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)"[1] 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.

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

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

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.