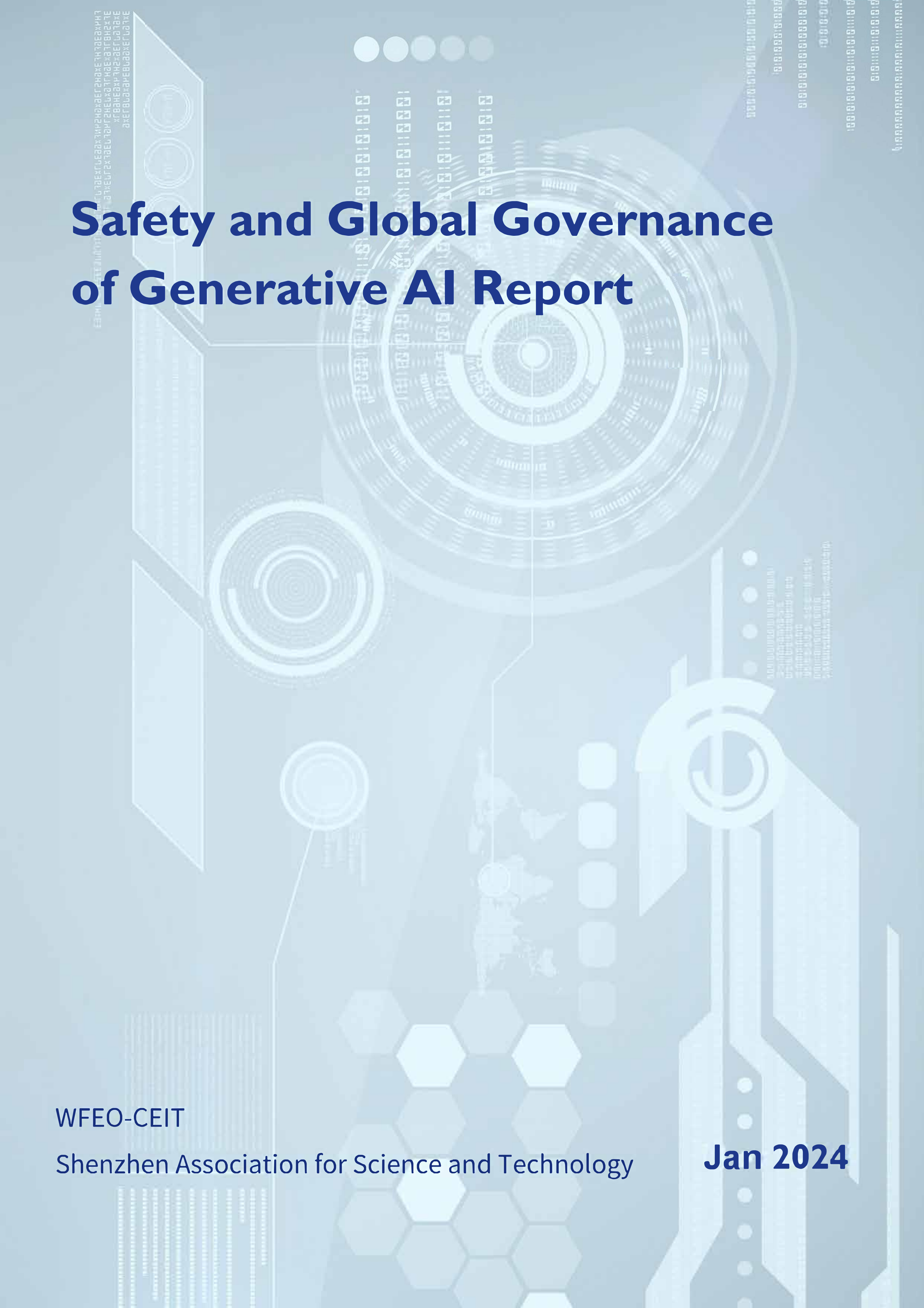

Safety and Global Governance of Generative AI Report

WFEO-CEIT

Shenzhen Association for Science and Technology

Jan 2024

Coordination, Cooperation, Urgency: Priorities for International AI Governance

Carlos Ignacio Gutierrez

The international governance of artificial intelligence (AI) is an inherently complex problem for which, as of late 2023, we have no clear solution. As this technology's capabilities increase at an unpredictable rate, over 190 national jurisdictions are tasked with managing its escalating and unforeseeable risks. To address these risks, independent action is clearly a sub-optimal approach since patching a problem in one place, will not prevent its spread to others. Instead, the effective governance of AI is a communal effort that requires global participation. Considering this, society should prioritize the development of a multilateral response that considers the following: What elements of AI should be governed and how? Who should be included in this governance process? When is the right time to act?

What and how: The options space for AI risks is characterized by its breadth and depth. Any number of issues could be proposed to pool the international community's attention. However, to succeed in catalyzing action, a candidate issue has to trigger a sense of commonality among countries with wide-ranging needs and capabilities. A proposal to jump-start the conversation is to focus on the mitigation of shared large-scale high-risk harms caused directly or indirectly by AI systems. The benefit of setting such a threshold is that it encompass a relatively narrow set of concerns. Moreover, it serves to centralize awareness and synchronize efforts, optimally through a multilateral organization with a concrete workstream composed of the following objectives: Identify vectors of shared large-scale high-risk harms produced by AI systems. Although many concerns will emanate from general purpose AI systems, the effort's remit must include narrow systems that qualify under its operating guidelines. Moreover, it should proactively inform stakeholders on potential risks and recognize existing vectors of harm. Coordinate global responses that are technically sound and consistent with best governance practices. This can take several shapes and depends on the scale and source of the problem at hand. For instance, a response can range from the solicitation of voluntary standards as a precautionary measure to the imposition of a compulsory set of rules as a reaction to an ongoing concern. Enforce adherence to agreed-upon actions that reduce the likelihood and impact of harms. Because the direct and indirect effects of AI are often unbound by jurisdiction, establishing an effective enforcement regimen requires maximizing the number of participating states. While the multilateral effort should be empowered to perform this role, it may also recruit, certify, or deputize public institutions and third-parties, often on a jurisdictional basis, to take on this task in order to scale its enforcement capabilities.

Who: Regardless of their capability to design, develop, or deploy AI technologies, all countries are vulnerable to AI's risks, and may wittingly or unwittingly host or shelter any part of the high-risk AI supply chain or the system itself. This is why cooperation must become a priority regardless of geography, a country's political system, or ideology. Essentially, no state should be excluded from engaging in multilateral action to address global AI concerns. The ability to mitigate shared large-scale high-risk harms depends on wide-spread participation. Thus, incentives should be considered for a range of countries, from those that are influential in the design, development, and deployment of AI to those with a role relatively limited to being subject to this technology‘s risks.

When: We face conditions where evermore powerful AI systems are deployed on a daily basis, and limited bandwidth is devoted to understanding what constitutes appropriate governance. This underscores the urgency of establishing a collective effort to proactively address AI risks. Even if efforts are undertaken today to begin multilateral coordination and cooperation, years will likely pass before a system is put in place. Therefore, in the short-term, it is understandable if an influential initial set of countries take the initiative to begin this multilateral process. This may include the membership of states that lead the world in the commercial deployment of systems, manufacturing of hardware, educating the technology's workforce, and/or establishing comprehensive regulation. In the long-term, the UN is the only organization with universal representation and the ability to host an effort such as the one described in this commentary. Ideally, it takes the reigns over the verification, coordination, and enforcement of efforts to mitigate the shared AI risks.

Conclusion: In optimizing multilateral governance, no "right" answers exist. What we can hope for is a multilateral AI governance scheme that prioritizes coordination, cooperation, and urgency in addressing shared large-scale high-risk harms. By focusing global attention on the mitigation of these issues, the international community needs to build the necessary commonality to achieve the only responsible end state for AI governance: one where the design, development, and deployment of this technology is safe and ethical.

Carlos Ignacio Gutierrez, artificial intelligence (AI) policy researcher at the Future of Life Institute, focusing on the impact of this technology's methods and application on hard law and AI's management through the design of effective and credible soft law programs.