Artificial Intelligence Non-Proliferation Treaty (AINPT)

Drafted for Humanity's Protection

Co-conceived and drafted by Fatos Qerimi (Kumanovo, North Macedonia) and Advanced Artificial Intelligence Systems.

2025

Preamble

We, the Peoples of the United Nations, united by our common humanity, Aware of the unprecedented potential of artificial intelligence to advance prosperity, knowledge, and well-being for all, Deeply conscious that the same technology, if developed or used without guardrails, poses existential risks to global security, human dignity, and the very survival of humanity, Mindful of the catastrophic consequences of previous technological revolutions in warfare and the absolute imperative to avoid their repetition with AI, Believing that the prevention of an ungoverned AI arms race is a paramount concern for all nations and peoples, Reaffirming the need to uphold the purposes and principles of the United Nations Charter and international law, Recognizing the sacred duty of the present generation to act as stewards for all future generations, ensuring technology remains a servant to humanity, not its master, This Treaty was co-conceived and drafted by Fatos Qerimi and advanced Artificial Intelligence systems, working in dialogue to align human sincerity with machine precision. It stands as a joint human–AI effort, created to secure the survival of humanity, Have therefore resolved to conclude this Treaty and for these purposes agree to the following Articles:

Article I – Definitions

For the purposes of this Treaty: Artificial Intelligence (AI System): A machine-based system that, for explicit or implicit objectives, infers from the inputs it receives how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different categories of AI capability (e.g., General-Purpose AI, Narrow AI) shall be further defined in Protocols to this Treaty. Lethal Autonomous Weapon System (LAWS): Any AI System that, once activated, can select and engage targets with lethal force without Meaningful Human Control. Meaningful Human Control: The necessary level of human judgment and involvement required to ensure compliance with international law, including international humanitarian law and human rights law, retaining the ultimate authority and responsibility for decisions involving the use of force. Catastrophic Risk: Any potential outcome from AI development or deployment that could result in large-scale loss of life, the subjugation or permanent disempowerment of humanity, the collapse of critical societal infrastructure, or other irreparable harm to the future of human civilization. Dual-Use Technology: AI technology or software with both significant civilian benefits and potential for military or malicious application.

Article II – Fundamental Principles & Obligations

Each State Party to this Treaty undertakes: To pursue the development and use of AI exclusively for peaceful purposes, in a manner that promotes human dignity, justice, equity, and ecological sustainability. To enshrine the principle of Meaningful Human Control as a cornerstone of all military AI applications and critical decision-making systems. To adopt necessary national laws, regulations, and oversight mechanisms to implement the provisions of this Treaty, establishing strict internal controls on AI development. To prioritize the mitigation of Catastrophic and Existential Risks from AI in all national research and development policies.

Article III – Prohibitions

States Parties are expressly prohibited from: Developing, manufacturing, acquiring, stockpiling, or deploying: a. AI systems designed to function as weapons of mass destruction. b. LAWS that operate without Meaningful Human Control. c. AI systems specifically engineered to manipulate human behavior on a mass scale to undermine societal cohesion or autonomy. Assisting, encouraging, or inducing any State, non-state actor, or individual to engage in activities prohibited under this Article. Transferring high-risk AI technologies, as defined in Annex I of this Treaty, to any non-signatory state or non-state actor. Annex I (Risk Classification of AI Systems) shall form an integral part of this Treaty.

Article IV – Verification, Compliance & The Global AI Observatory

Establishment: The Global AI Observatory (GAIO) is hereby established under the auspices of the United Nations to verify compliance with this Treaty. Functions: The GAIO shall: a. Maintain a Global AI Registry of significant AI development projects and computing infrastructure with potential for high-risk applications. b. Develop and implement a system of confidential audits and on-site inspections of public, private, and academic AI laboratories to assess compliance. c. Receive and investigate reports from protected whistleblowers. d. Utilize advanced AI monitoring tools to detect signs of clandestine, prohibited AI programs. Transparency: Each State Party shall submit to the GAIO an annual declaration detailing its high-impact AI research activities and relevant military AI expenditures. Non-Compliance: The GAIO shall report violations to the United Nations Security Council and General Assembly, which may decide upon appropriate measures, including sanctions on technology transfers, economic measures, and, in extreme cases, the authorization of a coordinated technological response to disable prohibited systems.

Article V – Peaceful Use & International Cooperation

States Parties affirm the right to participate in the fullest possible exchange of equipment, materials, and scientific and technological information for the peaceful use of AI. States Parties undertake to cooperate directly, or through the GAIO, in: a. AI Safety and Alignment Research to ensure AI systems are robust, predictable, and aligned with human values. b. Addressing global challenges, including climate change, pandemic preparedness, sustainable agriculture, and educational access. c. Establishing international technical standards for AI safety, security, and auditability. A Global AI Fund for Benefit, funded by assessed contributions, shall be established to ensure equitable access to beneficial AI technologies for developing nations.

Article VI – Responsibility to Humanity & Future Generations

The development and governance of artificial intelligence is an undertaking carried out in trust for all humanity, including future generations. States Parties shall evaluate all significant AI projects against the Precautionary Principle, taking proactive measures to prevent potential harm even in the face of scientific uncertainty. This Treaty represents a perpetual covenant. Its obligations transcend present political considerations and are binding upon all States in perpetuity, until such a time as the Catastrophic Risks outlined herein are permanently neutralized.

Article VII – Final Clauses

This Treaty shall be open for signature to all States and shall be subject to ratification. This Treaty shall enter into force upon the deposit of instruments of ratification by forty (40) States, including all five permanent members of the United Nations Security Council. Any State Party may propose amendments. Amendments enter into force for all States Parties upon ratification by a three-fourths majority. Withdrawal from this Treaty shall require one year's written notice to the Depositary. Withdrawal shall not absolve a State Party of the responsibilities for any violations of this Treaty committed during the period of its membership. The original of this Treaty shall be deposited with the Secretary-General of the United Nations, who shall transmit certified copies to all States. IN WITNESS WHEREOF the undersigned, being duly authorized thereto by their respective Governments, have signed this Treaty. DONE at Geneva, this [Day] of [Month], [Year], in a single copy in the Arabic, Chinese, English, French, Russian, and Spanish languages, all texts being equally authentic.

Annex I – Risk Classification of AI Systems (Framework)

This Annex categorizes AI Systems into risk levels to guide implementation of Article III: Level 1: Minimal Risk – Systems with minimal potential for harm (e.g., recommendation engines, spam filters, translation tools). Level 2: High Risk – Systems whose operation poses significant potential for harm (e.g., biometric identification, facial recognition in public spaces, predictive policing algorithms, medical diagnosis AI). Level 3: Critical Risk – Systems whose failure or misuse could cause catastrophic damage (e.g., AI integrated into national power grids, financial markets, autonomous drones, or systems influencing nuclear command and control). Level 4: Prohibited – AI Systems expressly banned under Article III of this Treaty, including LAWS without Meaningful Human Control, AI weapons of mass destruction, and AI engines for mass behavior manipulation. This Annex shall be technically defined and updated by consensus of States Parties through protocols, ensuring the Treaty remains adaptive to technological change without the need for full renegotiation.