

THE HIROSHIMA AI PROCESS: LEADING THE GLOBAL CHALLENGE TO SHAPE INCLUSIVE GOVERNANCE FOR GENERATIVE AI

Amid the growing global debate over advanced artificial intelligence (AI) systems such as generative AI, the Hiroshima AI Process was launched by the G7 under Japan's presidency in May 2023, with the aim of promoting safe, secure, and trustworthy AI. This article delves into the significance of its output, the Hiroshima AI Process Comprehensive Policy Framework—the world's first international effort toward that end.

The rapid development and spread of advanced AI systems such as generative AI have become an important global issue in recent years. The Hiroshima AI Process was launched in May 2023 for international discussion, following the Leaders' direction at the G7 Hiroshima Summit under Japan's presidency. Recognizing the need to build up and promote inclusive global governance on AI in order to maximize its innovative opportunities while mitigating the risks and challenges from advanced AI systems, work began on establishing international rules to serve as the foundation of such governance. And in December of last year, agreement was reached on the world's first international framework, known as the Hiroshima AI Process Comprehensive Policy

Framework.

The framework presents a set of elements, including both the "Hiroshima Process International Guiding Principles for All AI Actors" and the "Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems." The International Guiding Principles are the set of principles that should be applied to all actors across the AI lifecycle. In addition to principles mainly geared toward AI developers, such as publicly reporting advanced AI systems' capabilities and domains of inappropriate use, and protecting intellectual property, they also

include those that call on users to improve digital literacy dealing with such risks as disinformation.

The International Code of Conduct, on the other hand, lists in more detail the actions that AI developers must abide by. A certain level of abstraction is employed in the wording of both the International Guiding Principles and the International Code of Conduct so that stakeholders from various perspectives can agree on them. Yet at the same time, the International Code of Conduct also contains some references that are more detailed, ranging from examples of risks requiring the attention of developers to the

Prime Minister Kishida participated online in the inaugural AI Safety Summit hosted by the U.K. in November 2023. He stated that he would like to work closely with such efforts to drive the creation of rules that would benefit the international community as a whole.



At the G7 Hiroshima Summit in May 2023, leaders confirmed the need for generative AI governance and agreed to establish the Hiroshima AI Process as a forum for further discussion.

development and deployment of technology enabling users to identify AI-generated content. "While prioritizing international consensus building, it is also very important to be as specific as possible. In that sense, it was a great achievement to have been able to include a certain level of specificity here," said Professor MATSUO Yutaka of the Graduate School of Engineering at the University of Tokyo, who also chairs the Government of Japan's AI Strategy Council.

Matsuo also pointed out that it was of paramount significance that this international framework, which could serve as a common foundation, was realized just seven months after the launch of the Hiroshima AI Process. "Speed is of the essence here because technology is advancing at such a rapid pace."

From now on, the G7 plans to expand outreach to partner governments to broaden support

for the International Guiding Principles and the International Code of Conduct, and to further advance the Hiroshima AI Process by intensifying coordination and cooperation across multilateral forums including at the Organisation for Economic Co-operation and Development (OECD) and the United Nations (UN). Japan is slated to inaugurate the AI Safety Institute, whose roles will include conducting research on AI safety evaluation methods. The country additionally plans to establish the Tokyo Center of the Global Partnership on AI (GPAI)—an international public-private partnership—to conduct research and analysis on generative AI.

According to Matsuo, "Overall, Japan tends to take a positive stance on the development and use of AI. Simultaneously, it is actively engaged in ensuring safety, thus placing itself in the middle ground between those in

the world who promote the use of AI and those who lean toward regulation. It will be important to demonstrate Japan's leadership in the future while continuing to balance international discussions on AI."

With the goal of promoting safe, secure, and trustworthy AI, Japan will promote the outcomes of the Hiroshima AI Process while continuing to strengthen global dialogue and cooperation. ●



As AI capabilities improve further, discussions on an international framework will become increasingly important, said Professor MATSUO Yutaka of the University of Tokyo's Graduate School of Engineering, who chairs Japan's AI Strategy Council.