clear limits for private sector governance.[90]

Justice and equity must remain core treaty objectives. Current governance structures often replicate structural inequalities by excluding marginalised voices and privileging dominant actors. As Siminyu emphasises, community-based governance models that reflect local needs and knowledge systems are essential for meaningful legitimacy.[91] The treaty should enshrine procedural justice principles, including inclusive standard-setting, linguistic accessibility, and redistributive capacity-building funding. This could be operationalised through rotating regional consultations, multilingual negotiation documents, and a funding mechanism akin to the Green Climate Fund, tailored to AI capacity-building. As Monasterio Astobiza and others show, the Global South bears disproportionate costs of AI development through the extractivism of resources and hidden labour, even as governance dependency entrenches Northern dominance.[92] Okolo adds that infrastructural deficits and reliance on Big Tech risk a form of "algorithmic colonisation", unless Southern governments secure fair representation in global governance fora.[93] Vijayakumar further distils these concerns into the four "E"s: extractivism, exclusion, ethnocentrism, and enforcement, arguing that without enforceable obligations, Southern states remain unable to check exploitative practices.[94] This aligns with calls by Roberts and others for democratising international rule-making to avoid governance by a technocratic elite.[95]

Finally, embedding treaty goals within the framework of the Sustainable Development Goals (SDGs) offers a way to integrate AI governance into broader development priorities. AI has the potential to accelerate progress on health, education, and environmental goals, but only if deployed equitably and accountably. A binding treaty should promote "equitable innovation," ensuring that AI enhances, rather than undermines, human dignity and sustainable development.[96] Embedding enforceable obligations tied to SDG outcomes, such as equitable access to AI in education and health, can ensure the treaty has a measurable, development-oriented impact. This approach reframes the treaty not as a constraint on innovation, but as its ethical foundation.

Conclusion

Global AI governance stands at an inflexion point, where the gap between rapidly evolving technological risks and inadequate international regulation has become untenable. Despite the proliferation of ethical guidelines, voluntary principles, and soft law instruments across multilateral and private sectors, their collective impact remains inconsistent, largely unenforceable, and often symbolic. As AI systems embed themselves across critical domains such as healthcare, law enforcement, labour, and warfare with transboundary effects and existential implications, the continued reliance on fragmented, non-binding mechanisms is no longer defensible. The moment demands a decisive shift: from voluntary pledges to binding, accountable global governance.

A compelling case for a binding international treaty on AI rests on three interdependent pillars: enforceability, adaptability, and multilateral legitimacy. Enforceability is essential to prevent states and corporations from bypassing or selectively interpreting norms to serve narrow interests. As demonstrated by the current patchwork of soft regulation rife with forum shopping, jurisdictional gaps, and unequal protections, only binding international law can anchor procedural justice, human rights, and rule-of-law principles across borders.[97] Adaptability, far from being antithetical to law, can be embedded through tiered obligations, function-based risk frameworks, and periodic review mechanisms that ensure responsiveness to emerging harms.[98]

Multilateral legitimacy requires that such a treaty be negotiated inclusively within trusted global institutions, particularly the United Nations, and reflect diverse legal traditions, regional priorities, and socio-technical realities.[99]

Crucially, any such treaty must confront the deep asymmetries embedded in current governance arrangements. This article advances the academic debate by moving beyond descriptive accounts of governance fragmentation to a prescriptive treaty design. Its originality lies in synthesising comparative lessons from arms control, climate accords, and data protection into a normative architecture for AI governance. In doing so, it bridges theoretical discourse with actionable institutional proposals, showing how enforceability, adaptability, and legitimacy can be embedded in a binding treaty. Today, AI development and global standard-setting remain dominated by a small group of powerful states and private tech conglomerates, while much of the Global South is left as a rule-taker. This dynamic risks replicating historical patterns of domination and dependency, now cloaked in algorithmic design and digital infrastructure rather than territorial control. As scholars such as Siminyu and Gwagwa emphasise, inclusive AI governance must be grounded in participatory norm-making, community-based knowledge systems, and equitable access to AI resources.[100] Civil society, academia, and historically marginalised voices must play a central role in shaping not only the content of international regulation but the very processes through which it is crafted.

This moment must not be treated as a pause before regulation, but as the window for action. A just, inclusive, and enforceable global treaty on AI is not a distant aspiration; it is a present necessity. Legal imagination, institutional courage, and moral clarity must converge to confront the structural imbalances of the current regime. The responsibility lies with all, but most of all with those who hold the diplomatic, political, and normative capital to lead.

References

[1] Arthur Gwagwa, "Resisting Colonialism: Why AI Systems Must Embed the Values of the Historically Oppressed," in *AI Governance Essay Collection*, Chatham House, 2024.

[2] Tamar Vardiashvili, "The Council of Europe's Framework Convention on Artificial Intelligence," *AIRE Centre Report*, September 2024

[3] Ibid.

[4] Mira Lane and Stacey King, "Common Goals and Cooperation – Towards Multi-Stakeholderism in AI," in *AI Governance Essay Collection*, Chatham House, 2024.

[5] António Guterres, "Remarks at the UK AI Safety Summit," United Nations, November 1, 2023.

[6] Floridi, Luciano, and Josh Cowls. "A Unified Framework of Five Principles for AI in Society." *Harvard Data Science Review*, no. 1 (2019).

[7] Alex Krasodomski, "Introduction – The Need to Future-Proof AI Governance," in *AI Governance Essay Collection*, Chatham House, 2024.

[8] António Guterres, "Remarks at the UK AI Safety Summit," United Nations, November 1, 2023.

[9] Mira Lane and Stacey King, "Common Goals and Cooperation – Towards Multi-Stakeholderism in AI," in *AI Governance Essay Collection*, Chatham House, 2024.

[10] Kathleen Siminyu, "Community-Based AI," in *AI Governance Essay Collection*, Chatham House, 2024.

[11] Alek Tarkowski, "Open Source and the Democratization of AI," in *AI Governance Essay Collection*, Chatham House, 2024.

[12] Feldstein, Steven. *The Rise of Digital Repression: How Technology is Reshaping Power, Politics, and Resistance*. Oxford University Press, 2021.

[13] Alex Krasodomski, "Introduction – The Need to Future-Proof AI Governance."

[14] A.H. Niazi, "Digital Proxies of Humanity," in *Digital Pathways for Humanity*, 2024.

[15] Rahul Tongia, "Why Regulating AI Can Be Surprisingly Straightforward," World Economic Forum, 2024.

[16] Uyangoda and Rajendran, "AI and the Global South: Towards Equity by Design," in *AI Governance Essay Collection*, Chatham House, 2024.

[17] Ibid.

[18] Mohamed, Shazeda, et al. "Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence." *Philosophy & Technology* 33, no. 4 (2020): 659–684.

[19] Aníbal Monasterio Astobiza and others, 'Ethical Governance of AI in the Global South: A Human Rights Approach to Responsible Use of AI' (2022) 81 *Proceedings* 136

[20] Chinasa T Okolo, 'AI in the Global South: Opportunities and Challenges Towards More Inclusive Governance' (Brookings Institution, 29 June 2021)

[21] Thomas Schneider, "Regulating AI and Digital Technologies – What the New Council of Europe Convention Can Contribute," 2024.

[22] Tamar Vardiashvili, "The Council of Europe's Framework Convention on Artificial Intelligence," *AIRE Centre Report*, September 2024.

[23] Mira Lane and Stacey King, "Common Goals and Cooperation."

[24] Tom Wheeler, *"The Three Challenges of AI Regulation"*, Brookings Institution, June 15, 2023.

[25] Ibid.

[26] Shriya Yarlagadda, "Global AI Regulation and the Asilomar Analogy," *Harvard International Review*, 2024.

[27] UNESCO. *Recommendation on the Ethics of Artificial Intelligence*. Paris: UNESCO, November 2021.

[28] UNESCO, *Recommendation on the Ethics of Artificial Intelligence*, 2021.

[29] OECD, *Recommendation of the Council on Artificial Intelligence*, 2019.

[30] G7, *Hiroshima AI Process: International Guiding Principles for Organizations Developing Advanced AI Systems*, 2023.

[31] Tamar Vardiashvili, "The Council of Europe's Framework Convention on Artificial Intelligence," AIRE Centre Report, September 2024.

[32] Laney Zhang, "China: Generative AI Measures Finalized," *Law Library of Congress*, July 19, 2023.

[33] Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, White House, 30 October 2023.

[34] Alek Tarkowski, "Open Source and the Democratization of AI," Chatham House, 2024.

[35] Mira Lane and Stacey King, "Common Goals and Cooperation – Towards Multi-Stakeholderism in AI," Chatham House, 2024

[36] Alex Krasodomski, "Introduction—The Need to Future-Proof AI Governance," Chatham House, 2024.
[37] Thomas Schneider, "Regulating AI and Digital Technologies – What the New Council of Europe Convention Can Contribute," Chatham House, 2024.
[38] Kathleen Siminyu, "Community-Based AI," Chatham House, 2024.
[39] Thomas Schneider, "Regulating AI and Digital Technologies," Chatham House, 2024.
[40] António Guterres, "Remarks at the UK AI Safety Summit," United Nations, November 1, 2023.
[41] Rahul Tongia, "Why Regulating AI Can Be Surprisingly Straightforward," World Economic Forum, 2024
[42] Mira Lane and Stacey King, "Towards Multi-Stakeholderism in AI," Chatham House, 2024.
[43] Kathleen Siminyu, "Community-Based AI," Chatham House, 2024.
[44] Arthur Gwagwa, "Resisting Colonialism – Why AI Systems Must Embed the Values of the Historically Oppressed," Chatham House, 2024.
[45] Alex Krasodomski, "Introduction—The Need to Future-Proof AI Governance," Chatham House, 2024.
[46] Thomas Schneider, "Regulating AI and Digital Technologies—What the New Council of Europe Convention Can Contribute," *Chatham House*, 2024.
[47] Huw Roberts et al., "Global AI Governance: Barriers and Pathways Forward," *International Affairs* 100, no. 3 (2024): 540
[48] Mira Lane and Stacey King, "Common Goals and Cooperation—Towards Multi-Stakeholderism in AI," in *AI Governance Essay Collection* (Chatham House, 2024).
[49] Huw Roberts et al., "Global AI Governance," 542.
[50] Lee, Floridi, and Taddeo, "Regulating Artificial Intelligence: What, When and How?," in *The Oxford Handbook of AI Governance*, ed. Markus D. Dubber et al. (Oxford University Press, 2023), 570–572.
[51] Ibid., 573
[52] Stahl, Bernd Carsten. *Artificial Intelligence for a Better Future: An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies.* SpringerBriefs in Research and Innovation Governance. Cham: Springer Cham, 2021.
[53] Huw Roberts et al., "Global AI Governance," 541.
[54] Roberts et al., "Global AI Governance," 543
[55] Alex Krasodomski, "Introduction—The Need to Future-Proof AI Governance," in *AI Governance Essay Collection* (Chatham House, 2024)
[56] Lee, Floridi, and Taddeo, "Regulating Artificial Intelligence," 566.
[57] Roberts et al., "Global AI Governance," 544.
[58] Seán S. ÓhÉigeartaigh et al., "Overcoming Barriers to Cross-cultural Cooperation in AI Ethics and Governance," *Philosophy & Technology* 33, no. 3 (2020): 455–457.
[59] Ibid., 460.
[60] António Guterres, "Remarks at the UK AI Safety Summit," *United Nations*, November 1, 2023
[61] Tom Wheeler, "Artificial intelligence is another reason for a new digital agency," *Brookings*, July 2023.
[62] ÓhÉigeartaigh et al., "Overcoming Barriers," 462–463.
[63] Kathleen Siminyu, "Community-Based AI," *AI Governance Essay Collection*, Chatham House, 2024
[64] Ibid.
[65] Huw Roberts et al., "Global AI Governance: Barriers and Pathways Forward," *International Affairs* 100, no. 3 (2024): 741–763.
[66] António Guterres, "Remarks at the UK AI Safety Summit," United Nations, November 1, 2023.
[67] Roberts et al., "Global AI Governance," 747–750.
[68] Ibid., 751.
[69] Regulation (EU) 2016/679, General Data Protection Regulation, Art. 3.
[70] Bernd Carsten Stahl, *Artificial Intelligence for a Better Future* (Springer, 2021), 65–66.
[71] Lee, Floridi, and Taddeo, "Regulating Artificial Intelligence: What, When and How?" in *The Oxford Handbook of AI Governance*, 2023.
[72] Ibid.
[73] Arthur Gwagwa, "Resisting Colonialism—Why AI Systems Must Embed the Values of the Historically Oppressed," *AI Governance Essay Collection*, Chatham House, 2024.
[74] Ibid.
[75] Roberts et al., "Global AI Governance," 753.
[76] Lee, Floridi, and Taddeo, "Regulating Artificial Intelligence," 7–9.

[77] Jordan, Andrew, and Dave Huitema. 2023. "Polycentric Governance." In *Routledge Handbook of Environmental Policy*, edited by H. Jörgens, C. Knill, and Y. Steinebach, 55–67. Abingdon, UK: Taylor & Francis.
[78] Kathleen Siminyu, "Community-Based AI," AI Governance Essay Collection, Chatham House, 2024.
[79] Tamar Vardiashvili, "The Council of Europe's Framework Convention on Artificial Intelligence," AIRE Centre Report, 2024.
[80] European Commission, Directorate-General for Communications Networks, Content and Technology. "General-Purpose AI Code of Practice." *Shaping Europe's Digital Future*.
[81] Lee, Floridi, and Taddeo, "Regulating Artificial Intelligence: What, When and How?" in The Oxford Handbook of AI Governance (2023).
[82] António Guterres, "Remarks at the UK AI Safety Summit," United Nations, November 1, 2023.
[83] Lee, Floridi, and Taddeo, "Regulating Artificial Intelligence."
[84] Thomas Schneider, "Regulating AI and Digital Technologies—What the New Council of Europe Convention Can Contribute," Chatham House, 2024.
[85] Mira Lane and Stacey King, "Common Goals and Cooperation—Towards Multi-Stakeholderism in AI," AI Governance Essay Collection, Chatham House, 2024.
[86] Huw Roberts et al., "Global AI Governance: Barriers and Pathways Forward," International Affairs 100, no. 3 (2024): 713–730.
[87] Ibid.
[88] Ibid.
[89] Rahul Tongia, "Why Regulating AI Can Be Surprisingly Straightforward, When Teamed with Eternal Vigilance," World Economic Forum, 2024.
[90] Vardiashvili, "The Council of Europe's Framework Convention."
[91] Siminyu, "Community-Based AI."
[92] Monasterio Astobiza and others, 'Ethical Governance of AI in the Global South'
[93] Okolo, 'AI in the Global South'.
[94] Anupama Vijayakumar, *AI Ethics for the Global South: Perspectives, Practicalities, and India's Role* (RIS Discussion Paper No 296, October 2024)
[95] Roberts et al., "Global AI Governance."
[96] Alex Krasodomski, "Introduction—The Need to Future-Proof AI Governance," AI Governance Essay Collection, Chatham House, 2024.
[97] Huw Roberts et al., "Global AI Governance: Barriers and Pathways Forward," *International Affairs* 100, no. 3 (2024): 713–730.
[98] Lee, Floridi, and Taddeo, "Regulating Artificial Intelligence: What, When and How?" in *The Oxford Handbook of AI Governance* (2023).
[99] Thomas Schneider, "Regulating AI and Digital Technologies—What the New Council of Europe Convention Can Contribute," Chatham House, 2024.
[100] Kathleen Siminyu, "Community-Based AI," and Arthur Gwagwa, "Resisting Colonialism—Why AI Systems Must Embed the Values of the Historically Oppressed," *AI Governance Essay Collection*, Chatham House, 2024.