Policy Perspectives

PP-26-03

AI-as-a-Service: Securing the Generative AI Dividend for ASEAN+3

September 2026

[This page is intentionally left blank]

AI-as-a-Service: Securing the Generative AI Dividend for ASEAN+3

Prepared by QUACH Toan Long and Xianguo (Jerry) HUANG [1] [2]

Reviewed and approved by Abdurohman (Deputy Director for Functional Surveillance and Research).

September 2026

Abstract

Generative artificial intelligence (GenAI) is moving rapidly from experimentation to routine use across the ASEAN+3 region. Yet, for most member economies, access to frontier AI capability is obtained as a recurring service from a small number of foreign providers rather than through domestically controlled models and infrastructure. This paper maps the concentrated global AI value chain and ASEAN+3's uneven position within it and examines the macroeconomic implications of this access model. It shows that AI-as-a-service extends and intensifies the existing shift from ownership to revocable access in the technology sector: access can be repriced, modified, restricted, or withdrawn, while the associated payments are recurrent and often not separately identifiable in official statistics. Reliance on foreign-supplied AI could therefore affect the durability and distribution of productivity gains, labor-market adjustment, external accounts, and operational resilience, although the outcomes will depend on domestic absorptive capacity and each economy's position in the AI value chain. To guide deployment choices, the paper develops a capability–control–cost framework, based on which we propose that the policy objective should be neither technological self-sufficiency nor passive dependence, but to make necessary dependence productive, diversified, measurable, and reversible. Regional cooperation can reduce the costs of building absorptive capacity, maintaining fallback capability, developing common standards, improving measurement, and expanding access to compute.

JEL classification:	F02, F14, F15, L86, O33, O38
Keywords:	Generative artificial intelligence; AI-as-a-service; ASEAN+3; AI value chain; digitally delivered services; technological dependence; productivity; provider concentration; digital resilience; regional cooperation

---

[1] Authors' emails: Quach.ToanLong@amro-asia.org and Jerry.Huang@amro-asia.org.

[2] The authors would like to thank Yin Fai Ho, Wee Chian Koh (both AMRO), and Shu Yu (World Bank) for their comments and suggestions; Jade Vichyanond, Yuventus Effendi, Anthony Chia Kiat Tan, Guohua Huang, and Heung Chung (Andrew) Tsang for their valuable inputs and feedback. All remaining errors are the responsibility of the authors.

Abbreviations

ADB	Asian Development Bank
AI	Artificial intelligence
AIPI	AI Preparedness Index
AMRO	ASEAN+3 Macroeconomic Research Office
API	Application Programming Interface
ASEAN	Association of Southeast Asian Nations
ASEAN+3	ASEAN, China, Japan, and Korea
BIS	Bank for International Settlements
BPO	Business Process Outsourcing
CBDC	Central bank digital currency
DRAM	Dynamic random-access memory
ECI	Economic Complexity Index
EU	European Union
EUV	Extreme ultraviolet
FSB	Financial Stability Board
GenAI	Generative artificial intelligence
GPU	Graphics processing unit
HAI	Institute for Human-Centered Artificial Intelligence, Stanford University
ICT	Information and communication technology
ILO	International Labour Organization
IMF	International Monetary Fund
IoT	Internet of Things
ISEAS	Institute of Southeast Asian Studies
ISO	International Organization for Standardization
IT	Information technology
Lao PDR	Lao People's Democratic Republic
LLM	Large language model
METR	Model Evaluation and Threat Research
NACE	Statistical Classification of Economic Activities in the European Community
OECD	Organisation for Economic Co-operation and Development
R&D	Research and Development
TSMC	Taiwan Semiconductor Manufacturing Company
US	United States
USD	US dollar
VLSI	Very large-scale integration
WTO	World Trade Organization

Contents

Figures

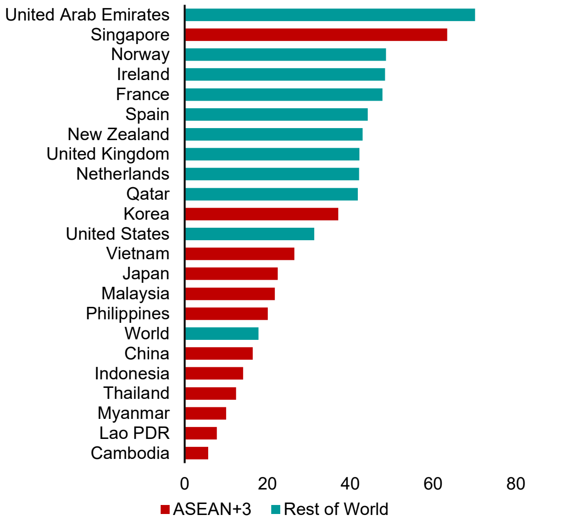
Tables

Appendix Figures and Tables

I. Introduction

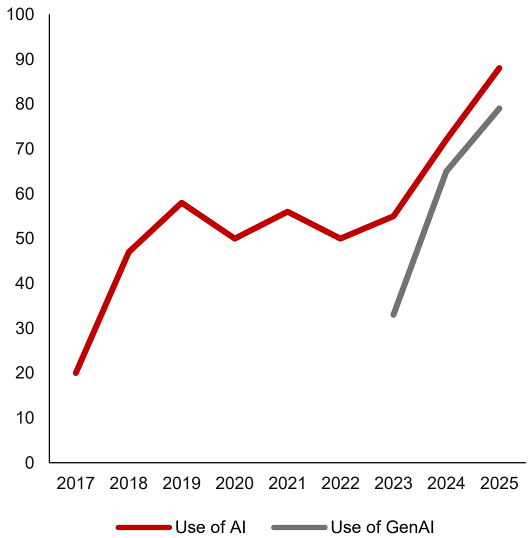
1. AI adoption has accelerated rapidly over the past few years, towards mainstreaming. The total number of ChatGPT users had reached 100 million users within two months of its launch in November 2022, making it one of the fastest-growing consumer applications at the time.[3] As of March 2026, around 18 percent of working-age adults worldwide were estimated to have used generative AI (GenAI) (Figure 1). Meanwhile, 88 percent of organizations reported using AI in at least one business function in 2025, up from 78 percent a year earlier (Figure 2). With such rapid adoption, AI tools have moved beyond experimentation to productive use in offices, factories, public services, and consumer applications.

Figure 1. Working-age Adults Using GenAI



Source: Our World in Data (2026)
Note: Estimated shares of working-age adults who have used GenAI tools are estimates derived from anonymized usage signals collected by Microsoft which does not distinguish between occasional and frequent users, as of March 2026.

Figure 2. Global: Use of AI by Organizations
(Percent of survey respondents)



Source: McKinsey Global Survey on the State of AI (2025)
Note: Share of respondents whose organizations use AI, or GenAI, in at least one business function, 2017–2025.

2. AI adoption could deliver large productivity gains through various channels. AI can automate routine components of service work, augment workers on more complex tasks, reduce the time and cost of information processing, drafting, coding, and analysis, improve output quality, and expand access to expertise where skilled labor is scarce. The World Development Report 2026 estimates that 16.2 percent of jobs in developing economies could receive a meaningful productivity boost from AI, close to 18.7 percent in high-income economies (World Bank 2026). Empowering less-experienced workers to perform more complex tasks with AI could narrow the productivity differences across workers. Beyond individual tasks, AI can accelerate the generation and testing of ideas, support scientific discovery and product innovation, and lower production and transaction costs. These gains can then propagate through the economy as lower costs and higher-

[3] Guardian (2023). https://www.theguardian.com/technology/2023/feb/02/chatgpt-100-million-users-open-ai-fastest-growing-app

quality output stimulate demand, while reduced communication and trade costs expand firms' access to markets and economic opportunities.

3. However, the gains from AI are likely to be uneven across the region. In this paper, the AI dividend refers to the durable economic gains that economies can realize from AI adoption, particularly through higher productivity and potential output. Almost 40 percent of global employment could be exposed to AI, with advanced economies more likely to see AI complement workers rather than replace them (Cazzaniga and others 2024). AI is projected to add 0.4 to 1.3 percentage points to annual labor productivity growth over the next decade in the most exposed advanced G7 economies, but gains can be up to 50 percent smaller where sectoral composition and adoption are less favorable, such as Japan (OECD 2025a). Exposure to AI in the rest of the region is more modest. The ILO finds that while 22.9 percent of ASEAN workers, or nearly 80 million people, are exposed to GenAI, most face only moderate or low exposure (ILO 2026). The same report shows that exposure tracks the share of clerical and office-based employment. Exposure was highest in service-oriented economies such as Singapore, at 42.2 percent, and the Philippines, at 28.1 percent. The AI dividend may therefore emerge earlier in economies with larger digitally enabled service sectors, while elsewhere, its realization will depend on whether AI capability can be adapted for use beyond a relatively narrow range of exposed occupations, especially through re-engineering processes and business-model innovation.

4. The ASEAN+3 economies have increasingly adopted national strategies, roadmaps, and governance frameworks to support AI development and adoption. Governments have adopted AI strategies and expanded support for computing infrastructure, research, skills, domestic models, and the adoption of AI across businesses and public services. By the end of 2025, 11 member economies had introduced a national AI strategy or equivalent plan, while one other incorporated AI into a broader long-term digital-economy strategy (Appendix Table I and OECD.AI Policy Navigator). These initiatives generally combine investment in skills, data, computing infrastructure and research, with measures to encourage sectoral adoption and promote responsible and trustworthy AI. At the ASEAN bloc level, regional frameworks for responsible deployment, including the ASEAN Guide on AI Governance and Ethics, its extension to GenAI, and the ASEAN Responsible AI Roadmap for 2025–2030, have complemented these national efforts. Private investment has expanded alongside these initiatives, particularly in data centers, cloud infrastructure, and semiconductors. Together, these efforts are widening access to AI and strengthening selected parts of the regional ecosystem, although their scale and depth remain uneven across economies.

5. However, a forward-looking strategic question is how ASEAN+3 economies can secure the AI dividend while building capability and managing potential risks arising from greater AI use in a sustainable manner. Although the region collectively possesses important capabilities in models, semiconductors, memory, equipment, and materials, most economies access leading AI systems through providers elsewhere. As of July 2026, popular frontier models were hosted and governed mainly by firms headquartered outside most member economies. From the perspective of an individual member economy, AI is therefore often imported as a service, or "AI-as-a-service". Throughout this note, "foreign" AI refers to services supplied by providers headquartered outside the purchasing economy and "AI-as-a-service" refers to access to frontier GenAI capability through hosted channels—consumer subscriptions, APIs, and cloud-bundled enterprise agreements—where the model is trained, hosted, updated, and governed by the provider.

6. This question is especially pressing as most economies would have to obtain productivity gains from AI without controlling either a frontier model or a domestically controlled AI stack. Estimates show that 62 percent of AI output is from the US, China, and Western Europe, and that imports could meet up to 78 percent of AI demand for net AI importer economies (Bekkers and others 2025). In this context, most members can therefore access frontier AI largely as final consumers and price takers, with limited influence over key aspects of the underlying services such as model standards, safeguards, and reliability. Gains from AI are conditional on access remaining affordable, reliable, and aligned with domestic needs. In a market as concentrated as the AI market, providers can reprice services, change access rules, retire a model, restrict features, or withdraw. These decisions may be commercially or geopolitically driven, but either way they can affect economies that depend on rented AI capacity. Developing and sustaining a comparable "sovereign AI"—a domestically developed and hosted AI model—may provide greater control but requires substantial and sustained resources and offers no assurance of frontier-level performance in a highly competitive market.

7. This note examines the implications of the current AI service access model seen in ASEAN+3 economies against this backdrop. It examines what the reliance on externally controlled frontier AI means for growth and economic resilience. It highlights gaps that might hinder ASEAN+3 economies from securing a long-term and durable dividend from AI development. The rest of the paper is structured as follows: Section II maps the global AI value chain, while Section III assesses the region's position within it. Section IV explains why purchasing access has become the default model and examines the terms on which that access is provided. Section V considers the implications for growth, external accounts, and labor markets, and Section VI discusses potential policy responses.

II. The Current State of the Global AI Value Chain

8. The AI value chain has several linked layers, encompassing specialized materials and equipment, computing infrastructure, training data, and hosting capabilities. In a simplified form, the chain begins with semiconductor design, equipment, and materials, followed by the fabrication of advanced logic and memory chips. Advanced packaging then integrates AI accelerators with high-bandwidth memory and other components. These chips and components are assembled into data-center systems that require networking, storage, power, and cooling. Cloud platforms, in turn, make this computing capacity available remotely to AI developers, which use it to train, deploy, and distribute foundation models. Firms and consumers then access these models through cloud services, application programming interfaces, or downstream applications.

9. A prominent feature of the AI value chain is its geographic dispersion alongside a high degree of concentration in critical layers. For example, as shown in Figure 3, at the semiconductor-equipment layer, Japanese firms are near-exclusive suppliers of the coater/developer systems used in EUV lithography and dominant suppliers of advanced photoresists.[4] The EUV lithography machines that project circuit patterns onto photoresist-coated wafers are, in turn, solely produced by ASML in the Netherlands.[5] At the manufacturing stage, TSMC holds about 73 percent of global foundry revenue and 80

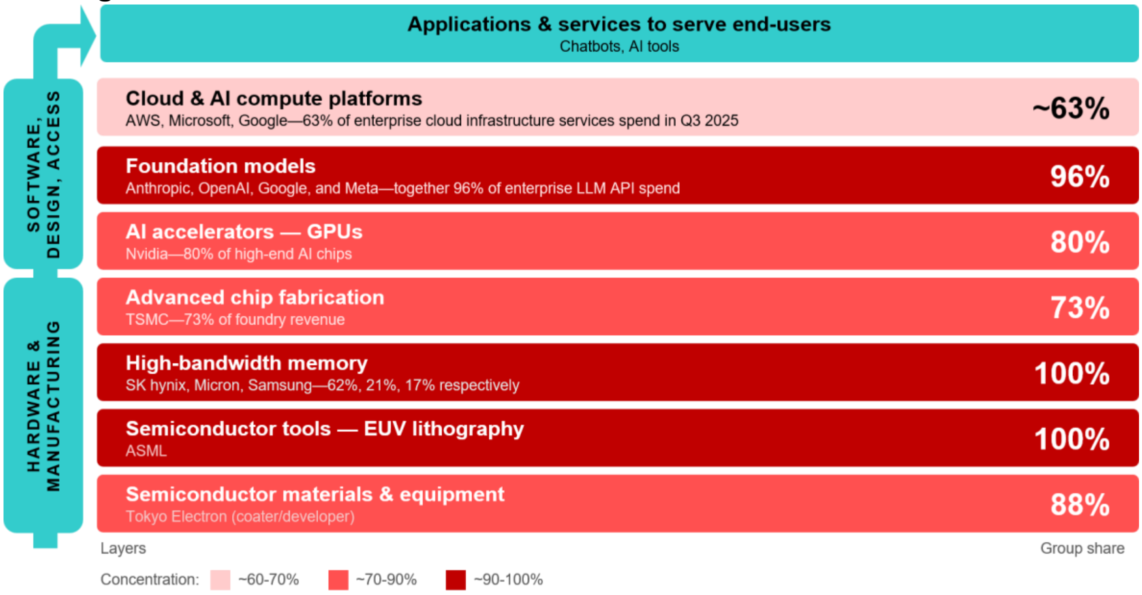
[4] For instance, according to a 2024 Brookings report, Japanese firms hold 88 percent of the global market share in coater/developers; 53 percent in silicon wafers, and 50 percent in photoresists. https://www.brookings.edu/articles/the-renaissance-of-the-japanese-semiconductor-industry/

[5] ASML 2022 Annual Report. https://www.asml.com/en/investors/annual-report/2022

percent or more of advanced semiconductor manufacturing.[6] High-bandwidth memory, another key component for developing AI infrastructure, is also concentrated among three suppliers: SK Hynix, Micron, and Samsung.[7] At the chip design layer, NVIDIA holds about 80 percent of the high-end chip market.[8] The concentration extends further downstream. A relatively small number of firms develop leading frontier foundation models, while cloud and computing capacity is concentrated among a few major infrastructure providers.

10. These concentrated positions are difficult to contest because the economics of the AI market work against new entrants. The AI market typically contains several features contributing to and reinforcing its high concentration (OECD 2025b). For instance, entry into several critical upstream and frontier layers of the AI value chain would require exceptionally high capital expenditure, large-scale research and development, long construction and production lead times, and access to proprietary technical ecosystems. These high fixed costs imply a large minimum efficient scale, making it difficult for new entrants to compete without securing substantial market share. For incumbent firms, the established scale advantages then cumulate across layers: firms operating at the largest scale obtain inputs earlier and on better terms, set the technical standards their downstream users build against, and reinvest the resulting margins in the next round (Korinek and Vipra 2025; CMA 2024). Limited interoperability, ecosystem lock-in, and long-term commercial commitments further raise switching costs for downstream users, reinforcing the market position of established providers.

Figure 3. Selected Concentration Across the Global AI Production Stack



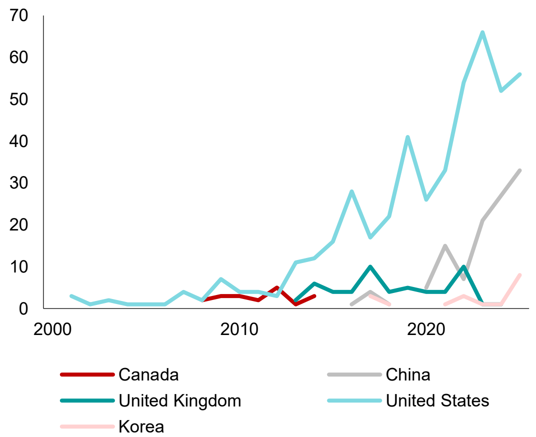
Source: AMRO staff illustration
Note: The figure highlights selected examples of market concentration across the AI stack, from semiconductor materials and manufacturing equipment to compute platforms and foundation models. It is not intended to provide a comprehensive mapping of all concentrated segments. The percentages indicate the market share of the leading firm or group of firms identified in each segment. The estimates are intended to highlight the key concentration at each layer group and are not directly comparable across layers due to differences in markets, metrics, and reference periods. Coverage of China-based providers is uneven across layers. For instance, the enterprise cloud and LLM API spend figures are based on markets outside China, where domestic providers hold substantial shares; the semiconductor and hardware figures are global.

[6] Counterpoint 2026 and CSIS 2024.
https://counterpointresearch.com/en/insights/global-semiconductor-foundry-market-share; and
https://www.csis.org/analysis/strategy-united-states-regain-its-position-semiconductor-manufacturing
[7] Counterpoint Research 2026.
https://counterpointresearch.com/en/insights/global-dram-and-hbm-market-share
[8] World Bank 2025.
https://openknowledge.worldbank.org/bitstreams/d52c2c9e-0bfe-4574-83d1-f6f6e78140cf/download

11. This concentration creates a distinction between participation in the AI value chain and the degree of depth economies have built within it. Participation includes adopting AI tools within the value chain. On the other hand, economies can establish varying degrees of depth in particular layers by having the broader domestic capacity to produce, adapt, govern, and export AI-related goods and services through local capabilities in data, compute, models, software, semiconductors, standards, and skilled labor.[9] Participation captures the value by raising efficiency, but digital depth determines how much more value an economy adds and how much control it has over pricing, availability, standards, and terms of use.

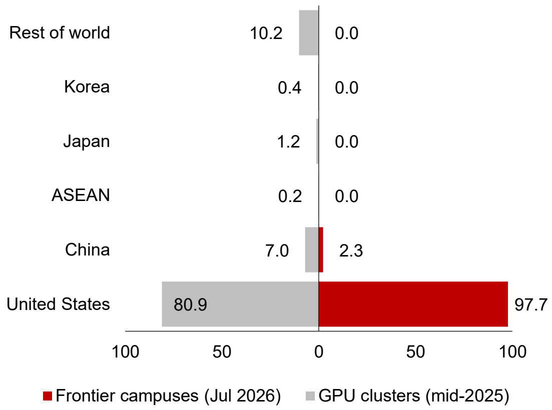
12. US firms have substantial depth across both the AI production stack and the platform development layers. In 2025, the US industry led the world in producing 56 notable AI models, followed by China with 33, and Korea with eight, while other economies were far behind (Figure 4). The US also hosts more than three quarters of GPU clusters and 98 percent of the global frontier computing power (Figure 5), as well as almost half of the world's data centers (Figure 6). This concentration expands the scale of training runs that US-based developers can undertake, strengthening their capacity to develop more capable frontier models. In the access layer, four US firms—Anthropic, OpenAI, Google, and Meta—accounted for 96 percent of enterprise LLM API spend (Menlo Ventures 2025).[10] The US also has a strong presence in other parts of the stack: US Big Tech such as Amazon Web Services, Microsoft, and Google account for 63 percent of enterprise cloud infrastructure services spend in Q3 2025,[11] while US firms such as NVIDIA also dominate chip design.

Figure 4. Selected Economies: Number of Notable AI Models Released



Source: Epoch AI, Data on AI Models dataset, accessed on July 30, 2026; AMRO staff calculations.
Note: Notable models meet at least one of the following criteria: more than 5,000 citations; an estimated training-compute cost exceeding USD1 million; or more than 1 million monthly active users. Refer to Epoch.ai for a detailed description.

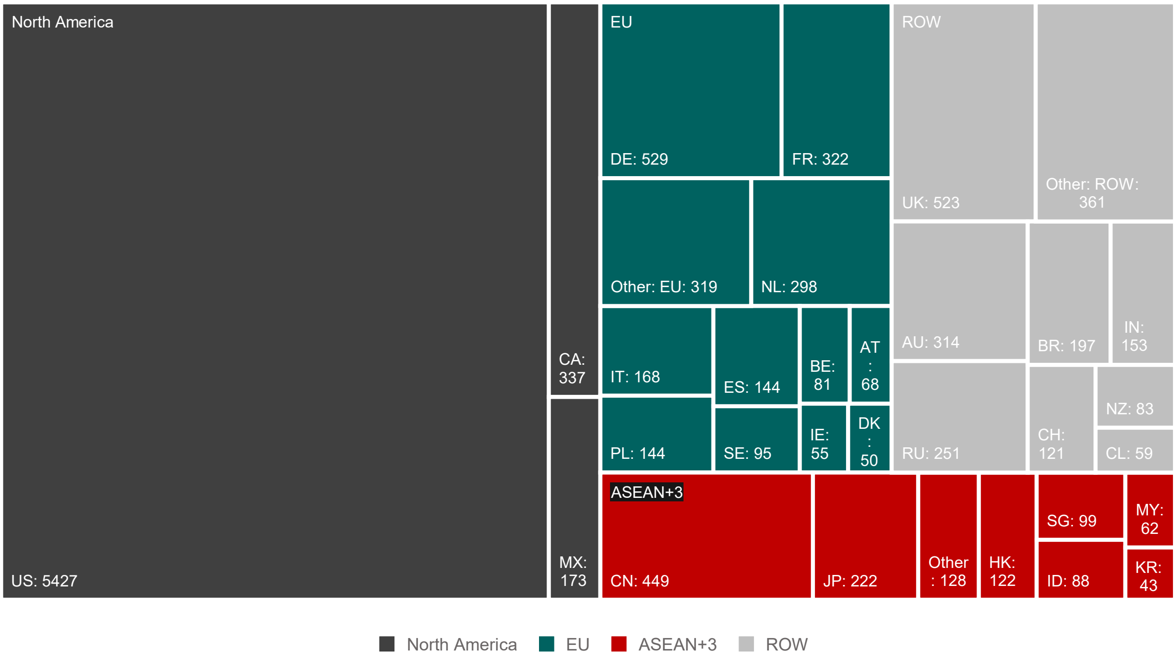
Figure 5. Selected Economies: AI Data Center Total Computing Power
(Share of global GPU capacity)



Source: Epoch AI, AI Data Centers dataset and Epoch AI, GPU Clusters dataset, accessed on July 30, 2026; AMRO staff calculations.
Note: Computing capacity is measured in H100-equivalent GPUs.

[9] In comparison, Udomsaph (2026) uses the concept of "digital depth" in the context of firm-level AI adoption in developing economies. https://thedocs.worldbank.org/en/doc/c749b541398f6b28ab8a64250b52e5fa-0280032026/original/udomsaph-AIAdoption-20260120.pdf
[10] Menlo 2025. https://menlovc.com/perspective/2025-the-state-of-generative-ai-in-the-enterprise/
[11] Synergy Research 2025 Estimates. https://www.srgresearch.com/articles/cloud-market-share-trends-big-three-together-hold-63-while-oracle-and-the-neoclouds-inch-higher

Figure 6. Geographic Locations of Data Centers



Source: Cloudscene, as of August 30, 2026; AMRO staff calculations.
Note: CN = China; HK = Hong Kong, China; ID = Indonesia; JP = Japan; KH = Cambodia; KR = Korea; LA = Lao PDR; MM = Myanmar; MY = Malaysia; PH = Philippines; SG = Singapore; TH = Thailand; VN = Vietnam; US = United States; CA = Canada; MX = Mexico; ROW = Rest of the World. All other economies follow the ISO 3166-1 alpha-2 coding standard, except the United Kingdom, which is shown as UK. Other ASEAN+3 comprises PH, VN, KH, MM and LA.

III. ASEAN+3 Positions in the AI Value Chain

13. **ASEAN+3 is key to the global AI value chain, but capabilities are uneven across members and layers.** Collectively, the region hosts significant suppliers in several stages of the AI stack, such as open-weight AI models, memory, fabrication, equipment, and materials. These regional strengths, however, do not eliminate the dependence of individual economies on foreign AI capability, especially those outside the region. AMRO staff's preliminary compilation suggests that some members hold strong upstream positions, while others participate mainly through hosting, adaptation, and adoption.

- China is strengthening capabilities across the full AI stack, from chips to models. It accounted for 74.2 percent of AI patents granted globally in 2024 and released the second-highest number of notable AI models in 2025, the highest in ASEAN+3 and second globally only to the United States (Stanford HAI 2026 and Epoch AI Notable Model dataset). China is also expanding its domestic hardware industry. For instance, Huawei reportedly plans to roughly double its 2025 die output in 2026 to up to 1.6 million dies.[12] At a higher stage of the semiconductor value chain, China is also building capacity to produce EUV lithography tools domestically to reduce reliance on imported EUV lithography systems, according to Reuters.[13]

[12] Bloomberg 2025. https://www.bloomberg.com/news/articles/2025-09-29/huawei-to-double-output-of-top-ai-chip-as-nvidia-wavers-in-china
[13] Reuters 2025. https://www.reuters.com/world/china/how-china-built-its-manhattan-project-rival-west-ai-chips-2025-12-17/

- Korea retains a strong position in memory, with domestic firms accounting for nearly 67 percent of the global DRAM market.[14] SK Hynix alone supplied 58 percent of high-bandwidth memory, and Samsung another 21 percent in Q1 2026.[15] This strength is further reflected by the talent concentration—around 20 percent of Korea's AI talent pool specializes in hardware, VLSI, and IoT (Stanford HAI 2026). Korea has also developed domestic capabilities in advanced fabrication, domestic cloud platforms, and Korean-language foundation models.[16]

- Japan retains strategic influence through AI compute infrastructure and its renewed efforts in leading-edge fabrication. Domestic capabilities include NTT's tsuzumi 2,[17] government-supported GPU cloud projects, and public AI infrastructure such as ABCI 3.0. Rapidus is also targeting 2-nanometer mass production in 2027.[18] Japan's clearest strength, however, lies in semiconductor equipment and materials (Figure 3).

- ASEAN participation extends beyond downstream adoption into selected electronics and computing segments of the AI value chain. Malaysia, Vietnam, and Thailand were among the largest exporters of AI-enabling products among the developing economies in 2025 (World Bank 2026). These exports include semiconductors and integrated circuits, computer and server components, storage devices, networking equipment, and other inputs used in data-center infrastructure. Malaysia, in particular, has an established semiconductor manufacturing base and is seeking to move further upstream into chip design, illustrating how participation in existing segments can support deeper domestic capability over time.

- Singapore has developed depth in AI talent, adoption, and digital infrastructure, rather than in upstream semiconductor production. It is among the world's leading economies in AI authors and inventors per capita, second only to Switzerland, according to the Stanford AI Index (Stanford HAI 2026). Singapore has also established itself as a regional hub for data centers, cloud services, and AI deployment.

14. Foreign AI-related investment into ASEAN+3 has been overwhelmingly concentrated in physical infrastructure and hosting rather than transferring AI capability. Announced cross-border investment into ASEAN+3 in selected AI-related industries totaled about USD286 billion over 2022–25. Manufacturing accounted for approximately USD150 billion of this sum, and data centers for USD123 billion, together representing about 95 percent of the total (Figure 7). Projects classified as research and development accounted for only USD1.3 billion in inward investment, or less than 1 percent (Figure 8). The investment commitment in manufacturing and data centers can bring construction activity, land and power revenue, and some skilled jobs; however, the higher-margin parts of the stack such as chips, models, cloud platforms, and application

---

[14] Counterpoint 2026
https://counterpointresearch.com/en/insights/global-dram-and-hbm-market-share
[15] Counterpoint 2026.
https://counterpointresearch.com/en/insights/global-dram-and-hbm-market-share
[16] For instance, HyperCLOVA X, a Korean LLM model, was released in 2023 with strong performance in several Korean language benchmarks.
https://clova.ai/en/tech-blog/introducing-hyperclova-x-our-state-of-the-art-ai-models-optimized-for-the-korean-language
[17] Refers to NTT tsuzumi 2 for more detailed information.
https://group.ntt/en/newsrelease/2025/10/20/251020a.html
[18] Refers to Rapidus for details.
https://www.rapidus.inc/en/

ecosystems usually remain with the foreign owners. Although data centers nevertheless represent a form of participation in the AI value chain by enabling digital services to be supplied to domestic and potentially regional markets, the extent of capability transfer depends on whether projects can establish local engineering, research, and model-development functions (Huang and Ho, forthcoming).

Figure 7. ASEAN+3: Announced Cross-border Investment in Selected AI-related Industries, by Selected Industry Category
(Billions of US dollars)

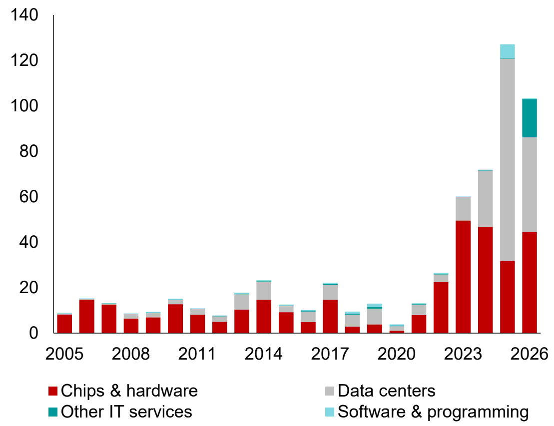
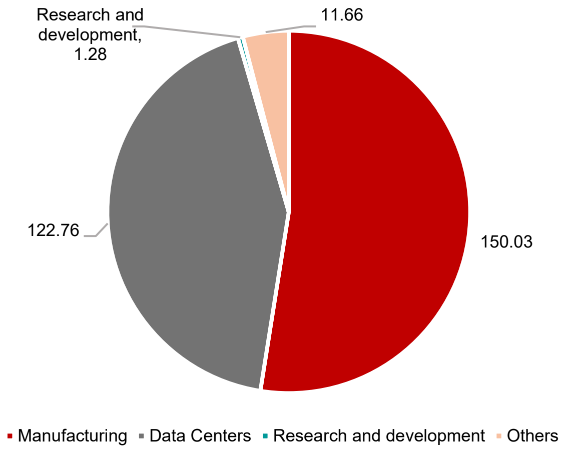


Figure 8. ASEAN+3: Announced Cross-border Investment in Selected AI-related Industries, by Business Functions, 2022-2025
(Billions of US dollars)

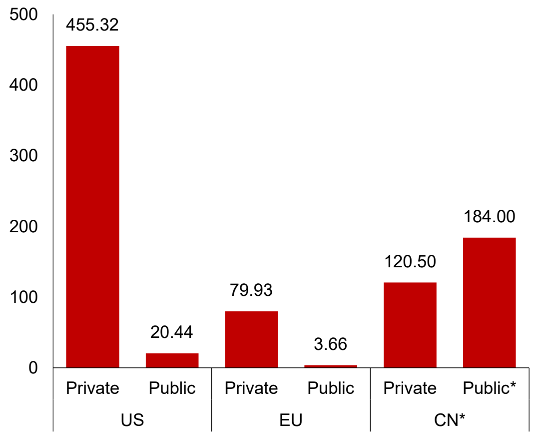


Source: Moody's/ Bureau van Dijk Orbis Crossborder Investment; AMRO staff calculations
Note: The dataset includes announced and completed cross-border projects by destination market, extracted on 28 July 2026. AI-related investment is proxied by projects whose primary destination NACE code is 6311 (data processing and hosting), 6201 (computer programming), 2611 (electronic components) or 2620 (computers and peripherals). Data centers are the subset of NACE 6311 projects explicitly identified as such in the project headline or business function, so the total is a lower bound. Values for multi-destination projects are split evenly across destinations. Projects are dated by announcement date. The industry-based proxy may include projects within the selected NACE industries that are not directly related to AI and should therefore be interpreted as an approximate upper bound on AI-related investment. Business Functions are based on Orbis dataset's categories.

15. At the same time, domestic AI investment in ASEAN+3 remains small compared with the US. Between 2013 and 2024, private AI investment in the US alone reached USD455 billion, while available estimates put combined public and private AI investment in China and Europe at roughly USD388 billion over the same period (Figure 9). The gap in private investment widened in 2025 as the US increased by another USD240 billion, compared with USD10 billion and USD17 billion for China and the EU respectively. The institutions driving AI investment also differ markedly. In the US, the public sector contributed only 4.3 percent of investment in AI, whereas in China, the state has long played a more active role through grants, budget allocations, and state investment funds. Elsewhere in ASEAN+3, governments have also begun committing public resources to AI through national strategies, research, computing infrastructure, and fiscal support, although the scale and coverage vary considerably across economies (Appendix Table II). China nevertheless remains by far the region's largest source of private AI capital—its cumulative private AI investment of about USD131 billion over 2013–2025 is more than triple that of the rest of the region combined (Figure 10).

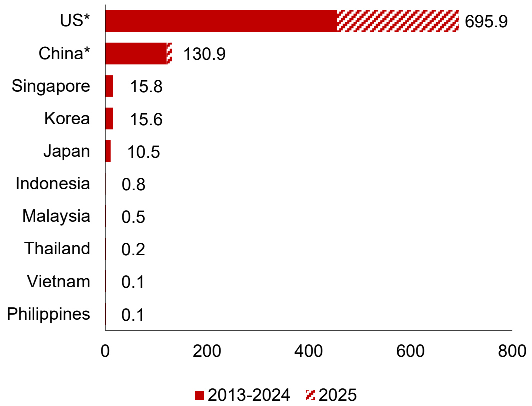
16. Funding gaps among member economies, and between much of the region and the US, suggest a widening gap in AI investment, which could translate into persistent gaps in future AI development capability. Frontier AI development requires not only long-term funding but also a sustained pipeline of inputs, including large amounts of curated training data, a deep pool of specialized technical talent, and computing infrastructure at a scale that few markets can amortize. The financial burden is compounded by the rapid technological obsolescence of AI computing assets, which would require recurrent investments to maintain competitive capability as new generations become available. As frontier AI products improve, the minimum investment needed to catch up rises—making external AI services the more economical choice and limiting the short-term value of building domestic capability. Investment funding must be sustained across successive model generations. The financing challenge is particularly acute for public-sector-led efforts. While private firms can accept such costs in pursuit of future revenue and market share, public efforts to catch up face fiscal constraints and may incur substantial costs without any assurance of achieving commercially competitive capability.

Figure 9. US, EU, and China: Available AI Investment, 2013-2024
(Billions of US dollars)



Source: HAI AI 2026 index; Haver Analytics
Note:
*China's public investment estimate is between 2000 and 2023 only, based on China's government guidance funds into AI firms.

Figure 10. US and Selected ASEAN+3: Total Private AI Investment, from 2013-2024/2025*
(Billions of US dollars)

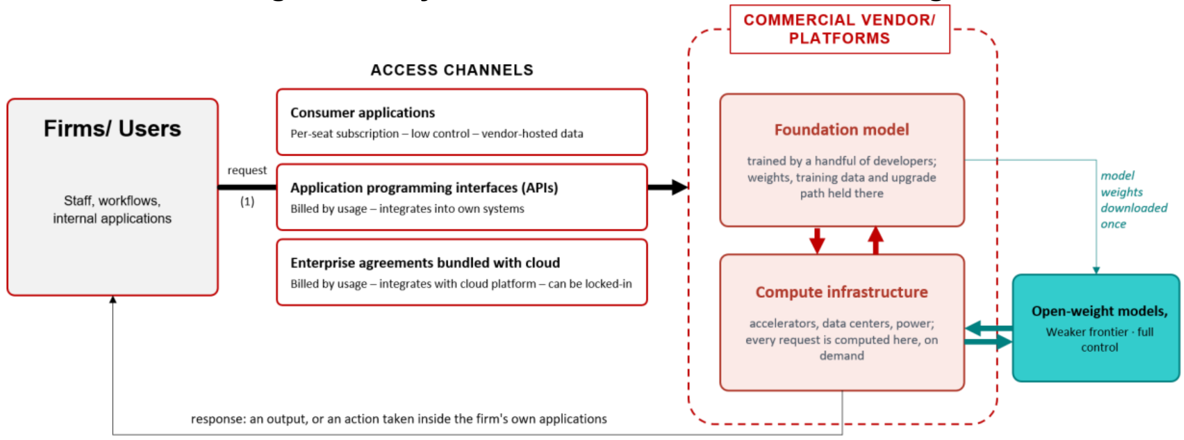


Source: Haver Analytics; AMRO staff calculations
Note: The dataset only covers private-market investment such as venture capital. It excludes non-equity financing, such as debt and grants, and publicly traded companies, including major Big Tech firms. Significant investments from public companies, corporate R&D, government funding, and broader infrastructure costs are not captured.
* Due to data availability, data for 2025 are available only for the US and China.

IV. AI as an Imported Service

17. Firms need access to a foundation model to obtain potential productivity gains. As simplified in Figure 11, an AI foundation model is a compute-intensive application that must be hosted on specialized computing infrastructure to process user requests. A user sends a request to an AI and receives either digital output or the AI acting directly on their applications. Access is available through several channels: consumer applications on per-seat subscriptions, application programming interfaces (APIs) typically billed by usage, enterprise agreements bundled with cloud contracts, or open-weight models downloaded and run on cloud or local hardware. The channels differ in cost, control, and data exposure, but in every case the capability comes from a base model that was trained in advance. For most ASEAN+3 firms, the immediate decision is not whether to build a frontier model, but which model and access channel best balance functionality, cost, compliance, and regulatory requirements.

Figure 11. Stylized Flow of AI Access and Training



Source: Authors' illustration

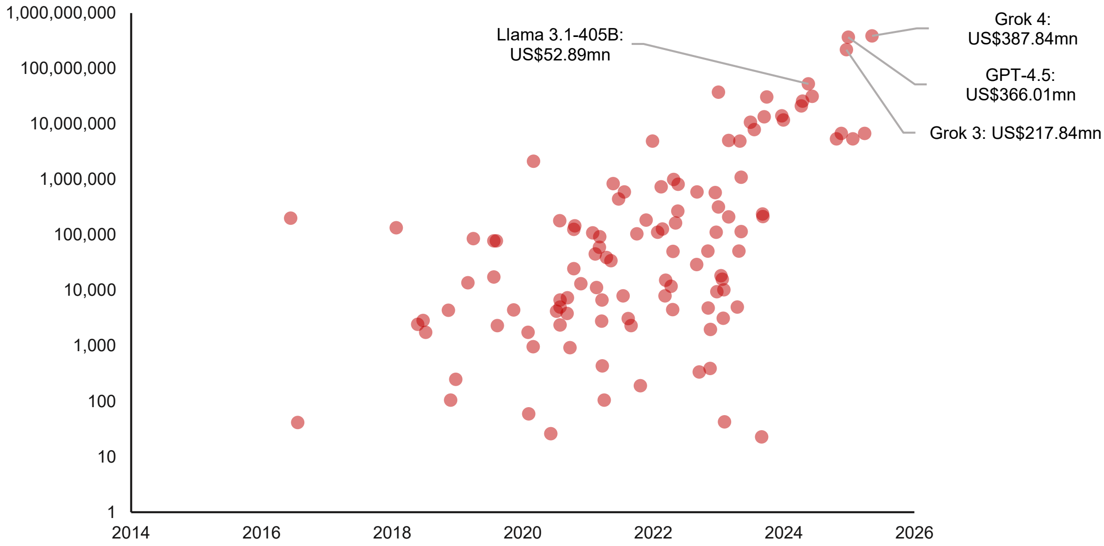
18. Yet, building a frontier model requires a comprehensive system of funding, training data, talent, and economies of scale. A frontier model requires trillions of tokens of carefully curated training input data, such as text, code, image, audio, and video; a system of advanced accelerators linked to high-bandwidth networks; and a large team spanning machine learning, data engineering, distributed systems, evaluation, and safety. A pre-training run alone could require a large computing cluster for months. For instance, Meta's Llama 3.1 405B, released in July 2024, was trained on about 15 trillion tokens using more than 16,000 H100 GPUs.[19] This means the compute cost of the final training run alone would have been on the order of USD50-90 million, excluding the costs of failed experiments, data preparation, post-training, personnel, and infrastructure. Researchers at Epoch AI suggest that the cost of training frontier AI models has been increasing at a rate of 2 to 3.1 times each year since 2016, and model training runs could cost more than USD1 billion by 2027 (Cottier and others 2024). Indeed, a few months after Llama 3.1 was rolled out, subsequent models such as GPT-4.5 and Grok 4 were estimated to have training-compute costs of more than USD350 million, more than six times the estimate for Llama 3.1 (per Figure 12).

19. Open-weight GenAI models offer an alternative route to advanced capability, providing greater control and more durable access without requiring users to bear the full cost of model development. Open-weight models are those that allow users to download and retain the model weights and run the models on locally deployed infrastructure. Because users retain the model parameters (hereafter model weights), access to those model weights and their capabilities cannot be revoked by the producers. In addition, users may also fine-tune those parameters, adapting the model to proprietary data and specialized domains through low-rank adaptation, continued training, distillation, or model merging, at a fraction of the cost incurred in training from scratch. Although fine-tuning improves a model's performance on a defined task, such model adaptation does not confer new general reasoning capacity and cannot raise the capability ceiling of the released checkpoints. Moving beyond that ceiling requires either updates from the developer or training by oneself, which requires own capital and expertise.

---

[19] Meta 2024 Introduction blog post, accessed on July 15, 2026. https://ai.meta.com/blog/meta-llama-3-1/

Figure 12. Selected GenAI Models: Training Compute Costs
(US dollars, logarithmic scale)



Source: Epoch AI; AMRO staff calculations
Note: Estimates are expressed in constant 2023 US dollars. The vertical axis uses a logarithmic scale.

20. Despite their accessibility, there are hurdles facing actual adoption of open-weight models. Even without the expense of a training run, users must secure sufficient infrastructure to store the model weights and serve concurrent requests. The binding constraint is often GPU memory, and the requirement increases with model size and precision. DeepSeek V3/R1, for example, has 671 billion total parameters. NVIDIA's deployment documentation estimates that its FP8 weights alone require about 671 GB of GPU memory, before accounting for activation tensors and the key-value cache.[20] Deployment therefore requires a multi-GPU system and specialized expertise in model sharding, serving, networking, and optimization. Access to these high-end accelerators may be constrained by their availability, particularly amid strong demand from large AI developers and cloud providers.

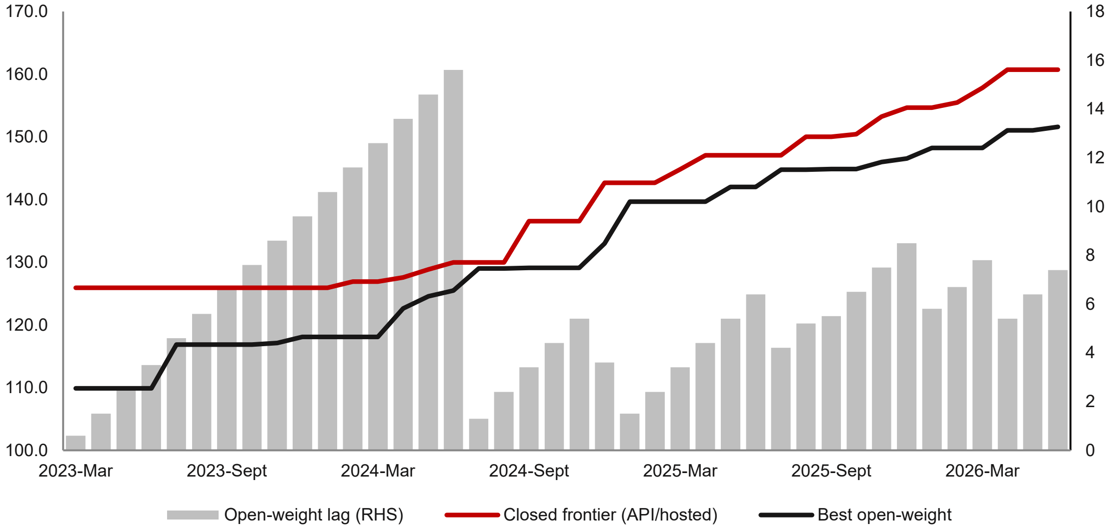
21. Firms can rent this infrastructure from cloud providers, but that relocates dependence and does not remove it. The dependence is different, however: model weights retained by the firm are portable across providers, so reliance shifts from a single vendor's model-access terms to the availability of computing infrastructure. The firm still depends on hardware supply and cloud capacity but retains a fixed model capability that can be redeployed elsewhere.

22. As of now, frontier commercial models have been advancing fast enough to maintain a measurable performance lead over the best open-weight alternative. Frontier capabilities have improved roughly 15 index points per year since April 2024, nearly twice the pace recorded over the preceding two years (Epoch AI 2026c). Although open-weight releases track frontier commercial counterparts at measurable distances—with the

[20] Per installation guide on NVIDIA. NVIDIA, "DeepSeek-V3, DeepSeek-R1, and DeepSeek-V3.2-Exp," *TensorRT-LLM*, "Hardware Requirements," GitHub repository, commit 5b6548439a1b0c431611f72c408d25a011166502, July 17, 2026, accessed August 2, 2026 https://github.com/NVIDIA/TensorRT-LLM/blob/5b6548439a1b0c431611f72c408d25a011166502/examples/models/core/deepseek_v3/README.md

best available open-weight model lagging state-of-the-art closed models by roughly four to six months, the gap between the two has never closed (Figure 13). A self-hosted open model, therefore, presents a capability discount of several months even if a firm redeploys and retrains the open model continuously.

Figure 13. Frontier AI Models: Closed versus Open-weight Capability
(Epoch Capabilities Index; lag in months)

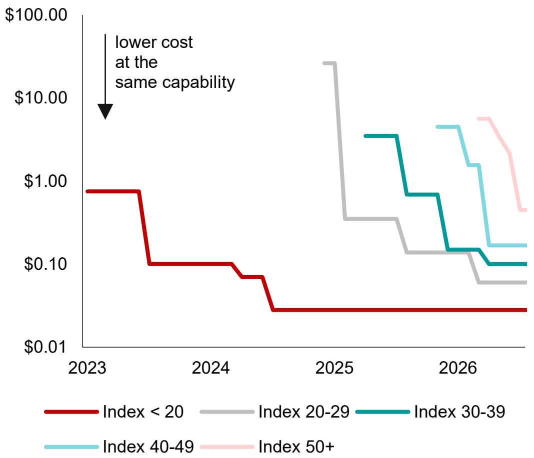


Source: Epoch AI Capabilities Index; AMRO staff calculations
Note: The lag is the number of months elapsed since the closed frontier first reached the current best open-weight score. Equal index scores indicate comparable aggregate benchmark performance, not equivalence on every task.
Data as of June 2026

23. At the same time, using externally provided models also offers users greater flexibility in terms of choices due to the rising competition in quality and costs between leading frontier developers. Switching costs between models are relatively low for individuals, as they can shift between models with minimal interruptions, although firms would still face costs arising from institutional considerations such as workflow integration, security requirements, procurement arrangement, fine-tuning and other customizations. The market is also witnessing increasing competition among new AI research laboratories with declining prices and higher-quality output—the token costs fell by more than 90 percent across model tiers (Figure 14) and the quality-adjusted AI prices fell about 80 percent over two years (André and others 2025). The same study also shows that despite this high contestability, the leading positions remain oligopolistic. A few developers—including OpenAI, Google, Meta, Anthropic, DeepSeek and Mistral—have alternated at the price-performance frontier, with others following closely. For most of the ASEAN+3 region, however, switching models changes the external supplier rather than bringing model development onshore.

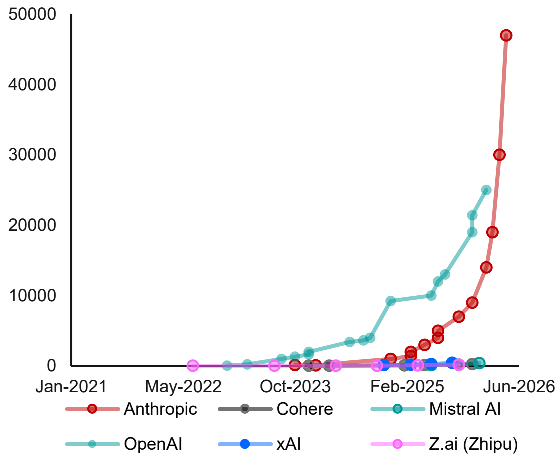
24. As a result of these costs and benefits, renting productivity services from AI providers remains the most affordable means of accessing frontier models for most firms and individuals for now. In this arrangement, the provider manages hosting, query processing, and model updating. Individuals may opt for a free tier, a fixed monthly subscription, or metered token usage. Retail users reach the service through web applications or dedicated desktop and mobile clients. For developers and applications, access can be acquired via API, in which code and applications connect directly to the centralized servers. As suggested by revenue data from some of these leading AI providers, spending on this access model is expanding. OpenAI reached an annualized run rate of roughly USD25 billion in February 2026, around seven times its June 2024 level of around USD3.4 billion. Anthropic, meanwhile, reported about USD47 billion in annualized revenue as of May 2026, up from just about USD9 billion at the end of 2025 (Figure 15).[21] For paid hosted services, access generally continues only while the subscription or usage agreement remains in force.

Figure 14. Selected Models: Inference Price by Intelligence Category
(US dollars per 1 million tokens)



Source: Artificial Analysis; and AMRO staff calculations
Note: Data last accessed on 16 July 2026. Token costs were calculated assuming a 3:1 input-to-output token ratio.

Figure 15. Selected AI Firms: Annualized Revenue
(Millions of US dollars)



Source: Epoch AI, "Data on AI Companies"; and AMRO staff calculations.
Note: Data last accessed on 16 July 2026.

25. Concentration in subscription models, access, and AI production stack presents resilience risks as AI usage scales. For economies that remain mainly digital participants, productivity gains depend on imported models, cloud infrastructure, and foreign platform governance. Access to these models, especially those at the frontier levels, is not guaranteed. In June 2026, the US Department of Commerce issued an export-control directive requiring Anthropic to suspend access to its newly released Fable 5 and Mythos 5 models for all foreign nationals, whether inside or outside of the US, on national-security grounds. The restriction was imposed three days after the models were released and remained in effect until the export controls were lifted on June 30, with access restored from July 1.[22, 23] Although the event was a temporary national-security related restriction, it nevertheless illustrated how access to a rented input can be altered or withdrawn by decisions taken elsewhere. The potential exposure is compounded by the fact that most of the key providers are headquartered in the same country and are therefore exposed to the same national regulatory regime.

26. Unlike oil or other strategic inputs that can be stored, commercial frontier AI largely cannot be stockpiled. A country can hold strategic petroleum reserves to ride out a

[21] Epoch AI. https://epoch.ai/data/ai-companies.
[22] Al Jazeera. https://www.aljazeera.com/news/2026/6/13/us-orders-anthropic-to-disable-ai-models-for-all-foreign-nationals.
[23] Although Fable 5 was commercially available, Mythos 5 was accessible only to selected cybersecurity partners and vetted critical-infrastructure operators.

supply disruption, but for most current commercial AI services, access is a live service that stops when the contract stops, and a model an economy depends on can be retired or changed by its provider. Although firms can stockpile tokens, these represent a right to submit future queries, not ownership of a model or guaranteed access to a particular model version. Their value also varies with provider pricing and model-specific token requirements. As shown by the case of the Fable model access halt, prepaid credits do not constitute a strategic reserve: if a provider retires a model or suspends access, those credits may no longer deliver the expected capability.

27. Using foreign-supplied AI-as-a-service extends and intensifies the existing shift to revocable access already seen in software and digitally delivered services. Earlier packaged software was delivered from a few foreign suppliers but was purchased outright as completed capital goods or under perpetual licenses. If the vendor support ceased or the supplier relationship ended, the users would only lose access to support and future updates but still generally retain and continue using a local copy of the software, preserving the productivity capability embodied in it. Commercial AI services more closely resemble digitally delivered services such as cloud computing and subscription-based software, where continued access depends on continued service subscription. Meanwhile, existing digitally delivered services provide relatively well-defined and bounded functions within an organization, such as computing infrastructure, digital storage, or specific applications. However, frontier AI models provide a more general-purpose capability that can be applied across the organization and embedded across multiple applications and institutional workflows. Moreover, in many use cases, the underlying capability may be tied directly to an imported model specifically—with different capabilities, behaviors, and outputs. Therefore, while losing access to a particular model can be offset by the substitution of another one, the functionality and associated operations could still be disrupted.

28. For firms and institutions, GenAI usage also extends the existing dependence on cloud computing infrastructure, often also provided by US firms. Before the emergence of commercial AI, firms had already shifted to cloud computing to avoid the fixed costs of owning and maintaining infrastructure and to scale computing capacity according to their actual usage. On this front, GenAI is creating additional cloud computing demand for model training, hosting, usage, data processing and deployment. In many cases, cloud computing providers also act as a "distributor," enabling firms to connect their cloud resources directly to GenAI services.

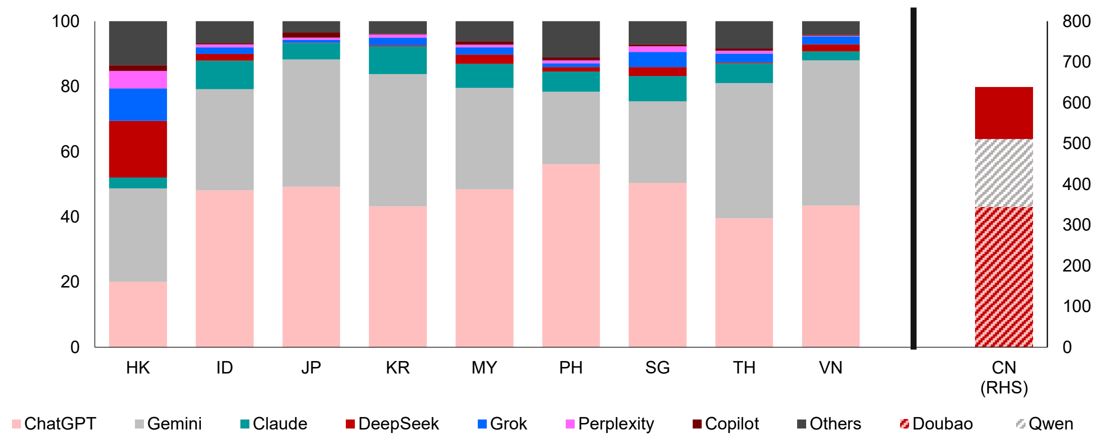
29. The concentration of providers poses additional risks relating to third-party dependencies in the financial sector. As discussed in FSB (2024), the dominance of a few AI providers can lead to increases in market correlations due to common modeling and input data used by pre-trained AI models. It should also be noted that because only a few firms supply these models, a failure or cyber-attack at one provider can affect multiple institutions relying on it (BIS 2024).

V. Implications for the ASEAN+3 Region

30. AI can create economic value for ASEAN+3 through both the adoption of advanced AI capabilities and the establishment of depth in the AI value chain. Importing AI allows firms and individuals to access frontier AI capabilities developed elsewhere without bearing the cost of reproducing the full stack of models, training data, compute, and supporting infrastructure. This model offers an opportunity to deploy AI across existing economic activity at an affordable price. At the same time, regional economies can gain additional sources of economic growth by participating in the AI value chain. As discussed above, several members already occupy important positions, including Korea in high-bandwidth memory, Japan in advanced semiconductor materials and equipment, and Malaysia in semiconductor packaging and other AI-enabling products. These benefits form an AI dividend for regional economies, even for those without a frontier model domestically.

31. As of July 2026, platforms outside the region—predominantly US-based—remain the main channel through which AI capabilities are accessed from most ASEAN+3 economies, except China. Although official data on direct purchases is lacking, web-traffic data used as a proxy for retail platform use shows that US-based AI platforms accounted for roughly 70 to 96 percent of the chatbot platform traffic in most member economies (Figure 16). The provider mix is more diverse in Hong Kong, China, where DeepSeek accounted for about 17 percent of traffic and ranked third, while the Chinese market remains predominantly supplied by domestic platforms.[24] Regardless of AI channels, AI can automate routine components of service work, augment workers on more complex tasks, improve output quality, and expand access to expertise where skilled labor is scarce. The World Development Report 2026 estimates that 16.2 percent of jobs in developing economies could receive a meaningful productivity boost from AI, close to 18.7 percent in high-income economies, and argues that the largest near-term opportunity for developing economies lies in augmenting rather than replacing workers (World Bank 2026).

Figure 16. ASEAN+3: Retail AI Chatbot Traffic Shares
(Percent, millions of monthly active users)



Sources: SimilarWeb and The Straits Times; AMRO staff compilations and calculations
Note: The data included traffic from both desktops and mobile phones. SimilarWeb data are not available for Brunei Darussalam, Cambodia, China, Lao PDR, and Myanmar.
Data are as of Q2 2026 except China which is up to March 2026.
CN = China; HK = Hong Kong, China; ID = Indonesia; JP = Japan; KR = Korea; MY = Malaysia; PH = Philippines; SG = Singapore; TH = Thailand; VN = Vietnam.

32. However, the reliance on foreign AI-as-a-service also introduces three macroeconomic considerations for the region. Given the early stage of adoption, these should be understood as potential channels and downside scenarios rather than established regional outcomes. First, on the growth front, while access to frontier models can potentially raise productivity by lowering the fixed cost of deploying advanced capability, such gains require certain readiness levels that some members lack. Second, if domestic adoption

[24] However, the platform maintained a smaller share in China as other domestic platforms such as Doubao and Qwen gained more popularity.

increases, this demand for imported AI can increase recurring payments for models, cloud computing, software, and related services, and raise imports of digital services. Third, AI adoption may weaken demand for entry-level workers and for routine digitally delivered services currently exported by the region.

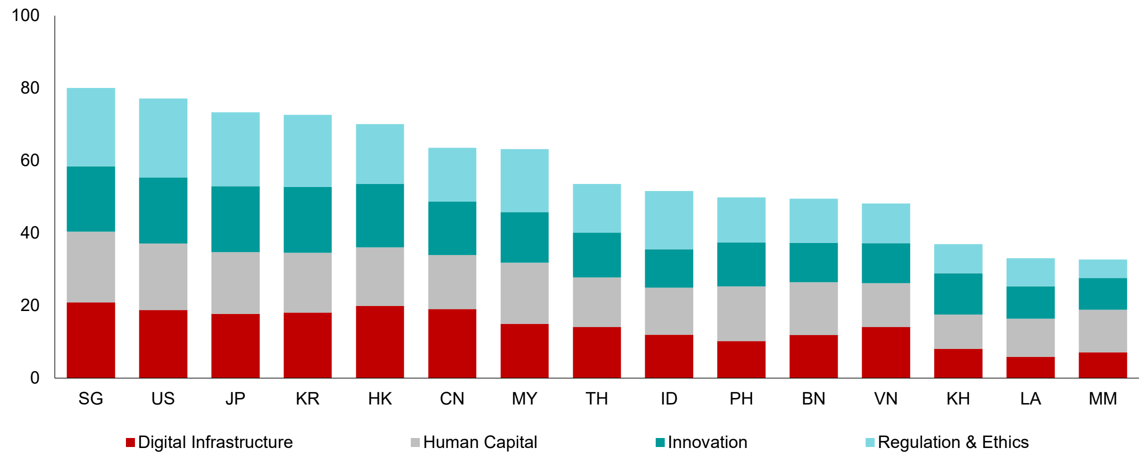
33. Consequently, the net macroeconomic impact of AI imports on ASEAN+3 will depend on three key questions. First, whether economies have sufficient absorptive capacity to translate access to imported AI into productivity gains. Second, whether worker augmentation and creation of higher-value activities can outweigh displacement in routine and entry-level tasks, including services currently exported by the region. Third, whether the productivity, competitiveness, and AI-related export gains generated by greater AI adoption can outweigh the associated increase in payments for imported models, cloud computing, software, and related services, as well as any weakening of existing services exports. These balances are likely to differ considerably across economies depending on their readiness and position in the AI value chain. Given the still-early stage of AI adoption, they should therefore be viewed as conditions shaping potential outcomes rather than as evidence of a common regional effect.

34. First, access to frontier AI does not by itself generate an AI dividend, which relies on whether an economy can integrate AI into production and diffuse its use across firms to create economic impacts. The World Bank (2025) identifies four foundations—its "4Cs": connectivity, including reliable electricity and digital networks; compute, including affordable cloud and processing capacity; context, including locally relevant data, languages, and applications; and competency, covering the skills needed to integrate AI into production. Economies do not need to develop all these capabilities domestically to benefit from AI. In particular, smaller economies can import models and computing services rather than build their own advanced infrastructure. Nonetheless, the combination and depth of capabilities required increase with the level of technological ambition. Adopting off-the-shelf tools mainly requires connectivity and foundational digital skills; adapting models requires local data and specialized talent; and developing new models requires advanced compute, research capacity, and technical expertise. Depending on the economy's technological ambition, weaknesses in any of the four foundations can therefore limit the extent to which access can translate into broad-based productivity gains.

35. ASEAN+3 members differ considerably in their broader preparedness to capture and manage these gains. Complementing the World Bank's focus on the technological foundations for AI adoption, the IMF's AI Preparedness Index assesses four wider dimensions: digital infrastructure; human capital and labor-market policies; innovation and economic integration; and regulation and ethics. The index puts Singapore, Japan, Korea, and Hong Kong, China above 70 on a 0–100 scale, within the same group as the US. By contrast, six of the region's 14 members score below 50: Brunei Darussalam, Cambodia, Lao PDR, Myanmar, the Philippines, and Vietnam (Figure 17).[25] Members that lag in AI readiness may therefore only be able to capture smaller gains later even when frontier models and funding are available.

[25] IMF AI Preparedness Index (AIPI) dataset, accessed on July 28, 2026. https://www.imf.org/external/datamapper/datasets/AIPI.

Figure 17. ASEAN+3 and the US: AI Preparedness Index and Sub-pillars



Source: IMF AI Preparedness Index Dataset; AMRO staff calculations
Note: Data are for 2023. For brevity, the names of the sub-pillars have been shortened. “Human Capital” refers to “Human Capital and Labor Market Policies,” while “Innovation” refers to “Innovation and Economic Integration.” BN = Brunei Darussalam; CN = China; HK = Hong Kong, China; ID = Indonesia; JP = Japan; KH = Cambodia; KR = Korea; LA = Lao PDR; MM = Myanmar; MY = Malaysia; PH = Philippines; SG = Singapore; TH = Thailand; US = United States; VN = Vietnam.

36. Second, greater AI adoption could reshape ASEAN+3 labor markets through both worker augmentation and task substitution, with the balance likely to differ across levels of seniority, economies, and occupations. Productivity gains can expand output and employment when firms use AI to complement workers, while automation can reduce labor demand for tasks that models can perform directly. Both job creation and displacement could occur simultaneously, with transitional unemployment rising as workers move across sectors (ADB 2026). Unlike previous waves of automation that were concentrated in manufacturing, GenAI exposure can extend to cognitive and digitally delivered tasks in finance, information and communication, education, and professional services, all of which carry larger weights in the economic structure of higher-income countries. Gauging beyond exposure, a cross-country study using job posting data finds that postings in more AI-vulnerable occupations declined by 5.8 percent on average relative to less vulnerable ones in high-income economies, but less so in non-high-income countries (Huang and others 2026). The domestic outcome depends less on the number of occupations exposed than on whether firms use AI primarily to augment or replace their workers (Huang 2025).

37. The same two-sided effects extend to the region’s digital export sectors, in which AI adoption can help workers enhance productivity while allowing foreign clients to automate tasks that were previously outsourced to the region. Digitally delivered service work is more exposed than most domestic services, and clerical occupations—common in many outsourced service activities—are among the most exposed (Cazzaniga and others 2024; ILO 2025). Many highly exposed activities overlap with commonly outsourced tasks, including customer support, data processing, document preparation, translation, and junior back-office functions. At the same time, AI can help these workers handle more requests or address higher value work for clients. The net effect on service exports will therefore depend on whether productivity upgrading and movement into higher-value work outpace substitution for routine tasks.

38. **Early evidence suggests that some signs of substitution are already occurring, although broader displacement has not yet materialized**. US firms that relied most heavily on online outsourced labor before ChatGPT were among the earliest AI adopters and have subsequently cut spending on contracted labor, while only spending a small fraction of the savings on AI services (Stevens 2026). Similarly, evidence of erosion appears on freelance platforms, where workers in exposed occupations saw about two percent fewer contracts and five percent lower earnings after ChatGPT's release, with demand for writing and translation falling 20 to 50 percent relative to trend (Hui and others 2024; Teutloff and others 2025). These developments could shift outsourcing away from labor-cost arbitrage toward capability arbitrage: clients could look to automate the routine layer of a task directly and retain fewer, higher-skilled workers to supervise, verify, and integrate the output.

39. **Case studies from the Philippines illustrate both the continued resilience of the sector and emerging pressures facing the business process outsourcing (BPO) sector.** Employment in the Philippines' IT and business-process-management sector continued to grow after the commercial introduction of GenAI, reaching about 1.9 million in 2025, around 20 percent above 2022. However, new labor demands have shown signs of a marked slowdown. IT and BPO new job postings in the country have declined sharply since their 2022 peak, dropping to less than half of that level by 2024 (Bergman and others 2025). Although some services such as managing complex customer interactions and handling high-value queries remain difficult to automate by AI, routine and simpler tasks could be automated, resulting in a net reduction in hiring of additional employees.[26] Going forward, the impact of AI could be more profound as suggested by the forecast by the IT and Business Process Association of the Philippines—its 2028 revenue target was reduced from USD59 billion to USD43.3–50.5 billion and its employment target from 2.5 million to 1.85–2.14 million.[27]

40. **A risk of longer-term capability loss could emerge if AI substitutes for entry-level work faster than firms and workers can move into higher-value activities.** ICT and business-process services are not only sources of income and foreign currency in the case of exported services, but they also form part of the broader ecosystem through which economies accumulate digital and technical capabilities. The World Bank notes that tacit capabilities such as project management, judgment, and other experience-based skills develop through learning by doing and warns that reduced demand for entry-level workers could weaken these opportunities over time (World Bank 2026). If clients automate routine and entry-level work faster than firms and workers can move into more complex activities, some of these learning and upgrading pathways could narrow. Over time, this could leave the region with a thinner domestic talent base for adapting, integrating, and eventually developing AI, reinforcing reliance on imported AI services. Evidence from the US suggests that employment among young workers aged 22–25 has weakened disproportionately in occupations highly exposed to AI. After controlling for firm-level shocks, employment for young workers in the most AI-exposed occupations stood about 19 percent below that for young workers in the least-exposed occupations (Brynjolfsson and others 2026).

---

[26] Per report by The Economist (2026). https://www.economist.com/asia/2026/08/06/the-philippines-big-offshoring-industry-is-growing-despite-ai

[27] GMANetwork. https://www.gmanetwork.com/news/money/economy/994797/ibpap-cuts-2028-it-bpm-revenue-jobs-forecasts-after-roadmap-review/story/.

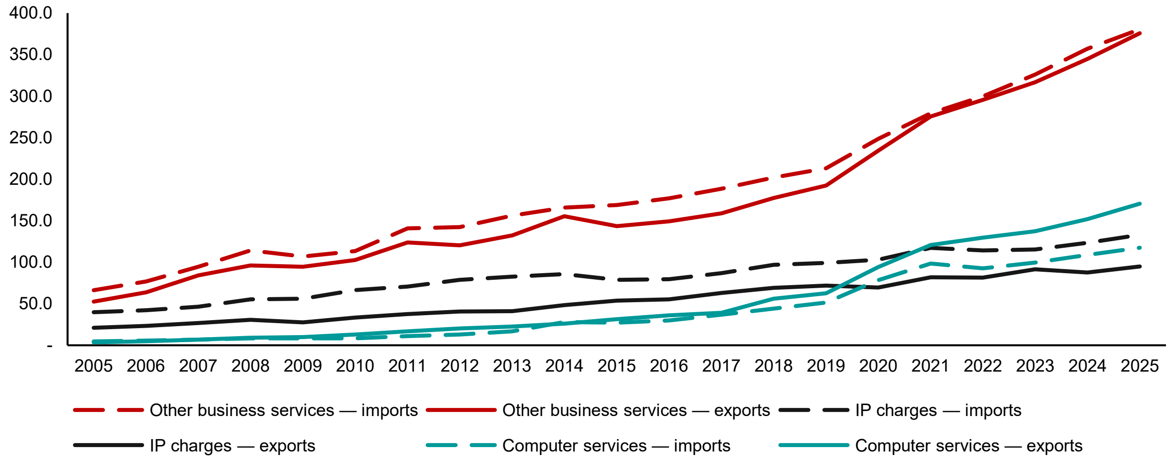
41. Finally, for the external sector, the net effect of imported AI would depend on increases in recurring AI payments alongside potential changes in productivity, competitiveness, and the region's exports of AI-related and digital services. Domestic adoption creates recurring payments for models, cloud computing, software, and related services, while adoption abroad may reduce demand for some business-process and digitally delivered services exported by ASEAN+3. In 2025, although ASEAN+3 remained a net exporter of computer services, with the surplus increasing since 2020, imports of intellectual-property services continued to exceed exports, while the balance in other business services was weakening (Figure 18). As discussed earlier, AI could place pressure on both positions: reduced demand for BPO and other digitally delivered services could weaken the other-business-services balance, while a reliance on proprietary models, software, and cloud infrastructure could widen the deficit in related AI charges. On the other hand, these pressures can be offset where AI raises the competitiveness of domestic firms or where economies export AI-enabling products, data-center services, or other parts of the AI value chain. As discussed earlier, Malaysia, Vietnam, and Thailand, for example, already export significant AI-enabling electronics and computing products.

42. External-sector pressures would be greater where increases in AI-related payments are not matched with sufficient productivity gains or AI-related exports. Economies that are weakly integrated into the AI value chain and have limited capacity to translate imported AI into productivity gains, higher-value employment, and export competitiveness could face a downside scenario. In such cases, AI-related domestic value creation or exports fail to offset the recurring import bills for AI models, cloud computing, software, and related services while domestic firms could lose competitiveness to foreign firms that adopt AI more effectively. The predominance of US-based providers may also add a potential dollar channel, although the actual size of such flow would depend on settlement currencies and the offsetting effects of the domestic AI-related exports, AI-related services exports, and productivity gains from AI.

43. The magnitude of these external-sector effects is difficult to establish because AI-related transactions are not separately identifiable in official statistics. Payments may be bundled into broader cloud contracts, routed through regional headquarters, or supplied by local affiliates, making the AI component difficult to separate from other digital expenditure.[28] The broader categories in which these transactions may be recorded are already among the region's largest digitally delivered services imports: charges for the use of intellectual property, for example, are the largest import subsector for Singapore, Vietnam, Malaysia, and Thailand, reflecting reliance on proprietary foreign digital products (Tham 2025). Detailed balance-of-payments data therefore need to be complemented by payment settlement data, tax records, enterprise and ICT-use surveys, cloud-procurement records, and provider-level disclosures.

[28] Classification may vary depending on whether users purchase consumer subscriptions, enterprise software, cloud computing, API access, or consulting services. The recorded transaction may also differ from the underlying economic relationship. For instance, a local Singaporean firm using services from Amazon Web Services (AWS) Singapore may pay in Singapore dollars and such transactions are not recorded in the balance of payments. However, AWS Singapore may subsequently make cross-border payments for services, cloud capacity, intellectual-property rights, equipment, or profit remittances.

Figure 18. ASEAN+3: Selected Digitally Delivered Service Trade
(Billions of US dollars)

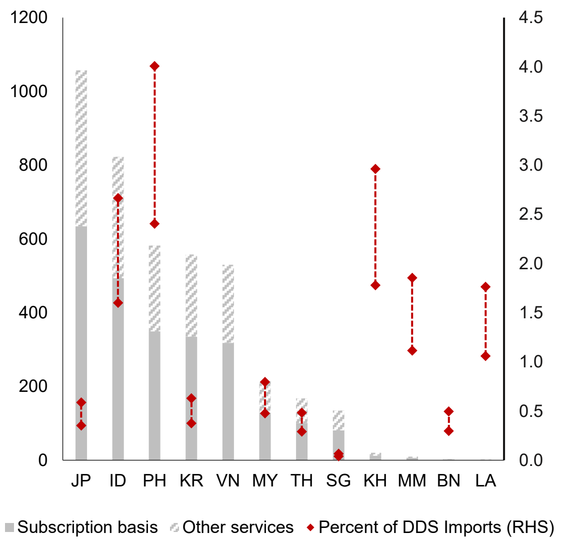


Source: WTO Digitally Delivered Services dataset; AMRO staff calculations
Note: Selected service categories through which AI-related transactions or effects may be reflected are other business services, charges for the use of intellectual property not included elsewhere, and computer services. Partner = world.

44. **Rough estimates of AI-related payments suggest that the imports of AI services could be non-negligible in some economies, although the exact value is still difficult to ascertain.** AMRO staff's back-of-the-envelope calculations suggest that annual expenditure attributable to OpenAI services in ASEAN+3 excluding China in 2025 could amount to as much as USD4.1 billion, depending on whether the calculation allocates only consumer-subscription revenue or OpenAI's total revenue using each economy's share of global ChatGPT web traffic (Figure 19). This represents about 0.7 percent of digitally delivered services (DDS) imports for the selected economies but could reach roughly 4 percent in the Philippines and 3 percent in Cambodia.[29] The estimate excludes expenditure on Gemini, Claude, API usage, and AI services embedded in cloud and enterprise-software contracts. It should therefore be interpreted as an order-of-magnitude expenditure proxy rather than a measured import flow.

[29] Vietnam's DDS share is not reported because available estimates differ materially across sources, reference years, and methodologies and are not directly comparable. Using the WTO DDS figure, the share for Vietnam is around 6.3 - 10.5 percent.

Figure 19. ASEAN+3 ex China: Estimated Annual "Imports" of ChatGPT
(Millions of US dollars, Percent of total digitally delivered service imports)



Economy	ChatGPT traffic share (percent)	Imports subscription basis (USD mn)	Imports full OpenAI revenue (USD mn)	Percent of 2025 DDS Imports (percent)
Japan	4.23	634.50	1,057.50	0.35 - 0.59
Indonesia	3.29	493.50	822.50	1.6 - 2.67
Philippines	2.33	349.50	582.50	2.4 - 4.01
Korea	2.23	334.50	557.50	0.38 - 0.63
Vietnam*	2.12	318.00	530.00	-
Malaysia	0.86	129.00	215.00	0.48 - 0.8
Thailand	0.67	100.50	167.50	0.29 - 0.48
Singapore	0.54	81.00	135.00	0.04 - 0.07
Cambodia	0.08	12.00	20.00	1.78 - 2.96
Myanmar	0.04	6.00	10.00	1.11 - 1.86
Brunei	0.01	1.50	2.50	0.3 - 0.5
Lao PDR	0.01	0.75	1.25	1.06 - 1.76

Source: OpenAI, SimilarWeb, AMRO staff calculations
Note: The figure and table provide a rough estimate of AI service imports from OpenAI.
The subscription component is calculated as the percentage of OpenAI's reported annualized revenue multiplied by each economy's share of global ChatGPT web traffic from January to June 2026. As inferred from the company announcement in 2026, roughly 60 percent of OpenAI revenue is from consumer subscriptions. Other OpenAI revenue represents the remaining roughly 40 percent of reported annualized revenue, allocated using the same traffic shares.
The calculation assumes equal paid-user conversion rates and average revenue per paid user across economies. Web traffic does not directly measure enterprise or API usage; allocating those revenues using retail web traffic therefore introduces additional uncertainty. The estimates should be interpreted as order-of-magnitude expenditure proxies rather than measured imports. DDS imports are calculated by the WTO from balance of payments services statistics, using the Eurostat–WTO model to estimate the share delivered cross-border through digital networks. DDS import data refer to 2025, while the OpenAI revenue and traffic inputs refer to 2026, reflecting data availability. Values refer to the latest available annual observations. Refer to the WTO DDS trade dataset for more information. BN = Brunei Darussalam; ID = Indonesia; JP = Japan; KH = Cambodia; KR = Korea; LA = Lao PDR; MM = Myanmar; MY = Malaysia; PH = Philippines; SG = Singapore; TH = Thailand; VN = Vietnam.
* Available international estimates of Vietnam's digitally deliverable services imports are internally inconsistent and are not used.

45. **Relying on foreign models could pose threats beyond the withdrawal of access.** Even when access is uninterrupted, the terms on which it is supplied remain outside the users' control. The dependency is therefore not only about whether the service continues, but about what content is delivered and how, at what price, and how well it fits local conditions. Some of these challenges apply to commercial and open-weight models alike:

- **Models trained mainly on foreign data misread local context.** Performance may degrade for the region's lower-resource languages and scripts, and models carry weaker knowledge of domestic law, institutions, and administrative practice. The gap is larger for members with the least capacity to correct it through fine-tuning or local evaluation.

- **Content and use rules embedded in the models are outside users' control.** Providers' policies, safety filters, and refusal behavior may not align with a user's own priorities. Training data are not disclosed even for most open-weight models, so these embedded rules cannot readily be audited.

Commercial models create additional exposure because providers can update or retire models. The same prompt may produce a different answer from one month to the next, weakening reproducibility and traceability in processes built on the model. Pricing and token terms are also set unilaterally, while the handling of usage data raises data residency and confidentiality concerns.

VI. Policy Discussion and Conclusion

46. High concentration in the AI supply chain and the current ease of access suggest that foreign AI will remain an important source of AI adoption and productivity gains for most ASEAN+3 economies. Frontier models and their supporting infrastructure are concentrated among a small number of global firms, but they remain widely accessible and, for now, offer the most practical means for public institutions and firms to experiment with and deploy advanced AI capabilities. By contrast, replicating the entire stack—from advanced semiconductors and compute to cloud platforms and foundation models—would be prohibitively costly and technically unrealistic for most ASEAN+3 members. Attempting to achieve self-sufficiency across every layer could therefore divert scarce resources, raise adoption costs, and delay the productivity gains available from existing foreign technologies.

47. The policy objective should therefore be neither technological self-sufficiency nor passive dependence on foreign providers, but a balance between strategic participation in the AI value chain and efficient access to capabilities developed elsewhere. As discussed, ASEAN+3 members occupy diverse positions across the AI value chain: some possess globally significant capabilities in semiconductors, memory, equipment, cloud, or foundation models, while others participate mainly through infrastructure hosting, model adaptation, and downstream use. Economies need not develop capability across the entire stack to capture the AI dividend; instead, building competitive capabilities in selected segments can secure added value and strengthen resilience while relying on external sources to fill gaps elsewhere. This access is particularly important for the foundation model layer, which serves as a key interface through which upstream advances in compute, data, and algorithms are translated into downstream applications and productivity gains. The appropriate balance between domestic capability and external dependence will consequently differ across economies and use cases.

48. These considerations give rise to a trade-off among AI capability, control, and cost. As summarized in Table 1, different modes of AI access involve distinct trade-offs across these three dimensions.

- At present, the most capable AI systems are concentrated among a small number of leading private firms that possess vast user bases, extensive data resources, advanced technological infrastructure, and scarce technical talent. Users can access these systems with minimal upfront investment and pay largely according to usage. However, the continuity and terms of access are not fully guaranteed, while the systems' underlying design, training processes, and behavior remain largely beyond the control of end-users.

- Owning and developing a proprietary AI model offers the greatest degree of control and potentially the most reliable access. However, it requires substantial and sustained investment in computing infrastructure, data, technical expertise, and model maintenance, with no assurance that the resulting system will match or surpass the capabilities of frontier models.

- Deploying foreign open-weight models represents an intermediate option. They can provide greater control and customization than reliance on proprietary external services, while requiring less investment than developing a model from scratch. Nevertheless, users remain dependent on design choices made by the original model developers, and open-weight models may continue to lag the most capable frontier systems.

Table 1. Ladder of Control over AI Capability

Mode of access	Capability	Cost	What users can control	What remains outside users' control	Best suited to
Public interface or API to a frontier model	Frontier	- Lowest - Metered payment or subscription	- Prompt instruction - Workflow built around the AI	- Model version and its parameters - Price and availability - Training content rules	- Low-stakes, high-volume work Individual users and developers
Frontier-model API combined with a proprietary retrieval layer or vendor-supported fine-tuning	Frontier	- Operating cost - Data engineering and storage	*All of the above, plus:* - The context and usage data - Partial behavior direction	- Model version and its parameters - Price and availability - Training content rules	- Institutions with proprietary or confidential data - Interruption risk still tolerable
Open-weight model run on a foreign cloud	Near frontier	*Operating costs at the above level, plus:* - Cloud rental	*All of the above, plus:* - The model behavior - Fine-tuning ability - Version control	- The origin of the weights and training data - Inference infrastructure - Data and model hosting jurisdiction - Chip supply	- Users needing a stable model version - Protection from repricing or deprecation - Foreign hosting acceptable
Open-weight model run on domestic compute	Near frontier	- High and largely fixed cost - Accelerators, power, staffing costs - Operating cost above	*All of the above, plus:* - The inference system - Full input and output data governance	- The origin of the weights and training data - Chip supply	*Same needs as above, plus:* - Full security and control over the data - Institutions with proprietary or confidential data - Payments, official statistics, public records
Domestically trained and hosted model	Depending on investment	- Very high - Training infrastructure, data science and R&D - Training run costs - Operating cost above	*All of the above, plus:* - Training inputs - Behavior rules - All layers above the chip supply	- Chip supply - The capability gap with the frontier models	- National capability programs - Building domestic capability and strategic autonomy

Source: Authors
Note: The table describes the position of the deploying user, not the ownership of the underlying technology.

49. The AI deployment trade-off provides a practical starting point for considering how member economies should manage dependence on imported AI. For most members, the realistic approach will be to balance the three dimensions rather than maximize any one of them. The appropriate position within this framework will depend on the function being performed, the resources available, and the risks that firms and authorities are prepared to bear. These choices should also be reassessed regularly as model capabilities, deployment costs, and the strategic value of control evolve.

50. First, dependence should be made productive by strengthening the domestic capacity to convert AI access into economic gains and ensuring a long-lasting value for the workforce. Members with stronger connectivity, compute access, locally relevant data, technical skills, and digitally capable firms are better positioned to capture the gains, while less-prepared members could incur similar dependence without receiving a comparable dividend. This could produce a two-speed AI region in which the same technology widens rather than narrows intra-regional income gaps (Abdurohman and Huang 2026). Investment in absorptive capacity should therefore be a common priority across all members: skills, data readiness, organizational change, and model-integration capabilities are domestic capabilities that every member can develop and retain. Inward AI-related investment should similarly be encouraged on terms that deepen domestic capability, including through workforce-development commitments, partnerships with local firms and research institutions, access to compute, and the establishment of engineering and model-adaptation functions, rather than infrastructure hosting or manufacturing alone. These measures to strengthen absorptive capacity should also preserve pathways for junior workers to acquire experience and advance into higher-value roles. Authorities could monitor changes in entry-level vacancies and hiring, support AI-enabled apprenticeships and training partnerships, and encourage firms to redesign junior roles around verification, judgment, client interaction, and other complementary tasks rather than eliminate entry-level positions altogether.

51. **Second, from a resilience perspective, dependence should be diversified across providers, models, infrastructure, and modes of access.** Competition among frontier providers gives users some scope to avoid lock-in. However, subscribing to several providers does not by itself ensure meaningful diversification if data, workflows, and application layers cannot be transferred. Governments, regulated institutions, and firms with material AI exposure should therefore promote interoperable application layers, retain data and evaluation records in portable formats, avoid embedding critical workflows irreversibly within a single proprietary system, and periodically test alternative providers or deployment arrangements. Authorities can also support the development of domestic AI applications, middleware, and integration capabilities that are designed to work across multiple models and providers. This can lower switching costs, widen the range of usable AI systems, and reduce dependence on any single upstream provider without requiring domestic replication of frontier models. Public procurement can support this objective through model-neutral technical standards, portability requirements, transparent model version management, and credible exit provisions.

52. **Third, dependence should be made measurable.** As discussed above, AI subscriptions, API usage, and related cloud charges are generally recorded within broad balance of payments categories such as computer services and charges for the use of intellectual property and may be bundled into larger contracts or booked through offshore regional headquarters. Few members could currently estimate their AI import bill, identify their provider concentration, or assess the exposure of economically important activities to a material repricing, degradation, restriction, or withdrawal of access. A dependence that cannot be measured cannot be adequately monitored, priced, or managed. Authorities should therefore begin developing indicators covering AI-related service imports, provider concentration, usage in critical sectors, exposure of employment and services exports, domestic readiness, and the availability of substitutes. Such monitoring should distinguish between productive dependence, where imported AI generates greater domestic value added and capability, and fragile dependence, where important activities rely on concentrated services without credible alternatives.

53. **Fourth, dependence should remain reversible, particularly for economically or institutionally critical uses.** Reversibility does not require an economy to reproduce frontier

capability domestically; rather, it means preserving the ability to migrate workflows, substitute providers, or maintain essential functions when access conditions change. Members could therefore consider maintaining a "controllable productivity floor" through open-weight models, assured access to compute, locally controlled data and retrieval systems, and the technical expertise needed to operate them. Although such a floor would likely be less capable than leading proprietary systems, it could preserve part of the productivity gains in the event of disruption, withdrawal, or substantial repricing of frontier access. Its effectiveness would depend not merely on possessing model weights, but on integrating the model into relevant workflows and periodically testing whether it can serve as a practical fallback.

54. The development of a controllable productivity floor can coexist with continued access to the best available model and support for local or regional models that better reflect linguistic, cultural, and public sector context. The region already contains several possible templates. AI Singapore's SEA-LION[30] is an open-weight family of models covering Southeast Asian languages and evaluated through the SEA-HELM benchmark; Indonesia's Sahabat-AI[31] extends this approach through cross-border collaboration among local firms and AI Singapore. Malaysia's ILMU[32] represents a further national effort to develop locally relevant capability. Such initiatives can support language coverage, public sector applications, skills development, and domestic experimentation without necessarily seeking to compete directly with the largest frontier models. Where frontier-model development is pursued, it should be justified by sufficient scale, clearly defined use cases, sustained funding, and transparent external benchmarks of performance.

55. Regional cooperation could help reduce the fixed costs of capability development, improve the measurement of dependence, and expand the options available to smaller members. Cooperation should initially focus on public goods for which scale, linguistic diversity, and shared standards are especially valuable. These could include multilingual datasets, common evaluation and safety benchmarks, shared or reciprocal access to research compute, model-portability and procurement standards, technical training, and information-sharing on service disruptions and provider concentration. A regional approach could also support the tracking of AI-related digitally delivered services, whose payments may pass through cloud providers, resellers, payment intermediaries, and regional headquarters across several jurisdictions before reaching the ultimate service provider. Harmonized definitions, reporting templates, and information-sharing among statistical and regulatory authorities can help reduce double counting, identify the origin and destination of such flows, and improve estimates of members' AI-related imports and provider exposure. Meanwhile, coordinated procurement or aggregated access arrangements could likewise improve the terms available to public institutions and researchers without requiring each member to develop a national frontier model. Members with more advanced infrastructure could contribute computing power and technical

---

[30] Raymond Ng and others, "SEA-LION: Southeast Asian Languages in One Network," arXiv:2504.05747, April 8, 2025.
https://arxiv.org/abs/2504.05747.

[31] Indosat Ooredoo Hutchison and GoTo Group, "Indosat Ooredoo Hutchison and GoTo Launch Sahabat-AI: Indonesia's Open-Source LLM for Empowering Digital Sovereignty," November 15, 2024.
https://www.gotocompany.com/en/news/press/indosat-ooredoo-hutchison-and-goto-launch-sahabat-ai-indonesias-open-source-llm-for-empowering-digital-sovereignty.

[32] YTL Power International, "YTL AI Labs Launches ILMU—100% Malaysian AI, Built by Malaysians for Malaysians," August 12, 2025.
https://www.ytlpowerinternational.com/press-releases/ytl-ai-labs-launches-ilmu-100-malaysian-ai-built-by-malaysians-for-malaysians/.

expertise, while others could contribute languages, local data, sectoral knowledge, and use cases.

56. The region should employ a careful approach to balance domestic capacity building and technological autarky. Support should focus on encouraging innovation alongside consumer protection instead of introducing blanket restrictions on foreign models. Policy should therefore scale with the sensitivity of the use case. Where interruption would carry material economic or public consequences, controls and continuity requirements should be stricter. Domestic and regional capability development should complement continued access to global innovation rather than attempt to replace it across the whole AI stack.

57. ASEAN+3 does not need to own every layer of the AI stack to secure an AI dividend. Foreign models and infrastructure will remain the most capable and economical option for many uses, and continued openness to them will be important for productivity. The durability of the dividend, however, will depend on whether members can convert access into domestic value added, avoid excessive concentration, understand the scale and criticality of their exposure, and preserve credible alternatives for essential functions. The appropriate objective is therefore neither complete technological self-sufficiency nor passive dependence, but open access supported by domestic absorptive capacity, diversified suppliers, better measurement, and a minimum level of controllable capability.

References

Abdurohman and Xianguo Huang. 2026. “Can ASEAN+3 Economies Still Escape the Middle-Income Trap in the AI Era?” *BusinessWorld*, January 27, 2026. https://bworldonline.com/opinion/2026/01/27/726498/can-asean3-economies-still-escape-the-middle-income-trap-in-the-ai-era/.

André, Christophe, Manuel Bétin, and Peter Gal. 2025. “Dynamism in Generative AI Markets Since the Release of ChatGPT.” VoxEU, Centre for Economic Policy Research, July 12. https://cepr.org/voxeu/columns/dynamism-generative-ai-markets-release-chatgpt

Asian Development Bank (ADB). 2026. “AI Readiness and Economic Impacts in Asia and the Pacific.” Special Topic, *Asian Development Outlook, April 2026*, Manila, April. https://www.adb.org/publications/asian-development-outlook-april-2026

Bank for International Settlements (BIS). 2024. “Artificial Intelligence and the Economy: Implications for Central Banks.” BIS Annual Economic Report 2024, Chapter III, Basel, June 25. https://www.bis.org/publ/arpdf/ar2024e3.htm

Bekkers, Eddy, Lee Humphreys, Hryhorii Kalachyhin, Karolina Wilczynska, and Danchen Zhao. 2025. “Through the Looking Glass: Artificial Intelligence, International Trade, and Economic Growth in the Long Run.” WTO Staff Working Paper ERSD-2025-09, Geneva. https://www.wto.org/english/res_e/reser_e/ersd202509_e.pdf

Bergman, Johan Bjurman, and He Wang. 2025. “Co-worker, Coach, or Competitor? How AI Is Transforming the Future of Digitally Delivered Services.” Digital Progress and Trends Report 2025, Case Study 3, World Bank, Washington, DC. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099112525160528750

Brynjolfsson, Erik, Danielle Li, and Lindsey R. Raymond. 2023. “Generative AI at Work.” NBER Working Paper 31161. https://doi.org/10.3386/w31161

Brynjolfsson, Erik, Bharat Chandar, and Ruyu Chen. 2026. “Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of Artificial Intelligence.” Stanford Digital Economy Lab, Stanford, CA, revised August 12, 2026. https://digitaleconomy.stanford.edu/publication/canaries-in-the-coal-mine-six-facts-about-the-recent-employment-effects-of-artificial-intelligence/

Cazzaniga, Mauro, Florence Jaumotte, Longji Li, Giovanni Melina, Augustus J. Panton, Carlo Pizzinelli, Emma J. Rockall, and Marina Mendes Tavares. 2024. “Gen-AI: Artificial Intelligence and the Future of Work.” IMF Staff Discussion Note SDN/2024/001, Washington, DC, January 14. https://www.imf.org/en/publications/staff-discussion-notes/issues/2024/01/14/gen-ai-artificial-intelligence-and-the-future-of-work-542379

Competition and Markets Authority (CMA). 2024. “AI Foundation Models: Technical Update Report.” London, April 16. https://www.gov.uk/government/publications/ai-foundation-models-update-paper

Cottier, Ben, Robi Rahman, Loredana Fattorini, Nestor Maslej, Tamay Besiroglu, and David Owen. 2024. “The Rising Costs of Training Frontier AI Models.” arXiv:2405.21015, May 31.
https://arxiv.org/abs/2405.21015

Epoch AI. 2026a. “Data on AI Models.” Notable AI models dataset, accessed in July 2026.
https://epoch.ai/data/ai-models

______. 2026b. “AI Data Centers.” Dataset, accessed in July 2026.
https://epoch.ai/data/ai-data-centers

______. 2026c. “AI Capabilities Progress Has Sped Up.” Data insight, accessed on July 3, 2026.
https://epoch.ai/data-insights/ai-capabilities-progress-has-sped-up

Financial Stability Board (FSB). 2024. “The Financial Stability Implications of Artificial Intelligence.” Report to the G20, Basel, November 14.
https://www.fsb.org/2024/11/the-financial-stability-implications-of-artificial-intelligence/

Huang, Jingyun, Yan Liu, He Wang, and Shu Yu. 2026. “Who Takes the Hit? The Uneven Impacts of Generative AI on Labor Demand Across Countries.” World Development Report 2026: Background Papers, July 27.
https://thedocs.worldbank.org/en/doc/07f7fe676b8331172af31de2c3fc11e1-0050062026/original/Cross-country-Labor-Demand-in-the-age-of-GenAI-0727.pdf

Huang, Xianguo. 2025. “Labor Market Exposure to AI: From GenAI to Future AGI.” AMRO Working Paper. ASEAN+3 Macroeconomic Research Office. November 10, 2025.
https://amro-asia.org/labor-market-exposure-to-ai-from-genai-to-future-agi.

Huang, Xianguo, and Yin Fai Ho. Forthcoming. “Global Cross-Border Data Center Investment - An ASEAN+3 Perspective.” ASEAN+3 Macroeconomic Research Office.

Hui, Xiang, Oren Reshef, and Luofeng Zhou. 2024. “The Short-Term Effects of Generative Artificial Intelligence on Employment: Evidence from an Online Labor Market.” *Organization Science* 35 (6).
https://doi.org/10.1287/orsc.2023.18441
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4527336

International Labour Organization (ILO). 2025. “Generative AI and Jobs: A Refined Global Index of Occupational Exposure.” ILO Working Paper 140, Geneva, May 20.
https://www.ilo.org/sites/default/files/2025-05/WP140_web.pdf

______. 2026. “Generative AI and Labour Markets in ASEAN: Significant Exposure, Limited Disruption, Uneven Preparedness.” Bangkok, July 8.
https://www.ilo.org/publications/generative-ai-and-labour-markets-asean-significant-exposure-limited

International Monetary Fund (IMF). 2026. “AI Preparedness Index (AIPI).” IMF DataMapper, accessed on July 28, 2026.
https://www.imf.org/external/datamapper/datasets/AIPI

Korinek, Anton, and Jai Vipra. 2025. “Concentrating Intelligence: Scaling and Market Structure in Artificial Intelligence.” *Economic Policy*, Volume 40, Issue 121, 225–256, January.
https://academic.oup.com/economicpolicy/article-abstract/40/121/225/7905140

McKinsey & Company. 2025. “The State of AI in 2025: Agents, Innovation, and Transformation.” QuantumBlack, AI by McKinsey, November 5. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai

Menlo Ventures. 2025. “2025: The State of Generative AI in the Enterprise.” December 9. https://menlovc.com/perspective/2025-the-state-of-generative-ai-in-the-enterprise/

METR. 2025. “Measuring AI Ability to Complete Long Tasks.” March 19. https://metr.org/blog/2025-03-19-measuring-ai-ability-to-complete-long-tasks/

Organisation for Economic Co-operation and Development (OECD). 2025a. “Macroeconomic Productivity Gains from Artificial Intelligence in G7 Economies.” OECD Artificial Intelligence Papers No. 41, Paris. https://oecd.ai/en/ai-publications/macroeconomic-productivity-gains-from-artificial-intelligence-in-g7-economies

______. 2025b. “Competition in Artificial Intelligence Infrastructure.” OECD Roundtables on Competition Policy Papers, Paris, November 13. https://www.oecd.org/en/publications/competition-in-artificial-intelligence-infrastructure_623d1874-en.html

______. n.d. “OECD.AI Policy Observatory: Policy Navigator.” Accessed in July 2026. https://oecd.ai/en/dashboards/overview

Our World in Data. 2026. “Estimated Share of Working-Age Adults Who Use Generative AI.” Based on Microsoft AI Diffusion Report data, accessed in July 2026. https://ourworldindata.org/grapher/estimated-share-people-generative-ai

Stanford Institute for Human-Centered Artificial Intelligence (Stanford HAI). 2026. *Artificial Intelligence Index Report 2026*. Stanford, CA. https://hai.stanford.edu/ai-index/2026-ai-index-report

Stevens, Ryan. 2026. “Payrolls to Prompts: Firm-Level Evidence on the Substitution of Labor for AI.” arXiv working paper, January 28. https://arxiv.org/abs/2602.00139

Teutloff, Ole, Johanna Einsiedler, Otto Kässi, Fabian Braesemann, Pamela Mishkin, and R. Maria del Rio-Chanona. 2025. “Winners and Losers of Generative AI: Early Evidence of Shifts in Freelancer Demand.” *Journal of Economic Behavior & Organization*. https://doi.org/10.1016/j.jebo.2024.106845

The Straits Times. 2026. “TikTok Parent ByteDance Plans Tiered Subscriptions for AI Chatbot Doubao.” Singapore, May. https://www.straitstimes.com/asia/tiktok-parent-bytedance-plans-tiered-subscriptions-for-ai-chatbot-doubao

Tham, Siew Yean. 2025. “Cross-Border Digitally Delivered Services in Southeast Asia: Growing amid Challenges.” ISEAS Perspective 2025/32, Singapore. https://www.iseas.edu.sg/articles-commentaries/iseas-perspective/2025-32-cross-border-digitally-delivered-services-in-southeast-asia-growing-amid-challenges-by-tham-siew-yean

Udomsaph, Charles. 2026. "AI Adoption in MENAAP: Evidence from Enterprise Surveys." Paper presented at the 5th Annual Central Bank Conference on Development Economics in the Middle East, North Africa, Afghanistan and Pakistan, World Bank, January 20. https://thedocs.worldbank.org/en/doc/c749b541398f6b28ab8a64250b52e5fa-0280032026/original/udomsaph-AIAdoption-20260120.pdf

World Bank. 2025. *Digital Progress and Trends Report 2025: Strengthening AI Foundations*. Washington, DC, November. https://www.worldbank.org/en/publication/dptr2025-ai-foundations

World Bank. 2026. *World Development Report 2026: The Promise of Artificial Intelligence*. Washington, DC, July. https://www.worldbank.org/en/publication/wdr2026

Appendix Table I. Selected AI Policy Initiatives in ASEAN+3 Economies

Economy	National AI Strategy	Note
Brunei Darussalam	The Guide on AI Governance and Ethics for Brunei Darussalam (2024).	
Cambodia	National Artificial Intelligence Strategy 2025–2030 (2025)	In draft as of 2026.
China	National New Generation AI Plan (2017)	
Indonesia	Stranas KA (Strategi Nasional Kecerdasan Artifisial), or Indonesia's National AI Strategy from 2020 to 2045 (2020)	
Japan	7th Basic Plan for the Promotion of Science and Technology (2026) AI Basic Plan (2025) AI Strategy of Japan (2019)	
Korea	National Strategy for AI (2020)	
Lao PDR	-	AI is identified in the National Digital Economy Development Strategy 2021–2030 and its Five-Year Plan 2021–2025 (per MTC reports).
Malaysia	AI Technology Action Plan 2026–2030 (2026) National Artificial Intelligence Roadmap (AI-Rmap) (2021)	
Myanmar	-	-
Philippines	National AI Strategy Roadmap 2.0 (NAISR 2.0) (2024)	
Singapore	National AI Strategy 2.0 (2023) National AI Strategy 1.0 (2019)	
Thailand	Thailand National AI Strategy and Action Plan (2022–2027) (2022)	
Vietnam	National Strategy on R&D and Application of AI (2021)	

Source: AMRO staff compilation from the OECD AI Policy Observatory, Policy Navigator and national dashboards, accessed 30 July 2026; national authorities.

Appendix Table II. Selected ASEAN+3 Public and State-Backed AI Investment Commitments

Economy	Instrument / program	In Local currency	In USD	Period	Note	Source
Brunei Darussalam	-	-	-	-	-	-
Cambodia	-	-	-	-	-	-
China	National Artificial Intelligence Fund	CNY60 bn	USD8.2 bn	From 2025	A national artificial intelligence fund to fast-track strategic investments in AI infrastructure and cutting-edge technologies	China Daily 2025
Indonesia	A Sovereign AI Fund is planned under the 2026–2029 AI roadmap	-	-	-	-	Reuters 2026
Japan	Fiscal support from Ministry of Economy, Trade and Industry	JPY1.5 tn	USD7.9 bn	From Apr 2026	Included both cutting-edge semiconductors and artificial-intelligence development.	Japan Times 2025
Korea	AI allocation in the 2026 national budget	KRW10.1 tn	USD7.3 bn	2026	Tripled the 3.3 trillion won allocated for 2025	Korea Times 2025
Lao PDR	-	-	-	-	-	-
Malaysia	Sovereign AI Cloud and other initiatives such as allocations, tax deductions, and investment.	MYR2 bn	USD488 mn			Malaysia MOF 2026
Myanmar	-	-	-	-	-	-
Philippines	Center for AI Research (CAIR) allocation	PHP350.5 mn	USD5.7 mn	2024		UNESCO Global AI Ethics and Governance Observatory
Singapore	National AI Research and Development (NAIRD) Plan (2025)	SGD1bn	USD786 mn	2025 - 2030		EDB 2026
Thailand	National AI Development Framework	THB25 bn	USD750 mn	2026 - 2027		Bangkok Post 2025
Vietnam	National Artificial Intelligence Development fund	VND30 tn	USD1.1 bn	2026 - 2027	Planned	LuatVietnam 2026

Source: AMRO staff compilation.
Note: Last accessed on August 12, 2026.

Address: 10 Shenton Way, 15-08
MAS Building, Singapore 079117
Website: www.amro-asia.org
Tel: +65 6323 9844
Email: enquiry@amro-asia.org
LinkedIn | **Twitter** | **Facebook**