Student Competition for drafting treaty on moratorium of large-scale AI capabilities R&D

International Treaty on the Pause and Governance of Artificial Intelligence Development

The States in this Treaty, which will be referred to as the “Parties to the Treaty”

Regarding the potential risks and detrimental impacts that may ensue upon mankind by the exponential and uncontrollable levels of growth of AI and the consequent obligation to make every effort to prevent the possibility of human destruction and ensure the protection of peoples,

Operating under the belief that the uninhibited development of AI would ameliorate the possibility of human displacement and further unforeseen risks posed to human safety,

Recognizing the rapid advancements and the vast potential of AI technology, the Parties to the Treaty should understand the emphasis on how international cooperation and coordination are vital in addressing the ethical and safety implications related to the development of AI,

Asserting the principle that the benefits of safe, controlled application and development of Artificial Intelligence, including its uses in increasing general human welfare globally, should be available to all and used to expand humankind’s productive capacity,

Whilst acknowledging that unregulated and uncontrolled growth of AI models will likely have catastrophic consequences on all peoples in the future,

Stating the need to ensure responsible and beneficial use of AI for the safety of humanity, we must recognize the need for a temporary pause in large-scale AI capabilities research and development until we have sufficient legal mechanisms in place,

Recognizing the need for an international coordination in this treaty, The United Nations and its principal organs have been designated as the international body responsible for overseeing the implementation, enforcement and monitoring of the treaty,

Have agreed as follows:

Article I

A general moratorium shall be enacted upon the development and establishment of large-scale AI capabilities and research until all Parties to the Treaty and the United Nations have agreed that the risks and negative implications of using large-scale AI are no longer present. Parties to the treaty shall suspend all large-scale AI capabilities research- a group of significant, specific areas involved in large-scale AI capabilities and research will be addressed in separate articles throughout the treaty on the understanding that the research and risk management that will be conducted for each specific area will progress at a different rate and therefore will have separate conditions for expiration.

For areas of large-scale AI capabilities and research which are not specifically addressed in an article of this treaty, the moratorium will remain in place until all Parties to the Treaty and the United Nations have unanimously agreed that it safe and ethical for the general operations of large-scale AI capabilities and research to resume.

Article II

1. Each of the parties to the treaty will immediately undertake the suspension of the operation of all GPU and TPU clusters that are within their jurisdiction following the implementation of this treaty.
2. This suspension of GPU and TPU clusters shall be in effect until it has been universally agreed upon by all parties based upon the recommendation of the United Nations that it is safe, ethical and necessary for GPU and TPU clusters to resume their operation.
3. All parties to the treaty will reconvene every two years, having fully evaluated the implications and risks involved in the development of AI within these environments, to discuss whether all parties agree on continuing to act under this treaty.
4. Only when parties unanimously agree that sufficient research has been carried out and the necessary legal framework is in place in all jurisdictions such that GPU and TPU clusters are reasonably deemed to be able to be run safely and ethically, can parties vote to resume the operation of GPU and TPU clusters.
5. Parties shall undertake research and evaluation to determine the potential consequences of AI development within GPU and TPU clusters, outlining the risks of their operation and the potential solutions that can be implemented to eradicate these risks.
6. The parties shall undertake continuous research and evaluation of the impact that this suspension has made on the development of AI technology within their own jurisdictions. Parties will also research and evaluate the impact of this suspension of GPU and TPU clusters on all industries for which AI is used within each jurisdiction.
7. Parties shall provide annual reports on their research and progress to the United Nations, including all ways in which the suspension of GPU and TPU clusters have made significant impacts within their jurisdictions.
8. All parties will put the appropriate mechanisms in place to guarantee full compliance with these provisions stated in Article I and to prevent covert operations of GPU and TPU clusters.

Article III

1. Parties shall prohibit the training of ML (machine learning) models that exceed 500 million parameters, or multiple models combined together that exceed 500 million parameters collectively.
2. The treaty prohibits all the stages involved in ML model training. This includes pre-training, transfer learning, and any enhancements made to existing ML models.
3. Parties shall establish mechanisms for the purpose of verifying full compliance with the provisions in (1) and (2). This must include regular auditing procedures and testing procedures which are to be fully included in an annual report. Parties shall explain the nature of the mechanisms used to verify full compliance in the annual report which is referenced in (8).
4. The United Nations will offer guidelines and technical assistance to Parties to aid the implementation and enforcement of the prohibition.
5. Parties will undertake continuous research and evaluation of the impacts that the prohibition on training ML models with more than 500 million parameters have had on all industries which were previously using these models before the prohibition and parties shall include these impacts on industries in their annual report.
6. Parties will undertake continuous research and evaluation outlining the potential risks that training ML models with more than 500 million parameters may pose and the solutions to eradicate these risks in the event that the models are no longer prohibited.
7. Parties shall include complete information on all research and evaluation carried out in compliance with (6) in their annual report.
8. The Parties shall produce annual reports relating to the prohibition of training ML models with more than 500 million parameters. The reports shall contain the full information required in (3), (5), (6) and (7). These reports shall be submitted to the United Nations in the annual report and Parties shall reconvene every two years in order to engage in reviews and assessments of all reports provided by each Party to evaluate how effective the prohibition has been within the given time period and to analyse the progress made in the risk management of the models.
9. The prohibition on training ML models with more than 500 million parameters will be in effect until, under the guidelines of the United Nations, the expert analysis of the annual reports provided by the Parties and the existence of the appropriate legal framework in each jurisdiction, the Parties unanimously agree that it is safe and ethical for the operation of the models to resume.

Article IV

1. Parties shall prohibit the use of quantum computers in any AI-related activities within their jurisdiction, upon the implementation of this treaty. Parties shall not manufacture, develop or otherwise acquire quantum computers for the use of AI-related activities. The prohibition of the use of quantum computers for AI-related activities includes the prohibition of the use of quantum computing resources, algorithms or any other quantum methodology for any AI-related task.
2. Parties should recognize the necessity for extensive research, experimentation, and evaluation to develop a more comprehensive understanding of the implications and potential risks of the use of quantum computers for any AI-related activities and the ways in which these risks may be eradicated. Parties shall carry out this research, experimentation and evaluation continuously and shall include all findings of this research in their annual reports. Parties shall allocate sufficient resources to support these research initiatives and to provide funding for research initiatives focused on finding safer alternatives to their current systems which use quantum computers for AI-related activities.
3. Parties shall establish mechanisms for the purpose of verifying full compliance with the provisions in (1). This must include regular auditing procedures and testing procedures which are to be fully included in the annual report. Parties shall explain the nature of the mechanisms used to verify full compliance in the annual report.
4. The Parties shall produce annual reports on the prohibition of quantum computers in any AI-related activities. The reports shall contain the full information required in (2) and (3). These reports shall be submitted to the United Nations annually and Parties shall reconvene every two years in order to engage in reviews and assessments of all reports provided by each Party to evaluate how effective the prohibition has been within the given time period of two years and to analyse the progress made in the risk management of quantum computers in all AI-related activities.
5. The prohibition on the use of quantum computers in any AI-related activities will be in effect until, under the guidelines of the United Nations, the analysis of the annual reports provided by the Parties and the existence of the appropriate legal framework in each jurisdiction, the Parties unanimously agree that it is safe and ethical for the operation of the models to resume.

Article V

All Parties to the treaty shall, upon the enactment of this treaty, pass national laws criminalising the development of any form of Artificial General Intelligence (AGI) or Artificial Superintelligence (ASI), within their jurisdiction. These national laws which criminalise the development or manufacture of AGI or ASI shall include:

1. Prohibition of researching, experimenting or manufacturing efforts aimed at achieving ASI or AGI capabilities in violation of national law.
2. The prohibition of the provision of any form of support given to any projects or initiatives which aim to develop or manufacture AGI or ASI in violation of national law.
3. Establishment of strict legal consequences for any violation of the national laws corresponding to (a) and (b). These consequences should include sanctions, fines, criminal charges and the immediate seizure of any property being used for the development and manufacture of ASI or AGI.
4. The establishment of regulatory bodies and authorities which to be responsible for the complete oversight, monitoring and enforcement of all laws concerning the development and manufacture of AGI and ASI.
5. National governing bodies can be permitted to give license to entities of their choosing to be exempted from (a) and (b) if these entities are performing these acts for the purpose of research and evaluation outlining the potential risks that ASI and AGI may pose and the solutions to eradicate these risks in the event that the models are no longer prohibited. To be considered for an exemption, entities must allow full monitoring by the national government. Entities must clearly state that which they plan to research on, and the full details of the plans, the operational mechanisms of the entity and the safeguards in place to allow the entities to conduct their research safely- all the information must be submitted by the national governing body to the United Nations. Following the submission of an entity’s research plans, the United Nations will conduct an evaluation before allowing the National Governing Body to give license to an entity to be exempted from the laws outlined in (a) and (b) for the aforementioned research purposes. The National Governing Body shall monitor and report the actions of an entity which has been given license to be exempted from the laws outlined in (a) and (b) and shall send a quarterly report of the entity’s research activities to the United Nations. The quarterly reports submitted to the United Nations will be open to viewing for all Parties to the Treaty and shall be evaluated by the United Nations in order to determine whether the research activities being conducted are being done so in a safe way. If the United Nations is not satisfied with the level of safeguarding in place, the United Nations can strip an entity of its license to conduct these research activities, without the intervention of the National Governing Body.

The Parties shall produce annual reports on the impacts of the laws criminalising the development and manufacture of ASI and AGI. The reports shall contain full information on research, experimentation, and evaluation to develop a more comprehensive understanding of the implications and potential risks of the use of quantum computers for any AI-related activities and the ways in which these risks may be eradicated- conducted as per clause (e) of this treaty. These reports shall be submitted to the United Nations annually and Parties shall reconvene every two years in order to engage in reviews and assessments of all reports provided by each Party to evaluate how effective the prohibition has been within the given time period of two years and to analyse the progress made in the risk management of the development and manufacture of AGI and ASI. The laws criminalising the development and manufacture of ASI and AGI will be in effect until, under the guidelines of the United Nations, the analysis of the annual reports provided by the Parties and the existence of the appropriate legal framework in each jurisdiction, the Parties unanimously agree that it is safe and ethical for the development and manufacture of ASI and AGI to resume.

Article VI

Recognizing the necessity for the responsible, controlled and minimized development of Artificial Intelligence, all Parties to the treaty shall take the necessary measures to restrict access to their AI-related data for countries that have not agreed to be bound to this treaty. For the purposes of this treaty, AI-related data should be considered as any data that can be used by countries to further develop any AI-related technology or activities. However, this restriction of access to data does not apply to countries sharing information on safeguarding precautions that can be taken in the aim of making AI development safer and better regulated.

Article VII

Parties to this treaty shall allocate adequate resources to support research initiatives and ensure compliance with the provisions outlined in this treaty. These resources shall be dedicated to the following purposes:

1. Research Initiatives: Parties shall allocate sufficient financial, technological, and human resources to support the research projects which are aimed at ameliorating our understanding of AI’s implications, risks and potential solutions.
2. Compliance Mechanisms: Parties shall allocate resources in order to establish and maintain adequate compliance mechanisms at the national level.

Article VIII

Parties to the treaty shall put enforcement mechanisms in place to ensure the compliance of all individuals and entities within their jurisdiction, to all prohibitions, laws and suspensions stated in Articles I, II, III, IV and V. These enforcement mechanisms must include:

1. Regular audits and inspections to verify compliance with this treaty.
2. The enaction of legislation that establishes penalties and sanctions for the violations of this treaty.
3. Full reporting and transparency of the efforts made to enforce the treaty.
4. The establishment of mechanisms and legislation used to protect any individual or entity who reports violations of the treaty, by ensuring confidentiality and non-retaliation.

All enforcement mechanisms addressed in this article and all other articles of this treaty must be fully included in the annual reports which are addressed in Article X.

Article IX

1. In the event of any type of dispute arising between parties over the application, implementation or interpretation of this treaty, Parties shall make every effort to resolve the dispute through peaceful negotiations.
2. If Parties cannot reach an agreement through the process of direct negotiations, Parties shall utilize these mechanisms sequentially:
   1. Parties shall engage in mediation over the dispute. The mediation can be carried out by a mutually agreed mediator or a mediator which has been recognized and accepted by the United Nations as suitable for the matter.
   2. In the case that mediation does not lead to a satisfactory outcome, Parties may agree to refer the dispute to an independent panel of experts in AI governance and related fields, provided that the United Nations recognizes and accepts the panel as suitable to provide an assessment of the dispute.
   3. If the dispute remains unresolved after the expert review, Parties shall submit the dispute to binding arbitration, conducted by an internationally recognized arbitration institution, provided that the institution has been deemed acceptable to arbitrate the matter by the United Nations.

Article X

Parties to the treaty will submit annual reports to the United Nations including all information required by and addressed in Articles I, II, III, IV, V and VI, containing all information on the impacts of the treaty and the research and evaluation conducted by each party on all areas of large-scale AI capabilities and research. These reports will be continuously reviewed by the United Nations and the other parties in order to assess the effects of the treaty and the progress achieved in risk mitigation. Parties shall reconvene every two years to discuss the contents of the reports and evaluate whether the prohibitions and suspensions addressed in this treaty are still necessary. The specific areas of the large-scale AI addressed in Articles II, III, IV, and V, will be treated separately to the general moratorium of large-scale AI capabilities and research. Nevertheless, all preceding articles will be subject to the expiration conditions- upon review by all Parties to the Treaty and the United Nations, if it is unanimously agreed that it is safe and ethical to resume operation.

Article XI

The treaty will be open to the signature of all states and any state that does not sign the treaty before its enactment may sign at any time. The treaty shall enter into force after its ratification by the United Nations.

Article XII

Each party, exercising its national sovereignty, has the right to withdraw from the treaty if it determines that exceptional circumstances which are directed related to the matters of this Treaty, pose a serious threat to its national interests.

IN WITNESS WHEREOF, the draft of this treaty has been prepared and presented for presentation.

Cover Page, Index and Explanatory Note for the Moratorium Treaty Competition

International Treaty on the Pause and Governance of Artificial Intelligence Development

Index:

1. Introductory Section
   1. Preamble
   2. Purpose and Objectives
   3. Establishment of International Oversight Body
2. Article I: General Moratorium of Large-scale AI capabilities and Research
   1. Temporary Pause and Governance
3. Article II: Shutting down large GPU and TPU clusters
   1. Suspension of Operation
   2. Evaluation and Resumption
4. Article III: Prohibition of training ML models with more than 500 million parameters
   1. Prohibition and Compliance
   2. Reporting and Evaluation
5. Article IV: Prohibition of the use of quantum computers in any AI-Related Activities
   1. Prohibition and Evaluation
   2. Reporting and Oversight
6. Article V: pass national laws criminalising the development of any form of Artificial General Intelligence (AGI) or Artificial Superintelligence (ASI)
   1. Criminalisation and Enforcement
   2. Research Exemptions and Reporting
7. Article VI: Restriction of Access to AI-Related Data
8. Article VII: Resource Allocation and Research Initiatives
   1. Research Initiatives
   2. Compliance Mechanisms
9. Article VIII: Enforcement Mechanisms
   1. Audits and Inspections
   2. Penalties and Sanctions
   3. Reporting and Transparency
   4. Protection of Whistleblowers
10. Article IX: Dispute Resolution
    1. Negotiations
    2. Mediation
    3. Expert panel review
    4. Binding Arbitration
11. Article X: Annual Reports and Evaluation
12. Article XI: Treaty Signature and Entry into Force
13. Article XII: Right to Withdraw from the Treaty

Explanatory Note:

I believe that the majority of that which I have included in my treaty is quite self-explanatory and therefore, there is only one article for which I desire to provide extra explanations. In Article VI: Restriction of Access to AI-Related data, the stance I have taken of requiring Parties which have agreed to the treaty to restrict access of all data that can be used to improve AI systems in other countries which have not agreed to the treaty, may seem rather polarising but nevertheless I felt as though this provision was necessary to include as it is rather likely that not all countries would agree to a treaty for the suspension of large-scale AI capabilities and research and as a result these countries may see short-term benefits that the countries which did sign the agreement would not see. Accordingly, it seems that the most just course of action to take would to be to minimize the level of AI development that these countries can carry out.