

Submission to the Global Digital Compact

Contact
carlos@futureoflife.org

Dr. Amandeep Singh Gill
Secretary-General's Envoy on Technology
United Nations
New York, NY

Dear Dr. Amandeep Singh Gill,

The Future of Life Institute (FLI) is an independent non-profit organization that works on fostering the benefits of technology and mitigating its risks. FLI created one of the earliest sets of AI governance tools - the Asilomar Artificial Intelligence (AI) principles - and maintains a large network of global AI researchers. In addition, we participated in the United Nations (UN) Secretary General's Digital Cooperation Roadmap as the civil society champion for AI.

FLI welcomes the opportunity to provide feedback on the Global Digital Compact (GDC) organized by the Office of the Secretary-General's Envoy on Technology. In preparation for the Summit of the

Core principles

- Risk identification and mitigation: Stakeholders involved in the design, development, and

Risk identification and mitigation

AI systems have the potential to generate direct and indirect harms that derail the achievement of the UN's sustainable development goals. Increasingly, the geographic, demographic, or socio-economic borders that protect individuals from this technology's effects are falling away. This should prompt the GDC to invite stakeholders involved in the design, development, and deployment of AI systems to identify and mitigate their negative impact.

Our suggestions for identifying and mitigating the risks of AI systems are divided into research commitments and actions concerning deployment. First, we strongly believe that the existing gap between the resources allocated to AI safety research and the commercialization of AI systems needs to be addressed. Second, stakeholders involved in any aspect of the AI lifecycle must prioritize the implementation of an effective risk management system to prevent harms.

Prioritizing AI safety research

AI Safety is the field of research into “techniques for building AI that is beneficial for humans.”¹ The challenges in harnessing AI systems in a manner that minimizes harms to individuals are continuously evolving.² Yet resources to support this field pale in comparison to those dedicated to the rapid deployment of AI-based products and services. In other words, society is prioritizing short-term benefits over long-term resilience and stability. Thus, our recommendation is to request that stakeholders re-evaluate the importance of AI Safety funding in their research portfolios.

All of society benefits from devoting resources to a field whose main goal is to minimize the harms and improve the benefits of AI technology. Therefore, stakeholders must work together to identify the most pressing issues in AI safety and channel sufficient funding to address them. Importantly, when funding is allocated to this field, it must be distributed in a manner that builds research capacity around the world. This not only improves the diversity of viewpoints able to generate solutions, but also represents an opportunity to address the global AI safety research funding gap.

geographies and socio-economic characteristics through their technology.

Our first recommendation here centers on the products and services generated by stakeholders. In this regard, the UN should have a role in connecting technology companies to its sustainability agenda via the publication of concrete challenges. Secondly, there is a segment of for-profit firms willing to commit their excess profits for the common good. Along these lines, the UN should assess mechanisms such as the “Windfall Clause” to fund the attainment of its sustainable development goals.

