AI Governance in Japan Ver. 1.0

INTERIM REPORT

January 15, 2021

Expert Group on Architecture for AI Principles to be Practiced

Table of Contents

1. Introduction

There has been discussion around the world on AI[1] governance, or how systems comprising AI as an element, AI services that make the systems available, other related services, and developers, users and service providers[2] should be governed. In Japan, the AI Strategy 2019, revised based on follow-ups this year, and the Integrated Innovation Strategy 2020 are requesting relevant ministries to "discuss ideal approaches to AI governance in Japan, including regulation, standardization, guidelines, and audits, conducive to the competitiveness of Japanese industry and increased social acceptance, for the purpose of operationalizing the AI Principles, taking domestic and international AI trends into account." Similar discussion can also be seen in Europe and the United States, where basic policies about regulations on AI systems and more specific regulations have been discussed and published.[3] Global Partnership on AI (GPAI), launched in June 2020, is an international effort for operationalizing OECD AI Principles.[4]

While the discussion on AI governance[5] is developing in Japan and around the world, it is not easy to design actual AI governance. On one hand, some may think that horizontal regulation can address issues unique to AI such as lack of explainability. On the other hand, solution to the issues can be sector-specific or use-case-specific because AI, versatile technology that can apply to various fields and uses, can raise different issues in each application. Discussion on the design of appropriate monitoring and enforcement mechanisms is also required in order to make the governance effective. We need to structure this complex and multi-layered governance while avoiding hindering innovation as well as addressing concerns regarding AI systems and services. In addition, the development of governance for AI that may be involved horizontally in digital transformation facilitates the utilization of digital technologies in the with/post COVID-19 world. AI governance is an urgent issue that we cannot solve without the knowledge and experience of experts from various fields.

In this regard, AI governance has been discussed by the Expert Group on Architecture for AI Principles to be Practiced, comprising not only experts in the intersection between AI and the

---

[1] This interim report recognizes that Weak AI has currently reached the stage of practical application and uses the term "AI" to mean "Weak AI" and, in particular, the academic discipline (research topic) related to machine learning. See Ministry of Economy, Trade and Industry, "Contract Guidelines on Utilization of AI and Data Version 1.1."

[2] Guided by the definition in "AI Utilization Guidelines" by the Conference toward AI Network Society, the Ministry of Internal Affairs and Communications.

[3] For example, the US federal government and European Commission indicated their following policies. Executive Office of the President, Office of Management and Budget, Memorandum to the Heads of Executive Departments and Agencies, "Guidance for Regulation of Artificial Intelligence Applications" (November 17, 2020). European Commission, White Paper on Artificial Intelligence – A European approach to excellence and trust (February 19, 2020).

[4] METI News Release, "Global Partnership on Artificial Intelligence Founded" (June 16, 2020). https://www.meti.go.jp/english/press/2020/0616_001.html.

[5] In reference to the definition under discussion by the Study Group on a New Governance Model in Society 5.0, this interim report defines AI governance as "design and operation of technological, organizational, and social systems by stakeholders for the purpose of managing risks posed by the use of AI at levels acceptable to stakeholders and maximizing their positive impact."

constitution, civil law, the Act on the Protection of Personal Information, and private and public co-regulation but also an expert on explainable AI, well-experienced executives from AI developers and AI users, and practitioners of insurance and audits for AI. In the discussion, the members have explored multi-layered governance, looking to goal-based governance suggested by the Study Group on a New Governance Model in Society 5.0 (Governance Model Study Group).[6]

Chapter 2 of this interim report discusses trends in AI governance in Japan and around the world. First, the chapter summarizes the chronological development of the AI governance discussion, followed by a summary of the discussion on risks that can serve as the basis for further elaboration. Then, the chapter defines a universal structure comprising goals, intermediate rules, and rules focusing on specific areas such as sectors. Finally, it maps AI principles, horizontal regulations, guidelines, standards, and regulations on specific applications, specific sectors and use by the government in the universal structure with a monitoring and enforcement mechanism.

Chapter 3 discusses ideal approaches to AI governance in Japan by taking domestic and international AI trends into account. First, relying on an approach guided by the Governance Model Study Group and other suggestions, it discusses AI governance generally required in the era of Society 5.0. Then, it discusses ideal approaches to AI governance based on stakeholders' opinions, including those of the members of the Expert Group. Based on the above discussion, it proposes the AI governance architecture that is ideal in Japan at the moment. Finally, it shows some issues that have not been fully discussed by the Expert Group.

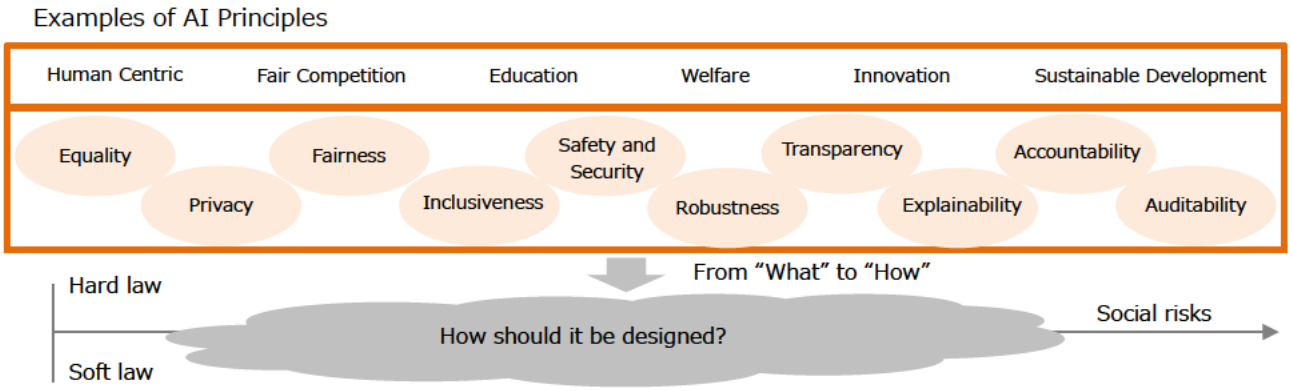
AI governance requires multi-stakeholder engagement and diversified views must be taken into account in the discussion. Although the Expert Group consists of experts with various backgrounds as mentioned above, the Group here discloses this interim report for public opinion to seek further diversity and inclusiveness.

[6] The study group has held discussions on "the need for new governance models to pursue both "promoting innovations" and "achieving social value," as well as ideal approaches thereto as Japanese society faces dramatic changes brought about by big data, IoT, AI and other digital technologies." METI News Release "Report "GOVERNANCE INNOVATION: Redesigning Law and Architecture for Society 5.0" Compiled" (July 13, 2020). https://www.meti.go.jp/english/press/2020/0713_001.html. Hereinafter this Report is referred to as "Governance Innovation Report."

2. AI Governance Trends in Japan and around the World

A. From AI principles to governance

AI principles have been published by not only governments and international organizations/groups but also companies and multi-stakeholders. According to a chronological map on principles for AI by Jessica Field, et al., starting with Tenets of Partnership on AI in September 2016 followed by Asilomar AI Principles and AI principles by Microsoft and Google, AI principles and the like have been published in Japan, Europe and so on.[7] Then, the OECD published AI Principles in May 2019, which multiple countries agreed to for the first time. The OECD AI Principles have five pillars: inclusive growth, sustainable development and well-being; human centric values and fairness; transparency and explainability; robustness, security and safety; and accountability. In June of the same year, the G20 Leaders' Summit adopted AI Principles that draw from the OECD AI Principles in the Annex of the leaders' declaration.



As the discussion on what AI principles are is moving toward a general consensus, it is said that the theme of the discussion is shifting from AI principles to AI governance that operationalizes AI principles in society.[8] In the context of AI governance, which can comprise various elements such as legally binding regulation, non-binding guidelines, and self-regulation, how to mix such elements to spur innovation while preserving basic values is discussed. A goal-based governance framework proposed by the Governance Model Study Group is a good reference for the discussion (see the column on the next page about the report of the Governance Model Study Group).

Please note that the discussion on AI principles is still ongoing. UNESCO is compiling a report on comprehensive AI ethics for a global standard. It considers comments in multi-stakeholder consultation and from member states, and will finalize the report in April 2021 and submit it to UNESCO's General Conference in 2021.[9]

---

[7] Fjeld, Jessica, Nele Achten, Hannah Hilligoss, Adam Nagy, and Madhulika Srikumar, "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI" Berkman Klein Center for Internet & Society (2020).

[8] Arisa Ema, "How to walk in AI society: how to go with AI" (February 28, 2019). Chapter 2, 6-3 (Shift in discussion from what to how). Available only in Japanese.

[9] UNESCO, "Elaboration of a Recommendation on the ethics of artificial intelligence," https://en.unesco.org/artificial-intelligence/ethics#recommendation.

Column: Governance Innovation

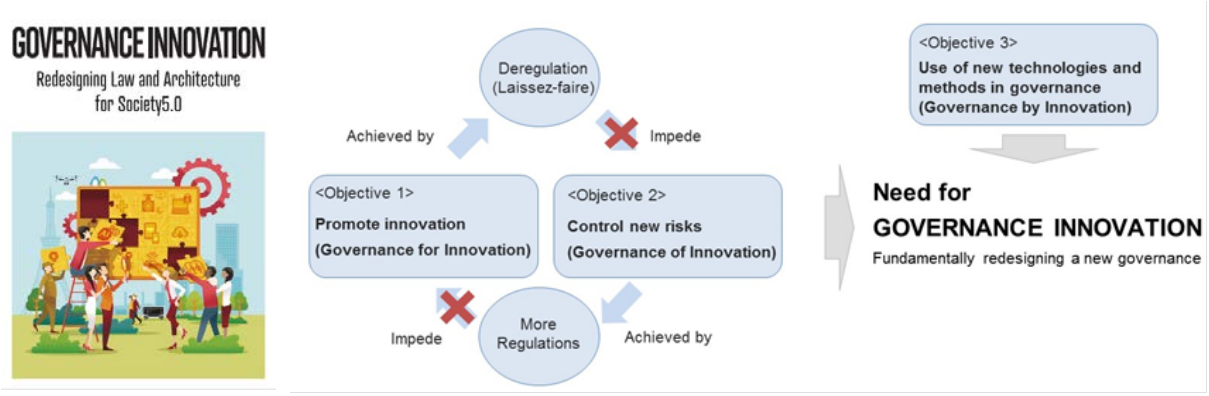
In an era where our society is facing dramatic changes derived from digitalization, the conventional governance models placing laws and regulations at the core face difficulties in keeping up with the speed of innovation. Accordingly, such governance models have, on one hand, been causing problems where laws cannot control new risks that may be brought about by innovations, while, on the other hand, hindering the development of innovations. This problem awareness was supported by the G20 member countries at the G20 Osaka Summit meeting in June 2019. Additionally, the G20 Ministerial Meeting on Trade and Digital Economy declared, under the title "Governance Innovation," that member countries would "strive for innovation-friendly policies and look to remove barriers to innovation accordingly."

Against this background, the Governance Innovation Report stated that Japan should break away from the conventional governance models in which the government plays a leading role in the process from designing to supervising to enforcing rules, and that instead, companies should also take the lead in designing, monitoring and enforcing rules, and proposed that in each process of governance, i.e., rule-making, monitoring and enforcement, governments should ensure the active involvement of businesses that design and implement cyber-physical architectures as well as the communities and individuals that use them.

Rule-making. Shift from rule-based regulations that specify detailed duties of conduct to goal-based regulations that specify value to be attained ultimately, in order to overcome the problem of laws not being able to accommodate the speed and complexity of society. Establish non-binding guidelines and standards with a wide range of stakeholders so that they can achieve the goals. Continuously evaluate the effects and impacts of guidelines/standards, and arrange opportunities for frequent reviews by referring to data collected during monitoring and the claims of parties involved in the enforcement phase. As the information required for governance is concentrated in the private sector (information asymmetry), design an incentive mechanism to promote self-regulation by businesses so that businesses will utilize the information they have in their governance. Oblige or incentivize information disclosure (transparency rules) so that discipline based on market and social norms will work effectively.

Monitoring. Encourage businesses to take innovative approaches to achieving goals provided by laws (compliance), and focus on accountability for their activities (comply and explain). Further, in order to maintain public trust, utilize various forms of assurance depending on the risk, such as self-check, peer review, internal audit, agreed procedures, third-party review and external audit. Consider technologies and mechanisms that enable each stakeholder, such as businesses, the government and individuals, to access real-time data and conduct efficient and effective monitoring. Conduct "monitoring and reviews" by stakeholders on a regular basis, in order to evaluate the results of monitoring among stakeholders, which will lead to the revision of rules and improvement of systems.

* As for enforcement, see the Governance Innovation Report.



B. Risk-based approach

A risk-based approach or idea that the degree of regulatory intervention should be proportionate to the impact of risks for AI governance is very much an international common ground. The European Commission stated in the AI White Paper that "the Commission is of the view that it should follow a risk-based approach" because "the new regulatory framework for AI should be effective to achieve its objectives while not being excessively prescriptive."[10] In the United States, when considering regulation, agencies are to take a risk-based approach and determine which risks are acceptable while considering potential benefits, and they think "it is not necessary to mitigate every foreseeable risk" and do not favor prescriptive regulations.[11]

Industry and consumer protection organizations support a risk-based approach.[12] U.S.-Japan Business Council has stated that "any efforts by the two governments in this area should be mindful of existing rules and regulations, incorporate risk-based approaches to AI governance ..."[13] DIGITALEUROPE has stated that multi-stakeholder discussion, together with agile, evidence and risk-based policy-making should therefore be the foundation of the European Union's AI policy landscape.[14] Orgalim, an industry group representing tech companies in Europe, has stated that "a regulatory approach to AI in Europe must be compatible with a risk-based approach."[15] It seems that BEUC, a consumer protection organization in Europe, does not disagree with a risk-based approach, although they object to the specific risk assessment of the European Commission.[16]

C. Assessment and classification of risks

Although there is a common ground as to the basic idea of a risk-based approach, countries, regions and other stakeholders have not necessarily reached a consensus on the specific risk assessment and classification. For example, it is worth noting that even member states of the EU may have different views on them.[17]

---

[10] See *supra* 3.

[11] See *supra* 3.

[12] Keidanren stated, in their opinion on the AI White Paper by the European Commission, that a risk-based approach is important.

[13] 56th U.S.-Japan Business Conference, "Supplement on Digital Economy of Joint Statement" (September 18, 2019). https://www.uschamber.com/file/25475/download.

[14] DIGITALEUROPE, "DIGITALEUROPE's Recommendations on Artificial Intelligence Policy" (November 13, 2019). https://www.digitaleurope.org/resources/digitaleuropes-recommendations-on-artificial-intelligence-policy/.

[15] Orgalim, "Orgalim Manifesto: a European Agenda on Industrial AI" (January 15, 2020). https://www.orgalim.eu/sites/default/files/attachment/Orgalim%20Manifesto%20for%20a%20European%20Agenda%20on%20Industrial%20AI%2015.01.2020.pdf.

[16] Bureau Européen des Unions de Consommateurs (BEUC), "BEUC'S RESPONSE TO THE EUROPEAN COMMISSION'S WHITE PAPER ON ARTIFICIAL INTELLIGENCE" (June 12, 2020). https://www.beuc.eu/publications/beuc-x-2020-049_response_to_the_ecs_white_paper_on_artificial_intelligence.pdf. "The proposed risk-based approach for the development of the new legal framework on AI and ADM should be revised and broadened", p2.

[17] Position paper on behalf of Denmark, Belgium, the Czech Republic, Finland, France Estonia, Ireland, Latvia,

Some issues are revolving around the assessment and classification of risk. First, how to classify AI risks is discussed and proposed. The AI White Paper by the European Commission takes a binary approach, where, on one hand, high-risk AI should be under legally binding regulations and on the other hand, non-high-risk AI should not be subject to such regulations. In contrast, the Data Ethics Commission, which the German federal government set up and gave a one-year mandate, classified AI risks into five levels and proposed a general regulatory approach to each level.[18] The consumer protection organization in Europe has criticized that such a binary approach as proposed in the AI White Paper would limit the scope of regulation and the regulation would not be effective.[19]

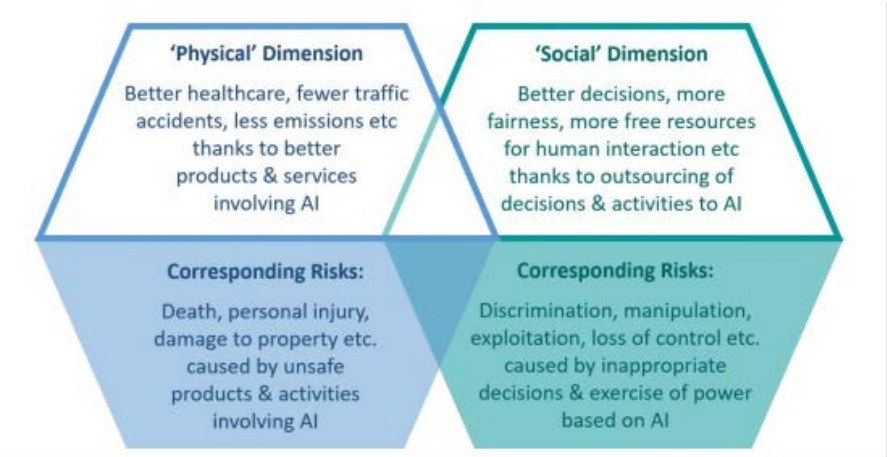


Figure from the opinion of the University of Vienna, et al. cited in footnote 20

Another perspective is classification by usage.[20] The University of Vienna, the University of Freiburg, and the European Law Institute classify AI risks into (1) the physical dimension of risks such as death, injury and damage to property caused by unsafe products and activities and (2) the social dimension of risks such as discrimination, manipulation, exploitation, etc. and general loss of control caused by inappropriate decisions and the exercise of power based on AI and discuss how to regulate them. They propose that a new AI regulation is necessary to address the social dimension of risks, although the physical dimension of risks should be addressed by assessing the applicability of existing regulations, such as the Product Safety Directive, the

---

Luxembourg, the Netherlands, Poland, Portugal, Spain and Sweden, "INNOVATIVE AND TRUSTWORHTY AI: TWO SIDES OF THE SAME COIN" (October 8, 2020). https://em.dk/media/13914/non-paper-innovative-and-trustworthy-ai-two-side-of-the-same-coin.pdf.

[18] The Data Ethics Commission, "Opinion of the Data Ethics Commission" (December 2019). https://www.bmjv.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/Gutachten_DEK_EN_lang.pdf. An outline of five risk levels is as follows: Level 1: Zero or negligible potential for harm → No regulation, Level 2: Some potential for harm → Regulated on an as-needed basis, Level 3: Regular or significant potential for harm → License, Level 4: Serious potential for harm → Oversight and transparency obligations, Level 5: Untenable potential for harm → Ban.

[19] See *supra* 16.

[20] The University of Vienna, the University of Freiburg, and European Law Institute, "Response to the public consultation on the White Paper: On Artificial Intelligence – A European approach to excellence and trust, COM(2020) 65 final" (June 16, 2020). https://www.europeanlawinstitute.eu/fileadmin/user_upload/p_eli/News_page/2020/ELI_Response_AI_White_Paper.pdf.

Product Liability Directive, and sectoral legal instruments and that the existing regulations should adopt the risks by expanding their scope if necessary.

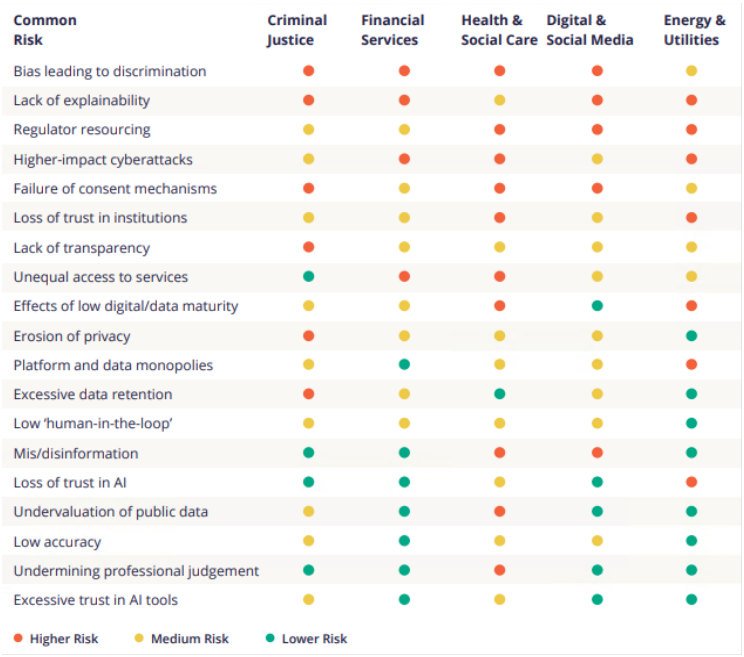


Figure cited from AI Barometer in footnote 22

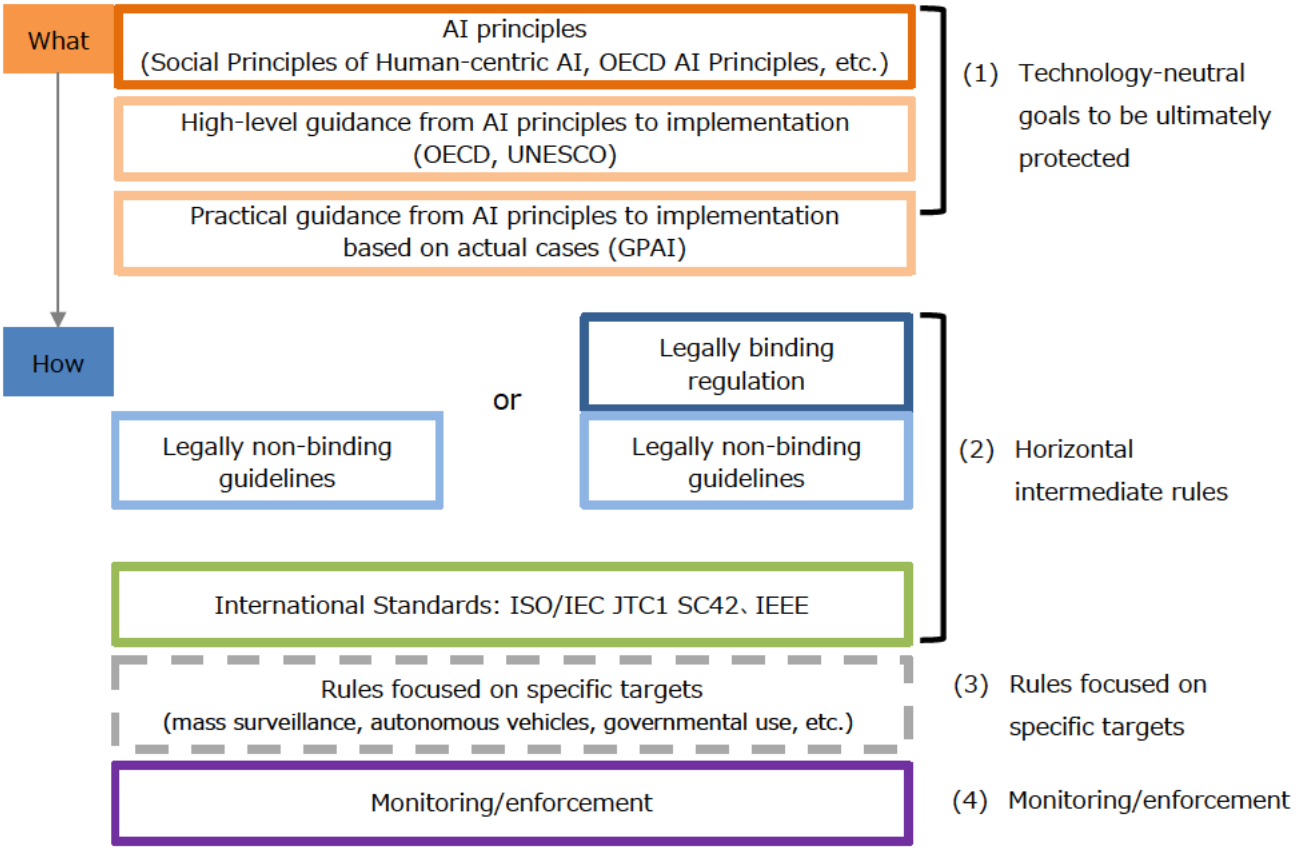
In addition to the abstract and theoretical analyses like the above, there is also case-specific risk analysis. The Center for Data Ethics and Innovation, established in the Department for Digital, Culture, Media and Sport in the UK government[21], published a report titled "AI Barometer."[22] The report evaluates various risks including bias leading to discrimination, lack of explainability, higher-impact cyberattacks, lack of transparency, and erosion of privacy in criminal justice, financial services, health and social care, digital and social media and energy and utilities, and then classifies them into three levels: higher, medium and low risk.

[21] An independent advisory body to connect policymakers, industry, civil society, and the public to develop an appropriate governance regime for data-driven technologies.

[22] UK Center for Data Ethics and Innovation, "AI Barometer" (June 18, 2020). https://www.gov.uk/government/publications/cdei-ai-barometer.

D. Architecture of AI governance

This interim report relies on the universal framework to describe the governance structure in the era of Society 5.0 suggested by the Governance Model Study Group and organizes in layers elements of AI governance, which are found in discussions and proposals in Japan and around the world: (1) technical-neutral goals to be ultimately protected, (2) horizontal intermediate rules such as horizontal regulations and guidelines and international standards, (3) rules focusing on specific targets such as a sector, (4) monitoring and enforcement.



While the discussion on what AI principles are is moving toward a general consensus, what a universal structure of AI governance should look like has not been discussed very much, at least not explicitly, since each country or region has discussed AI governance for operationalizing AI principles mainly in the context of its legal framework. It is our mission to discuss the image of internationally ideal approaches to AI governance and disseminate our idea as Japan did in the discussion of AI principles, while paying full respect to domestic and foreign discussions. It may be natural to adopt such attitude since AI technology can be used beyond national borders.

(1) Goals: AI principles

As mentioned above, the discussion on what AI Principles are is moving toward a general consensus. Japan adopts seven principles in "Social Principles of Human-centric AI": 1. Human-

Centric, 2. Education/Literacy, 3. Privacy Protection, 4. Ensuring Security, 5. Fair Competition[23], 6. Fairness, Accountability, and Transparency, and 7. Innovation. The OECD principles on AI, agreed to among countries including Japan, prescribe 1. Inclusive growth, sustainable development and well-being, 2. Human centric values and fairness, 3. Transparency and explainability, 4. Robustness, security and safety and 5. Accountability.[24] The G20 Trade Ministers and Digital Economy Ministers agreed to G20 AI Principles drawn from the recommendation of OECD council in the Annex of the ministers' declaration, and subsequently, the G20 Leaders' Summit adopted the AI Principles in the Annex of the leaders' declaration. It should be noted here that the High-Level Expert Group on Artificial Intelligence in Europe specified four principles: 1. Respect for human autonomy, 2. Prevention of harm, 3. Fairness and 4. Explainability and proposed seven key requirements: 1. Human agency and oversight, 2. Technical robustness and safety, 3. Privacy and Data governance, 4. Transparency, 5. Diversity, non-discrimination and fairness, 6. Societal and environmental well-being and 7. Accountability.

According to a study that classifies AI principles published by a government, region, international body or group, company and multi-stakeholder, AI principles around the world can be classified into eight themes: privacy, accountability, safety and security, transparency and explainability, fairness and non-discrimination, human control of technology, professional responsibility and promotion of human value[25], although AI principles do not necessarily correspond to each other because how to organize them and whom to ask to respect them are different.[26] Attention must be paid to the global discussion on AI principles when we discuss AI governance because it should be aligned internationally.

(2) Horizontal intermediate rules

(a) Legally non-binding guidelines

Several countries provide measures to encourage respect for AI principles. These measures can be generally divided into two categories: a commentary approach that explains each AI principle and an integration approach that intertwines AI principles with practices by companies and other entities. One of the exemplary measures of a commentary approach is an initiative of the High-Level Expert Group on Artificial Intelligence in Europe. They included the pilot Assessment List for Trustworthy AI for companies and other entities in their report "The Ethics Guidelines for Trustworthy AI" published on April 8, 2019, and then they finalized and published the list on

---

[23] Fair Competition is sometimes listed as an example of the difficulty of responding to principles in Europe and the United States. See *supra* 7, p17.

[24] In OECD's AI Observatory, a rationale is presented for each principle. https://oecd.ai/ai-principles.

[25] Japan's "Social Principles of Human-centric AI" states that "In order for AI to be accepted and properly used by society, we will systemize these basic principles into "Social Principles of AI" to which society (especially state legislative and administrative bodies) should pay attention, and "R&D and Utilization Principles of AI" to which developers and operators engaged in AI R&D and social implementation should pay attention (emphasis added)." Therefore, it can be interpreted that the R&D and Utilization Principles of AI that companies are required to follow do not have to fully cover all seven elements. Especially for Fair Competition and Innovation, it is believed that expectations are placed on state legislative and administrative bodies.

[26] See *supra* 7. Another research called "Linking Artificial Intelligence Principles" also compiles AI principles. http://www.linking-ai-principles.org/.

July 17, 2020.[27] The Assessment List comprises a series of questions or checklists, which are compiled on a requirement by requirement basis, i.e., which are placed under each requirement: human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, societal and environmental well-being and accountability. The Draft AI R&D Guidelines for International Discussions, and AI Utilization Guidelines by the Conference toward AI Network Society, the Ministry of Internal Affairs and Communications, can be categorized as a commentary approach because it explains points to pay attention to on a principle by principle basis.

One of the exemplary measures of an integration approach is the Model Framework of Singapore. The Model Framework briefly explains principles underlying the framework at the beginning[28] and then provides practical issues to be considered using practical examples from four perspectives: internal governance structures and measures, determining the level of human involvement in AI-augmented decision-making, operations management and stakeholder interaction and communication.[29] Principles such as explainability, transparency and fairness are intertwined in the framework. In addition, the self-assessment guide of Singapore compiled together with the World Economic Forum is also organized as a list under the same four perspectives.[30] The UK's guidance, "Explaining decisions made with AI," is another good example of an integration approach.[31] The guidance provides a different guide for Data Protection Officers and compliance teams, technical teams and senior management. The guidance sets forth the following principles: be transparent, be accountable, consider the context you are operating in and reflect on the impact of your AI systems on the individuals affected, as well as wider society.

(b) Legally binding horizontal regulations

Although, as mentioned later, some have focused regulation on a specific use of AI, no government, including a regional one, has adopted a legally binding horizontal regulation covering a certain group with respect to AI defined with criteria other than a specific use of AI. While governments around the world appear to be cautious about such a horizontal regulation, the European Commission is considering it. After publishing the AI White Paper in February

---

[27] The High-Level Expert Group on AI (AI HLEG), "THE ASSESSMENT LIST FOR TRUSTWORTHY ARTIFICIAL INTELLIGENCE (ALTAI) for self assessment" (July 17, 2020). https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment.

[28] The two principles underlying the Model Framework are as follows: a. Organizations using AI for decision making should make the decision-making process explainable, transparent, and fair. b. AI-driven solutions should be human-centric.

[29] Info-communications Media Development Authority and Personal Data Protection Commission, "Model Artificial Intelligence Governance Framework Second Edition" (January 21, 2020). https://www.pdpc.gov.sg/Help-and-Resources/2020/01/Model-AI-Governance-Framework.

[30] World Economic Forum, "Companion to the Model AI Governance Framework –Implementation and Self-Assessment Guide for Organizations" (January 2020). https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgisago.pdf.

[31] UK ICO and The Alan Turing Institute, "Explaining decisions made with AI" (May 20, 2020). https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-ai/.

2020, the commission proposed the following options in the Inception Impact Assessment published in July of the same year: 1. soft law, 2. voluntary labelling, 3.a mandatory requirements on remote biometric identification systems, 3.b mandatory requirements on "high-risk AI" applications, 3.c mandatory requirements on all AI applications, 4. combination of any of the options above. If proposals 3.a to 3.c are adopted, a certain set of requirements, for example, training data, record-keeping, information provision, robustness and accuracy and human oversight illustrated in the White Paper, can be imposed on designated AI applications.[32]

It should be noted that it appears as though EU member states as well as stakeholders have not fully reached a consensus about the legally binding horizontal regulation. 14 countries, comprising Denmark, Belgium, the Czech Republic, Finland, France, Estonia, Ireland, Latvia, Luxembourg, the Netherlands, Poland, Portugal, Spain and Sweden, have stated that we should turn to soft law solutions.[33]

(c) International Standards

SC42 (subcommittee 42) was established in October 2017 under ISO/IEC JTC1 (Joint Technical Committee 1), a standardization body in the information technology field, to discuss international standards on AI. SC42, with various breakout WGs including a trustworthiness WG, is making steady progress with some milestones such as six plenaries. In Japan, the Information Technology Standards Commission of Japan in the Information Processing Society of Japan established the Technical Committee for SC42, which hears opinions and organizes them in Japan and represents Japan in international arenas. As one of the driving forces, National Institute of Advanced Industrial Science and Technology (AIST) leads initiatives regarding SC42 as well as AI research and development.

SC42 has 6 working groups (WGs): AI governance, foundational standards, data, trustworthiness, use cases and applications, and computational approaches and computational characteristics of AI systems, in two of which Japan chairs discussions as a convener. In addition to the WGs, some ad-hoc groups such as AI management, quality of big data and AI lifecycle have been set up. Japan is leading the discussion on an AI quality evaluation method for trustworthiness, and the Machine Learning Quality Management Guideline published by AIST in June 2020 is understood as a fundamental milestone for a proposal for the quality evaluation agenda.

The Technical Committee for SC42, which is in charge of the standardization of AI, reinforces cooperation with CEN/CENELEC, the EU's standardization bodies, in addition to contributing to ISO/IEC JTC 1. Both European and Japanese standardization bodies set themes, discussed them, and held in September 2020 the EU-Japan Workshop on Trustworthy AI Standardization and

[32] GDPR Article 13 (2)(f) stipulates that the controller should, at the time when personal data are obtained, provide the data subject with information necessary to ensure fair and transparent processing regarding the existence of automated decision-making. The requirements for information provision here seem to be taking into consideration these provisions. See *supra* 3, European Commission, White Paper on Artificial Intelligence, p20.

[33] See *supra* 17.

R&D for government officers responsible for AI policy and standards.

ISO/IEC JTC 1 is not the only forum for AI standardization. IEEE is also discussing standards on AI, especially issues relevant to AI ethics. It published "Ethically Aligned Design: Prioritizing Human Wellbeing with Autonomous and Intelligent Systems (Version 1)" in December 2016.[34] Based on the vision, IEEE is creating the IEEE P7000 series of standards. Their themes include Model Process for Addressing Ethical Concerns During System Design, Transparency of Autonomous Systems, Data Privacy Process and Algorithmic Bias Considerations. A Japanese expert serves as Secretary of Transparency of Autonomous Systems.

(3) Rules focused on specific targets

(a) Regulations on specific use

Regulations on a specific use case utilizing personal identification and profiling capability enabled by AI have been discussed or introduced in Europe and the United States. Some examples of remote biometric identification and application to the hiring process are provided in this subsection.

The European Commission indicated in their AI White Paper and Inception Impact Assessment that they would possibly introduce regulation on remote biometric identification using AI. The AI White Paper deems "AI applications for the purposes of remote biometric identification and other intrusive surveillance technologies" as high-risk applications for European fundamental rights.[35] The paper explains that it is necessary to deliberate in what case the GDPR's public interest exception, which is one of the exceptions that permit personal identification with biometric information, applies and other relevant issues. In the United States, relevant bills were introduced to congress at the federal level to oblige companies to first obtain use consent in using any facial recognition technology commercially (Commercial Facial Recognition Privacy Act of 2019) and to prohibit the use of remote biometric recognition including facial recognition tools (Facial Recognition and Biometric Technology Moratorium Act of 2020).[36] Municipal governments such as San Francisco established an ordinance to prohibit the use of facial recognition technologies by the police to address concerns in society about privacy invasion, for example, the Stop Secret Surveillance Ordinance in San Francisco. The state of Washington introduced regulation on facial recognition.[37] Partially against this background, major

[34] IEEE, "Ethically Aligned Design: Prioritizing Human Wellbeing with Autonomous and Intelligent Systems" (December 2016). https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e.pdf. Then, Version 2 was published in December 2017.

[35] See *supra* 3, European Commission, White Paper on Artificial Intelligence, p18.

[36] The Security Industry Association (SIA) has voiced its strong objection to the introduction of the Moratorium Act above. SIA, "Security Industry Association Strongly Opposes the Facial Recognition and Biometric Technology Moratorium Act Citing Immeasurable Benefits of the Proven Technology" (June 26, 2020). https://www.securityindustry.org/2020/06/26/security-industry-association-strongly-opposes-the-facial-recognition-and-biometric-technology-moratorium-act-citing-immeasurable-benefits-of-the-proven-technology/.

[37] Reuters, "Washington State signs facial recognition curbs into law; critics want ban" (April 1, 2020), https://www.reuters.com/article/us-washington-tech-idUSKBN21I3AS.

technology companies such as IBM, Amazon and Microsoft stopped providing the police with facial recognition systems.

The state of Illinois introduced regulation on interviews using AI, the Artificial Intelligence Video Interview Act. Under the regulation, there is an obligation to give notice and obtain an interviewee's consent prior to the interview if AI technology is to be used with respect to the video recording of the interview. At the federal level, the Algorithmic Accountability Act of 2019 was introduced to congress, which would authorize the FTC to require AI operators to conduct impact assessment to address the issue of bias. In New York City, a bill was introduced to the council, which would regulate the sale of automated employment decision tools that filter employment candidates.

(b) Regulations on specific sector

Vertical rules are discussed or introduced to address issues arising from the use of AI in a sector where regulations have been developed by taking into account circumstances unique to each sector. In the automotive sector, the Road Traffic Act and Road Transport Vehicle Act were partially revised[38] and safety standards on automatic operation equipment were prescribed to address issues required for autonomous vehicles at SAE level 3.[39] On June 24, 2020, the UNECE World Forum for Harmonization of Vehicle Regulations (WP29), a forum for international regulations on autonomous vehicles, finalized some international regulation on the lane keeping feature of autonomous vehicle systems operating on a highway under the condition of traffic congestion slower than 60 km/h and so on.[40] In the United States, states set requirements for the testing and deployment of autonomous vehicles[41] and the federal government gives guidance to the states.[42] The European Commission noted in the AI White Paper that the transportation sector could be subject to regulation because it can be categorized as high-risk.

The health care sector can be high-risk, according to the AI White Paper by the European Commission, and some requirements could be imposed on it. In Japan, no person except a medical practitioner shall engage in medical diagnosis and treatment under Article 17 of Medical Practitioners' Act, which requires a medical practitioner, who plays a responsible role in medical diagnosis and treatment, to be responsible for the final decision even if AI is used in the medical

---

[38] As for the Road Traffic Act, refer to the National Police Agency's website related to autonomous vehicles. https://www.npa.go.jp/bureau/traffic/selfdriving/index.html. Available only in Japanese.

[39] Ministry of Land, Infrastructure, Transport and Tourism: "Safety standards for autonomous vehicles have been developed along with the design of autonomous vehicle stickers" (March 31, 2020). https://www.mlit.go.jp/report/press/jidosha07_hh_000338.html. Available only in Japanese.

[40] Ministry of Land, Infrastructure, Transport and Tourism: "Enactment of the first international standards for automatic operation equipment (level 3)" (June 25, 2020). https://www.mlit.go.jp/report/press/jidosha07_hh_000343.html. Available only in Japanese.

[41] California DMV, AUTONOMOUS VEHICLES, https://www.dmv.ca.gov/portal/vehicle-industry-services/autonomous-vehicles/.

[42] US Department of Transportation, USDOT Automated Vehicles Activities, https://www.transportation.gov/AV.

care.[43] [44] The Department of Health and Social Care of the UK government published a code of conduct for data-driven health care technology.[45]

The government of Singapore published Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector to complement the horizontal Model Framework.[46]

(c) Regulations on Use of AI by Government

Regulations and guidelines on the use of AI by the government have been proposed and/or disclosed. In the AI White Paper, the European Commission illustrates areas such as asylum, migration and border controls that could be categorized as high-risk applications and could be regulated. The UK government published the Data Ethics Framework for appropriate and responsible data use in government and other areas of the public sector in June 2018[47] and then, for the purpose of complementing the Data Ethics Framework, the Alan Turing Institute published Understanding Artificial Intelligence Ethics and Safety for the public sector.[48] In the same month, the UK government published a guide to using artificial intelligence in the public sector, which comprises a set of guidance on understanding AI, assessing whether AI is the right solution, planning and preparing for AI implementation and managing AI projects for the public sector.[49] On top of these efforts, the UK government prescribed the Guideline for AI

---

[43] The notification by Director of Health Policy Bureau, Ministry of Health, Labour and Welfare (Health Policy Publication 1219 No.1 (December 19, 2018)) states that "even if medical care is provided using an artificial intelligence (AI)-based program to support diagnosis/treatment, the main actor making diagnosis, providing care, etc. is doctors and doctors are to assume responsibilities for the final judgment. Due consideration should be given to the fact that the said treatment is to be conducted as medical practice defined in Article 17 of the Medical Practitioners Act (Act No. 201 of 1948)." Available only in Japanese.

[44] For evaluation indicators related to AI-based medical image diagnosis support systems and other efforts, detailed information is provided in the materials posted on the following website. https://www.mhlw.go.jp/content/10601000/000590652.pdf. Available only in Japanese.

[45] UK Department of Health and Social Care, "Code of conduct for data-driven health and care technology" (18 July 2019). https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology. It includes not only elements for ensuring equality, transparency, and accountability related to data usage but also elements for understanding users and deciding results and how to use technologies to obtain the results. Rather than AI principles, it is regarded as being closer to a management guide.

[46] Monetary Authority of Singapore, Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector (November 12, 2018). Paragraph 1.4 has been updated to reflect the details of the Model Framework announced on February 7, 2019. https://www.mas.gov.sg/-/media/MAS/News-and-Publications/Monographs-and-Information-Papers/FEAT-Principles-Updated-7-Feb-19.pdf.

[47] The Government Digital Service, "Data Ethics Framework" (June 13, 2018). https://www.gov.uk/government/publications/data-ethics-framework.

[48] The Alan Turning Institute, "Understanding artificial intelligence ethics and safety" (June 10, 2019). https://www.turing.ac.uk/research/publications/understanding-artificial-intelligence-ethics-and-safety.

[49] UK Government Digital Service and Office for Artificial Intelligence, "A guide to using artificial intelligence in the public sector" (June 10, 2019). https://www.gov.uk/government/collections/a-guide-to-using-artificial-intelligence-in-the-public-sector.

Procurement.[50] Looking to further prospective uses of AI, Canada stipulated Directive on Automated Decision-Making and made it effective in April 2020 to ensure that Automated Decision Systems are deployed in a manner that reduces risks to Canadians and federal institutions, and lead to more efficient, accurate, consistent, and interpretable decisions made pursuant to Canadian law.[51] The Directive requires algorithmic impact assessment, ensuring transparency, quality assurance, providing opportunities for objection, and reporting on the effectiveness. In Japan, Advisors to Government Chief Information Officer, et al. have discussed issues to be considered and basic approaches required in AI systems, when AI is used in the government information systems or services by the government, for example, the transparency of training data, elimination of bias, history of data processing and issue of rights, those issues being taken into account in light of the level of risks.[52] There are some initiatives by non-governmental organizations. For example, AI Now Institute provides a practical framework on algorithmic impact assessment for public organizations[53] and New Zealand Law Foundation released a report about governmental use of predictive modeling using AI.[54]

(4) Monitoring/enforcement

(a) Monitoring

The European Commission stated in the AI White Paper that adequate documentation should be required so that an ex-post monitoring authority can monitor high-risk AI applications.[55] This is a proposal for incentivizing a company to abide by legally binding regulation by monitoring its compliance and imposing sanctions if it does not comply with the regulation. On the other hand, Governance Model Study Group proposes a mechanism to facilitate a company to explain that they make efforts towards goals, to encourage them to improve their efforts through feedback from stakeholders including the government and reinforce mutual trust among companies, users and the government in society.

---

[50] UK Office for Artificial Intelligence, "Guidelines for AI Procurement" (June 8, 2020). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/890699/Guidelines_for_AI_procurement__Print_version_.pdf.

[51] The Government of Canada, "Directive on Automated Decision-Making" (February 5, 2019). https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592.

[52] Advisors to Government Chief Information Officer: Kenzaburo Tamaru, Hisafumi Mitsushio, Takeshi Nishimura, Akihiro Umegai, Masanori Kusunoki, Tsutomu Hosokawa; Representative Director and CEO of Ridge-i Inc., Takashi Yanagihara; Senior Research Fellow at Research Study Division, Institute for Information and Communications Policy, Ministry of Internal Affairs and Communications, Koichi Takagi; "Characteristics of data use and key points in handling AI-based systems" (June 2020). https://cio.go.jp/dp2020_01. Available only in Japanese.

[53] AI Now Institute, "ALGORITHMIC IMPACT ASSESSMENTS: A PRACTICAL FRAMEWORK FOR PUBLIC AGENCY ACCOUNTABILITY" (April 2018). https://ainowinstitute.org/aiareport2018.pdf.

[54] New Zealand Law Foundation, "GOVERNMENT USE of ARTIFICIAL INTELLIGENCE in NEW ZEALAND" (2019). https://www.data.govt.nz/assets/data-ethics/algorithm/NZLF-report.pdf.

[55] See *supra* 3, European Commission, White Paper on Artificial Intelligence, F. COMPLIANCE AND ENFORCEMENT, pp23-24.

(b) Enforcement

In the AI White Paper of the European Commission, prior conformity assessment is shown as an example of enforcement if high-risk AI applications are regulated. It is mentioned that the prior conformity assessment may include test, inspection, and authentication procedures and when establishing these procedures, the existing conformity assessment is to be used as a basis. The White Paper also states that "effective judicial redress for parties negatively affected by AI systems should be ensured," and Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics published on the same day elaborates on the issues.[56] This aims at ensuring what is required by law through governmental power.

On the other hand, Governance Model Study Group emphasizes that "an environment should be created in which a penalty is imposed that creates sufficient incentive for businesses to comply with regulations, taking into account the impact of the behavior on society and the extent of the risk." The study group points out that it is rather appropriate to ask a company to do its best to reduce risk arising from an AI system and encourage it not to hesitate to conduct R&D if it cannot avoid risk due to uncertainty of the AI system even if it makes its best effort, because additional sanctions would not give it an incentive to avoid the risk. The group shows examples where "an incident investigation committee may be established that focuses on the investigation of the cause and prevention measures based on the identified cause" and "when such incident investigation committee is established and conducts an investigation, the committee could compel related parties and businesses to submit necessary information by adopting systems of deferred prosecution/suspension of an indictment."

E. International harmonization and alignment between layers

AI governance construction is in progress globally in parallel at multiple layers of AI governance architecture, for example, high-level discussion at GPAI, OECD and UNESCO and mid-level rule-making such as guidelines and AI standards. As described in the AI Strategy 2019, revised based on follow-ups this year, and the Integrated Innovation Strategy 2020, it is required to "discuss ideal approaches to AI governance in Japan, including regulation, standardization, guidelines, and audits, conducive to the competitiveness of Japanese industry and increased social acceptance, for the purpose of operationalizing the AI Principles, taking domestic and international AI trends into account (emphasis added);" in other words, it is required to discuss AI governance, taking international harmonization into account. In designing AI governance, alignment between layers such as AI principles, intermediate rules, and AI standards is also necessary because they are closely related to each other.

An interesting trend can be found among P4 countries, which are known to be a basis of TPP. New Zealand, Chile and Singapore signed the Digital Economy Partnership Agreement (DEPA)

[56] The European Parliament has made a recommendation on civil liability to the European Commission. European Parliament resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence (2020/2014(INL)). https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_EN.pdf. The European Commission is expected to consider this recommendation in proposing legislation but is not obligated to accept it.

in June 2020[57]. This agreement will complement the WTO negotiations on e-commerce and build on the digital economy work underway within APEC, the OECD and other international forums. Article 8.2 deals with AI, sub article 3. of which encourages the parties to promote the adoption of ethical and governance frameworks that support the trusted, safe and responsible use of AI technologies (AI Governance Framework). Australia and Singapore signed the Digital Economy Agreement in August 2020[58]. While this agreement will upgrade the digital trade arrangements between Australia and Singapore under the Comprehensive and Progressive Agreement on the Trans-Pacific Partnership and the Singapore-Australia Free Trade Agreement, it was signed with MoU on Cooperation on Artificial Intelligence, one of the purposes of which is to support the development and adoption of ethical governance frameworks for the trusted, safe and responsible development and use of AI technologies. Although it is too early to tell how significant these digital economic agreements and the AI articles and the MoU are at this point, it can be said that they are a different form of international cooperation from a multi-lateral framework like GPAI and OECD and bilateral dialogue.

[57] New Zealand Foreign Affairs & Trade, Overview of the Digital Economy Partnership Agreement is a new initiative with Chile and Singapore, https://www.mfat.govt.nz/en/trade/free-trade-agreements/free-trade-agreements-concluded-but-not-in-force/digital-economy-partnership-agreement/overview/.

[58] Australian Government, Department of Foreign Affairs and Trade, Australia-Singapore Digital Economy Agreement, https://www.dfat.gov.au/trade/services-and-digital-trade/Pages/australia-and-singapore-digital-economy-agreement.

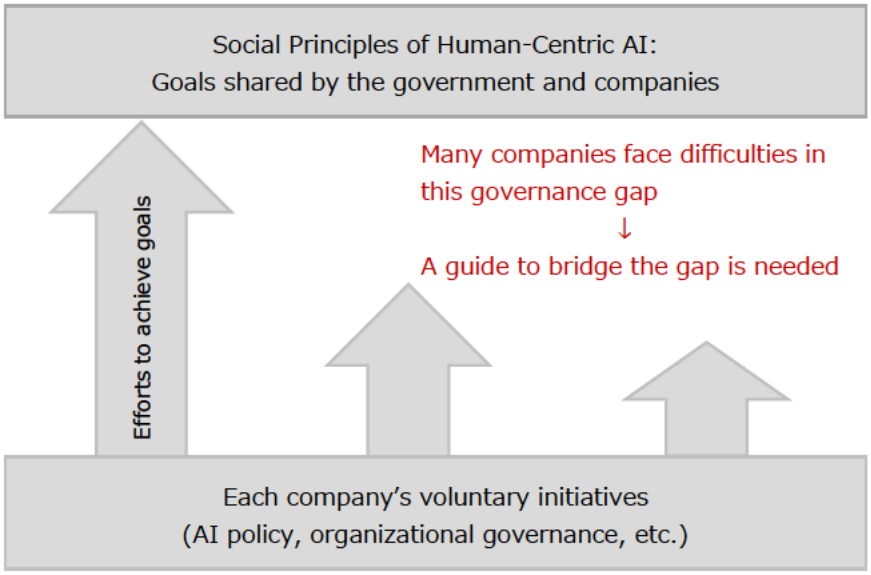
3. Ideal approaches to AI Governance in Japan

A. Suggestion of Governance Innovation

On one hand, laws and regulations face difficulties in keeping up with the speed and complexity of AI innovation and deployment. On the other hand, prescriptive regulation or rule-based regulation can hinder innovation. To address these conflicting problems, it is necessary to change governance models from conventional rule-based ones to goal-based ones that can guide entities such as companies to the value to be attained.[59] Because our society shares the Social Principles of Human-Centric AI, which state the goals for the use of AI, and because principles on AI are slowly but steadily reaching a consensus globally, it can be said that we are finalizing the building of a foundation for goal-based governance.

On the other hand, one of the issues of goal-based regulations is that they could create a big gap between goals shared by society and operations to achieve the goals at the corporate level[60]. In fact, members of an industry group for AI business promotion tried to implement AI principles in in-house rules for operations and then concluded that they need intermediate and practical guidelines to embody the AI principles in in-house rules.

In order to address this issue, it is ideal to set an intermediate rule like a guideline with multi-stakeholders[61]. At the same time, we need to take note of potential harm to innovation in the case where the intermediate rule is operated in a legally binding manner like regulations. The above suggestions and notes point to non-binding intermediate goal-based guidelines to promote AI innovation and deployment.



[59] Governance Innovation Report 5.1.1
[60] Governance Innovation Report 5.1.1
[61] Governance Innovation Report 5.1.2

B. Opinions of stakeholders

It is indispensable for multi-stakeholders to have a dialogue for cooperation in discussing AI governance.[62] Since the subject of AI governance has been discussed domestically and internationally, our discussion must be put in the context of prior discussions around the world. For example, in addition to opinions we heard from the Expert Group and on other occasions, opinions from companies, industry groups and other entities of Japan and other countries on the AI White Paper and Inception Impact Assessment for AI regulation of the European Commission should be taken into account. We also need to look to discussions based on consumer opinions in Japan at the AI Working Group under the Committee on Challenges to Consumer Digitalization of Consumer Affairs Agency.

(1) Opinions of industries (opinions provided to the European Commission)

The European Commission published the Inception Impact Assessment for AI regulation in July 2020, in which the commission proposed the following options: 1. soft law, 2. voluntary labelling, 3.a mandatory requirements on remote biometric identification systems, 3.b mandatory requirements on high-risk AI applications, 3.c mandatory requirements on all AI applications, and 4. combination of any of the options above. Public comments on the inception impact assessment are excellent materials to understand the opinions of industries on AI governance.[63] [64]

First, it sounds as though industries generally agree that some regulation, including soft law, is necessary for AI applications. A comment that "AI is too important not to be regulated – the only question is how to approach it successfully" describes industry's general attitude toward AI regulation.

Many industry groups and companies generally agree with soft law. Many of the industry groups and companies that agree with soft law have commented on how to realize it. One of the proposals is that soft law should support self-regulation by industries. Some mentioned private and public co-regulation. The other proposal is that regulators should take diverse applications of AI into account. Some opine that regulators should support the development of guidance by industry or by sector by respecting their initiatives and expertise.

Not many agree with a voluntary labelling scheme. Some negative comments sound modest, like it seems premature and it is likely to create a heavy administrative burden, and others sound harsh, like "labelling systems would not add value." There is a company that expresses concern over using the Assessment List for Trustworthy AI (ALTAI) as a checklist for a labelling scheme because the list does not take much variation across applications into account.

---

[62] Cabinet Office "Social Principles of Human-Centric AI" (March 2019) P.2.

[63] The progress of discussions after public comments on the AI White Paper in February 2020 is deemed to be reflected in the public comments on the inception impact assessment of AI legislation in July 2020.

[64] European Commission, Feedback received on: Artificial intelligence – ethical and legal requirements. https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements/feedback?p_id=8242911.

Many opine that the scope of regulation should be carefully defined in introducing regulation for high-risk AI applications, although they theoretically agree with the idea of regulation. There is a company that argues that not only risk but also opportunity cost should be considered in assessing risk. Some argue that the scope should be limited to certain applications, for example, an application that substantially replaces human decision making with AI. Among binding regulation options, regulation on remote biometric recognition seems to capture generally favorable comments, although they come with some concerns that rules would need to be clear and avoid being overly prescriptive. It may be natural that there are no favorable comments on the proposal for legally binding regulations on all AI applications.

Some favor a combination of multiple approaches to regulation. For example, it seems like common sense to ensure proportionality according to risk level by allowing a combination of approaches because potential risks are different from one AI application to another.

Some argue that industry specific contexts should be taken into account in AI application to autonomous vehicles and medical equipment. European Automobile Manufacturers' Association argues that regulators should leave regulatory discussions to the automotive industry because the industry has already been rigidly regulated, for example, by ex-ante type approval, and because the international discussion on AI application in UNECE/WP29 has started. MedTech Europe has stated that no new regulations are required because Medical Devise Regulation and CE Marking, which are rigid regulations, apply even to AI application to medical equipment.

(2) Opinions heard from the Expert Group and on other occasions

A company and industry group point out that an intermediate guideline is necessary to bridge the gap between AI principles and corporate-level practice mentioned in 3. A.[65] At the same time, it is argued that it is not wise to make the intermediate guideline function as a checklist. Some are concerned that they would have to accept the list as a de facto obligation even if it comes with a note that it is legally non-binding intermediate guidance and that it would result in increasing administrative cost. Members pointed out that an intermediate guideline should rather have contents to support the development of processes and organizational structures to achieve goals. The background to this is the awareness that even if companies establish AI policies, that in itself does not facilitate behavior that is in accordance with AI principles.

Others opine that it is important to build a common understanding between companies. It is rare for the development and operation of an AI system, including the provision of data, to be completed by a single company. Therefore, companies are required to share views about the process of AI system development and operation. In this regard, the Ministry of Economy, Trade and Industry published "Contract Guidelines on Utilization of AI and Data," which organizes

[65] With regard to "Fair Competition" in Japan's "Social Principles of Human-centric AI," other efforts through laws and regulations, etc. are being made to correct unfair competition. While discussions on AI governance in the context of AI ethics are requested in the Integrated Innovation Strategy 2020, etc., the Expert Group primarily discusses governance designed to supplement the low explainability of AI itself and pays close attention to the promotion of information disclosure related to AI-based services only as a related field. Examples of other countries to which the latter applies include: Guidelines on ranking transparency pursuant to Regulation (EU) 2019/1150 (December 8, 2020).

major problems and issues, exemplary clauses in contracts, issues to be considered in drafting clauses regarding a contract for software development and the use of software, but it does not explicitly deal with the quality and management issues of AI. It is said that an AI developer is sometimes required to assure and/or quantify the quality of an AI system itself, despite the difficulty of doing so due to the black box issue of quality evaluation. Quantifying the quality of AI is an important issue, but the quantification is not the only element for quality assurance. It can be said that it is important for companies to have shared understanding that a broader sense of quality or trust in AI systems needs to be reinforced by other elements such as corporate governance and dialogue among stakeholders. It is argued that corporate governance in Japan still values the interests of stakeholders such as employees, business partners and local society, although it is focusing more on stockholders recently, which guides us to the idea that it is necessary to support initiatives to ensure trust in AI systems through the entire supply chain with an intermediate guideline by making the most of these characteristics.[66]

It is argued that it is necessary to pay attention to differences between BtoB and BtoC companies.[67] Even if a BtoB company provides an AI product that meets the requirements of a customer company, inappropriate operation of the product might have an adverse effect on others. It would be too harsh to ask the BtoB company to avoid such risks. Therefore, some argue that one-size-fits-all guidance requiring the same would not be appropriate.[68] This argument sounds similar to the one mentioned in 3. B. (1) that the ALTAI does not necessarily pay attention to the differences in AI applications.

It is also argued that intellectual property needs to be respected. The AI White Paper by the European Commission lists the legal requirements of keeping records and data and making them available upon request by competent authorities. Although the white paper does not explicitly mention the production of trade secrets, it is better to guide a company to meet the expectation of accountability by urging a company to provide a voluntary explanation rather than asking it

---

[66] Soichiro Kozuka “AI Age and Law” Iwanami Shinsho (November 20, 2019) P.214- for an explanation about the need to explore the direction that was argued for by Professor Simon Deakin of the University of Cambridge as “Japanese companies’ style to operate business with due consideration to the interests of stockholders and employees while paying attention to relationships with society may be rather suited for the age of technological innovation.” Simon Deakin “Corporate Governance, Enterprise Reform and Digitisation” NBL 1153 (September 1, 2019) P.41-. Similar opinions were expressed by the Expert Group. The cited documents are available only in Japanese.

[67] Takashi Matsumoto, Arisa Ema “RCModel, a Risk Chain Model for Risk Reduction in AI Services” (July 6, 2020). “If all processes such as data acquisition, AI model development, and service delivery have been conducted by one company like Google and Facebook, companies can take an optimal approach to all risk factors. However, in Japan, there are many cases in which service providers, developers of AI models, providers of execution environments, and data providers are different. Therefore, in order for AI service providers to comprehensively examine all risk factors, a framework for dialogue among relevant stakeholders is necessary.” The article was available online first in Japanese on June 4, 2020.

[68] For example, some have pointed out that the questions “Could the AI system have a negative impact on society at large or democracy? Did you assess the societal impact of the AI system’s use beyond the (end-) user and subject, such as potentially indirectly affected stakeholders or society at large?” in THE ASSESSMENT LIST FOR TRUSTWORTHY ARTIFICIAL INTELLIGENCE (ALTAI) are inappropriate for BtoB companies since they depend on the uses of end users or BtoB customers. Since the use of this list is voluntary, it is reasonable to understand that it is likely that consideration of unrelated items is assumed to be omitted.

to submit trade secrets such as source codes and algorithms.[69]

(3) Consumer Perspective

Consumers need more information about AI. Statistical research has revealed such characteristics of consumers. According to a survey conducted by Fujitsu Research Institute in 2017 regarding expectations and concerns about robots and artificial intelligence (AI) by gender/generation, the number of respondents who chose "I still don't have any specific image" was the highest. According to consumer opinions about AI disclosed by the European Consumer Organisation (BEUC), 21% of consumers either have not heard about AI or do not understand what AI is actually used for and 43% of them feel that consumers are not provided with sufficient information[70]. The Consumer Affairs Agency's analysis that consumers (AI service users) may not be fully aware of the specific details of technologies, such as robots and AI, and the status of their introduction, accurately describes the current situation[71].

In addition, consumers feel insecure about AI. An item related to concerns: "Concerned about potential damage caused by system errors" was ranked high in the above survey by Fujitsu Research Institute. In addition, opinions were expressed in BEUC's survey that information may be manipulated or personal information may not be used appropriately.

Meanwhile, consumers also express their expectations about AI. In the above survey by Fujitsu Research Institute, relatively high expectations are placed on areas where wide spread use of services can be easily imagined, such as healthcare/nursing care and automatic translation. In BECU's survey, "AI is useful for predicting traffic accidents and disease" garnered a high percentage of responses. The Consumer Affairs Agency's analysis that expectations and concerns are mixed is very much to the point.

Under these circumstances, the Consumer Affairs Agency has concluded that they expect appropriate responses from business providers and that it is desirable for AI to be used smartly

---

[69] Examples of publicly disclosed opinions about concerns over the forced disclosure of trade secrets are as follows: Keidanren "Comments on 'White Paper on Artificial Intelligence: a European approach to excellence and trust' by European Commission" (June 10, 2020). "When the authorities implement prior conformity assessment, they should not unduly seek the disclosure of information that could be sources of competitiveness (e.g., algorithms, details of data sets). Disclosure should not include confidential information and must be limited to the minimum required, and the reason for seeking such disclosure must be clarified." JEITA Position Paper on European Commission's "White Paper on Artificial Intelligence: a European approach to excellence and trust" (June 13, 2020). "We believe that source codes and algorithms, a source of competitive strength for products and services, should not be required to be submitted by authorities. If necessary, clear reasons should be given, required information should be limited to the minimum extent possible, and sufficient consideration should be given to the handling of confidential information."

[70] See *supra* 16.

[71] The Consumer Affairs Agency "The Committee on Challenges to Consumer Digitalization, First AI Working Group Material 4: Current Status of AI" (January 31, 2020). https://www.caa.go.jp/policies/policy/consumer_policy/meeting_materials/assets/review_meeting_004_200206_0007.pdf. Available only in Japanese.

by improving customer literacy. In order to achieve this purpose, the agency released “Handbook on Use of AI—Keys to Effectively Using AI” in July 2020 (available only in Japanese). However, these issues cannot be resolved by simply improving consumer literacy. This direction of improving consumer literacy is based on the assumption that companies, etc. using AI provide appropriate information to consumers. Therefore, an intermediate guideline to bridge the gap between AI Principles and corporate-level practice is deemed necessary in the sense of facilitating such information provision.

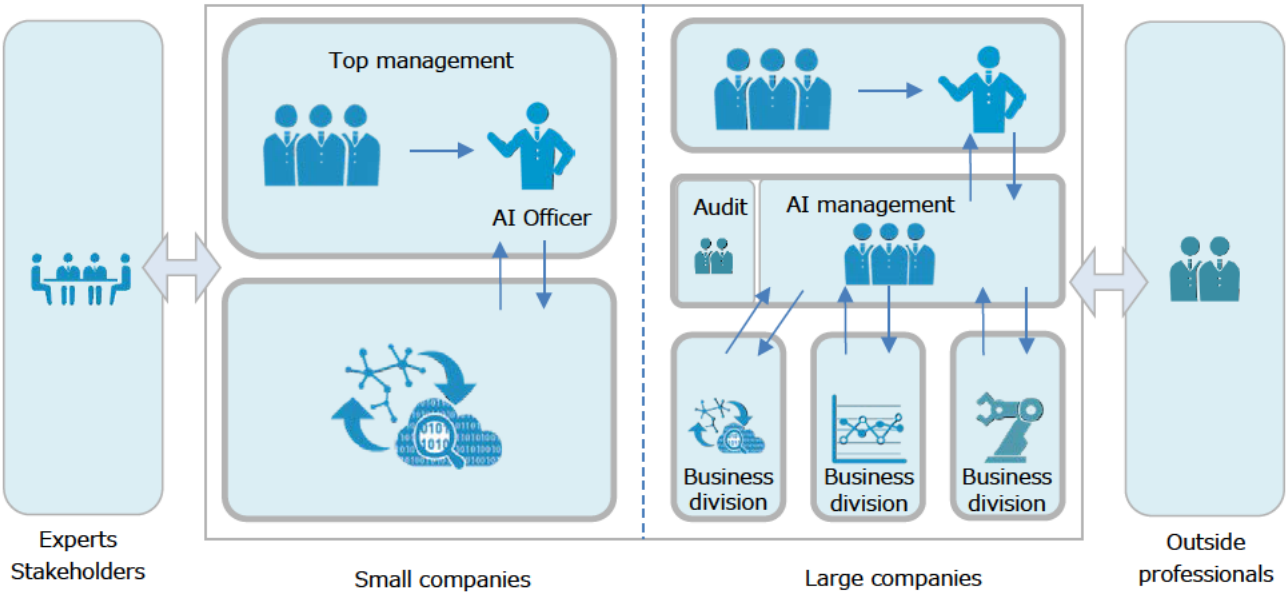
C. Ideal approaches to AI governance in Japan

This section makes a recommendation on ideal approaches to AI governance in Japan based on the above analysis and the architecture of AI governance presented in 2.D.

(1) Legally non-binding corporate governance guidelines

Principles, guidelines, articles, etc. that are released and discussed domestically and internationally illustrate various risks related to the use of AI and summarize points to note and consider regarding how to address these risks. However, these points to note and consider are often summarized by principle and it is not easy to translate them into corporate governance. In addition, while these principles, guidelines, articles, etc. do not seem to be intended to provide a one-size-fits-all checklist that is to be applied across the board to all AI users, since making a choice about each item is left to each user's consideration, companies with limited experience in using AI may feel obligated to respond to all items[72]. Further, since guidelines, etc. that are released and discussed domestically and internationally do not provide sufficient guidance on risk assessment, companies may be faced with potential risks more than necessary and the use of AI may be hindered as a result[73].

To support flexible responses to risks, etc., it is effective to provide certain guidance on desirable responses and organizations according to the size of the risk, etc. Such risk-based management



[72] For example, the following is widely known as a comprehensive self-checklist for using AI profiling functions. Personal Data +α Association "Interim Report Attached to Draft Recommendations on Profiling" NBL No. 1137 (January 2019). Available only in Japanese.

[73] Keidanren "AI Utilization Strategy For an AI-Ready Society" (February 19, 2019) P.34 states, "For example, for the government to hold AI users accountable across the board without considering any background, such as the scope of AI use and social impact, to ensure AI reliability may inhibit the use of AI and thus is inappropriate." The full text is available only in Japanese, although the outline is available in English. Consideration should be given to the fact that even if they are legally non-binding guidelines, they have a certain degree of influence.

is natural consequence of goal-based governance, which is based on the assumption of flexible responses by companies. Desirable responses to risks and organizations for small companies in the use of AI are necessarily different from those for large companies. In addition, risks vary according to specific situations in which AI is used. The provision of guidelines that include desirable risk assessment and management is expected to facilitate deployment of AI.

In developing guidelines, not only companies that use AI but also a wide range of stakeholders, including users, engineers, academics, and law/audit experts, should engage in the discussion. It is desirable that the government functions as a facilitator in the discussion and objectively evaluates whether companies satisfy the guidelines developed, thereby enhancing society's trust in the companies that meet the guidelines[74]. In addition, attention should be paid to differences in the level of experience in using AI. There should be companies with extensive experience in using AI on the one hand and those that are thinking of actively using it in the future on the other[75]. Semi-forcibly applying sophisticated governance of companies that have used AI on a large scale for a long time to companies that have just started using AI on a small scale may stagnate the use of AI. For companies with little experience, it is also effective to provide specific guidance by using examples.

Given these elements, in addition to the opinions of stakeholders provided in 3.B., a non-binding intermediate guideline should give consideration to the following points: Avoid using standards that are based on a specific level of experience in using AI[76], avoid one-size-fits-all application to all companies, do not prevent leading companies in governance from doing something new, support the improvement of AI risk management, etc., function as a benchmark of trustworthiness of AI system in inter-company transactions, include useful good practices for companies that have just started using AI, and facilitate the provision of explanations to consumers, etc. A possible outline of a non-binding intermediate guideline is as provided below.

- Creation of a foundation for use of AI: Dissemination of activities throughout the company, raising awareness of AI governance, and improving AI literacy
- Development/introduction of AI systems: Development of principles, creation of a management framework[77], establishment of an escalation process, development of a risk management process[78]
- Operation of AI systems: Monitoring, internal audits, use of external evaluation,

---

[74] Governance Innovation Report 5.1.2

[75] See *supra* 74 for the consideration of this argument.

[76] Keidanren's "Guidelines for AI-Ready Society" provide rough standards for the level of use of AI.

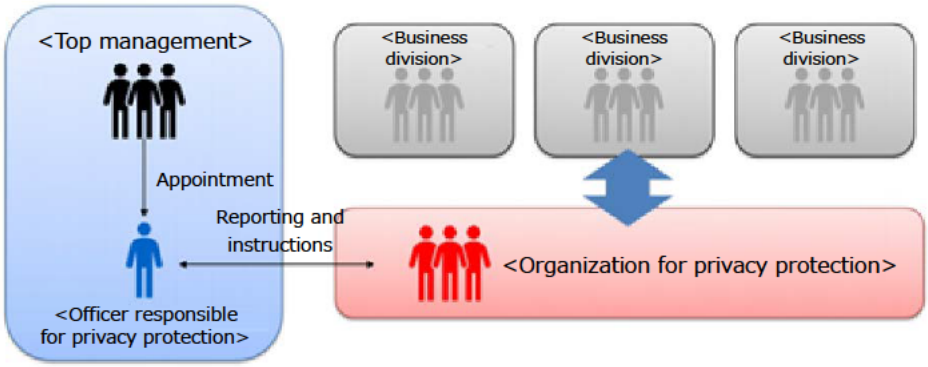
[77] Since the majority of AI use is positioned downstream of data governance ("AI strategies are designed to advance society as a whole using AI based on the premise that there is a large volume of high-quality data. Therefore, data strategies are positioned to be responsible for the assumption of AI strategies. "(Data Strategy Task Force Preliminary Report (draft), available only in Japanse), it is not a must to handle data governance in AI guidelines. However, since the use of AI is closely related to data governance, attention should at least be paid to consistency with discussions on data governance.

[78] In AI systems, failure to achieve intended equality and other situations in addition to conventional safety are related to risks. This is why the definition of risk—impact of uncertainties about various purposes (ISO31000)—is believed to be suitable. Meanwhile, in the context of safety, it seems easier to manage if the traditional definition—combination of harm with probability of its occurrence—is adopted.

development of relationships with stakeholders, improvement and progress management

In developing a non-binding intermediate guideline, attention should be paid to characteristics of corporate governance in Japan while referring to activities in foreign countries. There is research suggesting that Japanese companies that have valued their relationships with employees and society have the upper hand in AI risks compared to European and US companies[79]. The said guideline is required to provide creative solutions to elicit the advantages of Japanese companies.

In developing guidelines on AI, consistency must be ensured with other guidance on governance in the era of digitalization to facilitate the use of the guidelines by companies promoting DX[80].



Cited from the Guidebook on Corporate Governance for Privacy in Digital Transformation (DX)

(2) International standards

The ISO/IEC JTC1 Technical Committee for SC 42 (the Technical Committee for SC 42 has been established under the Information Technology Standards Commission of Japan in the Information Processing Society of Japan) is taking the lead in negotiations, such as chairing discussions as a convener in various WGs, etc. As mentioned in the previous chapter, the Technical Committee is building positive relationships with CEN/CENELEC, the EU's standardization bodies. These positive relationships is resulting in a positive effect on activities of the AI joint committee established between the DG CONNECT, European Commission, and the Ministry of Economy, Trade and Industry. This committee is expected to serve as a forum

---

[79] Simon Deakin "Corporate Governance, Enterprise Reform and Digitisation" NBL 1153 (September 1, 2019). The three AI risks (bias risk, lack of transparency, technological overreach) from the perspective of companies pointed out in this discussion serve as a useful reference for the development of a guideline.

[80] One example is to cite applicable sections related to the development of relationships with stakeholders from the Guidebook on Corporate Governance for Privacy in Digital Transformation (DX) in order to ensure consistency with privacy governance. In the amendment of the Act on the Protection of Personal Information, the following disciplines were introduced to address concerns about so-called "profiling": (1) relaxing of requirements for cease of utilization, deletion and others, (2) prohibition of improper use, (3) disclosure of records of third party provision, and (4) confirmation of a principal's consent regarding information that is supposed to become personal data on the recipient's side (The Amendment Act of the Act on the Protection of Personal Information, etc. (Overview)). Attention should also be paid to discussions on the handling of profiling in identifying the purpose of use of personal information (155th Personal Information Protection Commission).

for dialogue between the policy side and the standard side. It is important for the Technical Committee for SC 42 to cooperate with the Japanese government so that Japan's proposals will be reflected.

(3) Legally binding horizontal regulation

Based on the opinions of industries and the direction of improvement of literacy through the Handbook on Use of AI, legally-binding horizontal requirements for AI systems is deemed unnecessary at the moment. Even if discussions on legally-binding horizontal requirements are held in the future, risk assessment should be implemented in consideration of not only risks but also potential benefits. In doing so, consideration should be given to the possibility that certain risks may be eliminated due to the development of technologies.

In addition, specific AI-based technology itself should not be included in the scope of mandatory regulations. Even when mandatory regulations are needed, the scope of application and use of AI should be decided carefully in order to prevent regulations from having an impact on unintended areas. This is because the possible benefits and damage to society differ depending on the specific use of technologies, etc. (e.g., areas, purposes, size, and situations of use, whether it affects specified or unspecified users, feasibility of prior notification, and feasibility of opt-out). For example, while AI-based facial recognition technology is used for smartphones and other devices, the technology is effective in preventing unauthorized login by third parties, and many users are aware that AI is used and are able to choose not to use it if the performance is poor. In the case of monitoring the behavior of an unspecified number of people through cameras installed everywhere in urban areas, it may be necessary to impose restrictions from the perspective of protecting privacy even for legitimate purposes of crime prevention. For reference, the discussion above is relevant to the discussion of "Regulations focused on specific targets" in the next section.

(4) Regulations focused on specific targets

In certain areas, there may be cases where it is better for organizations responsible for industry laws to be involved in regulations rather than to be intervened from the information technology side. There may be cases where one should not think that even if AI is introduced, everything will be changed completely to such extent that the concept of safety that has been built up over the years will no longer apply. For example, in the automotive and healthcare sectors, it is deemed desirable to respect rule-making in the respective sectors by making the most of the existing concept of regulations and design philosophy.

D. Future Issues

(1) Ensuring incentives to use the non-binding intermediate guideline

Since the non-binding intermediate guideline is not legally mandatory, there may not be sufficient incentive to use the guideline, and thus it may fall short of the expectations that the guideline facilitates deployment of AI while respecting AI principles.

An example of solution to the issues of lack of incentive may be public relations efforts for raising awareness of value of the guideline in business or a new mechanism that associates the use of the guideline with benefits. For instance, an idea that a company that respects AI principle receives additional points in government procurement process of AI system may suggest a possible occasion for the guideline.

(2) Guidance on the use of AI by the government

As was seen in the domestic and international trends, progress has been made in the development of rules related to the use of AI by the government, especially in the UK. In Japan, Advisors to Government Chief Information Officer have discussed characteristics and points to be considered in the use of data in the government's information systems or AI systems that are used for services, etc. provided by the government. However, it is difficult to say that progress has been made in the development of rules in this area in Japan. AI systems may be actively introduced in the government in line with the promotion of digitalization by the government. Guidance for the government, as an end user of AI systems, may be necessary.

(3) Harmonization with other countries' governance

Since companies are able to seek sales channels for AI systems throughout the world, international harmonization is indispensable in developing AI governance. Japan has participated in discussions at/on GPAI, OECD, UNESCO, and AI standards and must continue leading these discussions. In addition to multilateral discussions, Japan should facilitate bilateral relationships, such as the EU-Japan AI joint committee and discussions with CEN/CENELEC, when effective. Through these activities, efforts should be made to make sure that Japan's AI governance is in harmony with international AI governance.

(4) Coordination between policies and standards

In the discussion of AI governance, multiple layers, such as intermediate guidelines and standards, are closely related to each other in a multi-layered way. In order to respond to these situations, the AI joint committee is exploring ways to facilitate coordination between the policy side and the standard side. The InDiCo Project operated by ETSI with the EU's budget conducts activities that focus on the importance of coordination between the policy side and the standard side. Efforts should be made in effective areas to facilitate coordination between the policy side and the standard side in order to ensure consistency in AI governance over multiple layers.

(5) Monitoring and enforcement

An issue of a non-binding intermediate guidelines is monitoring the status of its usage. Since the said guidelines is legally non-binding, it may be better to loosely grasp the status of use. Examples in the near term may include conducting a survey on the status of use and identifying reasons why the guidelines are not widely used when it is found.

Enforcement should also be discussed. For example, with regard to civil liabilities of businesses, the Governance Model Study Group has presented the following general guidance: “If it is unclear with whom the liability for negligence related to damage caused by an AI system lies, remedies for victims cannot be ensured under the liability for tortious acts prescribed by the civil law, the principle of which is liability for negligence. There is also a limit on the scope of application of the Product Liability Act, which reduces the burden of proof for victims, and, in any case, it is impossible to provide remedies to the victims if providers of AI systems do not have sufficient financial strength.” It is necessary to have elaborate discussion about the need for horizontal responses and whether responses by specific area or by usage are appropriate under the guidance provided by the Governance Model Study Group and in accordance with the AI governance system of this interim report while paying attention to international trends and discussions.

4. Concluding Remarks

This interim report has clarified the trends in AI governance in Japan and around the world and discussed AI governance that is ideal in Japan at the moment. From the perspective of balancing respect for AI principles and promotion of innovation and at least at this moment, except for some specific areas, AI governance should be designed mainly with soft laws, which is favorable to companies that respect AI principles. However, it is necessary to keep discussing AI governance in Japan since concrete discussions on the subject have just begun in the world and such discussions are most likely becoming more active in the future. In addition, multi-stakeholder engagement is indispensable for discussions on AI governance. The Expert Group on Architecture for AI Principles to Practice aims to deepen discussions by publishing this interim report and seeking a wide spectrum of opinions on a direction of discussions and future issues.

5. List of Expert Meeting Members (Expert Group on Architecture for AI Principles to be Practiced)

Toshiya Watanabe	Professor, The University of Tokyo Institute for Future Initiatives (Chair)
Takenobu Aoshima	General Manager, Data Analysis Department, Digital & AI technology Center, Technology Division, Innovation Promotion Sector, Panasonic Corporation
Shunichi Amemiya	Head of Research and Development Headquarters, NTT DATA Corporation
Naoto Ikegai	Associate Professor, Department of Policy Studies, Faculty of Economics, Toyo University
Katsuya Uenoyama	Representative Director, PKSHA Technology Inc.
Takayoshi Kawakami	Partner/Managing Director, Industrial Growth Platform, Inc. Board Member, Japan Deep Learning Association
Tomokazu Saito	Partner, LAB-01 Law Office
Mihoko Sumida	Professor, Graduate School of Law, Hitotsubashi University
Kenzaburo Tamaru	National Technology Officer, Microsoft Japan Co., Ltd.
Yoshihiro Tsuchiya	General Manager, Commercial Lines Underwriting Dept., Tokio Marine & Nichido Fire Insurance Co., Ltd.
Satoshi Hara	Associate Professor, The Institute of Scientific and Industrial Research, Osaka University
Shinnosuke Fukuoka	Partner, Nishimura & Asahi
Kazuya Miyamura	Partner, PricewaterhouseCoopers Aarata LLC
Tatsuhiko Yamamoto	Professor, Keio University Law School