

G7 Hiroshima AI Process
G7 Digital & Tech Ministers' Statement
(September 7, 2023)

1. Foreword

1. We, the G7 Digital and Tech Ministers, and partners, including the Organisation for Economic Co-operation and Development (OECD) and the Global Partnership on Artificial Intelligence (GPAI), met virtually on September 7th, 2023 for discussions on the opportunities and challenges of advanced artificial intelligence (AI) systems, with a focus on foundation models and generative AI as part of the G7 Hiroshima AI Process established by G7 Leaders. Building on the discussions on responsible AI and AI governance at the G7 Digital and Tech Ministers' meeting on April 29-30, and the work of the G7 Hiroshima AI Process, we endorse the following to be presented to G7 Leaders: (1) a report by the OECD summarizing a stocktaking of priority risks, challenges and opportunities of generative AI based on priorities highlighted in the G7 Hiroshima Leaders' Communiqué, (2) work towards international guiding principles applicable for all AI actors, (3) developing a code of conduct for organizations developing advanced AI systems and (4) project-based cooperation in support of the development of responsible AI tools and best practices.

2. We affirm our commitment to fostering an environment where trustworthy AI systems are designed, developed and deployed for the common good worldwide including in emerging and developing economies in furtherance of democracy, human rights, the rule of law, and our shared democratic values and interests. We oppose the misuse and abuse of AI in ways that undermine democratic values, suppress freedom of expression, and threaten the enjoyment of human rights. We reaffirm our commitment to promote the development and adoption of international standards and interoperable tools for trustworthy AI that enables innovation. To these ends, we reaffirm our commitment to promote human-centric and trustworthy AI based on the OECD Recommendation on Artificial Intelligence.

3. Most immediately and urgently, we recognize the need to manage new risks and challenges for individuals, society and democratic values and harness benefits and opportunities brought by advanced AI systems in particular foundation models and generative AI given the rapid advancement of these technologies. We commit to

the development and deployment of safe, secure, and trustworthy advanced AI systems that facilitate respect for human rights, promote inclusion, mitigate risk, and help solve society's greatest challenges including the climate crisis and achieving the Sustainable Development Goals (SDGs). We share the recognition that promotion of design, development, deployment, and responsible use of advanced AI systems requires appropriate guardrails, based on risks, and international cooperation with like-minded partners, including like-minded emerging economies and developing countries.

4. To these ends, we commit to develop guiding principles and an international code of conduct for organizations developing advanced AI systems to be presented to leaders for discussion.
5. We aim by the end of the year to develop a comprehensive policy framework, including overall guiding principles for all AI actors in the AI eco-system. The comprehensive framework will seek to support responsible AI innovation and help guide the development of regulatory and governance regimes, in line with our respective domestic approaches, and will benefit from stakeholder outreach and consultation. The comprehensive policy framework including guiding principles should be seen as a living document, which can be updated and complemented in the light of technological developments. Stakeholder outreach and consultation will help ensure the elaboration and implementation of the principles by key stakeholders around the world, including emerging economies and developing countries.
6. We commit to fostering an environment where safe, secure, and trustworthy AI systems are developed and deployed for the common good in furtherance of democracy, human rights, the rule of law, and our shared values and interests.

1. Understanding of Priority Risks, Challenges and Opportunities from the OECD's Report

7. From a report compiled and drafted by the OECD in July/August 2023, a range of risks and opportunities were identified as priorities and as the basis for our consideration on our common understanding, position, and future action, including for collective work on generative AI. For example, the report identified transparency,

disinformation, intellectual property rights, privacy and protection of personal data, fairness, security and safety, amongst others, as key areas of concern across G7 members. Opportunities such as productivity gains, promoting innovation and entrepreneurship, improving healthcare, and helping to solve the climate crisis were also identified. The risks and opportunities identified in the report will help inform our future work on advanced AI systems. We plan to also engage with stakeholders in academia, civil society, government, and industry to seek their views on these issues as part of our work.

8. We also recognize that further work and multi-stakeholder engagement is required on these complex issues and that an iterative process is needed to build approaches that can adapt as the technologies and risk landscape evolves.
9. As a first step, we welcome the OECD's Report Towards a G7 Common Understanding on Generative AI, developed by the OECD Secretariat based on input from the G7 Hiroshima AI process working group.

II. International Guiding Principles and Code of Conduct for Organizations Developing Advanced AI Systems

10. We recognize that various AI actors in the AI eco-system share the responsibility for addressing existing and emerging AI risks, and that developing guiding principles for all AI actors including AI developers, deployers, and users will be important. In particular, we recognize that organizations developing advanced AI systems, such as foundation models and generative AI play a critical role in this moment, and we see developing a code of conduct for organizations developing advanced AI systems is one of the most urgent priorities for global society given the rapid pace of improvement in these technologies.
11. From this perspective, we commit to develop guiding principles for organizations developing, deploying, and using advanced AI systems, in particular foundation models and generative AI. These guiding principles would form the basis for an international code of conduct for organizations developing advanced AI systems. We will also continue to consider issues around intellectual property rights, such as copyright protection and issues around data protection as part of these principles.

Principles for the development of advanced AI systems could include but are not limited to:

- a. Take appropriate safety measures and consider societal risks prior to deployment including placing on the market
 - b. Endeavor to identify and mitigate vulnerabilities after deployment including placing on the market
 - c. Publicly report models' capabilities, limitations and domains of appropriate and inappropriate use, ensuring sufficient transparency
 - d. Work toward responsible information sharing among organizations developing AI and governments, civil society, and academia
 - e. Develop and disclose risk management plans and mitigation measures including privacy policies and AI governance policies
 - f. Invest in robust security controls, including cybersecurity and insider threat safeguards
 - g. Develop and deploy mechanisms such as watermarking or other techniques to enable users to identify AI-generated content
 - h. Prioritize research to mitigate societal, environmental and safety risks and prioritize investment on mitigations
 - i. Prioritize the development of advanced AI systems to address the world's greatest challenges notably but not limited to the climate crisis, global health or education
 - j. Advance the development of and alignment with internationally recognized technical standards
12. Guiding principles for deployment and use of advanced AI systems will be also be developed through the G7 Hiroshima AI Process.
13. Different jurisdictions may take a unique approach to those guiding principles, from legal frameworks, to voluntary commitments, and various other instruments, or a combination of these.

III. Project-based cooperation

14. We plan to cooperate in promoting project-based activities which could be undertaken with international organizations such as the OECD, GPAI, and the United Nations Educational, Scientific and Cultural Organization, in advancing evidence-based policy discussions. These project-based activities could include those

identified by G7 members in the OECD's Report on G7 Common Understanding on Generative AI such as advancing research and understanding of state-of-the-art technical capabilities for distinguishing AI-enabled mis/disinformation including foreign information manipulation to increase trust in AI and support the information environment.

15. We also welcome the development of the Global Challenge on Trust in the Age of Generative AI, to be launched later this year.

ANNEX 1: OECD's Report Towards a G7 Common Understanding on Generative AI

(Please follow the link provided below.)

https://www.oecd-ilibrary.org/science-and-technology/g7-hiroshima-process-on-generative-artificial-intelligence-ai_bf3c0c60-en