IFI Policy Recommendation

IFI Policy Recommendation No.21 March 2023

Pathways to AI Governance Coordination

: Policy Recommendations for the G7 Summit

AI Governance Project, Technology Governance Policy Research Unit
Institute for Future Initiatives, The University of Tokyo

東京大学未来ビジョン研究センター
Institute for Future Initiatives, The University of Tokyo

Pathways to AI Governance Coordination: Policy Recommendations for the G7 Summit

AI Governance Project, Technology Governance Policy Research Unit
Institute for Future Initiatives, The University of Tokyo

Executive summary

Artificial intelligence (AI) is having a profound impact on the way we live and work. While more innovative breakthroughs to benefit people and society are highly anticipated, we must also acknowledge the risks.

In 2016, as the chairing country of the G7 Ise-Shima Summit, Japan proposed a draft guideline[1] that would serve as a list of principles for AI R&D at the Kagawa-Takamatsu Ministerial Conference on Information and Communications. The guideline catalyzed international attention on the need for AI governance. Discussions continued in subsequent G7 summits, and international organizations, countries and regions, companies, industry associations and civil society groups developed AI principles, such as the OECD AI principles.[2]

In addition to numerous risk assessment tools, several major countries and regions are now proposing regulations for AI in order to advance AI principles into practice. As the number of countries, regions, and application areas utilizing AI expands, international coordination is required to ensure that these tools and regulatory frameworks are developed in a coherent manner that is also consistent with the context of each country, region, and application domain.

Japan is the G7 summit host in 2023. Just as the draft guidelines proposed by Japan in 2016 led to important international discussions, this summit will play an important role in charting the future of AI governance and increasing international coordination and action. Recent developments such as generative AI have sparked legal and ethical debates and highlighted the need for ongoing attention to AI. As the use of AI grows, the public will be much more interested in the governance mechanisms in place. The central question for the G7 at this time is, how do we promote AI innovation to advance national and planetary sustainability and resilience[3] while reducing risks to people and society.

This paper proposes two foundational AI governance policy recommendations that the G7 should respect, and four actions for international coordination by the G7.

Foundational policy recommendations

1. **Compliance with shared fundamental values:** The development and use of AI

[1] Draft AI Development Guidelines for International Discussion, AI Network Society Promotion Council, Ministry of Internal Affairs and Communications, https://www.soumu.go.jp/main_content/000507517.pdf

[2] OECD AI principles overview, OECD.AI Policy Observatory, https://oecd.ai/en/ai-principles

[3] The AI Strategy 2022 by Japan's Cabinet Office mentions the need for Resilient and Responsible AI to achieve resilience to national and planetary crisis, to improve resilience against attacks such as strengthening cybersecurity, and to increase reliability of AI (https://www8.cao.go.jp/cstp/ai/aistratagy2022en.pdf). The promotion of a resilient society through AI deployment was also raised at the GPAI Summit held in Japan in November 2022 (https://www.meti.go.jp/press/2022/11/20221124002/20221124002-1. pdf)

technologies should comply with the fundamental human rights and democratic values underlying existing AI principles

2. **Respect for the context of the application:** In putting AI principles into practice, the institutional, social, and cultural context of the application, including the field and region in which the AI technology will be used, should be respected while ensuring commonality as much as possible

Actions for international coordination

i. **Establish and adopt international standards:** International standards for risk assessment and management frameworks across the AI lifecycle should be established and adopted where possible, in addition to sharing best practices and case studies to promote responsible and safe AI design, development, use and operation across countries, regions and organizations

ii. **Promote human-AI collaboration research, policy and practice**: Research, policy and practice on the beneficial characteristics of collaborative work between humans and machines should keep pace with rapid AI technology advances and the expansion of human-AI collaborative work

iii. **Increase literacy on AI use and governance:** Efforts should be strengthened to educate and train people with the skills needed to develop and use AI responsibly, to discuss and practice AI governance, and to empower citizens through literacy

iv. **Support discussion forum:** In response to the rapid development of AI technologies, multi-stakeholder forums for knowledge-sharing and coordinated action-planning on AI governance in an agile manner should be supported

Process for formulating this policy recommendation

The content of this proposal was developed by the AI Governance Project, Technology Governance Research Unit, Institute for Future Initiatives, The University of Tokyo, based on the findings of experts from industry, academia, and government in Japan and overseas experts. The discussion at the roundtable held 6 March 2023 was also incorporated into this report. See the Appendix for a list of people who provided feedback on this report.

Foundational policy recommendations

1. **Compliance with shared fundamental values:** The development and use of AI technologies should comply with the fundamental human rights and democratic values underlying existing AI principles

The AI principles established by international organizations, national and regional governments, civil society organizations, and corporations include the values that each organization holds. The main values listed in the survey comparing each item are human dignity, human-centric, equity, transparency, accountability, and privacy.[4] Common to these principles is the importance of upholding fundamental human rights and democratic values. In light of the "Strong Democracy Statement" confirmed by the G7 in 2022[5], the G7 should pursue the development and use of AI technologies for the benefit of humanity in line with democratic principles.

2. **Respect for the context of the application:** In putting AI principles into practice, the institutional, social, and cultural contexts of the application, including the field and region in which the AI technology will be used, should be respected while ensuring commonality as much as possible

Systems and services using AI technologies are deployed virtually and as systems and services integrated into the real world, including in medicine and welfare, agriculture, forestry, fishing, manufacturing, construction, service, transportation, disaster prevention, education, and the arts.[6] Therefore, AI systems, and services should be deployed in harmony with existing technological systems, regulations, customs, and cultures of the applied fields, and in inclusive and equitable ways. In addition, when AI technologies are designed, developed, used, and operated across countries, regions, and organizations, they should respect the systems, customs, cultures, local laws and regulations of each country, region, and field of application, in compliance with fundamental human rights, and democratic values.

Actions for international coordination

i. **Establish and adopt international standards:** International standards for risk assessment and management frameworks across the AI lifecycle should be established and adopted where possible, in addition to sharing best practices and case studies to promote responsible

---

[4] Anna Jobin, Marcello Ienca & Effy Vayena: The global landscape of AI ethics guidelines, Nature, Machine Intelligence, 1, 389-99, 2019.

[5] 2022 Resilient Democracies Statement, G7, 2022, https://www.consilium.europa.eu/media/57543/2022-06-27-g7-resilient-democracies-statement-data.pdf

[6] The Japanese government has proposed the concept of Society 5.0 as a system where cyberspace and physical space merge (https://www8.cao.go.jp/cstp/english/society5_0/index.html)

and safe AI design, development, use and operation across countries, regions and organizations

AI technology is characterized by the difficulty of prior performance assurance and ex-post verification. To make the use of AI trustworthy and safe, there is a need for greater transparency in the companies and organizations that use it.[7] Therefore, legally binding regulations[8] and framework conventions[9] are being discussed, contractual guidelines[10] are being developed, and non-binding guidelines and tools for risk management frameworks,[11] risk assessment,[12] verification,[13] auditing,[14] quality control,[15] business governance guide[16] are being promoted for AI technologies.

In addition, many AI systems and services, not only AI technologies, are being developed and utilized across countries, regions, and organizations. Therefore, interoperability of not only AI technologies, AI systems and services, but also organizational management frameworks and methodologies is an important issue. To realize this, first, it is important to identify the ecosystem of organizations involved in AI governance practices[17] and establish a mechanism

---

[7] While there is an importance of increasing transparency, there are also concerns about risk management. For example, some companies have voiced concerns that there is a risk of more serious security incidents if a source code disclosure requirement is imposed to ensure transparency. There are also concerns about increasing risks from an (economic) security perspective.

[8] In 2021 the European Commission released an Artificial Intelligence Act, in 2022 the U.S. Algorithmic Accountability Act of 2022 was introduced in both houses of Congress, and in the same year Canada proposed the Artificial Intelligence and Data Act (AIDA) that would mandate for risk management and information disclosure regarding high-impact AI systems.

[9] The Council of Europe is discussing the development of an AI Treaty.

[10] Japan's Ministry of Economy, Trade and Industry has published guidelines for AI contracts (https://www.meti.go.jp/press/2019/12/20191209001/20191209001.html). The Ministry of Economy, Trade and Industry and the Japan Patent Office have also published a "Model Agreement for Promoting Open Innovation between R&D Startups and Business Companies"

[11] The Singapore government's AI Governance Framework (https://www.pdpc.gov.sg/ help-and-resources/2020/01/model-ai-governance-framework) and the OECD's Framework for the Classification of AI Systems (https://oecd.ai/en/ classification) and others are publicly available.

[12] The Government of Canada has released the Algorithmic Impact Assessment tool (https://www.canada.ca/en/government/system/digital-government/digital-government-innovations /responsible-use-ai/algorithmic-impact-assessment.html), and the Council of Europe is also exploring a common methodology for risk impact assessment of AI called Human Rights, Democracy and the Rule of Law Impact Assessment (HUDERIA). UNESCO is also discussing the need for an Ethical Impact Assessment (EIA) to promote AI recommendations.

[13] The Singapore government released an AI governance testing toolkit called "AI Verify" in 2022, which will allow organizations to demonstrate in an objective and verifiable way that they are using AI appropriately.

[14] The AI system audit of the Automated Employment Decision Tool (AEDT) is being discussed in New York City, USA (https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=4344524&GUID=B051915D- A9AC-451E-81F8-6596032FA3F9&Options=ID%7CText%7C&Search=).

[15] Japan's National Institute of Advanced Industrial Science and Technology (AIST) has released "Machine Learning Quality Management Guideline" to manage the quality of AI-based products and services (https://www.digiarc.aist.go.jp/en/publication/aiqm/).

[16] Japan's Ministry of Economy, Trade and Industry has published "Governance Guidelines for Implementation of AI Principles" (https://www.meti.go.jp/english/press/2022/0128_003.html)

[17] The report of the Japan Deep Learning Association's "AI Governance and its Evaluation" study group proposes the creation of an AI governance ecosystem that includes not only AI service providers, but also governance for service providers such as auditing and insurance, and third-party organizations such as internal reporting and third-party committees in case of accidents, under the concept of the AI Governance Ecosystem. The report proposes the creation of a system (https://www.jdla.org/en/en-about/en-studygroup/en-sg01/)

to share best practices[18] and accidents and incidents[19]. To this end, it is necessary to design incentives that allow organizations, which have shared information, to be evaluated and receive certifications, and establish an accident investigation system. In addition, scenarios[20] regarding past, present and future risks and harms to humanity and the environment should be created, and findings from social experiments such as the regulatory sandboxes system[21] should be shared.

When there is consensus among stakeholders to respect the context of the application but transcend it, the establishment and adoption of international standards should proceed. AI technologies, services, systems, and organizational management frameworks are being promoted by standardization bodies such as ISO,[22] IEEE[23], NIST[24] and CEN/CENELEC.[25] As international standards facilitate interoperability, the number of companies and organizations participating in the AI lifecycle (design, development, operation, use, decommissioning, and system retirement) will diversify. Establishing international standards through a transparent and fair formulation process is not an easy task. However, creating interoperable standards that are easy to adopt not only for specific companies but also for vulnerable people, countries and regions, start-ups, SMEs, and organizations through a transparent and fair development process will help create a level playing field and prevent monopoly, oligopoly, or AI governance failures by certain companies.[26]

ii. **Promote human-AI collaboration research, policy and practice**: Research, policy and practice on the beneficial characteristics of collaborative work between humans and machines should keep pace with rapid AI technology advances and the expansion of human-AI collaborative work

When AI is used to support human action and decision making, the degree of human and AI involvement and the division of their roles can be categorized.[27] For example, while utilizing

---

[18] The OECD.AI website (https://oecd.ai/en/catalogue/tools) has a collection of practical examples

[19] The AI Incidents Database (https://partnershiponai.org/workstream/ai-incidents-database/) has a collection of AI incidents.

[20] The risk chain model (RCModel) developed by the University of Tokyo identifies risk scenarios for each AI application and organizes countermeasures (https://ifi.u-tokyo.ac.jp/projects/ai-service-and-risk-coordination/overview/).

[21] The New European Innovation Agenda includes a regulatory sandbox in Europe (https://research-and-innovation.ec.europa.eu/strategy/support-policy- making/shaping-eu-research-and-innovation-policy/new-european-innovation-agenda_en?ct=t(EMAIL_CAMPAIGN_1_12_2023_15_17)), and Spain has implemented a regulatory sandbox on AI (https://digital-strategy.ec.europa.eu/en/events/launch-event-spanish-regulatory-sandbox-artificial-intelligence).

[22] ISO/IEC JTC 1/SC42 is discussing international standardization of terms and basic concepts related to AI, and ISO/IEC 38507 is discussing international standardization of AI governance

[23] The IEEE's P7000 series and others discusses standards for practical issues in AI

[24] NIST discusses a unified risk-based framework for AI that is interoperable with ISO/IEC management standards and concepts, as well as OECD AI recommendations

[25] CEN/CENELEC discusses standards for AI in Europe

[26] To promote fair trade in data and AI services and systems, it is necessary to establish a mechanism to monitor whether fair trade is taking place. From the perspective of fair competition, Japan's Fair-Trade Commission has analyzed the risk of cartels and unfair trading conducted by AI algorithms from the perspective of fair competition and concluded that many of these issues can be addressed under existing antitrust law.

[27] For example, the University of Tokyo's proposed classification of medical AI types suggests that it is important to select

AI for action and decision making previously performed by humans will enable more efficient and innovative actions and decisions, complete dependence on AI may make it difficult to make ethical decisions. In addition, if the operation is one in which humans can be involved in AI decision making, it is necessary to set an appropriate speed for the AI's output that humans can oversee.[28] By clarifying the nature of collaborative work between humans and AI in this way, an appropriate design can be made according to identified operational needs.

While considering that, the nature of human-AI collaboration changes with the rapid technological progress of AI, it is important that the required criteria differ according to the ethical, legal, social, and economic contexts. Therefore, international coordination should support research on classifications that can be referenced according to technological progress and vertical challenges in different application areas and regions.[29]

iii. **Increase literacy on AI use and governance:** Efforts should be strengthened to educate and train people with the skills needed to develop and us AI responsibly, to discuss and practice AI governance, and to empower citizens through literacy

As the use of AI technology expands, concerns have arisen about the misuse and abuse of AI services and systems, and the risks and harms it may inflict on humanity and the environment. It is important not only to develop responsible AI, but also to provide the public at large with access to necessary information, materials on responsible AI use, and opportunities to reskill in order to help users make positive use of AI. As the technology evolves, it is also important to develop literate personnel who can promote AI governance and provide resources for those in responsible positions in the organization so they can properly understand the potential and risks of AI.[30]

Identifying and providing learning opportunities to acquire the necessary competencies needed by personnel involved in the AI lifecycle is also important. Developing and sharing these human resources will promote discussion on international standards and research.

iv. **Support discussion forum:** In response to the rapid development of AI technologies, multi-stakeholder forums for knowledge-sharing and coordinated action-planning on AI governance in an agile manner should be supported

In promoting international standards, research, and human resource development, a path

---

the appropriate type of medical AI rather than a level (https://ifi.u-tokyo.ac.jp/en/news/4638/) .

[28] "WHITE PAPER on Artificial Intelligence - A European approach to excellence and trust" discusses various human oversight (https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en).

[29] For example, the "Future of Work" working group of the Global Partnership on AI is also discussing how people and AI should cooperate, and conducting research on the impact of AI on employment.

[30] The World Economic Forum offers a toolkit to help executives understand the risks and opportunities of AI (https://www.weforum.org/reports/empowering-ai-leadership-ai-c-suite-toolkit, https://express.adobe.com/page/RsXNkZANwMLEf/)

should be established for constructive discussions to bring together vertical issues in each field and link them to horizontal issues across fields and internationally. As the development of AI technology and the diffusion of its use are advancing quickly, such discussion should be conducted in an agile and responsible manner.

Multi-stakeholder discussions are currently underway in various institutions and organizations, including OECD.AI and the Global Partnership on Artificial Intelligence to bridge the gap between theory and practice on AI. There are also nonprofit organizations, such as the Partnership on AI, established to convene industry, media, policy, civil society and research organizations across borders to develop voluntary norms and collective action on responsible AI. In December 2022, the EU Trade and Technology Council (TTC) reached an agreement to promote common terminology, taxonomy, and risk management tools for AI in Europe and the United States.[31] The G7 should support the experts, institutions, and citizens involved in these discussions, and establish a system to consolidate and continuously evaluate the knowledge gained there, prepare for emerging risks and harms to humanity and the environment, and utilize it appropriately in international discussions and standards. For the future of humanity, it is necessary for the G7 to do so within a framework of collaboration that transcends national boundaries.

Addendum: The impact of generative AI and the need for collaboration on AI governance

In 2022-2023, a number of large AI content generation (including text, audio, images and video) models have emerged, presenting a variety of possibilities and challenges. Many of them are related to the items proposed in this policy recommendations document and it is likely that new research, policy, practice and international coordination will be required.

For example, Generative AI has raised legal issues including copyright law, regarding large amounts of training data and generated content used in the development of genitive AI. In addition, problems of inappropriate output results, wrong answers, misuse, privacy, and leakage of secrets due to flaws in learning methods and algorithms, data bias in the input and training data, and incomplete training have also captured public attention. Therefore, it is necessary to have more appropriate literacy on the use of generative AI models. Furthermore, experts pointed towards a cycle in which a large amount of data generated by a generative AI model is then relearned by the generative AI, further entrenching biases and errors. Further research and discussion are needed on the evolution of relearned AI and its impact.

At this time, models are being released by both large corporate and start-up entities, and it remains to be seen whether responsible AI principles identified in this policy recommendations will be adhered to. It is important that we determine what new capabilities are emerging that

[31] FACT SHEET: U.S.-EU Trade and Technology Council Advances Concrete Action on Transatlantic Cooperation, https://www.whitehouse.gov/briefing-room/statements-releases/2022/12/05/fact-sheet-u-s-eu-trade-and-technology-council-advances-concrete-action-on-transatlantic-cooperation/

may challenge existing risk assessment and mitigation strategies. However, there is a concern that regulation that is too restrictive in the emerging stage of the technology will become a barrier to entry for smaller companies and lead to market oligopoly, and may also hinder the free use of information that is in the public domain. Therefore, there is a need to promptly and continually discuss the possibilities, challenges, and risks posed by generative AI models through already established multilateral and multistakeholder forum for international and interdisciplinary sharing of knowledge to develop risk mitigating measures.

Publication of this policy recommendation and support

Institute for Future Initiatives (IFI), The University of Tokyo was established in 2019 to create a sustainable future society, make policy and social recommendations on future society issues, and to pursue research in collaboration with society to these ends. It also serves as an international network hub integrating university knowledge related to future societies and as a platform for collaborative creation with industry, government, academia, and citizens to provide research-based alternatives for creating our future society and to help develop the human resources necessary to achieve it.

This policy recommendation was released as part of the research and activities of the AI Governance Project, a project of the Technology Governance Policy Research Unit at the Center, which studies perspectives and methodologies to properly control the process of science and technology research and its application.

This research is part of the flagship projects of the Institute for Future Initiatives "Designing visions of the future in an AI society", as well as the Toyota Foundation D18-ST-0008 "Formation of a platform for Ethics and Governance of Artificial Intelligence" and Grant-in-Aid for Scientific Research (A) 18H03620 "International Governance of New Information Technology and Biotechnology - Information Sharing and Role of Private Actors". The results of joint research with companies and government officials in the AI Governance Project are also included as part of the recommendations.

Appendix: List of people who provided feedback on this policy recommendation

Takashi Akoshima, Internal Auditor, Japan Digital Design,Inc.
Gregory C. Allen, Executive Director of the CSIS AI Council
Junichi Arahori, Head, AI Ethics and Governance Office, Fujitsu Limited
Kuan-Wei Chen, Graduate Schools for Law and Politics, The University of Tokyo
Takashi Egawa, National Institute of Advanced Industrial Science and Technology (AIST)
Arisa Ema, Associate Professor, Institute for Future Initiatives, the University of Tokyo; Visiting researcher, RIKEN - Rapporteur
Rebecca Finlay CEO, Partnership on AI
Shinnosuke Fukuoka, Attorney-at-Law (Japan & New York), Nishimura & Asahi.
Naohiro Furukawa, Attorney-at-Law, ABEJA,Inc.

Danit Gal, Associate Fellow at the Leverhulme Centre for the Future of Intelligence at the University of Cambridge

Hiroki Habuka, Research Professor, Graduate School of Law, Kyoto University / CEO, Smart Governance Inc.

Yuko Hararyama, Professor Emeritus, Tohoku University

Tagui Ichikawa, Professor, Institute of Innovation Research, Hitotsubashi University

Chihyung Jeon, Korea Advanced Institute of Science and Technology (KAIST)

Toshiya Jitsuzumi, D.Sc., Professor, Chuo University

Kit Kitamura, The Head of AI Legal Group, CDLE (Community of Deep Learning Evangelists)

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

Fumiko Kudo, Project Strategy Lead, World Economic Forum Centre for the Fourth Industrial Revolution Japan

Satoshi Kurihara, Professor, Keio University, Faculty of Science and Technology / Director, Center of Advanced Research for Human-AI Symbiosis Society

Takashi Matsumoto, Visiting researcher, The University of Tokyo, Institute for Future Initiatives; Deloitte AI Institute

Yutaka Matsuo, Professor, Graduate School of Engineering, The University of Tokyo

Youichiro Miyake, Specially Appointed Professor, Graduate School of Artificial Intelligence and Science, Rikkyo University

Hiroshi Nakagawa, Team leader, RIKEN Center for Advanced Intelligence Project

Takehiro Ohya, Professor, Faculty of Law, Keio University

Takako Ouchi, Research assistant, Institute for Future Initiatives, the University of Tokyo

Karine Perset, Head of AI Unit and OECD.AI, OECD Digital Economy Policy Division, OECD

George Shishido, Professor, Graduate Schools for Law and Politics, The University of Tokyo

Hideaki Shiroyama, Professor, Institute for Future Initiatives, The University of Tokyo

Roy Sugimura, Supervisory Innovation Coordinator, National Institute of Advanced Industrial Science and Technology (AIST)

Masashi Sugiyama, Director, RIKEN Center for Advanced Intelligence Project / Professor, The University of Tokyo

Akira Tajima, Yahoo Japan Corporation

Junichi Tsujii, AIST fellow, Artificial Intelligence Research Center; Professor, University of Manchester

Kyoko Yoshinaga, Non-Resident Senior Fellow, Institute for Technology Law & Policy, Georgetown University Law Center

Tatsuya Yoshizawa, Professor, Kanagawa University

Toshiya Watanabe, Professor, Institute for Future Initiatives, The University of Tokyo

*Arisa Ema edited this document, and Takako Ouchi provided editorial support.