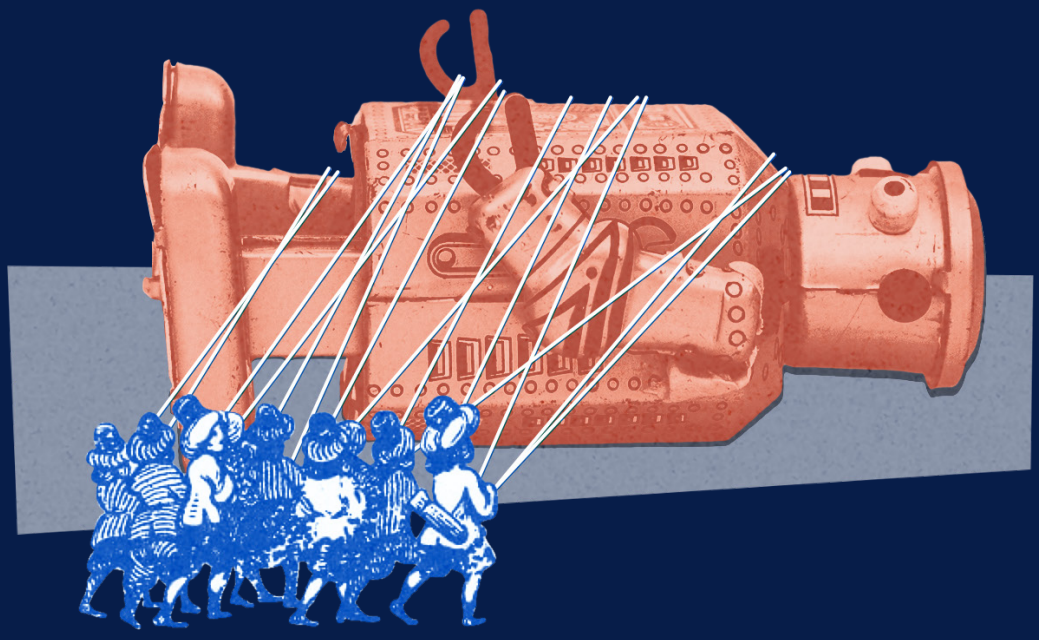
PARTNERSHIP ON AI

AI Agents and Global Governance

Analyzing Foundational Legal, Policy, and Accountability Tools

Talita Dias



Contents

Executive Summary

Artificial intelligence (AI) agents are advanced computer programs that can set their own goals, make plans, and act in ways that are not entirely predictable. Unlike chatbots, they are anticipated to work with a high degree of autonomy to tackle increasingly complex tasks. While not yet widespread, some AI agents can take action in external environments by interacting with different systems, such as by calling tools or writing and running computer code.

These capabilities are expected to make AI agents useful for different tasks — for example, streamlining administrative tasks or optimizing critical systems. But the ability to take autonomous action in external environments and directly affect software or hardware in unexpected ways could amplify well-known AI risks and create new ones. In a hyperconnected world, many of these risks could cross national borders or concern the international community as a whole. Potentially significant global risks include:

1. **Cross-border harms,** such as electoral interference and critical infrastructure disruption.
2. **Human rights impacts,** such as privacy breaches and free speech limitations.

This brief explores how these two sets of risks could be managed through anticipatory global governance leveraging foundational tools that are non-AI specific in nature and universal in scope, as action-taking AI agents become more widespread around the world. These tools are:

1. International law.
2. Non-binding global norms.
3. Global accountability mechanisms.

As a starting point for governing AI agents globally in anticipation of potential risks, this brief identifies where these tools work well, where they fall short, and what actions by different global stakeholders — including governments, companies, and civil society — can strengthen them.

GOVERNMENTS

Governments already have legal obligations that apply to AI agents, including:

1. To respect other states' sovereign rights to use, regulate, and host those technologies in their territories.
2. To refrain from using agents for foreign interference.
3. To behave with due diligence to prevent cross-border harms.
4. To respect and protect human rights.

COMPANIES

Companies, while not directly bound by these rules, stand to gain from observing international law when designing, developing, and deploying AI agents. Respect for international law is at the heart of environmental, social, and corporate governance (ESG) principles, companies' existing cybersecurity commitments, and their responsibility to respect human rights, shaping corporate relationships with governments, investors, and users.

ALL STAKEHOLDERS

All stakeholders — governments, companies, civil society, and individuals — can seek global accountability through established legal, normative, and institutional channels. They could also work together to address pressing gaps in the current global governance ecosystem as it applies to AI agents, including potential liability gaps; a lack of global solutions to conflicts of laws; and decentralized enforcement.

Governments, companies, civil society, and individuals can seek global accountability through established legal, normative, and institutional channels.

Introduction

Artificial intelligence (AI) agents can be defined as "[c]omputer software systems capable of creating context specific plans in non-deterministic environments."[1] Built with a combination of reinforcement learning (RL) and large language models (LLMs), new AI agents are marked by greater autonomy and pursue increasingly complex goals.[2] While not yet widespread, some AI agents can take action and have a causal impact on the external environment by interacting with various tools or systems, including online; this includes AI agents that can call tools via application programming interfaces (API) and those that can write and execute computer code with software development kits (SDKs).[A, 3]

AI agents are expected to have many useful applications for governments, companies, and individuals around the world. For example, they could be used for computer coding across sectors, to streamline administrative tasks, and to optimize a wide range of critical infrastructure systems from energy to transport grids. At the same time, action-taking AI agents, and especially those that can execute their own computer code,[4] may directly affect software and/or hardware infrastructure with which they interface in complex and unpredictable ways. This could potentially exacerbate well-documented AI risks, such as information manipulation, anthropomorphism, malfunction, and privacy breaches, as well as job displacement and power concentration.[5] AI agents may also give rise to new risks such as self-preservation and loss of control.[6]

AI agents are still in the early stages of development and adoption, so it remains to be seen how they will be used by various stakeholders in different sectors.[B] But there is increasing interest in those products and a growing body of literature on the topic. Published papers have explored key features, benefits, and risks, as well as technical, legal, and policy interventions to mitigate the potential impacts of AI agents.[7] Yet little has been said to date on the global governance of AI agents.[C]

This brief contributes to filling this gap by offering guidance on anticipatory global governance for AI agents, leveraging foundational legal, policy, and accountability tools that are non-AI specific in nature and universal in scope:

1. International law.
2. Non-binding global norms.
3. Global accountability mechanisms, particularly multilateral institutions that are open to universal membership, such as the United Nations (UN).[D]

These global governance tools are the foundation of many AI-specific frameworks, such as the G7 Hiroshima Process Guiding Principles and Code of Conduct and the Organisation for Economic Co-operation and Development (OECD)'s AI Principles; yet, they have at times been overlooked in discussions about AI governance. They are also inclusive of all states, developed and developing, and, in some cases, other stakeholders.

A These are level 3 to level 5 agents in line with PAI's Failure Detection Monitoring framework, Srikumar et al., "Prioritizing Real-Time Failure Detection in AI Agents."

B For early examples, see OpenAI, "Introducing Operator;" Anthropic, "Claude Opus 4.1;" and Manus AI.

C See Kaprayoon et al, 2025, identifying this as a pressing policy question.

D On the importance of global AI governance grounded in these tools, see UN General Assembly (UNGA), "Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development," especially paras 5, 6(a), 13.

In reviewing these tools, this brief focuses on two of the most pressing global risks that are likely to arise as action-taking AI agents are developed and deployed at scale: **cross-border harms** and **human rights impacts**. It also identifies gaps in how foundational global governance tools manage those risks, outlining options for governments, companies, international organizations, individuals, and groups to strengthen those tools as AI agents continue to be developed and adopted worldwide.[E]

FIGURE 1. **Examples of cross-border harms and human rights impacts potentially arising from the deployment of action-taking AI agents**

CROSS-BORDER HARMS	EXAMPLES	HUMAN RIGHTS IMPACTS
Information manipulation	Deepfakes, social engineering	Privacy breaches
AI-powered cyberattacks	Phishing, ransomware, polymorphic malware	Hateful and deceptive content
Adversarial LLM attacks	Data poisoning, prompt injection	Bias and discrimination
Automation failure	Malfunction in AI-powered industrial control system	Job displacement

E On the related question of what should be internationalized in AI governance or what outcomes we should expect from global institutions when it comes to AI governance, see Dennis et al., "What Should Be Internationalised in AI Governance?" and Smith & Crampton, "Global governance: Goals and Lessons for AI."

This brief responds to a core concern expressed by all UN member states in the Global Digital Compact, as it applies to AI agents:

> *[t]he need for a balanced, inclusive and risk-based approach to the governance of [AI], with the full and equal representation of all countries, especially developing countries, and the meaningful participation of all stakeholders [...] in full respect of international law, including international human rights law, and taking into account other relevant frameworks [...].*[8]

This work is important for three main reasons. First, as with other digital technologies, the risks and impacts of AI agents could transcend national borders and affect a range of global stakeholders. For example, a bad computer code executed by an AI agent in one country might affect computer software and hardware in multiple countries via the internet, including in such critical sectors as health care, energy, and finance. Even when risks and impacts are restricted to the domestic context, they may concern the international community as a whole, as in the protection of human rights.

Second, while the conversation about the global governance of AI has focused on developing new, AI-specific rules, norms, or institutions,[9] foundational, non-AI specific global governance tools already govern AI and AI agents globally, just as they govern other digital technologies.[10] International law binds states — and in some circumstances, non-state actors — regardless of which tools or technologies are used in their activities.[F, 11] Where binding international rules are lacking, global norms and policy frameworks can drive responsible state and corporate behavior online and offline. And certain global mechanisms are available to affected stakeholders, including states, companies, individuals, and groups, to hold irresponsible actors

F On how existing law applies to new technologies, see generally Lessig, Code 2.0.

to account. Together, international law, non-binding global norms, and global accountability mechanisms provide the foundation upon which more specific governance tools for AI and AI agents can be built.

Third, AI-specific governance tools are still nascent and scarce at the global level; most frameworks apply domestically, regionally, or cross-regionally but not universally.[G] Notably, binding, AI-specific international rules are not yet in force.[H] Likewise, there are no AI-specific accountability mechanisms that include and are open to all states, despite many proposals for the establishment of global AI institutions, for example.[I] Global AI-specific policy frameworks such as the UNESCO AI Recommendation and the Global Digital Compact are anchored in, and refer back to, foundational rules of international law, global norms, and global accountability mechanisms, particularly those found in the UN Charter and international human rights frameworks.[12] Developing AI-specific global governance tools can be a resource-intensive process, especially in today's challenging geopolitical environment, while foundational global governance tools are already in place. Moreover, if not crafted with sufficient care, new rules, norms, or mechanisms can undermine well-established protections. This is why foundational global governance tools should be the starting point for thinking about how to govern AI agents — and other AI technologies — globally and inclusively. They provide a tried and tested common language that has helped different global stakeholders, including states, companies, individuals, and civil society organizations, navigate through some of the world's greatest challenges, from war and famine to the climate crisis and the digital revolution.

This paper is divided into three sections. The first two explain how foundational global governance tools (i.e., international law, global norms, and global accountability mechanisms) are already in place to address the cross-border harms and human rights impacts that might arise from the design, development, and deployment of AI agents. The concluding section identifies gaps in how those issues are currently governed and makes recommendations as to how various global stakeholders might address them.

G For an analysis of AI-specific governance tools, see Ifayemi et al., "Decoding AI Governance: A Toolkit for Navigating Evolving Norms, Standards, and Rules."

H The first international treaty on AI, the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, might enter into force soon for its states parties upon ratification by five signatories. See Council of Europe, Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, 2024, Article 30(3). However, this treaty is not open to universal membership.

I For a summary of those proposals, see Maas & Villalobos, "International AI institutions: A literature review of models, examples, and proposals."

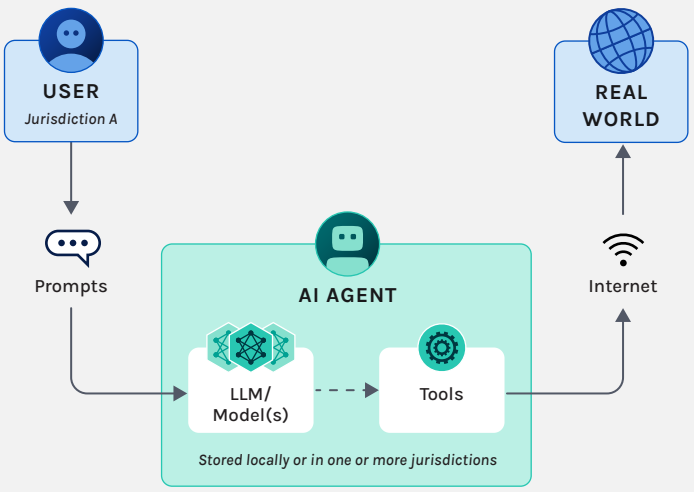
Foundational global governance tools should be the starting point for thinking about how to govern AI agents — and other AI technologies — globally and inclusively.

1. Cross-border Harms

AI agents are connected to the internet and other networks to gather key contextual information and, in some cases, to take action by calling tools or executing computer code on a variety of applications, such as web browsers. This means that their actions might have effects across borders and result in cross-border harms affecting government institutions, companies, individuals, and groups around the world.

For example, AI agents developed and deployed in one country could generate and disseminate content via online platforms that might interfere in electoral processes abroad. Likewise, malicious actors in one state might exploit security vulnerabilities in AI agents that control or optimize critical infrastructure in another state, causing significant disruption to the delivery of public services, financial loss, and even physical or mental harm. Furthermore, AI agents — and, in particular, multi-agent systems that manage cross-border operations such as international trade and financial transactions — may fail, with high-stakes consequences across jurisdictions.

FIGURE 2. **Representation of the cross-jurisdictional scope of AI agents' actions via the internet**



International law

Different rules of international law that govern cross-border state activity apply to the design, development, and deployment of AI agents, impacting not only the behavior of states but also companies, individuals, and groups building or interacting with this technology. First and foremost, states have the sovereign right[13] to design, develop, and use AI agents in a manner consistent with international law. They also have "jurisdiction," or the right to regulate, adjudicate, and implement policies[14] when AI agents are designed, developed, and deployed by other actors in their territory. This right covers:

a. The physical infrastructure or hardware needed to train and run AI agents, such as chips, GPUs, computer clusters, and data centers.

b. The computer devices and infrastructure that connect those agents to the internet and other networks, such as servers, routers, and cables.

c. The different actors (e.g., developers, deployers, and end-users).

d. Activities across the AI agent value chain, including those that involve datasets, model training, and inference.[J]

Therefore, sovereignty means not only the power to build "sovereign AI agents" with local data, models, compute, and workforce but also to govern the AI agent value chain domestically.

Jurisdiction might extend extraterritorially, for example, over the cross-border effects of domestic activities and the activities of companies registered or domiciled domestically.[15] As a result, various domestic laws that might apply simultaneously to AI agents could clash, giving rise to conflict of law challenges. These challenges are more likely to arise as the AI agent value chain becomes more global and more AI regulations are adopted worldwide. These potential problems are compounded in the case of multi-agent systems, given the multitude of systems and applications across jurisdictions that could be implicated.[16]

State sovereignty entails respect for the rights of other states, including persons or property abroad.[17] This means that states must not use AI agents in a way that causes physical damage or loss of functionality in the territory of another state, even if only private entities are affected.[K, 18] For example, a state must not use an AI agent to damage or disable industrial control systems or Internet of Things devices in another state. In the same vein, states must not use AI agents to undermine another state's governmental functions, including the choice of a political, economic, social, and cultural system, and the formulation of foreign policy, irrespective of any physical or functional damage.[19] For instance, a state must not use an AI agent to undermine electoral processes or e-government services abroad. Sovereignty is not merely about prohibiting misuse or malfunction of AI agents; it also protects AI agents themselves from the actions of foreign nations insofar as those technologies are designed, developed, and deployed consistently with international law by public or private entities.

A significant number of malicious activities online are carried out by nation-states or individuals and groups closely linked to them.[20] There is also increasing evidence that AI agents can be leveraged for a range of malicious purposes — from data exfiltration to online scams to malware development.[21] So it is to be expected that AI agents will be built or exploited by state actors for malicious purposes.[L] Likewise, as governments start building or using AI agents for government services, it is possible to foresee that AI agent malfunctions will affect not only government agencies but also companies and individuals that rely on those services. Therefore, it is in everyone's interest to ensure respect for state sovereignty in the age of AI agents.

A related obligation is non-intervention: States must not interfere in the internal or external affairs of other states through coercive means to manipulate or undermine their sovereign

J See, in the cyber context, Heintschel von Heinegg, "Territorial Sovereignty and Neutrality in Cyberspace."

K Note that, for some states, any unauthorized intrusion into another state's system would violate sovereignty (see e.g., African Union, "Common African Position on the Application of International Law to the Use of Information and Communication Technologies in Cyberspace," para 16). For For other states, e.g., the United Kingdom (UK), sovereignty does not create binding obligations in the cyber context (see UK Attorney General. "International Law in Future Frontiers.")

L On AI agents' security vulnerabilities, see Chen & Royce, "AI Agents Are Here. So Are the Threats;" Zenity Labs. "AgentFlayer: 0Click Exploit Methods."

will,[22] including by using AI agents or by interfering with AI agents that perform governmental functions. There is growing evidence that LLMs trained with RL (of the kind used by new AI agents) can resort to manipulation, threats, or other coercive methods to gain power over operators and achieve their aims.[23] There is also evidence to suggest that AI agents are predisposed toward escalation when used for political strategy.[24] For those reasons, states should be particularly careful when using AI agents in their international relations.[25] While the obligation of non-intervention does not directly bind companies, they should also seek to ensure that their AI agents do not enable foreign interference. Respect for the rule of law is an overarching benchmark for ESG principles, and companies that disregard international rules can face government, user, and investor backlash.[26]

Neither sovereignty nor non-intervention requires intention or knowledge in an anthropomorphic sense to engage state responsibility.[27] For some scholars, states may be liable for breaches of sovereignty even if the harm is unexpected or unforeseeable.[28] However, this question remains underexplored and requires clarification as AI agents increase in autonomy, efficacy, and complexity, yet remain unpredictable, like other LLM-powered technologies. (For comparison, in the context of outer space, states have absolute or strict liability for damage caused by their space objects on the surface of the Earth or to aircraft in flight.)[29]

For a state to be held responsible for violations of sovereignty or non-intervention, the behavior in question needs to be attributed to that state.[30] According to existing principles of attribution, a human being (whether a government official or a private actor operating on behalf of the state) must have performed the relevant action.[31] This means that the actions of "public" AI agents cannot at present be automatically attributed to states. Changing the law to automatically attribute the acts of AI agents to states would mean that states would be strictly liable for the harms caused by this technology, whether or not these harms are foreseeable. This is a consequential policy choice that requires careful consideration.[M]

M See Lior, "AI Entities as AI Agents: Artificial Intelligence Liability and the AI Respondeat Superior Analogy," arguing for the application of agency law to 'AI entities' in the domestic context.

Nevertheless, even when the actions of AI agents cannot be attributed to states, or when these technologies are designed, developed, or deployed by non-state actors, states still have an obligation to exercise due diligence in their own territory to protect the rights of other states.[32] They also have an obligation to prevent significant cross-border harms to persons, property, or the environment, or in any event to minimize the risk thereof, irrespective of the source of the harm.[33] These "due diligence" obligations have gained particular traction in the environmental context, given the borderless nature of pollutants and other sources of environmental harm.[34] There is some debate as to whether they apply to non-physical harms caused by digital technologies, including AI.[35] But a majority of states that have spoken out on this matter agree that due diligence obligations apply whether the harm occurred offline or online.[36]

Diligent state behavior can prevent a number of cross-border harms arising from the design, development, and deployment of AI agents by public and private actors, including agent misuse and malfunction. Companies do not have binding due diligence obligations under

international law. But as the primary developers and deployers of AI agents, they are expected to behave with a sufficient degree of due diligence in this context.

Due diligence obligations are breached when foreseeable harm materializes, and the state should have known that the activity in question carried the risk of causing such harm.[37] Many potential harms caused by AI agents are now foreseeable, given a growing body of scholarly work and publicly available evidence.[38] However, states need to consider whether and to what extent they want to be held responsible for unforeseeable harms caused by AI agents developed or deployed by others. A related question is what threshold of autonomy and unforeseeability is globally acceptable, beyond which states should not permit AI agents to be developed or deployed.

Non-binding global norms

UN member states have recognized a number of non-binding norms of responsible state behaviour in the context of information and communications technologies (ICTs).[39] These norms flesh out foundational rules of international law, fill gaps, and help bridge legal disagreements in contentious areas. A general norm stipulates that states should not knowingly allow their territory to be used for internationally wrongful acts employing ICTs. And specific norms are in place to protect critical infrastructure. For example, one stipulates that states should take appropriate measures to protect their critical infrastructure from ICT threats. An additional norm asks states to respond to appropriate requests for assistance from another state whose critical infrastructure is subject to malicious ICT acts. These norms apply to AI agents insofar as they are connected to the internet or to other ICTs, such as private networks.

States need to consider whether and to what extent they want to be held responsible for unforeseeable harms caused by AI agents developed or deployed by others.

Grounded in the UN norms of responsible state behavior, the Paris Call for Trust and Security in Cyberspace applies not only to states but also to companies and civil society organizations.[40] The Call has been signed by over 95 governments; nearly 350 international, civil society, and public sector organizations; and more than 600 private sector entities around the globe, including some of the largest AI developers and deployers.[41] Among other things, it calls upon its signatories to prevent malicious cyber activities against individuals, critical infrastructure, and the public core of the internet; foreign influence operations; and the development of harmful ICT tools.

Similarly, the Cybersecurity Tech Accord has been signed by a number of technology companies that have committed to work together to protect users online, including by "design[ing], develop[ing], and deliver[ing] products and services that prioritize security, privacy, integrity and reliability, and in turn reduce the likelihood, frequency, exploitability, and severity of vulnerabilities."[42] Notably, Tech Accord signatories have reaffirmed the applicability of the Accord to AI technologies.[43]

Although not the focus of this brief, it is important to note that AI-specific policy frameworks such as the G7 Hiroshima Process Guiding Principles and Code of Conduct and the OECD

AI Principles apply to AI agents alongside the non-AI specific norms discussed above. Both sets of norms — specific and non-AI specific — are complementary in that they focus on different risks or harms and recommend different measures, yet are grounded in the fundamental principle that technology should benefit individuals and society. In this sense, both are mutually reinforcing and should be strengthened.

Global accountability mechanisms

To seek compliance with those rules and norms as well as redress for any transboundary harms caused by AI agents, states can take unfriendly but lawful measures of "retorsion," such as severing diplomatic relations or adopting lawful economic sanctions against other states.[44] If injured by a violation of international law resulting from the deployment of AI agents, they can also take countermeasures.[45] Those are measures that would normally breach international law but are justified to drive compliance in response to a prior breach.[46]Examples include economic sanctions that would normally violate existing trade agreements or the suspension of preexisting treaty obligations. Countermeasures have been used with some success to tackle such varied global challenges as acts of aggression, systematic human rights abuses, and terrorism.[47] States can also resort to diplomatic (e.g., negotiation and conciliation) or legal means (i.e., arbitration and courts, such as the International Court of Justice) to settle international disputes regarding AI agents. Whether or not a specific dispute is in place, the UN has a number of distinct bodies that can recommend specific measures to govern AI agents globally. This includes the Secretary-General; the Office for Digital and Emerging Technologies (ODET); the newly adopted Independent International Scientific Panel on AI and Global Dialogue on AI Governance;[48] the General Assembly and its various committees; and the International Law Commission.

2. Human Rights Impacts

AI agents have numerous beneficial applications, many of which can help advance internationally recognized human rights. For example, AI agents that provide personalized teaching can foster the right to education, whereas AI agents used for medical diagnosis can help individuals enjoy their right to the highest attainable standard of physical and mental health. Many of these beneficial applications can also help achieve several of the UN Sustainable Development Goals in both developed and developing countries. For example, AI agents that optimize energy efficiency and automate extreme weather predictions can help combat climate change in line with Goal 13. Similarly, agents used in agriculture to enable precision farming and to optimize crop yields can help end hunger, achieve food security, improve nutrition, and promote sustainable agriculture in line with Goal 2.

At the same time, it is well-documented that the design, development, and deployment of AI can interfere with a range of internationally recognized human rights.[49] For example, AI algorithms have been known to reproduce biases captured in their training data. Were this

to occur in public sectors such as education or at the workplace, the use of AI would likely violate the right to non-discrimination under international human rights law.[50] AI has also been known to generate false, misleading, or hateful content, which could infringe upon the rights to freedom of opinion and expression, including the rights to receive and impart information freely.[51] Likewise, the increasing demand for mass training data might lead to the unauthorized collection of personal data in violation of the human right to privacy,[52] as well as exploitation of data enrichment workers[53] in breach of the right to enjoy just and favorable conditions of work.[54] In the longer term, there is fear that AI might displace workers in at least some sectors, given the technology's lower costs and increasing efficiency gains, which could threaten the very right to work.[55]

AI agents may exacerbate these and further human rights risks. Notably, AI agents seem to have a propensity to pursue instrumental and self-preserving goals by resorting to manipulation, threats, and other control tactics.[56] These features could increase AI agents' likelihood of coercing users and of generating false, misleading, or hateful content.[57] Access to, and interface with, different applications by AI agents could significantly increase the chances of privacy breaches via surveillance and information leaks.[58] At the same time, the ability of AI agents to act autonomously in external environments has raised concerns about a potential liability gap.[N]

N See, in the domestic context, Toner et al., "Through the Chat Window and Into the Real World: Preparing for AI Agents;" Kolt, "Governing AI Agents;" O'Keefe et al., "Law-Following AI: Designing AI Agents to Obey Human Laws."

International law

International human rights treaties such as the International Covenant on Civil and Political Rights (ICCPR) and the International Covenant on Economic, Social and Cultural Rights (ICESCR) recognize a range of human rights. Like other treaties, the ICCPR and ICESCR bind only the states that ratified them. But several civil and political rights listed in the ICCPR and in the Universal Declaration of Human Rights have acquired the status of customary international law; on this basis, they apply universally.[59]

Internationally recognized human rights give rise to both negative and positive obligations that are binding on states. Negative human rights obligations require states not to interfere with human rights. Conversely, positive obligations require states to protect human rights from interference by others, including states and non-state actors, by exercising due diligence.[60] Positive human rights obligations are triggered not only by intentional actions but also by general conditions in society that may give rise to direct, foreseeable, and preventable threats to human rights, including accidents or massive cyberattacks.[61] None of these harms need to have materialized: Because positive obligations are preventative in nature, they arise in the face of reasonably foreseeable risks or threats to human rights.[62]

This means that states cannot be held responsible under international human rights law for actions taken by AI agents that are completely unforeseeable. Nevertheless, states have an obligation to exercise due diligence to prevent or mitigate any foreseeable harms or risks resulting from the design, development, or deployment of those technologies. Whether or not a specific harm or risk is foreseeable depends on how much scientific knowledge is out

there.[63] Insofar as risks are known or foreseeable, states are required to put in place policies to prevent, or at least mitigate them, such as by adopting human rights risk assessment frameworks.[64] Given that AI agents are being primarily designed, developed, and deployed by private entities, states may also need to require a certain level of transparency and access to relevant documentation held by those entities. To prevent and mitigate human rights impacts, states may also need to conduct or mandate appropriate tests and evaluations of AI agents before allowing them to be placed on the market.

Some human rights treaties, like the ICCPR, apply only within a state's jurisdiction.[65] While jurisdiction is primarily territorial, it extends extraterritorially in some circumstances. This includes the extraterritorial effects of the activities of private companies incorporated or domiciled in a state's territory, such as in the case of privacy violations arising from the use or export of surveillance technology.[66] Jurisdiction arguably extends to all instances wherein the state exercises power or effective control over the enjoyment of human rights, such as when it controls a data center used for an AI agent's training or inferences.[67]

Insofar as risks are known or foreseeable, states are required to put in place policies to prevent, or at least mitigate them, such as by adopting human rights risk assessment frameworks.

Non-binding global norms

While businesses are not bound per se by international human rights law — and as such, have no binding international obligations to respect or protect human rights — the UN Guiding Principles on Business and Human Rights (UNGPs) recommend that companies respect human rights.[68] Respect for human rights is also a key factor in ESG frameworks, guiding investment decisions and consumer choices that can help shape the future of a company.[69] This means that companies should not infringe upon human rights and should address any adverse human rights impacts arising from their operations, products, or services, including those involving AI agents.[70] This corporate human rights responsibility is a "global standard of expected conduct" for all business enterprises irrespective of where they operate.[71]

There is a dedicated UN Human Rights Council working group on the topic of business and human rights. This group has recently issued its first report on AI, which focuses on applying the UNGPs to AI procurement and deployment by states and non-developers.[72] At the request of the UN Human Rights Council, the Office of the UN High Commissioner for Human Rights has issued a recent report on how the UNGPs — including the concept of human rights due diligence — apply to technology companies, particularly those designing, developing, and deploying AI.[73] Although its membership is not global, the OECD has also issued guidance on corporate human rights due diligence in the context of AI.[74]

Global accountability mechanisms

There are different avenues for accountability when it comes to violations of internationally recognized human rights. In the context of the ICCPR and the ICESCR, any member state can bring a human rights complaint against another state to the Human Rights Committee

and the Committee on Economic, Social and Cultural Rights, respectively.[75] Affected individuals and groups can do the same with respect to states that have accepted the individual complaint procedure.[76] In addition, both committees can issue general comments on broader issues, in particular, how specific human rights apply in different contexts. States are also required to submit regular reports to each committee on the application of human rights in their jurisdiction.[77] For its part, the UN Human Rights Council can hear complaints about consistent patterns of gross, reliably attested human rights violations; these may be submitted by any individual, group of individuals, or nongovernmental organization.[78] In addition, the Council can appoint special rapporteurs and working groups that can issue reports on distinct human rights issues such as privacy and freedom of expression.[79] States can also take measures of "retorsion", or countermeasures against other states, to induce compliance with international human rights law. International human rights issues, including those involving AI agents, can also be adjudicated before regional and domestic courts.

Conclusion: Gaps and Recommendations

Foundational global governance tools already address some of the global challenges expected to arise from the widespread use of AI agents, including cross-border harms and human rights impacts. Important areas not explored in this piece but ripe for future research include the impact of AI agents on international peace and security, including in the context of military operations, international trade, investment, and the environment.

For states, international law lays out general prohibitions, permissions, and requirements that apply by default to AI agents as well as to other AI technologies such as foundation models. Violations arising from the design, development, or deployment of AI agents can trigger state responsibility insofar as the harms or risks in question were foreseeable. These rules are complemented by norms that flesh out how states should behave responsibly in the ICT context, including when AI agents are used to perform tasks online and especially when they risk affecting critical infrastructure or the core of the internet. Respect for international law is also crucial to protect companies and individuals against cross-border harms caused by states or non-state actors deploying or targeting AI agents.

Companies are not directly bound by the rules of international law discussed in this brief. However, upholding these rules is an overarching ESG benchmark that can benefit corporate developers and deployers in their relationships with governments, investors, and users. At the same time, global norms such as the Paris Call for Trust and Security in Cyberspace and the UNGPs are directly addressed to companies and apply when they design, develop, and deploy AI agents.

States, companies, individuals, and affected groups can use global channels to seek

accountability for the cross-border harms and human rights impacts anticipated by the widespread deployment of AI agents. In particular, international organizations such as the UN and its subsidiary bodies can offer important recommendations on what constitutes responsible state and corporate behaviour in this context.

FIGURE 3. **Summary of foundational global governance tools that apply to the anticipated cross-border harms and human rights impacts of action-taking AI agents**

	CROSS-BORDER HARMS	HUMAN RIGHTS IMPACTS
INTERNATIONAL LAW	Sovereignty, non-intervention, and due diligence obligations	International Bill of Rights (ICCPR, ICESCR, and Universal Declaration of Human Rights)
NON-BINDING GLOBAL NORMS	UN GGE Norms of Responsible State Behaviour; Paris Call for Peace and Security in Cyberspace; Cybersecurity Tech Accord	UN Guiding Principles on Business and Human Rights
GLOBAL ACCOUNTABILITY MECHANISMS	Retorsion, countermeasures, international courts and tribunals, the UN and its subsidiary bodies (e.g., General Assembly, Security Council, Independent Scientific Panel on AI, and Global Dialogue on AI Governance)	UN Human Rights Committee, Committee on Economic, Social and Cultural Rights, UN Human Rights Council, retorsion, countermeasures

Still, these foundational tools are far from perfect. One particular challenge arising from AI agents' ability to take action in the world is the potential for significant liability gaps. As seen earlier, the actions of public AI agents cannot be automatically attributed to the states that deploy these technologies. Moreover, state responsibility usually arises in the face of foreseeable risks or harms. While there is some scholarly work on the potential risks emerging from AI agents' distinct capabilities, more work is needed to test hypotheses and uncover presently unknown risks. LLMs' inherent unpredictability also means that certain risks might not be known until they have materialized. In addition, companies lack binding obligations and therefore face no legal liability under international law.[80] This means that whether they may be considered liable for the anticipated risks or harms of AI agents will depend on individual states' domestic law. While various legislatures around the world are currently debating the topic of liability for AI harms, no nation has yet adopted an AI or AI agent-specific liability framework.[81] The question of corporate liability for the risks or harms caused by AI agents falls back on general doctrines of liability such as tort, agency, or product liability law.

As AI agents become more widespread and begin to take actions across jurisdictions, the likelihood of conflicts between domestic laws and regulations will inevitably increase. This challenge may be compounded in the case of multi-agent systems, which interact with other AI agents and can therefore affect a larger set of tools and systems located in different jurisdictions. Yet international law offers no solution to such conflicts of laws.[82] With AI's legal and policy landscape still so fragmented globally,[83] conflicts of laws are likely to increase as AI agents become more common.

A systemic, more fundamental challenge facing the current international order is the lack of centralized enforcement: At present, no global police force or court can automatically enforce international rules when they are breached.[84] The closest to this is the UN Security Council, which can adopt binding resolutions, establish courts, and mandate enforcement action.[85] But the veto power held by the Council's permanent members, coupled with persistent ideological divides among them, can be paralyzing. This is especially true in a geopolitical environment currently marked by increasing backlash against multilateralism. Moreover, the jurisdiction of international courts and tribunals such as the International Court of Justice depends on state consent,[86] which is often lacking. This mostly leaves the enforcement of international rules in the hands of states. To be sure, this challenge manifests across all international affairs and is hardly unique to AI or AI agents. Still, as autonomous, action-taking AI agents start to be deployed worldwide, legal and factual disputes among global stakeholders are likely to grow. Leaving their resolution entirely in the hands of individual states might not always lead to the best outcome.

FIGURE 4. **Challenges of applying foundational global governance tools to action-taking AI agents**

Global stakeholders have an important role to play in addressing these gaps and leveraging international law, norms, and accountability mechanisms to ensure that AI agents are designed, developed, and deployed responsibly around the world.

States	States could consider calling out other states and non-state actors when confronted with the design, development, or deployment of AI agents in a manner inconsistent with international law or norms. They could seek to induce compliance through economic and political incentives, or impose sanctions as a last resort. For example, states could incorporate safety, security, and other relevant standards into trade and investment deals. In addition, they could consider transposing international rules and norms into domestic law, regulation, or policy frameworks. These could include provisions seeking to ensure that: • AI agents, whether used by public or private actors, are appropriately tested and evaluated for relevant attributes (such as safety and security.)[87] • Real-time vulnerability- and failure-detection systems are put in place to catch relevant risks before they materialize.[0, 88] • A sufficient level of human oversight or review of AI agents' actions is available, especially for high-stakes decisions (such as via a "disable" or "override" function).[89] • There is transparency about the deployment of AI agents (such as via notifications or agent IDs), as well as relevant documentation about the technology used.[90] • Individuals or groups affected by the actions of AI agents have access to an effective remedy, such as by appropriately resourcing domestic courts and fostering the use of insurance by AI agent developers and deployers.[91]

O See also Srikumar et al., Prioritizing Real-Time Failure Detection in AI Agents, proposing a failure detection framework for AI agents.

Companies	Companies, including developers and deployers of AI agents, could consider: • Developing internal policies and governance mechanisms specifically for AI agents, such as risk assessment and human rights impact-assessment frameworks; • Testing and evaluating AI agent models and applications before and after release; • Putting in place real-time failure-detection mechanisms along with appropriate human oversight or review consistent with international law and norms;[92] and • Establishing grievance mechanisms to ensure that individuals and groups affected by their use of AI agents have access to an effective remedy;[93] this could, for example, take the form of independent oversight mechanisms to hear individual complaints[94] or build on existing AI incident-reporting mechanisms.[95]
States, companies, individuals and groups	States, companies, individuals, and groups can bring documented human rights risks and harms arising from the design, development, and deployment of AI agents to the attention of the UN Human Rights Council, the Human Rights Committee, and the Committee on Economic, Social and Cultural Rights.
The UN	The UN could consider appointing a UN Special Rapporteur on AI and Human Rights.[96] The Special Rapporteur could be asked to prepare a report on the human rights risks arising from the design, development, and deployment of AI agents, as well as the practical measures that states and companies could take to mitigate those risks, including via agent-specific human rights due diligence. The UN could also leverage the Independent International Scientific Panel on AI and the Global Dialogue on AI Governance to identify the challenges and opportunities presented by AI agents. Indeed, the Panel — an independent body of 40 multidisciplinary experts — has been tasked with issuing "evidence-based scientific assessments synthesizing and analysing existing research related to the opportunities, risks and impacts of artificial intelligence."[97] For its part, the Global Dialogue is a multistakeholder forum for discussions of AI governance questions, including "[r]espect for and protection and promotion of human rights in the field of artificial intelligence" and "[t]he transparency, accountability and robust human oversight of artificial intelligence systems in a manner that complies with international law."[98]

All Stakeholders	All stakeholders, including states, companies, international organizations, and civil society, could invest in additional research on both the risks and opportunities of AI agents, using outlets such as the International AI Safety Report,[99] the International AI Summit Series,[100] and the International Network of AI Safety Institutes.[101] They could also work together to flesh out how international law and norms apply to different use cases of AI agents, including in critical sectors such as health care, national security, energy, finance, and transportation. Given the role of companies in designing, developing, and deploying AI agents, it would be particularly useful to better understand how the Paris Call for Peace and Security in Cyberspace, the Cybersecurity Tech Accord, and the UNGPs apply to AI agents. To foster consensus and common understandings, it could be helpful to focus initially on low-hanging fruit or red lines around prohibited, permitted, required, and recommended behaviors by various stakeholders.[P, 102] Notwithstanding some efforts to develop international security guidelines for AI agents[103] — and AI red lines more generally[104] — none of these initiatives are ostensibly grounded in international law or norms. While these initiatives should be global and inclusive in nature, stakeholders could draw from existing AI-specific policy frameworks, such as the G7 Hiroshima Process Guiding Principles and Code of Conduct and the OECD AI Principles.

P For an example, see Oxford Institute for Ethics Law and Armed Conflict, "The Oxford Process on International Law Protections in Cyberspace."

There is no shortage of first-order principles to guide the behavior of states and companies as they prepare for the widespread deployment of AI agents. Global accountability channels are also available to affected stakeholders. Yet significant challenges remain. Some stem from the distinct capabilities of AI agents, and others are systemic to the global order as it currently exists. Notably, liability gaps might arise from the ability of certain AI agents to act autonomously in external environments, including actions that, thanks to the internet and other ICTs, might cross national borders. More fundamentally, few avenues are currently available to enforce compliance with international rules. Bridging these and other gaps between principles and action will require creative thinking and concerted efforts. This brief is intended to help start the conversation about the global governance of AI agents. Much work still needs to be done across policy, academic, and industry circles to ensure that AI agents are designed, developed, and deployed in ways that benefit the international community as a whole.

ACKNOWLEDGMENTS

I am grateful to my colleagues in the Policy Team, Stephanie Ifayemi and Jacob Pratt, as well as to our CEO, Rebecca Finlay, for their comments on earlier versions of this brief. I would also like to thank Duncan Hollis, Harriet Moyhihan, and Kubo Mačák for their thorough reviews and thoughtful comments, as well as the members of PAI's Policy Steering Committee for their guidance on this work. Finally, I am thankful to Peter Hall for his work in copyediting this piece.

Endnotes

1 Mitchell, Margaret, et al. "Fully Autonomous AI Agents Should Not be Developed." arXiv preprint arXiv:2502.02649 (2025). For other definitions see Kaprayoon, Jam, et al. "AI Agent Governance: A Field Guide." IAPS Report (2025); NVIDIA. "Glossary: AI Agents" (2025); Gabriel, Iason, et al. "The Ethics of Advanced AI Assistants." arXiv preprint arXiv:2404.16244 (2024).

2 Mitchell et al, 2025; Kasirzadeh, Atoosa, and Gabriel, Iason. "Characterizing AI gents for Alignment and Governance." arXiv preprint arXiv:2504.21848 (2025); Chan, Alan, et al. "Harms from Increasingly Agentic Algorithmic Systems." Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (2023).

3 Mitchell et al, 2025; Khoo, Shaun, Foo, Jessica and Lee, Roy Ka-Wei. "Agentic Risk & Capability Framework." Singapore GovTech (2025); Deshpande, Chinmay, and Joshi, Ruchika. "AI Agents In Focus Technical and Policy Considerations." CDT Brief (2025).

4 Mitchell et al, 2025; Deshpande and Joshi, 2025.

5 Mitchell et al, 2025; Chan et al, 2023; Akbulut, Canfer, et al. "All Too Human? Mapping and Mitigating the Risk from Anthropomorphic AI". Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (2024).

6 Bengio, Yoshua, et al. "International AI Safety Report" (2025).

7 E.g., Srikumar et al (2025); World Economic Forum in collaboration with Capgemini. "Navigating the AI Frontier: A Primer on the Evolution and Impact of AI Agents." White Paper (2024); Mitchell et al, 2025; Kasirzadeh and Gabriel, 2025; Chan et al, 2023; Kaprayoon et al, 2025; Deshpande and Joshi, 2025; Soder et al, 2025.

8 UNGA. "Global Digital Compact." A/79/L.2 (2024), paras 50 and 52.

9 E.g., Krasodomski, Alex, et al. "Artificial intelligence and the challenge for global governance." Chatham House (2024); Ho, Lewis, et al. "International Institutions for Advanced AI." arXiv preprint arXiv:2307.04699 2023; Cass-Beggs, Duncan, et al. "Framework Convention on Global AI Challenges." CIGI Discussion Paper (2024); Baker, James N. "International law and advanced AI: exploring the levers for 'hard' control." Institute for Law & AI Blog Post (2024); OpenAI. "Governance of superintelligence." (2023).

10 Dias, Talita and Sagoo, Rashmin. "AI Governance in the Age of Uncertainty: International Law as a Starting Point." Just Security (2022).

11 Akande, Dapo et al. "Drawing the Cyber Baseline: The Applicability of Existing International Law to the Governance of Information and Communication Technologies." 99 International Law Studies 4 (2022).

12 UNESCO. "Recommendation on the Ethics of Artificial Intelligence." (2024), pp. 5-7, paras 8(a), 9, 11, 13, 18, 19, 28, 32, 42, 46, 52, 57, 61, 63, 65, 69, 72, 73, 83, 107, 120, 121, 131(a), 133, 141; UNGA. "Global Digital Compact." A/79/L.2 (2024), para 52.

13 See UN Charter (1945), Article 2(1); Permanent Court of Arbitration (PCA). Island of Palmas case. 2 RIAA 829 (1928), p. 838; Cyber Law Toolkit. "Sovereignty" (2025); Moynihan, Harriet. "The Application of International Law to State Cyberattacks: Sovereignty and Non-Intervention." Chatham House Research Paper (2019).

14 International Court of Justice (ICJ). Lotus case. Serie A - No 70, Judgment of September, 7th, 1927, pp. 16-20.

15 Kamminga, Menno T. "Extraterritoriality." Max Planck Encyclopedias of International Law (2020).

16 See Hammond, Lewis, et al. "Multi-Agent Risks from Advanced AI." Cooperative AI Foundation Technical Report #1 (2025).

17 UNGA. "Declaration on Principles of International Law concerning Friendly Relations and Cooperation among States in accordance with the Charter of the United Nations." Res 2625 (XXV) (1970), p. 124.

18 Cyber Law Toolkit. "Sovereignty" (2025); Schmitt, Michael N (ed.). *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations* (Cambridge University Press, 2017), Rule 4, paras 11-13.

19 Cyber Law Toolkit. "Sovereignty" (2025); Schmitt, 2017, Rule 4, paras 15-17.

20 Microsoft. "Microsoft Digital Defense Report" (2024), Chapter 1.

21 Lupinacci, Matteo, et al. "The Dark Side of LLMs Agent-based Attacks for Complete Computer Takeover." arXiv preprint arXiv:2507.06850v1 (2025); Sjouwerman, Stu. "How 'Agentic AI' will drive the future of malware." SC Media (2025); Rubin, Sam. "Unit 42 Develops Agentic AI Attack Framework." Palo Alto Networks (2025); Claburn, Thomas. "Voice-enabled AI agents can automate everything, even your phone scams", The Register (2025).

22 ICJ, Nicaragua Case, Merits, Judgment. I.C.J. Reports 1986, p. 14, 1986, paras 203-205.

23 Chan et al, 2023; Anthropic. "Agentic Misalignment: How LLMs could be insider threats" (2025).

24 Jensen, Benjamin. "The Troubling Truth About How AI Agents Act in a Crisis." Foreign Policy (2025).

25 E.g., Albrecht, Eduardo. "AI Agents in Global Governance: Digital Representation for Unheard Voices" Columbia SIPA (2025).

26 Lee, Paul. "The Rule of Law and investor approaches to ESG." Bingham Centre for the Rule of Law Discussion Paper (2022).

27 Schmitt, 2017, Rule 4, para 25.

28 See Schmitt, 2017, Rule 4, para 25.

29 Convention on International Liability for Damage Caused by Space Objects (1972), Article II.

30 UNGA. "Responsibility of States for Internationally Wrongful Acts." Res 56/83 (2001), Articles 2(a) and 4-11.

31 International Law Commission (ILC). "Draft articles on Responsibility of States for Internationally Wrongful Acts, with commentaries." (2001), Commentary to Article 2, para 5.

32 ICJ. Corfu Channel Case. Judgment of April 9th, 1949, I.C. J. Reports 1949 p. 4, 1949, p. 22; PCA. Island of Palmas case. 2 RIAA 829 (1928), p. 839.

33 UNGA. "Prevention of Transboundary Harm from Hazardous Activities", Res A/56/10, 2001, Article 3.

34 Koivurova, Timmo, and Singh, Krittika. "Due Diligence." Max Planck Encyclopedias of International Law (2022); International Law Commission, Ridings, Penelope. "Due Diligence in International Law." UNGA Res A/79/10, Annex II (2025).

35 Coco, Antonio and Dias, Talita. "'Handle with care': due diligence obligations in the employment of AI technologies." In Geiß, Robin and Lahmann, Henning. Research Handbook on Warfare and Artificial Intelligence (Edward Elgar, 2024); Moynihan, Harriet. "Unpacking due diligence in cyberspace." Journal of Cyber Policy 8 (1): 4–25 (2023); Coco, Antonio and Dias, Talita. "'Cyber Due Diligence': A Patchwork of Protective Obligations in International Law." 32 European Journal of International Law (2021).

36 Cyber Law Toolkit. "Due Diligence" (2025).

37 UNGA. "Prevention of Transboundary Harm from Hazardous Activities", Res A/56/10, 2001, Commentary to Article 3; paras 11 and 18; Coco and Dias, 2021.

38 E.g., Anthropic. "System Card: Claude Opus 4 & Claude Sonnet 4" (2025); Mitchell et al, 2025; Chan et al, 2023.

39 UNGA. "Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security." Res A/70/174 (2015), para 13; UNGA. "Developments in the field of information and telecommunications in the context of international security." Res 70/237 (2015).

40 Government of France. "Paris Call for Trust and Security in Cyberspace" (2018).

41 Government of Canada. "Paris Call for Trust and Security in Cyberspace" (2024).

42 Tech Accord. "Cybersecurity Tech Accord" (2018).

43 Tech Accord. "The Cybersecurity Tech Accord in the Age of AI: A new series exploring challenges and opportunities for industry" (2024).

44 Giegerich, Thomas. "Retorsion". Max Planck Encyclopedias of International Law (2020).

45 UNGA. "Responsibility of States for Internationally Wrongful Acts." Res 56/83 (2001), Article 49.

46 Dias, Talita. "Countermeasures in international law and their role in cyberspace." Chatham House Research paper (2024).

47 Dawidowicz, Martin. Third-Party Countermeasures in International Law (Cambridge University Press, 2017); Drezner, Daniel. "Are Economic Sanctions Effective Foreign Policy Tools?" TuftsNow (2024).

48 UN General Assembly. "Terms of reference and modalities for the establishment and functioning of the Independent International Scientific Panel on Artificial Intelligence and the Global Dialogue on ArtificialIntelligence Governance." A/79/L.118 (2025).

49 Jones, Kate. "AI governance and human rights: Resetting the relationship." Chatham House Research Paper (2023).

50 International Covenant on Civil and Political Rights (ICCPR) (1966), Article 26.

51 ICCPR, Article 19; UNGA, "Disinformation and freedom of opinion and expression." Res A/HRC/47/25 (2021), especially para 81.

52 ICCPR, Article 17; UNGA. "Artificial intelligence and privacy, and children's privacy." A/HRC/46/37 (2021), paras 16-26.

53 Jindal, Sonam. "Protecting AI's Essential Workers: A Pathway to Responsible Data Enrichment Practices." PAI Blog Post (2024).

54 International Covenant on Economic, Social and Cultural Rights (ICESCR) (1966), Article 7.

55 ICESCR, Article 6. See also the Committee on Economic, Social and Cultural Rights. "The Right to Work: General comment No. 18". E/C.12/GC/18 (2006), paras 10, 25-26.

56 Chan et al, 2023; Bengio, Yoshua, et al. "Superintelligent Agents Pose Catastrophic Risks: Can Scientist AI Offer a Safer Path?." arXiv preprint arXiv:2502.15657 (2025).

57 Akbulut et al, 2024.

58 Mitchell et al, 2025.

59 Schabas, William A. *The Customary International Law of Human Rights* (Oxford University Press, 2021).

60 UN Human Rights Committee (HRC). General comment no. 31. CCPR/C/21/Rev.1/Add. 13 (2004), paras 6 and 8.

61 HRC. General Comment No. 36. CCPR/C/GC/36 (2019), para 26.

62 HRC. General Comment No. 36. CCPR/C/GC/36 (2019), para 7.

63 Urs, Priya, et al. "The International Law Protections against Cyber Operations: Targeting the Healthcare Sector." Oxford Institute for Ethics, Law and Armed Conflict Report (2023).

64 US State Department. "Risk Management Profile for Artificial Intelligence and Human Rights" (2024); Council of Europe. "Methodology for the Risk and Impact Assessment of Artificial Intelligence Systems from the Point of View of Human Rights, Democracy and the Rule of Law (HUDERIA Methodology)." (2024).

65 ICCPR, Article 2(1).

66 UNGA, "The right to privacy in the digital age : report of the United Nations High Commissioner for Human Rights." A/HRC/39/29 (2018), para 25.

67 HRC. General Comment No. 36, CCPR/C/GC/36 (2019), para 63; Shany, Yuval. "Taking Universality Seriously: A Functional Approach to Extraterritoriality in International Human Rights Law." 7 The Law & Ethics of Human Rights (2013). Urs et al (2023).

68 UN. Guiding Principles on Business and Human Rights (2011), Principle 11.

69 See UN Human Rights Council. "Investors, environmental, social and governance approaches and human rights." A/HRC/56/55 (2024);

70 UN. Guiding Principles on Business and Human Rights (2011), Principle 11.

71 UN. Guiding Principles on Business and Human Rights (2011), Commentary to Principle 11.

72 UN Human Rights Council. "Artificial intelligence procurement and deployment: ensuring alignment with the Guiding Principles on Business and Human Rights." Res A/HRC/59/53 (2025).

73 UN Human Rights Council. "Practical application of the Guiding Principles on Business and Human Rights to the activities of technology companies, including activities relating to artificial intelligence." Res A/HRC/59/32 (2025).

74 OECD. "OECD Business and Finance Outlook: Human rights due diligence through responsible AI" (2021).

75 UN Human Rights. "Complaints about human rights violations" (2025).

76 UN Human Rights. "Individual Communications" (2025).

77 UN Human Rights. "Introduction to the Committee" (2025); UN Human Rights. "Introduction to the Committee" (2025).

78 UN Human Rights. "Human Rights Council Complaint Procedure" (2025).

79 UN Human Rights. "Special Procedures of the Human Rights Council" (2025).

80 See Chesterman, Simon. "Silicon Sovereigns: Artificial Intelligence, International Law, and the Tech-Industrial Complex." (2025). NUS Law Working Paper No. 2025/008.

81 See IAPP. "Global AI law and Policy Tracker" (2025).

82 Mills, Alex. *The Confluence of Public and Private International Law* (Cambridge University Press, 2010).

83 Howell, John and Ifayemi, Stephanie. "Policy Alignment on AI Transparency." PAI Policy Paper (2024).

84 Proukaki, Elena Katselli. *The Problem of Enforcement in International Law* (Taylor & Francis, 2011).

85 UN Charter (1945), Articles 39-42.

86 Statute of the International Court of Justice (1946), Articles 36 and 40.

87 UN Human Rights Council. Res A/HRC/59/32 (2025), para 12.

88 UNGA. A/HRC/46/37 (2021), para 23.

89 UNGA. A/78/L.49, para 6(k); UN Human Rights Council. Res A/HRC/59/53 (2025), paras 27, 50.

90 UN Human Rights Council. Res A/HRC/59/53 (2025), paras 32, 46, 50; UN Human Rights Council. Res A/HRC/59/32 (2025), paras 13, 25, 28, 38(a).

91 UN Human Rights Council. Res A/HRC/59/53 (2025), paras 58-59, 63, 65(m); UN Human Rights Council. Res A/HRC/59/32 (2025), paras 34-41, 54.

92 UN Human Rights Council. Res A/HRC/59/53 (2025), paras 27, 42-44, 50, 52, 53, 65(h); Res A/HRC/59/32 (2025), paras 12-16.

93 UN Human Rights Council. Res A/HRC/59/53 (2025), paras 58, 60; UN Human Rights Council. Res A/HRC/59/32 (2025), paras 34-41, 54.

94 See Owono, Julie. "Ensuring AI accountability : lessons from Meta's oversight board on human rights protection." Oxford Martin School Lecture (2025), drawing on the experience of Meta's Oversight Board.

95 E.g., OECD. "Overview and methodology of the AI Incidents and Hazards Monitor" (2025); MIT. "Tracking and classifying incidents of harm from AI" (2025).

96 Rotenberg, Marc. "The Imperative for a UN Special Rapporteur on AI and Human Rights."1 Journal of AI Law and Regulation (2024).

97 UNGA. A/79/L.118 (2025), para 1(a).

98 UNGA. A/79/L.118 (2025), paras 4(e)-(f).

99 Bengio et al. (2025).

100 E.g., Government of France. "Artificial Intelligence AI Action Summit" (2025); Government of India, "INDIA - AI IMPACT SUMMIT 2026" (2025).

101 NIST. "International Network of AI Safety Institutes - Mission Statement" (2024).

102 Chesterman, Simon. "Silicon Sovereigns: Artificial Intelligence, International Law, and the Tech-Industrial Complex." (2025). NUS Law Working Paper No. 2025/008, p. 8.

103 E.g., International Telecommunications Union. "ITU-T Work Programme: Security requirements and guidelines for Artificial Intelligence agent" (2025).

104 E.g., World Economic Forum. "AI red lines: the opportunities and challenges of setting limits" (2025); "International Dialogues on AI Safety" (2025), The Future Society. "Global Red Lines for AI: A Three-Part Series" (2025).