Japan's AI Guidelines for Business: Key Updates in Version 1.2 Thereof

Yuka Minoda
yuka.minoda@ohebashi.com

I. Introduction

In recent years, the use of AI in business operations, particularly generative AI, has expanded rapidly. While the scope of applications of AI in corporate activities continues to broaden, responding to legal and compliance-related risks has also become an increasingly important issue. Against this backdrop, the Ministry of Internal Affairs and Communications and the Ministry of Economy, Trade and Industry formulated and published the AI Guidelines for Business in April 2024 with the objective of promoting innovation while ensuring the safety and reliability of AI. The said AI Guidelines are considered a "living document" and are therefore subject to continuous updates.

The latest version, AI Guidelines for Business Ver. 1.2 (the "**AI Guidelines**"), which was published on March 31, 2026, expands upon recent technological developments and practical trends. In addition, supplementary support tools, including the draft Handbook for Utilizing the AI Guidelines for Business and a chatbot, are currently under consideration, thereby further enhancing the practical use of the AI Guidelines. This article outlines the key updates introduced in the AI Guidelines compared with prior versions of the AI Guidelines and presents practical approaches for utilizing the AI Guidelines in corporate practice.

II. Basic Structure of the AI Guidelines for Business Ver. 1.2

Before reviewing the updates introduced in the revised AI Guidelines, it is helpful to first briefly summarize the basic structure of the updated AI Guidelines.

1. Overall Framework of the Updated AI Guidelines

The updated AI Guidelines consist of the main part[1] and several appendices.[2] The main part first sets forth the necessary definitions and terminology (Part 1) and then explains the basic philosophies of AI governance, that is, the type of society the AI Guidelines aim to realize while considering the expectations of stakeholders (the "why") (Part 2). Based on these basic philosophies, the AI Guidelines further organize the guiding principles regarding actions that should be taken in relation to AI (the "what") according to the categories of AI business actors described in Section 2 of this article, namely, the AI Developers, AI Providers, and AI Business Users (Parts 3 to 5).

The appendices to the AI Guidelines present practical implementations (the "how"), including specific practical examples and reference materials for implementing the guiding principles based on the foregoing basic philosophies. The appendices consist of Appendices 1 to 9 and include examples of AI systems

1. Available at https://www.soumu.go.jp/main_content/001064305.pdf.
2. Available at https://www.soumu.go.jp/main_content/001064306.pdf.

and services and specific use examples, examples of AI governance structures, practical examples categorized by AI business actor, contractual considerations, checklists, comparisons with overseas guidelines, and other reference materials.

2. Categorization of AI Business Actors

The AI Guidelines classify the entities subject to the AI Guidelines into three principal categories based on the AI value chain: "AI Developers," "AI Providers," and "AI Business Users," the main actors responsible for AI-related business activities. As discussed in Section 1 of this article, the actions expected to be taken by AI business actors are organized according to each category thereof. Accordingly, companies should first determine which category applies to their own activities before reviewing the contents of the AI Guidelines.

In addition, depending on how AI is utilized, a single AI business actor may simultaneously fall within multiple categories. Companies should therefore avoid viewing their role in a fixed or overly simplistic manner and should instead assess their position from multiple perspectives based on the actual way AI is being utilized by them.

III. Major Updates in AI Guidelines Ver. 1.2 and Specific Practical Applications in Corporate Business

1. Addition of Provisions Relating to AI Agents and Physical AI

One of the most significant updates introduced in Ver. 1.2 of the AI Guidlines is the addition of provisions on AI agents and physical AI. In light of the technological advancement and increasing social adoption of AI agents and physical AI, together with the growing number of cases in which AI-related risks have materialized and the emergence of new risks requiring consideration (including social transformation), the AI Guidelines newly define relevant terms, related benefits, risks, matters requiring attention, and examples of AI systems and services regarding AI agents and physical AI.

The AI Guidelines define an "AI agent" as "an AI system that perceives its environment and acts autonomously to achieve a specific goal" (AI Guidelines, Part 1). Examples of services utilizing AI agents include travel destination suggestions and AI booking agents, as well as AI sales and customer support assistance agents (AI Guidelines, Appendix 1). The AI Guidelines identify as one of the benefits of AI agents their ability to improve operational efficiency in areas such as coordination, analysis, and decision-making through autonomous actions while interacting with multiple systems (*Id.*).

The AI Guidelines define "physical AI" as "a system that takes in environmental information through sensors, processes that information using an AI model, autonomously infers and determines strategies for achieving objectives given by humans, and acts on those strategies without human intervention" (AI Guidelines, Part 1). Physical AI offers benefits such as addressing labor shortages through autonomous actions in physical environments, improving safety, and supporting nursing care and daily living. Examples include autonomous driving systems and autonomous mobile robots (AI Guidelines, Appendix 1).

The AI Guidelines further note that AI agents and physical AI involve risks such as unintended actions resulting from autonomous behavior, the increase in attack surfaces and methods, increased difficulty in control due to highly complex mechanisms, and the generation of malicious code (*Id.*). Considering these risks, attention must be paid to matters such as implementing mechanisms that incorporate human judgment, applying least-privilege settings, and

considering residual data stored in hardware (AI Guidelines, Appendix 3).

From a corporate perspective, the benefits and examples of services relating to AI agents and physical AI cited in the AI Guidelines may serve as useful references when considering the development, provision, or use of new products and services. Companies developing, providing, or using AI agents or physical AI must also pay careful attention to the risks identified in the AI Guidelines and implement measures that appropriately take relevant considerations into account.

2. Updates to AI-Related Risks

The updated AI Guidelines reorganized and expanded the comments relating to AI-related risks to enable AI business actors to better identify and appropriately respond to risks arising from the use of AI. More specifically, the updates include: (i) the addition of content intended to facilitate a risk-based approach; (ii) updates to AI-related risks; and (iii) revisions to the classification of discriminatory outputs.

The AI Guidelines adopt the concept of a "risk-based approach," under which the priority of countermeasures is determined based on factors such as the purpose of the AI use, the relevant stakeholders, and the magnitude and likelihood of potential risks. Under this approach, AI business actors should conduct risk assessments for each category of AI use, classify risks according to levels such as high, medium, and low, and design corresponding response measures in stages. These updates provide useful guidance for AI business actors seeking to translate the risk-based approach into practical implementation and design appropriate response measures.

(a) Addition of Content Intended to Facilitate a Risk-Based Approach

To further enhance AI business actors' understanding of the risk-based approach, the updated AI Guidelines additionally reference, as helpful materials for risk assessment and risk classification methodologies, *"AI Jidai no Keiei Ishikettei to Gabanansu: Seme no AI Gabanansu Jitsugen no tame no Senryaku Repōto Ver 1.0 (the Management Decision-Making and Governance in the AI Era: Strategic Report for Realizing Offensive AI Governance Ver 1.0)"* published by the AI Governance Association ("**AIGA**"),[3] as well as the supplementary material to the EU AI Act entitled *"Artificial Intelligence Act Annex III: High-Risk AI Systems Referred to in Article 6(2)"*[4] (AI Guidelines, Appendix 2).

(b) Updates to AI-related Risks

The discussion of AI-related risks has been updated considering recent developments. These updates include additional discussions of examples of attacks against AI systems; risks of infringement of privacy rights arising from the use of multimodal generative AI, cameras, and voice recognition technologies; the possibility that hallucinations may also generate certain benefits; risks associated with the use of AI in the education sector; risks of becoming victims of financial loss; and risks relating to the infringement of licenses, qualifications, and other rights (AI Guidelines, Appendix 1).

(c) Updates to the Classification of Discriminatory Outputs

"Discriminatory outputs," which constitute one category of AI-related risks, was previously classified as an output-stage risk within the category of technical risks (i.e., risks primarily associated

3. Available at https://cdn.prod.website-files.com/66e98b87b115812d1af8fc1c/69285da091ec71dde1ae3c71_management-strategy-report-ver1.0.pdf (in Japanese).
4. Available at https://artificialintelligenceact.eu/annex/3.

with AI systems). However, based on the view that whether a particular issue constitutes a risk should not be determined solely by technical characteristics, but rather by legal and ethical evaluation as well, discriminatory outputs have been reclassified as an ethical and legal risk within the category of societal risks (i.e., existing risks that may also arise in AI or be amplified by AI). Please refer to Table 1 below for details regarding the systematic classification of AI-related risk examples, including the foregoing updates.

Table 1: Systematic Classification of AI-related Risk Examples (tentative version)[5]

Major categories	Subcategories	Risk examples
Technical Risks (=risks primarily associated with AI systems)	Risks during the learning and input stages of AI	Attacks on AI systems such as data poisoning attacks
	Risks during the output stage of AI	Biased outputs and inconsistent outputs
		Incorrect outputs due to Hallucinations and similar issues
	Risks during the post-response stage	Black-boxing and inadequate explanations of decisions
Societal Risks (=existing risks that may also arise in AI or be amplified by AI)	Risks related to ethics and law	Inappropriate use of personal information
		Occurrence of accidents related to lives, etc.
		Discriminatory outputs
		Excessive dependence
		Misuse
	Risks related to economic activities	Infringement of intellectual property rights, etc.
		Financial loss
		Leak of confidential information
		Unemployment of workers
		Concentration of data and profits
		Infringement of qualifications, etc.
	Risks related to the information space	Distribution and diffusion of disinformation
		Negative influence on democracy
		Filter bubble and echo chamber phenomena
		Loss of diversity and inclusion
		Reproduction of biases
	Risks related to the environment	Energy consumption and environmental load

3. Categorization of AI Business Actors

The updated AI Guidelines revised the categorization of each AI business actor through: (i) supplemental clarification of the definition of AI Developers; (ii) revisions to the "Correlation between AI business actor and general AI use flow"; and (iii) revisions to the role of each AI business actor. As discussed in Section II-2 above, when using the AI Guidelines, companies must determine which category or categories apply to their own activities. These updates are expected to facilitate such determination.

(a) Supplemental Clarification of the Definition of AI Developers

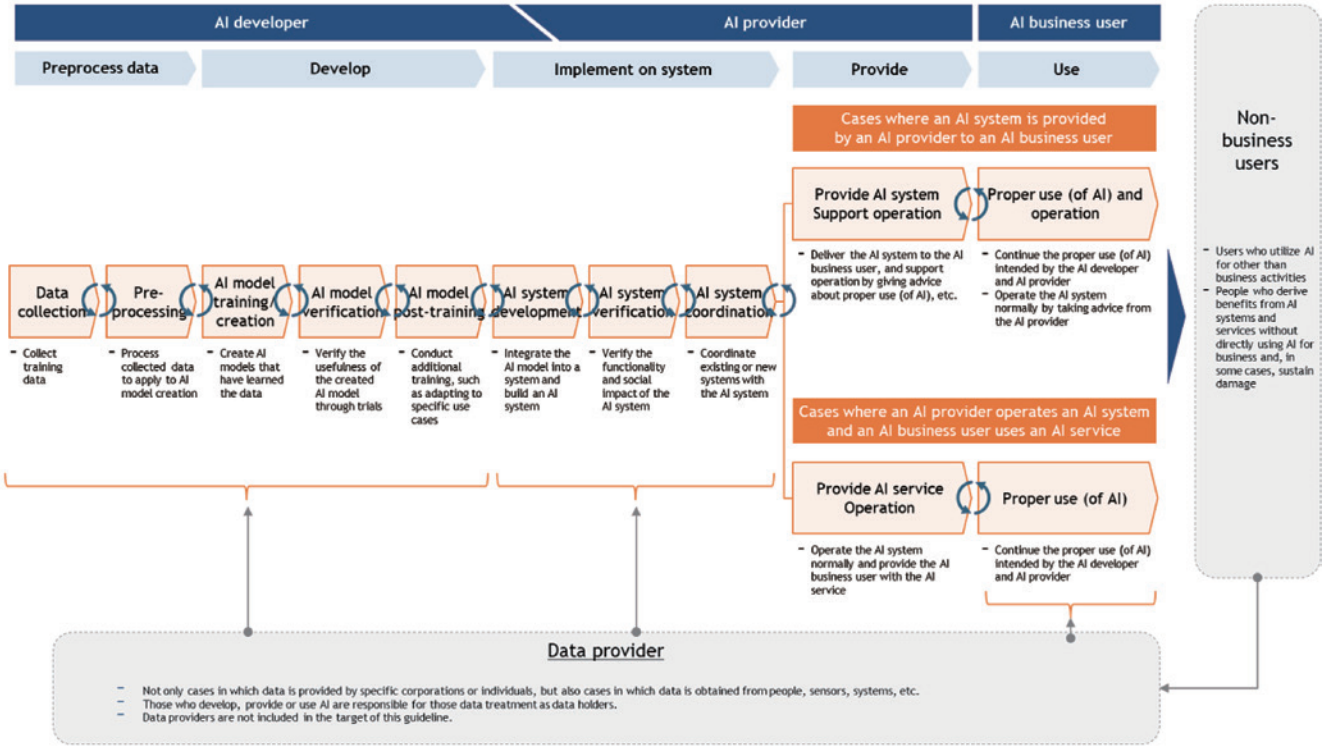
The definition of AI Developers was expanded to clarify that AI Developers do not necessarily undertake all aspects of AI system development and that their role includes post-development model adjustments (post-training), such as fine-tuning, following the development of AI models (AI Guidelines, Part 1).

5. *Appendix to the AI Guidelines for Business*, p. 24, available at https://www.soumu.go.jp/main_content/001064306.pdf.

(b) Revisions to "Correlation between AI Business Actor and General AI Use Flow"

The "Correlation between AI Business Actor and General AI Use Flow" below was revised to expressly include "post-training of AI models" and reorganize the processes relating to "data preprocessing" and "implementation into systems" (AI Guidelines, Appendix 1).

Correlation between AI Business Actor and General AI Use Flow[6]



(c) Revisions to the Roles of AI Business Actors

The role of each AI business actor was also revised as follows:

- where the same AI business actor performs fine-tuning and provides AI systems, it is categorized as both an AI Developer and an AI Provider (AI Guidelines, Appendix 1);
- where an AI system is provided with the assistance of code-generation AI or similar technologies, the relevant actor is categorized as an AI Business User and an AI Provider (*Id.*);
- the AI Guidelines clarified that the definition of API specifications and similar matters fall within the role of AI Developers, while implementation processing relating to APIs falls within the role of AI Providers; and
- the respective roles of AI Developers and AI Providers in relation to alignment (i.e., adjustments intended to realize behavior consistent with human intentions and values) and Retrieval-Augmented Generation ("**RAG**") were reorganized and revised (AI Guidelines, Part 1, and Appendices 3 and 4).

6. *Id.*, p. 9, available at https://www.soumu.go.jp/main_content/001064306.pdf.

4. Revision of Definitions and Terminology for "Training," "Inference," and "Data"

In the updated AI Guidelines, the terminology relating to "training" and "inference," which may be interpreted in multiple ways, as well as the classification of "data," has been reorganized and revised to improve clarity and facilitate understanding.

First, because the term "training" may be interpreted in different ways, the AI Guidelines added a clear definition (AI Guidelines, Part 1). By expressly clarifying that training refers to "the process of determining or improving the parameters of an AI model," the AI Guidelines make clear that "In-Context Learning" does not constitute training because it does not involve the determination of model parameters.

The updated AI Guidelines also added a clear definition of the term "inference," considering the expanding scope of data handled or processed during AI utilization through technologies such as RAG (*Id.*). Under the said guidelines, "[i]nference is the process of providing new data to a trained model and computing outputs (such as predictions, classifications, and generation)."

Furthermore, the updated AI Guidelines added the table below on data used in AI training and utilization, thereby clarifying the relevant definitions. In addition, the terminology relating to data used throughout the AI Guidelines has been revised to align with terminology more commonly used in practice (AI Guidelines, Appendix 1).

Table 2: Classification of AI-related Risks[7]

	Process	Overview	Data Type	Overview	Specific Examples
Flow of AI Training and Utilization	Training (Machine Learning)	A process of determining or improving model parameters using large volumes of data	Training data	Data used to optimize model parameters. The learning algorithm minimizes errors based on this data and learns the relationship between inputs and outputs.	• Internal data • Large-scale open data (e.g., CIFAR-10, MNIST) • Data from stakeholders • Data collected from sensors and systems (refer to Figure 6 "Concepts of data Provision" of the Appendix)
			Validation data	Data used separately from training data during the model training process. It is not used to update model parameters, but is primarily used for purposes such as hyperparameter tuning, detection of overfitting, and model evaluation and judgment.	
			Test data	Data used to evaluate the final performance of a model after the training has been completed. As it has not been used for training or validation, it serves as an appropriate basis for indicators of generalization performance.	
	Inference	A process of providing new input data to a trained machine learning model and computing outputs (such as predictions, classifications, and generation)	Data for inference	Data for inference refers to data used when a trained model generates outputs in response to new inputs. This includes not only user instructions that the model directly processes and data from production environments, but also additional information (such as internal databases and external knowledge) referenced for contextual supplementation and accuracy improvement. By combining these elements, the accuracy and consistency of the model's responses are expected to be enhanced.	• User inputs (prompts, images, audio) • Data acquired in operational environments (sensor-acquired data) • Internal information (FAQs, knowledge bases) • External information (web search results, external APIs) • Contextual information (past conversation history, session information) • Outputs from other models (text and analysis results generated by other AI models)

5. Practical Support Tools for Using the AI Guidelines

To support the use of the AI Guidelines, the following have been made available: (i) the draft Handbook for Utilizing the AI Guidelines for Business, and (ii) a chatbot.

(a) Draft Handbook for Utilizing the AI Guidelines for Business

The draft Handbook for Utilizing the AI Guidelines for Business was prepared to support the practical use of the AI Guidelines. It introduces, among other

7. *Id.*, p. 7, available at https://www.soumu.go.jp/main_content/001064306.pdf.

things, the foundational concepts underlying the use of the AI Guidelines, matters that are advisable to address at the initial stage as preparation and groundwork for establishing AI governance, methods for referring to the AI Guidelines when implementing AI governance in practice, and examples of practical use. In particular, the handbook is intended for organizations and individuals that are beginning to establish and implement AI governance frameworks and is expected to serve as a useful practical tool. However, an English version of the handbook has not yet been released.

(b) Chatbot

A rule-based AI chatbot has been made to enable users to review information relating to the AI Guidelines by two methods: (i) selecting the relevant AI business actor and the matters to be confirmed; and (ii) free-text input. Where questions arise regarding the content of the AI Guidelines, the chatbot provides an accessible first option for consultation and is expected to facilitate understanding of the AI Guidelines. The chatbot is available in several languages, including English.

6. Inclusion of Domestic and International Trends and Corporate Initiatives

The updated AI Guidelines incorporated recent developments relating to AI governance, including the latest domestic and international trends and examples of corporate initiatives that warrant attention. Although this article does not discuss these matters in detail, the updates provide valuable insight into both institutional developments and practical implementation efforts.

IV. Conclusion

The contents of the AI Guidelines were reorganized and expanded considering recent technological developments and the advancement of AI use in practice. As such, the AI Guidelines may be regarded as a useful framework for companies considering how AI should be used and governed in corporate practice.

We hope that this article will contribute to a better understanding of the updated AI Guidelines and assist readers in using them in practice.

In addition, because the AI Guidelines are positioned as a "living document" and are expected to undergo continuous review and revision, companies should not merely refer to the AI Guidelines on a one-time basis. Rather, it is important for companies to continue considering how future updates should be reflected in practice while monitoring developments relating to the AI Guidelines.