Deloitte.

Southeast Asia's data centres and AI infrastructure imperative

Capitalising on a once-in-a-generation opportunity

September 2025

MAKING AN IMPACT THAT MATTERS since 1845

Contents

Southeast Asia's data centres and AI infrastructure imperative

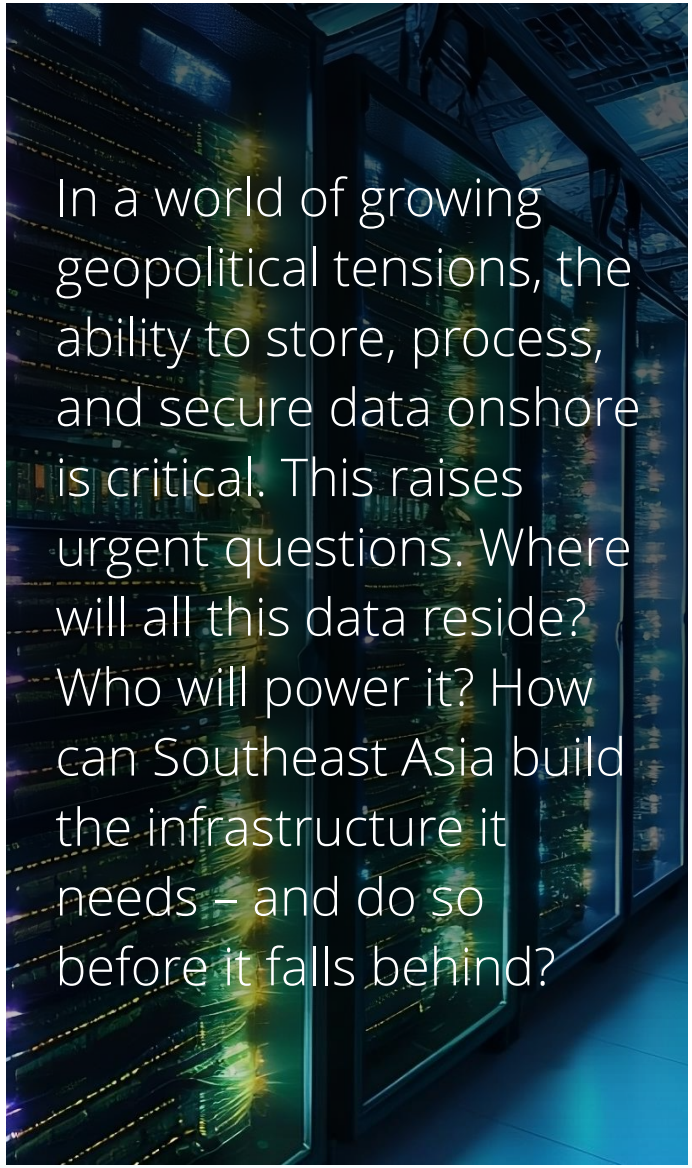
Artificial intelligence (AI) is fuelling an unprecedented surge in data demand – and Southeast Asia is not yet ready to meet this challenge.

Across industries such as manufacturing, mobility, and logistics, next-generation AI applications are starting to replace traditional sensors with high-resolution images, videos, and other data-intensive inputs.

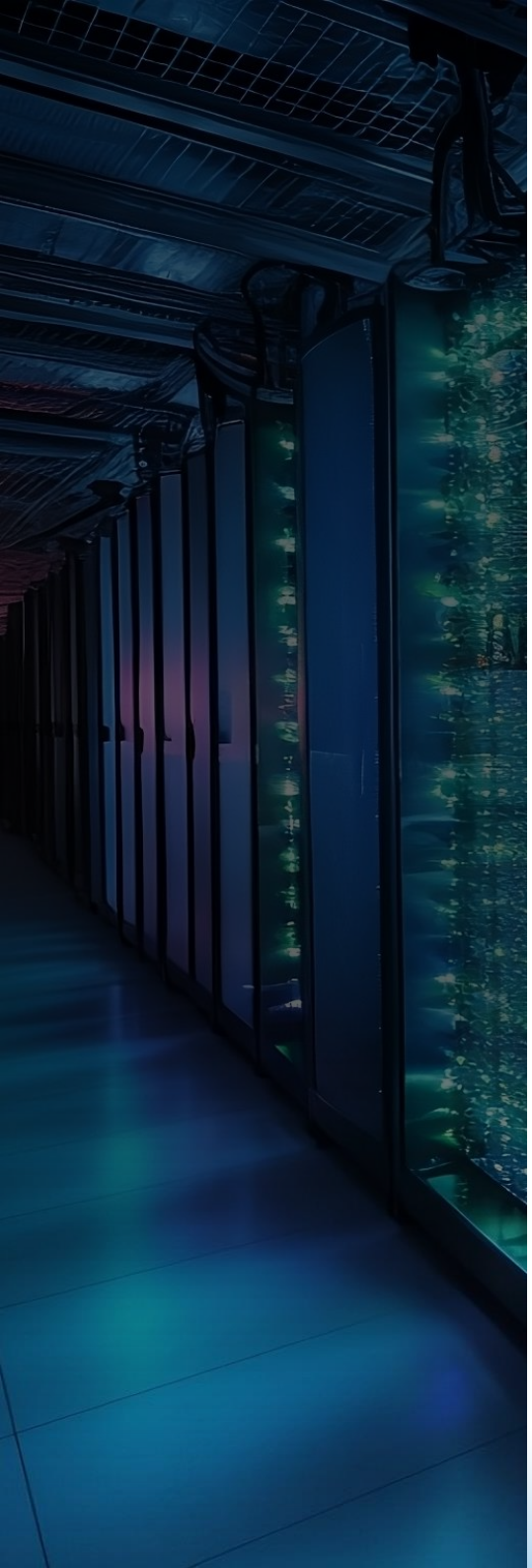
These visual applications require immense computing power and low latency networks to work in real time. As a result, the region is seeing a massive spike in data capacity requirements, far beyond what its existing infrastructure can support.

In a world of growing geopolitical tensions, the ability to store, process, and secure data onshore is also critical. Governments are tightening data regulations, enterprises need faster processing closer to the source, and consumers expect always-on, AI-powered experiences delivered seamlessly and instantaneously.

With data becoming a strategic asset and currency, the global sovereign cloud market, currently valued at approximately US$37 billion, is expected to expand at a compound annual growth rate (CAGR) of 36% to reach US$169 billion in 2028[1]. NVIDIA, one of the world's largest chipmakers, also expects its sovereign AI revenue for fiscal 2026 to be more than double that of the previous year[2].



1 "Forecast analysis: Sovereign cloud IaaS, worldwide". Gartner. 25 October 2024.
2 "NVIDIA Corporation (NVDA) Q2 FY2026 earnings call transcript". Yahoo Finance. 27 August 2025.



This raises several urgent questions. Where will all this data reside? Who will power it? How can Southeast Asia build the infrastructure it needs – and do so before it falls behind?

Meeting this challenge will require action across the ecosystem. National governments must rethink policies and incentivise infrastructure investment. Local telecommunications players must develop go-to-market strategies to realise the market opportunity, while accelerating broadband rollout, 5G deployment, and private network capabilities.

For their part, energy providers must ensure grid stability and develop scalable power sources to meet rising demand. Meanwhile, investors must move quickly to fund next-generation data infrastructure that will form the backbone of Southeast Asia's digital economy.

In this paper, we unpack the emerging AI value chain and its three segments – Application, Platform, and Infrastructure – to reveal the specific areas Southeast Asia can capitalise on in this once-in-a-generation opportunity.

While all three segments matter, the centre of gravity is now rapidly shifting to infrastructure. At the heart of it are data centres – the critical foundation on which our AI readiness depends.

With AI expected to add US$1 trillion to regional GDP by 2030[3] to position Southeast Asia as the world's fourth-largest economy, it suffices to say that the cost of inaction now far outweighs the cost of action.

3 "Is Southeast Asia the next frontier for AI?". Economic Research Institute for ASEAN and East Asia. 29 November 2024.

AI is not a hype – it is at scale

AI is now a multi-trillion-dollar game, but not all AI is equal. Next-generation AI, in particular, goes beyond mere pattern recognition capabilities of traditional AI to transform how images are generated, processed, and stored. This defining feature – the use of images, not just words – as core inputs is transformative.

Across industries such as manufacturing, mobility, and logistics, next-generation AI is replacing physical sensors with high-resolution image recognition. Unlike traditional sensors, image-based AI delivers far greater resolution, speed, and sensitivity – enabling the real-time detection of subtle environmental changes with unprecedented precision.

The exponential growth of the global computer vision in AI market offers a clear sign of what lies ahead. Valued at US$42.44 billion in 2025, the market is on track to hit US$124.08 billion in 2029 – expanding at a CAGR of 30.8%[4].

As demand for image-based AI continues to accelerate, so too will the need to support it. This is driving urgent demand across Southeast Asia not just for advanced algorithms, but also infrastructure – particularly data centres – capable of supporting high-resolution, image-intensive workloads and real-time processing at scale.

AI'S CONTRIBUTION TO THE GLOBAL ECONOMY[5]

US$19.9 trillion
through 2030

3.5%
of global GDP in 2030

AI CAPITAL EXPENDITURE BY FOUR BIG TECH COMPANIES[6]
(Alphabet, Amazon, Meta, and Microsoft)

US$350 billion
in 2025

US$402 billion
in 2026

4 "AI in Computer Vision Global Market Report 2025". The Business Research Company. January 2025.
5 "IDC: Artificial intelligence will contribute $19.9 trillion to the global economy through 2030 and drive 3.5% of global GDP in 2030". IDC. 2024.
6 ""Big Tech's $400 billion AI spending spree just got Wall Street's blessing". The Wall Street Journal. 1 August 2025.

Similar but different: The emerging AI value chain

Along with the rise of next-generation AI applications is a value chain that is emerging to support it (see Figure 1). On the surface, it looks similar to the traditional AI value chain – with the exception of foundational models. But it is this very difference that drives differential value creation across the value chain.

Foundational models, in particular, are central to enabling advanced image recognition capabilities for a variety of use cases in automation, mobility, and manufacturing. Training and leveraging such foundational models for inference, post-training, and test time scaling, in turn, requires massive amounts of data – and with these come highly demanding specifications for the construction of data centres and other AI infrastructure.

Figure 1: The emerging AI value chain

Application	Platform	Infrastructure
Applications create AI-generated content to drive consumer personalisation, boost enterprise productivity, and enhance government services.	Platforms enable the creation and deployment of AI-generated content, driving innovation and automation across applications.	Infrastructure provides the hardware and energy required to power and scale AI platforms and their applications.
Consumer applications	**Foundational models**	**GPU and GPU clouds**
Enterprise applications	**DevOps tools and platforms**	**Data centres**
Government applications	**Storage and IT infrastructure**	**Energy infrastructure**

Stack it up: Value creation in the age of AI

Given the rapid pace of AI's evolution, all players must continuously reassess their value creation models while adapting and responding to disruptions that can occur at any time. The introduction of models like DeepSeek, for example, has shocked the world by opening up the possibility for low-cost and lightweight models[7].

To this end, we believe that Southeast Asian players should consider their value creation activities across all three segments of the AI value chain: Application, Platform, and Infrastructure[8] (see Figure 2). Based on our estimates, the total addressable AI market in Southeast Asia will be worth a whopping US$65 billion in 2035.

Figure 2: Segments and sub-segments of the AI value chain

Application

Consumer applications
Applications focused on the generation of text, code, image, and video, as well as AI-powered search (e.g., *ChatGPT*, *Perplexity*)

Enterprise applications
Applications focused on the summarisation of complex documentation, retrieval augmented generation (RAG) search, and natural language processing (NLP) chat agents (e.g., *Cohere*, *Harvey*)

Government applications
Applications focused on waste, parking, security management, smart cities, urban mobility, and defence (e.g., *Hayden AI*)

US$30 billion

Platform

Foundational models
Language models, diffusion models, and world models to power applications and autonomous vehicles/robots (e.g., *OpenAI*, *DeepSeek*, *Garuda LLM*)

DevOps tools and platforms
IT infrastructure and software tools for model development, orchestration, and observability (e.g., *PineCone*, *LangChain)*

Storage and IT infrastructure
Storage infrastructure dedicated to high capacity, low latency high-performance computing, as well as AI workloads and data lakes (e.g., *DDN*, *Pure Storage)*

US$10 billion

Infrastructure

GPU and GPU clouds
Compute hosted on-premise and on clouds dedicated to AI applications and models (e.g., *CoreWeave*, *Lambda*, *YTL*, *Singtel*, *IOH*)

Data centres
Data centres customised for high rack density AI GPU workloads (e.g., *QTS*, *Crusoe*, *Gulf*, *AIS*, *PLDT*, *Telkom)*

Energy infrastructure
Energy solutions with colocation and reduced transmission losses for AI loads (e.g., *Crusoe*, nuclear small modular reactors (SMRs))

US$25 billion

US$65 billion

Total addressable AI market in Southeast Asia in 2035[9]

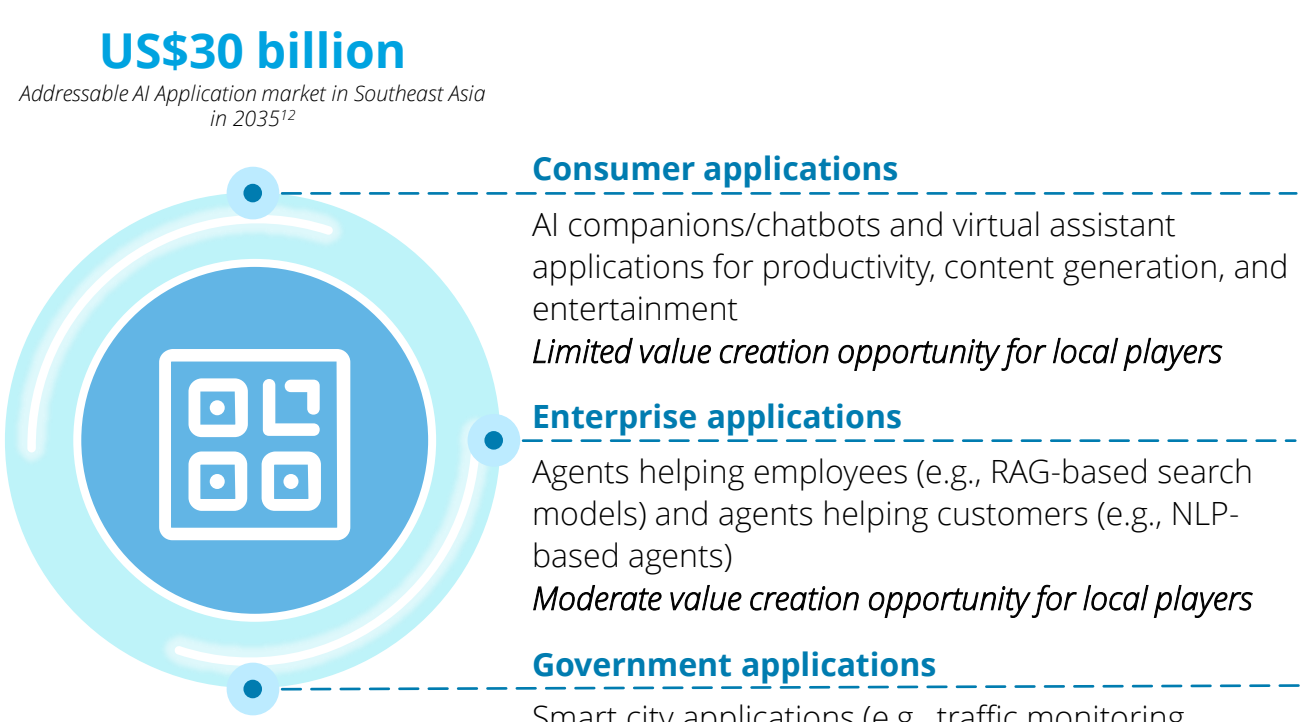
7 "DeepSeek R1's implications". IoT Analytics. 5 February 2025.
8 Capitel analysis.
9 Deloitte analysis.

Application

Consumer AI applications have experienced exponential growth in recent years. Of note are applications like Perplexity, which has amassed 10 million monthly active users (MAU)[10], with one in four based in Indonesia[11]. Such patterns reveal significant value leakage and the need to better retain consumer value within the region.

However, as Consumer applications have high development costs and are dominated by global players, most of the value for Southeast Asian players will likely lie in developing Enterprise and Government applications for specific local use cases (see Figure 3). These applications could also include distinct service level differentiation across the dimensions of security, reliability, and latency.

Figure 3: Value creation in the Application segment



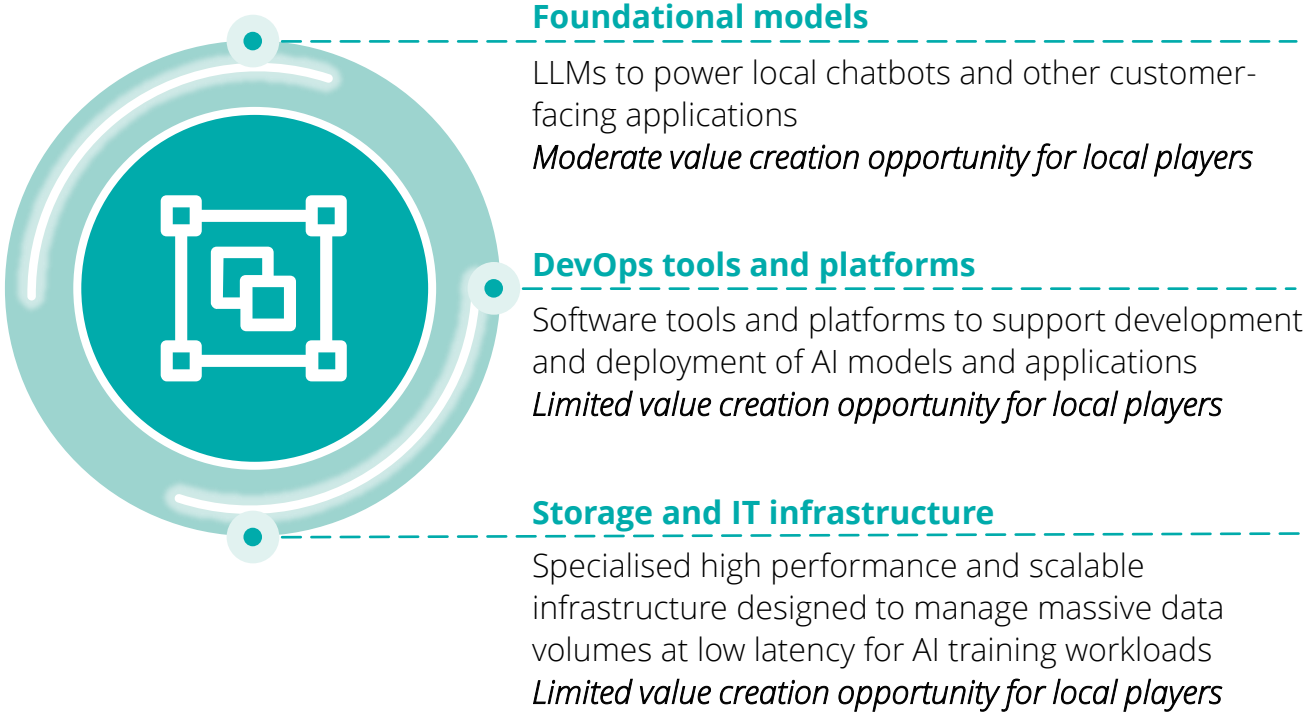
10 "Report: Perplexity business breakdown and founding story". Contrary Research. 2024.
11 "The latest perplexity AI stats (2024)". Exploding Topics. Accessed on 25 February 2025.
12 Deloitte analysis.

Platform

Next-generation AI applications run on foundational models, which are used to generate tokens and deliver predictions. For Southeast Asia, local opportunities in the Platform segment are mainly in the running of inference models, rather than training of models (see Figure 4).

There is clear value capture opportunity in leveraging available open-source models, fine-tuning them, and developing national large language models (LLMs) and their associated applications. Singapore, for example, recently launched Southeast Asia's first LLM ecosystem initiative, known as the National Multimodal LLM Programme (NMLP)[13].

Figure 4: Value creation in the Platform segment



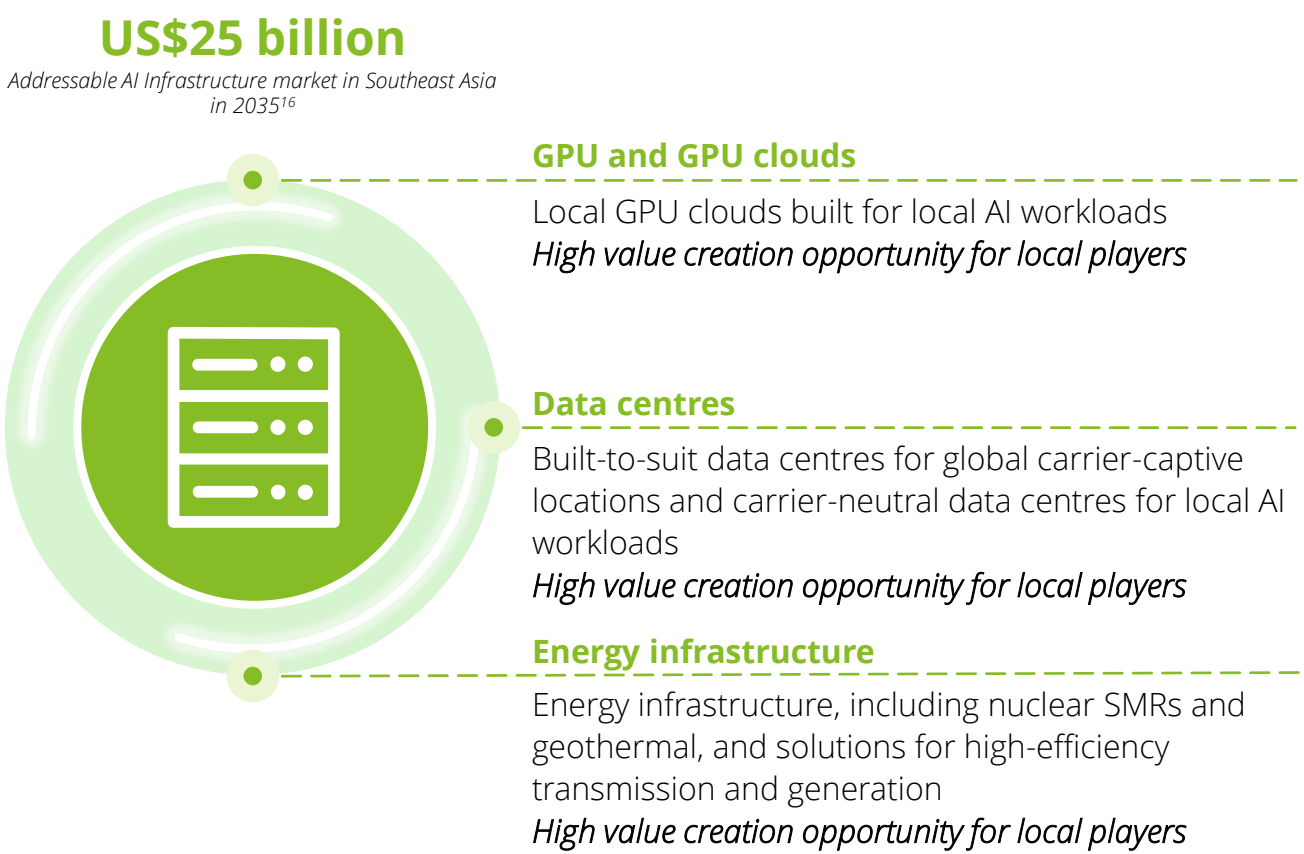
13 "Singapore pioneers S$70m flagship AI initiative to develop Southeast Asia's first large language model ecosystem catering to the region's diverse culture and languages". Infocomm Media Development Authority. 2023.
14 Deloitte analysis.

Infrastructure

Next-generation AI applications require high-performance graphics processing unit (GPU) computing systems for large-scale, intensive workloads[15]. This means that many of Southeast Asia's existing data centres cannot be used for their training and inference. Given that data centres are highly resource-intensive in nature, they will also require a massive water supply and reliable electricity supply 24/7.

In this discussion, however, we will focus on energy infrastructure and the uphill task of keeping up with its demand. Accordingly, three Infrastructure sub-segments – GPU and GPU clouds, data centres, and energy infrastructure – will be critical future 'toll roads' driving investment and value in Southeast Asia (see Figure 5).

Figure 5: Value creation in the Infrastructure segment



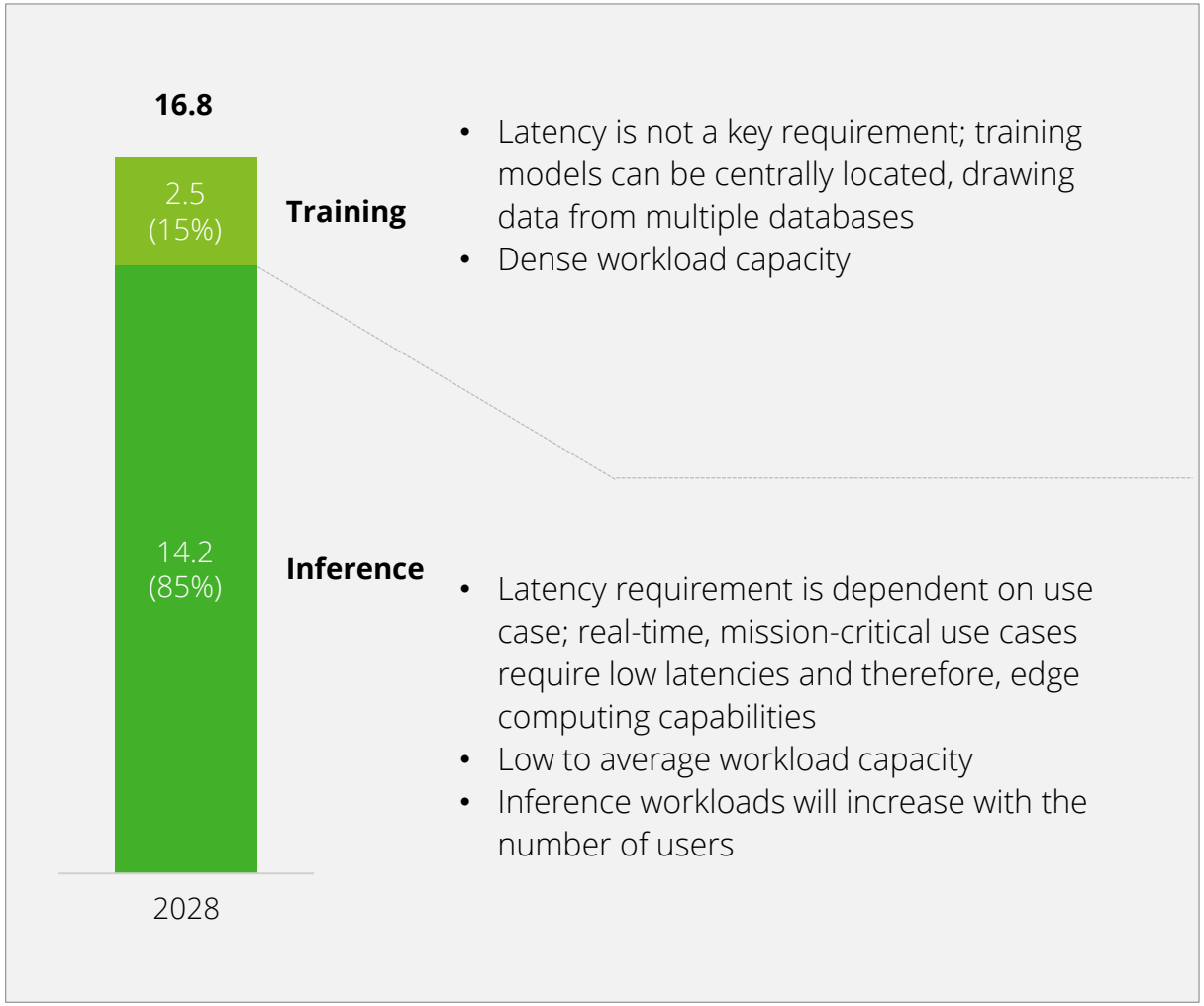
15 "What generative AI means for data centres". Equinix. 2023.
16 Deloitte analysis.

GPU and GPU clouds

Globally, 85% of global GPU demand is expected to be from AI inference[17] (see Figure 6). Due to low latency requirements for some use cases, these may need to be served locally. Export controls have resulted in GPU clouds experiencing high utilisation rates – and with these, high returns.

Within Southeast Asia, we have witnessed significant recent developments, such as the building of local GPU clouds to support local AI workloads by data centre operators in Indonesia[18] and Malaysia[19], in a move from colocation-based business models to higher margin, services-based GPU cloud models, and the launch of GPU-as-a-service (GPUaaS) offerings in Singapore and Southeast Asia[20].

Figure 6: Global AI workload breakdown, GW (2028)



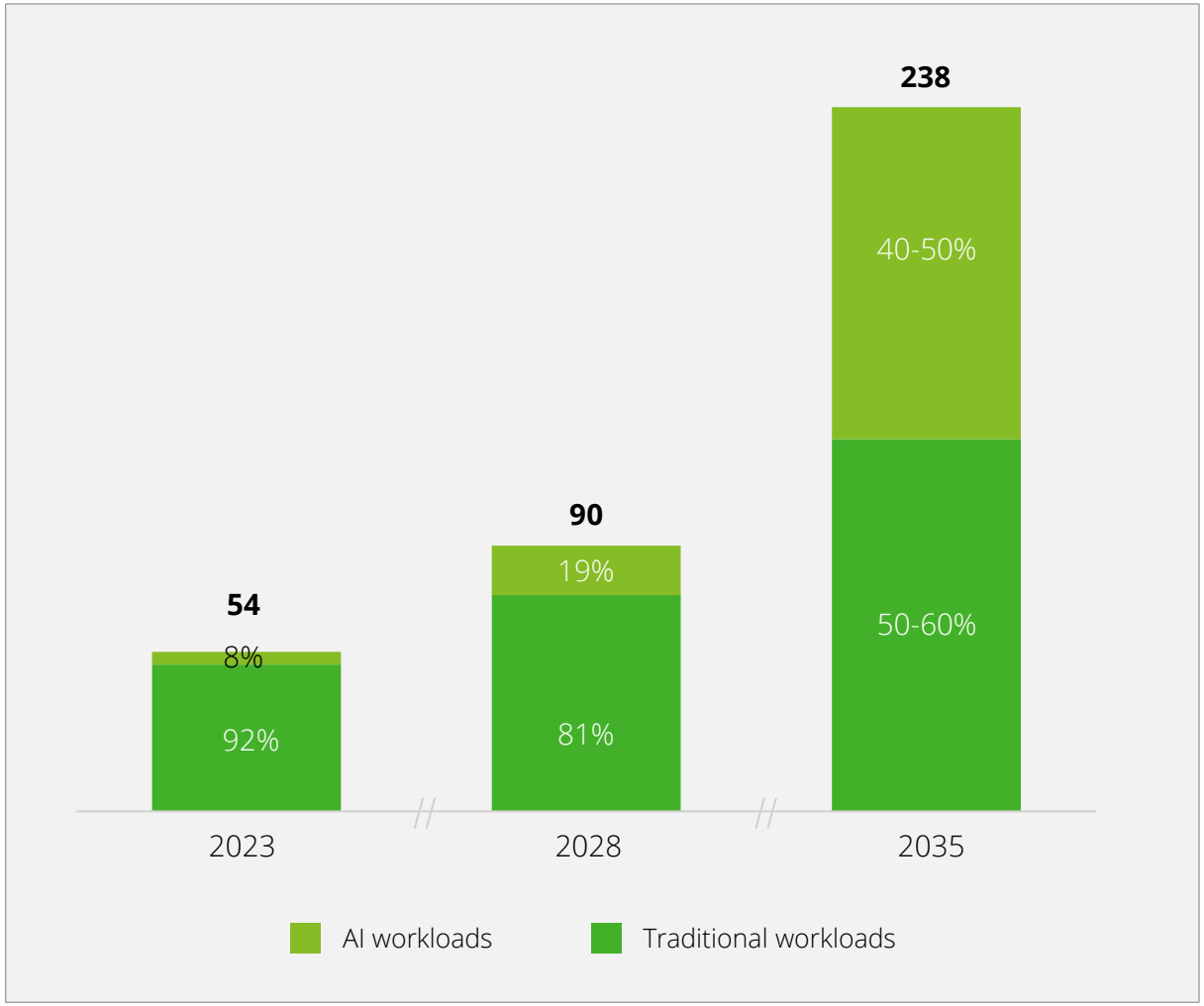
17 "The AI disruption, challenges and guidance for data centre design". Schneider Electric. September 2023.
18 "Indosat partners with Nvidia for $200m AI centre in Indonesia". Data Centre Dynamics. 5 April 2024.
19 "Nvidia & YTL Power partner for $4.3bn AI data centres in Malaysia". Data Centre Dynamics. 11 December 2023.
20 "Singtel to introduce GPU-as-a-Service powered by NVIDIA accelerated computing". Singtel. 19 March 2024.

Data centres

By 2035, 40-50% of total IT workload demand is expected to be AI-driven[21] (see Figure 7). This underscores the urgency for Southeast Asian players to build AI-ready data centres, not only to retain value within local markets, but also ensure that data and infrastructure remain on shore in a world of growing geopolitical tensions where data is the new currency.

To serve local inference workloads within Southeast Asia, data centres will need to be purpose-built to handle high density workloads. There will also be clear build-to-suit opportunities for captive AI data centres located locally within markets or within regional hubs to serve this demand.

Figure 7: Global IT workload breakdown, GW (2023-2035)

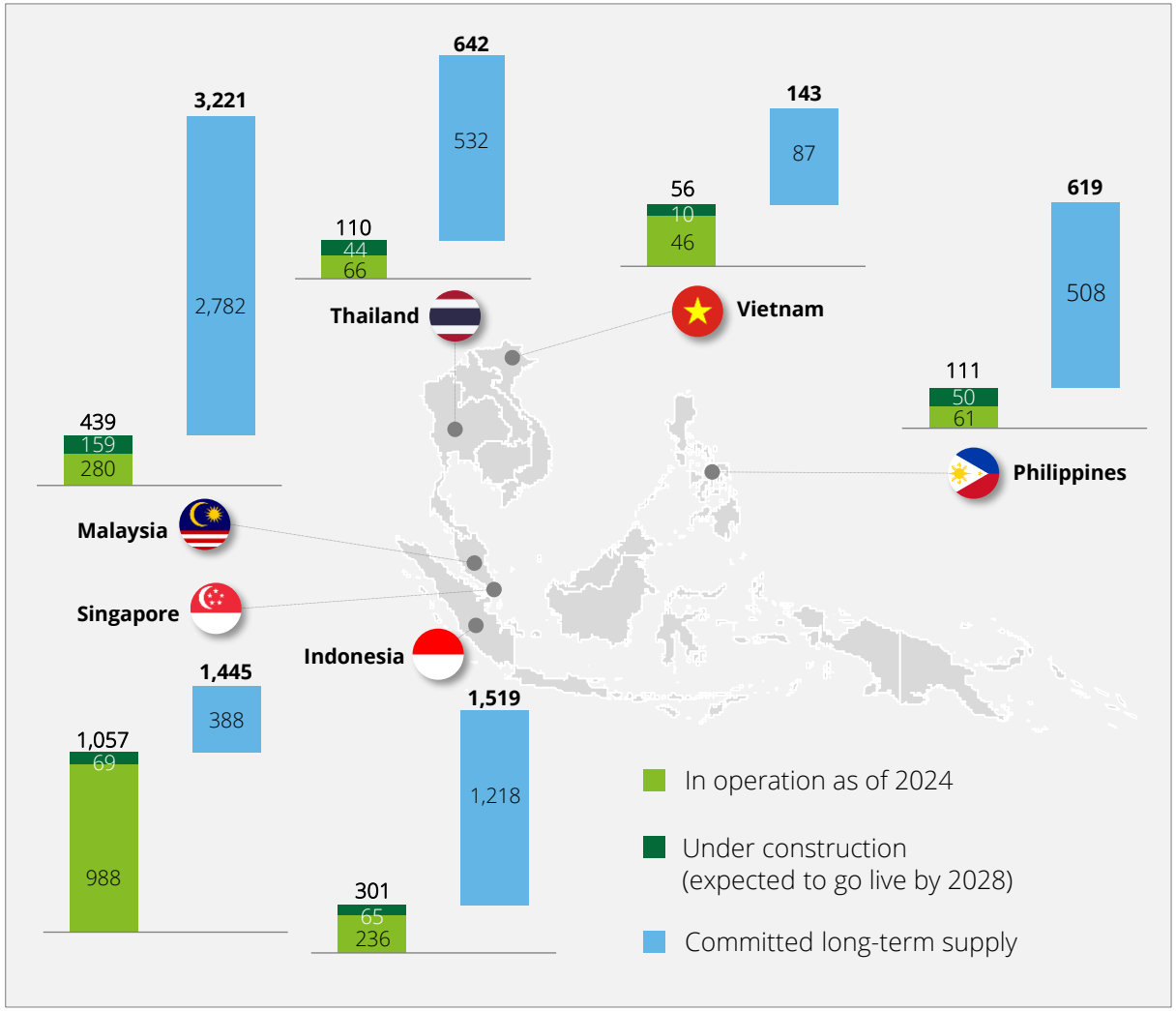


21 "The AI disruption, challenges and guidance for data centre design". Schneider Electric. September 2023.

Across Southeast Asia, approximately 397 MW of data centre capacity is expected to go live by 2028. Total capacity will more than quadruple from 1.68 GW in 2024 to 7.59 GW in the long term (see Figure 8)[22].

Notably, several Southeast Asian hubs are gaining prominence. Malaysia, for example, is exhibiting a robust future supply pipeline with its state of Johor becoming one of the fastest-growing markets. While Singapore currently possesses the largest data centre capacity, its pole position is likely to be supplanted by Malaysia and Indonesia, who are at least in part beneficiaries of the limited land capacity for supply expansion in neighbouring Singapore.

Figure 8: Data centre capacity across Southeast Asia, MW



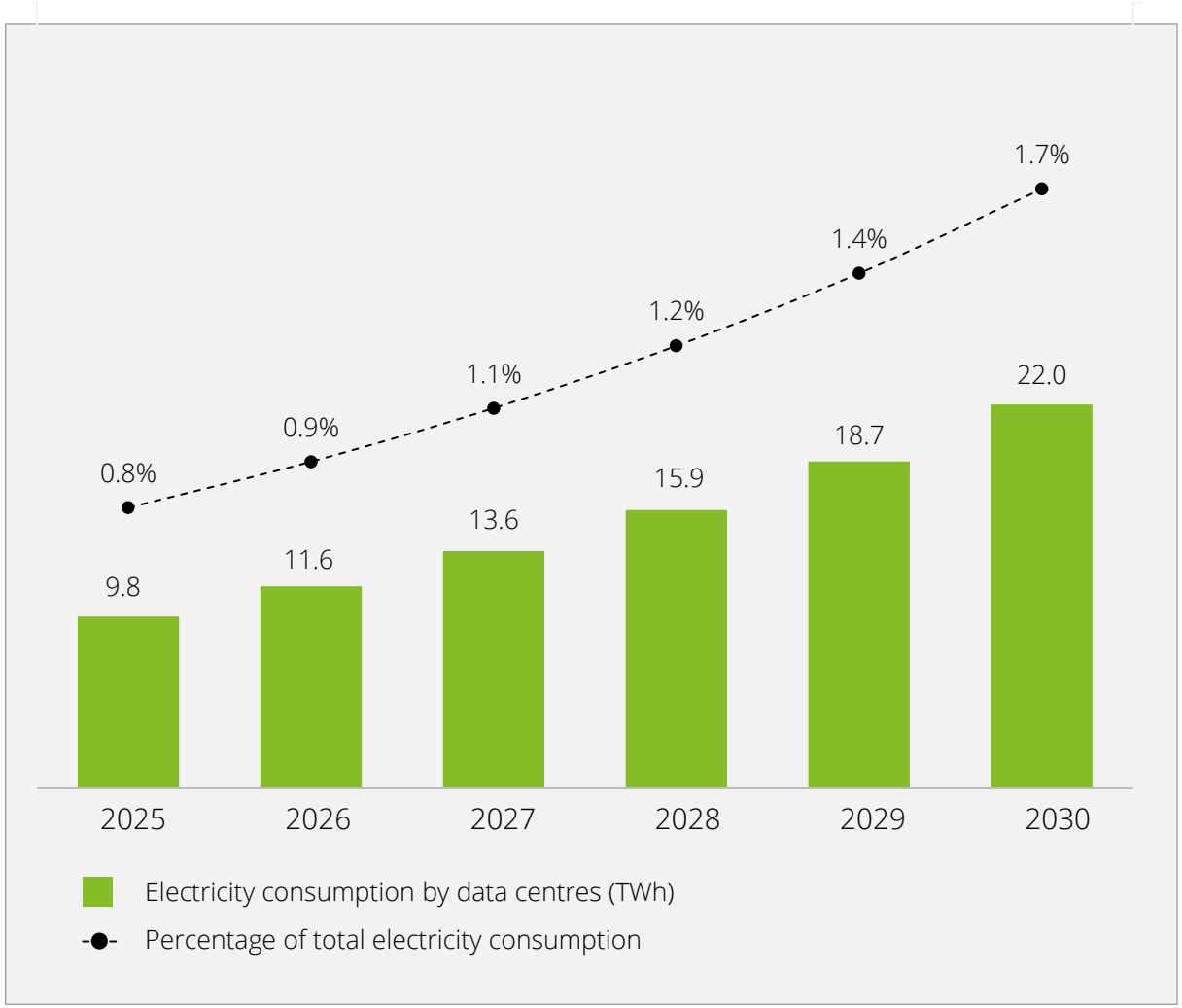
22 Deloitte analysis.

Energy infrastructure

On the back of power-intensive AI requirements, electricity consumption by data centres in Southeast Asia is expected to more than double between 2025 and 2030[23] (see Figure 9).

Notably, AI data centres leveraging GPUs require markedly more power and produce more heat per square metre than traditional data centres leveraging central processing units (CPUs). A CPU runs at only 150-200 W per chip; in contrast, AI GPUs run at more than double at 500 W, while generative AI (GenAI) GPUs and next-generation chips run at 700 W and 1,200 W respectively.

Figure 9: Electricity consumption by data centres in Southeast Asia (2025-2030)

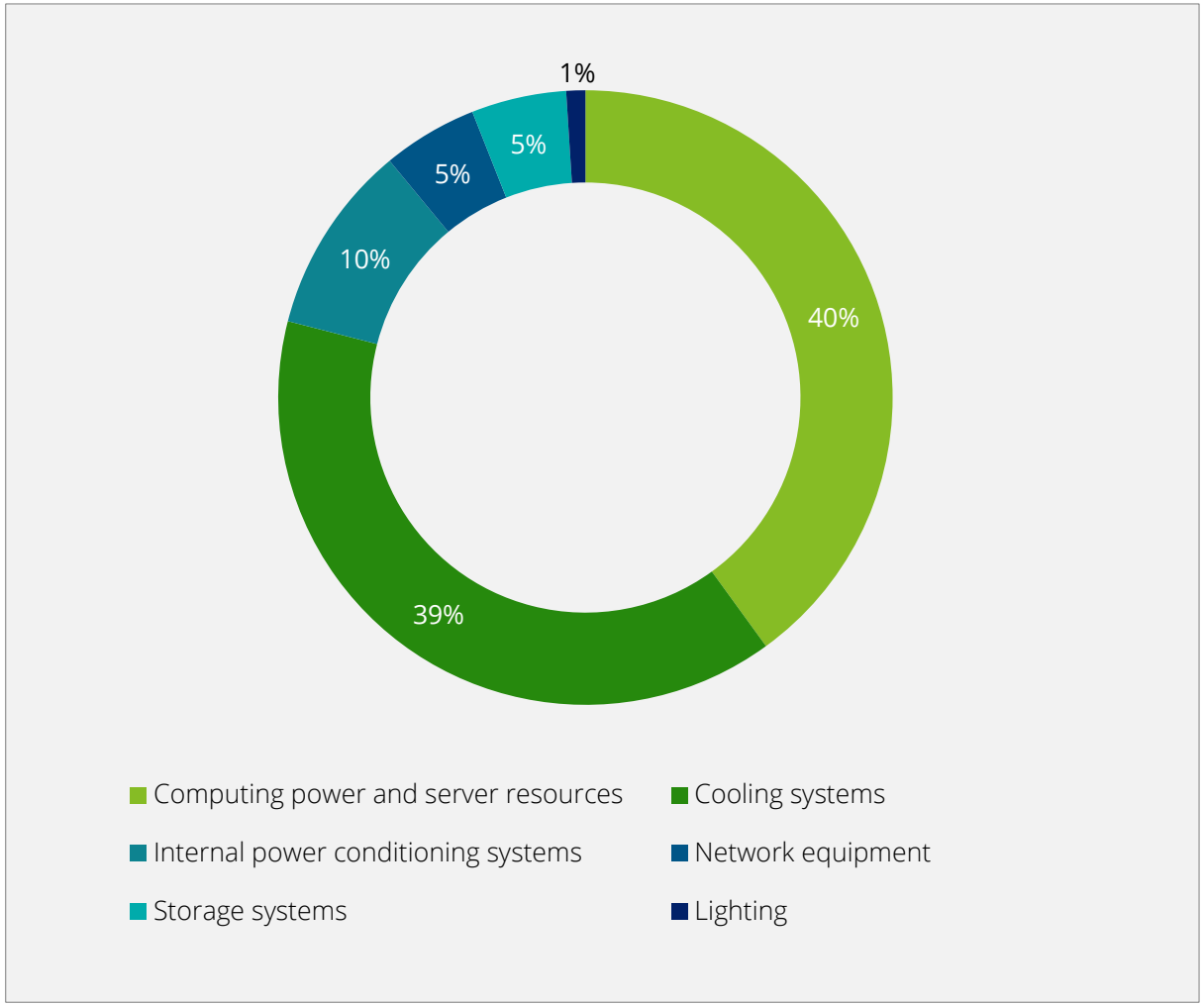


23 Deloitte analysis.

High-performance AI computing environments also require high-density infrastructure to support computing power. As of early 2024, data centres typically support rack power requirements of 20 kW or higher. Current state of the art AI server racks can consume up to 120 kW per rack, and megawatt-scale racks are expected by 2030.

Broadly, electricity consumption in data centres is primarily driven by two components – computing power and server systems, as well as cooling systems – that together account for nearly 80% of total needs[24] (see Figure 10). These two are the most energy-intensive components, and will continue to fuel growth in power consumption for the foreseeable future.

Figure 10: Breakdown of electricity consumption in data centres



24 "As generative AI asks for more power, data centres seek more reliable, cleaner energy solutions". Deloitte Insights.

A supply and demand view

Meeting the energy requirements of Southeast Asia's data centres will require a two-pronged approach

Supply

Reliable and uninterrupted supply

Significant improvements in grid capacity, transmission, and distribution will be required to support data centre operations. Uninterruptible power supply (UPS) systems can also play a role in minimising outages, service disruptions, data loss, and/or hardware damage.

Carbon-reduced energy supply

While existing energy infrastructure should be leveraged to enhance data centre competitiveness, 'clean coal' and carbon capture technologies can be deployed in tandem to reduce environmental impacts.

Affordable procurement of green energy

Solar and wind sources could meet up to 30% of Southeast Asia's data centre energy requirements in 2030[25]. However, the development of supporting policies, market reforms, and innovative procurement mechanisms will be key to ensuring the cost-effectiveness and affordability of these clean energy sources.

Demand

Energy efficiency

By providing real-time visibility into power consumption, energy-efficient technologies, such as advanced metering infrastructure, can help to improve demand forecasting and outage response, as well as promote energy efficiency in data centres.

Energy demand response

Predictive analytics tools powered by AI and machine learning (ML) can be leveraged for load requirement forecasts and power distribution optimisation[26]. While these applications are becoming mainstream, effectively leveraging them will require collaboration between data centre players and grid managers. By dynamically scheduling compute-intensive tasks to coincide with periods of high renewable energy availability (e.g., peak solar or wind generation), data centres can reduce their reliance on fossil fuels and better balance grid loads.

On-site and distributed renewable energy integration

Data centres can invest in on-site solar generation or enter into power purchase agreements (PPAs) for renewable energy to reduce grid demand. Smaller operators can also leverage virtual PPAs or green tariffs to gain access to clean energy without high upfront investments. However, there is a need to alleviate restrictions on such agreements as some markets may lack regulatory and commercial certainty on these strategies.

25 "Solar and wind could power up to a third of ASEAN's data centres in 2030, without needing batteries". Ember. 27 May 2025.
26 "The Power Play: How data centres and utilities are reinventing energy strategies". Data Center Frontier. 28 March 2025.

All systems go: Move now, and move quickly

National governments, local telecommunications and energy players, and investors must recognise data centres and other AI infrastructure as critical assets of tomorrow – and move now, and move quickly, to build these assets on their shores.

National governments

National governments have arguably the most pivotal and important role to play in enabling the development of these assets and spurring growth in local and regional ecosystems by:

Regulating and incentivising investments by ensuring clarity in policies, standards, data security classification, and data localisation regulations, as well as providing infrastructure grants and incentives

Attracting global players including hyperscalers and cloud providers, by streamlining approvals, achieving policy stability, and ensuring infrastructure readiness

Developing local and regional ecosystems with investors, local talent pool, and infrastructure players, including sovereign data ecosystems based on population size; economies with smaller populations should consider joining other ecosystems with similar profiles

Overseeing data governance and managing risks such as technology availability (particularly for GPUs in an export-controlled system), commercial (through demand forecasting), consumer safety, data privacy, and cyber (e.g., Singapore's Digital Infrastructure Act[27]) risks

Developing competitive and sustainable energy policies for data centres by setting minimum efficiency standards, formalising data centre energy demands in power plans, facilitating cross-border PPAs, and fostering public-private partnerships (PPPs)

27 "IMDA launches guidelines for cloud services and data centres ahead of Digital Infrastructure Act". The Straits Times. 25 February 2025.

Local telecommunications and energy players

Local telecommunications and energy players should take concerted steps to consider how they can best support the build-out of data centres and other AI infrastructure on their shores by:

Understanding where to play
including size of opportunity, segment of value chain to target, technology choices (e.g., 5G mobile edge computing, fibre, or satellite), and amount of investment to commit

Developing go-to-market strategies
to realise the market opportunity (e.g., through partnerships/alliances to monetise investments in infrastructure)

Considering funding or co-funding models
including whether to go it alone or find a co-investor, and structuring the asset to unlock value

Managing commercial risks
by developing accurate demand forecasts

Enhancing access to diversified energy sources
to balance accessibility, affordability, and sustainability, and supplement conventional sources (e.g., coal, solar) with new energy technologies (e.g., clean coal, SMRs, storage); construction costs could be reduced by reusing existing infrastructure, localising engineering, procurement, construction, and commissioning (EPCC), and leveraging innovative financing mechanisms

Investors

The capital needs to build data centres and other AI infrastructure are staggering. Building a 100-MW AI-ready data centre costs at least US$1 billion; this is not to mention the more than US$2 billion worth of GPUs and associated supercomputers and connectivity that it will be housing.

The upside for investors is that data centres offer stable cashflows. One MW of AI-ready power can generate US$1.5 to 2 million per year and likely comes with a 10-year commitment or more, while one MW of GPU power can generate about US$15 million in annual revenue under a GPU cloud model. As they consider how best to deploy their capital, investors and private equity players should think about:

Understanding where to play including size of investment opportunity and potential targets

Weighing their partnership or consortia options (e.g., BlackRock, Global Infrastructure Partners, Microsoft, and MGX recently launched a new partnership to invest in data centres and power infrastructure[28])

Assessing new methods to reduce financing costs (e.g., real estate investment trust (REIT) management and partial investment)

Managing commercial risks by considering a modular or phased build approach to reduce upfront costs and time-to-revenue

Investing in a broad range of energy solutions from mature (e.g., solar, wind) to new technologies (e.g., battery storage, microgrids, biomass, nuclear SMRs) and energy efficiency projects, including through participation in PPPs

28 "BlackRock, Global Infrastructure Partners, Microsoft and MGX launch new AI partnership to invest in data centres and supporting power infrastructure". Microsoft. 17 September 2024.

The bottomline

The future of AI is visual, data-intensive, and infrastructure-driven. With this shift underway, the risk of underinvesting in data centres and AI infrastructure is dramatically greater than the risk of overinvesting in them – and continues to grow by the day.

1

AI is a once-in-a-generation opportunity for Southeast Asia

2

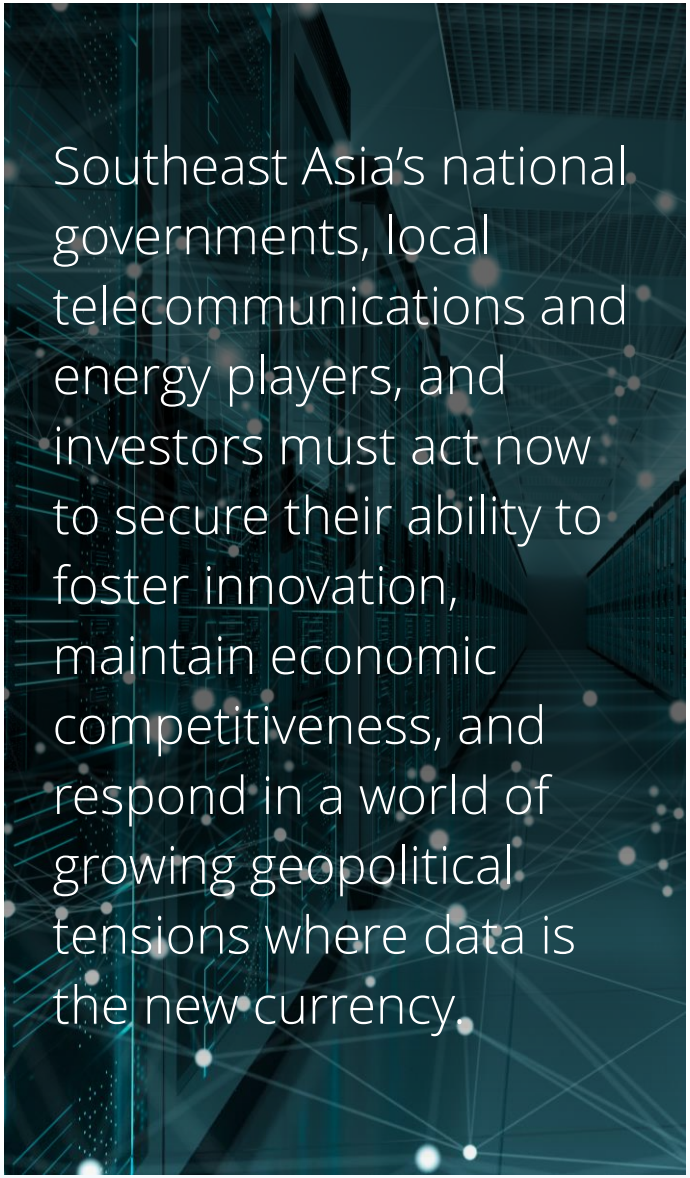
Data centres and AI infrastructure are linchpins to AI participation

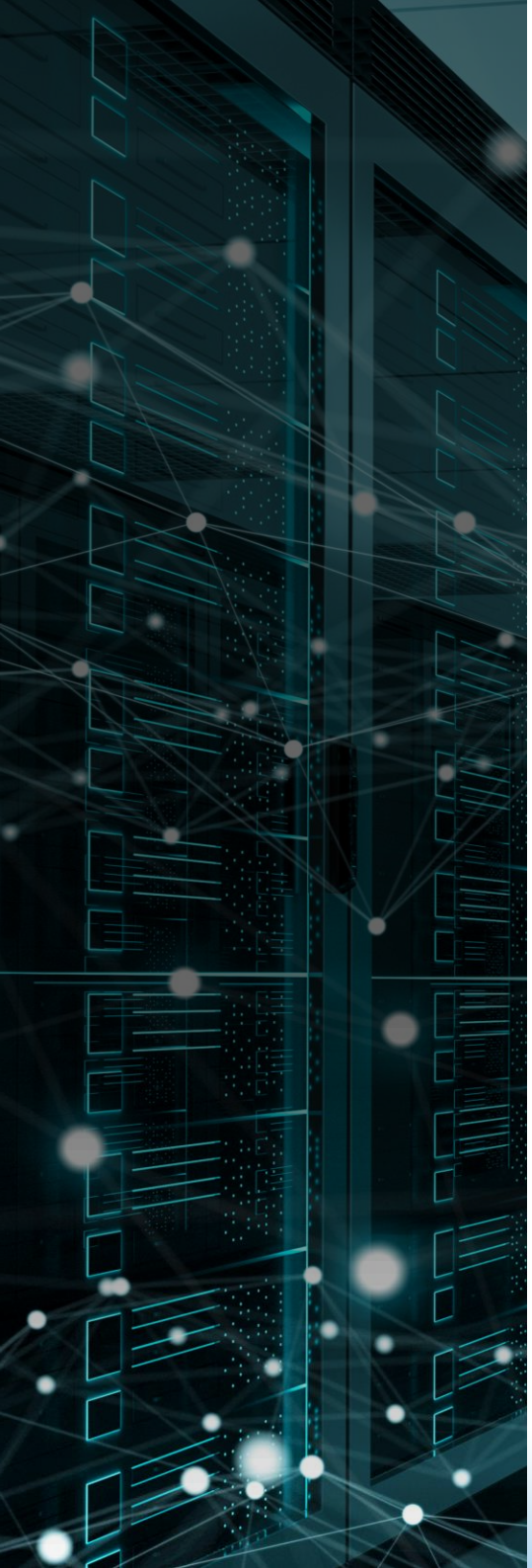
3

Value preservation, sovereignty, and security considerations underscore the need for onshore infrastructure

4

The cost of inaction far outweighs the cost of action





Researched and written by

Yang Chi Chih
Technology, Media & Telecommunications Industry Leader
Deloitte Southeast Asia

Piyush Jain
Strategy, Risk & Transactions Leader for Technology, Media & Telecommunications
Deloitte Southeast Asia

Haridas Kanagasabai
Audit & Assurance Leader for Technology, Media & Telecommunications
Deloitte Southeast Asia

Other contributors

Raj Kannan
Power, Utilities & Renewables Sector Leader
Deloitte Southeast Asia

Keith Davies
Technology & Transformation Leader for Energy, Resources & Industrials
Deloitte Southeast Asia

Samrat Bose
Strategy, Risk & Transactions Partner
Deloitte Singapore

Farhan Rashid
Strategy, Risk & Transactions Director
Deloitte Singapore

Chong Chor Ming
Strategy, Risk & Transactions Senior Manager
Deloitte Singapore

Contact us

Southeast Asia Technology, Media & Telecommunications industry practice

Southeast Asia Technology, Media & Telecommunications Leader
Yang Chi Chih
chiyang@deloitte.com

Audit & Assurance
Haridas Kanagasabai
kaharidas@deloitte.com

Strategy, Risk & Transactions
Piyush Jain
pijain@deloitte.com

Technology & Transformation
Nicholas Chan
nickchan@deloitte.com

Tax & Legal
Ng Lan Kheng
lkng@deloitte.com

Deloitte.

Designed by CoRe Creative Services. RITM1993654