Draft Treaty:

Pause and Governance of Artificial Intelligence Development

2023

Index

1. Preamble 04

2. Definitions 05

3. General Provisions 05-07

3.1 Purpose

3.2 Scope

3.3 Objectives

3.4 Principles

4. Provisions Regarding the Moratorium on Large-scale AI Capabilities 07-09

Research and Development

5. Rights and Obligations 09

5.1 Parties' Rights

5.2 Parties' Obligations

6. Institutional Framework 10

6.1 Establishment of Regulatory Body

6.2 Functions and Powers of the Regulatory Body

6.3 Reporting and Accountability Mechanisms

7. Regulatory Measures 10-11

7.1 Licensing and Permitting

7.2 Technical Standards and Specifications

7.3 Monitoring and Compliance

7.4 Reporting and Information Exchange

8. Dispute Resolution 11

9. Potential Enforcement Procedures 11

10. Duration and Review 11

11. Ratification and Entry into Force 12

*The States Parties to this Treaty,*

*Recognizing the rapid advancements and potential risks associated with Artificial Intelligence (AI) technologies,*

*Acknowledging the need for international cooperation and governance to ensure the safe and ethical development of AI,*

*Mindful of the concerns regarding the uncontrolled growth and potential unintended consequences of large-scale AI capabilities research and development,*

*Considering the importance of a cautious and responsible approach in addressing the ethical, societal, and safety implications of AI,*

*Recognizing the significance of comprehensive risk assessment, oversight mechanisms, and international collaboration in mitigating the potential risks of large-scale AI capabilities research and development,*

*Emphasizing the importance of striking a balance between the promotion of innovation and the protection of human rights, privacy, and the integrity of democratic societies,*

*Aware of the need to address public concerns and build public trust in the development and deployment of AI technologies,*

*Convinced that a temporary moratorium on large-scale AI capabilities research and development, accompanied by appropriate governance and international cooperation, is necessary to ensure the long-term benefits and responsible use of AI,*

*Have agreed as follows:*

*Article I*

For the purpose of this treaty, the following terms and expressions shall have the meanings ascribed to them below:

1. Large GPU and TPU Clusters: Refers to the computer farms comprising large-scale graphics processing units (GPUs) and tensor processing units (TPUs) that are used for refining and enhancing powerful Artificial Intelligence (AI) systems.
2. Machine Learning (ML) Models: Refers to algorithms and computational models that enable AI systems to learn and make predictions or decisions based on data patterns.
3. Quantum Computers: Refers to computing devices that utilize the principles of quantum mechanics to perform complex calculations, potentially offering significant advancements in AI capabilities.
4. Moratorium: Refers to a temporary suspension or pause on specific activities, in this case, the large-scale research and development of AI capabilities.
5. Artificial General Intelligence (AGI): Refers to AI systems capable of performing any intellectual task that a human being can do.
6. Artificial Superintelligence (ASI): Refers to AI systems that surpass human intelligence and possess capabilities far beyond human comprehension.
7. AI-enabled Crimes: Refers to criminal activities facilitated or perpetrated through the use of Artificial Intelligence technologies, including but not limited to fraud, cyberattacks, malicious manipulation, and other harmful actions.
8. Redress: Refers to the act of addressing and compensating individuals or entities who have suffered harm or damages due to the use of AI technologies.
9. AI Governance Council (AIGC): Refers to the international oversight body established under this treaty, composed of representatives appointed by the participating parties, responsible for ensuring the effective implementation, enforcement, and monitoring of the treaty provisions.

*Article 2*

2.1 The purpose of this article is to establish the general provisions that govern the implementation of this treaty. It sets out the overarching goals and principles that guide the treaty's application and enforcement.

2.2 This treaty applies to all activities related to the development, deployment, and use of Artificial Intelligence (AI) technologies. It encompasses AI research, design, implementation, testing, and any other relevant AI-related activities, regardless of the sectors or industries involved.

2.3 The objectives of this treaty are:

2.3.1 To ensure the safe, ethical, and responsible development and deployment of AI technologies, taking into account the potential risks and societal impact they may pose.

2.3.2 To foster international cooperation and collaboration among the participating parties, promoting the exchange of information, best practices, and expertise in AI governance.

2.3.3 To establish mechanisms for the protection of human rights, privacy, and the integrity of democratic societies in the context of AI development and deployment.

2.3.4 To promote transparency, accountability, and inclusivity in AI governance, involving stakeholders from diverse backgrounds, including academia, industry, civil society, and affected communities.

2.3.5 To encourage innovation and research in AI technologies while ensuring that their development aligns with societal values, diversity, and the common good.

2.4 The following principles shall guide the implementation of this treaty:

2.4.1 Ethical and Human-Centric Approach: AI development and deployment shall prioritize the well-being, rights, and dignity of individuals, respecting human values, and avoiding harm or discrimination.

2.4.2 Transparency and Explainability: AI systems and their decision-making processes should be transparent, understandable, and explainable to foster trust, accountability, and meaningful human oversight.

2.4.3 Fairness and Non-Discrimination: AI technologies should be designed and used in a manner that avoids unjust biases, discrimination, or unfair outcomes, ensuring equal treatment and opportunities for all.

2.4.4 Privacy and Data Protection: The privacy and personal data of individuals should be safeguarded throughout the lifecycle of AI systems, adhering to applicable laws and international standards.

2.4.5 Accountability and Liability: Parties involved in AI development and deployment shall be accountable for the actions and consequences of AI systems, and mechanisms for legal and ethical accountability should be established.

2.4.6 International Cooperation: Participating parties shall cooperate in the exchange of information, research findings, best practices, and capacity-building initiatives to foster global collaboration in AI governance.

2.4.7 Risk Assessment and Mitigation: Robust risk assessment procedures should be implemented to identify and mitigate potential risks and adverse impacts associated with AI technologies.

2.4.8 Human Rights and Social Impact: AI development and deployment shall respect and uphold fundamental human rights, ensuring that AI systems contribute positively to societal well-being, inclusivity, and sustainable development.

2.4.9 Proportionality and Precaution: AI technologies should be developed and deployed in a manner that considers the potential risks, societal impact, and the precautionary principle, ensuring that benefits outweigh potential harm.

2.4.10 Adaptability and Responsiveness: The governance framework established by this treaty should be adaptable to technological advancements and evolving societal needs, promoting continuous monitoring, evaluation, and updates as necessary.

2.4.11 Non-Maleficence: AI development and deployment shall prioritize the prevention of harm, minimizing potential risks and negative consequences to individuals, communities, and the environment.

2.4.12 Public Engagement and Participation: The public should be actively engaged in discussions, decision-making processes, and the governance of AI technologies, ensuring inclusivity, transparency, and democratic values.

*Article 3*

3.1 All parties shall immediately suspend the operation and refinement of large GPU and TPU clusters, which are known to support the development of the most powerful AIs, subject to stringent risk assessment procedures based on their proposed approach.

3.2 Parties shall provide detailed justifications, demonstrating how their approach significantly reduces AI risks, while considering political feasibility. The AI Governance Council (AIGC) established in Article 6 shall review and assess the justifications provided.

*Article 4*

4.1 Parties shall adopt regulations and guidelines to limit the parameters of Machine Learning (ML) models, subject to stringent risk assessment procedures that consider the proposed approach of each party.

4.2 Parties shall provide detailed justifications for their chosen parameter limits, demonstrating how they effectively reduce AI risks. The AIGC shall review and assess the justifications to ensure a balance between comprehensiveness and effectiveness.

*Article 5*

5.1 Parties shall enact and enforce national legislation prohibiting the use of quantum computers for AI-related activities, subject to stringent risk assessment procedures that take into account the proposed approach of each party.

5.2 Parties shall provide detailed justifications for their prohibition of quantum computers, demonstrating how it effectively reduces AI risks. The AIGC shall review and assess the justifications to ensure a balance between comprehensiveness and effectiveness.

*Article 6*

6.1 Parties shall implement a moratorium on large-scale AI capabilities research and development, subject to stringent risk assessment procedures that consider the proposed approach of each party.

6.2 Parties shall provide detailed justifications for their chosen scope and duration of the moratorium, demonstrating how it effectively reduces AI risks. The AIGC shall review and assess the justifications, ensuring a balance between comprehensiveness and effectiveness.

*Article 7*

7.1 Parties shall enact and enforce national laws criminalizing the intentional development, deployment, or use of AI for criminal activities, subject to stringent risk assessment procedures that align with the proposed approach of each party.

7.2 Parties shall provide detailed justifications for their chosen criminalization measures, demonstrating how they effectively address AI-enabled crimes. The AIGC shall review and assess the justifications, ensuring a balance between comprehensiveness and effectiveness.

*Article 8*

8.1 Parties shall establish legal frameworks and procedures to provide redress for individuals or entities harmed by the use of AI, subject to stringent risk assessment procedures that consider the proposed approach of each party.

8.2 Parties shall provide detailed justifications for their chosen redress mechanisms, demonstrating how they effectively address harm caused by AI. The AIGC shall review and assess the justifications, ensuring a balance between comprehensiveness and effectiveness.

*Article 9*

9.1 The AI Governance Council (AIGC) shall be established as an independent international oversight body, composed of representatives appointed by each party, to ensure the effective implementation and enforcement of this treaty.

9.2 The AIGC shall provide guidance and flexibility to parties, allowing them to focus on necessary provisions aligned with their proposed approach, while ensuring the overall effectiveness and coherence of the treaty.

*Article 10*

10.1 Parties have the right to engage in responsible AI research, access and share information, participate in decision-making processes, and benefit from AI advancements.

10.2 Parties shall adhere to the principles of this treaty, establish regulatory frameworks, conduct risk assessments, ensure transparency and accountability, protect privacy, cooperate internationally, enforce legal and ethical accountability, monitor impacts, promote sustainability, protect human rights, and support research and education.

*Article 11*

11.1 An independent regulatory body, the AI Governance Commission (AIGC), shall be established to oversee treaty implementation.

11.2 The AIGC shall develop guidelines, conduct risk assessments, foster international cooperation, and monitor compliance with the treaty.

11.3 Parties shall provide regular reports to the AIGC, which will review compliance, issue recommendations, and ensure transparency and accountability.

*Article 12*

12.1 Licensing and Permitting:

12.1.1 Parties shall establish a licensing and permitting framework to regulate the development, deployment, and use of AI technologies, ensuring compliance with ethical, safety, and legal requirements.

12.1.2 The licensing and permitting process shall include the evaluation of technical expertise, risk assessments, and adherence to established guidelines, to ensure responsible and accountable AI development and deployment.

12.2 Technical Standards and Specifications:

12.2.1 Parties shall develop and implement technical standards and specifications for AI technologies, promoting interoperability, transparency, and compatibility.

12.2.2 The technical standards and specifications shall cover areas such as data privacy, security, fairness, explainability, and accountability to ensure the ethical and reliable operation of AI systems.

12.3 Monitoring and Compliance:

12.3.1 Parties shall establish monitoring and compliance mechanisms to ensure the effective enforcement of regulatory measures related to AI technologies.

12.3.2 The monitoring and compliance mechanisms shall include regular inspections, audits, and assessments of AI systems to verify compliance with regulatory requirements and standards.

12.4 Reporting and Information Exchange:

12.4.1 Parties shall establish reporting mechanisms to facilitate the exchange of information on AI development, deployment, incidents, and lessons learned.

12.4.2 The reporting mechanisms shall ensure timely sharing of relevant information, research findings, and best practices to foster international cooperation and knowledge sharing in AI governance.

*Article 13*

13.1 Parties shall resolve disputes through peaceful means, such as negotiation or mediation, utilizing risk assessment procedures that consider political feasibility and align with established legal principles and frameworks.

131.2 The AIGC shall establish monitoring and compliance mechanisms that take into account the proposed approaches of parties, providing flexibility while ensuring the overall effectiveness and coherence of the treaty.

*Article 14*

14.1 The AIGC shall recommend appropriate measures to address non-compliance, utilizing risk assessment procedures that consider political feasibility and are proportionate to the nature and severity of the non-compliance.

14.2 Parties shall consider the recommendations of the AIGC in good faith and take appropriate actions to rectify any non-compliance, while maintaining the balance between global cooperation and national autonomy.

*Article 15*

15.1 This treaty shall remain in effect until it is universally agreed that it is safe and ethical to resume large-scale AI capabilities research and development, based on a comprehensive assessment of risks and benefits, considering the proposed approaches of parties.

15.2 Parties shall conduct periodic reviews of the treaty, at least once every five years, to assess its effectiveness, consider adjustments, and evaluate the alignment with established legal principles, frameworks, and risk assessment procedures.

*Article 16*

16.1 This treaty shall be open for signature and ratification by all nations.

16.2 The treaty shall enter into force upon the deposit of instruments of ratification by a sufficient number of parties, as determined by the AIGC.