Access Partnership

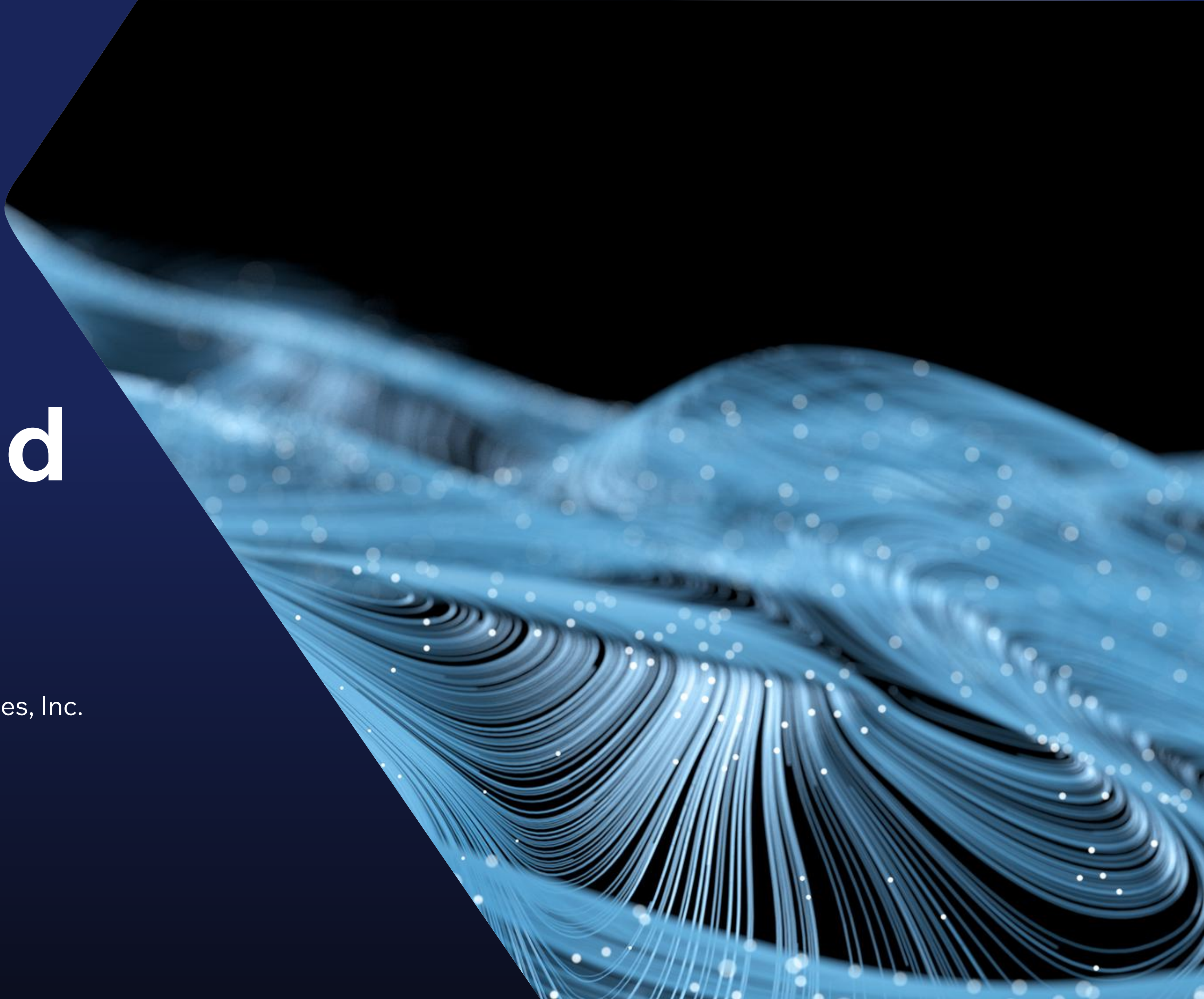
From Cloud to Edge

The value of distributed AI in Japan

Commissioned by Qualcomm Technologies, Inc.

September 2026

Prepared by Access Partnership



Disclaimer

All information in this report is derived or estimated by Access Partnership analysis using both non-proprietary and publicly available information. Qualcomm has not supplied any additional data, nor does it endorse any estimates made in the report. Where information has been obtained from third-party sources and proprietary research, this is clearly referenced in the endnotes.

With the exception of desktop researched claims which have been footnoted, all claims in the report have been derived based on Access Partnership modelling.

All calculations were done in USD and are based on the latest available data as of time of analysis in 2026.

Access Partnership (https://accesspartnership.com/), the frontier technology consultancy getting innovation into market faster, works with the public sector, corporations, NGOs, not- for- profits, and investors to identify opportunities and develop strategies to grow, improve service delivery, support economic prosperity, and generate tangible impact..

Access Partnership's Economics Strategy team is a leading economic and strategy consulting practice with deep experience across a number of topics in the digital economy and the future of skills, sustainability, and economic development.

Content

Executive summary and key highlights

Japan's opportunity in AI will be shaped by how effectively it deploys AI across its high-priority sectors. Distributed and physical AI can accelerate productivity, resilience, and innovation across manufacturing, mobility, and critical infrastructure, generating more than US$90B in annual economic value by 2035. Japan already has the foundation to claim this leadership through its world-class technological capabilities. Next is translating these strengths into widespread, real-world deployment through targeted policies that accelerate adoption and investment.

Japan can unlock opportunities with distributed AI…

Distributed AI can **boost automated manufacturing** with applications such as robotics and predictive maintenance

>$40B
of economic value in 2035 from **smart manufacturing and robotics**

Edge-enabled applications such as smart power girds and intelligent traffic systems can enable **smarter urban networks**

>$39B
of economic value in 2035 from **smart urban and utility network management**

Japan's transportation system can benefit from **smarter mobility systems** from AI-powered driver assistance systems and real-time detection

>8.8B
of economic value in 2035 from **intelligent, safe mobility**

…especially with the country's distinct advantages across the AI stack

Government-driven initiatives and strong **domestic hardware capabilities** create an opportunity to push for distributed AI

Highly advanced ecosystem of **software and hardware layers** of AI development infrastructure

Development of AI-native connectivity network is already underway

Government-led push for globally interoperable **standards**

Targeted policies will be key to unlock this

1. Diversify – Embed vendor diversity and healthy competition across the AI stack
2. Deploy – Make physical AI and robotics Japan's flagship national opportunity
3. Demonstrate – Establish a testbed framework for distributed AI
4. Lead – Convert standards and 6G leadership into global influence on the AI stack

CHAPTER 1

Japan's AI and technology landscape

An imperative for distributed AI

Japan is a global technological and industrial powerhouse whose next frontier is AI deployment

1.1 Ensuring Japan's economic leadership will require the country to lean toward AI – how the country will deploy it will be key

Japan is a top global economic and technology powerhouse

- Ranks **12th out of 139** economies in the Global Innovation Index (GII)
- Exports US$102 billion-worth of high-tech goods, **ranking top 11 globally** in volume
- Supplies **50% of global output** of robotics

However, socio-economic pressures are creating domestic challenges

- Firms are operationally constrained – **66% say labor shortfalls are impacting them**
- Japan's workforce is rapidly shrinking, projected to **decline by 31% by 2060**
- It's aging population further creates labor strain, with **1 elderly person for every 2 working-age adults** today (OECD average is 1 elderly per 3 working-age)

Accelerating AI adoption is will be key to securing Japan's future

- To ensure Japan's innovation leadership endures and to stem its labor shortage issues, new national strategies are mobilizing behind AI (**AI Promotion Act**, **AI Basic Plan**)
- AI uptake is already surging, with a **214% YoY growth in visits** to genAI tools in 2025

Japan is pursuing AI not from a position of experimentation alone but from economic necessity. **The question is no longer whether to deploy AI, but *how***

Japan is a global technological and industrial powerhouse whose next frontier is AI deployment

1.2 Distributed AI is best fit to deploy AI in the real-world environments that matter most to Japan

Key risks	Limitations of AI today	How distributed AI matters (and examples)
Energy demand by data centers	Data center energy use could **rise 80%** by 2030 and drive **60% of power demand growth** by 2034 – limiting larger scale deployment for cloud AI	Moves inference off data centers, easing power pressure. On-device AI can cut power use by **up to 90%** by removing network transmission.
Economy-wide labor shortage	Japan's **working-age population is projected to decline from 2027**, yet AI adoption lags on data-sensitivity concerns, limited in-house expertise, and low trust	Helps the existing workforce do more, with secure data and sector-specific tools. (E.g., Preferred Robotics' autonomous mobile robot (AMR) Kachaka works without connectivity, keeping data in-house while transporting items across factories, homes, or commercial spaces)
Demographic and healthcare risks	An ageing population **needs personalised, timely care** – limited by cloud AI's latency and data-transfer constraints	Enables personalised care with less exposure to data or network failures. (E.g. Cyberdyne and Tsukuba University's world's first wearable robotics suit Hybrid Assistive Limb (HAL))
Limited AI in manufacturing	Just **9% of Japan's manufacturing executives used AI** in 2022, as cloud AI can't serve time-sensitive, hazard-management uses	Enables worker safety and real-time line optimization with no network dependency. For example, on-device vision boosts quality checks and machine monitoring at ultra-low power without the need for cloud.
Ageing infrastructure	Much of Japan's infrastructure (**730,000 bridges, 11,000 tunnels and and an ageing power grid needing ¥6-7.9T in upgrades by 2050**) is near end-of-life – too vast for centralised defect management	Enables predictive maintenance and early detection. (E.g., Toda Construction's AI-powered drones reduced tunnel inspection time from 2 hours to 17 minutes and Aramco Digital is deploying edge AI over a private 5G network for asset monitoring across energy infrastructure)
Disaster resilience	Japan faces **one-tenth of the world's earthquakes**; disasters sever cellular and fibre networks – exactly when embedded AI must keep working	Keeps critical systems running without connectivity. (E.g., ACSL develops specialized drones that enable rapid disaster relief even when roads, bridges and connectivity are shut down)

Japan is a global technological and industrial powerhouse whose next frontier is AI deployment

1.3 Japan has the specific assets needed to lead in distributed and physical AI

Key capabilities today for distributed AI

A global leader in edge AI hardware – from materials to devices

Japan holds 53% of the global silicon wafer market; 88% of coater/developers; and 50% of photoresists – the inputs to every edge AI chip. Its OEMs also ship on leading edge AI platforms: Toyota's new RAV4 will deliver AI-driven, personalized in-vehicle experiences

An industrial base primed for physical AI at scale

Physical AI already runs in live, high-stakes settings including automotive production lines and logistics warehouses. FANUC, Mitsubishi Electric, and Kawasaki have already deployed physical AI in their machines for predictive maintenance, vision-guided picking, and automation.

A distributed AI network backbone already at national scale

NTT's IOWN delivers network-controlled AI inference fast enough for real-time, 6G-era robotics, while KDDI's "Power to Connect 2028" initiative extends low-latency edge infrastructure nationwide.

Government priorities and industry are aligned to push for distributed AI

The second AI Basic Plan (July 2026) elevates Vertical AI and Physical AI to Japan's primary strategic focus, with execution driven by the PM-chaired AI Strategy Headquarters and industry pushing in the same direction. Both priorities depend on AI running where field data is generated: at the edge.

The foundation is in place. The question is no longer whether Japan can lead in distributed AI, but where to deploy it first: the real-world settings where cloud-only AI falls short.

CHAPTER 2

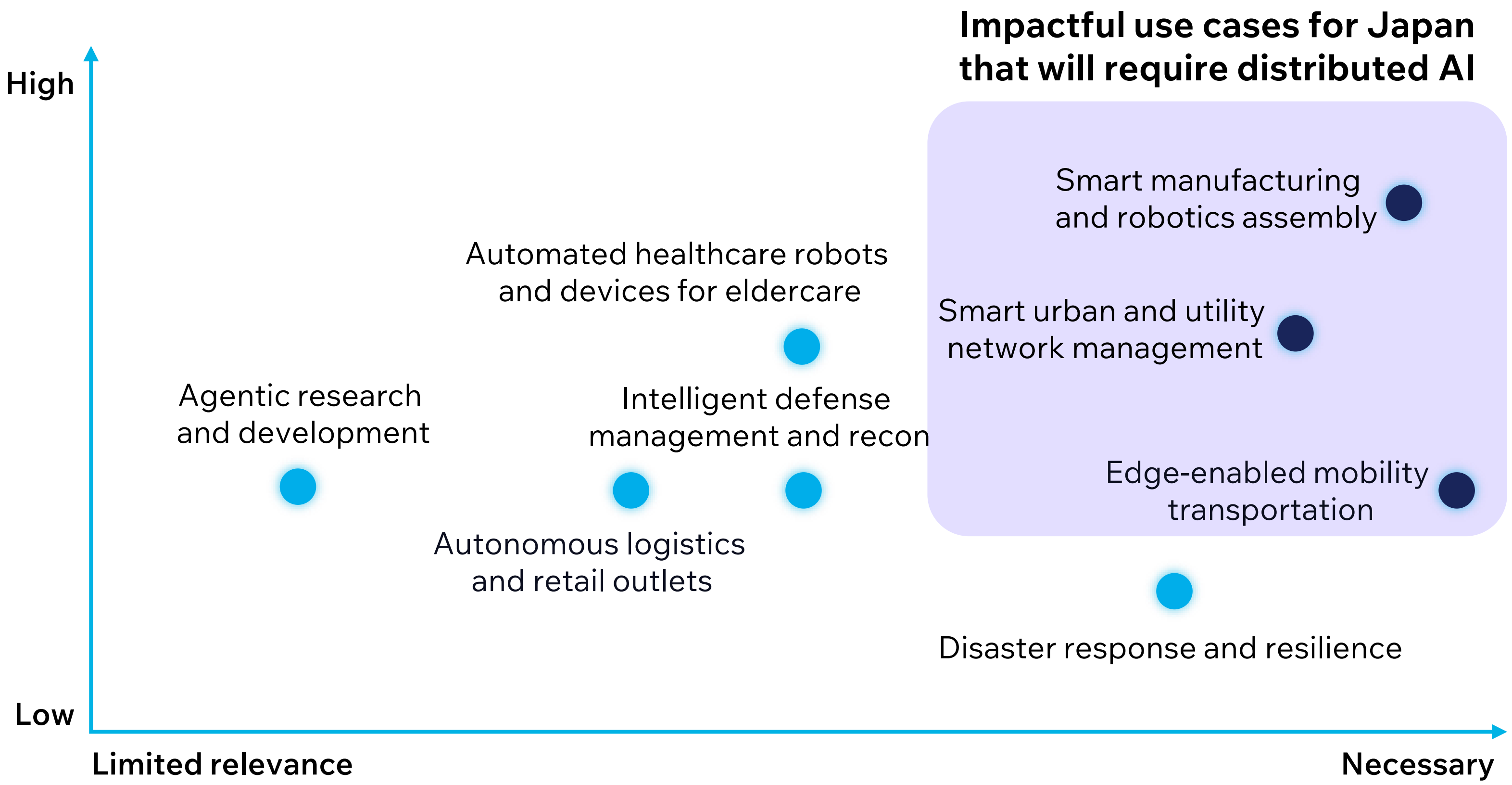
Impact of distributed AI on Japan’s economy and society

Distributed AI is expected to play a significant role in unlocking value through manufacturing, infrastructure, and automotive in Japan

Impact on economy

This dimension analyzes the scale of a use case's impact using public data, such as:

- Impact on the socio-economic challenge
- Market size
- Breadth of impact
- Quality of current infrastructure
- Governmental priority



Prevalence of distributed AI

This dimension relies on assessments of the scale of distributed AI deployment due to factors such as real-time or near-instant processing, reliance on sensitive data, level of personalization needed, or ability to leverage lightweight AI models

Japan's factories and cities are set to derive up to US$ 90B worth of economic value from distributed AI applications

Impact of **three use-cases** - driven by automating work to promote industrial productivity, more efficient infrastructure, and safer cities and factories

	>$40B **Smart manufacturing and robotics**	**>$39B** **Smart urban and utility network management**	**>$8.8B** **Intelligent, safe mobility**
Equivalent to:	**2.8%** of manufacturing contribution to GDP	**18.1%** of utility and infrastructure maintenance costs	**4.3%** of productivity value of time on the road
Examples:	Applications such as Industrial robotic arms, construction robots, humanoid robots for hazardous tasks, XR smart glasses for hands-free guidance, autonomous cleaning and inspection robots, and predictive maintenance	Applications such as intelligent traffic systems, sensor-based traffic light management and infrastructure failure detection, automated power distribution across the energy grid	Applications such as AI-enabled advanced driver assistance systems (ADAS) and autonomous emergency brakes (AEB), augmented through real-time detection and processing
Impact created by:	• Lower downtime • Higher labor productivity • Fewer safety incidents	• Energy efficiency • Reduction in traffic congestion • Remedying infrastructure failures preemptively	• Reduction in traffic accidents and loss of life • Reduction in insurance liabilities

Access Partnership

Note: Economic Value refers to the increased revenue or cost savings that businesses generate from increased market reach or increased productivity. It is not directly attributable to GDP.
Source: Access Partnership analysis

CHAPTER 3

Paving the way for distributed AI leadership with governmental support

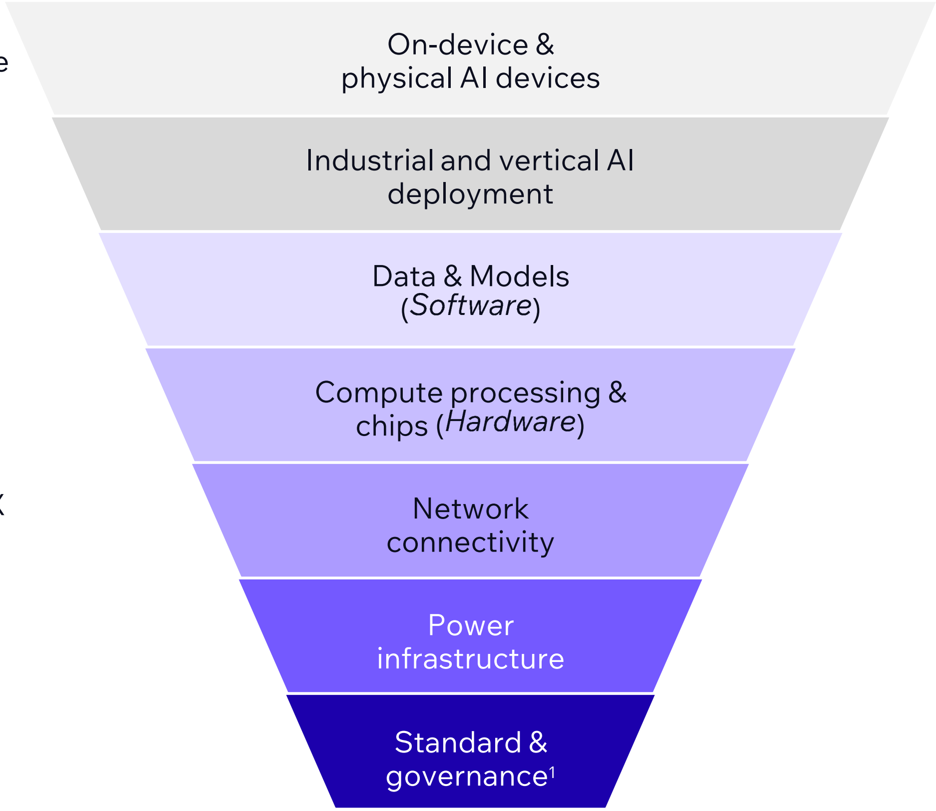
Japan has policy coverage across the full distributed AI stack – three principles can convert coordination into deployment at scale

In Japan, policy is increasingly system-level, linking AI, chips, networks, energy, and industrial deployment into one agenda, capped by the second AI Basic Plan, which makes vertical AI and physical AI Japan's primary strategic bets. Distributed AI is the architecture these bets require, as field data will be generated, and actions taken, at the edge.

Japan's policy foundations

	Japan's policy foundations
AI deployment	**Vertical AI and physical AI deployment now statutory strategy** under the 2nd AI Basic Plan; the LDP's AI White Paper 2.0 reinforces physical AI, secure compute, and trust in AI
Hardware & software	**~JP ¥4 trillion committed through 2030** – Government backing 30+ foundational model projects; Rapidus targets 2027 mass chip production with ¥250 billion state support under METI's System-to-Silicon strategy
Infra-structure & utilities	**6G and clean power on one roadmap** – the NICT-led 6G Promotion Project targets 2030 commercialization on an all-photonics network (NTT, KDDI, Fujitsu, NEC, Rakuten Mobile); the GX 2040 Vision prioritizes clean power co-location
Regulation	**Responsible agile governance** – Strengthened AISI, rules for AI agents in government, paired with continued Hiroshima AI Process leadership and plans to host an AI Summit

The distributed AI stack



Policy principles to advance distributed AI

Promoting multi-vendor diversity and portability – Keep AI models, data and applications portable across vendors and environments, operationalizing the Plan's open AI sovereignty and multi-modal approach

Establishing trusted and resilient systems – Embed security and transparency from the design stage, including for on-device and agentic AI, extending responsible agile governance and AISI evaluation work

Facilitating open and interoperable ecosystems – Extend the global nexus ambition into technical architecture with open interfaces from cloud to edge

1. Cuts across all layers

Source: Access Partnership analysis

Access Partnership

Targeted investment is needed to advance and solidify Japan’s position in the distributed AI ecosystem

We conservatively estimate that the incremental investment required is around $17.6B over 2026-2035

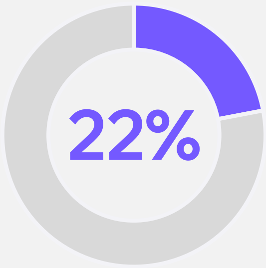
Estimated investment required for effective distributed AI deployment (2026-2035)

~$17.6B
(or ~1.7B per annum over a decade)

Includes funding for:
- Semiconductor & distributed AI hardware ecosystem
- Robotics & automation application development
- Industrial AI adoption incentives
- Distributed AI infrastructure & connectivity
- Workforce & skills transition

If the required investment were spread over a decade (~1.7B per annum), the annual investment…

Equates to



of Japan’s total AI budget announced for 2026

Generates

of total annual economic returns – based on the three major use cases (Page 8)

Japan can advance its distributed AI capabilities through targeted policy and investment

3.3 Four priorities – focusing on deployment and strategic positioning – will allow Japan to leverage its strengths to establish global leadership in distributed AI

1. Diversify – Embed vendor diversity and healthy competition across the AI stack	Japan should hard-wire multi-vendor requirements into public AI procurement and GENIAC-linked programs, so that models, data, and applications stay portable across vendors and environments – operationalizing the AI Basic Plan's open AI sovereignty and multi-model approach. Opening the ecosystem to multinational investment, with DFFT as the cross-border layer, keeps applications vendor-agnostic, cost-competitive, and resilient to lock-in effects.
2. Deploy – Make physical AI and robotics Japan's flagship national opportunity	Japan builds 38% of the world's industrial robots, and the AI Basic Plan now elevates Physical AI to primary strategic focus. Japan should convert this into deployment at scale: accelerate the AI Robotics Strategy, strengthen system integrators and component supply chains, and run targeted adoption programs that bring distributed AI to SMEs, manufacturers, and eldercare providers.
3. Demonstrate – Establish a testbed framework for distributed AI	Japan should extend the Plan's commitment to AI evaluation infrastructure and build a standing testbed network: designate real-world zones (e.g., a factory line, a care facility, a city district, a disaster response corridor) where physical, distributed, and agentic AI operate under regulated conditions; shared evaluation infrastructure with AISI to certify safety, security, and interoperability (ISO/IEC 42001 as baseline); and open publication of performance results so adopters and investors can act on evidence.
4. Lead – Convert standards and 6G leadership into global influence on the AI stack	Japan should use its convening role, already established through the Hiroshima AI Process, G7, and OECD, to anchor industry-led global standards for distributed, physical and agentic AI as well as 6G, and to extend the Japanese-influenced open interfaces by design into third markets through bilateral and multilateral partnerships – particularly to lead in promoting supplier diversity and interoperability.

Source: Access Partnership analysis

Appendix

Key definitions: the AI computing continuum

Cloud AI	AI training and inference executed in hyperscale data centers. Cloud AI provides the capacity needed for model training, global model hosting, orchestration, and the heaviest workloads, ensuring consistency across large user populations, but requires data to travel to and from centralized servers.
Edge AI	AI inference executed at or near the network edge (i.e., factory gateways, hospital edge servers, or micro-data centers) rather than in centralized cloud. Edge AI delivers real-time responsiveness, reduces backhaul costs, and keeps data within local or organizational boundaries.
On-device AI	A subset of edge AI, on-device AI operates inference directly on end-user hardware (smartphones, PCs, robots, industrial endpoints) without connectivity, which enhances privacy and offline resilience with the lowest latency.
Distributed AI	The hybrid computing continuum in which AI workloads are dynamically partitioned across on-device, edge, on-premises, and cloud tiers based on latency, privacy, bandwidth, energy, and cost requirements. No single tier can meet the demands of real-time, multimodal AI at scale; distributed AI treats the tiers as one tightly integrated system, with each handling the workloads it serves best.
Sovereign AI	A nation's ability to independently build, deploy, and govern AI using its own infrastructure, data, and talent that advances its laws, cultures, and values.
Physical AI	Unlike the compute tiers above, physical AI describes a class of applications: AI systems that perceive, reason, and act in the physical world in real time – robots, autonomous machines, drones, and connected vehicles. Physical AI must operate, learn, and scale reliably under real-world conditions, on-device and at the edge, coordinated with cloud-based learning across fleets.

Source: BCG (2026), 6G: The Network for the Future of AI and Immersive Connectivity, and Access Partnership analysis

Top-down framework for estimating AI compute inference

Overall description of approach taken

- To describe **total compute available**, the total compute capacity[1] available in data centers (public and on-premise) dedicated to AI up to 2030 was used as proxy (including compute used for training and inference) and modelled based on publicly available data.
- To measure the **potential on-device AI compute capacity, the overall AI compute capacity demanded** was also estimated – the "gap" between demand and current supply represents the total compute that will happen away from data centers.

Compute capacity (AI DC supply)

Key assumptions for projecting supply of data center (DC) compute capacity

- Projected compute capacity of DCs dedicated to AI workloads using: (a) total DC power capacity in gigawatts (GW) and; (b) total AI compute[2] to estimate a *compute-per-GW ratio* that changes over time
- **Assumption**: Of total cloud workloads. the share dedicated to AI would grow from 14% in 2025, estimated by Goldman Sachs, to 70% by 2030 based on McKinsey's projections. Additionally, around 40% of workloads are dedicated to training-related compute based on analysts from Centre for the Governance of AI.

Compute demanded (AI demand)

Key assumptions for projecting demand for AI compute

- Estimated total demand for AI by leveraging estimates by McKinsey that projects "annual FLOP demand" from use of **generative AI** to be 0.2 QFLOPs in 2024 – growing to be 25 QFLOPs in 2030
- It also estimates that by 2030, 40% of total compute demanded by businesses and consumers will be toward generative AI applications. Additionally, we assume that of total demand by businesses and consumers, demand for training applications will range between 10% and 25% of total compute.

1. Given limited research on TOPs (tera operations per second) used to describe neural processing units, compute is measured in terms of FLOPs (or floating point operations per second) as a proxy of potential capacity to complete calculation.; 2. Estimates on the total AI compute were estimated in April 2025 by the AI Futures Project, a team of AI developers and researchers. See AI Futures Project (2025), *AI 2027*. Available at: https://ai-2027.com/about

Source: BCG (2026), 6G: The Network for the Future of AI and Immersive Connectivity, and Access Partnership analysis

Top-down framework for country-specific AI compute projections

Overall description of approach taken

- Existing modelled figures on global compute capacity and demand were apportioned into country-specific figures. This illustrates the relative levels of installed DC capacity (AI DC supply) and AI demand for each country, with the assumption that DC capacity available within each country is dedicated to its domestic users.

Country-level compute capacity (AI DC supply)

Key assumptions for projecting country-level supply of data center (DC) compute capacity

- Existing modelled figures were apportioned into country-level figures based on a count of DCs (Data Center Map) and DC energy usage (IEA). These two factors served as proxies for supply of DC compute capacity.
- Growth in Japan's supply of DC compute capacity is based on growth of the domestic DC market (Arizton).

Country-level compute demanded (AI demand)

Key assumptions for projecting country-level demand for AI compute

- Existing modelled figures were apportioned into countries based on the relative usage of AI tools (aitools.xyz). This is measured by utilising web traffic analysis software to track the number of average visits[1] to a database of over 10,500 AI tools, applications, and platforms. This serves as a proxy for demand from both B2B and B2C users.
- Growth in Japan's demand for AI compute is based on global growth rates (McKinsey). Global growth rate was further adjusted based on the respective CAGR of AI spending as measured by market research firms, including IDC.

1. The number of average visits refers to the monthly average visits for each month between January and December 2025.
Source: BCG (2026), 6G: The Network for the Future of AI and Immersive Connectivity, and Access Partnership analysis

Impact of distributed AI use cases

Overall description of approach taken

- Three scenarios were developed to assess the potential economic value of edge AI adoption across various distributed AI use cases in 2035.
- The conservative scenario assumes that edge AI adoption in Japan follows the baseline AI adoption in Japan (i.e., SME AI adoption rates) (OECD). The base scenario assumes moderate edge AI adoption driven by adoption in industrial use cases (PROM WAD).

Smart manufacturing and robotics

Key assumptions for projecting the economic value of smart manufacturing and robotics

We assume that economic value is due to 3 value drivers:
- Productivity gains (based on productivity gains or increased output per worker)
- Downtime reduction (based on the reduction in downtime and the corresponding reduction in monetary losses)
- Safety improvements (based on reduction in safety incidents and the corresponding reduction in cost of workplace incidents)

Smart urban and utility network management

Key assumptions for projecting the economic value of smart urban and utility network management

We assume that economic value is due to 3 value drivers:
- Energy efficiency gains (based on improved distribution loss rate in energy grid through AI adoption)
- Congestion cost reduction (based on reduction in annual urban congestion economic cost from edge AI traffic management)
- Infrastructure maintenance savings (based on reduction in annual maintenance spend)

Edge-enabled mobility

Key assumptions for projecting the economic value of edge-enabled mobility

We assume that economic value is due to 2 value drivers:
- Road safety improvements (based on reduction in annual passenger-vehicle crashes from ADAS)
- Insurance and liability cost reduction (based on reduction in insurance claims from ADAS and AEB)

Source: BCG (2026), 6G: The Network for the Future of AI and Immersive Connectivity, and Access Partnership analysis

Access Partnership

accesspartnership.com

Need a problem solved?

Our dedicated experts, located around the world, are here to help.

Europe

London

Thomas House, 84 Eccleston Square
Office 5.06, 5th floor,
SW1V 1PX, London
United Kingdom

+44 20 3143 4900
london@accesspartnership.com

Brussels

8th Floor, Silversquare Europe
Square de Meeûs 35
B-1000 Brussels
Belgium

brussels@accesspartnership.com

North America

Washington DC

1300 Connecticut Avenue NW,
Suite 250
Washington, DC 20036
USA

+1 202 503 1570
washingtondc@accesspartnership.com

Asia

Bangkok

188 Spring Tower
11th Floor, Unit 106, Phayathai Road
Thung Phayathai, Ratchathewi, 10400
Thailand

+ 66 (2)-8216148

Hanoi

19th floor, Tower 1
Capital Place Building
No 29 Lieu Giai Street
Ngoc Khanh Ward, Ba Dinh District
Vietnam

Jakarta

Revenue Tower 21st Floor
Unit 104 SCBD Lot 13, Jl. Jend. Sudirman
Kav. 52-53
Provinsi DKI Jakarta, 12190
Indonesia

+62 21 5020 0949

Kuala Lumpur

Common Ground Q Sentral
Level 39, Unit 39-02 (East Wing), 2A,
Jalan Stesen Sentral 2, Kuala Lumpur
Sentral, 50470
Malaysia

Manila

CG8ROCKWELL Level 21, 8 Rockwell,
Hidalgo Dr., Rockwell Center, Bgy.
Poblacion, 1210
Makati City, Metro Manila
Philippines

Singapore

12 Marina View
#11-20, Asia Square Tower 2
Singapore 018961

+65 8323 7855
singapore@accesspartnership.com

Follow us