Hiroshima AI Process G7 Digital & Tech Ministers' Statement (December 1, 2023)

1. We, the G7 Digital and Tech Ministers, and knowledge partners, the Organization for Economic Co-operation and Development (OECD), the Global Partnership on AI (GPAI), met virtually on December 1st, 2023, to continue and develop our discussions on advanced artificial intelligence (AI) systems, with a focus on examining the opportunities and challenges throughout the AI lifecycle, as part of the Hiroshima AI Process established by G7 Leaders.

2. We endorse the attached Hiroshima AI Process Comprehensive Policy Framework (the "Comprehensive Policy Framework") as the culmination of our work in the Hiroshima AI Process under Japan's G7 Presidency. The output of the Hiroshima AI Process under Japan's G7 Presidency represents the first successful international framework comprising of guiding principles and a code of conduct to address the impact of advanced AI systems on our societies and economies. It shows that democracies can act quickly to lead the way in responsible innovation and in the governance of emerging technologies, aligning the development of advanced AI systems with our shared values. We encourage actors throughout the AI lifecycle to follow the Hiroshima Process International Guiding Principles for All AI actors (the "International Guiding Principles") (annex1) as relevant, as appropriate, and in particular we call on organizations developing advanced AI systems to commit to the application of the Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems (the "International Code of Conduct").

3. We continue to stress the importance of international discussions on AI governance and interoperability between AI governance frameworks, while we recognize that like-minded approaches and policy instruments to achieve the common vision and goal of trustworthy AI may vary across G7 members.

Ⅰ. Hiroshima AI Process Comprehensive Policy Framework

4. The Comprehensive Policy Framework presents a comprehensive set of elements including **(1) the OECD's Report towards a G7 Common Understanding on Generative AI (the "OECD's Report"), (2) International Guiding Principles, (3) International Code of Conduct and (4) project-based cooperation on AI.**

5. We consider these elements to be complementary and an important foundation for potential future AI work through the G7 and in our respective countries.

6. We intend to continuously promote the implementation of the International Guiding Principles and the International Code of Conduct and keep them reviewed and updated

as necessary, including through ongoing inclusive multi-stakeholder consultations, in order to ensure they remain fit for purpose and responsive to this rapidly evolving technology. The International Guiding Principles and the International Code of Conduct are complementary to each other and closely inter-related.

7. We acknowledge the OECD's Report for its important role as input to discussions on the International Guiding Principles and International Code of Conduct. We thank the numerous stakeholders around the globe that participated in consultations. We also welcome the Roundtable of G7 Data Protection and Privacy Authorities Statement on Generative AI.

Ⅱ.*Work Plan to advance Hiroshima AI Process*

8. We plan to continue our work on AI under Italy's G7 presidency next year in the following aspects;

i. Expand outreach to partner governments to broaden support for the International Guiding Principles and the International Code of Conduct by taking opportunities at international forums and others.

ii. Intensify efforts to encourage AI actors to follow the International Guiding Principles and call on organizations developing advanced AI systems to commit to the application of the International Code of Conduct. We will, in consultation with the OECD and other stakeholders, develop proposal to introduce monitoring tools and mechanisms to help these organizations stay accountable for the implementation of these actions.

iii. Continue collaboration on project-based cooperation with the OECD, GPAI and United Nations Educational, Scientific and Cultural Organization (UNESCO), through the Global Challenge and other potential opportunities, to explore measures and practices to counter disinformation, transparency challenges, and other challenges related to generative AI

9. We plan to launch a dedicated web site for the Hiroshima AI Process to provide updates on policy developments in relevant countries and a list of organizations that commit to the International Code of Conduct.

10. We are dedicated to promoting Hiroshima AI Process outcomes, in particular by promoting our dialogue with the multi-stakeholder community. We encourage the OECD to consider the Hiroshima AI Process outcomes in its existing AI efforts, such as the OECD workstreams on implementing trustworthy AI and risk management.

11. We seek to further advance the Hiroshima AI Process by intensifying our coordination and cooperation together across multilateral forums to promote our vision for advanced AI systems, including at the OECD, GPAI, and the United Nations.

12. We look forward to continuously working together under Italian Presidency.

(Attachment)

Hiroshima AI Process Comprehensive Policy Framework

1. The Hiroshima AI Process Comprehensive Policy Framework aims to promote safe, secure and trustworthy AI worldwide. This Comprehensive Policy Framework includes a number of elements, some of which have already been published. As part of the G7 Leaders' Statement on the Hiroshima AI Process, issued October 30, 2023, Leaders welcomed the Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI Systems[1] and the Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems[2]. Additionally, as part of the G7 Digital & Tech Ministers' Statement on the Hiroshima AI Process, issued September 7, 2023, Ministers welcomed the OECD's Report Towards a G7 Common Understanding on Generative AI[3], developed by the OECD Secretariat based on input the Hiroshima AI process working group. Today's G7 Digital & Tech Ministers' Statement introduces new elements of the Comprehensive Policy Framework: the Hiroshima Process International Guiding Principles for All AI Actors, and Project-Based Cooperation. All elements of this Comprehensive Policy Framework are complementary to each other and closely inter-related.

I. OECD's Report towards a G7 Common Understanding on Generative AI

2. The OECD's Report towards a G7 Common Understanding on Generative AI was developed by OECD secretariat based on inputs by G7members and identified a range of risks and opportunities as priorities and as the basis for the consideration on a G7 common understanding, and future action, including for collective work on generative AI.

II. Hiroshima Process International Guiding Principles for All AI Actors and for Organizations Developing Advanced AI Systems

3. These Guiding Principles are non-exhaustive lists of principles discussed and elaborated as living documents to build on the existing OECD AI Principles in response to recent developments in advanced AI systems. Hiroshima Process International Guiding Principles for All AI Actors incorporate and build upon International Guiding Principles for Organizations Developing Advanced AI systems, which, as previously stated, should be applied to all AI actors, when and as applicable, and appropriate, to cover the design, development, deployment, provision and use of advanced AI systems. They are meant to help all AI actors seize the benefits and address the risks and challenges brought by these technologies. Application of these principles should be commensurate with the risk of the AI system and the role of the AI actor. These Guiding Principles can inform our

---

[1] https://www.soumu.go.jp/main_content/000912746.pdf
[2] https://www.soumu.go.jp/main_content/000912748.pdf
[3] https://www.oecd-ilibrary.org/science-and-technology/g7-hiroshima-process-on-generative-artificial-intelligence-ai_bf3c0c60-en

policymaking efforts, following a risk-based approach, as we develop more enduring and/or detailed regulatory and governance regimes on advanced AI systems.

4. Users of open-sourced advanced AI are not exempt from the intended application of the relevant principles.

III. Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems

5. The Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems is a non-exhaustive list of actions and recommendations discussed and elaborated as a living document to build on the existing OECD AI Principles in response to the recent developments in advanced AI systems and is meant to help seize the benefits and address the risks and challenges brought by these technologies. We welcome those organizations that have already given statements to support the International Code of Conduct, and we will continue to reach out to more organizations to encourage broad endorsement.

IV. Project-Based Cooperation

6. We welcome the coordinated efforts of the OECD, GPAI, UNESCO and other partners to advance the Global Challenge on Trust in the Age of Generative AI (globalchallenge.ai), which aims to put forth and test innovative ideas to promote trust and counter the spread of disinformation. We also welcome projects on generative AI which will contribute to supporting the implementation of outcomes of Hiroshima AI process, including the projects supported by forthcoming GPAI Tokyo Center.

7. In particular, we expect that this project-based cooperation will contribute to promoting evaluation of the following technologies and concepts:
    ● reliable content authentication and provenance mechanisms to enable users to identify the originators of contents
    ● labeling mechanism to help users understand that the content is generated by AI

8. We plan to promote and deepen collaboration on our joint initiatives to maximize and share the benefits of this technology for the common good worldwide with partners beyond G7, while mitigating its risks. This non-exhaustive list of initiatives includes for now:
    ● U.S.-EU Administrative Arrangement on AI
    [The European Union and the United States of America strengthen cooperation on research in Artificial Intelligence and computing for the Public Good | Shaping Europe's digital future (europa.eu)]

    ● U.S.-UK-Canada Donor Partnership on AI for Development
    [UK unites with global partners to accelerate development using AI - GOV.UK (www.gov.uk)]

(Annex 1)

Hiroshima Process International Guiding Principles for All AI Actors

1. We emphasize the responsibilities of all AI actors in promoting, as relevant and appropriate, safe, secure and trustworthy AI. We recognize that actors across the lifecycle will have different responsibilities and different needs with regard to the safety, security, and trustworthiness of AI. We encourage all AI actors to read and understand the "Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI Systems (October 30, 2023)"[4] with due consideration to their capacity and their role within the lifecycle.

2. The following 11 principles of the "Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI Systems" should be applied to all AI actors when and as relevant and appropriate, in appropriate forms, to cover the design, development, deployment, provision and use of advanced AI systems, recognizing that some elements are only possible to apply to organizations developing advanced AI systems.

    I. Take appropriate measures throughout the development of advanced AI systems, including prior to and throughout their deployment and placement on the market, to identify, evaluate, and mitigate risks across the AI lifecycle.

    II. Identify and mitigate vulnerabilities, and, where appropriate, incidents and patterns of misuse, after deployment including placement on the market.

    III. Publicly report advanced AI systems' capabilities, limitations and domains of appropriate and inappropriate use, to support ensuring sufficient transparency, thereby contributing to increase accountability.

    IV. Work towards responsible information sharing and reporting of incidents among organizations developing advanced AI systems including with industry, governments, civil society, and academia.

    V. Develop, implement and disclose AI governance and risk management policies, grounded in a risk-based approach – including privacy policies, and mitigation measures, in particular for organizations developing advanced AI systems.

    VI. Invest in and implement robust security controls, including physical security, cybersecurity and insider threat safeguards across the AI lifecycle.

---

[4] https://www.soumu.go.jp/main_content/000912746.pdf

VII. Develop and deploy reliable content authentication and provenance mechanisms, where technically feasible, such as watermarking or other techniques to enable users to identify AI-generated content.

VIII. Prioritize research to mitigate societal, safety and security risks and prioritize investment in effective mitigation measures.

IX. Prioritize the development of advanced AI systems to address *the world's greatest challenges, notably but not limited to the climate crisis, global health and education.*

X. Advance the development of and, where appropriate, adoption of international technical standards.

XI. Implement appropriate data input measures and protections for personal data and intellectual property.

3. In addition, AI actors should follow the 12th principle.

XII. Promote and contribute to trustworthy and responsible use of advanced AI systems

AI actors should seek opportunities to improve their own and, where appropriate, others' digital literacy, training and awareness, including on issues such as how advanced AI systems may exacerbate certain risks (e.g. with regard to the spread of disinformation) and/or create new ones.

All relevant AI actors are encouraged to cooperate and share information, as appropriate, to identify and address emerging risks and vulnerabilities of advanced AI systems.