Deloitte.

Southeast Asia's data centres and AI infrastructure imperative

Capitalising on a once-in-a-generation opportunity

March 2025

MAKING AN IMPACT THAT MATTERS since 1845

Contents

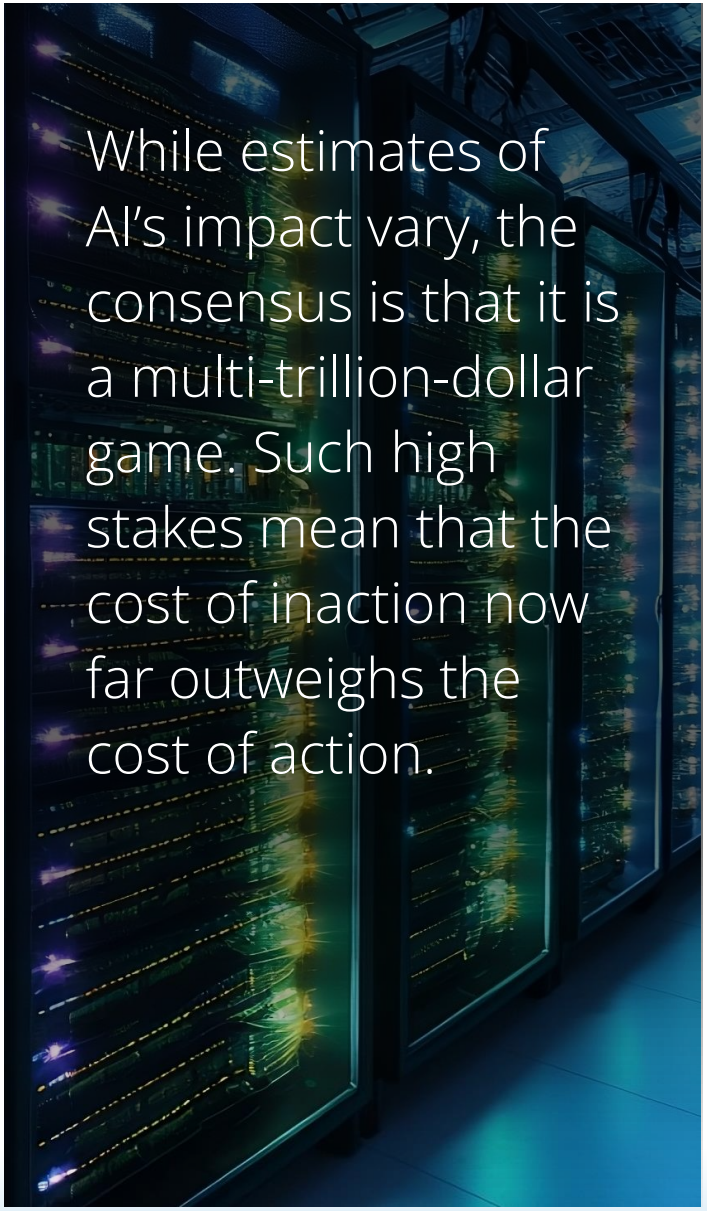
Preface

Artificial intelligence (AI) is amongst the most defining megatrends of this century, permeating our daily lives and driving substantial growth across economies.

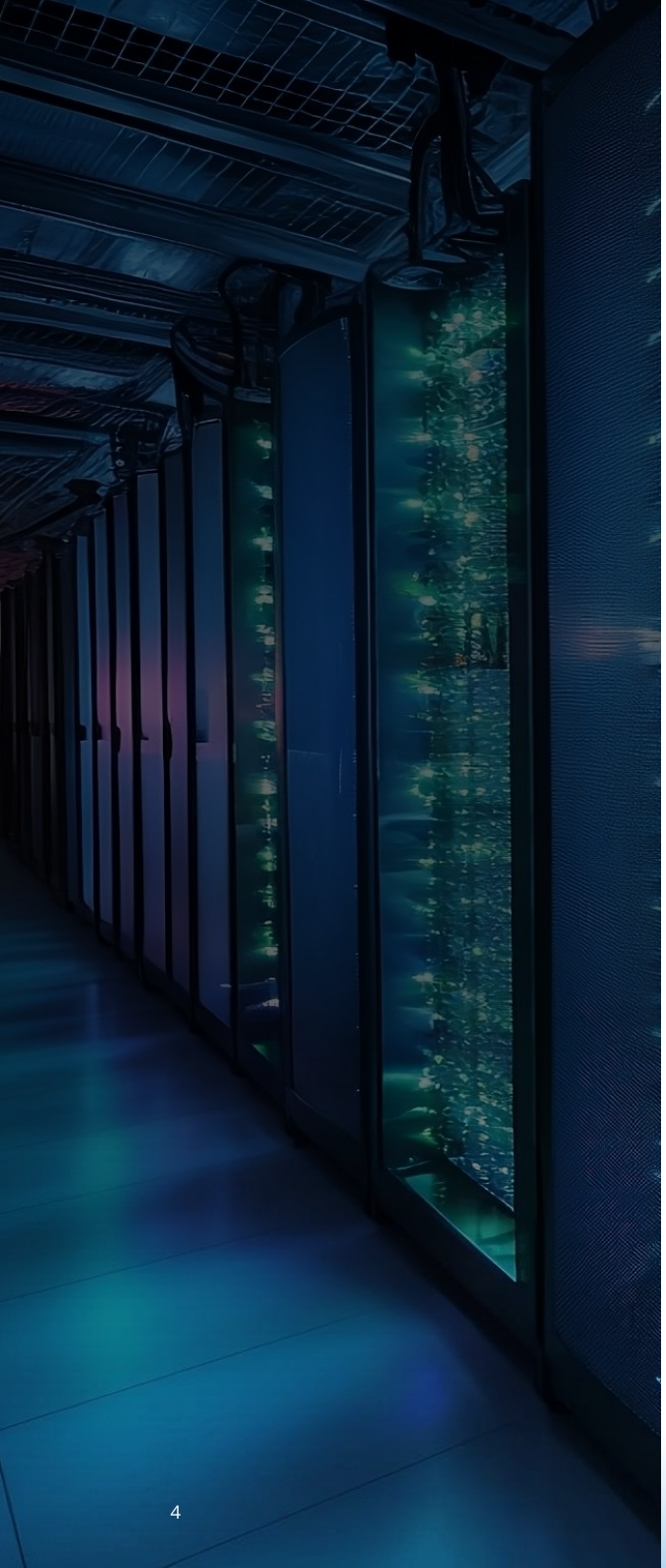
While estimates of AI's impact vary, the consensus is that it is now a multi-trillion-dollar game. On a global level, major players such as the US and China are accelerating investments to position themselves as leaders in the game.

In January 2025, the US issued its export control framework, known as the interim final rule (IFR) on AI diffusion[1], and its Executive Order 14179[2] in a clear sign that it intends to remove barriers to US leadership in AI for national security and economic reasons. Within the same month, China also announced the launch of its RMB60 billion (US$8.2 billion) national AI industry investment fund[3].

The bottomline is that AI is real, and it is at scale. Closer to home in Southeast Asia, AI is expected to add US$1 trillion to its regional gross domestic product (GDP) by 2030[4] to position it as the world's fourth-largest economy. Such high stakes mean that the cost of inaction now far outweighs the cost of action.



1 "Framework for artificial intelligence diffusion". US Bureau of Industry and Security, Department of Commerce. 15 January 2025.
2 "Executive Order 14179 – Removing barriers to American leadership in artificial intelligence". The American Presidency Project. 23 January 2025.
3 "In China, the domestic AI race intensifies as Chinese go gaga over DeepSeek". The Straits Times. 9 February 2025.
4 "Is Southeast Asia the next frontier for AI?". Economic Research Institute for ASEAN and East Asia. 29 November 2024.



This is especially in light of the advent of next-generation generative AI (GenAI) applications, many of which are rapidly gaining scale even as we speak. GenAI applications with image recognition capabilities, for example, are expected to replace sensors in many use cases, with significant and widespread repercussions for automation, mobility, and manufacturing, amongst others.

Underpinning these GenAI applications are foundational models, which require massive amounts of data for training and inference. These, in turn, must be supported by highly specific and demanding specifications for data centres and other AI infrastructure – most of which cannot currently be met in Southeast Asia today.

To secure their nation's ability to foster innovation, maintain economic competitiveness, and safeguard national security, national governments must therefore prioritise as a matter of urgency policies and investments to incentivise national players, global technology companies, and investors to collaborate with them on building their nation's data centres and other AI infrastructure on their very own shores.

GenAI is not a hype – it is at scale

As the global AI arms race heats up, the sheer size of investments being poured into it is staggering. In 2025 alone, the four largest global technology companies – Alphabet, Amazon, Meta, and Microsoft – are expected to allocate US$320 billion to AI-related capital expenditures.

But not all AI is equal. Of note is GenAI, which goes beyond mere pattern recognition capabilities of traditional AI to transform how images are generated, processed, and stored. In recent years, we have witnessed consumer GenAI applications explode onto the scene with skyrocketing user engagement rates.

The result has been a growing ubiquity of such applications in our daily lives. In 2024, consumers across the globe downloaded AI applications 17 billion times and spent a mind-boggling 7.7 billion hours using these applications. Suffice it to say that GenAI is not a hype; on the contrary, it is at scale.

To capture the value of this opportunity and remain competitive on the global arena, Southeast Asian players must move now – and move quickly.

AI'S CONTRIBUTION TO GLOBAL ECONOMY[5]

US$19 trillion through 2030

3.5% of global GDP in 2030

SPENDING ON AI BY BIG TECH

Alphabet[6]
US$75 billion

Amazon[7]
US$100 billion

Meta[8]
US$65 billion

Microsoft[9]
US$80 billion

US$320 billion

EXPLOSION OF CONSUMER GENAI APPS[10,11,12]

ChatGPT
350 million MAU
4.7 billion visits (Jan 2025)

Character.ai
29 million MAU
98 minutes/day per user

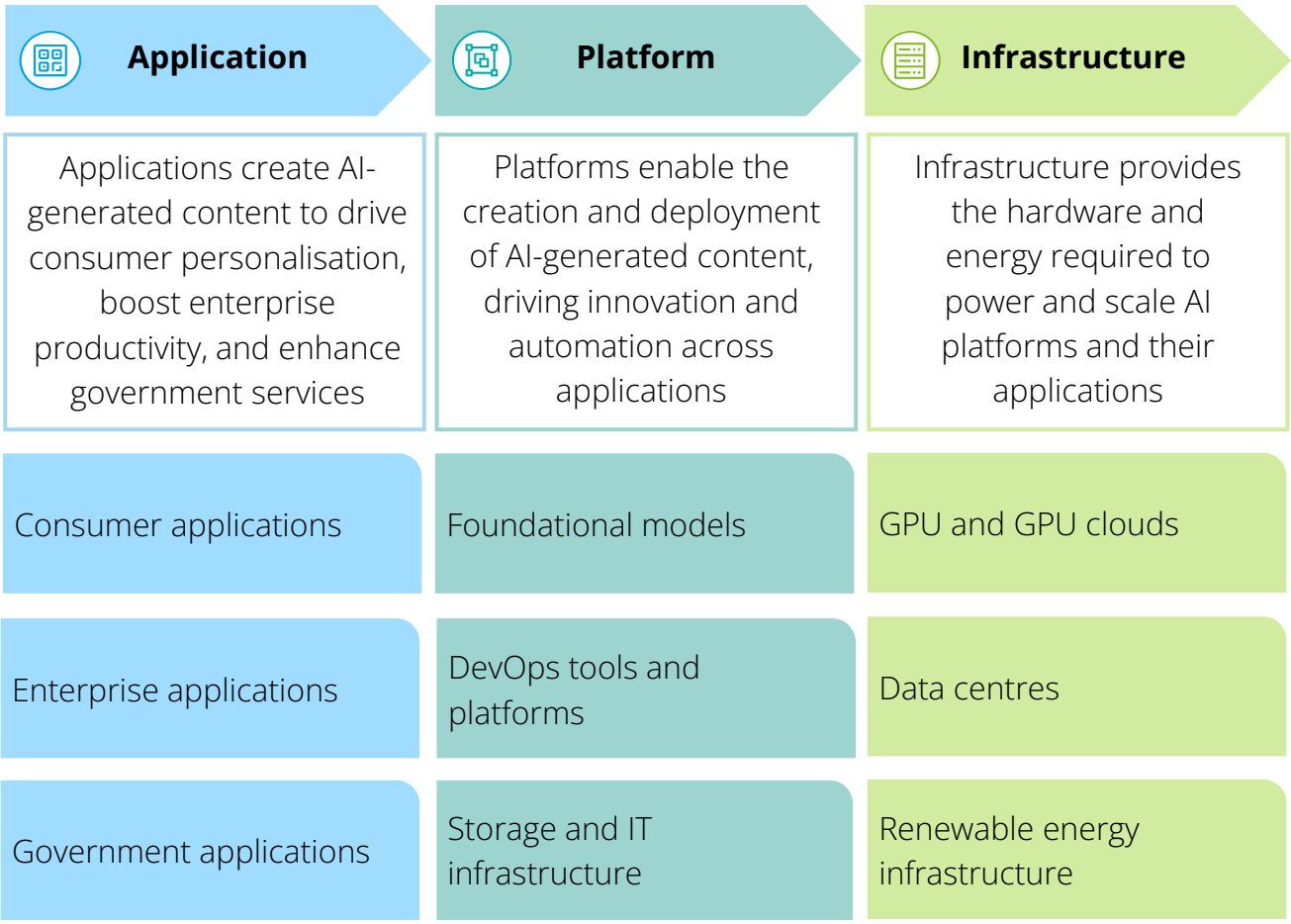
5 "IDC: Artificial intelligence will contribute $19.9 trillion to the global economy through 2030 and drive 3.5% of global GDP in 2030". IDC. 2024.
6 "Alphabet unveils $75b AI investment plan for 2025". Tech in Asia. 5 February 2025.
7 "Amazon plans to spend $100 billion this year to capture 'once in a lifetime opportunity' in AI". CNBC. 6 February 2025.
8 "Meta to spend up to $65 billion this year to power AI goals, Zuckerberg says". Reuters. 25 January 2025.
9 "Microsoft reiterates plan to invest $80 billion in AI, but may 'adjust our infrastructure in some areas'". CNBC. 24 February 2025.
10 "Most popular AI apps". Backlinko. 25 February 2025.
11 "Character AI, an AI-focused startup, explores web-based gaming features". Motivo Media. 18 January 2025.
12 "Number of ChatGPT users". Exploding Topics. 22 February 2025.

Similar but different: The emerging GenAI value chain

Along with the rise of GenAI 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 GenAI image recognition capabilities for a variety of use cases in automation, mobility, and manufacturing. Training and leveraging such foundational models for inference, 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 GenAI value chain



Stack it up: Value creation in the age of GenAI

From end-users to suppliers, nothing is yet set in stone in the GenAI value chain. Given the rapid pace of its 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[13].

To this end, we believe that Southeast Asian players should consider their value creation activities across all three segments of the GenAI value chain: Application, Platform, and Infrastructure[14] (see Figure 2).

Figure 2: Segments and sub-segments in the GenAI value chain

			Examples
Application	Consumer applications	Applications focused on the generation of text, code, image, and video, as well as AI-powered search	• *ChatGPT* • *Perplexity*
	Enterprise applications	Applications focused on the summarisation of complex documentation, retrieval augmented generation (RAG) search, and natural language processing (NLP) chat agents	• *Cohere* • *Harvey*
	Government applications	Applications focused on waste, parking, and security management, smart cities, urban mobility, and defence	• *Hayden AI*
Platform	Foundational models	Language models, diffusion models, and world models to power applications and autonomous vehicles/robots	• *Open AI* • *DeepSeek* • *Garuda LLM*
	DevOps tools and platforms	IT infrastructure and software tools for model development, orchestration, and observability	• *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	• *DDN* • *Pure Storage*
Infrastructure	GPU and GPU clouds	Compute hosted on-premise and on clouds dedicated to AI applications and models	• *CoreWeave* • *Lambda* • *YTL* • *Singtel* • *IOH*
	Data centres	Data centres customised for high rack density AI GPU workloads	• *QTS* • *Crusoe* • *Gulf* • *AIS* • *PLDT* • *Telkom*
	Renewable energy infrastructure	Renewable solutions with colocation and reduced transmission losses for AI loads	• *Crusoe* • *Nuclear small modular reactors (SMRs)*

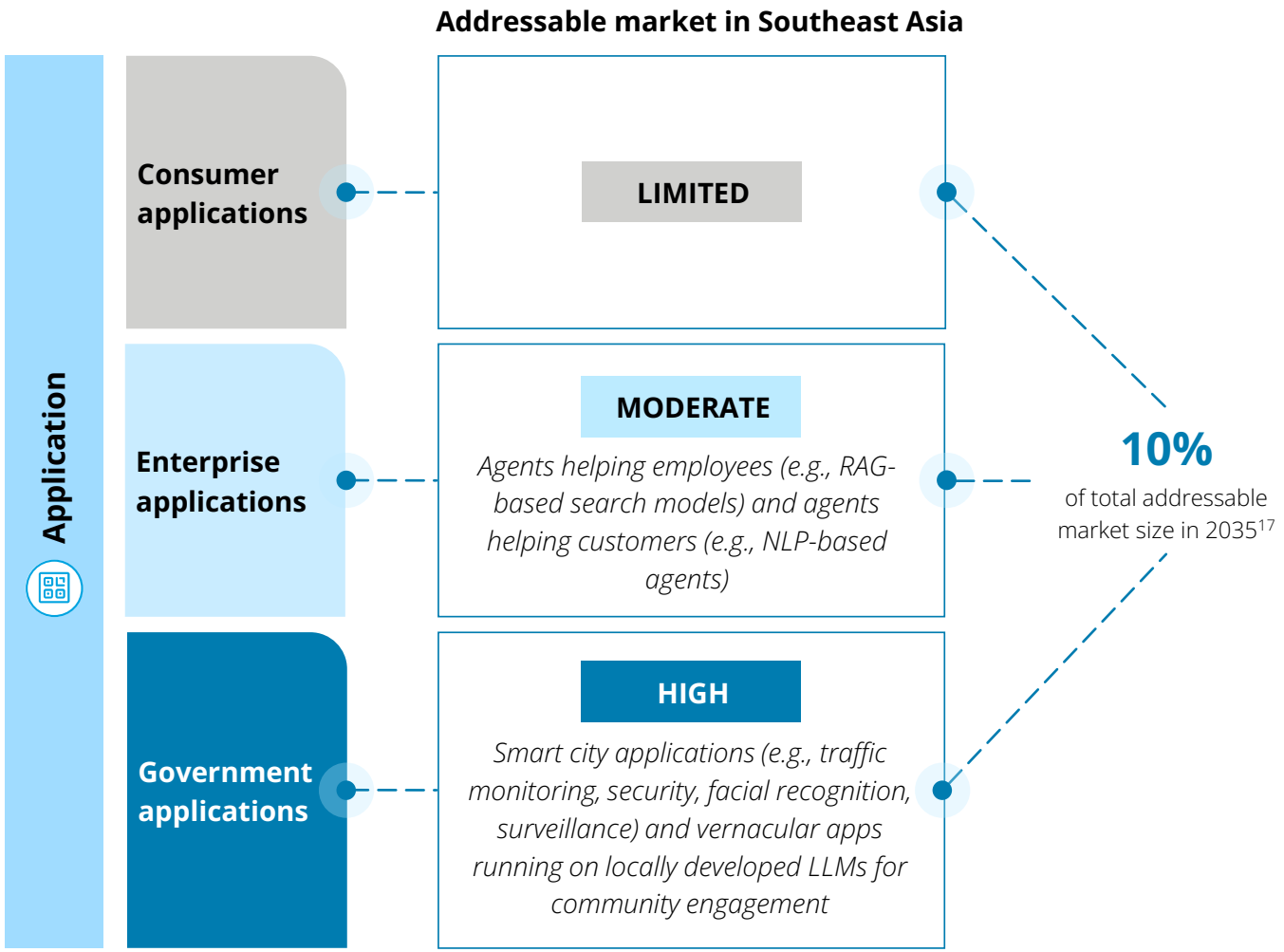
13 "DeepSeek R1's implications". IoT Analytics. 5 February 2025.
14 Capitel analysis.

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)[15], with one in four based in Indonesia[16]. Such patterns reveal significant value leakage and the need to better retain consumer AI value within the region.

However, as Consumer applications have high development costs and are dominated by global players, most of the value for Southeast Asia 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



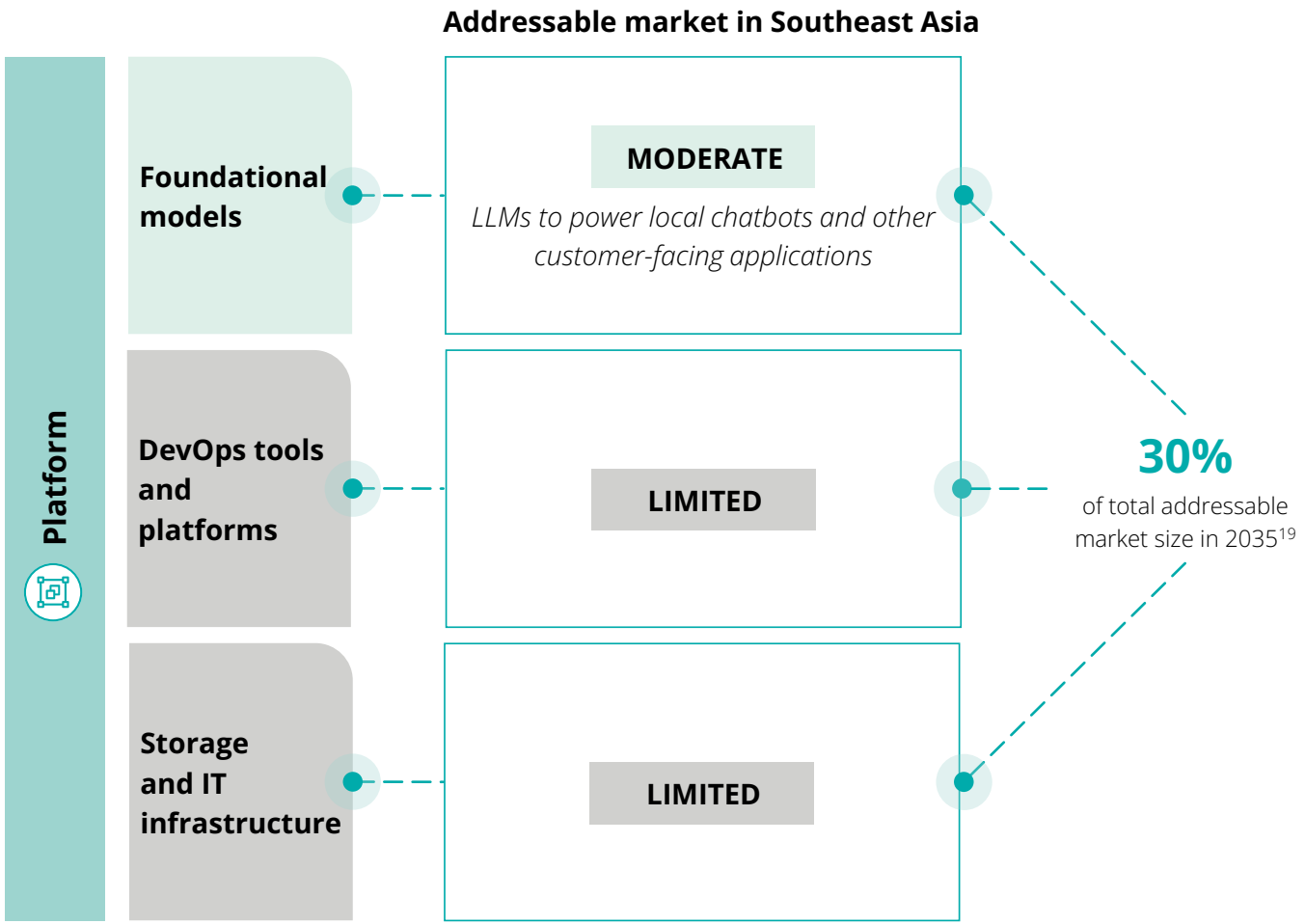
15 "Report: Perplexity business breakdown and founding story". Contrary Research. 2024.
16 "The latest perplexity AI stats (2024)". Exploding Topics. Accessed on 25 February 2025.
17 Deloitte analysis.

Platform

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)[18].

Figure 4: Value creation in the Platform segment



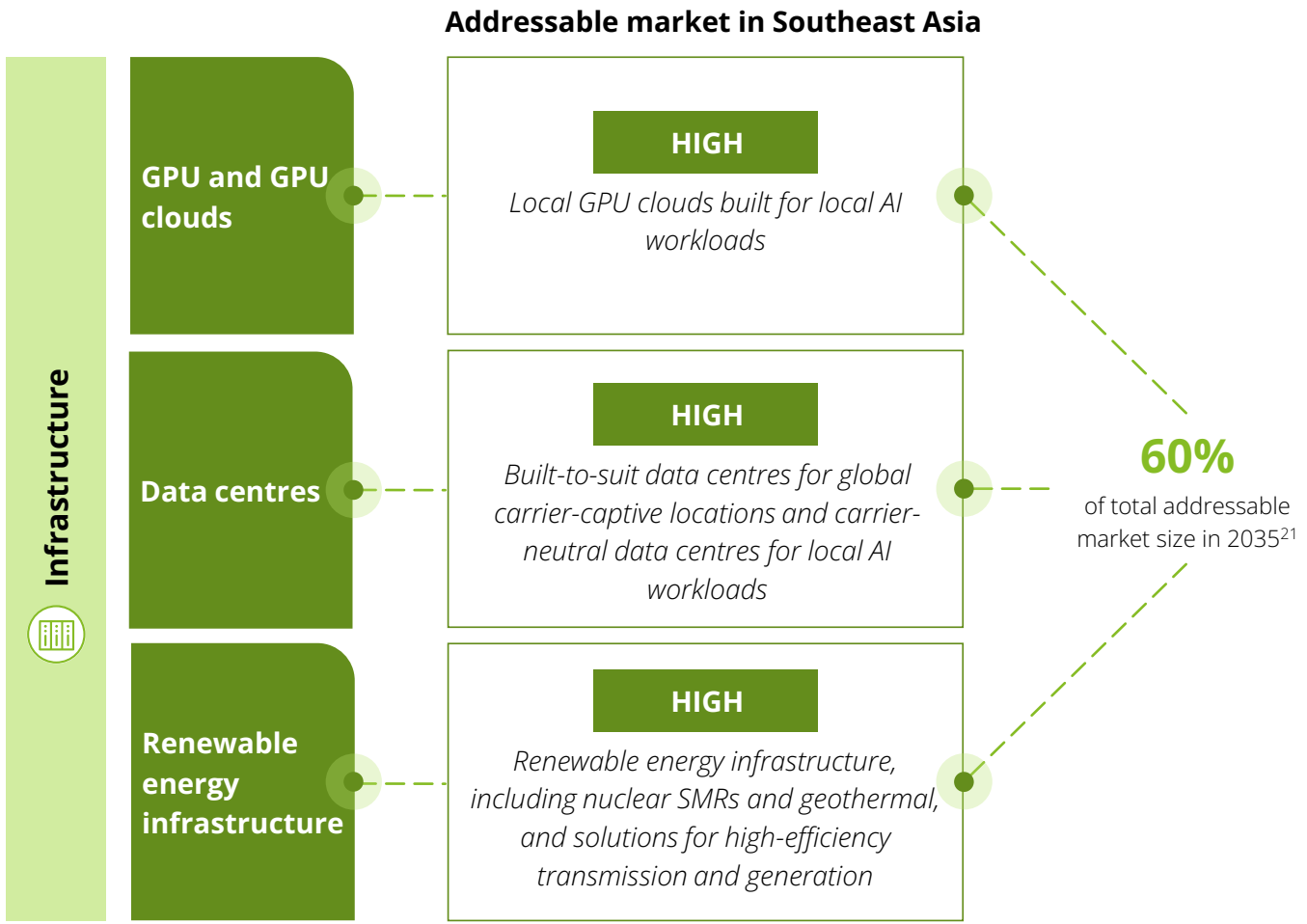
18 "Singapore pioneers S$70m flagship AI initiative to develop Southeast Asia's first large language model ecosystem catering to the region's diverse culture and languages". Infocomm Media Development Authority. 2023.
19 Deloitte analysis.

Infrastructure

GenAI requires AI clusters with neural networks in the range of billions or even trillions of parameters, as well as high-performance graphics processing unit (GPU) computing systems for large-scale, intensive workloads[20]. This means that many of Southeast Asia's existing data centres cannot be used to provide GenAI training and inference capabilities (see Figure 5).

Given that data centres are massive energy guzzlers and require reliable electricity supply 24/7 with high levels of redundancy, it will also be an uphill task to keep up with energy demand. Each of these Infrastructure sub-segments are therefore critical future "toll roads" driving investment and value across Southeast Asia.

Figure 5: Value creation in the Infrastructure segment



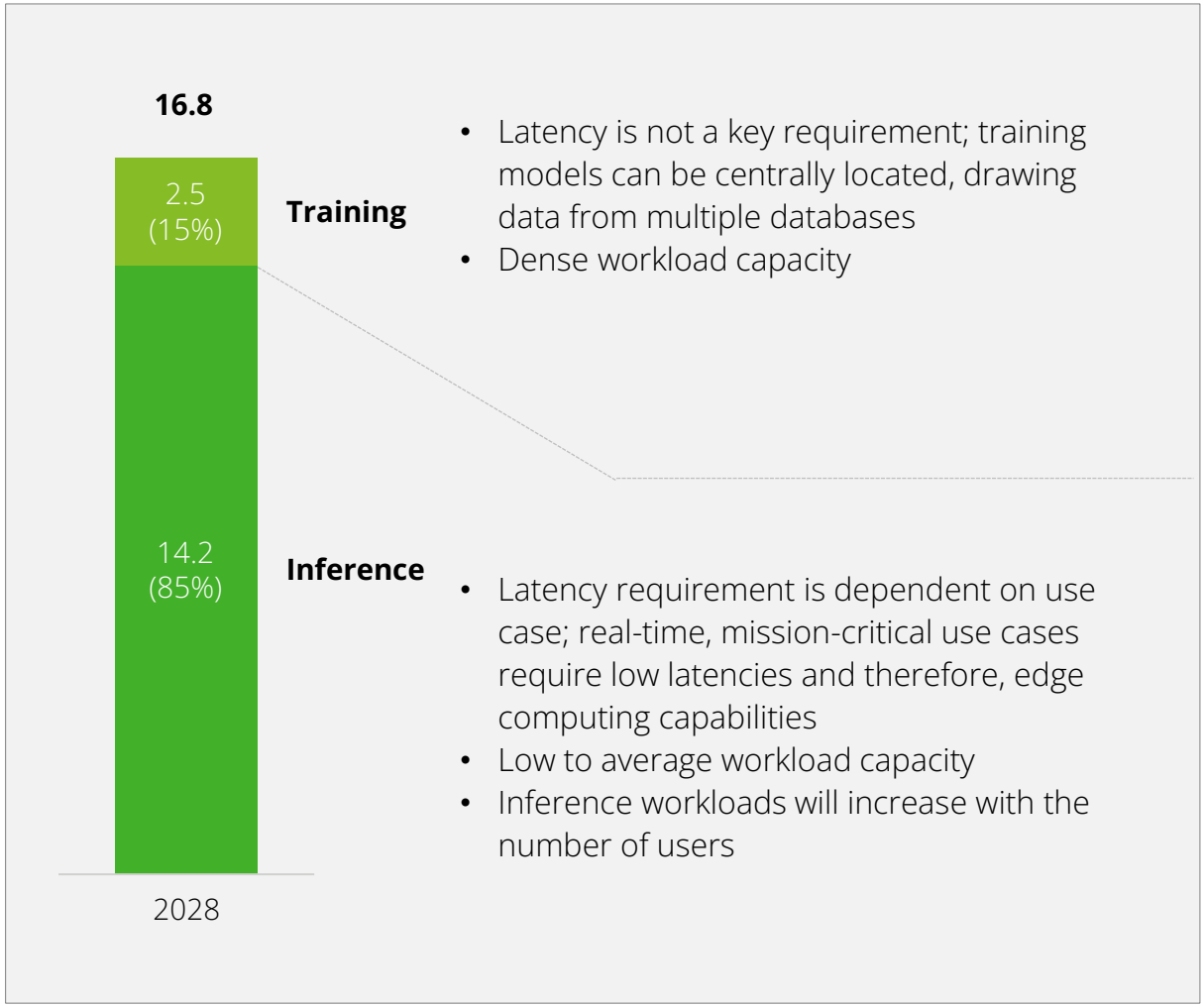
20 "What generative AI means for data centres". Equinix.2023.
21 Deloitte analysis.

GPU and GPU clouds

Globally, 85% of global GPU demand is expected to be from AI inference[22] (see Figure 6). Due to low latency requirements, these will 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[23] and Malaysia[24], 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[25].

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



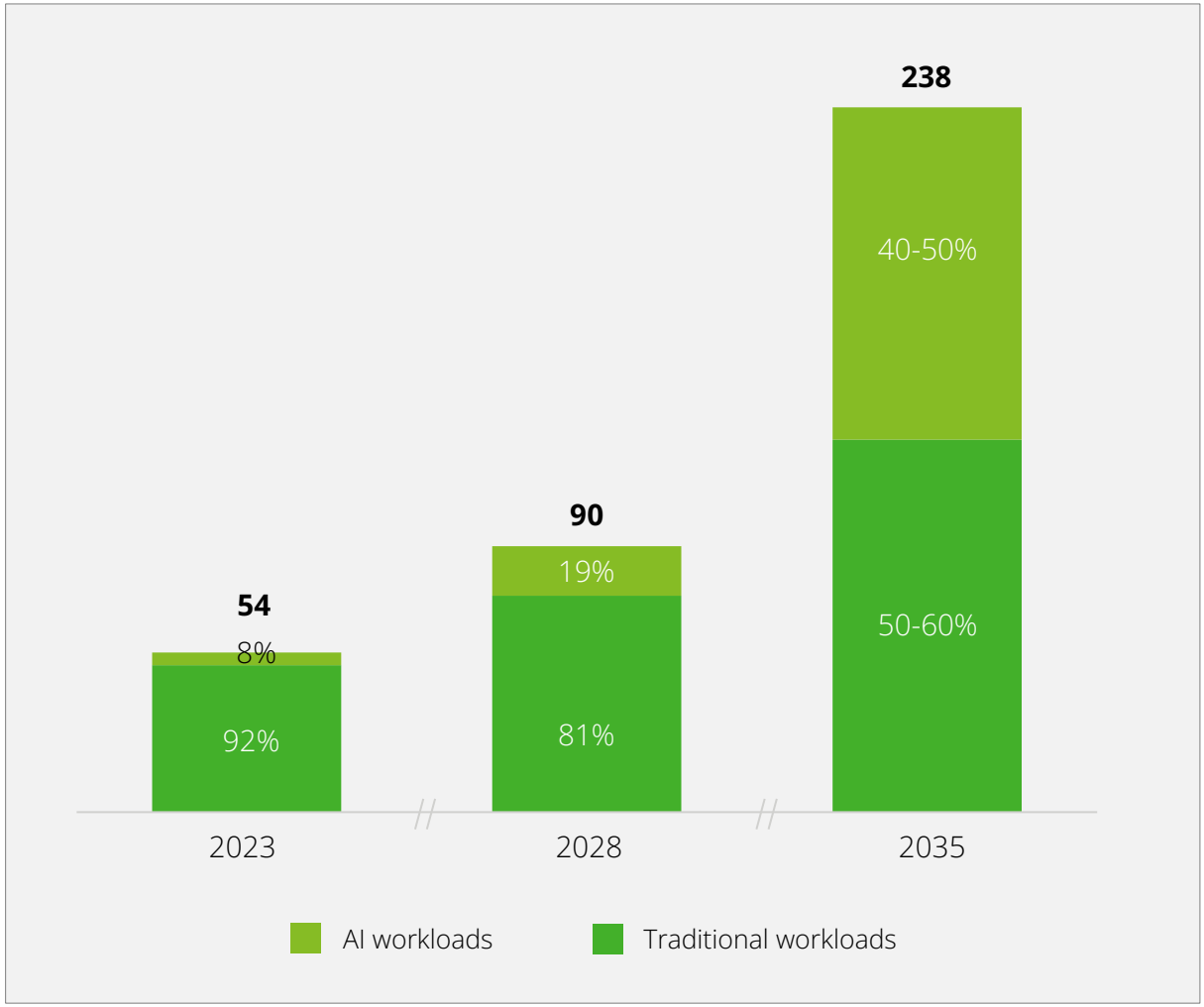
22 "The AI disruption, challenges and guidance for data centre design". Schneider Electric. September 2023.
23 "Indosat partners with Nvidia for $200m AI centre in Indonesia". Data Centre Dynamics. 5 April 2024.
24 "Nvidia & YTL Power partner for $4.3bn AI data centres in Malaysia". Data Centre Dynamics. 11 December 2023.
25 "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[26] (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 to ensure that data and infrastructure remain on their shores.

To serve local AI inference workloads, local data centres will need to be purpose-built to handle high density workloads, including higher floor loading and liquid cooling technologies. 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 within Southeast Asia.

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



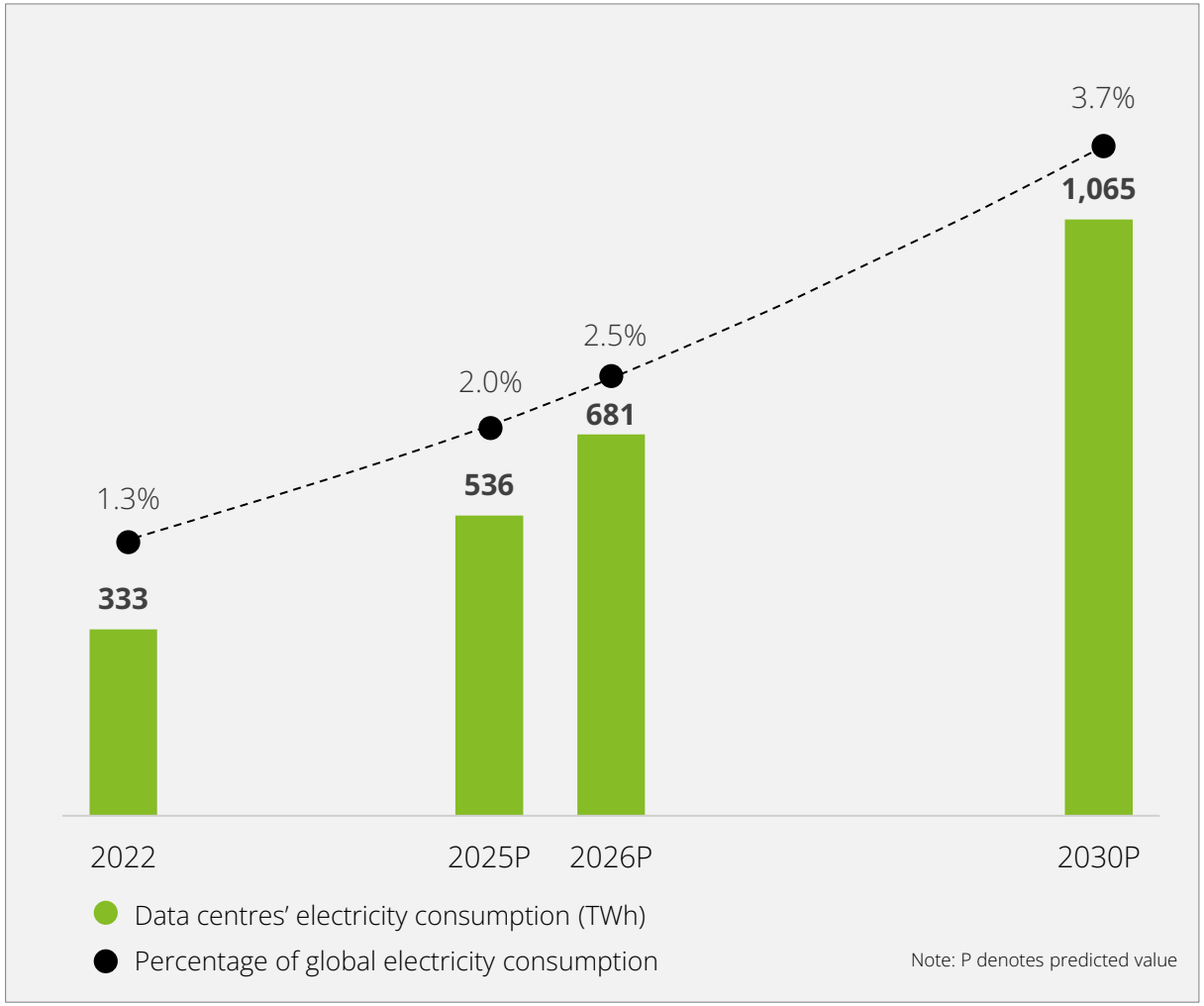
26 "The AI disruption, challenges and guidance for data centre design". Schneider Electric. September 2023.

Energy infrastructure

On the back of power-intensive GenAI training and inference, global data centre electricity consumption could roughly double in the five-year period from 2025 to 2030[27] (see Figure 8). For Southeast Asia, the demand for energy to power data centres for inference will always be local.

Governments should consider incentivising investments in energy infrastructure, including renewables – such as solar and biomass in Malaysia[28], nuclear SMRs in Singapore[29], and geothermal in Indonesia[30] – and leveraging specialised solutions for high efficiency transmission and generation now available on the market.

Figure 8: Global electricity consumption by data centres (2022-2030P)



27 "As generative AI asks for more power, data centers seek more reliable, cleaner energy solutions". Deloitte. 19 November 2024.
28 "UVCell Solar and Iozela Data Center collaborate to develop major projects in Penang and Pahang". The Sun. 29 December 2024.
29 "Small modular reactors a possible solution to power Singapore's data centres, says Energy Market Authority CEO". Channel NewsAsia. 21 October 2024.
30 "Indonesia's largest geothermal firm aims to power data centres". Bloomberg. 31 July 2024.

All systems go: Move now, and move quickly

National governments, national players, global technology companies, 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

Managing risks such as technology availability (particularly for GPUs in an export-controlled system), commercial (through close collaborations with national players on demand forecasts), consumer safety, data privacy, and cyber risks (e.g., Singapore's Digital Infrastructure Act[31])

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

National players

National 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, which part of the value chain to target, and how much 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

Investors

The capital needs to build data centre and other AI infrastructure are staggering. Building a 100-megawatt AI-ready data centre cost 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 megawatt 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 megawatt 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[32])

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

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

32 "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 risk of underinvesting in data centres and other 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

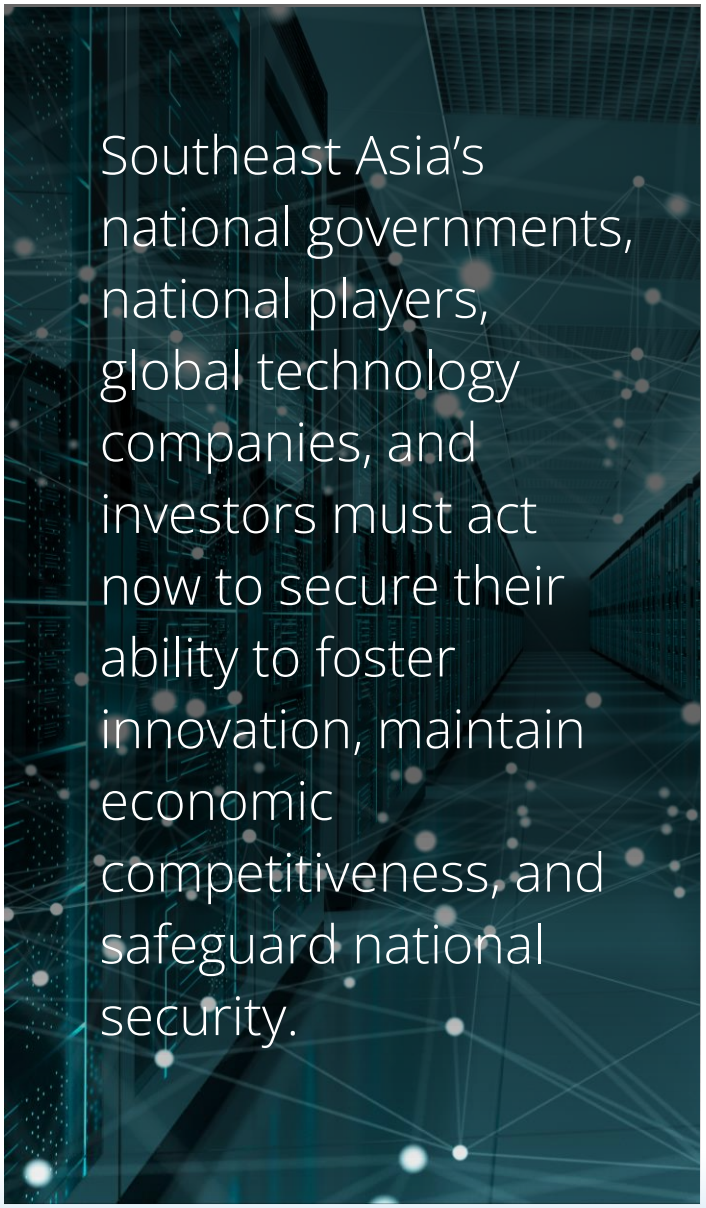
Data centres and other infrastructure are linchpins to AI participation

3

Value preservation, sovereignty, and security considerations underscore the need for on-shore 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
chiyang@deloitte.com

Piyush Jain
Strategy, Risk & Transactions Leader for Technology, Media & Telecommunications
Deloitte Southeast Asia
pijain@deloitte.com

Haridas Kanagasabai
Audit & Assurance Leader for Technology, Media & Telecommunications
Deloitte Southeast Asia
kaharidas@deloitte.com

Farhan Rashid
Strategy, Risk & Transactions Director
Deloitte Singapore
fmohamedrashid@deloitte.com

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