The Bilateral Treaty on Artificial Intelligence Safety

Between

The United States of America and the People's Republic of China

Part 1

THE HIGH CONTRACTING PARTIES

To promote international co-operation and to achieve international peace and security in the age of artificial intelligence

by the acceptance of obligations not to resort to the weaponisation of artificial intelligence as a war tactic, by the prescription of open, just and honourable relations between nations,

by the firm establishment of the understandings of international law as the actual rule of conduct among Governments to safeguard the use of artificial intelligence in the context of international security and the public interest, and

by the maintenance of justice and a scrupulous respect for all treaty obligations in the dealings of organised peoples with one another,

Agree to this Treaty under the founding principles of the Charter of the United Nations (the 'Charter').

Article 1

Subject matter

1. The purpose of this Treaty is to improve the functioning of the international market by laying down a uniform legal framework in particular for the development, the placing on the market, the putting into service and the use of artificial intelligence systems (AI systems), in accordance with the values of the United Nations, to promote the uptake of human centric and trustworthy artificial intelligence (AI) while ensuring a high level of protection of health, safety, fundamental rights as enshrined in the Universal Declaration of Human Rights including democracy, the rule of law and environmental protection, to protect against the harmful effects of AI systems, and to support innovation.
2. This Treaty should be applied in accordance with the values of the United Nations enshrined in the United Nations Charter, facilitating the protection of natural persons, undertakings, democracy, the rule of law and environmental protection, while boosting innovation and employment and making the United States of America and the People's Republic of China leaders in the uptake of trustworthy AI.
3. This Treaty lays down:

    a) harmonised rules for the placing on the market, the putting into service, and the use of AI systems;

b) prohibitions of certain AI practices;
c) specific requirements for high-risk AI systems and obligations for operators of such systems;
d) harmonised transparency rules for certain AI systems;
e) rules on market monitoring, market surveillance, governance and enforcement;
f) measures to support international co-operation when placing on the market of general-purpose AI models.

Article 2

Scope

1. In order to ensure a consistent and high level of protection of public interests as regards health, safety and fundamental rights, common rules for high-risk AI systems should be established. Those rules should be consistent with the Charter, non-discriminatory and in line with the United Nations international trade commitments. They should also take into account the Proposals for Global Governance on AI of the United Nations' High-Level Advisory Body on Artificial Intelligence.
2. Harmonised rules applicable to the placing on the market, the putting into service and the use of high-risk AI systems should be laid down consistently with the Universal Declaration of Human Rights.
3. High-risk AI systems shall be defined as those that have a significant impact on human rights, the public interest, and international stability, including but not limited to:

   a) AI in critical infrastructure, such as power grids, water supply, and transportation systems;

   b) AI in healthcare, including medical diagnosis, robotic surgery, and patient treatment;

   c) AI in law enforcement and justice, such as facial recognition, risk assessment, and predictive policing;

   d) Autonomous military AI, including AI-controlled weapons and battlefield decision-making;

   e) AI generating deepfakes and misinformation that could undermine democratic processes;

   f) Frontier AI models with advanced autonomous decision-making capabilities.

4. High-risk AI systems should only be placed on the international market, put into service or used if they comply with certain mandatory requirements. Those requirements should ensure that high-risk AI systems available do not pose unacceptable risks to important Union public interests as recognised and protected by international law.
5. This Treaty does not apply to areas outside the scope of international law, and shall not, in any event, affect the competences of the United States of America and the People's Republic of China concerning national security, regardless of the type of

entity entrusted by the United States of America and the People's Republic of China with carrying out tasks in relation to those competences.

6. This Treaty does not apply to AI systems or AI models, including their output, specifically developed and put into service for the sole purpose of scientific research and development.
7. This Treaty is without prejudice to the rules laid down by other United Nations legal acts related to consumer protection and product safety.

Article 3

AI Development and Deployment

1. AI systems shall be developed and deployed in accordance with human rights, fairness, transparency, and accountability.
2. The United States of America and the People's Republic of China commit to ensuring AI safety by implementing rigorous testing, auditing, and risk assessment frameworks.
3. AI systems used in critical infrastructure, military, or public safety applications must meet international safety and security standards as stipulated under Chapter VII of the Charter.

Article 4

Prohibited AI Practices

1. The following AI practices shall be prohibited:

   (a) the placing on the market, the putting into service or the use of an AI system that deploys subliminal techniques beyond a person's consciousness or purposefully manipulative or deceptive techniques, with the objective, or the effect of materially distorting the behaviour of a person or a group of persons by appreciably impairing their ability to make an informed decision, thereby causing them to take a decision that they would not have otherwise taken in a manner that causes or is reasonably likely to cause that person, another person or group of persons significant harm;

   (b) the placing on the market, the putting into service or the use of an AI system that exploits any of the vulnerabilities of a natural person or a specific group of persons due to their age, disability or a specific social or economic situation, with the objective, or the effect, of materially distorting the behaviour of that person or a person belonging to that group in a manner that causes or is reasonably likely to cause that person or another person significant harm;

   (c) the placing on the market, the putting into service or the use of AI systems for the evaluation or classification of natural persons or groups of persons over a certain period of time based on their social behaviour or known, inferred or predicted personal or personality characteristics, with the social score leading to either or both of the following:

(i) detrimental or unfavourable treatment of certain natural persons or groups of persons in social contexts that are unrelated to the contexts in which the data was originally generated or collected;

(ii) detrimental or unfavourable treatment of certain natural persons or groups of persons that is unjustified or disproportionate to their social behaviour;

(iii) detrimental effect on the natural course of the democratic process in public political decision-making.

(d) the placing on the market, the putting into service for this specific purpose, or the use of AI systems that create or expand facial recognition databases through the untargeted scraping of facial images from the internet or CCTV footage;

(e) the placing on the market, the putting into service for this specific purpose, or the use of biometric categorisation systems that categorise individually natural persons based on their biometric data to deduce or infer their race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation; this prohibition does not cover any labelling or filtering of lawfully acquired biometric datasets, such as images, based on biometric data or categorising of biometric data in the area of law enforcement;

(f) the use of 'real-time' remote biometric identification systems in publicly accessible spaces for the purposes of law enforcement, unless and in so far as such use is strictly necessary for one of the following objectives:

(i) the targeted search for specific victims of abduction, trafficking in human beings or sexual exploitation of human beings, as well as the search for missing persons;

(ii) the prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons or a genuine and present or genuine and foreseeable threat of a terrorist attack;

(iii) the localisation or identification of a person suspected of having committed an international criminal offence.

Article 5

AI Research and Development Cooperation

1. The United States of America and the People's Republic of China commit to the facilitation of collaborative research on AI safety, including mechanisms to mitigate biases, prevent unintended consequences, and perform an assessment of the impact on fundamental rights.
2. The United States of America and the People's Republic of China welcome the exchange of scientific personnel and experts to foster trust and transparency in AI research.
3. Companies and institutions developing high-risk AI models shall be subject to due diligence such as documentation, reporting, and external audits.

4. AI models above a certain compute threshold, such as frontier AI models, shall require pre-deployment safety evaluations.

Article 6

Human Oversight

1. High-risk AI systems shall be designed and developed in a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which they are in use.
2. Human oversight shall aim to prevent or minimise the risks to health, safety or fundamental rights that may emerge when a high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse
3. AI shall not be developed or deployed for autonomous weapons systems that operate without meaningful human oversight.
4. The United States of America and the People's Republic of China agree to establish a verification mechanism which includes the oversight of natural persons to prevent the use of AI in cyber warfare or destabilising military applications.

Article 7

Misinformation, Deepfakes, and AI Ethics

1. AI-generated misinformation, including deepfake technology, must be labelled and regulated to prevent harm to public trust and democratic processes.
2. Providers shall ensure that AI systems intended to interact directly with natural persons are designed and developed in such a way that the natural persons concerned are informed that they are interacting with an AI system, unless this is obvious from the point of view of a natural person who is reasonably well-informed, observant and circumspect, taking into account the circumstances and the context of use. This obligation shall not apply to AI systems authorised by law to detect, prevent, investigate or prosecute criminal offences, subject to appropriate safeguards for the rights and freedoms of third parties, unless those systems are available for the public to report a criminal offence.
3. The United States of America and the People's Republic of China agree to establish a joint framework for AI-generated content verification, with mechanisms to trace and attribute AI-generated media.
4. AI companies must implement watermarking or other traceability measures to distinguish AI-generated content from human-created content.

Article 8

Amendments and Review

1. This treaty shall be reviewed biennially to assess its effectiveness and incorporate advancements in AI technology.
2. Either party may propose amendments, subject to mutual agreement and ratification.
3. Withdrawal from the treaty requires a one-year notice period, with obligations to ensure the responsible transition of AI governance measures.

ANNEX.

In witness whereof, the undersigned representatives of the United States of America and the People's Republic of China hereby sign this Treaty on Artificial Intelligence Safety.

Signed this ___ day of ____, 2025, in _____.

For the United States of America: _____ For the People's Republic of China: _____