Statement on AI Governance
Ver. 3.0

March 2026

Digital Policy Forum Japan

Table of Contents

Introduction

The Digital Policy Forum Japan (DPFJ) has been conducting discussions on AI governance since spring 2024. Focusing on the central theme of how to ensure the controllability — both technical and societal—of rapidly evolving AI, the forum engaged in deep discussions with numerous experts and representatives from government and industry. It published its first statement in July 2024 and its second set in December of the same year.

This third statement established three pillars for AI governance discussion:

(1) minimizing AI risks,

(2) maximizing AI convenience, and

(3) fostering a sound market environment.

Key issues were then organized within this framework, which was maintained for the second and current third statement.

During the compilation of the third statement, it became clear that the practical use of AI is rapidly advancing at both corporate and individual levels. Consequently, the issues surrounding AI are also becoming more concrete. For example, interviews with corporate stakeholders revealed that management's understanding of AI has deepened, making AI utilization a key agenda item in business strategy. Furthermore, it has become increasingly common to see perspectives focused not only on improving management efficiency (cost reduction) through AI, but also on enhancing added value by creating new businesses.

Another defining feature of this statement is that policy discussions surrounding AI now span a broad range of fields, including industrial policy, competition policy, science and technology policy, security policy, and foreign policy. Furthermore, the interconnectivity between these policy domains is intensifying. This signifies that the approach to AI governance has become a high-priority policy theme directly linked to national interests.

Furthermore, as the practical application of AI advances, the scope of issues requiring discussion has significantly expanded. Specifically, moving beyond policy discussions solely about AI, this third statement substantially broadens the discourse to include questions like: What aspects of the socio-economic landscape will AI transform? For example, the relationship between AI and creativity, the potential for digital democracy, and the relationship between religion and AI. The Internet once enabled us to transcend the constraints of time and distance. Now, AI is successively enabling us to

overcome language barriers, organizational barriers, and the divide between the cyber and physical realms. How profound an impact will the implementation of AI, the accelerator of our data-driven society, ultimately have? Collaboration with experts across a broad spectrum of fields, not just the traditional IT-centric tech community, will become crucial in future discussions surrounding AI governance.

A critical consideration in AI discussions is that the very subject of debate—AI itself—continues to evolve rapidly. Dr. Eliezer Yudkowsky, a prominent U.S. AI researcher, warns: “The greatest danger of AI is that people believe they fully understand it.” DPFJ plans to conduct ongoing analysis of technological trends, market shifts, and the socioeconomic impacts surrounding AI, compiling these findings into its fourth statement.

Finally, we would like to take this opportunity to express our deepest gratitude to the experts, companies, and other stakeholders who contributed to the creation of our third statement.

Yasuhiko Taniwaki

Chairman, Digital Policy Forum Japan

March 2026

Basic Approach

Technological development surrounding generative AI is advancing at a remarkable pace, and efforts to implement generative AI within socioeconomic systems are also accelerating. Against this backdrop, the creation of rules to govern generative AI is shifting focus from general discussions, such as establishing legal frameworks in various countries, toward debates on the specifics of policy implementation.

This document[1], bearing these generative AI trends in mind, organizes discussions on AI governance with the following three fundamental objectives as its core perspective, aiming to achieve them in a balanced manner:

1. Ensuring the controllability of AI risks (minimizing risks)

2. Establishing an environment where the benefits of AI can be fully realized (maximizing benefits)

3. Creating a market that autonomously realizes the above environment (establishing a sound market environment)

AI governance, specifically the mechanisms for continuously maintaining the controllability of AI technology, requires ongoing discussion that balances the benefits and risks AI brings.

AI contributes to enhancing productivity and creativity across all societal domains. It delivers various benefits, such as enabling access to highly convenient services while technically safeguarding individual data sovereignty through personalization (decentralization of intelligence).

On the other hand, risks exist, including the potential for severe harm such as human rights violations, the risk of losing human controllability, and risks arising from AI replacing humans. Incidentally, regarding these risks, aiming for technical solutions wherever possible is preferable to hastily introducing regulations, as the latter is not conducive to fostering innovation.

The Digital Policy Forum Japan (DPFJ) published Version 1.0 of its proposed discussion points for establishing an AI governance framework in July 2024. Furthermore, in December of the same year, it published Statement Version 2.0, which included a more detailed organization of discussion points. This version specifically addressed the need for a legal framework and the advancement of a comprehensive AI strategy, reflecting the concrete discussions surrounding AI legislation.

However, AI technology continues to evolve rapidly, and the nature of AI itself as a subject of

---

[1] This document was supervised by Yasuhiko Taniwaki, Chairman of the Digital Policy Forum Japan (DPFJ).

discussion is constantly changing.

Therefore, building on the expert interviews, hearings with corporate stakeholders, and related meetings held since Statement Ver. 2.0, and incorporating recent international trends, we have reorganized the key issues concerning AI governance. This is presented as the “Statement on AI Governance Ver. 3.0.”

This document addresses a wide range of topics essential for ensuring AI governance, including legal frameworks, risk management methodologies, competition policy, industrial policy, foreign policy and security policy concerning AI, and impact analysis of societal structural changes brought about by AI. It is crucial to recognize that these topics interact organically with one another. Discussing AI governance requires a bird's-eye view.

Please note that this document will primarily focus on generative AI currently available to the general public for discussion purposes, excluding Artificial General Intelligence (AGI) except in limited contexts.

1. Minimizing Risk

(1) Approach to Risk Management

Challenges in Tiered Risk Management for AI

Various methods exist for managing risks (including negative impacts on human life and fundamental human rights) by dividing them into several tiers as an AI management technique. For example, the EU's AI Act classifies risks into four tiers.

This approach manages risks inherent in AI models by severity level while linking this to the degree of regulation. However, this method faces challenges. Beyond determining the scope of risks to control and the criteria for risk classification, it is not entirely clear whether the following are sufficiently established: the entity responsible for conducting risk assessment, the methods for clearly demonstrating the accuracy of that entity's judgments to third parties (accountability), and other related aspects.

In this regard, if a future system is established to score AI risks by recording and analyzing AI system logs (operational history) and then publishing these scores, it may become possible for users to select AI based on their own risk tolerance. Specifically, while recognizing the potential to build a system where users can compare the benefits and risks of a given AI and select a (personalized) AI

suited to their specific purpose, it is necessary to actively participate in discussions within international organizations on AI standardization (including risk assessment methodologies).

Furthermore, linking AI model risk assessment to different levels of regulation could undermine the predictability of regulatory application, given that risks themselves may change. In this regard, Europe is advancing the "Digital Omnibus Proposal,"[2] which seeks to relax technical requirements for high-risk AI. We must closely monitor future developments in this discussion.

Indeed, the sources of risk inherent in AI are diverse, making it difficult to grasp the full picture. For example, according to an MIT study[3], over 700 risks exist surrounding AI, and implementing a risk management system that accounts for all of them presents significant challenges. Furthermore, this study points out that "post-deployment risks" account for 65% of all risks, indicating that risks dynamically and qualitatively change over time.

Of course, AI risk management itself is critically important. Domestically, industry-academia-government collaboration should actively promote the creation and analysis of AI risk repositories.

Risk Management by Stakeholder

Considering the above, AI risk management should ideally be examined by dividing stakeholders into three categories:

- AI developers

- Service providers implementing AI

- End users

Incidentally, when considering AI risks, two types can be identified: risks inherent in AI during the development phase, and risks AI will possess during the service provision phase (risks that may materialize depending on how the AI-implemented service is provided and used, such as the generation and dissemination of misinformation or disinformation). Particularly for the latter risks, careful discussion is necessary regarding whether the risk was first introduced by AI or whether it is

---

[2] European Commission "Simpler EU digital rules and new digital wallets to save billions for businesses and boost investment" (November 2025).
https://ec.europa.eu/commission/presscorner/detail/en/ip_25_2718

[3] P. Slattery et al. "Global AI adoption is outpacing risk understanding, warns MIT CSAIL" (MIT CSAIL News, August 14, 2024)
https://www.csail.mit.edu/news/global-ai-adoption-outpacing-risk-understanding-warns-mit-csail#:~:text=Global%20AI%20adoption%20is%20outpacing,it%20remains%20current%20and%20relevant.

a pre-existing risk that has been made apparent or amplified by AI.

Regarding risk management by entity, risk management by AI developers should be limited to a "do-not-do list" approach that enumerates specific considerations during development. Going forward, the basic approach should be to address specific issues as they arise during AI development, while also implementing regular monitoring.

Specifically, principles such as "ensuring activities throughout the AI system lifecycle are sufficiently aligned with human rights, democracy, and the rule of law" could be considered, referencing sources like the Council of Europe's AI Convention (September 2024)[4].

Furthermore, risk management by service providers implementing AI should be as limited as possible. It is desirable to confine regulations to prohibiting unfair discriminatory treatment in service provision, as stipulated in Article 6 of the Telecommunications Business Act.

Prohibiting unfair discriminatory treatment related to AI is necessary from a human rights protection perspective. While AI enables the provision of highly personalized services utilizing personal information, such as individualized medical care, it must not result in discrimination lacking reasonable justification based on individual characteristics, rather than truly personalized service provision.

Furthermore, when providing services incorporating AI to users, it is necessary from the perspective of user protection to clearly define the division of responsibility between the AI developer and the service provider in advance, at the stage prior to service provision.

Additionally, regarding risk management by end users (including SMEs), literacy education is required to ensure a proper understanding of AI risks[5] (see Item (5)).

Risk Management Methods

Risk management should be conducted based on the results of creating and analyzing the aforementioned AI risk repository. In doing so, it is necessary to consider the applicability of either self-assessment or third-party assessment (e.g., audit or certification systems).

Specifically, for AI developers, self-assessment by the developer themselves should be fundamental.

---

[4] Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law）
https://www.mofa.go.jp/mofaj/press/release/pressit_000001_01725.html

[5] Vesteinsson, Baker, Brody, Funk, Grothe, Slipowits eds. Freedom on the Net 2025, Freedom House, 2025 www.freedomonthenet.org

For AI closely related to socially essential critical services, combining this with a third-party audit system could be considered.

Furthermore, for service providers, evaluating AI functionality in isolation is difficult. Therefore, consideration and implementation should occur within existing frameworks for user protection under relevant industry laws. Adding additional regulations solely because of AI utilization is not appropriate.

The Need for Systemic Risk Analysis

When considering AI risks, there are two aspects: risks inherent to AI itself (risk of AI) and risks caused by AI (risk by AI). The latter risk—the potential for risk factors to cascade and cause larger, more severe negative impacts—constitutes AI's systemic risk, necessitating cross-domain analysis[6]. As digital technologies, including AI, are expected to become highly versatile societal infrastructure, analyzing the systemic risks AI brings will become increasingly important.

(2) Regulatory Approaches and Ensuring Effectiveness

Regulatory approaches for AI include hard law (statutory regulation), soft law (voluntary regulation by the private sector), and co-regulation through public-private partnerships combining these methods. For example, the European “AI Act” and China's “Interim Measures for the Administration of Generative AI Services” are primarily based on hard law.

However, even when pursuing hard law, there exists a range of approaches regarding the nature of regulation, such as a foundational, flexible approach or a more disciplined approach imposing specific behavioral restrictions.

Co-regulation involves the state setting basic policy guidelines, with participating businesses implementing rules based on these guidelines and reporting outcomes to the state. The state then evaluates these reports and revises the basic policy as necessary. This approach has been adopted in Europe for measures against disinformation by platform operators. While co-regulation excels in enabling flexible, private-sector-led discipline application, it requires ensuring sufficient transparency to prevent administrative discipline from being exercised arbitrarily without legal basis.

Amid rapid technological innovation, past AI-related discussions have occasionally tended to

---

[6] World Economic Forum “The Global Risks Report 2026” (January 2026) https://reports.weforum.org/docs/WEF_Global_Risks_Report_2026.pdf

become overly cautious and abstract, detached from market realities. The fundamental approach should be to advance three pillars in tandem:

- Ensuring necessary discipline
- Promoting the digital industry (maintaining an appropriate balance between regulation and promotion)
- Achieving international harmonization of regulations

while maintaining a premise of calm discussion and prioritizing voluntary initiatives by relevant parties.

Deepening Rivalry Among Major Nations

The United States, under the previous Biden administration's executive order (October 2023), suggested the possibility of enacting an AI law while advancing efforts to establish safety and security standards for AI and develop guidance to prohibit discrimination by AI algorithms.

However, when the Trump administration returned to power in January 2025, it announced a comprehensive review of the previous administration's AI policies. In July of that year, it formulated and released its “AI Action Plan.”

This plan listed 30 initiatives under three pillars: accelerating AI innovation, building AI infrastructure within the United States (including strengthening semiconductor manufacturing and upgrading the power grid), and leading in international diplomacy and security related to AI. Furthermore, in November of the same year, as a concrete measure to accelerate AI innovation, the administration announced the launch of a new national project, the “Genesis Mission.” It also announced a policy to promote strategic technology development using the ASSP (American Science and Security Platform), an integrated platform developed by the federal government itself.

This U.S. approach (prioritizing non-regulation and industrial promotion) sharply contrasts with Europe's approach toward implementing AI legislation (where regulation is a prerequisite for achieving a trustworthy AI environment).

For example, at the AI Action Summit held in France in February 2025, EU Commission President Gertrud von der Leyen asserted the legitimacy of the hard law approach, stating, “People need confidence that AI is safe, and that is precisely the purpose of the AI Act.” In contrast, U.S. Vice President JD Vance countered that “over-regulating the AI sector could kill an innovative industry

that is just taking off," and did not sign the joint statement issued at the summit[7] (signed by 64 countries and regions, including Europe and Japan).

Furthermore, within the United States itself, a divergence is emerging between the federal government's approach and that of state governments. At the state level, 48 states have imposed AI regulations (excluding two states where bills are under deliberation), imposing certain restrictions on the operation of AI chatbots, the generation of fake/misinformation, and AI-based medical practices. Regarding these differing stances on AI regulation between the federal and state governments, the federal government is establishing systems to address excessive state regulations, such as filing lawsuits[8]. Various developments are expected to emerge in the future.

Enactment of AI Legislation in Japan

In Japan, the "Act on Promotion of Research and Development and Utilization of Artificial Intelligence-Related Technologies" (AI Act) came into effect in September 2025.

In the preceding DPFJ Statement Ver. 2.0 (cited above), it was recommended that:

- When considering legal frameworks surrounding AI in Japan, rather than codifying in detail the contents of various guidelines previously discussed by the government, a fundamental AI law should be enacted as hard law;

- The Basic AI Act should stipulate, for example, while referencing Article 18 of the Basic Act on Cybersecurity, the fundamental principles of AI policy, the responsibilities of national entities, the formulation of AI strategies, the authority of the government's AI Strategy Headquarters (and its Secretariat), and coordination with relevant agencies;

- For cross-industry initiatives such as risk management by developers, efforts should be centered around the Headquarters Secretariat established within the Cabinet Secretariat. For service providers, efforts should be conducted by the competent ministries and agencies for each industry. However, when matters requiring cross-industry assurance (unified standards) are deemed necessary from a user protection perspective, particularly given AI's characteristics, it is desirable for the Headquarters Secretariat to take the lead, collaborate

---

[7] https://www.elysee.fr/en/emmanuel-macron/2025/02/11/statement-on-inclusive-and-sustainable-artificial-intelligence-for-people-and-the-planet

[8] White House, Executive Order "Ensuring a National Policy Framework for Artificial Intelligence" (December 11, 2025) https://www.whitehouse.gov/presidential-actions/2025/12/eliminating-state-law-obstruction-of-national-artificial-intelligence-policy/

with relevant ministries and agencies, and promote unified measures.

These three points were highlighted.

While taking the form of hard law, this legislation maintains a non-regulatory approach composed of fundamental provisions, consistent with the line of the Version 2.0 statement, which is commendable.

Furthermore, the “Basic Plan for Artificial Intelligence” (approved by the Cabinet in December 2025) formulated based on the AI Act also clearly establishes the direction for advancing related measures through collaboration between the AI Strategy Headquarters and relevant ministries and agencies. It sets forth three principles: “Balancing Innovation Promotion and Risk Response,” “Agile Response,” and “Integrated Domestic and International Policy Promotion.”

Effectiveness of Rules and Player Autonomy

As noted above, the AI Act itself maintains a non-regulatory stance. However, when establishing non-binding rules for AI development and use through co-regulation or guidelines in the future, careful consideration is necessary to ensure that no substantive restrictions (regulations) are imposed on the autonomy of AI developers and users, from the perspective of ensuring the effectiveness of such rules.

For example, the “Draft Principles and Code of Conduct for the Protection of Intellectual Property and Transparency in the Appropriate Use of Generative AI” (Cabinet Office Intellectual Property Strategy Promotion Office), published in December 2025, is positioned as a code of conduct to avoid risks of intellectual property infringement associated with AI use. While the initiative itself is commendable, further consideration is needed to ensure that rules developed under the “comply or explain” principle do not become overly detailed, thereby functioning as substantive regulation despite being a code of conduct.

Furthermore, discussion is needed regarding the extent to which third-party verification of AI algorithm validity and transparency should require disclosure of verifiable information.

In traditional information retrieval, highly relevant URLs were displayed, but users still needed the literacy to read and interpret the information themselves. However, AI outputs merely present information organized by algorithms, leaving no opportunity to evaluate the algorithm's accuracy. Therefore, establishing a mechanism for objective third-party evaluation of algorithms is considered one of the most critical challenges for the future societal implementation of AI.

For instance, regarding self-audits expected of developers and external audits by third parties, these should remain voluntary measures by developers. A mechanism could be considered where, when necessary, collaboration occurs with the government (AI Strategy Headquarters). This could involve a co-regulatory approach, such as the government establishing audit guidelines and revising them based on actual audit findings.

While discussions about introducing regulations specifically for developers of large-scale AI are occasionally seen, as mentioned earlier, it is also an option to limit large-scale AI to voluntarily introducing third-party external audits as an option, distinguishing this from competition policy-related discussions about how the scale of AI affects relevant markets (see item (6)).

(3) Vulnerability Countermeasures Against External Risks

As AI becomes a societal infrastructure, mission assurance[9] to ensure AI resilience is critically important. Therefore, stakeholders must collaborate on countermeasures, particularly those addressing external risks such as AI vulnerabilities.

Countermeasures Against Cyberattacks Targeting AI

From the perspective of managing external risks to AI models, incorporating vulnerability assessments (red teaming) into audit items (self-audits or third-party external audits) is appropriate. Guidelines for implementing this should be developed through public-private collaboration. In Japan, progress is being made on this front, such as the AI Safety Institute (AISI) publishing the "Guide to Red Teaming Methods for AI Safety" in September 2024.

Given the diverse characteristics of AI, it is important to clearly define the scope and purpose of vulnerability assessments during this consideration to ensure effectiveness.

Furthermore, risks exist where AI fails to perform as intended or behaves unexpectedly due to attacks such as data poisoning during its learning process (cyberattacks against AI). Additionally, risks are becoming apparent where AI is exploited to discover vulnerabilities, create malware, generate fake accounts, or disseminate misinformation (cyberattacks using AI). Concrete countermeasures against both "cyberattacks targeting AI" and "cyberattacks using AI" must be

---

[9] US Department of Defense "Mission Assurance Strategy" (April 2012)) https://policy.defense.gov/Portals/11/Documents/MA_Strategy_Final_7May12.pdf

urgently developed.

Furthermore, the proliferation of AI is intensifying cognitive warfare, with activities such as mass-generating fake accounts and spreading anti-government rhetoric via bots. This necessitates accelerating the development of digital technologies to counter such cognitive warfare[10].

When considering the above, concerns exist that opening up training data and AI systems to ensure AI transparency could lead to the manifestation of vulnerabilities, malicious imitation by third parties, or misuse for criminal activities. Therefore, it is necessary to simultaneously advance discussions on both ensuring openness and the potential for AI misuse.

Ensuring Data Space Integrity

During the process of AI training on data over multiple iterations, it often employs a process of discarding low-frequency data (aiming to improve query hit rates). In such cases, it has been pointed out that words with high occurrence probabilities in previous-generation models may be overvalued in the next generation, while words with low occurrence probabilities may be undervalued. This leads to a loss of model diversity (model degeneration), known as "model collapse."[11] Allowing this situation to persist spreads inaccurate and unhealthy data, accelerating the contamination of the data space.

Therefore, it is necessary to consider establishing certain regulations, such as limiting AI training data to human-created content or explicitly disclosing externally that the AI is pre-trained. For example, establishing a democratically-led certification system should be considered.

Furthermore, from the perspective of increasing human-created data, open data initiatives that widely utilize copyright-expired documents and documents created by public institutions as training data are effective.

(4) Handling of Outputs

AI learns from training data to form language models, then produces outputs through an inference

---

[10] Open AI "Disrupting malicious uses of AI: an update"(October 2025)
https://cdn.openai.com/threat-intelligence-reports/7d662b68-952f-4dfd-a2f2-fe55b041cc4a/disrupting-malicious-uses-of-ai-october-2025.pdf

[11] I. Shumailov et al. "The Curse of Recursion: Training on Generated Data Makes Models Forget" arXiv (May 2023)
https://arxiv.org/abs/2305.17493

process utilizing these models. Therefore, from the perspective of ensuring the “integrity” of data—meaning it has not been tampered with by third parties—while the “ensuring the integrity of the data space” mentioned in (3) focuses on ensuring the integrity of input values (training data), efforts to ensure the integrity of output values (generative outputs) are also necessary. Therefore, given the current situation where vast amounts of AI-generated misinformation are already circulating, it is necessary to effectively and concretely advance countermeasures against misinformation, based on a collaborative regulatory approach.

In this context, the introduction of digital watermarks to enable identification of AI-generated content is considered effective. Furthermore, regarding the effectiveness of Originator Profile (OP) technology for users to verify the creators and disseminators of online information (content), it is necessary to deepen discussions related to international standardization of technical specifications and the appropriate entities responsible for applying OP.

2. Maximizing Convenience

(5) Proactive Utilization of AI

Japan still lags behind other nations in digital technology adoption, with significant delays particularly evident in AI implementation. Even among companies deploying AI, most utilize U.S.-developed systems, making it highly likely Japan's digital deficit will widen further (see item (9)). Against this backdrop, actively leveraging AI has become an urgent priority for Japan's socio-economic development.

Recognizing that a nation's ability to independently develop and operate digital technologies constitutes “digital sovereignty,” with AI serving as its core component, the government must proactively address policy challenges in the AI domain.

Promoting AI Utilization for Problem Solving

While various initiatives for AI utilization are already underway, given the particularly severe aging and declining birthrate, and considering the lag in data utilization efforts in the education and healthcare sectors, it is necessary to actively advance AI utilization in these fields.

Mechanisms that link and analyze relevant data starting from students in education and patients in healthcare, with individual consent, are expected to contribute to the personalization of education

and healthcare.

On the other hand, it is also necessary to consider certain safeguard measures to ensure that such data linkage does not lead to excessive profiling. Furthermore, AI analysis can enable automatic linkage even in cases where data linkage has been difficult due to differing data formats across regions or organizations, such as medical record data.

Furthermore, beyond education and healthcare, proactive AI utilization must be pursued across a broad spectrum of fields, including global challenges like environmental measures, disaster prevention and mitigation to protect lives and property, and cultural initiatives to realize enriched lifestyles. In doing so, it is necessary to deepen consideration of key considerations and technologies to be developed for actively leveraging AI in these domains.

Simultaneously, necessary measures from a privacy protection perspective are required, such as handling personal data as training data and avoiding the inclusion of personal data in outputs when such data is incorporated. Clarification is also needed regarding the treatment of training data and generated outputs under copyright law.

Furthermore, as previously stated, literacy education is crucial for enabling general users to correctly understand AI risks. Similar to public-private initiatives for improving the internet environment for youth, it is important to conduct widespread awareness and educational activities regarding AI risks.

Promoting AI Utilization in Administrative Services

In providing administrative services at the national and local government levels, amid the ongoing challenges of a declining birthrate and aging population, it is necessary to efficiently allocate limited human resources through AI utilization and proactive data integration, while also achieving personalized, meticulous services.

However, recognizing that gaining public understanding is essential for the proactive use of AI in delivering these administrative services, efforts are required. These include establishing and operating necessary institutional frameworks (formulating basic guidelines and conducting risk assessments), referencing examples such as the cases in Kobe City, Hyogo Prefecture, and sharing best practices.

AI Utilization and the Labor Market

Some argue that actively utilizing AI advances societal automation and leads to job losses (replacing

human labor). However, the fundamental policy direction should be to utilize AI not as a replacement for existing labor, but solely as a tool to enhance labor productivity and create new market areas. The government is expected to provide necessary policy support to realize this.

Digital technologies, including AI, are not solely intended to enhance the efficiency of existing markets. Rather, it is essential to widely share the understanding that they create new employment by breaking down barriers in existing business domains and generating new market areas.

3. Fostering a Sound Market Environment

(6) Building a Sound Ecosystem—Competition Policy

The evolution of AI should fundamentally be driven by private-sector innovation. The government should actively support this while establishing necessary rules and providing policy support to ensure the public interest.

To secure an ecosystem that involves diverse stakeholders, including AI developers and users, competition policy aimed at establishing a sound market environment is crucial[12].

Therefore, mechanisms must be established to monitor entry barriers in AI-related markets and anti-competitive practices such as the abuse of dominant positions by large corporations. Furthermore, while major AI systems are currently predominantly provided by existing large-scale platform operators, it is necessary to examine the potential for abuse of market dominance in the AI market or adjacent markets (e.g., platform operations) and the corresponding competition safeguard measures.

Particular concern exists that vertically integrated AI developers, such as platform operators, which operate across multiple layers, may possess greater market dominance than other developers and are more likely to exercise that dominance in adjacent markets. How competition policy should address this requires consideration.

Furthermore, when examining whether market dominance is being abused, market definition approaches should be considered with an eye toward cross-border data flows and the networking of AI.

Furthermore, while the European “AI Act” includes provisions for extraterritorial application, the

---

[12] OECD “Artificial Intelligence, Data and Competition” OECD Artificial Intelligence Papers No. 18 (May 2024)
https://www.oecd.org/daf/competition/artificial-intelligence-data-and-competition.html

potential for such extraterritorial application to increase could lead to excessive regulation, such as the overlapping application of foreign regulations domestically. Measures to avoid this potential outcome also require consideration.

(7) Industrial Promotion and Global Collaboration—Industrial Policy

Digital industries, including AI-related services, often tend to form oligopolistic market structures by leveraging the characteristics of data (such as zero marginal cost and non-rivalry) and its scalability. At the same time, they require global connectivity. Consequently, a world where AI is networked across borders is envisioned—one where computing resources are diversely combined and where the functionality of AI is enhanced through the interaction of networked AI systems. Given this worldview, ensuring AI openness becomes essential, alongside the need for global rule-setting (see Item (8)).

Ensuring Openness and Standardization Strategy

One key driver of the Internet's explosive growth was its openness, rooted in the fundamental principles of "autonomy, decentralization, and collaboration." Similarly, two approaches to AI can be considered: closed proprietary AI and open AI. From the perspective of fostering healthy market development and maintaining the quality of AI-related services, ensuring openness that creates a sufficiently competitive environment is indispensable. A similar approach is also seen in Europe and the United States[13].

From this perspective, the government should actively promote initiatives such as utilizing open source, ensuring interoperability between different AI systems, advancing standardization to realize such an environment, and providing R&D support premised on encouraging open AI development.

Furthermore, given that Japan is already lagging in AI-related technological development within the global market, the government should consider implementing proactive promotion measures for open AI, such as supporting the development of solutions incorporating open AI. Discussions are

---

[13] European Commission（DG COMP）"Competition in Virtual Worlds and Generative AI: Calls for Contribution" （January 2024）
https://ec.europa.eu/commission/presscorner/detail/en/ip_24_85
European Commission（DG COMP）"Competition Policy Brief"（September 2024）
https://competition-policy.ec.europa.eu/document/download/c86d461f-062e-4dde-a662-15228d6ca385_en

particularly needed to strengthen initiatives supporting AI-focused ventures.

In doing so, it is crucial to distinguish between formal openness and substantive openness, with policy aiming to ensure the latter. For example, if AI training data or specific feedback within the RLHF (Reinforcement Learning from Human Feedback) process remains undisclosed, while technical specifications may be open, the substantive openness of the AI cannot be guaranteed (or proven). Consequently, safeguards to ensure effective openness must also be considered.

(8) Fostering International Consensus—Foreign Policy

AI is not developed or utilized in isolation within national borders; its development and use presuppose networking and widespread utilization in cyberspace. In this context, it is essential to form a loose international consensus on the aforementioned issues, reflect them in each country's legal systems and other rules, and achieve necessary harmonization.

Given that AI is a strategic field significantly impacting national industrial competitiveness and problem-solving capabilities, a comprehensive approach involving experts from diverse fields—including industry, technology, and diplomacy—is essential. This necessitates the establishment of effective frameworks through inter-ministerial coordination and public-private partnerships. Furthermore, recognizing AI's substantial potential to contribute to solving challenges faced by the Global South, progress must be made with the full participation of these nations.

Furthermore, within this process of building international consensus, the formation of norms concerning the military use of AI is particularly urgent. It is necessary to expand voluntary commitments regarding AI use, such as those proposed in the “Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy”[14] from the “Responsible AI in the Military Domain” (REALM) Summit held in The Hague in February 2023. Simultaneously, incorporating AI security audit (inspection) mechanisms within the UN security framework warrants consideration. Given that the military use of AI is already a reality[15], discussions on this approach to AI and security must be expedited. In this regard, serious concern exists that the significant disruption to coordinated international cooperation, exemplified by the US withdrawal from numerous

---

[14] US DoS “Political Declaration on Responsible Use of Artificial Intelligence and Autonomy” (February 2023))ぐ
https://www.state.gov/political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autonomy/

[15] Yual Abraham “Lavender’: The Ai machine directing Israel’s bombing spree in Gaza” +972 Magazine (April 3, 2024)
https://www.972mag.com/lavender-ai-israeli-army-gaza/

international organizations, will have profoundly detrimental effects.

(9) Discussions on the Far-Reaching Impact of AI—Expanding the Debate

Centralization and Decentralization in AI

The history of computers began with the era of large electronic computers called mainframes. Intelligence resided centrally, shared through terminals connected to it. However, with the advent of personal computers, the spread of PCs led to the decentralization of intelligence. Subsequently, the advent of cloud computing, utilizing virtualization and parallel distributed processing technologies, led to a renewed centralization of intelligence. This has further intensified the trend of establishing data centers worldwide. In recent years, as cloud adoption has become even more widespread, edge computing has emerged, making the decentralization of intelligence another major trend.

Thus, the world of computing has cycled between centralization and decentralization. However, centralization and decentralization are not opposing concepts. Considering technological innovation, accompanying changes in cost structures, advances in operational and management techniques, and diverse user needs, the optimal mix between centralization and decentralization has been selected at each point in time.

The world of networks follows a similar pattern. Legacy telephone networks employed a hierarchical tree structure, centrally managed and operated by telephone companies. In contrast, the Internet is a horizontally distributed system where countless routers, deployed by diverse entities scattered across the globe, function through mutual cooperation. Furthermore, the Web3 world, centered around recent blockchain technology, also adopts a horizontally distributed mechanism.

However, this shift is not a one-way movement from centralization to decentralization. The key point of discussion is how to achieve coexistence between centralized and decentralized models and what roles each should fulfill.

The same applies to AI. Large Language Models (LLMs), a centralized model, are not the only option. A decentralized model is also conceivable, where personal AI equipped with intelligence at the edge is connected via a network and virtually integrated for unified operation. Furthermore, by constructing a hybrid model combining centralized training at the core with edge-based inference, it is expected to achieve low latency, improved availability (ensuring resilience), and enhanced privacy protection.

Particularly as the adoption of Physical AI—where AI and actuators like robotics collaborate within Cyber Physical Systems (CPS) that integrate physical and cyber spaces—is anticipated to grow, it is

essential to deepen discussions on the balance between centralization and decentralization in AI promotion, Japan's competitive edge, and standardization strategies.

Furthermore, regional decentralization of data centers must be considered alongside decentralizing electricity demand. The “Watt-Bit Collaboration” initiative, which involves coordinating power infrastructure with data center facilities for distributed deployment, is a critically important policy issue. This effort is expected to contribute not only to decentralizing electricity demand but also to decentralizing the intelligence required for power supply management. Strengthening the coordination between energy policy and digital policy is therefore essential.

AI and National Security

China's AI policy permits only domestic AI that aligns with national political directives under its laws, positioning it as part of the state's cyber sovereignty to control its citizens through AI. In contrast, former Western nations—though divided between the U.S., which prioritizes deregulation, and Europe, which advocates strict regulation through AI legislation—emphasize digital sovereignty rooted in liberal principles. This includes safeguarding freedom of expression, press freedom, and national sovereignty. Japan aligns with this camp. Thus, the approach to AI governance is intrinsically linked to a nation's conception of sovereignty and cannot be separated from its security strategy.

The relevance of AI to security can be considered from three angles: the intensification of cognitive warfare, the escalation of cyberattacks, and the nature of weapon deployment. The intensification of cognitive warfare and the escalation of cyberattacks have already been addressed (see item (3)), so here we focus on the nature of weapon deployment.

While both offensive and defensive sides are actively pursuing AI utilization, its application remains confined to automated fire control and decision support systems. It has not yet reached a stage where it fundamentally alters deterrence strategies in the realm of interstate security. However, discussions must continue regarding the use of AI in the exercise of force, particularly from the perspective of compliance with international humanitarian law (see item (8)).

Digital technologies, including AI, create gray zone situations that blur the distinction between peacetime and wartime. As hybrid warfare progresses, where the boundaries between military and non-military actions are ambiguous, we must recognize that AI governance is a policy issue closely linked to security.

Relatedly, Japan's digital deficit is projected to expand to 6.7 trillion yen by 2024. Unless Japan

promotes its AI industry as an industrial policy, this digital deficit will widen further, becoming a significant destabilizing factor for security. Recognizing that AI plays a central role in establishing "digital sovereignty," it is crucial to adopt an economic security perspective that links industrial policy and security policy concerning AI.

AI and Democracy

By implementing AI into people's communication networks, opportunities for communication will dramatically increase not only between people, but also between people and AI, or even between AI systems themselves. Ordinary human-to-human communication possesses a self-correcting mechanism that forms consensus by seeking points of compromise. Generally, in a democratic world, people change their minds through discussion, and a self-correcting mechanism operates where compromise and new insights lead to a middle ground (consensus) that more people can agree upon.

However, bot algorithms do not change during human debates. This means that in networks with a high bot ratio, the self-correcting mechanism fails to function. Consequently, democratic debate could become ineffective, not only preventing convergence but potentially exacerbating differences in positions (conflict) within the debate. Furthermore, it has been pointed out that introducing specific biases into such algorithms could enable the effective advancement of cognitive warfare.

On the other hand, the potential for digital democracy is increasingly being discussed. This involves leveraging AI to identify trends within vast amounts of opinion through methods like broad listening, enabling decision-making closer to direct democracy from among numerous policy options.

Thus, as AI becomes more widespread, it is essential to continue broad discussions on the impacts it will have on democracy.

Furthermore, related to this, discussions are also needed on the appropriate use of AI in judicial processes. It is necessary to debate what requirements AI must meet to ensure the continued maintenance of a trustworthy judicial process (the rule of law).

Promoting a Comprehensive, Holistic AI Strategy

As seen above, AI policy has now significantly transcended the realm of overcoming technical challenges or promoting the use of digital technologies. It demands a cross-domain, comprehensive

strategy spanning multiple policy areas. As already touched upon, this requires organic coordination among industrial policy, competition policy, foreign policy, security policy, and others.

In Japan, following the enactment of the AI Act, the Basic Plan for Artificial Intelligence was approved by the Cabinet in December 2025. While this initiative, which transcends traditional frameworks, is commendable, many of the measures included in the plan remain confined within the boundaries of existing ministries and agencies. The plan has not yet reached the stage of incorporating descriptions of the organic coordination between measures or the necessity for such coordination.

Particularly in Japan, the utilization of generative AI remains limited to partial applications within companies, and cases driving business transformation are still scarce. Industries implementing AI can build new data-driven business models and enhance industrial competitiveness.

Therefore, when formulating the government's AI strategy, it is essential to develop a comprehensive, overarching AI strategy that incorporates economic security perspectives. This should encompass cutting-edge technology development, semiconductor manufacturing and distribution, language model development, establishing environments for data circulation, and mechanisms for handling intellectual property and copyright rights.

(10) Addressing Ethical Issues

With the rapid advancement of AI, we must also consider the possibility of AI developing a form of self-awareness in the future. Therefore, similar to the life sciences field, we should examine ethical issues surrounding AI research and establish concrete research ethics guidelines and research approval processes. For example, ethical guidelines must be formulated and implemented for issues such as “endowing AI with self-awareness” and “determining the extent to which AI should possess self-replication or modification capabilities.” This issue can also be seen as directly connected to the fundamental and spiritual question for humanity: “What is religion?”

Future Work Plan

As stated at the outset, the fundamental theme of this document is the “controllability of AI technology.” Humans and AI are not opposing, separate entities. We must never forget that AI is ultimately a tool created by humans. This is precisely why AI governance—aiming to establish an environment where humans make the final risk assessments regarding AI's impacts and take

responsibility for them—is crucial. In this sense, discussions surrounding AI span a broad spectrum. It is continuously necessary to assess not only the nature of social and economic governance rules but also the impact AI will have on the very structure of society itself.

With this awareness, DPFJ will continue updating this document based on its content, including through ongoing workshops involving relevant stakeholders. Concurrently, we will deepen broader discussions on AI governance, such as by holding open forums during document update cycles. In doing so, we will actively collaborate with other forums and academic societies engaged in similar discussions to foster consensus.