Report on International Trends in AI Governance: How do we go from AI principles to practice?

On March 7, 2023, the Institute for Future Initiatives, University of Tokyo, held a seminar titled "International Trends in AI Governance: How do we go from AI principles to practice?". Held as a hybrid of SMBC Academia Hall on the 4th floor of the University of Tokyo's International Academic Research Building and online, the seminar attracted a little over 30 participants.

Over the past decade, there has been much discussion on how artificial intelligence (AI) should be developed in relation to society. In particular, the debate has now entered a new phase with the emergence of technologies, such as generative AI and large-scale language models (LLM). The significant impact of the newest AI technologies on society has been highlighted, and discussions have focused on the need for AI governance.

The main purpose of this event was to review the global trend of AI governance and delve into the role that Japan should play in this trend, welcoming overseas guests who are leading discussions on AI governance.

Panelist Speakers

Gregory C. Allen: Director of AI governance project, Center for Strategic and International Studies
Rebecca Finlay: CEO at Partnership on AI
David Leslie: Director of Ethics and Responsible Innovation Research at The Alan Turing Institute and Professor of Ethics, Technology and Society, Queen Mary University of London
Karine Perset: Head of AI Unit and OECD.AI, OECD Digital Economy Policy Division – OECD
Hideaki Shiroyama: Professor, Institute for Future Initiatives, The University of Tokyo
Facilitator: Arisa Ema: Associate Professor, Institute for Future Initiatives, The University of Tokyo

Trends in AI Governance Debate

First, Prof. Hideaki Shiroyama of the Institute for Future Initiatives, University of Tokyo, introduced the purpose of this event and explained why AI governance is being discussed. Prof. Shiroyama stated that he would like to understand the current global trend of AI governance discussions and consider the role Japan can play, particularly at the time of the G7 Summit to be held in Japan.

Presentations from panelists began with Mr. Gregory Allen. Mr. Allen is the Director of the Center for Strategic and International Studies (CSCI) AI Council.

Mr. Allen began his presentation by asking how these principles could be put into practice in a robust manner, stating that flexible approaches should be considered to move from AI principles to practical applications.

First, "adhering to AI principles" and "demonstrating those principles as practice" are two very different things, and the place for that practice is very different in a larger organization, with thousands or hundreds of thousands of employees running billions of dollars in various operations, than in a smaller organization.

Allen stated that it is important to remember that there are two aspects of AI: its usefulness and risks. With a declining population, continued economic growth requires increased productivity, and AI is a technology that holds great promise in this regard. However, it is also true that various risks exist in the use of AI. For example, an algorithmic risk case was reported in the Netherlands. An inappropriate algorithm that encourages racial discrimination was used in the child allowance application system, affecting thousands of people.

When using AI systems, one must be very careful regarding the values that should be included in the system. Shared values are essential for democracies. The principles must then be transformed into organizational procedures, technical procedures, and sometimes even government policies to implement AI principles.

Mr. Allen cites three key areas: regulation and governance, practices, and technology strategy. However, this does not necessarily imply that AI-specific regulations must be created immediately. For example, AI-specific regulations may not exist for the inclusion of machine learning systems in commercial aircraft operations. However, the existing safety regulations are in effect in some areas. Because of the different natures of traditional software and AI systems, it is important to consider new regulations and legislations; however, it is important to remember that the regulations already in place for software are useful.

Various organizations are working on collecting and disseminating practices. The CSIS AI Council will be issuing a paper on this topic. He also noted that there are various phases of technology strategies, and that all issues need to be discussed with both the government and the private sector.

At the end of his presentation, Mr. Allen mentioned the role he expects Japan to play in contributing to "non-binding guidelines," "binding regulations," and "the development of standards themselves." Among these, the development of standards is crucial. To this end, it is important that all institutions and organizations, including businesses, governments, and academia, participate in and support the standards development process.

Ms. Rebecca Finlay, CEO of the Partnership on AI, then gave a presentation. The Partnership on AI is a global AI community founded in 2016 by the heads of AI Research at the six large technology companies and now contains over 100 non-profit organizations, academic, industry, and media partners.

There are several approaches to building a healthy AI system, one of which Ms. Finlay noted, is PAI's multi-stakeholder approach. This is based on the understanding that norms created within a group have a more significant impact on best (or better) practices. When setting norms as a community, the role of smaller countries and middle powers such as Canada and Japan are also influential.

To have impact as a multistakeholder community, the goals are fourfold: building a community, spreading literacy, promoting policy innovation, and facilitating change. The core is developing collective standards, and as a community, responding to emerging AI development and innovation trends. It is also important that this is done through international gatherings of individuals and organizations. Multi-stakeholder gatherings will examine both the utility and risks of emerging technologies. Developing worldwide norms with best practices, putting them into practice, developing them, and bringing resources together must also be considered by expert groups.

The safety of technologies, such as LLMs, should also be carefully examined. Developers, researchers, and the general public should be included in a responsible manner with regards to standardization. She said that she looked forward to Japan playing a leadership role in this area on the global stage.

Ms. Karine Perset, who heads the OECD AI Unit, asserts that the time is not yet right to create an environment for the successful spread of AI as a technology in society. The OECD developed AI principles in 2019, and is attempting to put these principles into practice. An AI Policy Observatory was launched. It is an online platform that gathers a variety of resources, including data, trend movements, policies, and evidence, and covers cases from more than 60 countries.

The OECD AI Unit's approach to risk management comprises four streams: First, the risk context is defined. Second, AI risks are assessed. Third, AI risks are addressed. Fourth, AI risks are governed. This governance component is the most important and includes incident monitoring tools. In other words, a risk value chain for AI systems should be assembled and risk assessment should be conducted in an interoperable manner.

The utility of interoperability is to lower the burden on those who implement this framework and increase its effectiveness, efficiency, and enforceability. This framework should be agreed upon by various countries in the upstream part of policymaking. Interoperability will allow countries in democracies and market economies to communicate with each other. This was the role of the OECD in building a large consensus that included non-member countries as stakeholders.

In terms of the governance of AI risks, the report points to the importance of corporate governance beyond the national and regional levels. The focus is not simply on drafting legislation, but also on the usefulness of developing specific codes of conduct that make the management of AI risks operationally viable. Ms. Perset used the OECD Due Diligence Guidance for Responsible Corporate Behavior released by the OECD as an example here, and said that by developing these existing guidelines in an AI-specific manner, AI risks could be addressed in the short term without having to create a governance system from scratch. She also noted that Japan's role in AI risk management was to provide a framework for addressing AI risks on an individual basis.

She also said that Japan should participate in international discussions even on an individual basis.

The last panelist presentation was by Professor David Leslie of the Alan Turing Institute in the UK. Prof. Leslie is the Director of Ethics and Responsible Innovation Research.

The Alan Turing Institute has been working to develop a framework from principles to practice on a variety of issues, including AI, working with the Ministry of Justice and the government's Digital Services AI Office, which released its AI Ethics guidelines in 2019. In it, they present a process-based governance framework to connect with actual governance practices.

Prof. Leslie pointed out that the movement in the discussion of AI ethics and governance is a growing emphasis on the connection between moral concepts and social practices. What is needed, he says, is a change in organizational culture, technological approaches, and individual attitudes. For responsible AI innovation, it is important to re-conceptualize AI systems throughout the lifecycle of a social project. Every step of the engineering and design process should be ethically engaged, considering the real-world impact of technology on these processes and the roles of purpose, value, and benefit. This is a practice-driven process-based approach.

Prof. Leslie also states that the concept of Responsible Research and Innovation (RRI) and the habits of national and international trends should be incorporated into all research innovations. This RRI perspective makes researchers and innovators aware that all processes of scientific discovery and problem solving have sociotechnical aspects and ethical stakes. The Alan Turing Institute created the Care & ACT Framework as an RRI tool to support this.

Panel Discussion: From Principles to Practice, From Practice to Principles

In the second part, Professor Shiroyama joined Mr. Allen, Dr. Finlay, Ms. Perset, and Professor Leslie in a panel discussion. Dr. Ema moderated these sessions.

When it comes to AI governance, discussions typically revolve around "From principles to practice" in terms of how to prepare for a society coexisting with AI. In this panel discussion, however, the discussion was also developed from the opposite perspective: “Is there anything to be learned from principles to practice?”

Currently, new AI technologies such as generative AI and LLM are spreading rapidly. One characteristic of these technologies is that they have suddenly entered the phase of practice without interrupting the process of reviewing whether they are in line with the principles. If this is the case, principles can be updated through such practices. Dr. Ema raised the issues of how such technology would be civilized, what would change when many people use it as a tool, and what principles should be considered.

As the discussion examined the impacts and risks of ChatGPT or generative AI, Professor Leslie expressed the view that these new AI technologies must be approached in advance in terms of governance. For example, OpenAI holds the users of their technology (services) accountable. They place responsibility on users and not on those who develop and provide services. This is a major problem that must be addressed at a policy level. However, Mr. Allen cited the difficulty of assessing the true capability of such systems as a challenge, and said that the first step is to learn how to assess them properly.

Regarding Japan's leadership in the global governance of AI, many panelists expressed hope for Japan's G7 presidency. Now that numerous countries and regions are preparing AI regulations, G7 countries must make commitments to ensure the interoperability of these regulations. There is a need to explore ways to mitigate these risks and find a landing place, not just a broad range of benefits, for advanced AI systems. It will be very important for the G7 to ensure the interoperability of these regulatory frameworks and standards, as this could be a barrier to collaboration, depending on how it is done.

An audience member said that AI is a tool and gave hints; an interactive relationship between AI and humans who use AI should be established, and humans are ultimately responsible for making decisions. The panelists agreed that AI is a useful tool. Therefore, it is necessary to obtain a realistic understanding of the limitations of the system. Ms. Finlay pointed out that workers need to be educated about how to use AI so that they can document, determine and decide clearly what the system will and will not do.

Finally, Prof. Shiroyama emphasized the importance of continuing the discussion, considering the fact that society itself is changing rapidly when considering the actual transition from principle to practice and that some aspects of operation will have to be considered within the framework of democracy. The event ended with the statement that we must recognize that society itself is changing rapidly when considering actual principles and practices.

International Trends in AI Governance:
How do we go from AI principles to practice?
AIガバナンスを巡る国際的な潮流：AI原則をどのように実践に落とし込むのか
Hideaki Shiroyama
Reserved

(From left to right) Dr. Ema, Mr. Allen, Dr. Finlay, Ms. Perset, Prof. Leslie, Prof. Shiroyama