**International Agreement on the Regulation and Oversight of Artificial Intelligence Development**

**Preamble:**

The participating nations, recognizing the potential benefits and risks associated with the development and deployment of Artificial Intelligence (AI) technologies, acknowledging the need for comprehensive international governance and legal structures, and inspired by the movement to pause further AI development, have come together to draft this treaty. The treaty aims to foster a safe, ethical, and globally responsible approach to AI development, ensuring the protection of humanity and the environment.

Conscious of the transformative power of AI technologies and their potential to revolutionize various sectors, including healthcare, transportation, finance, and communication, the participating nations recognize the imperative of establishing a robust framework that guides the development, deployment, and use of AI systems. This treaty, grounded in principles of precaution, transparency, accountability, and human-centricity, seeks to address the challenges associated with the rapid advancement of AI, while leveraging its potential for the betterment of society.

Emphasizing the risks of uncontrolled AI development, particularly in the realm of highly autonomous systems, the participating nations endorse a moratorium on large-scale AI capabilities research. This pause allows for comprehensive risk assessments, regulatory framework development, and responsible advancement of AI technologies.

Acknowledging the significance of large GPU and TPU clusters in refining AI systems, the participating nations recognize the need for temporary shutdowns. This enables thorough evaluation, risk mitigation, and the establishment of stalwart regulatory mechanisms.

Furthermore, the participating nations highlight the potential of quantum computing in enhancing AI capabilities. They stress the importance of careful evaluation and international cooperation to integrate quantum computing responsibly, guided by globally recognized standards and guidelines.

Appreciating the risks of uncontrolled AGI and ASI development, the participating nations commit to national legislation criminalizing unauthorized AGI and ASI activities.

Recognizing the dynamic nature of AI technologies, the participating nations commit to periodic reviews, ensuring adaptability to scientific advancements and emerging global consensus. Inclusive decision-making processes and stakeholder input inform treaty adjustments.

With the intent of establishing a robust and internationally accepted framework for the regulation and oversight of AI development, the participating nations hereby adopt this treaty as a cornerstone of their commitment to the responsible and ethical advancement of AI technologies.

**Article 1: Scope and Definitions**

1.1 Scope: This treaty applies to all activities related to the development, deployment, and use of AI technologies, encompassing both narrow AI systems and potentially disruptive technologies such as Artificial General Intelligence (AGI) and Artificial Superintelligence (ASI). The scope includes AI systems developed for commercial, governmental, or research purposes, regardless of their intended application domain.

1.2 Definitions: For the purposes of this treaty:

(a) "AI" refers to systems or machines that demonstrate capabilities for autonomous decision-making or learning based on data or algorithms. It includes machine learning, neural networks, expert systems, and other related technologies. AI encompasses both software-based systems and hardware components designed specifically for AI purposes.

(b) "Large GPU and TPU clusters" refers to the high-performance computing infrastructures used for large-scale AI training, incorporating Graphics Processing Units (GPUs) and Tensor Processing Units (TPUs). These clusters consist of interconnected nodes specifically optimized for parallel processing, enabling the rapid refinement of AI models through intensive computational operations.

(c) "ML models" refers to machine learning models, which are mathematical representations trained on data to make predictions or decisions. These models encompass various techniques such as deep learning, reinforcement learning, and supervised learning, and they form the foundation for many AI applications.

(d) "Parameters" refers to the number of adjustable variables or weights within an ML model, influencing its complexity and capabilities. Parameters determine the model's capacity to learn and generalize from data, with larger parameter counts often associated with more powerful and complex AI systems.

(e) "Quantum computers" refers to advanced computing systems leveraging principles of quantum mechanics, which may enhance certain AI algorithms. Quantum computers operate based on the principles of superposition and entanglement, offering potential computational advantages over classical computers for specific AI tasks, such as optimization, simulation, and cryptography.

(f) "Artificial General Intelligence (AGI)" refers to AI systems capable of outperforming humans in most economically valuable work, exhibiting general-purpose intelligence. AGI systems possess cognitive abilities that rival or surpass human intelligence across a broad range of tasks, enabling them to adapt and learn in various contexts.

(g) "Artificial Superintelligence (ASI)" refers to AI systems surpassing human capabilities in virtually all domains, including problem-solving and decision-making. ASI systems exhibit levels of intelligence that far exceed human cognitive capacities, potentially leading to a transformative impact on society and requiring careful regulation and oversight.

**Article 2: Moratorium on Large-Scale AI Capabilities Research and Development**

2.1 Moratorium: Until it is universally agreed that it is safe and ethical to resume large-scale AI capabilities research and development, all participating nations shall impose a general moratorium on such activities. This moratorium encompasses research, funding, and resource allocation towards the development of AI systems capable of AGI or ASI. The moratorium seeks to provide a dedicated period for in-depth evaluation, risk assessment, and the establishment of comprehensive regulatory frameworks.

2.2 Research Focus: During the moratorium, participating nations are encouraged to prioritize research efforts towards AI safety, ethics, robustness, interpretability, and mitigating potential risks associated with AI technologies. Research initiatives shall aim to address challenges such as bias mitigation, explainability, fairness, privacy, accountability, and the prevention of unintended consequences.

2.3 Ethical Guidelines: Participating nations shall establish comprehensive ethical guidelines for AI development and deployment, addressing issues such as privacy, fairness, transparency, accountability, and the prevention of discriminatory practices. These guidelines shall be aligned with universally recognized human rights principles and take into account cultural, social, and economic contexts. Ethical considerations shall be integrated into AI research, development, deployment, and decision-making processes, promoting the responsible use of AI technologies.

2.4 International Research Collaboration: Participating nations shall actively promote and facilitate international collaboration in research projects related to AI safety and ethics. Collaborative efforts shall include knowledge sharing, joint funding initiatives, and the establishment of research networks to leverage diverse perspectives, expertise, and resources for addressing global challenges. International research collaborations shall prioritize the development of robust methodologies, tools, and frameworks for evaluating AI risks, ensuring the alignment of research practices with international ethical standards.

2.5 Responsible Innovation: Participating nations shall encourage responsible innovation in AI, promoting multidisciplinary approaches that involve not only AI experts but also professionals from fields such as philosophy, social sciences, law, and ethics. This interdisciplinary collaboration aims to anticipate and address potential ethical and societal implications of AI technologies. Participating nations shall foster a culture of responsible AI innovation, emphasizing values such as human well-being, transparency, fairness, and accountability throughout the AI development lifecycle.

**Article 3: Shuttering of Large GPU and TPU Clusters**

3.1 Shutdown: Recognizing the potential risks associated with the refinement of highly powerful AI systems, participating nations shall enforce the immediate shutdown of large GPU and TPU clusters used for AI refinement. This measure aims to prevent the development of AI systems with unintended consequences and to enable a comprehensive review of existing AI technologies. The shutdown shall be implemented within [specified time frame] from the entry into force of this treaty.

3.2 Asset Repurposing: Participating nations shall establish policies and mechanisms to repurpose the computing resources of large GPU and TPU clusters for AI safety research, infrastructure for AI education and training, public computing initiatives, or other socially beneficial purposes as identified by national authorities. Repurposing efforts should prioritize projects aligned with the responsible development and deployment of AI. Transparent and accountable processes shall be established to ensure fair resource allocation and effective utilization of repurposed assets.

3.3 Security Measures: Participating nations shall implement robust security measures to protect the shutdown and repurposed assets of large GPU and TPU clusters, ensuring that the computing infrastructure remains secure from unauthorized access, data breaches, and tampering. Security protocols shall be developed and enforced to safeguard the integrity, confidentiality, and availability of sensitive AI-related information and resources.

3.4 Resource Redistribution: Participating nations shall collaborate to develop mechanisms for the equitable redistribution of computational resources, including access to high-performance computing capabilities, to ensure that nations with limited resources can actively participate in AI research and development activities. Resource redistribution initiatives shall prioritize capacity building, knowledge transfer, and equitable opportunities for nations at different stages of AI development.

3.5 Assessment and Monitoring: Participating nations shall establish monitoring and assessment mechanisms to evaluate the impact and effectiveness of the shutdown of large GPU and TPU clusters. These mechanisms shall involve the collection and analysis of data on the use of repurposed assets, advancements in AI safety research, the availability of computing resources, and the extent to which the shutdown has contributed to the mitigation of risks associated with AI technologies. Regular assessments shall inform adjustments to resource allocation strategies and the optimization of AI development frameworks.

3.6 Transitional Support: Participating nations shall provide transitional support to stakeholders affected by the shutdown of large GPU and TPU clusters, including researchers, developers, and organizations involved in AI refinement. Support measures may include financial assistance, retraining programs, knowledge exchange initiatives, and infrastructure development to facilitate the transition towards alternative AI development methodologies.

**Article 4: Prohibition on Training ML Models**

4.1 Parameter Limitation: Acknowledging that extremely large ML models pose challenges in terms of interpretability, resource consumption, and potential risks, participating nations shall prohibit the training of ML models or combinations of models with more than 500 million parameters. This parameter limitation aims to strike a balance between innovation and responsible AI development, ensuring that AI systems remain manageable, explainable, and accountable.

4.2 Research and Development Exemptions: Exceptions to the parameter limitation may be granted for specific research and development projects deemed critical to AI safety, ethical considerations, or addressing societal challenges, subject to the approval of competent national authorities and review by the AI Governance Council. Requests for exemptions shall undergo rigorous evaluation, considering factors such as the scientific merit, potential benefits, and adherence to ethical guidelines. Approvals shall be accompanied by clear justifications and safeguards to prevent misuse or unintended consequences.

4.3 Model Explainability and Accountability: Participating nations shall promote research and development efforts focused on enhancing the explainability and interpretability of AI models, enabling individuals and stakeholders to understand and challenge the decision-making processes of AI systems. Transparent AI systems shall be encouraged to foster accountability and prevent the emergence of "black box" AI technologies. Participating nations shall support the development and adoption of methodologies, tools, and standards for model explainability, ensuring that AI systems provide meaningful and understandable explanations for their decisions and actions.

4.4 Bias Mitigation: Participating nations shall actively address and mitigate biases in AI systems, promoting fairness, non-discrimination, and equal opportunity. Measures should be taken to ensure that AI systems do not perpetuate existing biases or create new forms of bias in decision-making processes, particularly in areas such as hiring, lending, criminal justice, and healthcare. Participating nations shall invest in research and development efforts to understand, identify, and mitigate biases in AI algorithms, datasets, and decision-making processes, ensuring that AI technologies contribute to inclusive and equitable societies.

4.5 Data Quality and Diversity: Participating nations shall prioritize the availability of high-quality, diverse, and representative datasets for training ML models. They shall support initiatives that promote data collection practices adhering to ethical standards, privacy protections, and transparency requirements. Participating nations shall foster data sharing collaborations, ensuring equitable access to datasets and promoting initiatives that enhance the inclusivity and fairness of AI technologies.

4.6 International Standardization: Participating nations shall collaborate to establish internationally recognized standards and guidelines for ML model development and training processes, including parameter limitations, transparency requirements, bias mitigation strategies, and data quality assurance measures. These standards shall be developed through a transparent and inclusive process, involving participation from academia, industry, civil society organizations, and affected communities. International standardization efforts shall aim to promote consistency, transparency, and interoperability in AI systems, facilitating responsible AI development and deployment at a global scale.

**Article 5: Prohibition on Quantum Computers in AI Activities**

5.1 Prohibition: Recognizing the potential implications of quantum computing in accelerating AI capabilities, participating nations shall prohibit the use of quantum computers in any AI-related activities during the moratorium period. This measure intends to avoid unforeseen risks arising from the combination of advanced quantum computing and AI technologies. The prohibition applies to both research and development activities, as well as the deployment and use of AI systems enhanced by quantum computing.

5.2 Research and Evaluation: Participating nations are encouraged to invest in research and evaluation of the potential impact of quantum computing on AI development, including its benefits, challenges, and necessary safeguards. Such research should align with the principles of AI safety, ethics, and international cooperation. Participating nations shall collaborate to advance scientific understanding, risk assessment methodologies, and the development of guidelines that ensure the responsible integration of quantum computing into AI systems.

5.3 Quantum Computing Standards: Participating nations shall collaborate to develop internationally recognized standards and guidelines for the responsible integration of quantum computing into AI systems, addressing security, fairness, interpretability, and the prevention of unintended consequences. These standards shall consider the unique characteristics and challenges posed by quantum computing technologies. Standardization efforts shall involve the participation of experts from quantum computing, AI, cryptography, and relevant disciplines to ensure robust and comprehensive guidelines.

5.4 Risk Assessments: Participating nations shall conduct comprehensive risk assessments of the potential security threats and vulnerabilities associated with the combination of quantum computing and AI technologies. These risk assessments shall involve the evaluation of potential attacks, privacy breaches, data integrity risks, and the impact on existing cryptographic mechanisms. The findings of these risk assessments shall inform the development of safeguards and mitigation strategies to protect against emerging risks. Participating nations shall collaborate to exchange information on risk assessments, facilitating a comprehensive understanding of potential threats and the development of effective countermeasures.

5.5 Quantum Computing Education and Awareness: Participating nations shall prioritize the development of educational programs and initiatives to enhance understanding and awareness of quantum computing among AI practitioners, policymakers, and the public. These education initiatives shall address the underlying principles, capabilities, limitations, and potential applications of quantum computing in the context of AI technologies. Participating nations shall promote multidisciplinary collaborations between quantum computing experts, AI researchers, and policymakers to foster informed decision-making, ethical considerations, and responsible governance in the context of quantum-enhanced AI technologies.

**Article 6: Criminalization of AGI and ASI Development**

6.1 National Laws: Participating nations shall enact comprehensive national legislation criminalizing the development, deployment, or use of any form of Artificial General Intelligence (AGI) or Artificial Superintelligence (ASI). This measure emphasizes the collective commitment to prevent the uncontrolled emergence of highly autonomous AI systems with unknown consequences. National laws shall establish legal frameworks that deter and penalize violations of AGI and ASI development regulations.

6.2 Legal Frameworks: Participating nations shall establish legal frameworks outlining the specific offenses, penalties, and jurisdictional procedures regarding the development, deployment, and use of AGI and ASI technologies. These legal frameworks shall reflect the severity and potential consequences of unauthorized AGI and ASI development, ensuring proportionate penalties and deterrent measures. Legal frameworks shall also establish mechanisms to address emerging challenges, evolving technological landscapes, and the ethical considerations associated with AGI and ASI technologies.

6.3 International Cooperation in Law Enforcement: Participating nations shall enhance international cooperation in law enforcement, sharing information, intelligence, and best practices to detect, investigate, and combat illegal AGI and ASI development activities. Mutual legal assistance frameworks shall be strengthened to facilitate the prosecution of individuals or entities involved in the unauthorized development or malicious use of AGI or ASI technologies. Collaborative efforts shall involve the sharing of knowledge and expertise, joint investigations, and the exchange of forensic evidence to ensure effective enforcement of AGI and ASI development regulations.

6.4 Reporting Mechanisms: Participating nations shall establish reporting mechanisms to encourage individuals and organizations to report suspected AGI or ASI development activities that may violate this treaty. Whistleblower protection measures shall be implemented to safeguard individuals who come forward with credible information related to illegal AGI or ASI development. Reporting mechanisms shall provide secure channels for reporting, ensuring confidentiality and protection against retaliation for those reporting in good faith.

6.5 Public Awareness and Education: Participating nations shall launch comprehensive public awareness campaigns and educational initiatives to inform the general public about the risks, benefits, and ethical implications associated with AGI and ASI technologies. The campaigns shall emphasize the collective responsibility of society in shaping the development and deployment of AI technologies. Participating nations shall promote public dialogue, inclusive consultations, and citizen engagement to ensure that the concerns and perspectives of diverse communities are considered in the governance of AGI and ASI technologies.

**Article 7: Establishment of International Oversight Body**

7.1 Creation of Body: An international body, hereinafter referred to as the AI Governance Council, shall be established to oversee the implementation, regulation, and enforcement of this treaty. The AI Governance Council shall act as a central authority for coordinating international efforts in AI governance and promoting global cooperation. The council shall serve as a forum for discussion, exchange of information, and collaboration on AI-related matters.

7.2 Composition: The AI Governance Council shall consist of representatives from participating nations, selected in a fair and equitable manner, taking into account expertise in AI research, ethics, policy, legal frameworks, multidisciplinary perspectives, and regional representation. The council shall strive for balanced participation and diverse representation to ensure inclusive decision-making and the consideration of a wide range of perspectives.

7.3 Functions: The AI Governance Council shall have the following functions and powers:

* (a) Develop comprehensive guidelines, principles, and standards for AI development, deployment, and use, ensuring adherence to safety, ethics, human rights, and societal well-being. These guidelines shall reflect the evolving nature of AI technologies and incorporate inputs from relevant stakeholders, including academia, industry, civil society organizations, and affected communities. The guidelines shall provide a framework for responsible AI development, deployment, and use, guiding participating nations in the formulation of national policies and regulatory frameworks.
* (b) Monitor compliance with this treaty, including the enforcement mechanisms detailed in Article 8, through audits, inspections, and evaluations conducted by qualified experts. The AI Governance Council may request participating nations to provide relevant information, conduct site visits, or facilitate the assessment of AI systems to ensure compliance. The council shall establish procedures and guidelines for compliance monitoring, ensuring transparency, fairness, and accountability in the evaluation process.
* (c) Facilitate international collaboration and information sharing on AI-related research, best practices, emerging risks, and technological advancements. The AI Governance Council shall establish platforms for knowledge exchange, coordination of research initiatives, and the dissemination of AI-related insights and experiences. It shall foster a collaborative environment that encourages the sharing of lessons learned, data, and methodologies to advance the responsible development and deployment of AI technologies globally.
* (d) Foster capacity-building initiatives to enhance the technical, ethical, and legal expertise of participating nations in AI development and governance. These initiatives may include training programs, workshops, and joint research projects aimed at empowering nations to develop effective AI policies, strategies, and regulatory frameworks.
* (e) Review and update the treaty periodically to reflect technological advancements, evolving global perspectives, emerging challenges related to AI development and deployment, and the feedback received from participating nations and relevant stakeholders. Treaty revisions shall be based on rigorous scientific assessments, risk analysis, and consensus-building processes. The AI Governance Council shall establish mechanisms and timelines for the review and revision of the treaty, ensuring transparency, inclusivity, and the participation of all relevant stakeholders.

**Article 8: Mechanisms for Enforcement**

8.1 Compliance Monitoring: Participating nations shall establish robust mechanisms for monitoring and ensuring compliance with this treaty within their respective jurisdictions. These mechanisms shall include national AI regulatory bodies or designated authorities responsible for overseeing AI development and enforcing the provisions of this treaty. Participating nations shall empower these bodies with sufficient resources, expertise, and enforcement powers to effectively monitor compliance. Compliance monitoring shall include regular audits, inspections, and assessments to verify adherence to the treaty provisions.

8.2 Reporting Obligations: Participating nations shall regularly report to the AI Governance Council on their progress in implementing the provisions of this treaty, providing information on legislative measures, research initiatives, enforcement actions, emerging risks, and notable case studies. Reporting shall be conducted in a standardized format to facilitate comparative analysis and identification of best practices. The AI Governance Council shall establish reporting guidelines and templates, ensuring consistency and accuracy in the information provided by participating nations.

8.3 Sanctions and Remedial Measures: In cases of non-compliance with the treaty provisions, the AI Governance Council may recommend appropriate measures to the participating nations, including but not limited to diplomatic actions, economic sanctions, technical assistance, public awareness campaigns, or capacity-building initiatives to facilitate compliance. The severity of sanctions and remedial measures shall be proportionate to the nature and extent of the violation, with due consideration for the potential impact on international cooperation and the development of responsible AI technologies. The AI Governance Council shall establish procedures for sanctioning non-compliant nations, ensuring due process and the opportunity for affected parties to present their case.

8.4 Dispute Resolution: Participating nations shall resolve disputes arising from the interpretation or application of this treaty through peaceful means, including negotiation, mediation, or other mutually agreed-upon mechanisms. In the event of an unresolved dispute, participating nations may submit the matter to an independent arbitration body established by the AI Governance Council. The arbitration body shall operate in accordance with agreed-upon procedures and principles, ensuring impartiality, transparency, and timely resolution of disputes.

**Article 9: Treaty Duration and Reevaluation**

9.1 Treaty Duration: This treaty shall remain in force until it is universally agreed that it is safe and ethical to resume large-scale AI capabilities research and development. The AI Governance Council, in consultation with participating nations, shall periodically review the conditions necessary to resume such activities, considering scientific advancements, risk assessments, societal readiness, and the evolving landscape of AI technologies. The review shall be conducted at regular intervals.

9.2 Reevaluation: The AI Governance Council shall conduct regular reviews and reevaluations of the treaty's provisions to assess the progress made in ensuring AI safety, ethics, and responsible development. These reviews shall take into account advancements in AI technologies, scientific knowledge, societal needs, global consensus, and the inputs received from participating nations, academia, industry, civil society organizations, and affected communities. The reevaluation process shall be inclusive, transparent, and evidence-based, involving stakeholder consultations, expert assessments, and international cooperation.

9.3 Treaty Amendment: Participating nations may propose amendments to this treaty, subject to the approval of the AI Governance Council. Amendments shall be adopted by a majority vote of the participating nations and shall enter into force after the deposit of instruments of ratification by a majority of participating nations. Amendments shall respect the principles and objectives of the treaty, ensuring the continued effectiveness, relevance, and adaptability of the treaty in light of technological advancements and emerging challenges.

**Article 10: Ratification and Entry into Force**

10.1 Ratification: This treaty shall be subject to ratification by participating nations in accordance with their respective domestic procedures. Each participating nation shall deposit its instrument of ratification with the AI Governance Council.

10.2 Entry into Force: This treaty shall enter into force upon the deposit of instruments of ratification by a majority of participating nations. For nations that join the treaty after its entry into force, their instruments of ratification shall be deposited with the AI Governance Council, and the treaty shall enter into force for those nations thirty days after the deposit.

10.3 Amendments and Protocols: Participating nations may propose amendments or protocols to this treaty, subject to the approval of the AI Governance Council. Amendments or protocols shall be adopted by a majority vote of the participating nations and shall enter into force after the deposit of instruments of ratification by a majority of participating nations.

10.4 Review Conferences: The AI Governance Council shall convene periodic review conferences to assess the implementation and effectiveness of this treaty, facilitate knowledge sharing, and promote dialogue among participating nations. The outcomes of review conferences shall inform further actions and adjustments to enhance the treaty's impact. The AI Governance Council shall establish the frequency and format of review conferences, ensuring their accessibility, inclusivity, and relevance to the evolving challenges and opportunities in the field of AI.

In witness whereof, the undersigned, duly authorized by their respective governments, have signed this treaty.

**Signatures of Participating Nations:**