ROBOT TRANSFORMATION DRIVEN BY ROBOT FOUNDATION MODELS

— AI THAT SEES, THINKS, AND ACTS TRANSFORMS PHYSICAL-WORLD DATA INTO ASSETS —

Daisuke Yamada
Consumer Innovation Dept., Technology & Innovation Studies Div.
Mitsui & Co. Global Strategic Studies Institute

SUMMARY

- Structural labor shortages are progressing not only in Japan but worldwide. Even as digital transformation (DX) advances, the automation of physical work is likely to remain a major bottleneck to productivity improvement. To address this challenge, robot transformation (RX) is required. RX extends DX into the physical world and redesigns operations and business models on the premise of utilizing robots.

- Introducing robots into areas where conventional industrial robots struggle requires more general-purpose robots. One of the key enabling technologies supporting this shift is the robot foundation model (VLA/Physical AI).

- Japan has many operational environments with stringent quality and safety requirements, making it well-suited for refining AI robots through practical use. Whether real-world data from these settings can be transformed into assets will determine the adoption and competitiveness of RX.

1. BACKGROUND: STRUCTURAL LABOR SHORTAGES AND THE NECESSITY OF RX

1-1. DEEPENING LABOR SUPPLY CONSTRAINTS AND THE BARRIER OF PHYSICAL WORK

Declining birthrates, aging populations, and stagnant labor force growth are occurring globally, particularly in Japan, as well as Europe and the US.[1] Among OECD member countries, the working-age population (20–64) is projected to decline by 8% between 2023 and 2060, with decreases exceeding 30% in one-quarter of these countries, indicating that labor supply constraints are becoming structural, particularly in advanced economies.[2] While demand remains robust for on-site physical work in sectors such as food service, logistics, nursing care, and cleaning, labor shortages persist.[3] Amid ongoing wage increases across a wide range of industries in the US, average hourly wages continue to rise in service sectors, including logistics, food service, and accommodation.[4] Even as generative AI improves white-collar productivity, tasks such as dishwashing, plating, picking, unloading, and bed-making cannot be transferred to digital environments, making on-site operations

[1] https://population.un.org/wpp/assets/Files/WPP2024_Key-Messages.pdf
[2] https://www.oecd.org/en/publications/2025/07/oecd-employment-outlook-2025_5345f034/full-report.html
[3] https://www.oecd.org/content/dam/oecd/en/events/2025/02/labour-shortages--evidence-and-policy-implications/Causa_soldani_labour_shortages.pdf
[4] https://www.bls.gov/news.release/empsit.t19.htm

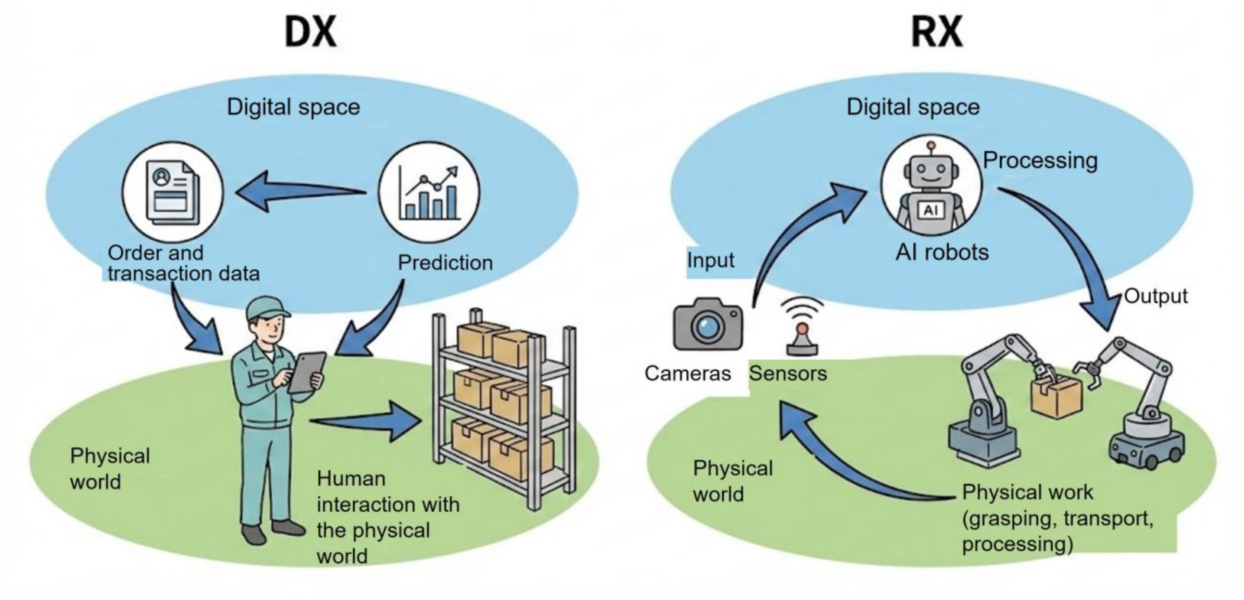
prone to becoming bottlenecks. Consequently, demand is rapidly increasing for the deployment of AI-powered robots as a means of directly intervening to supplement and substitute for human labor.[5]

1-2. EXPANSION FROM DX TO RX

The essence of introducing AI robots lies not merely in replacing human labor. Digital transformation (DX), the prevailing approach to date, refers to initiatives that leverage digital technologies to enhance operational and decision-making efficiency and drive business and organizational transformation. Robot transformation (RX) extends DX into the physical world by redesigning operations on the premise that robots will work, encompassing process design, layouts, shift planning, quality control, and safety management (Figure 1). The key is to create a cycle in which data accumulates as robots operate, enabling the optimization of models and processes. RX can be accelerated by embedding data logging and performance metrics into on-site operations and managing them with continuous improvement in mind, and once standardized, it can be readily rolled out across multiple sites.

This report distinguishes between "specialized robots," which are already widely adopted in operational settings, and "next-generation AI robots," which leverage foundation models (see Section 2-2) to add and improve tasks (Figure 2).

Figure 1: DX in digital space and RX intervening in physical space

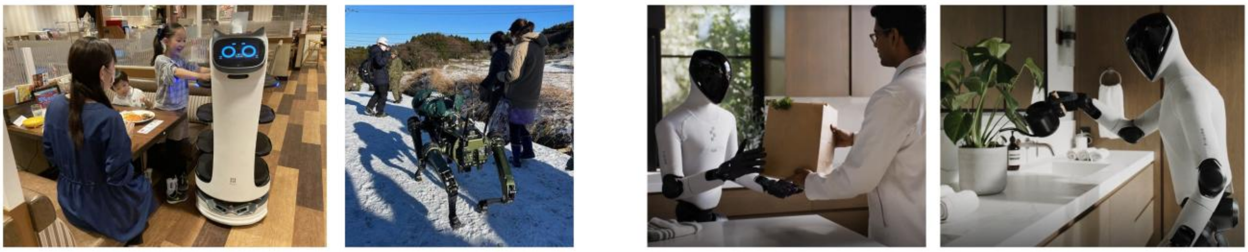


Source: Compiled by MGSSI

5 https://ifr.org/ifr-press-releases/news/service-robots-see-global-growth-boom

Figure 2: Differences between current robots and next-generation AI robots

Current robots (task-specific) Next-generation AI robots (general-purpose)



Serving robot Quadruped robot Humanoid

Source: https://www.watch.impress.co.jp/img/ipw/docs/1501/163/html/1_o.jpg.html, https://image.itmedia.co.jp/l/im/aiplus/articles/2601/13/l_tm1636144_01131_5_w490.jpg, https://www.figure.ai/ (Accessed January 22, 2025)

1-3. A USD 50 TRILLION UNTAPPED MARKET IN HIGH-MIX, LOW-VOLUME PRODUCTION

NVIDIA has stated that Physical AI, described as AI that can "see, think, and act," will transform industries such as manufacturing and logistics and create a USD 50 trillion opportunity.[6] In conventional industrial robotics, which has primarily focused on low-mix, high-volume production such as automobiles, objects were standardized, and the value lay in having robots repeatedly move along the same path at high speed and with precision. By contrast, operational settings in sectors such as food, services, and logistics involve high-mix, low-volume production, significant variation among individual items, and constantly changing work environments. In these settings, robots must operate at human-level speeds while safely and dexterously handling objects whose conditions vary, such as being soft, slippery, or wet. These areas present substantial room for automation. If untapped physical tasks can be replaced, the impact on productivity would be significant, positioning RX as a growth theme that extends beyond simply addressing labor shortages.

2. TECHNOLOGY AND BUSINESS TRENDS: AI ROBOTS AND ROBOT FOUNDATION MODELS

2-1. EXPANSION INTO HIGH-MIX, LOW-VOLUME PRODUCTION AND THREE DIMENSIONS OF GENERALITY

In high-mix, low-volume production environments, flexibility is required to accommodate variation in individual objects and changes in operating conditions. These requirements can be organized into three dimensions: (1) task generality, or the ability to switch between multiple tasks (e.g., grasp -> place -> wipe), (2) object generality, or the ability to handle items with different shapes and materials (e.g., boxes, bags, food items, trays), and (3) environmental generality, or the ability to operate despite changes in layout and traffic flow (e.g., people passing nearby, storage locations changing). Examples of tasks where deployment is advancing include unloading, depalletizing[7], and picking in logistics; multi-item picking in warehouses; transport of supplies and specimens in hospitals; and cleaning and surveillance patrols in facilities (Figure 3). Conventional teaching-centric approaches alone do not justify costs, making AI technologies that integrate perception, planning, and control through learning increasingly effective. Figure 4 summarizes representative examples of RX.

[6] https://blogs.nvidia.com/blog/nvidia-keynote-at-gtc-2025-ai-news-live-updates/

[7] The task of removing cargo from a pallet and sorting it by case or by product (palletizing refers to the reverse process)

Figure 3: Examples of operational settings where RX is being adopted and robots deployed

Logistics: Trailer unloading

Example: Boston Dynamics' Stretch
Depalletizing variable cases with both speed and safety

Warehouse: High-mix picking

Example: Covariant's Goods-to-Person Picking
Pick-and-place for diverse objects

Hospital: In-hospital transport

Example: Diligent Robotics' Moxi
Safely navigates busy hospital environments and assists with transport of supplies

Facilities: Autonomous cleaning

Example: SoftBank Robotics' Whiz
Floor cleaning that can adapt to environmental changes

Figure 4: Examples of RX

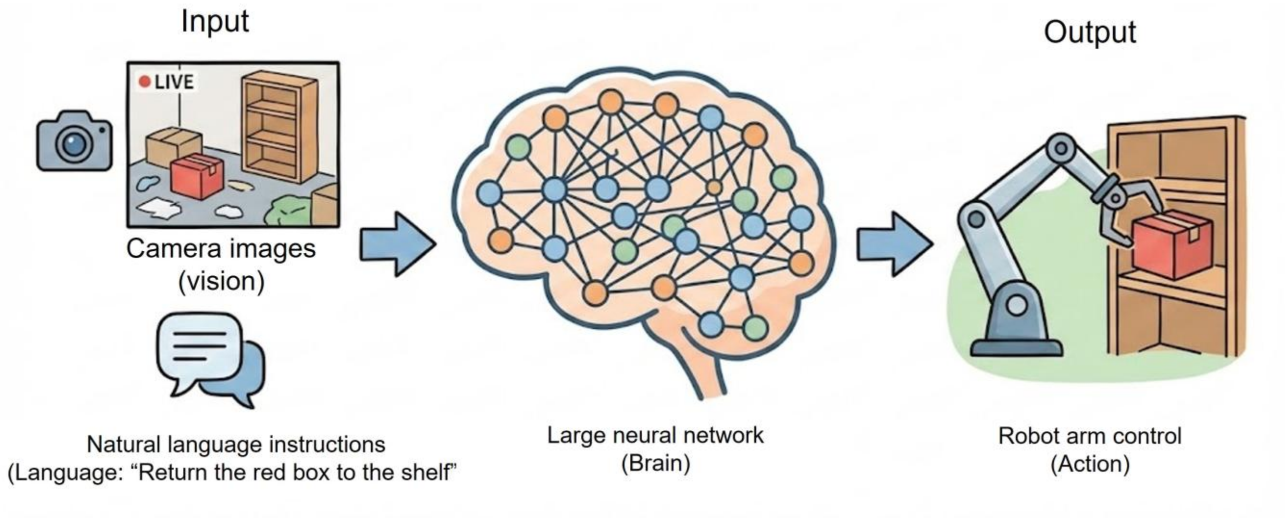
Industry	Developer	Robot	Robot Features (Including Deployment)
Logistics (3PL)	Boston Dynamics (US)	Stretch	Logistics robot for case handling. DHL Group is deploying an additional 1,000 units. MOU signed to accelerate automation across business units.
Warehouse (high-mix picking)	Covariant (US)	Goods-to-Person Picking (Covariant)	Targets high-mix SKU picking in warehouses. Aims to handle exceptions and achieve generalization using learning-based robotic intelligence such as Covariant Brain and RFM-1.
Hospital (in-hospital transport)	Diligent Robotics (US)	Moxi	In-hospital support robot that assists hospital staff with non-patient-facing tasks such as supplies and specimen transport, freeing time for nursing and clinical work.
Facilities (cleaning)	SoftBank Robotics (Japan)	Whiz	Commercial autonomous floor-cleaning robot. Supports standardization and labor savings in cleaning operations amid labor shortages.
Prepared food and food manufacturing	RT Corporation (Japan)	Foodly	Humanoid collaborative robot capable of working alongside humans on the same conveyor line. Described as the first deployment in the prepared food processing industry. Introduced at Ichibiki's Factory 2 (announced October 2021, operational March 2022). Pilot deployments were also conducted at Hirai and Fujimoto Foods under the framework of the Japan Ready-made Meal Association.
Food service and ready-to-eat (fried foods)	TechMagic (Japan)	F-Robo	Automates frying operations in store kitchens. Introduced at the first Real×Tech LAWSON store (announced July 2025), demonstrating automated preparation of Karaage-kun.
Retail (convenience store backroom)	Telexistence (Japan)	TX SCARA / Astra	AI robot that automates restocking of refrigerated beverages. FamilyMart officially announced plans to deploy at approximately 300 stores in conjunction with a task analysis system. Seven-Eleven also plans introduction.
E-commerce fulfillment	Amazon Robotics (US)	Sequoia/Sparrow/Proteus, etc.	Integrates inventory storage and retrieval, sorting, robotic grasping, and autonomous mobile robots (AMRs). Officially disclosed operation of more than 750,000 robots across its network.

Source: Compiled by MGSSI based on press releases from the companies

2-2. THE RISE OF ROBOT FOUNDATION MODELS (VLA/PHYSICAL AI)

Robot foundation models are referred to by various names, including Vision-Language-Action (VLA), Physical AI, and Embodied AI. They generalize the ability to see, think, and act by integrating and learning from visual information such as images and video, language instructions, and action data (Figure 5). By training on a combination of logs from teleoperation, on-site records of successes and failures, and synthetic data generated in simulators, these models expand the coverage of the learning space, making it easier to generalize to previously unseen failures and exceptions. Synthetic data also enables the intentional generation of low-frequency failures and edge cases, facilitating pre-training of fundamental recovery strategies such as re-grasping, posture adjustment, exploration, stopping, and handover to humans. Subsequently, a small amount of real-world data from actual machines is used to fine-tune for site-specific differences, including camera placement, jigs, and operational procedures, thereby adapting the model to individual environments. As a result, the focus of competition is shifting toward the ability to iterate the cycle of collecting, evaluating, and improving data. Over the next few years, robots may achieve a level of generality approaching that of a robotics version of ChatGPT. As foundation models mature, robots will evolve from static hardware products into systems that continuously learn from objects and environments, driving business models toward the ongoing capitalization of data from the physical world.

Figure 5: Robot foundation models integrating vision, language, and action (VLA/physical AI)



Source: Compiled by MGSSI

2-3. TECHNOLOGICAL HEGEMONY AND POLICY TAILWINDS IN THE US, CHINA, AND JAPAN

At present, the US is relatively ahead in models and software, while China leads in hardware mass-production capabilities and data collection. The axis of competition is shifting from standalone hardware to data and learning. Figure 6 shows examples of major robotics players in the US, China, and Japan.

Japan’s strengths lie not only in its on-site capabilities but also in its accumulated expertise in mechatronics component technologies such as reducers, actuators, motors, and sensors, as well as in quality and safety design. Japan can secure a competitive advantage in areas where robots are implemented in environments with stringent quality, hygiene, and safety requirements and where operational data can be accumulated and used to drive continuous improvement. In fact, domestic academic conferences such as RSJ and ROBOMECH place heavy emphasis on hardware related to mechatronics and sensor technologies, whereas international conferences such as CoRL, ICRA, and IROS place greater emphasis on software related to planning and perception technologies, particularly those based on machine learning. A classification of all papers presented at RSJ 2025 and CoRL 2025 shows that the former is roughly balanced between hardware and software, while the latter is dominated by software-focused research (Figure 7). Accordingly, Japan’s path to competitiveness lies in linking its strengths in mechatronics with operational data and securing an advantage in the implementation phase of foundation models.

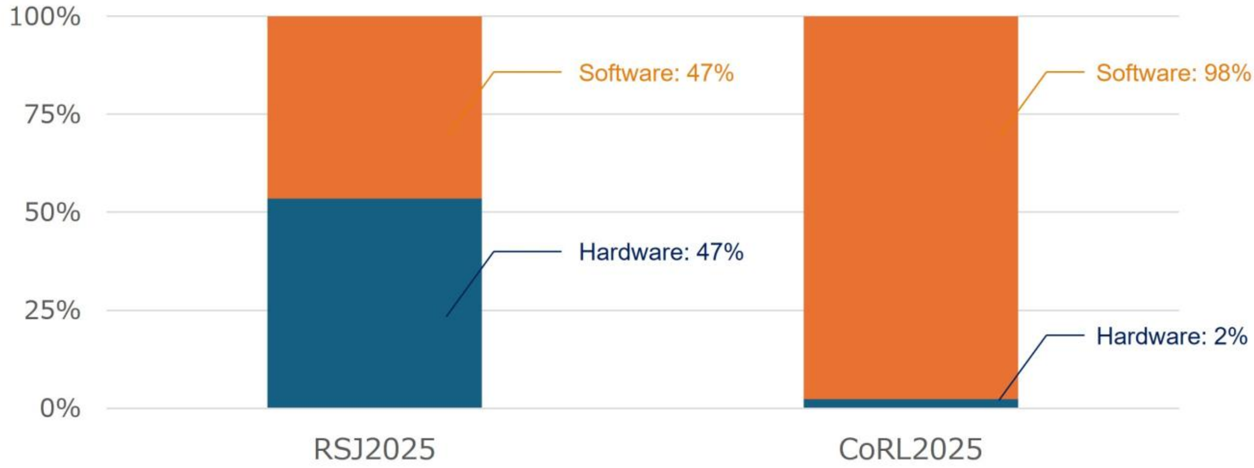
Furthermore, in Japan, the AI Strategic Headquarters set forth a policy in December 2025 to promote the integrated development of frameworks for development, utilization, and evaluation centered on trustworthy AI.

Figure 6: Examples of major robotics players

Domain	US	China	Japan
Brain (foundational models/software)	Physical Intelligence Skild AI Figure AI NVIDIA Google DeepMind Meta FieldAI etc.	AgiBot Tencent Robotics X Astribot X Square Robot etc.	Toyota Motor Corporation Telexistence etc.
Body (hardware/mass production)	Figure AI Boston Dynamics Apptronik Agility Robotics Tesla etc.	AgiBot UBTECH Robotics RealMan Robotics Galbot Astribot Unitree Robotics Pudu Robotics etc.	Yaskawa Electric Corporation Kawasaki Heavy Industries etc.

Source: Compiled by MGSSI

Figure 7: Comparison of technical domain classifications of papers at RSJ2025 and CoRL2025



Source: Compiled by MGSSI based on papers from RSJ2025 and CoRL2025

3. TOWARD THE ADOPTION OF RX: BUSINESS MODELS AND KEY ISSUES FOR JAPAN

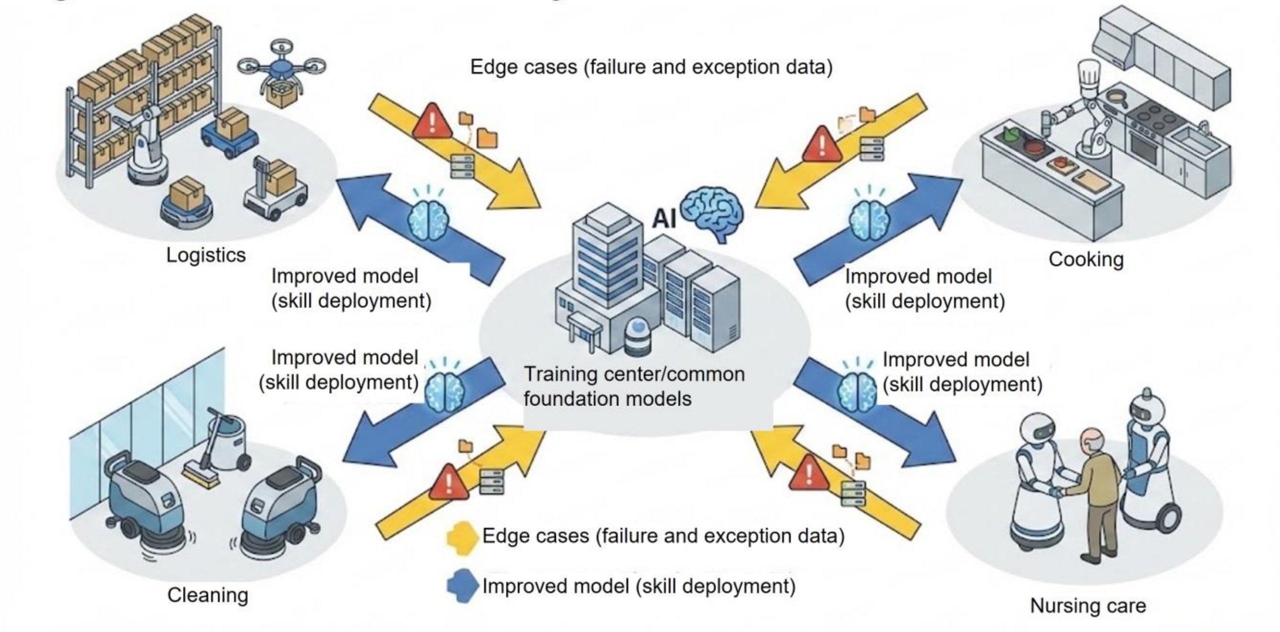
3-1. FROM HARDWARE SALES TO RAAS AND DATA BUSINESSES

When introducing robots, the initial investment forms a barrier to adoption at many operational sites. For example, robots for palletizing and depalletizing can cost JPY 10 million to JPY 80 million, depending on the configuration. Consequently, the focus is shifting from hardware sales to Robotics as a Service (RaaS), which also includes operation, maintenance, training, and improvements. From the user's perspective, RaaS enables verification of effectiveness while minimizing upfront investment. From the provider's perspective, it facilitates continuous improvement of models and processes based on operational data, making learning cycles more sustainable than one-time sales. During the initial deployment phase, operational design that incorporates human support to raise uptime is critical. The more robust the operational framework becomes, including remote monitoring, teleoperation, and continuous model updates, the faster deployment across sites can progress.

3-2. DATA AND TRAINING CENTERS: THE CORE OF COMPETITION

The performance of robot foundation models depends heavily on high-quality real-world data. As the volume of data increases, including failure cases and exception handling, robustness in operational settings improves. Training centers that accumulate data by combining physical machines, teleoperation, and simulators are therefore becoming a source of competitive advantage. Training centers must be designed not merely as sites for data collection but as operational platforms that incorporate evaluation procedures that meet safety requirements, root cause analysis in the event of accidents or near misses, and feedback to operational sites. Operational data may include highly sensitive information such as video, action logs, and process information, making practices that satisfy requirements for data rights management, confidentiality, security, and explainability essential. Specifically, it is necessary to establish clear rules for data handling, maintain records regarding which data were used for training and updates, and which models were deployed on-site, and make safety and performance verifiable through prescribed procedures. Going forward, the competitive focus will shift toward how data are collected, utilized, and evaluated (Figure 8).

Figure 8: RX value creation cycle based on real-world data



Source: Compiled by MGSSI

3-3. IMPLICATIONS FOR JAPAN: POTENTIAL AS AN IMPLEMENTATION HUB

Japan has many operational environments with stringent quality and safety requirements, making it well-suited for testing and improving AI robots. In sectors such as food, healthcare, and nursing care, requirements for hygiene, safety, and traceability are particularly rigorous, and systems proven in these settings can serve as strong benchmarks for deployment in other countries. If Japan establishes rules for the secure handling of data and frameworks for collaboration among companies, operational sites, and research institutions, and accumulates real-world data as assets, the practical adoption of foundation models will advance, supporting the spread of RX. Conversely, if operational data remains fragmented and learning and improvement cycles fail to function effectively, robot deployment will remain isolated, and RX will struggle to scale. The key issues ahead can be distilled into three points: (1) which operational settings can generate data, (2) where the economic viability threshold for RaaS lies, and (3) how to incorporate safety and reliability evaluations.

RX is not a competition to increase the number of robots but a race to accelerate learning cycles at operational sites. A practical implementation strategy in the era of foundation models is to begin with tasks where data can be readily collected, and benefits such as labor savings and quality improvements can be quantified, and then scale through standardization and horizontal deployment. Looking ahead, RX is expected to accelerate as data from the physical world is capitalized, simultaneously enhancing service quality and productivity.