



AI

Strengthening competitiveness through strategic investment in sovereign AI via public-private partnerships

March, 2026

Industry Research Department
Mizuho Bank

Glossary

- ✓ Generative AI: AI (and its core models/systems) primarily intended to generate content such as text, images, audio, and video. As a development approach, it leverages deep learning—particularly foundation models—which existed prior to the advent of generative AI.
- ✓ Foundation Models: Models that acquire general-purpose capabilities through training on large-scale data (primarily text) and are designed on the premise of being repurposed (applied) across a wide range of tasks. The stage of acquiring general-purpose capabilities is referred to as "pre-training," and the model at the point pre-training has been completed is referred to as the "base model."
- ✓ Specific Models: Models optimized through additional training of a base model for a specific domain (e.g., healthcare, finance, education) or task (e.g., classification, decision support, drafting, summarization, analysis).
- ✓ Sovereign AI: A state or concept in which a nation independently develops, operates, and manages the core elements of AI (e.g., compute resources, data, models) without relying on other countries or external organizations.
- ✓ Agentic AI: AI that autonomously makes decisions and takes actions to achieve goals on behalf of organizations or individuals.
- ✓ Physical AI: A design and development approach for integrating AI systems with physical devices (e.g., robots, vehicles, equipment, factory facilities) so that they function in the real world.

Envisioned landscape of the AI domain in 2040

AI: Pursuing performance improvement through scaling of data, models, and compute resources, and advancing the automation of society and the economy

Vision for society in 2040 shaped by AI

- ✓ Realizing capabilities surpassing humans in virtually all tasks, premised on the scaling of data, models, and compute resources
- ✓ Embedded into society and the economy as a key technology, becoming the infrastructure of socioeconomic activities

Current situation

- ✓ The U.S. and China are leading in the competition for AI development and deployment, backed by large-scale availability of compute resources and a deep pool of talent producing cutting-edge research outcomes
- ✓ Establishing and maintaining competitive advantage requires large-scale investment in compute resources and talent acquisition

Japan's strengths

- ✓ The depth of data accumulated primarily in the manufacturing sector, a core strength of Japan, constitutes a competitive advantage in AI development and deployment

Japan's challenges

- ✓ In addition to compute resources, data technology infrastructure as well as specialized talent and skills are in short supply
- ✓ Weaknesses in the technology base drive an expansion of the digital deficit

Strategy for Japanese industry

- ✓ Aiming to establish a technology base through the construction of Sovereign AI via public-private partnerships

Strategies

- ✓ Creating use cases expected to generate recoverable outcomes in order to secure the financial resources required for the construction of Sovereign AI

Envisioned landscape for Japan's future in the AI domain

- ✓ Establishing a technology base through the construction of Sovereign AI and deploying it across socioeconomic challenge domains such as healthcare to curb expenditures
- ✓ Although investment on the scale of several JPYT will be required, expenditure containment of over approx. JPYT10 is expected through addressing socioeconomic challenges. Indirect economic effects are also anticipated through embedding AI into autonomous driving and exporting AI systems overseas

Strategic investment via public-private partnerships envisioned to strengthen competitiveness in the AI era

■ Background and Issues

- With the advent of generative AI, AI is evolving into a general-purpose technology capable of handling a wide range of cognitive tasks such as decision-making, drafting, and analysis. By 2040, AI is expected to handle a wide range of cognitive tasks and become a key technology forming the foundation of economic and national security.
- Meanwhile, the U.S. and China are leading in cutting-edge AI development, backed by compute resources, talent, and capital. If domestic AI development does not advance, technological and economic competitiveness will decline and the digital deficit risks expanding further over the medium-to-long term. In addition to the economic perspective, given that AI underpins economic security and national security, excessive reliance on foreign sources is also an issue.

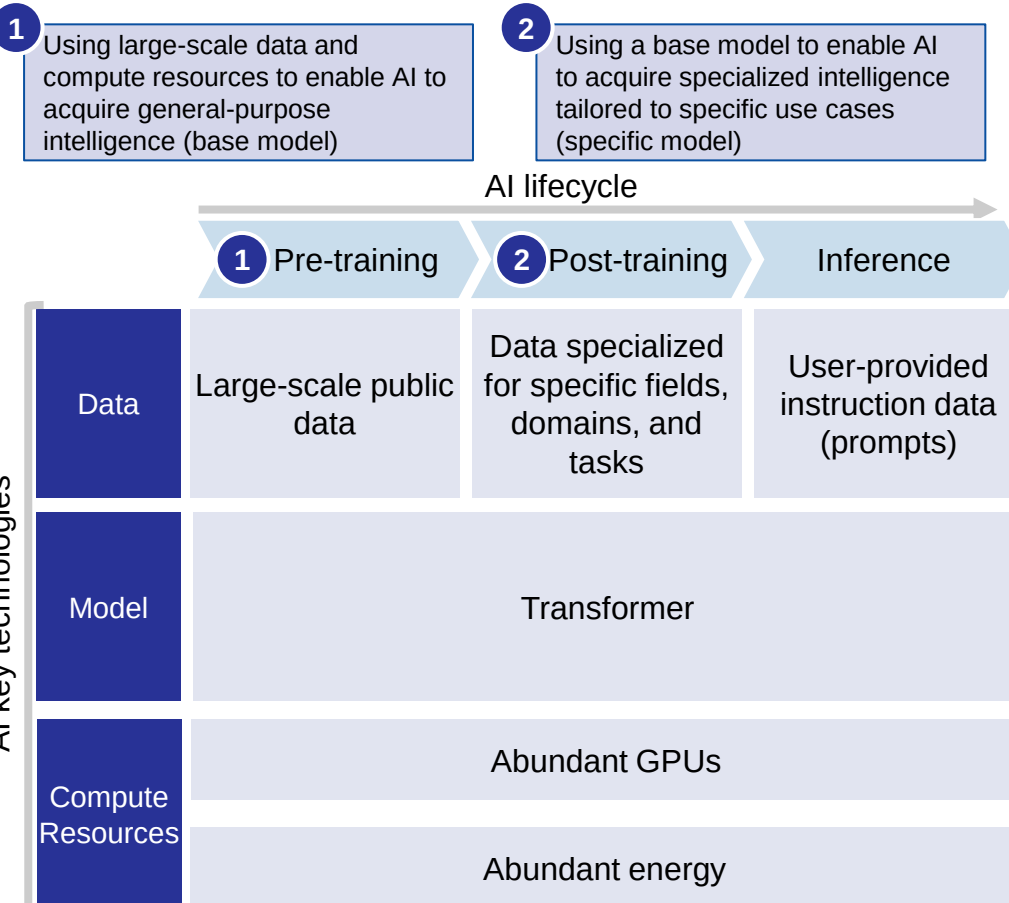
■ Key Points

- With AI expected to become a key technology, domestic AI development has emerged as a focal point, both from the perspective of the economic risk that the digital deficit could expand to as much as approx. JPYT36 by 2040 and from a security standpoint. In the current AI development race, the U.S. and China—possessing large-scale compute resources and top-tier talent—are leading, while Japan lags behind. On the other hand, Japan possesses data potential underpinned by a robust industrial base centered on manufacturing, as well as strengths such as friendly relations and industrial collaboration platforms with ASEAN that can be leveraged in the future overseas deployment of AI.
- To maximize these strengths and maintain and reinforce national interest and industrial competitiveness, the key lies in realizing "Sovereign AI"—a state in which the nation independently develops and operates the core elements of AI such as compute resources, data, and models. Public-private partnerships serve as the course of action for realizing "Sovereign AI," with the following role-sharing envisioned: (1) the public sector leads the development of base models providing general-purpose capabilities and their application to highly public use cases such as resolving social issues; and (2) the private sector applies base models to industrial use cases to strengthen industrial competitiveness.
- Regarding Japan's AI strategy, the government is also promoting various measures as one of the 17 strategic fields. While aligned with the government's policy direction, this report focuses on (1) clarifying the role-sharing between the public and private sectors, and (2) designing a recovery logic for upfront investments, with reference to the concept of GX Economy Transition Bonds.
- The cumulative investment scale for 2026–2040 is estimated at approx. JPYT2.8–4.8, yet investment recovery is considered feasible through outcome design that emphasizes direct impacts on expenditures such as social security cost containment. From the perspective of economic and national security in the AI Era, resolving social issues and strengthening industrial competitiveness are anticipated through strategic investment in "Sovereign AI" via public-private partnerships.

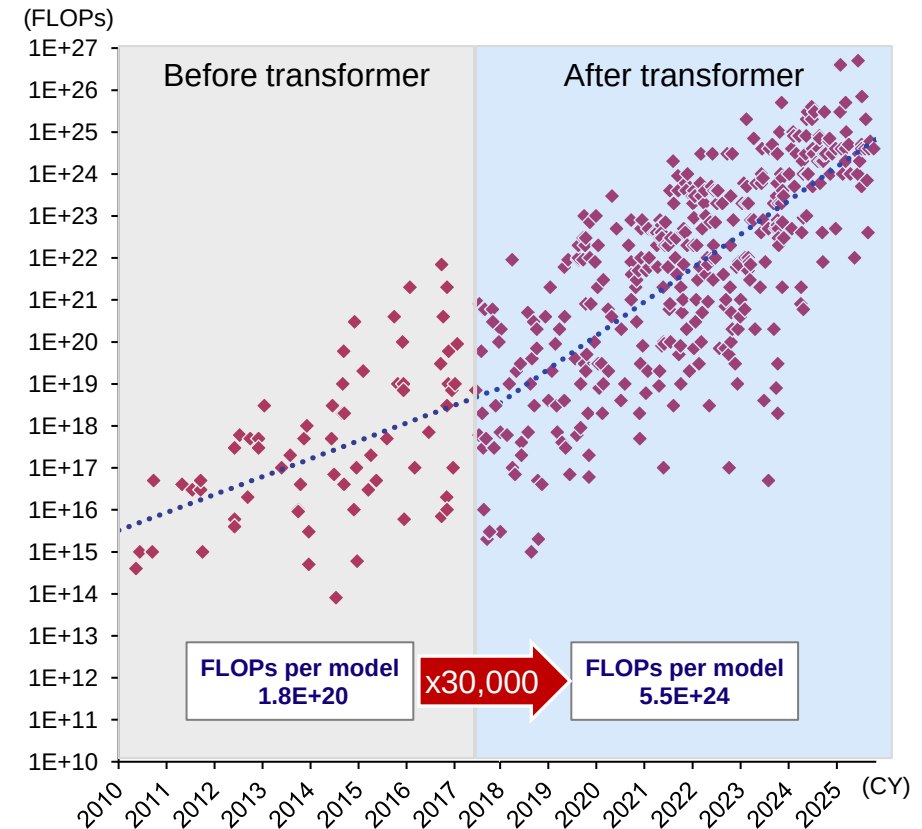
Mainstream AI development is scaling based on transformer

- Artificial intelligence (AI), a technology that emulates human intellectual activity, currently relies primarily on an approach called machine learning, which uses data and compute resources to "have machines learn how to solve tasks."
 - Since the method known as "Transformer" was proposed in 2017, the trend has been to pursue performance improvement by scaling data, models, and compute resources based on this method. In addition, by tuning for specific use cases, various tasks can be solved.

Overview of AI: key technologies and lifecycle



Trends in AI model compute volume before and after transformer



Note: FLOPs stands for Floating Point Operations and indicates the compute volume required for AI model training.

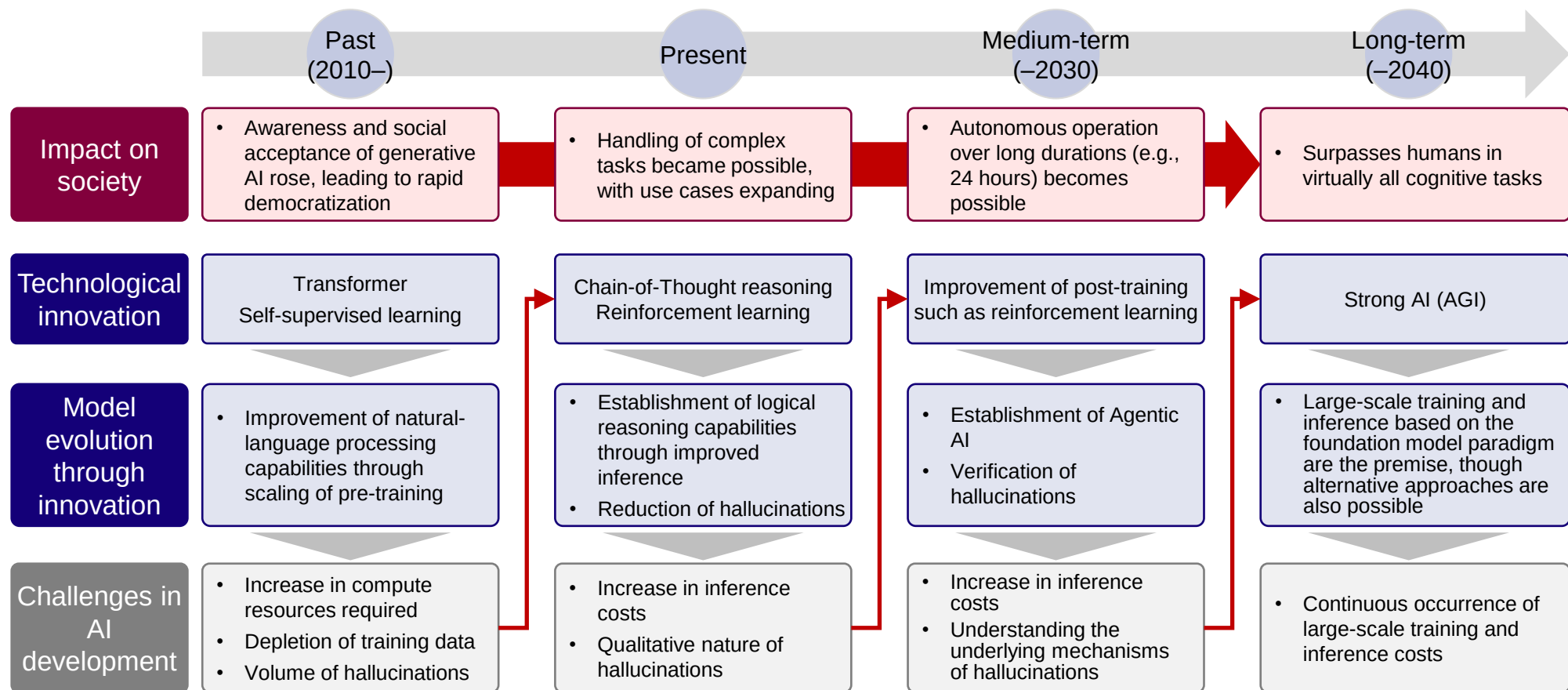
Source: Compiled by Industry Research Department, Mizuho Bank, Ltd. based on Epoch AI, 'Data on AI Models'. Published online at epoch.ai. Retrieved from <https://epoch.ai/data/ai-models> [online resource].

Source: Compiled by Industry Research Department, Mizuho Bank, Ltd.

AI is rapidly transitioning to practical deployment and is expected to impact society and the economy through automation of cognitive tasks

- Following Transformer, "generative AI" emerged, capable of sophisticated recognition and generation of text, images, and more. Research and development has been led by high-tech companies represented by OpenAI in the U.S., and AI is now capable of handling complex tasks.
 - Although technological challenges remain, by 2040 the emergence of strong AI (AGI) surpassing humans in virtually all tasks is anticipated. Going forward, automation of cognitive tasks such as decision-making, drafting, and analysis is expected to impact society and the economy.

Overview of AI: history and outlook

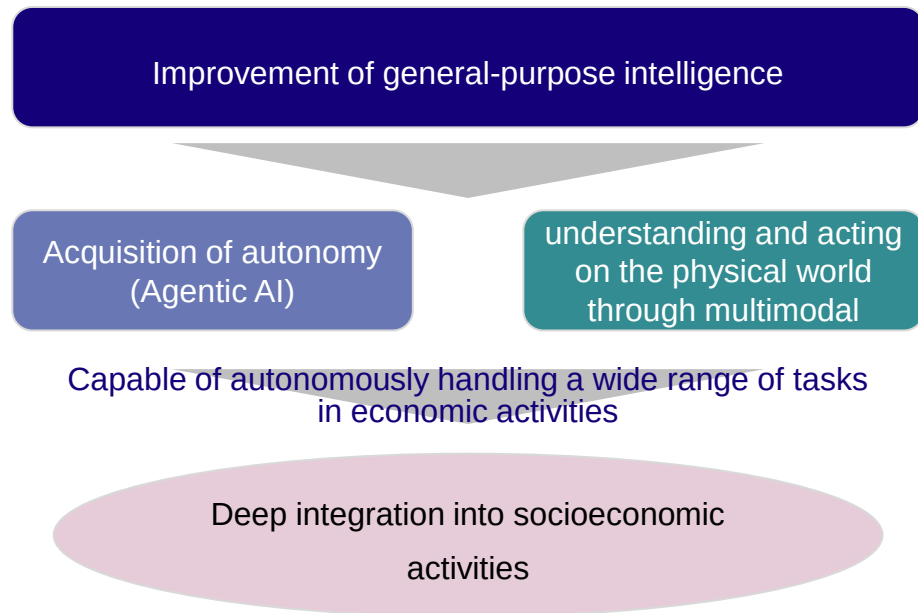


Source: Compiled by Industry Research Department, Mizuho Bank, Ltd.

AI will become a key technology toward 2040, functioning as the foundation of economic and national security

- By 2040, with AI expected to handle a wide range of cognitive tasks, AI is positioned as a key technology supporting decision-making in socioeconomic activities and is anticipated to form the foundation of economic security and national security.
- In a vision of society where AI becomes a socioeconomically indispensable technology, establishing in-house capabilities for the core elements of AI, such as compute resources, models, and data, is critical.

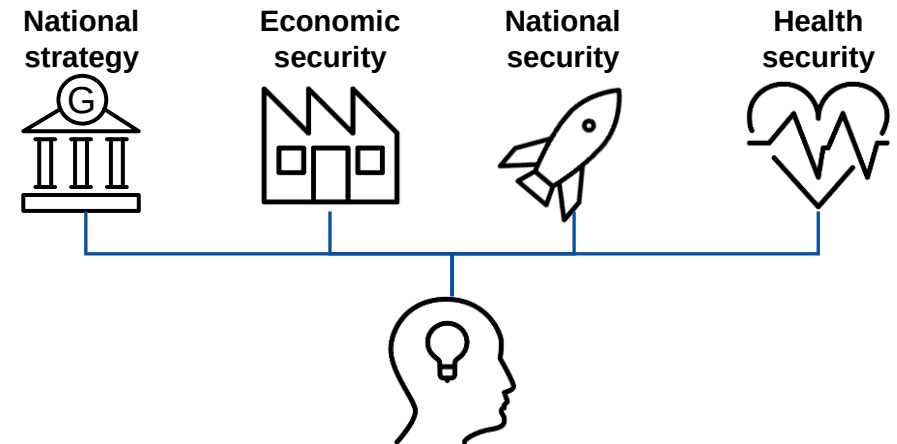
Real-world impact expected alongside improvement in AI performance



Alongside the improvement of general-purpose intelligence, AI is becoming a socioeconomically indispensable technology through the acquisition of autonomy and advancement in understanding and acting on the physical world

Source: Compiled by Industry Research Department, Mizuho Bank, Ltd.

The need to establish strategic autonomy and indispensability with a view to the vision of society in 2040



Strategic autonomy

- From the perspective of security and prevention of national wealth outflow, in-house capabilities are required for compute resources and models, avoiding excessive dependence on other countries

Strategic indispensability

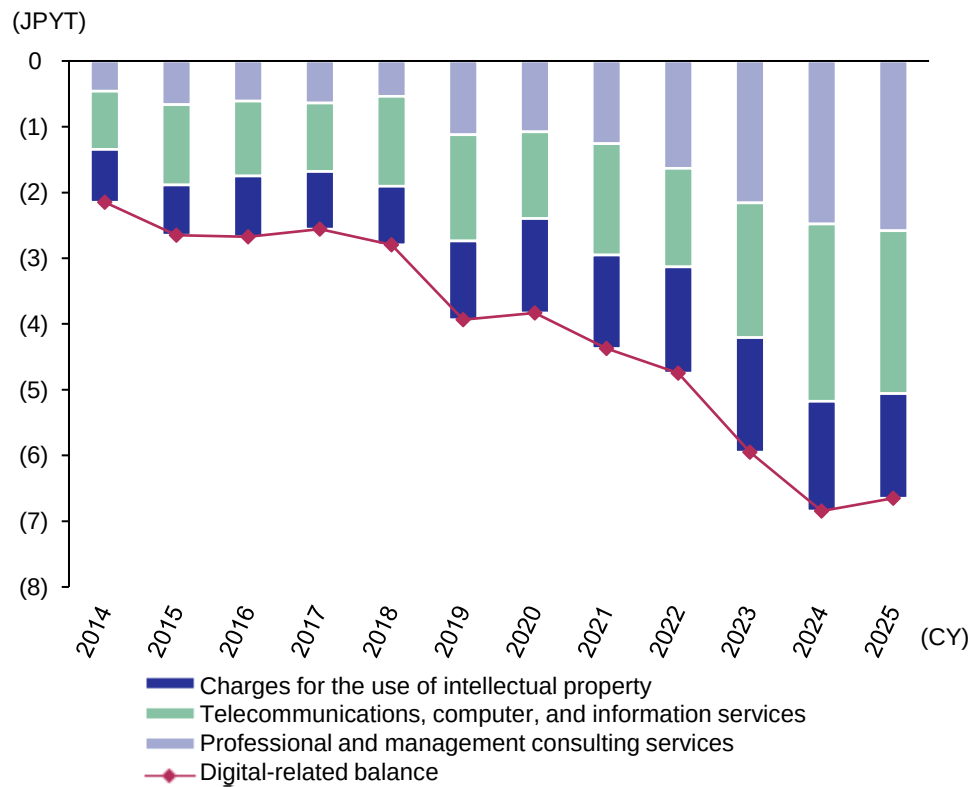
- Establishing capabilities to develop the AI technology stack and create use cases, and building advantages that can also appeal to other countries and regions, are required

Source: Compiled by Industry Research Department, Mizuho Bank, Ltd.

As AI becomes a key technology, there is a risk of further expanding the digital deficit

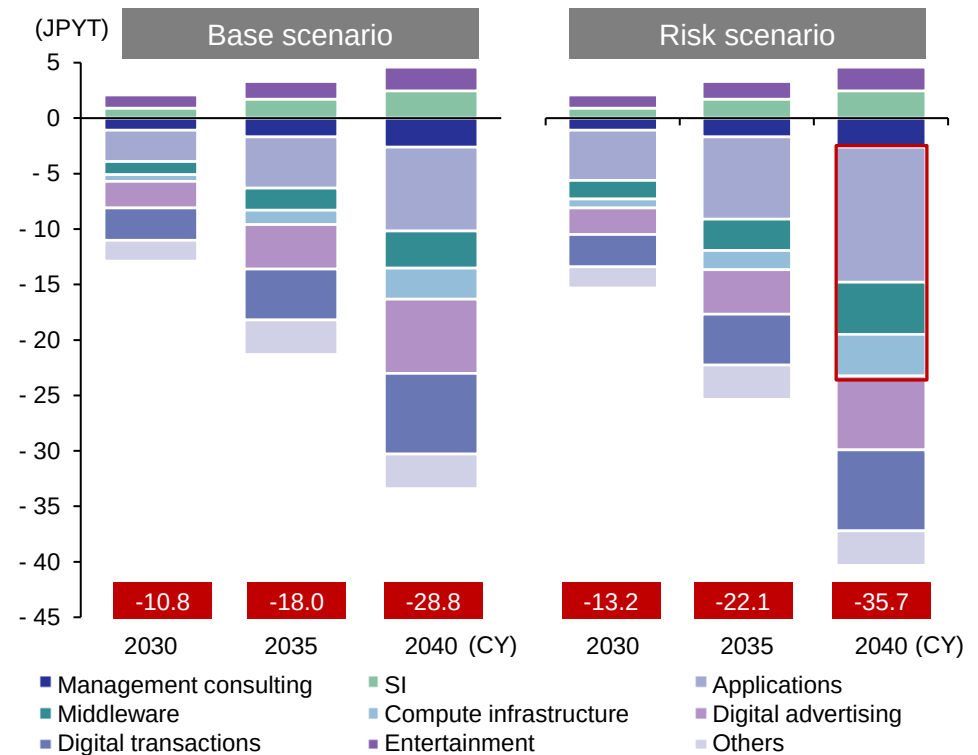
- Japan's digital-related balance has seen its deficit widening since the 2010s when cloud services began to spread, reaching approx. JPYT6.7 in 2025—more than three times the level of 2014, the earliest observable year. The deficit expansion is mainly attributable to increases in computer services and platform usage fees.
- If domestic development of foundation models does not advance, in addition to dependence on foreign cloud services, the digital deficit could expand further, risking outflow of national wealth.
 - According to estimates by the Ministry of Economy, Trade and Industry (METI), the digital deficit is expected to reach approx. JPYT18 by 2035. A scenario is also envisioned in which, if catch-up in foundation models lags, national wealth on the scale of nearly JPYT36 could flow out overseas by 2040.

Trends in Japan's digital-related balance



Source: Compiled by Industry Research Department, Mizuho Bank, Ltd. based on "Balance of Payments Statistics" by the Ministry of Finance and Bank of Japan

Estimates of Japan's digital deficit



Note: The 2040 projection for the base scenario and the 2030, 2035, and 2040 projections for the risk scenario are estimated by Industry Research Department, Mizuho Bank, Ltd. with reference to the projection logic in the "Digital Economy Report"

Source: Compiled by Industry Research Department, Mizuho Bank, Ltd. based on the "Digital Economy Report" by the Ministry of Economy, Trade and Industry (METI)

AI model development faces the same competitive environment as cloud services, requiring action on structural bottlenecks

- Regarding dependence on foreign cloud services—the primary driver of the recent expansion of the digital deficit—three structural factors are considered to be at work: "large-scale upfront investment," "economies of scale," and "developer ecosystem."
- The same structural factors are considered to be at work in AI model development as well, requiring action
 - In areas that are difficult for private-sector companies to address alone, the government needs to take the initiative in driving progress

Japan's challenges interpreted through the structural factors that led to dependence on foreign cloud services

	Examples of failures in cloud services		Anticipated pattern in AI model development
Large-scale upfront investment	➤ From the perspective of short-term revenue certainty, upfront investment until scaling and monetization cannot be tolerated ➤ Investment in data centers, networks, and other facilities is enormous, and no single company can bear the burden alone	➤	➤ Due to differences in capital market size and risk tolerance in asset management, it is difficult for unicorns to emerge ➤ Large-scale investment in compute resources for foundation model training cannot be undertaken at the individual-company level
Economies of scale	➤ Demand is dispersed in both private and government procurement (due to individual requirements and silos), resulting in a fragile home market ➤ On the other hand, global majors leverage economies of scale to enjoy first-mover advantage (placing late entrants at a disadvantage)	➤	➤ The absolute volume of Japanese training data lags, making it difficult to replicate scaling laws ➤ In inference as well, use cases may be limited for Japanese-specific models
Developer ecosystem	➤ APIs/SDKs, documentation, marketplaces, and communities become de facto standards as they grow, creating a structure that attracts even more developers; however, in Japan, contract-based business structures provides weak incentives to expand common platforms	➤	➤ Developers gather around major overseas frameworks, evaluation benchmarks, and open-source models, with domestic top talent absent

Source: Compiled by Industry Research Department, Mizuho Bank, Ltd.

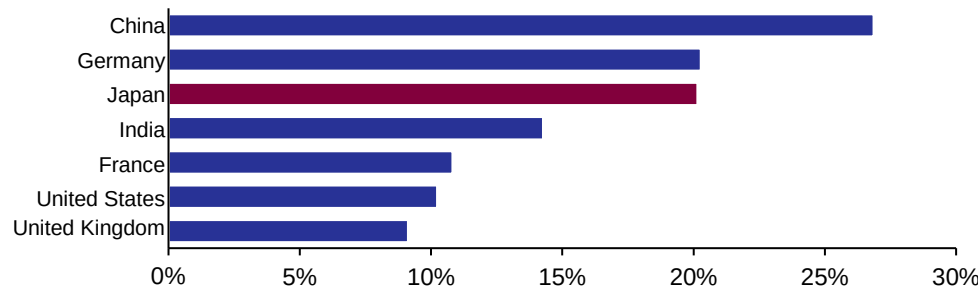
As expectations for AI shift toward real impact on society and the economy, Japan's industrial structure can become a strength

- While compute resources are important, as expectations for AI shift from benchmark performance toward real impact on society and the economy, the scale of data that can be generated and obtained in actual socioeconomic activities is expected to become more important.
 - Japan's structural tendency to accumulate substantial volumes of data critical to AI development and deployment, underpinned by its manufacturing industrial base, is a strength

Japan's strengths in AI development and deployment

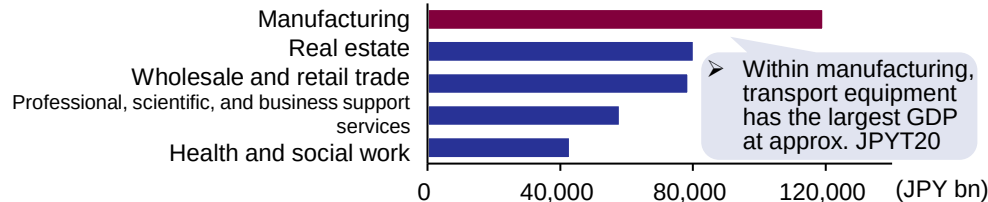
Share of manufacturing in GDP (2023, value-added basis)

- Japan is at a high level even in international comparison



Domestic GDP by industry (2024, top 5 sectors)

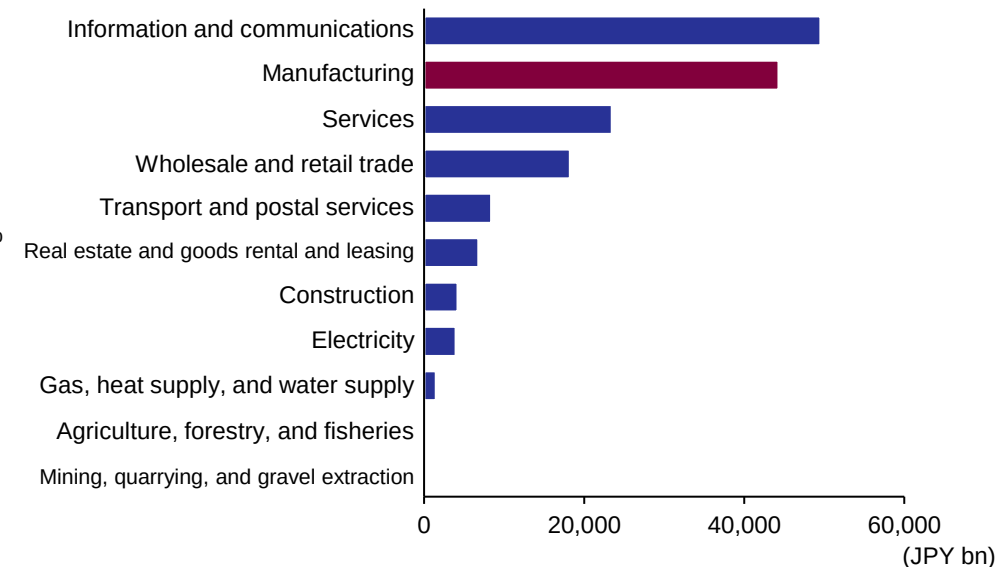
- The absolute scale of manufacturing, particularly the automobile industry as a leading industry, is large. Strengths in production capacity and technological capabilities in specific fields are evident.



- Within manufacturing, transport equipment has the largest GDP at approx. JPYT20

Software assets by industry (FY2024)

- Manufacturing has the second-largest software assets after information and communications, suggesting that data may be abundant compared with other industries



A structural tendency to accumulate substantial volumes of data critical to AI development and deployment, underpinned by the manufacturing industrial base

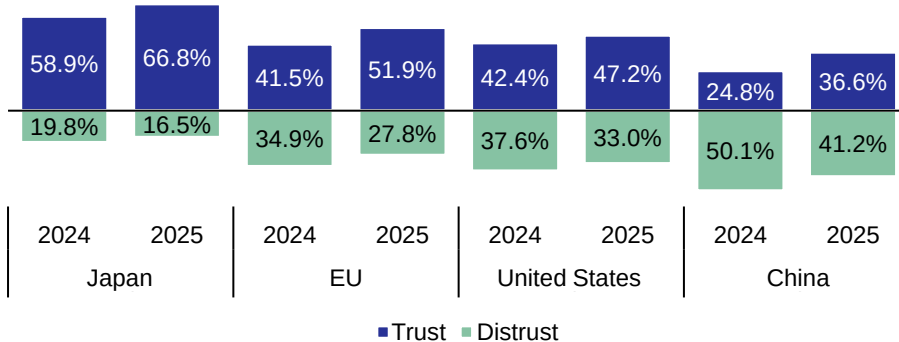
Source: Compiled by Industry Research Department, Mizuho Bank, Ltd. based on UNdata (upper left chart), "National Accounts" by the Cabinet Office (lower left chart), and "Financial Statements Statistics of Corporations by Industry" by the Ministry of Finance (right chart)

Building a co-creation framework with ASEAN in the AI domain, where industrial collaboration foundations are in place

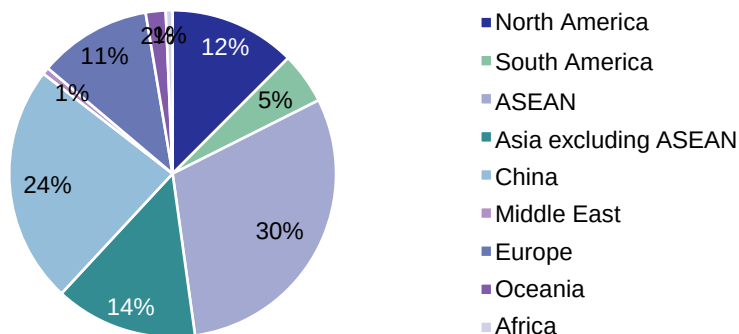
- Japan has built robust friendly relations with ASEAN through more than 50 years of dialogue, maintaining the No. 1 position for seven consecutive years in ASEAN's trust survey. In addition, ASEAN is the region with the largest number of local affiliates of Japanese companies, and the established foundation for industrial collaboration is a strength.
- In October 2025, the "Japan-ASEAN AI Co-Creation Initiative" was launched with the aim of strengthening government-to-government and industry-government-academia collaboration between Japan and ASEAN in the AI domain, advancing the framework toward building a safe, secure, and trustworthy AI ecosystem.
 - Given AI usage rates in ASEAN, the industrial collaboration foundation and government-to-government cooperation can potentially be leveraged as Japan's strengths

Japan's presence in ASEAN

Asean's trust in major countries



Number of Japanese corporate affiliates by region



Source: Compiled by Industry Research Department, Mizuho Bank, Ltd. based on ISEAS "The State of Asia 2025 Survey Report" and "Basic Survey on Overseas Business Activities" by the Ministry of Economy, Trade and Industry (METI)

Overview of the ASEAN-Japan AI co-creation initiative



Strengthening government-to-government and industry-government-academia collaboration between Japan and ASEAN in the AI domain

Pillar 1: Co-creation of AI solutions

- Launch of a co-creation platform for resolving social issues
- Promoting issue resolution by utilizing JICA's private-sector partnership programs and other initiatives

Pillar 3: Human resource development and capacity building

- Human resource development and job matching related to AI development and operation
- Agreements concerning safety (including cybersecurity, etc.)

Pillar 2: Institutional development and governance

- Cooperation on AI strategy and roadmap formulation, and on legal system development
- Strengthening cooperation through the Hiroshima AI Process Friends Group

Pillar 4: Infrastructure support

- Support for research and development of AI models tailored to local languages and cultures
- Support for data integration

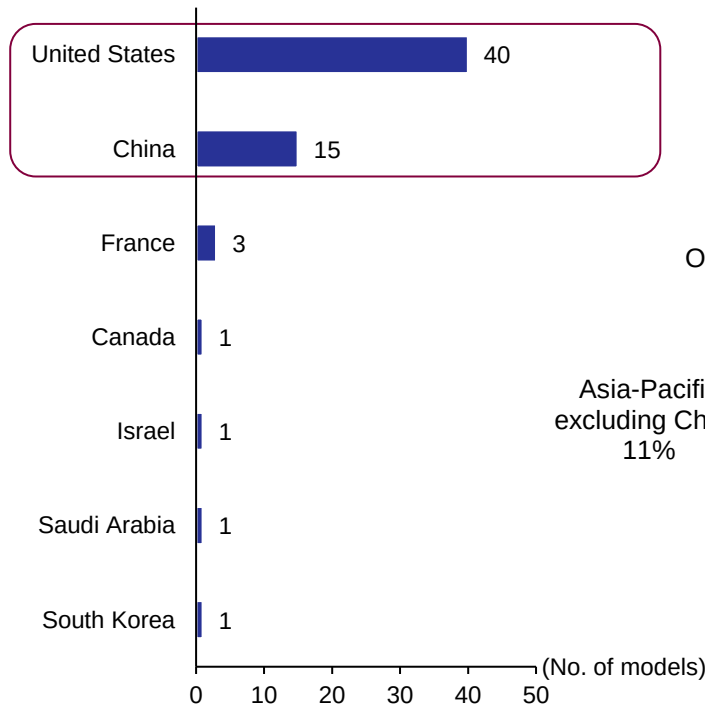
Source: Compiled by Industry Research Department, Mizuho Bank, Ltd. based on publicly available materials from the Ministry of Foreign Affairs of Japan

Japan lags behind the U.S. And China in high-performance AI model development; securing compute resources and talent is a challenge

- Japan has also embarked on AI model development; however, against the backdrop of the availability of large-scale compute resources—indispensable for AI development—and the depth of the talent pool producing cutting-edge research outcomes, the U.S. and China have a larger number of notable AI models.
 - In addition, Japan is in a position requiring catch-up even in comparison with other countries, and securing and reinforcing compute resources and talent for AI development is considered to be a challenge

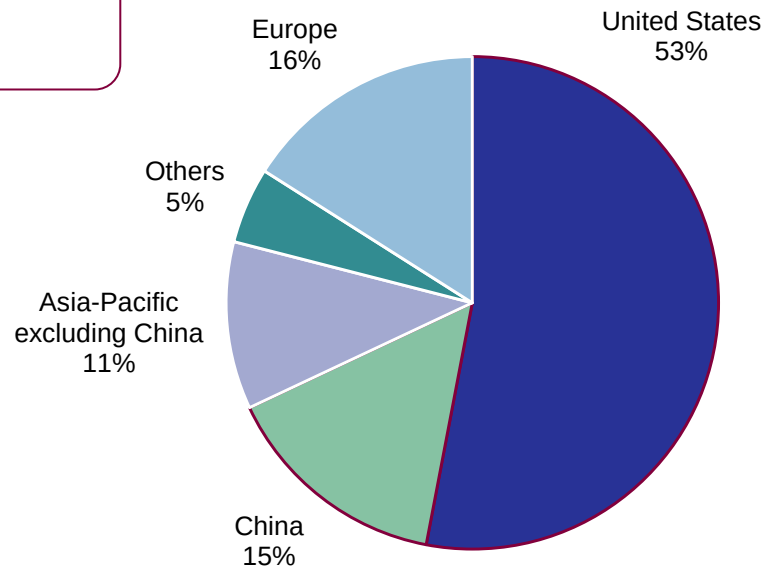
Comparison of regional characteristics in AI model development

Regional comparison of the number of notable AI models (2024)



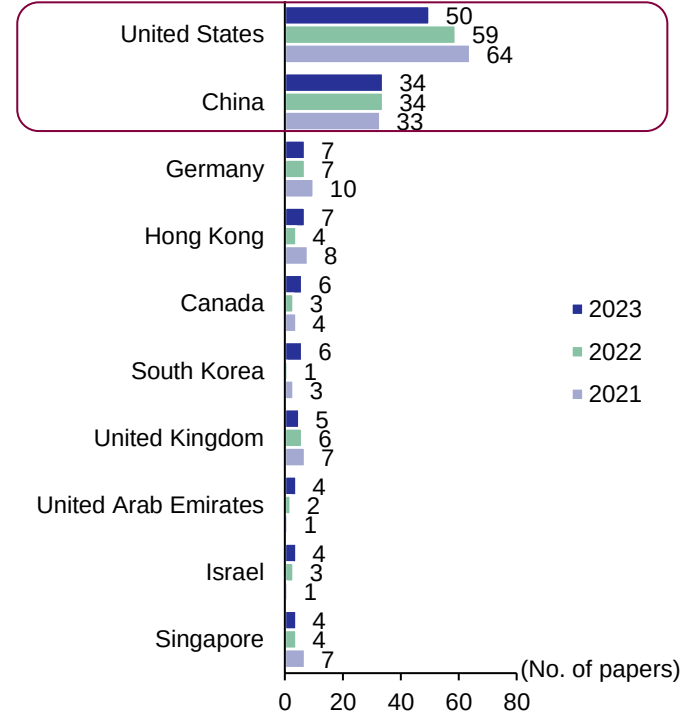
The U.S. and China have developed many notable models compared with other countries

Regional share of data centers (data capacity): Q2 2022



The U.S. and China have abundant compute resources

Regional comparison of the number of Top 100 most-cited AI papers



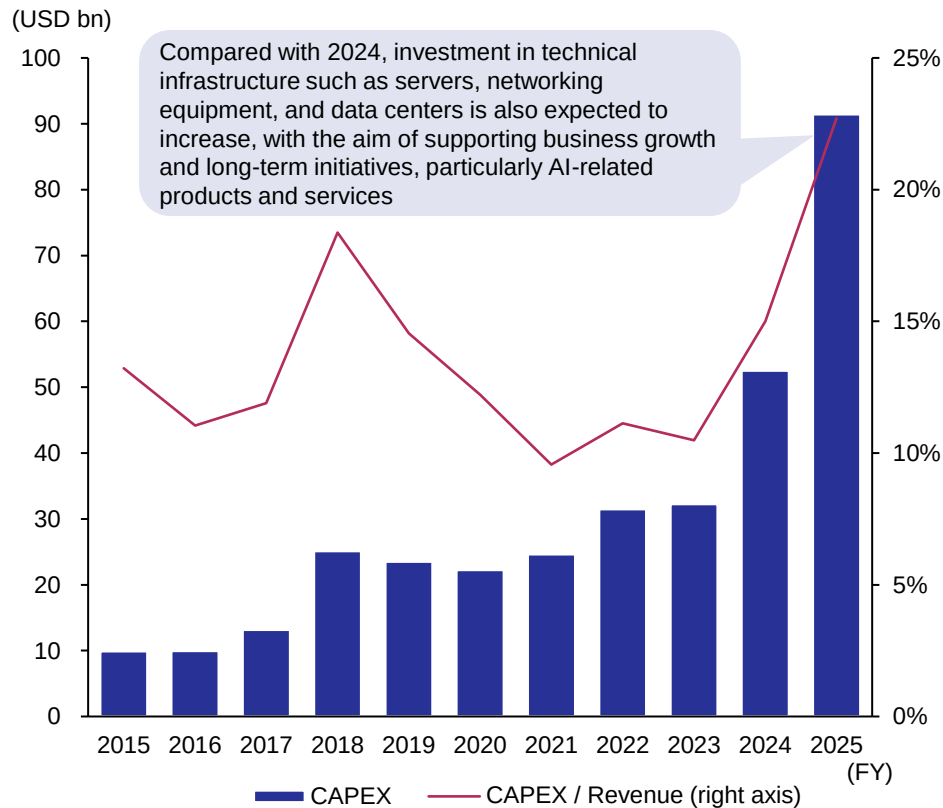
Talent is also abundant

Source: Compiled by Industry Research Department, Mizuho Bank, Ltd. based on "The AI Index 2025 Annual Report," AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2025 (left chart and right chart; full citation provided at the end of this report) and Synergy Research Group (center chart)

Capital-intensive development competition: large-scale investment is required to secure competitive advantage

- Among the elements required for development, compute resources are being built out by U.S. technology companies such as Google and Meta through massive capital expenditure, resulting in a capital-intensive competitive structure
 - Talent acquisition also requires substantial compensation; for Japan to establish and maintain international competitive advantage in AI development, large-scale investment in both compute resources and talent may be required

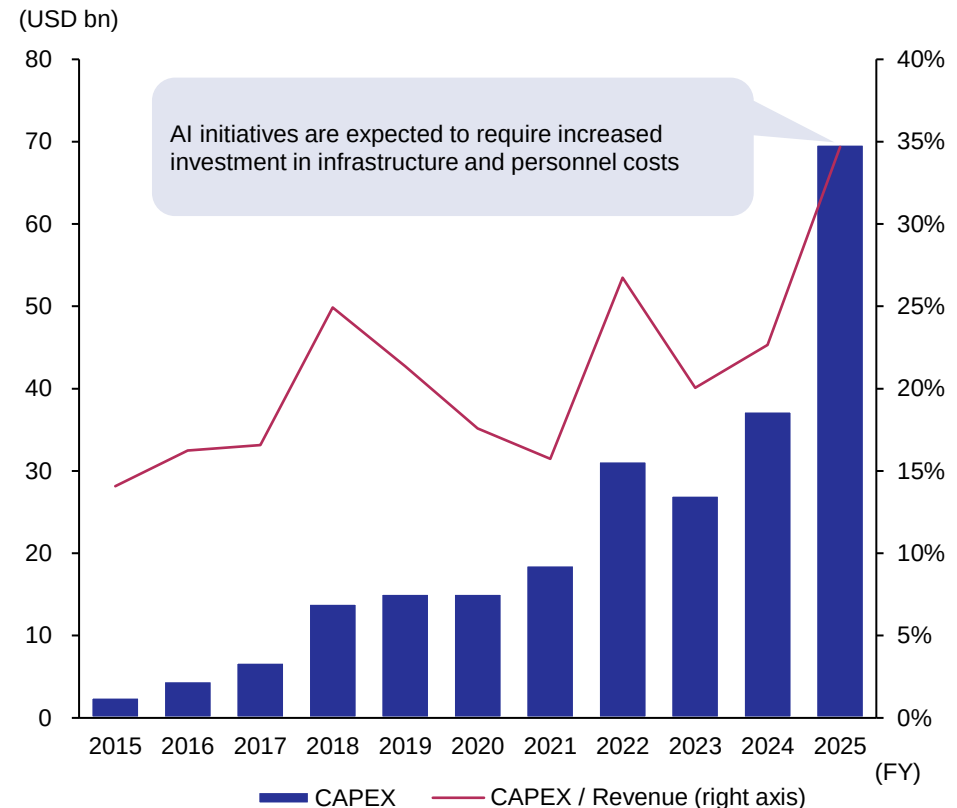
Trends in capital expenditure by google (alphabet inc.)



Note 1: CAPEX is based on the absolute value of "Acquisition and sale of property, plant and equipment" from SPEEDA

Source: Compiled by Industry Research Department, Mizuho Bank, Ltd. based on annual and quarterly reports of Alphabet Inc.

Trends in capital expenditure by meta



Note 1: CAPEX is based on the absolute value of "Acquisition and sale of property, plant and equipment" from SPEEDA

Source: Compiled by Industry Research Department, Mizuho Bank, Ltd. based on annual and quarterly reports of Meta Platforms, Inc.

Establishing a technology base through the realization of sovereign AI via public-private partnerships is necessary to maintain national interest and industrial competitiveness

- As AI becomes a key technology indispensable to national interest and industrial competitiveness, establishing a technology base through the realization of Sovereign AI is necessary
- An architecture is envisioned in which the government leads the development of base models presumed to have broad applicability across diverse use cases, while the application of these base models to specific use cases is advanced through public-private partnerships tailored to the characteristics of each field and domain.
 - In general, the public sector is expected to take the lead in highly public fields and domains such as defense and healthcare
 - On the other hand, for industrial fields and domains such as manufacturing, base models are specialized using industry-specific data to differentiate them from overseas AI. The intent is to deploy AI unique to Japan and thereby strengthen the international competitiveness of industries and businesses.

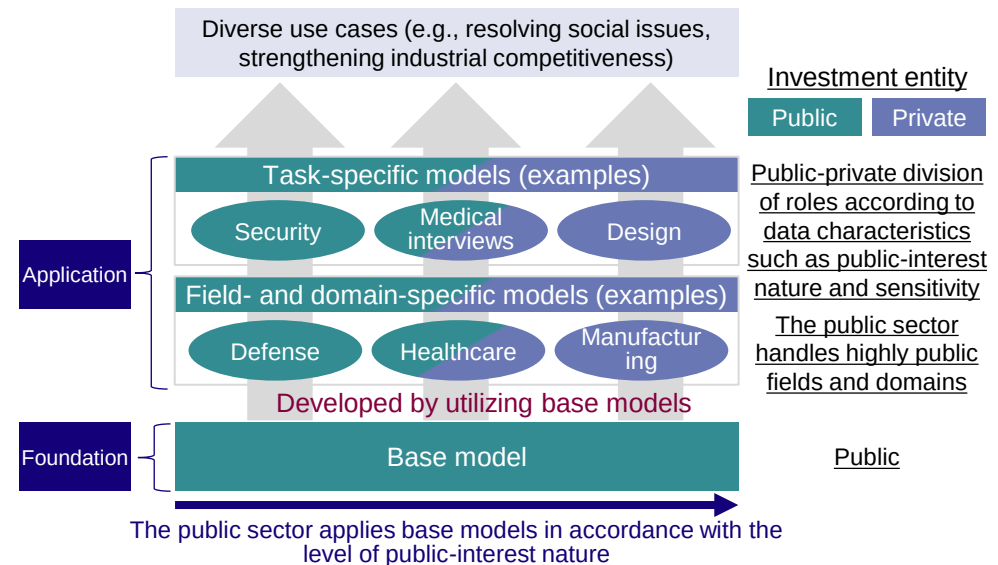
The concept of sovereign AI

Sovereign AI	
A state or concept in which a nation independently develops, operates, and manages the core elements of AI without relying on other countries or external organizations	
Core elements	Rationale for independent development, operation, and management
Data	➤ Data that is fragmented, unevenly distributed, and scattered across both the public and private sectors needs to be organized so that it can be effectively used for AI training and inference
Model	➤ A cutting-edge AI development environment may not be cultivated, talented AI researchers and engineers may flow out overseas, and the foundation for technology development and the ecosystem may fail to form
Compute resources	➤ If data center development does not progress domestically, the digital deficit risks expanding alongside AI proliferation ➤ As for sensitive data, domestic data needs to be managed under the authority of the home country and domestic organizations

Source: Compiled by Industry Research Department, Mizuho Bank, Ltd.

Basic philosophy for the realization of sovereign AI

Model development and application through public-private partnerships
To pursue catch-up with technological evolution and the development of highly competitive services while ensuring consistency with national policies such as security, investment in both the foundation and applications is promoted through public-private partnerships



Source: Compiled by Industry Research Department, Mizuho Bank, Ltd.

Strengthening private-sector AI capabilities and transforming business models through the construction of sovereign AI

- Given the highly public nature and investment risks involved, the public sector is expected to provide financial support and oversight for compute resource procurement and base model development. On the other hand, it is essential for the private sector to leverage public-sector support and aim to acquire and reinforce capabilities in specific model development and the data layer.
- The public sector establishes base models and compute resources as public goods, while also ensuring a safe and equitable utilization framework for data. On the private-sector side, to enable competition through value creation via on-the-ground application, open-source models and the like are utilized in the short term to enhance development and operational capabilities.

Concept of role-sharing between the public and private sectors for the construction of sovereign AI

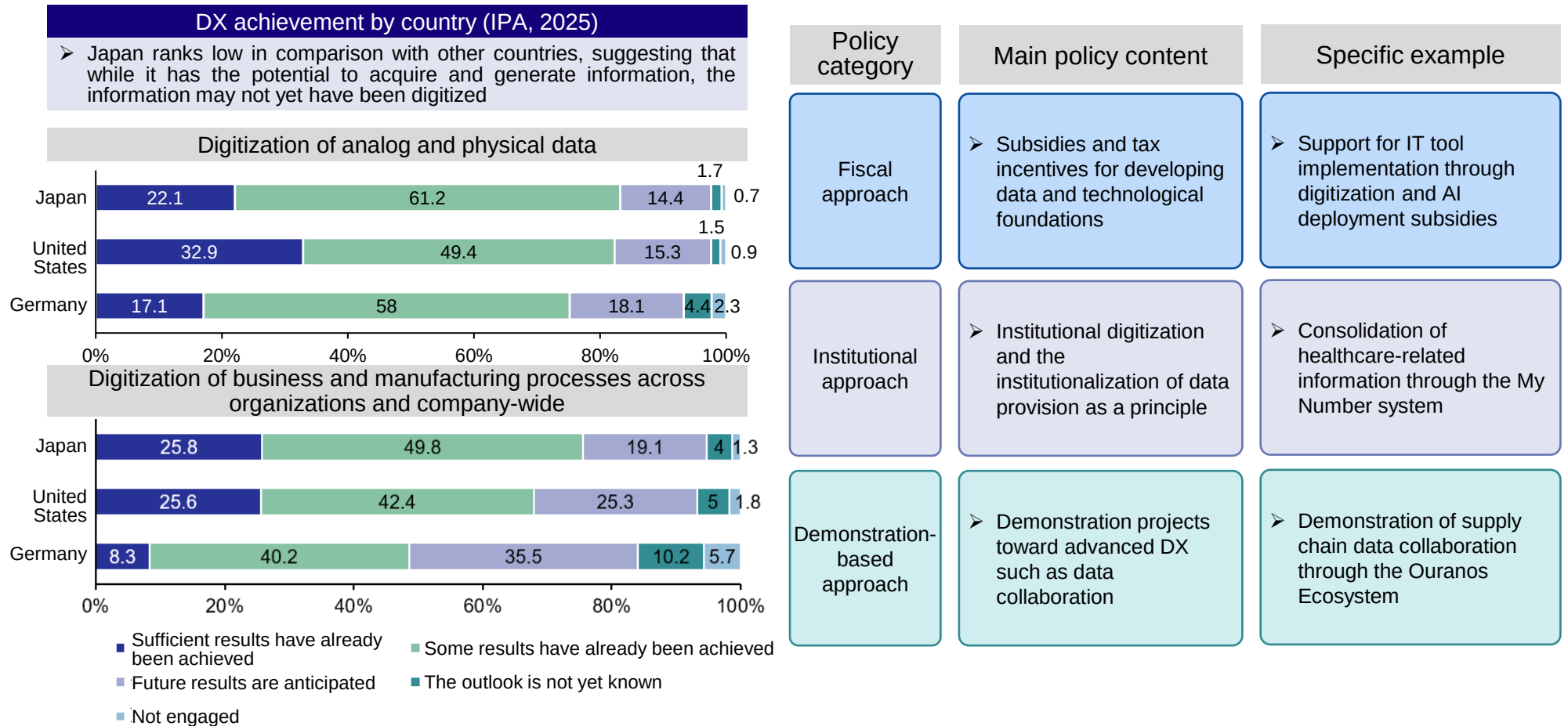
	Investment and oversight	Short term (–2030)	Medium term (2031–)
specific models (applications)	Private	➢ Develop domain-specific models by utilizing open-source software (OSS) models and explore use cases ➢ Through the development of domain-specific models optimized for each use case, cultivate digital skills and pursue transformation toward business models premised on AI	➢ Leveraging OSS model utilization capabilities to develop specific models built on Sovereign AI base models ➢ Establish de facto standard specific model groups by industry and translate this into export competitiveness of Japanese companies
Data		➢ [Private] Develop in-house data infrastructure and carve out shareable data in accordance with public-sector standard contracts ➢ [Public] Machine-readability conversion, metadata development, and update management for administrative documents, laws and regulations, guidelines, statistics, judicial precedents, and the like ➢ [Public] Develop standard contracts for data provision (licensing, division of responsibilities, audit, and re-provision permissibility)	➢ [Private] Review competitive and cooperative domains in light of industrial structural transformation, and build a cross-organizational data integration platform ➢ [Public] Expand a framework that balances privacy protection and protection of sensitive information with the usefulness of training data ➢ [Public] Build training datasets with ASEAN countries by utilizing the framework of the AI Co-Creation Initiative
Base model	Public	➢ Achieve both technological catch-up and human resource development through beta-version base model development utilizing foreign cloud services	➢ Develop release-version base models using domestic data centers
Compute resources		➢ Securing financial resources for upfront investment and promoting data center development	➢ Diversification of procurement to prepare for supply disruption and export restriction risks

Source: Compiled by Industry Research Department, Mizuho Bank, Ltd.

A policy-based approach can potentially be effective in establishing the data foundation

- Advancing data collection and utilization through public-private partnerships is important for establishing the data foundation; however, against the backdrop of delays in DX response and other factors, the utilization of data presumed to exist abundantly in the private sector remains a barrier
 - For data utilization, policy-based approaches such as financial support for the technology base and institutional digitization in sensitive domains such as social security can be effective

Current status of DX in Japan and policy examples for data foundation development



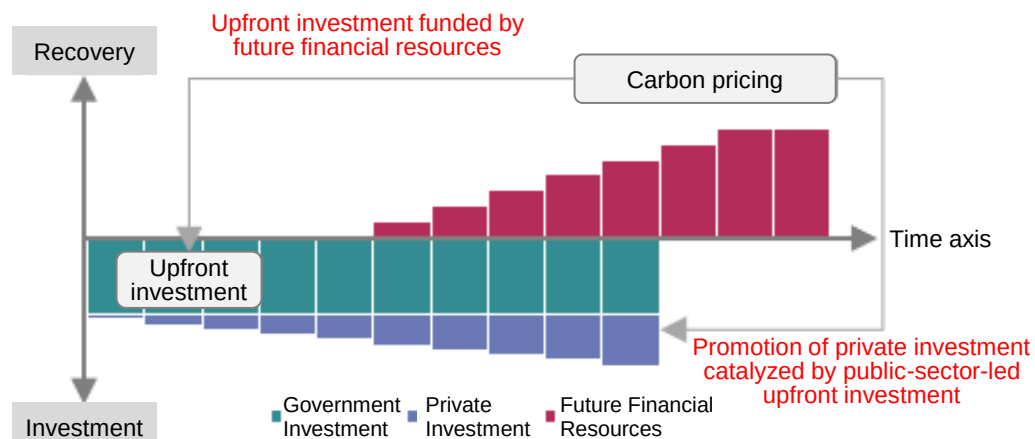
Source: Compiled by Industry Research Department, Mizuho Bank, Ltd. based on "DX Trends 2025" by IPA (Information-technology Promotion Agency, Japan)

Policy mechanisms and logic design serve as the course of action against barriers posed by large-scale upfront investment accompanying the construction of sovereign AI

- To advance Sovereign AI development under public-sector leadership, public-sector upfront investment in compute resources—the principal bottleneck—is required
- For the financial resources to justify upfront investment, one option is to regard the future outcomes generated by utilizing Sovereign AI as future financial resources, with reference to GX Economy Transition Bonds

Conceptual mechanism for generating investment financial resources with reference to GX economy transition bonds

GX economy transition bonds		Application toward the construction of sovereign AI	
P	■ Agreement on initiatives toward realizing GX globally, triggered by the Paris Agreement	P	■ The importance of AI is increasing as a key technology that affects economic and national security.
E	■ Realizing GX in Japan requires over JPYT150 in investment over the next 10 years, and the high risk involved is making private investment difficult	E	■ AI development is a capital-intensive development competition premised on scaling laws, requiring massive investment funds.
S	■ Since GX is expected to benefit future generations, intergenerational equity needs to be ensured	S	■ From the perspective of avoiding future digital deficits and geopolitical risks, benefits are expected to extend into the future.
T	■ Realizing GX is difficult through extensions of existing technology; research and social implementation of innovative technologies with high development risk are indispensable	T	■ AI is a next-generation infrastructure, and catch-up with innovative technologies in the AI ecosystem and their social deployment are indispensable.



Concept of GX economy transition bonds

By issuing "GX Economy Transition Bonds" backed by future financial resources obtained as a result of introducing carbon pricing, bold upfront investment is supported at the national level

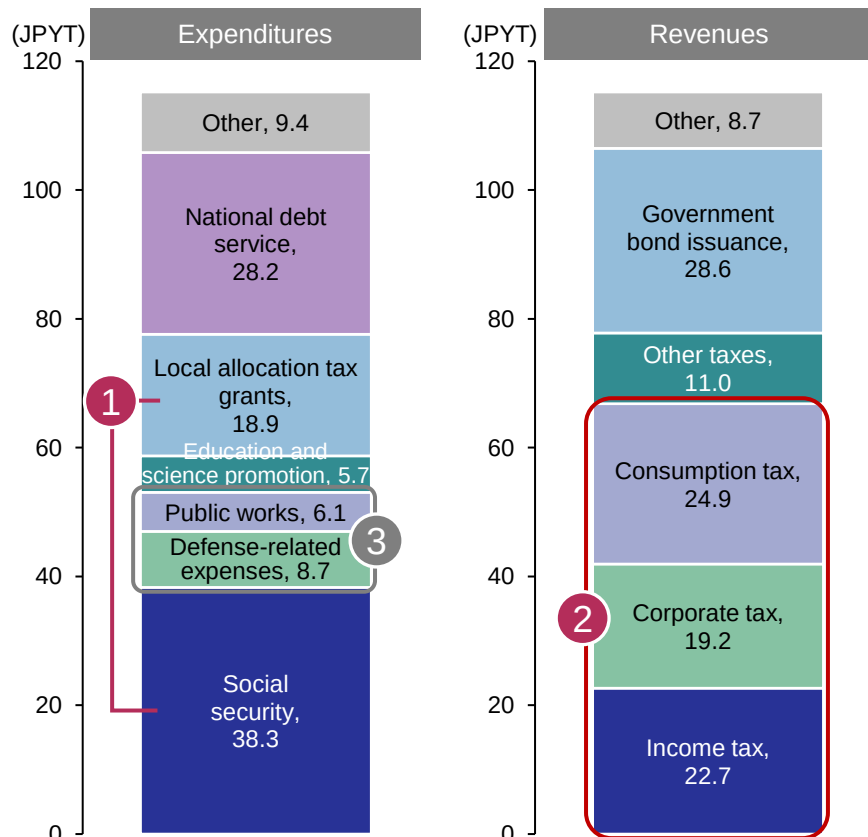
By treating the outcomes generated through the utilization of Sovereign AI as future financial resources, a mechanism similar to GX Economy Transition Bonds can be established to advance upfront investment

Source: Compiled by Industry Research Department, Mizuho Bank, Ltd. based on publicly available materials from the Ministry of Economy, Trade and Industry (METI)

For outcomes treated as future financial resources, emphasis is placed on direct impact on expenditures

- As both direct and indirect outcomes are anticipated for the Japanese government, institutional design that carefully assesses the certainty of such outcomes as future financial resources is important
 - Expenditure containment effects on items such as social security and local allocation tax grants can be recognized as direct outcomes
 - However, monitoring mechanisms (such as setting KPIs) to evaluate the containment amount attributable to Sovereign AI need to be examined
 - Strengthening industrial competitiveness through the utilization of Sovereign AI can be realized as an indirect outcome in the form of increased tax revenues such as corporate tax

Conceptual image of direct and indirect outcomes from sovereign AI



Source: Compiled by Industry Research Department, Mizuho Bank, Ltd. based on the "FY2025 Budget" by the Ministry of Finance

1 Direct economic impact

- Against the backdrop of demographic trends, as the burden of social security and local allocation tax grants increases, reducing and containing expenditures translates directly into economic value
- Through the containment of medical expenses and local administrative service maintenance costs by utilizing Sovereign AI, future direct outcomes can be generated

2 Indirect economic impact

- Indirect economic value such as increased tax revenues can be enjoyed through manufacturing—one of Japan's strengths—and particularly through the production of autonomous vehicles
- In addition, Sovereign AI is expected to contribute to containing the digital deficit projected to expand in the future, and ecosystem export to ASEAN through the AI Co-Creation Initiative is also anticipated

3 Impact on the national foundation

- Defense and public works are strategic investments that support the national foundation, including national security and infrastructure maintenance, contributing to the strengthening of national resilience
- As AI utilization advances in domains such as defense and social infrastructure, social outcomes can be generated by leveraging trustworthy Sovereign AI

Leveraging Japan's strengths and the capabilities accumulated through the construction of sovereign AI to deploy in ASEAN

- In its relationship with ASEAN, Japan possesses strengths including (1) a high level of trust between nations, (2) a robust foundation for industrial collaboration anchored by local subsidiaries, and (3) the AI Co-Creation Initiative. Know-how and institutional frameworks cultivated through the domestic construction of Sovereign AI can be horizontally deployed to ASEAN.
- With the guarantee of sovereignty—the basic philosophy of Sovereign AI—as the greatest differentiating factor, realize value co-creating AI deployment through public-private partnerships, contributing to both the growth of Japanese industry and the resolution of social issues on the ASEAN side while also utilizing ASEAN capital and assets
 - Promote initiatives in an integrated manner, from developing compute resources and jointly developing models to building applications through use-case exploration

Course of action for ASEAN deployment leveraging Japan's strengths

	Contributing strengths	Public	Private
Application	(2) Industrial collaboration foundation (3) Co-creation framework	➤ Within the Co-Creation Initiative, specifying industries—particularly manufacturing—where both Japan's and ASEAN's strengths can be leveraged as deployment domains, and establishing de facto standards	➤ With local subsidiaries serving as the deployment field, developing optimal applications tailored to localized workflows and deploying them through the industrial collaboration foundation as a sales channel
Data	(1) Trust between nations (2) Industrial collaboration foundation (3) Co-creation framework	➤ Within the Co-Creation Initiative, conducting rule-making on data standardization premised on AI operation and on the permissibility of cross-border data flows	➤ Horizontally deploying data development know-how cultivated through the domestic construction of Sovereign AI to local subsidiaries, accelerating the cycle through to operation
Model	(1) Trust between nations (3) Co-creation framework	➤ Jointly developing model requirements for sovereign operation (such as auditability and safety) and establishing a framework for joint evaluation and testing of models	➤ Developing an operational environment that includes peripheral technologies such as Agentic AI and RAG, with a view to early deployment of jointly developed models
Compute resources	(1) Trust between nations (3) Co-creation framework	➤ Providing packages premised on continuous operation, including security, BCP, and operational human resource development ➤ Economic support such as ODA is also an option	➤ In addition to the energy domain, which is a traditional strength, providing AI-optimal energy management through integrated operation with data centers

Source: Compiled by Industry Research Department, Mizuho Bank, Ltd.

Phased investment is required for the development of sovereign AI base models

- Development of Sovereign AI base models requires investment in compute infrastructure for training models and in talent and electricity for actually conducting modeling and training. In addition, electricity costs required for inference also arise as running costs.

Logic for estimating the investment amount required for the construction of sovereign AI

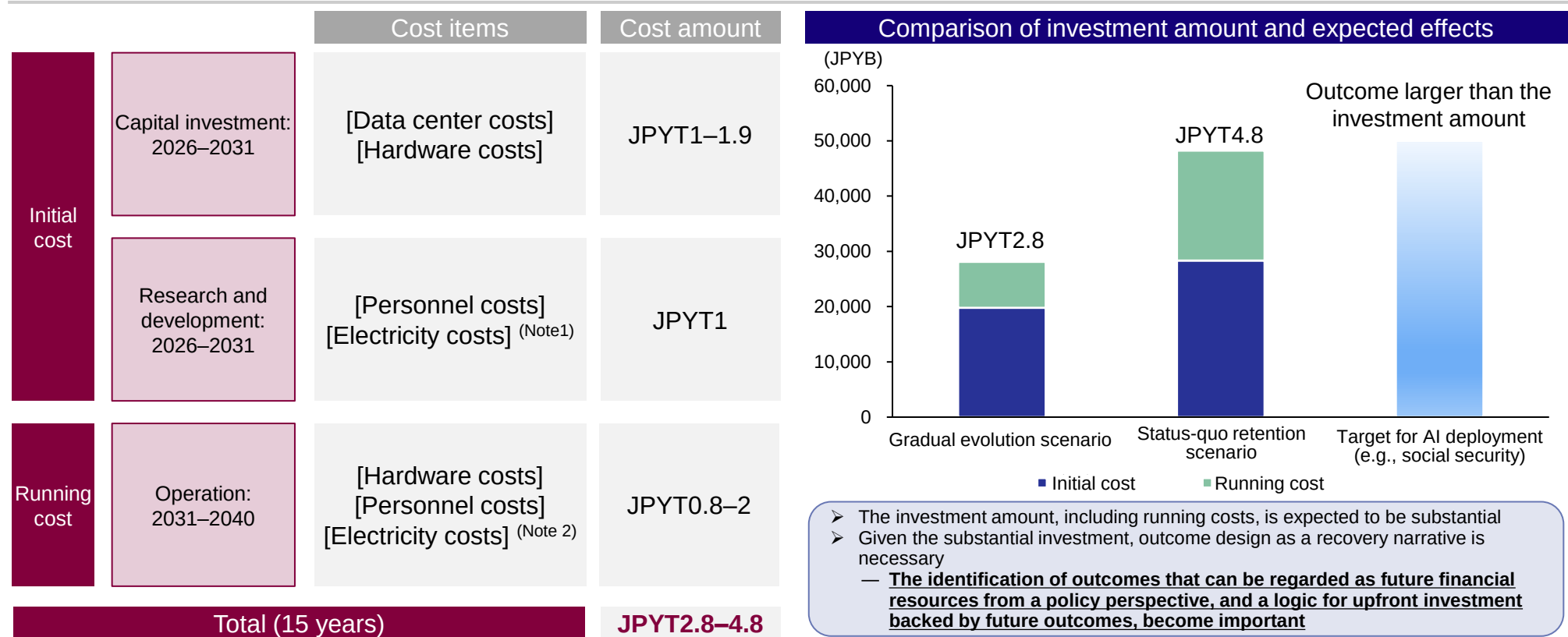
Assumptions for estimating development investment amount			
Model performance (parameter size)		Hardware scale	
➢ Develop a model (base model) with a parameter size of 5 trillion as a proxy variable for model performance: development from 2026 to 2031 is assumed, premised on the foundation model paradigm. Continued scaling of parameter size is assumed, with the effectiveness of new alternative paradigms expected to remain immature. ➢ The size assumed for notable models such as OpenAI's GPT-4 and later		➢ Estimate the total number of GPUs—the hardware required to perform the compute volume needed for base model development within a given period (e.g., three months) ➢ Two scenarios are set for the evolution of compute performance—status-quo retention and gradual (linear) evolution—to estimate a cost range	
Cost items		Logic for calculating costs	
Initial cost	Capital investment: 2026–2031	[Data center costs]	➢ Assumed to be the infrastructure construction cost required to support the electricity consumed by GPUs in data centers ➢ Electricity and cooling: Estimated based on the per-watt construction cost in Tokyo and the electrical load (in watts) of the GPU cluster ➢ Network: Assumed to be a fixed ratio of hardware costs
		[Hardware costs]	➢ Assumed to be the cost of deploying the GPU cluster ➢ Estimated based on the price of NVIDIA's H100 and the GPU cluster to be deployed (total number of GPUs)
	Research and development: 2026–2031	[Personnel costs]	➢ Assumed to be the cost required to acquire and retain talent (top talent) capable of carrying out base model development and the application to highly public fields and domains (specific models) using cutting-edge methods. This also includes dataset development. ➢ Estimated using OpenAI as the benchmark, based on the company's estimated number of employees and personnel cost per employee
		[Electricity costs]	➢ Estimate the electricity cost involved in base model development and the application to highly public fields and domains (specific models) ➢ Estimated based on the total electricity volume (kWh) required for training and the electricity price (JPY/kWh). Development takes experimentation into account, with the parameter size ultimately reaching 5 trillion. Application costs are assumed to be a fraction of the base model (e.g., 1/x).
Running cost	Operation: 2031–2040	[Hardware costs]	➢ Considering wear of GPUs used in development, assumed to be the cost of newly deploying GPU clusters and the like (assumed for 2036)
		[Electricity costs]	➢ Estimate the electricity cost required for model retraining (assumed for 2036) and for the operation of specific models in highly public fields and domains ➢ Estimate the theoretical maximum electricity consumed by the entire data center (total electricity consumption): calculated based on the total number of hardware units multiplied by the electricity consumed by GPUs at peak load. Electricity for servers, data center cooling, and the like is also taken into account. ➢ After applying an average utilization rate to total electricity consumption (kWh), the annual electricity cost is calculated based on the electricity price (kWh). For the electricity price, annual price increases due to inflation are taken into account.

Source: Compiled by Industry Research Department, Mizuho Bank, Ltd. based on publicly available information

Outcome design is important as the narrative justifying large-scale investment

- Investment funds, including running costs, are estimated at JPYT2.8–4.8 in total over the 15 years from 2026 to 2040
 - To avoid letting large-scale investment end merely as expense, the identification of outcomes that can be regarded as future financial resources from a policy perspective, and a logic for upfront investment backed by future outcomes, become important

Estimate of the investment amount required for the sovereign AI strategy



Note 1: This report assumes “defense” and “healthcare” as generally highly public fields. (1) For each, one field- and domain-specific model is trained. (2) In addition, as for task-specific models, one model is trained in the defense field, and three models are trained in the healthcare field for solution development in the “AI- and data-driven healthcare system” (refer to the individual Promising Domain). (1) and (2) are recorded in addition to the development cost of the base model.

Note 2: For operation, one model in the defense field and three models in the healthcare field are assumed (field- and domain-specific models are used only during training and therefore are excluded from active models). Of the annual electricity consumption calculated based on theoretical values, one-quarter is allocated to the defense field as electricity cost for operation. For the healthcare field, the electricity cost for operation is estimated based on the operational forecast of the “data-driven healthcare system.”

Source: Compiled by Industry Research Department, Mizuho Bank, Ltd. based on publicly available information

Economic effects exceeding upfront investment are expected in both loss containment and profit expansion

- Sovereign AI is expected to contribute on two fronts: addressing Japan's social issues and enhancing industrial competitiveness in sectors such as the automotive industry
 - The effect of containing economic losses such as social security costs and the digital deficit is estimated at approx. JPYT20, while the revenue expansion effect from the industrial competitiveness perspective is estimated at approx. JPYT11, suggesting that the upfront investment can be sufficiently recovered
 - The start of utilization of base models and specific models is assumed to be from 2031 onward. The government's medium-to-long-term support policy and investment framework have already been clarified (Note 1), suggesting consistency with policy.

An example of expected effects on Japanese society generated by sovereign AI (2040)

Loss containment effect			Revenue expansion effect		
Use case / outcome	[Healthcare] Containment of social security costs	JPYT13/year (rationalization of outpatient medical expenses)	Use case / outcome	[Vehicles] Establishing competitive advantage in the autonomous driving market	JPYT2.7/year (economic ripple effect of the automotive industry)
Logic for calculating outcomes	➢ The medical expenses contained by rationalizing the frequency of inpatient and outpatient visits, a structural challenge in the healthcare delivery system, through digital solutions such as medical-interview AI are assumed as the outcome • Outpatient visit frequency Japan: 11.1/person ⇄ G7 average: 5.7/person		Logic for calculating outcomes	➢ The domestic production of Japanese OEMs (for both domestic and overseas demand) maintained through catch-up with autonomous driving technology, and the resulting domestic economic ripple effect, are assumed as the outcome ➢ The new vehicle sales market size is calculated based on the autonomous driving penetration rate (2040) in each country, applied to Japan's domestic demand and exports to the U.S. (overseas demand)	
Use case / outcome	[Digital expenditure] Prevention of outflow of national wealth	JPYT6.9/year (containment of AI-related overseas expenditure)	Use case / outcome	[Model export] Model export to ASEAN	JPYT8.1/year (Note2) (productivity improvement through AI)
Logic for calculating outcomes	➢ The further expansion of the digital deficit anticipated as AI becomes a key technology is contained through the realization of Sovereign AI ➢ The difference in deficit amounts between the risk scenario and the base scenario is assumed as the loss avoidable through the realization of Sovereign AI		Logic for calculating outcomes	➢ For ASEAN-5 major industries (manufacturing, retail and wholesale, agriculture/forestry/fisheries), the annual AI deployment rate is estimated based on GDP and occupational composition ratios for each country and industry, taking into account the deployment difficulty by occupation ➢ The outcome is assumed to be the value-added amount for each ASEAN-5 major industry multiplied by the productivity improvement rate associated with the AI deployment rate	

Note 1: The "Framework for Strengthening the Foundation of the AI and Semiconductor Industries," formulated in November 2024, declares the implementation of over JPYT10 in public support for the semiconductor and AI sectors over the seven years through FY2030, the catalyzing of over JPYT50 in combined public and private investment over 10 years using such support as the trigger, and the targeting of approx. JPYT160 in economic ripple effects.

Note 2: This profit amount is not directly obtained by Japan as consideration for model export; rather, it refers to the productivity improvement expected as a result of ASEAN countries deploying the models. By assuming that the productivity improvement is allocated to future investment, this can be regarded as an expansion of TAM (Total Addressable Market) for Japan; therefore, this report categorizes it as an economic profit expansion effect.

Source: Compiled by Industry Research Department, Mizuho Bank, Ltd.

References

- Nestor Maslej, Loredana Fattorini, Raymond Perrault, Yolanda Gil, Vanessa Parli, Njenga Kariuki, Emily Capstick, Anka Reuel, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, Juan Carlos Niebles, Yoav Shoham, Russell Wald, Toby Walsh, Armin Hamrah, Lapo Santarlaschi, Julia Betts Lotufo, Alexandra Rome, Andrew Shi, Sukrut Oak. “The AI Index 2025 Annual Report,” AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2025. <https://doi.org/10.48550/arXiv.2504.07139>

Industry Research Department

Next-Generation Business Support Office
Strategic Project Team

Asia Office

Yuki Saito

Yu Maeshima

Midzuki Low

yuki.c.saito@mizuho-bk.co.jp

Mizuho Industry Research／80 2026

Published March 31, 2026

© 2026 Mizuho Bank, Ltd.

This document has been prepared solely for the purpose of providing information. This document is not recommendation for sales. This document has been prepared based on information believed to be reliable and accurate. The Bank accepts no responsibility for the accuracy or appropriateness of such information. Upon using this document, if considered appropriate, or if necessary, please consult with lawyers, CPAs and tax accountants. This document may not be altered, reproduced or redistributed, or passed on to any other party, in whole or in part, without the prior written consent of Mizuho Bank, Ltd.

Edited / issued by Industry Research Department, Mizuho Bank, Ltd

1-3-3 Marunouchi, Chiyoda-ku, Tokyo ird.info@mizuho-bk.co.jp