IFI Policy Recommendation

IFI Policy Recommendation No.23 September 2023

Towards Responsible AI Deployment

Policy Recommendations for the Hiroshima AI Process

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

Towards Responsible AI Deployment:
Policy Recommendations for the Hiroshima AI Process

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

Executive Summary

Alongside the proliferation of artificial intelligence (AI) technologies, particularly machine learning, generative AI—capable of producing diverse content—has gained widespread use, raised expectations for improvements and innovations in various aspects of life and work. However, it has also underscored the need for appropriate design, development, deployment, and utilization of AI. Given that AI evolves and operates within society, there is a pressing need to establish comprehensive AI governance that encompasses AI developers, providers, users, public institutions, and the society at large. Emphasizing transparency and equity, the call for a shared framework that fosters innovation while mitigating risk is important.

Foundational Policy Recommendations

1. Establish Forums for Responsible AI Deployment Discussions

To uphold shared values, such as fundamental human rights and democratic principles, it is imperative to facilitate the creation of forums for agile and ongoing multistakeholder deliberation.

2. Promote "Interoperability" Between Frameworks and Mutual Respect for Discipline on AI

The concept of "interoperability," as discussed at the G7 Hiroshima Summit, can be contemplated from two aspects: “standards” and “interoperability among frameworks.” Interoperability among frameworks involves an approach to achieving common objectives while respecting disciplines pertaining to AI, which may vary among different countries, regions, organizations, and domains. Disciplines encompass various classifications and it is not always imperative to establish entirely new regulations for every emerging technology within the AI spectrum. Nevertheless, in cases where the application relationship between AI services and existing laws remains ambiguous, or when the objective is to safeguard vulnerable segments of society, appropriate measures must be considered, including the potential enactment of new legislation within the respective countries, regions, and domains.

3. Stakeholders and Measures for Responsible AI Deployment

To advance the responsible deployment of AI, it is imperative to delineate the responsible actors and formulate appropriate measures. Particularly, in the continuum of processes spanning AI design, development, provisioning, and utilization, where various organizations and individuals may be involved, the locus of responsibility can become ambiguous. Consequently, in interorganizational transactions encompassing AI development to provisioning, ensuring appropriateness through contractual agreements is important. Furthermore, monitoring mechanisms should be established to ensure proper transactions. However, in transactions between AI providers and consumers, providers should not only take suitable preventive and corrective measures, but AI users can also leverage governance through disciplines other than regulations, such as market dynamics, investments, and reputation, by acquiring appropriate literacy. Additionally, it is advisable to consider establishing remedial measures such as compensation systems when accountability is unclear. Considering that the AI lifecycle extends beyond national, regional, and organizational boundaries, it is essential to promote discussions that enhance transparency regarding the responsibilities and measures of these stakeholders.

Process for formulating this policy recommendation

The content of this proposal is based on discussions between Arisa Ema, a member of AI Governance Project, Technology Governance Policy Research Unit, Institute for Future Initiatives, The University of Tokyo, and external experts Toshiya Jitsuzumi (Professor, Chuo University), Hiroshi Nakagawa (Team Leader, RIKEN Center for Advanced Intelligence Project), and Naonori Kato (Principal Research Supervisor and Director, Next Generation Fundamental Policy Research Institute). The draft was sent to experts in industry, academia, and the government, and an online feedback meeting was held on August 3, 2023, incorporating the feedback received via email and in person. See the Appendix for a list of the people who provided valuable feedback to compile this recommendation.

Towards Responsible AI Deployment

In general, the third AI boom, driven by breakthroughs in deep learning since the mid-2010s, has resulted in the proliferation of artificial intelligence (AI) technologies, particularly those centered around machine learning. In addition, between 2022 and 2023, many generative AI systems were capable of generating content that encompassed text, audio, images, and videos.[1] These AI systems have become familiar to many people, not only in professional contexts, but also in private use, thereby raising expectations for improvements and innovations in our daily lives and work routines. However, this proliferation has brought a spectrum of challenges to the forefront, including misinformation, privacy concerns, copyright issues, and fairness.

Discussions surrounding the governance of AI technologies, systems, services, and their responsible design, development, provisioning, and utilization have focused on AI developers and providers. Nevertheless, machine-learning technologies, encompassing generative AI, are imbued with the capability to train and evolve within society, accentuating the significance of human-machine interactions in various operational contexts.[2] Hence, a comprehensive perspective inclusive of not only AI developers and providers, but also users, public institutions, and society as a whole is warranted to deliberate on the responsible deployment of AI.

This recommendation advocates not imposing responsibility upon users, but rather promoting a shared understanding of transparent and equitable disciplines accessible to all actors. This approach aims to mitigate risks while simultaneously reducing entry barriers for small and medium-sized enterprises and emerging companies, thereby accelerating innovation. In light of these considerations, this policy recommendation advances the three foundational principles of the Hiroshima AI Process, with the objective of appropriately governing AI based on machine learning, including generative AI.

1. Establish Forums for Responsible AI Deployment Discussions

The AI principles developed by international organizations, nations, regions, civil society, and corporations encompass values such as human dignity, human-centrism, fairness, transparency, accountability, and privacy.[3] The inherent significance of safeguarding fundamental human rights and democratic values is common to these principles.[4]

---

[1] Generative AI is composed of a combination of multiple machine learning technologies, but it deserves special mention for its ability to generate human-like natural sentences and other content by creating foundation models and making alignments to individual purposes and fields.

[2] The policy recommendation released by the Institute for Future Initiatives at the University of Tokyo in March 2023 as a predecessor to this recommendation also includes the promotion of human-AI collaboration research, policy, and practice, https://ifi.u-tokyo.ac.jp/en/news/11267/

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

[4] The communiqué of the G7 Hiroshima Summit also affirms the realization of a human-centered, inclusive, and resilient world by promoting common values such as the rule of law, respect for human

The Hiroshima AI Process, initiated in the wake of the G7 Hiroshima Summit, is committed to engaging in international discourse on AI and examining its impact of generative AI. Its primary objective is to affirm the principles that guide the responsible deployment of AI. The risks associated with AI include such as reduced transparency (so-called “blac box” problems, wrong answers, etc) biases in input and training data, inappropriate output results, erroneous responses arising from imperfect training (issues related to fairness), stemming from its inherent characteristics of AI technology. Furthermore, risks that have arisen with existing information and communication technologies, such as misuse, privacy concerns, breaches of confidentiality, and security may also be exacerbated by AI. In particular, the copyright law and the personal information protection law, which are cross-cutting legal systems, have been under review, and risks that infringe upon third-party rights, such as copyright and portrait rights, with the emergence of generative AI has also been discussed in society.[5] necessitating the prompt consideration of intellectual property.[6]

While promoting AI utilization, existing regulations can impede AI design, development, operation, and usage. For instance, with regard to industry laws such as the Medical Practitioners Act and existing safety regulatory systems, while it is important to continue to ensure safety, the issue is that innovation may be hampered by unclear applicability of laws for new AI-based services. Furthermore, from the perspective of concerns about future unemployment and the so-called AI threat, there is an argument, especially overseas, that the precautionary principle should be applied to AI and that it should be placed in a regulatory framework in advance from the stage when the risks are unclear. However, overly stringent regulation of the uncertain future risks and concerns posed by the technology may create barriers to entry for companies, which may inhibit innovation, and even if entry is possible, it may result in market monopolization. In addition to legal challenges, harmonizing the use of generative AI in education and human resource development with existing systems and institutions is important.[7]

To address these complex issues, it is imperative to establish a framework that facilitates agile and ongoing discussions among experts, organizations, and citizens regarding the potential risks posed by AI including generative AI. Leveraging established that international and interdisciplinary multistakeholder forums are essential for supporting these discussions. Preparing for new risks and potential harm to humanity and the environment

---

rights, gender equality, and human dignity. https://www.mofa.go.jp/mofaj/files/100506909.pdf

[5] In Japan, for example, four organizations, including the Newspaper Publishers Association, have released a statement calling for consideration of copyright protection measures for the development of generative AI, and creators' organizations have also voiced their opinions.

[6] In Japan, the Secretariat of Intellectual Property Strategy Headquarters of the Cabinet Office released the "Intellectual Property Strategic Promotion Plan 2023" in June 2023, and has begun discussions on how intellectual property should be handled in the age of generative AI.

[7] Various guidelines have been published in Japan against the use of sentence generation AI and image generation AI by students to do school assignments, and it has also been pointed out that the method of assigning and evaluating assignments needs to be devised.

necessitates the construction of a collaborative framework that transcends national boundaries and fosters cooperation among the G7 nations to safeguard the future.

2. Promote "Interoperability" Between Frameworks and Mutual Respect for Discipline on AI

The term "interoperability," as referenced in the G7 Hiroshima Summit Communiqué and the Ministerial Declaration of the G7 Digital and Technology Ministers Meeting, is primarily discussed in the context of "standards" and "interoperability between frameworks."

Domestic and international standards enable the comparison and interoperability of AI systems developed, operated, and used across organizations and nations. Therefore, standardization discussions and building consensus concerning AI-related terminology, fundamental concepts, and AI governance and management frameworks are essential. Organizations such as ISO/IEC[8], IEEE[9], NIST[10], and CEN-CENELEC are advancing standardization efforts[11].

"Interoperability between frameworks" is introduced in Annex 5 of the G7 Digital and Technology Ministers Declaration as "interoperability between AI governance frameworks."[12] Distinguishing itself from mutual recognition or adequacy decisions involving the coordination of domestic processes, interoperability in the framework layer allows for the concurrent and cooperative existence of AI disciplines and responses from various countries, regions, and organizations.[13] In essence, it affords flexibility in regulatory design, operation, and adaptation by different countries, regions, organizations, and application domains, considering institutional, societal, customary, and cultural contexts while efficiently achieving common global objectives.[14]

---

[8] ISO/IEC JTC 1/SC42 has so far already created and published 20 international standards, including AI-related terms and basic concepts (ISO/IEC 22989) and AI governance (ISO/IEC 38507), and 31 international standards are under discussion.is discussing international standardization of terms and basic concepts related to AI, and ISO/IEC 38507 is discussing international standardization of AI governance.

[9] The IEEE's 7000 series and others discusses standards for practical issues in AI.

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

[11] CEN-CENELEC discusses standards for AI in Europe.

[12] G7 Gunma-Takasaki Digital and Technology Ministerial Meeting, Annex 5: AI Action Plan for promoting global interoperability between tools for trustworthy AI, https://www.meti.go.jp/press/2023/04/20230430001/20230430001-ANNEX5.pdf

[13] For example, the OECD has developed an itemized comparison of ISO, IEEE, and NIST standards, the European AI Act, and the Council of Europe's HUDERIA risk and impact assessment as an "interoperability framework". OECD.AI work promoting interoperability of AI risk management frameworks, IGF Policy Network on AI meeting #4, 18 July 2023, https://www.intgovforum.org/en/filedepot_download/282/25999. The OECD provides other "frameworks" that classify AI systems and the life cycles of AI systems that can be categorized and compared, https://doi.org/10.1787/2448f04b-en. In Japan, the Ministry of Economy, Trade and Industry (METI) has published "Governance Guidelines for Implementation of AI Principles ver. 1.1." https://www.meti.go.jp/english/press/2022/0128_003.html

[14] However, if the AI lifecycle spans across countries or organizations, it will need to conform to the legal framework of some country or organization.

Furthermore, establishing entirely new disciplines for emerging technologies is not always necessary. For instance, risks arising from AI-driven process efficiencies without fundamental changes in AI systems or service products can often be addressed through interpretation or extension of existing regulations. By contrast, excessive or redundant regulations may impede innovation and potentially result in societal losses.

However, in cases where the unclear applicability of laws to new AI-driven services hinders innovation or when entirely new AI systems or services with unique features that lack human intervention emerge, societal responses, including the establishment of mechanisms for pre-assessment through pilot testing, are warranted. In addition, considering the magnitude, frequency, and impact of the risks, it may be necessary to contemplate new legislation and other measures based on multi-stakeholder discussions. Even if some new discipline is to be introduced, given the speed at which AI technology is advancing, an agile system of reviewing the discipline in response to changes in technology should be considered.

Additionally, any discipline should provide transparency and fair opportunities that are accessible and understandable to everyone to ensure effectiveness. This approach not only reduces entry barriers for small and medium-sized enterprises and new companies but also accelerates further innovation.

Column: Diversity of Disciplines

The disciplines adopted by different countries, regions, and application domains exhibit diverse landscapes. Table 1 categorizes the entities responsible for formulating the disciplines and whether these disciplines are enforced. While some regulations are enforced by entities other than nations, such as markets, investments, and reputations,[15] there are instances in which nations do not enforce discipline themselves. These alternative disciplines can prove to be more efficient in rapidly evolving technologies that span nations and organizations. However, achieving this efficiency often requires improvements in people's literacy and awareness.

Table 1: Classification of Disciplines[16]

		Enforcement by nations or not	
		Nation enforces	**Nation does not enforce**
Entity Responsible for Disciplines	**Government**	Legislation / Some Guidelines	Guidelines[17] / Administrative Guidance
	Industry	Standardizations backed by legislation (Mandatory Standards)	Industry guidelines / Corporate policies / Industry standards
	Others (Civil organizations, Academic, etc.)	Customary Law	Market / Investment / Moral / Norms / Academic standards / Customs / Reputation

3. Stakeholders and Measures for Responsible AI Deployment

To promote the responsible deployment of AI, each stakeholder in the supply chain / life cycle of AI development to use must articulate the disciplines they have chosen and be accountable for the consequences. Therefore, it is essential to clarify the responsible entities and consider appropriate responses. The responsibility relationship of each stakeholder in such a supply chain is not limited to AI systems, but similar problems exist in conventional large-scale engineering systems as well, and have been addressed by existing laws. In such a situation, new measures may be required, especially in light of the unique characteristics of AI such as its black-box nature.

From the perspective of the AI lifecycle (design, development, provisioning, utilization, and decommissioning), this policy recommendation presents explanations for key

[15] For example, competition by market, ESG investment, risk of SNS blew up

[16] Created using the framework in the table on p. 6 of Tomohiro Fujita's "Basic Theory of Soft Law. " (in Japanese)

[17] Japan is considered to have a strong influence on businesses even with unenforceable guidelines.

stakeholders in Table 2[18] and summarizes the direction of responsibility in Figure 1.[19]

Table 2: Explanation of Stakeholders

Stakeholder	Explanation
AI Developer	Organizations (or individuals) that develop AI algorithms or models.
AI Trainer	Organizations (or individuals) to train additional pre-processed data (including fine tuning) to the developed AI algorithms or models.
Data Provider	Organizations (or individuals) that supply datasets used for AI algorithms, models, or fine-tuning AI.[20]
AI Provider	Organizations (or individuals) that offer AI systems or services in the market.
AI User (Legal entities)	Organizations that receive AI systems or services from AI providers, integrate them into their own services, and operate them, making them available to their own members or other organizations / individuals.
AI User (Individual)	Individuals who ultimately use AI systems or services provided by AI providers or AI users (Legal entities). They may use AI systems or services for business purposes or private purposes.
Society / General public	Organizations (or individuals) who may be affected indirectly in terms of their rights and interests by AI services or systems used by other organizations / individuals. When the impact of AI service or system use extends to society as a whole, it may also include AI developers, trainers, providers, users, and data providers.
Public Institutions	Organizations such as governments, including national and local authorities, responsible for policy formulation and execution.[21]

[18] This paper organizes the responsibilities of the entities involved along the AI lifecycle and does not cover all the other parties involved. For example, although not mentioned in this paper, recommendations and social communications by industry associations, academic, and others such as civil societies also play an important role in AI governance.

[19] The entities that are expected to be involved in the utilization of AI are classified in the Ministry of Internal Affairs and Communications "AI Utilization Guidelines" (https://www.soumu.go.jp/main_content/000658284.pdf). Figure 1 was created with reference to the guidelines, while taking into account the development and use of generated AI. In addition, while the Utilization Guidelines classify end-users among users as "business users" and "consumer users," in this paper business users are defined as "users" of EU AI Act ('user' means any natural or legal person, public authority, agency or other bodies using an AI system under its authority, except where the AI system is used in the course of a personal non-professional activity), and categorized them into AI Users (Legal entities) and AI Users (Individual).

[20] Although this paper is assuming organizations and individuals who create, collect and provide publicly available data, there are other types of data that AI developers and trainers can collect by crawling the web, or data collected from the behavior of AI users.

[21] Public institutions could also be AI developers, providers, and users (legal entities).

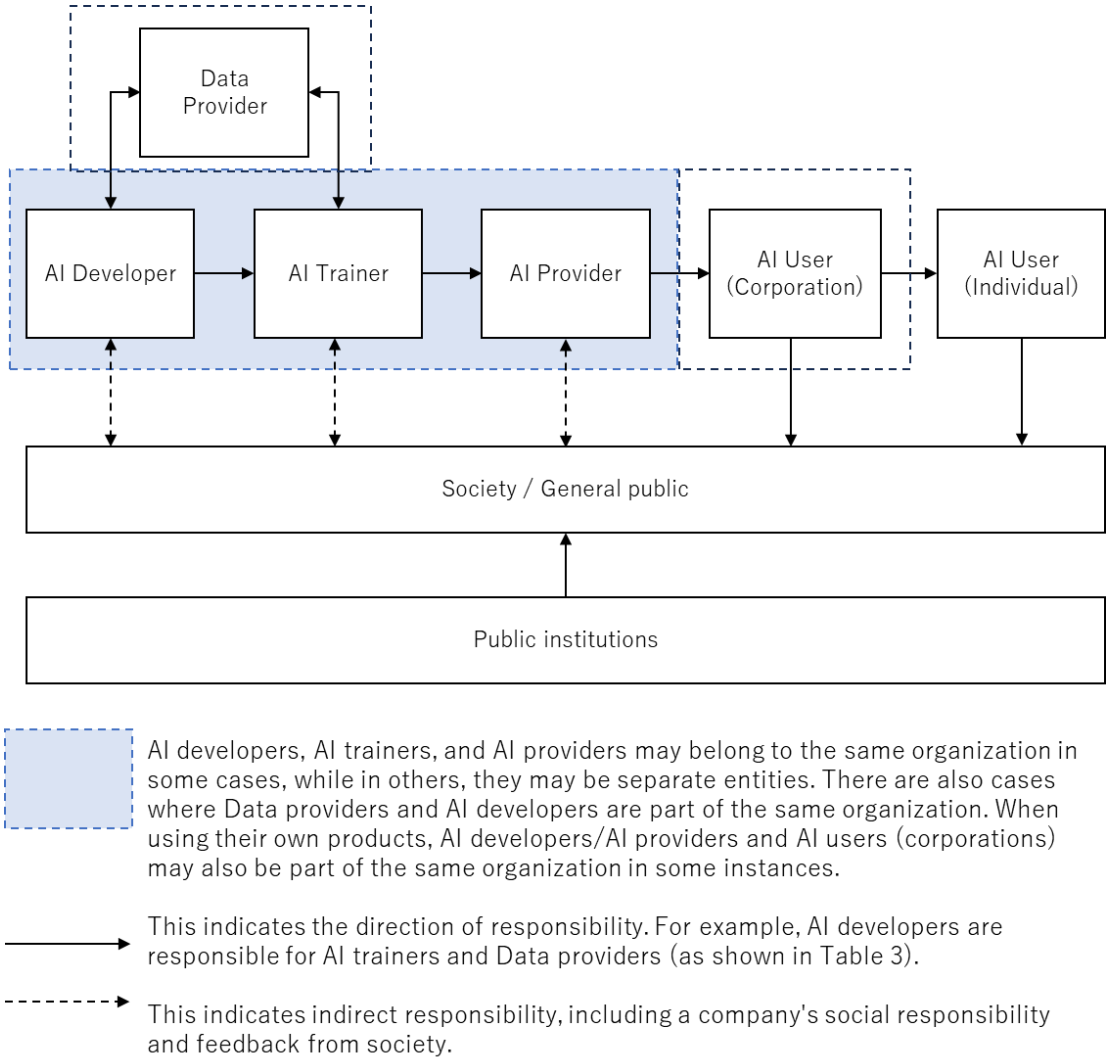

Figure 1: Direction of responsibilities related to AI systems and actors

As indicated in Figure 1, some organizations engage in the entire process of AI service development, from development to provision, and in some cases, even utilization. Conversely, in situations where supply chains span different organizations or services that are integrated by combining multiple AI systems, as illustrated in Figure 2, complexity arises. In such scenarios, where a single entity assumes multiple roles, it is necessary to fulfill the responsibilities required of all roles.

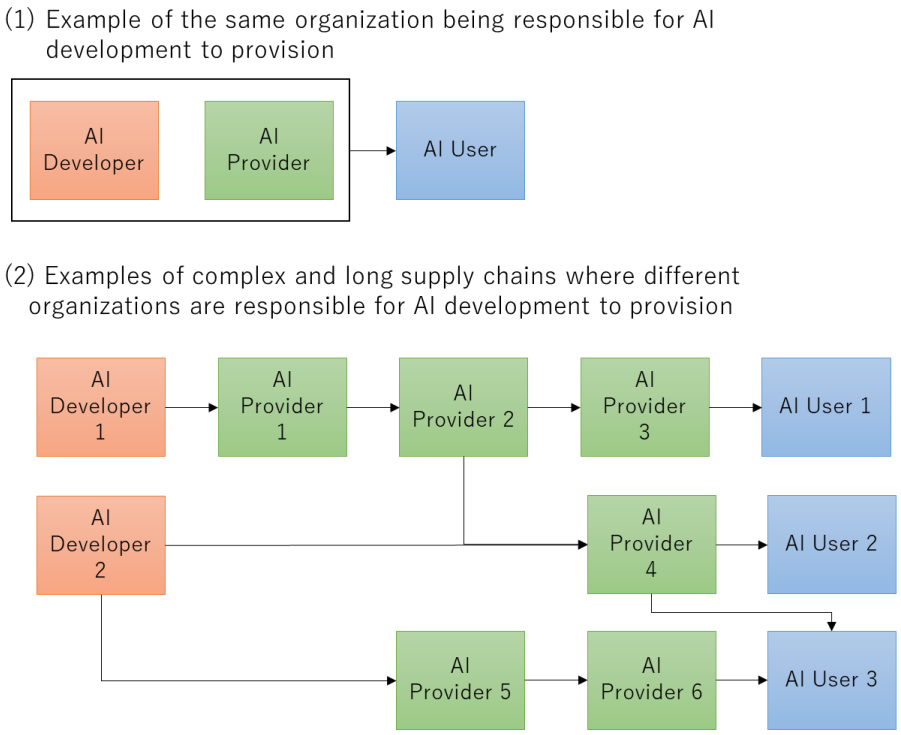

Figure 2: Example of supply chain from AI development to use

Figure 1 illustrates the various responsible entities and their directional responsibilities, which are represented by arrows. For each arrow, Table 3 presents examples of the factors to consider and responsible actions. Additionally, although not explicitly included in Table 3, there is an indirect responsibility depicted by dashed arrows, wherein AI developers, trainers, and providers engage in societal responsibilities, such as sharing their vision and disseminating information to society and the general public. Conversely, society and the public are expected to provide appropriate feedback and monitoring of AI developers, trainers, and providers.

Table 3: Responsible entities and examples of major responsible actions

Responsible entity	Responsibility recipient	Factors to consider	Examples of responsible actions
AI Developer	AI Trainer AI Provider	Transparency	Disclosure of information to a reasonable extent, information sharing[22]
		Filtering	Deterrence of inappropriate input/output
AI Trainer	Data Provider	Reward	Appropriate payment of compensation[23]
		Information control	Appropriate handling of copyrights, personal information
	AI Provider	Transparency	Disclosure of information to a reasonable extent, information sharing
Data Provider	AI Developer AI Trainer	Quality assurance	Quality assurance of provided data
AI Provider	AI User (Legal entities / Individual)	Transparency	Disclosure of information to a reasonable extent, information sharing[24]
		Dissemination of usage environment	Dissemination of information on the environment suitable for AI use and updates
AI User (Legal entities)	AI User (Individual)	Transparency	Disclosure of information to a reasonable extent, information sharing
		Dissemination of usage environment	Establishment of guidelines, response to problems, and dissemination of information on the environment suitable for AI use and updates
	Society / General Public	Proper Use	Prevention of inappropriate use such as abuse and misuse, use in an appropriate environment
AI User (Individual)	Society / General Public	Proper Use	Prevention of inappropriate use such as abuse and misuse, use in an appropriate environment

[22] The World Trade Organization (WTO) lists "prohibition of transfer and disclosure requirements for source code/algorithms," "personal data protection/consumer protection," and "prohibition of cryptographic disclosure requirements for ICT products" to ensure the security and safety of transactions (https://www.meti.go.jp/english/press/2022/0613_002.html). These international rules should also be taken into account when considering transparency in AI.

[23] A particular problem with generative AI is that the value of the information generated is not properly returned to the data creator.

[24] Even after an AI is decommissioned, the output results of that AI are expected to continue to be used everywhere. Therefore, it is important to ensure the transparency of AI as long as another system that uses the output results of AI is in use in society.

Public Institutions	Society / General Public	Market Surveillance	Prevention of oligopoly and monopoly, assurance of fairness in contractual relationships, protection of human rights and social legal interests
		Safety Net	Stabilization of employment and protection of socially vulnerable groups
		Institutional Design	Public-private sector collaboration to promote innovation and mitigate risk, and to support risk management for start-up companies with limited resources
		Education / Literacy Improvement of the Public	Public-private sector collaboration to improve AI literacy of society and the general public
		Improved International Alignment	Public-private sector collaboration to ensure sound foreign transactions, improve interoperability between frameworks

Table 3 illustrates the direction of responsibility and responsible actions from AI developers to users. However, it is essential to respect the diversity of disciplines and emphasize the values adopted among stakeholders based on the challenges faced and societal goals pursued in various countries, regions, and application domains.[25] Nevertheless, considering potential power imbalances among different entities, monitoring is required to ensure fair transactions. An entity's responses can be categorized into two scenarios: business-to-business and business-to-consumer transactions. In this case, AI ethics and guidelines are also important as a basis for promoting fair transactions.

In the case of business-to-business transactions:

1. Measures for Entities Involved in Business Transactions:

When AI developers and providers (as indicated by the blue frame in Figure 1) are not part of the same organization, they should engage in contracts[26] and societal

[25] For example, in terms of respecting diversity, there are various cut-off points such as race, gender, age, education, lifestyle, and the focus may differ depending on the field in which AI is applied. There are also technical indicators of fairness, such as "individual fairness," which refers to the state in which one individual is treated in the same way as other individuals regardless of group attributes, and "group fairness," which refers to fairness among sensitive groups within a group, such as men and women.

[26] 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

commitments.[27] It is crucial to consider the differences between AI model development and data provision/training compared with conventional software development when entering such agreements.[28] Contracts are important not only in discussing the responsibility between entities but also in considering intellectual property rights for training data and AI models.

2. Measures for Public Institutions:

Many contracts between AI developers, trainers, and providers occur between start-up companies (developers and trainers) and large corporations (providers). Therefore, when the parties involved cannot take responsibility through incentives, or when externalities arise, rendering private negotiations suboptimal, public institutions should consider actions such as market monitoring[29], safety nets, institutional design, literacy improvement, and improved international alignment, as presented in Table 3.

In the case of business-to-consumer transactions:

1. Measures for AI Providers:

In transactions between AI providers and AI users (individuals), as well as in cases where the rights or interests of organizations or individuals may be affected by the use of AI services provided by others, it is essential for AI providers to acquire the necessary information from AI developers and trainers and take appropriate measures. The measures are categorized in Table 4 under pre- and post-action and technical and organizational approaches.

---

and Industry and the Japan Patent Office have also published a "Model Agreement for Promoting Open Innovation between R&D Startups and Business Companies."

[27] In July 2023, the U.S. government and seven companies involved in the development of generated AI have made voluntary commitments to the White House to ensure the safety of AI.

[28] For example, it is difficult in deep learning for a vendor to guarantee the completion of the functionality required by the client (performance guarantee), because the content and performance of the model depends on the quality of the data, and unknown data being trained or input may exhibit behavior not expected during development. In addition, in some cases, it may be difficult to retrospectively verify the decisions and outcomes of an AI model.

[29] Internal and external audits to ensure that contracts are fair among contractors will also be important.

Figure 4: Examples of key pre / post- risk mitigation measures AI providers should take

	Examples of proactive/preventive measures	**Examples of post-event/reliability assurance measures**
Technical Measures	Development of AI to control abuse and misuse of inappropriate products, and establish a mechanism to ensure safety outside of the AI implementation section.[30]	Identification of causes for accidents or incidents.
Organizational Management Measures	Development and revision of internal policies, as well as implementation of literacy education for team members.	Establishment of mechanisms for the reasonable implementation, recording, and post-verification of data, training methods, utilization models, and human monitoring.

2. Measures for AI Users (Individuals):

Users can use AI services and systems appropriately by acquiring the necessary literacy, enabling them to navigate AI services and systems without falling prey to misinformation or misuse. Moreover, AI users can govern through means other than legal regulations such as market dynamics, investments, and reputation by becoming capable of properly assessing AI providers. Nevertheless, adequate evaluation requires essential information from AI providers.

In cases where AI users become victims of incidents or accidents caused by AI services or systems or even perpetrators, they should promptly share information with organizations providing compensation, relief measures, and support to prevent the expansion of damages.

3. Measures for Public Institutions and Others:

Compensation and relief measures can be provided by AI providers or insurance products. However, for this to occur, the probability of damage or accident must be estimated both theoretically and empirically. In cases where risk estimation is not feasible owing to new services or technologies, mechanisms need to be established to assess risks through empirical testing or other means, not only by a single organization or company, but also as a risk response for society as a whole. Support for accident cause investigations and the establishment of specialized public committees may also be necessary. Furthermore, for cases

[30] In addition to the implementation part of AI models and algorithms, there could also be a mechanism to secure AI systems, including basic control and other systems that stop in an emergency, AI Network Society Promotion Council (p.17) (in Japanese). https://www.soumu.go.jp/main_content/000637098.pdf

where damage or accidents are likely to occur with a certain probability, but substantial societal benefits can be expected from AI use, the establishment of compensation systems, such as victim support funds, should be considered.

Considering the expansion of the AI lifecycle across countries and organizations, it is important to promote discussions that enhance transparency regarding the responsibilities of various entities and their corresponding measures.

Challenges and Future Developments, and the Path to AI Governance in Japan

Japan initiated discussions on AI governance early in the wake of its third AI boom. At the 2016 G7 Ise-Shima Summit, the Ministerial Meeting on ICT proposed a draft of the principles for AI development.[31] Subsequently, this proposal contributed to international discussions on the necessity of AI governance.

As of 2023, Japan serves not only as the host of the G7 Hiroshima Summit but also as the chair of the OECD's Committee on Digital Economy Policy (CDEP)[32], chair of the Global Partnership on AI[33], and as a host country for the United Nations' Internet Governance Forum 2023 (scheduled for October 2023 in Kyoto).[34] Considering these roles, this recommendation outlines the foundational principles and issues that Japan should propose when organizing and leading discussions on AI governance as the G7 chair.

Beyond the G7, discussions on the responsible deployment of AI have occurred in various international forums. For example, Japan is a member country of the United Nations Educational, Scientific, and Cultural Organization (UNESCO) and the OECD, which propose principles and activities related to AI. Japan is also an observer on the Council of Europe[35], where the AI Convention has been discussed. Additionally, there are recommendations and activities related to AI provided by the Partnership on AI[36], a nonprofit organization for AI research that includes Sony and the University of Tokyo. In the Hiroshima AI Process, incorporating these international discussions and actively promoting principles and activities involving Japan's academia, industry, government, and civil society are expected to make international contributions.

Finally, while this policy recommendation aims to contribute to international discussions within the framework of the G7 Hiroshima AI Process, it is crucial to address domestic systems and frameworks urgently. This recommendation can also be applied to organize domestic issues and clarify policies, hopefully assisting future discussions.

---

[31] 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
[32] OECD-CDEP: https://oecdgroups.oecd.org/Bodies/ShowBodyView.aspx?BodyID=1837&Lang=en
[33] GPAI: https://gpai.ai/
[34] IGF 2023: https://www.intgovforum.org/en/content/igf-202
[35] Council of Europe, Committee on Artificial Intelligence: https://www.coe.int/en/web/artificial-intelligence/cai
[36] PAI: https://partnershiponai.org/

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

Owing to constraints related to time and organizational affiliations, it was not possible to include all the names of individuals. However, it is important to acknowledge that some individuals provided valuable feedback. We express our gratitude to those who contributed their feedback.

Junichi Arahori, Head, AI Ethics and Governance Office, Fujitsu Limited
Tagui Ichikawa, Specially Appointed Professor, Institute of Innovation Research, Hitotsubashi University
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
Takehiro Ohya, Professor, Faculty of Law, Keio University
Atsushi Okada, Attorney-at-Law, Mori Hamada & Matsumoto
Takafumi Ochiai, Attorney-at-Law, Atsumi & Sakai
Taichi Kakinuma, Attorney-at-Law, STORIA Law Office
Naonori Kato, Principal Research Supervisor and Director, Next Generation Fundamental Policy Research Institute (*)
Yoshihiro Kawahara, Professor, Graduate School of Engineering, the University of Tokyo
Kit Kitamura, The Head of AI Legal Group, CDLE (Community of Deep Learning Evangelists)
Fumiko Kudo, Visiting Academic Staff, Osaka University Research Center on Ethical, Legal and Social Issues
Revolution Japan
Jun Kuribayashi, Master Student, Graduate School of Public Policy, the University of Tokyo
Satoshi Kurihara, Professor, Keio University, Faculty of Science and Technology / Director, Center of Advanced Research for Human-AI Symbiosis Society
George Shishido, Professor, Graduate Schools for Law and Politics, The University of Tokyo
Hideaki Shiroyama, Professor, Institute for Future Initiatives, The University of Tokyo
Toshiya Jitsuzumi, D.Sc., Professor, Chuo University (*)
Roy Sugimura, Supervisory Innovation Coordinator, National Institute of Advanced Industrial Science and Technology (AIST)
Shoko Suzuki, Professor Emeritus, Kyoto University / Senior Visiting Scientist, RIKEN Center for Advanced Intelligence Project
Koichi Takagi, Group leader, Planning Group, Technology Affair Dept., KDDI Corp.
Hideaki Takeda, Professor / Director, Principles of Informatics Research Division, National Institute of Informatics
Kenzaburo Tamaru, National technology officer, Japan Microsoft Corporation
Hiroshi Nakagawa, Team leader, RIKEN Center for Advanced Intelligence Project (*)

Satoshi Narihara, Associate Professor, Faculty of Law, Kyushu University
Hiroki Habuka, Research Professor, Graduate School of Law, Kyoto University / CEO, Smart Governance Inc.
Yuko Hararyama, Professor Emeritus, Tohoku University
Shinnosuke Fukuoka, Attorney-at-Law (Japan & New York), Nishimura & Asahi.
Satoshi Funayama, Chief legal officer, rinna Co., Ltd.
Naohiro Furukawa, Attorney-at-Law, ABEJA,Inc.
Yutaka Matsuo, Professor, Graduate School of Engineering, The University of Tokyo
Toshiya Watanabe, Professor, Institute for Future Initiatives, The University of Tokyo

Contributors to the drafting of this policy recommendation are indicated by asterisks (*).