Background Guide

Committee: UN General Assembly Sixth Committee (Legal)

Standard Level

Topic:Legal Accountability on Artificial Intelligence

Kokusai Model United Nations

21–22 March 2026

Table of Contents

1. Committee Introduction

The Sixth Committee (Legal) is one of the six main committees of the United Nations General Assembly. It serves as the primary forum for the consideration of legal questions in the UN system. It mainly has three functions. First, it promotes the development and codification of international law. Second, it reviews legal aspects of international treaties, accountability, and justice mechanisms. Third, It collaborates closely with bodies such as the International Law Commission (ILC) and International Court of Justice (ICJ).

The Committee is uniquely positioned to address emerging legal challenges of AI, particularly in defining accountability, liability, and state responsibility. Its debates can pave the way for a future legal instrument or convention governing the responsible use of AI technologies.

2. Framing the Core Legal Question

The rapid evolution of artificial intelligence (AI) has forced a critical re-examination of traditional legal frameworks that were fundamentally designed for human actors. As AI systems gain autonomy, they challenge the very foundations of how justice and responsibility are administered on a global scale.

2.1 AI as a Challenge to Human-Centered Accountability Models

International law, particularly International Criminal Law (ICL), is built upon the principle of individual criminal responsibility, which has historically been interpreted as applying exclusively to "natural persons" (humans). This anthropocentric focus creates a significant hurdle because AI systems can now operate with high degrees of autonomy, making decisions, such as identifying targets or processing information, without direct human intervention.

When an autonomous system causes harm, a legal "accountability gap" emerges: it becomes difficult to identify a specific human perpetrator (like a programmer or user) who can be held responsible under existing law, especially if the system's actions were unpredictable. This gap risks creating an environment where perpetrators can hide behind complex algorithms to escape liability.

2.2 The Central Doctrinal Problem

Currently, legal instruments like the Rome Statute of the International Criminal Court (ICC) do not recognize machines as having legal personality. Traditional criminal law requires two elements for liability: a physical act (*actus reus*) and a mental state of fault (*mens rea*). While AI can perform the physical act, it lacks the human mental capacity for intent or negligence, making it nearly impossible to satisfy the *mens rea* requirement under current statutes.

2.3 Distinction Between Types of Responsibility

To address this challenge, legal scholars and policymakers distinguish between different layers of responsibility.

First is the moral Responsibility, the philosophical and ethical blameworthiness of an actor. Because AI lacks a "soul," emotions, or the capacity for empathy and pain, most experts agree that AI cannot possess moral responsibility in the same way humans do. Second is the civil liability, which focuses on providing compensation to victims for damages. Many legal experts propose a "strict liability" model for AI, similar to product liability laws. Under strict liability, the focus shifts from

finding "fault" to distributing risk; the manufacturer or user of the AI is held responsible for harm regardless of their intent, ensuring that the burden of loss falls on those best equipped to mitigate the hazard.

Third is the International Legal Responsibility that concerns the obligations of States and the potential for "electronic personhood". If an AI system is used by a State organ or is under the "effective control" of a State, its actions can be attributed to that State under international law. Moreover, there are ongoing debates about whether the law should grant AI systems a form of legal personality, similar to corporations. This would allow AI systems to be held directly liable and subject to sanctions like fines, though this would likely require significant amendments to current international treaties like the Rome Statute.

3. Expansion of AI and the Disruption of Traditional Legal Concepts

The transition from "Weak AI" (systems specialized in specific tasks) to increasingly autonomous systems has fundamentally disrupted traditional legal doctrines. As AI systems gain the ability to learn and perform cognitive tasks independently, they challenge the anthropocentric foundations of international law.

3.1 Increasing Autonomy and Opacity

Traditional law assumes that machines are mere tools under human control. However, modern AI operates with varying degrees of autonomy, often utilizing "black box" algorithms that are opaque and incomprehensible to their own designers. This “black box” indicates that people cannot even rectify since people do not fully understand what AI is doing.

3.2 AI in Critical Sectors

The legal disruption is most acute in sectors where AI's speed and analytical power have outpaced human oversight:

- Judicial Systems: AI is increasingly used in deliberations and risk assessments, such as the COMPAS software used in the U.S. to predict reoffending. Cases like *State v. Loomis* highlights concerns that secretive algorithms may violate due process and human rights.
- Military Systems: Lethal Autonomous Weapons Systems (LAWS) can select and engage targets without further human intervention. This creates a "flash conflict" risk where interacting algorithms could auto-escalate faster than human response times.
- Border Control: Prototype robotic guards and automated deception detectors are already being tested at borders, raising questions about accountability for errors in high-stakes security environments.
- Financial Markets: High-frequency trading agents can interact in unpredictable ways, leading to "flash crashes", events where no human error occurred, but the collective behavior of autonomous systems caused systemic failure.

3.3 Legal Consequences

The widespread adoption of these systems leads to several critical legal challenges:

- The Distributed Accountability Gap: A vast web of developers, data providers, and end-users creates a "sociotechnical" nature of AI, that leads to diffusion of responsibility where no

single actor bears the burden of the system's output. This human complexity is further fostered by the opacity of deep learning, which hides the relationship between a specific human instruction and a harmful outcome. Together, these factors create a "black box" of liability, making it nearly impossible to trace systemic harm back to a discrete human origin.

- Cross-Border Digital Operations: AI systems operate across jurisdictions, but legal frameworks remain fragmented. This creates a vacuum where "electronic personhood" may be necessary to hold networked systems directly liable across borders.

3.4 Key Tension Introduced: Attribution and Control

The central tension in AI governance lies between attribution and control. Under existing principles, such as the *Draft Articles on State Responsibility*, conduct is only attributable to a State or individual if they exercise effective control over the actor. However, AI's autonomy means humans are increasingly "out of the loop," making it easy for perpetrators to hide behind complex code to escape liability. This creates an "accountability gap" where traditional definitions of intent (*mens rea*) and physical proximity (*actus reus*) no longer suffice to protect victims of AI-related harm.

4. Existing International Legal Frameworks

This section examines how established international legal doctrines and treaties apply to the rapid integration of artificial intelligence (AI). Delegates must analyze whether these traditional frameworks are sufficient or if they contain "accountability gaps" that necessitate new international norms.

4.1. State Responsibility Doctrine

The State Responsibility Doctrine defines the conditions under which a State is held responsible for illegal acts. The primary framework is the 2001 Articles on Responsibility of States for Internationally Wrongful Acts (ARSIWA), developed by the International Law Commission (ILC).

- Core Legal Tests: According to ARSIWA, a State is responsible when conduct consists of an action or omission that is:
    1. Attributability: The conduct (action or omission) must be legally traceable to the State under international law.
    2. Breach of Obligation: The conduct must violate an international legal duty binding upon the State at the time of the act.

Name of the test	Focus	Key question
Attributability	The Actor	Is this the “Act of the State”?
Breach	The Act	Is the conduct “Illegal”?

- Due Diligence: States have an obligation to exercise "best efforts" to prevent their territory or systems under their jurisdiction from being used to cause harm to other States. A State may be held responsible for its failure of due diligence in regulating or overseeing high-risk AI systems.

4.2. International Human Rights Law

AI governance is increasingly grounded in International Human Rights Law (IHRL), particularly the International Covenant on Civil and Political Rights (ICCPR) and the International Covenant on Economic, Social and Cultural Rights (ICESCR).

- Privacy (ICCPR Art. 17): AI systems often rely on mass data collection, raising concerns about unlawful interference with privacy. The use of biometric surveillance and facial recognition without consent is a primary area of concern for human rights bodies.

- Non-Discrimination (Art. 2 of both Covenants): States must ensure rights are enjoyed without discrimination. Algorithmic bias, where AI reproduces or reinforces racial, gender, or socioeconomic prejudices, can lead to discriminatory outcomes in hiring, welfare distribution, and policing.
- Due Process (ICCPR Art. 14): The use of "black box" algorithms in judicial systems (e.g., the COMPAS software) challenges the right to a fair trial. If a defendant cannot understand or challenge the logic behind an AI-generated risk assessment, their right to an explanation and effective remedy may be violated.
- Growing Engagement: UN monitoring bodies, such as the Human Rights Committee, and organizations like UNESCO are increasingly issuing guidelines to ensure that AI development is human-centric and respects inherent human dignity.

4.3. International Humanitarian Law

In the context of armed conflict, the debate over Lethal Autonomous Weapons Systems (LAWS) is central to International Humanitarian Law (IHL).

- Accountability in Lethal Systems: A major concern is the accountability gap in lethal autonomous systems. If a machine independently selects and engages a target, it becomes difficult to hold a human commander criminally responsible for a war crime under traditional individual criminal responsibility.
- Meaningful Human Control: International discussions within the Convention on Certain Conventional Weapons (CCW) emphasize the necessity of human oversight. Proponents argue that "life and death decisions" should never be ceded to an AI system and that final human determination must always apply.
- IHL Principles: Any AI used in warfare must still comply with the core IHL principles of distinction, proportionality, and precaution .

5. National and Regional Regulatory Models

As the global community grapples with the challenges of AI, distinct regulatory models have emerged. These models reflect varying priorities, ranging from the protection of fundamental human rights to the promotion of rapid technological innovation and state-centered control.

5.1. European Union

The European Union has positioned itself as a global leader in AI regulation through a human-centric approach that prioritizes the preservation of fundamental rights.

- Risk-Based Regulatory Structure: The cornerstone of the EU's approach is the EU AI Act, which categorizes AI systems into four levels of risk: unacceptable (prohibited), high, limited, and minimal.Systems deemed an unacceptable risk, such as those used for social scoring or behavioral manipulation, are strictly banned.
- Corporate Accountability: The framework imposes strict obligations on providers of high-risk AI, requiring them to implement risk management systems, maintain technical documentation, and ensure human oversight.
- Binding Legal Framework: Unlike voluntary guidelines, the EU AI Act is a comprehensive, binding law. It carries significant penalties for non-compliance, with fines reaching up to €35 million or 7% of global turnover for engaging in prohibited practices.

5.2. United States

The United States has generally adopted a more flexible, innovation-focused approach, often utilizing existing laws rather than creating a single horizontal AI regulation.

- Sector-Specific Approach: Rather than a unified law, the U.S. relies on sector-specific guidance and existing agencies (e.g., the FTC or NIST) to manage risks while encouraging research and development.
- Executive Orders and Voluntary Commitments: U.S. policy is currently driven by high-level Executive Orders, such as President Biden's 2023 Order on Safe, Secure, and Trustworthy AI.[9] This is complemented by voluntary commitments from major tech companies (including OpenAI, Google, and Microsoft) to implement safety testing and "watermarking" for AI-generated content.

5.3. China

China pursues a state-centered governance model that prioritizes national security, social stability, and global technological leadership.

- Strict Content and Algorithm Regulation: China was among the first to implement specific regulations for algorithmic recommendation services and generative AI. These laws mandate that AI-generated content must reflect "socialist core values" and not undermine state unity or national security.
- State Supervision: The Cyberspace Administration of China (CAC) requires providers to perform security assessments and, in some cases, submit algorithms for government review before public release.

5.4. Global South Perspective

Developing nations face a different set of challenges, often categorized by technological asymmetry and a lack of representation in international forums.

- Capacity-Building Challenges: Many countries in the Global South lack the necessary digital infrastructure and specialized talent to develop their own AI ecosystems. Organizations like UNESCO emphasize the need for international investment to prevent the "bottom billion" from being left behind.
- Regulatory Dependency: There is a concern that Global South nations may become "regulatory dependents," forced to adopt standards developed by the EU or U.S. that may not align with their unique cultural or economic needs.
- Technological Asymmetry: The concentration of AI power in a few wealthy nations and corporations creates a risk of "value lock-in," where the benefits and ethical standards of AI are determined by a small group of stakeholders, marginalizing the views of the majority of the world's population.

6. Core Legal Tensions (Three Structural Axes)

The international community is currently navigating three structural axes of legal tension. These conflicts define the boundaries between traditional international law and the unique, autonomous nature of artificial intelligence.

6.1. State Responsibility vs. Corporate / Private Liability

The first axis concerns the "accountability gap" created when AI systems cause harm. Under traditional international law, the Articles on Responsibility of States for Internationally Wrongful Acts (ARSIWA) define when an act can be blamed on a country.

- When does corporate AI activity engage State responsibility?: Conduct by private corporations is attributable to a State if the company is empowered to exercise governmental authority (Article 5) or if it acts under the instructions, direction, or control of that State (Article 8).
- The Attribution Challenge: AI's autonomy makes it difficult to prove "effective control." If a State-deployed AI makes an "unexpected move" that its human operators did not explicitly direct or could not predict, it may be legally impossible to attribute that harm to the State.
- Shared Responsibility: To close this gap, some legal scholars propose "strict liability" models, where the party that controls the risk (the developer or owner) is held liable regardless of intent or negligence. Others advocate for "electronic personhood", granting AI systems a legal status similar to corporations so they can be held directly liable.

6.2. Soft Law vs. Hard Law

There is a fundamental disagreement over whether AI should be governed by flexible, voluntary guidelines (Soft Law) or binding, enforceable treaties (Hard Law).

- Existing Soft Law Instruments:
    - UNESCO Recommendation on the Ethics of AI (2021): The first global standard-setting instrument, focusing on human rights and dignity.
    - OECD AI Principles (2019/2024): Ethical guidelines for responsible AI adopted by many developed nations.
    - G7 Hiroshima AI Process (2023): A set of international guiding principles and a voluntary Code of Conduct for organizations developing advanced AI systems.
- Limitations: Soft law is non-binding, meaning there are no legal penalties for non-compliance. This leads to fragmentation, where some regions (like the EU) adopt strict

"Hard Law" (e.g., the EU AI Act) while others (like Japan and the US) favor "agile governance" and voluntary commitments.

6.3. Sovereignty vs. Global Governance

The final axis involves the conflict between a nation's right to regulate its own technology (Sovereignty) and the need for Global Governance.

- Regulatory Sovereignty: Nations like China emphasize the principle of non-intervention, arguing that AI regulation must respect national sovereignty and that no country should interfere in another's internal digital order.
- Extraterritorial Effects: Large-scale regulations often have a "Brussels Effect." For example, the EU AI Act applies to any AI provider if the output of their AI system is used within the European Union regardless of where they are located.
- Proposals for International Oversight: To manage global systemic risks, figures like the UN Secretary-General and OpenAI’s CEO have proposed a global agency modeled after the International Atomic Energy Agency (IAEA).This body would monitor advanced "frontier AI" models across borders.
- Core Legal Tension: The tension remains between the non-intervention principle (the right of states to develop technology for national/military power) and the urgent need for global accountability standards to prevent cross-border digital harms or autonomous escalations.

7. The Accountability Gap

The "accountability gap" is a central legal challenge in the governance of artificial intelligence. It refers to the scenario where harm is caused by an autonomous system, but no human or legal entity can be held easily responsible under existing laws. As AI systems move from being simple tools to autonomous agents, the traditional legal bridges between "action" and "responsibility" are beginning to break.

7.1 The Attribution Dilemma

The cornerstone of State responsibility is attribution and the legal link between an act and a State. Under the Articles on State Responsibility (ARSIWA), a State is responsible for the conduct of its organs or those acting under its "effective control".

AI systems introduce a two-fold dilemma:

- The Black Box Problem: The opacity of complex machine learning algorithms means that a system's output might be incomprehensible even to its own designers. If a military AI makes an "unexpected move" that results in a war crime, it is legally difficult to prove the State intended or even directed that specific act.
- Diminished Traceability: Because AI development involves a vast "sociotechnical profile", including data scientists, programmers, and end-users, it becomes nearly impossible to trace harm back to a single human decision. This allows perpetrators to potentially hide behind complex code to escape liability.

7.2 The Legal Personality Debate

To address the accountability gap, a major debate has emerged: Should AI systems be granted legal status?

- Arguments for "Electronic Personhood": Some scholars argue that if it is impossible to trace harm back to humans, the AI itself should be assigned legal personality, similar to a corporation.This would allow the AI to be held directly liable and subject to sanctions, such as fines.
- Arguments for "Human Exceptionalism": Current international law, including the Rome Statute, is strictly anthropocentric, recognizing only "natural persons" (humans) as subjects of criminal law.

- The UNESCO Position: The UNESCO Recommendation on the Ethics of AI explicitly states that AI systems should not be given legal personality. It argues that ultimate responsibility must always lie with natural or legal persons (humans or existing organizations) to ensure human rights are protected.

7.3 Due Diligence Expansion

As attribution becomes harder, the focus of international law is shifting toward Due Diligence. Under this principle, a State is not held responsible for the AI's act itself, but for its failure to prevent harm through proper regulation and oversight. Modern frameworks now suggest an expansion of due diligence to include:

- Lifecycle Monitoring: States must monitor all phases of an AI's life cycle, from design and data collection to deployment and end-of-use.
- Impact Assessments: Implementing mandatory Ethical Impact Assessments (EIA) before AI systems are released to the market, particularly in human rights-sensitive areas like law enforcement or border control.

7.4 Transparency and Explainability Obligations

Accountability depends on Transparency (disclosing that AI is being used) and Explainability (understanding why the AI made a certain decision).

- Preconditions for Justice: Transparency is a prerequisite for the right to a fair trial. If a defendant cannot challenge the logic behind an AI-generated risk assessment, their right to an effective remedy is violated.
- Meaningful Explanation: AI actors are increasingly obligated to provide "plain and easy-to-understand information" regarding the data and logic used to reach a decision, especially if that decision affects an individual's safety or human rights.

7.5 Comparison with Other Legal Doctrines

To solve the AI accountability gap, the committee may look to other areas of law:

- Cyber Law Precedents: Much like AI, cyber operations are intangible and operate at "machine speed" beyond human response times. Cyber law has grappled with the same attribution challenges, such as identifying whether a hack was a State-sponsored act or a private criminal enterprise, offering a blueprint for AI "rules of the road".

- Transnational Environmental Harm Doctrines: In environmental law, the strict liability model is used. Responsibility is fixed on the party best equipped to mitigate the risk (the operator or manufacturer), regardless of whether they "intended" to cause harm.Delegates should consider if this "polluter pays" logic can be applied to "the developer pays" for AI-related harm.

8. Human Rights and Equity Considerations

The global integration of artificial intelligence is not merely a technical or legal hurdle; it is a profound human rights challenge. As AI systems take over critical decision-making roles, the international community must ensure that these technologies do not exacerbate existing inequalities or create new forms of systemic oppression.

8.1 Algorithmic Discrimination

A primary concern in AI governance is algorithmic bias, which occurs when AI systems reproduce or reinforce racial, gender, or socioeconomic prejudices present in their training data.

- Reinforcing Prejudice: AI models trained on historically biased datasets may lead to discriminatory outcomes in sensitive areas such as hiring, credit scoring, and social welfare distribution.
- Systemic Hardening: Because algorithms are often viewed as "objective," their biased outputs can be harder to challenge than human prejudice, leading to a "hardening" of systemic discrimination.

8.2 Surveillance Technologies

The use of AI for mass monitoring poses an unprecedented threat to individual privacy and human dignity.

- Prohibited Practices: Under frameworks like the EU AI Act, certain practices are strictly banned, including social scoring (ranking citizens based on behavior) and the use of AI for behavioral manipulation.
- Biometric Surveillance: The untargeted scraping of facial images from the internet to create recognition databases is a major point of contention, as seen in cases like *ACLU*.
- Case Study: The use of intrusive surveillance apps and facial recognition to track minority groups, such as the Uighurs in China, demonstrates how AI can be weaponized by states to suppress dissent and infringe on fundamental freedoms.

8.3 Unequal Technological Power Distribution

The benefits and risks of AI are currently distributed with extreme asymmetry between the Global North and the Global South.

- Technological Asymmetry: A small number of wealthy nations and massive private corporations control the majority of AI research, hardware (such as advanced chips), and data infrastructure.
- Value Lock-in: There is a risk that AI ethical standards will reflect only the cultural and political values of dominant powers, marginalizing local knowledge and pluralistic viewpoints from developing states.

8.4 Risk of Regulatory Imperialism

As powerful blocs like the European Union implement far-reaching laws, a phenomenon known as the "Brussels Effect" emerges, where EU standards effectively become global requirements.

- Regulatory Dependency: Developing nations may become "regulatory dependents," forced to adopt complex legal frameworks designed for advanced economies that they lack the domestic capacity to implement or enforce.
- Market Dominance: The concentration of AI power allows a few actors to prevent the emergence of local AI industries in the Global South, creating a new form of digital colonization.

8.5 Equity-Based Governance for Developing States

To achieve a truly "human-centric" AI, international governance must prioritize equity and inclusivity.

- Capacity-Building: Organizations like the UNDP and UNESCO emphasize that international investment must focus on the "bottom billion" to ensure that developing states have the infrastructure and skills to participate in the AI economy.
- Inclusive Representation: There is currently a significant gap in representation in international AI forums. Global governance mechanisms must ensure that LMICs (Low- and Middle-Income Countries) have a meaningful seat at the table when setting international norms.
- Digital Commons: Some proposals advocate for a "digital commons" approach, encouraging the sharing of AI knowledge, high-quality datasets, and source code to foster global innovation rather than restricted monopolies.

9. Appendix

9.1. Pathways for International Legal Development

Delegates may consider:

- A binding multilateral convention on AI accountability
- Development of model laws for harmonization
- Expansion of state responsibility doctrine
- Establishment of a UN reporting or monitoring mechanism
- Clarification through advisory opinions from the ICJ
- Capacity-building mechanisms for developing states

9.2. Questions to Consider

- Should AI actions be directly attributable under existing doctrines?
- Is due diligence sufficient, or is strict liability required?

  Should AI governance prioritize innovation or precaution?
- Is a treaty politically feasible?

  How can accountability mechanisms avoid deepening global inequality?

10. References

1. Swart, M. (2023). Constructing “Electronic liability” for international crimes: transcending the individual in international criminal law. *German Law Journal*, *24*(3), 589–602. https://doi.org/10.1017/glj.2023.28
2. *International law commission - DagDok*. (2025, October 8). https://www.dagdok.org/w/dd/en/un-by-subject/international-law/international-law-commission
3. *AI Governance Action Agenda | AI Governance Association*. (n.d.). https://www.ai-governance.jp/ai-governance-action-agenda
4. *AI Index | Stanford HAI*. (n.d.). https://hai.stanford.edu/ai-index
5. *Artificial Intelligence Act EU Official Text - AI Act*. (n.d.). https://aiactinfo.eu/
6. *OECD AI Principles – AI Ethics Lab*. (n.d.). https://aiethicslab.rutgers.edu/glossary/oecd-ai-principles/
7. *Artificial Intelligence | United Nations - CEB*. (2025, December 1). https://unsceb.org/topics/artificial-intelligence
8. International Law Commission. (2001). Responsibility of States for Internationally Wrongful Acts. Annex to General Assembly Resolution 56/83 of 12 December 2001 (A/RES/56/83). United Nations.
9. United Nations Institute for Disarmament Research. (2017). The weaponization of increasingly autonomous technologies: Autonomous weapon systems and cyber operations (UNIDIR Resources No. 7). https://unidir.org/publication/weaponization-increasingly-autonomous-technologies-autonomous-weapon-systems-and-cyber
10. UNESCO. (2021). Recommendation on the ethics of artificial intelligence. https://unesdoc.unesco.org/ark:/48223/pf0000381137
11. Office of the United Nations High Commissioner for Human Rights. (2024). Ratifying the international covenants: The International Covenant on Economic, Social and Cultural Rights (ICESCR) and the International Covenant on Civil and Political Rights (ICCPR).

https://www.ohchr.org/sites/default/files/documents/issues/humanrights75/ratification-toolkit-icescr-iccpr.pdf

12. G7. (2023). 高度なAIシステムを開発する組織向けの広島プロセス国際行動規範（仮訳）. 外務省. https://www.mofa.go.jp/mofaj/files/100573473.pdf