



DataGateway Sdn Bhd (202401021123 – 1566972-V)
UOA Business Park, Tower 8, Level 2, Unit 2-1,
No 1 Jalan Pengatucara U1/51A
40150 Shah Alam, Selangor, Malaysia
T + 603 5590 9883, F + 603 5569 3355
Website: <https://areagroup.my>

Understanding and navigating the AI Diffusion Framework

Implications and Opportunities for Malaysia

Disclaimer: This document was first issued on 9 May 2025. The contents herein are subject to amendments or revisions at the discretion of the publisher without prior notice



DataGateway Sdn Bhd (202401021123 – 1566972-V)
UOA Business Park, Tower 8, Level 2, Unit 2-1,
No 1 Jalan Pengatucara U1/51A
40150 Shah Alam, Selangor, Malaysia
T + 603 5590 9883, F + 603 5569 3355
Website: <https://areagroup.my>

For more information on this publication, visit <https://areagroup.my>

About Us

AREA Market Intelligence (AMI) is a research and advisory platform under the AREA Group that provides insights to support strategic real estate, industrial, and digital infrastructure development. Our mission is to enable better decision-making through data-driven analysis, helping investors, developers, and policymakers unlock the full potential of locations across Malaysia and the region. AMI is independent, industry-informed, and focused on delivering real-world value through local expertise and global perspectives.

Research Integrity

At AMI, we are committed to providing accurate, unbiased, and actionable intelligence through rigorous data analysis, field research, and stakeholder engagement. We adhere to the highest standards of research integrity by ensuring transparency in our methodology, independence in our conclusions, and clarity in our communications. Our processes include independent peer review, source validation, and strict data governance policies to maintain the credibility and reliability of our findings.

AMI's publications do not necessarily reflect the views of AREA Group's commercial affiliates or partners.

Published by AREA Market Intelligence (AMI), a division of DataGateway Sdn Bhd, a subsidiary of the AREA Group of Companies, Malaysia

© 2025 AREA Group

Limited Print and Electronic Distribution Rights

This publication and its content are protected by copyright and are intended for non-commercial use only. Unauthorized reproduction or redistribution in any form, including online sharing, is prohibited unless explicitly approved by AMI. Excerpts may be cited with proper attribution and linked to the original publication.



DataGateway Sdn Bhd (202401021123 – 1566972-V)
UOA Business Park, Tower 8, Level 2, Unit 2-1,
No 1 Jalan Pengatucara U1/51A
40150 Shah Alam, Selangor, Malaysia
T + 603 5590 9883, F + 603 5569 3355
Website: <https://areagroup.my>

Table of Contents

Executive Summary	4
Understanding the Diffusion Framework.....	6
Chips Act: Economic Repatriation and Guardrails	6
Diffusion Framework: Global Compute Governance	7
Validated End User (VEU) Pathways – A Loophole for the Trusted.....	11
The New Frontier: Software-Based Controls	13
Foreign Ownership, Control, or Influence (FOCI)	14
White-zone (Inference) vs. Black-zone (Training) vs Grey Zone Chips.....	14
International Reaction	16
Why is Malaysia Categorized as Tier 2?	17
What does all this mean for Malaysia?	18
Are Tier 2 nations ready for Life 3.0?	21
Pathways, Innovation and Strategic Positioning - Malaysia.....	24
Pathway to Compliance and Growth in line with the Framework.....	24
Innovation Under Constraint.....	26
Building the Backbone	28
From Market of Convenience to Partner of Choice.....	29
Strategic Risks and Tactical Openings	31
Conclusions: From Compliance to a Competitive Edge	32
References	33

Executive Summary

In January 2025, the United States formalized the *AI Diffusion Framework*, a comprehensive export control regime regulating the global use of advanced U.S.-origin AI chips. Under this system, countries are categorized into three tiers based on trust, governance, and geopolitical alignment. Countries are grouped into three tiers:

- Tier 1 (e.g., U.S., Japan, UK): Full access
- Tier 3 (e.g., China, Russia): Complete bans
- Tier 2 (including Malaysia): Subject to strict limits and licensing

Malaysia is designated Tier 2, a middle-ground classification that permits limited access to high-end AI hardware under strict licensing rules and compliance oversight.

What's the Difference Between AI Training and Inference?

AI Training

- **What it is:** The process of teaching a machine learning model by feeding it vast amounts of data so it can learn patterns.
- **Requires:** Huge volumes of data, very powerful chips (e.g., NVIDIA H100), and weeks of continuous computation.
- **Use case:** Training a new language model like ChatGPT from scratch

AI Inference

- **What it is:** The process of using a trained model to make predictions or generate outputs.
- **Requires:** Far less computing power, works with older or export-compliant chips.
- **Use case:** Running a chatbot, recommending products, scanning X-rays, analyzing CCTV feeds

Why it matters:

Under the U.S. AI export controls, **training is restricted** due to national security concerns, but **inference is allowed**. Malaysia can legally build and scale AI services by focusing on inference.

As a Tier 2 nation, Malaysia faces a hard cap of approximately 50,000 high-end AI chips (e.g., NVIDIA H100) from 2025 to 2027. Any attempt to build large-scale training clusters requires formal U.S. export licenses. U.S. cloud providers must also retain at least 50% of global compute within the U.S. and are capped at deploying 7% per Tier 2 country, further limiting Malaysia's access to frontier AI resources.

Yet within these restrictions lies a strategic opening. The framework does not apply to inference-based workloads, allowing Malaysia to host export-compliant, commercially viable AI-as-a-Service platforms at scale. In addition, compliance pathways—such as Validated End-User (VEU) designations—allow for expanded GPU quotas if rigorous security and governance standards are met.

This paper presents a full architecture of the AI Diffusion Framework and outlines:

- The structural reasons Malaysia remains Tier 2: regulatory immaturity, re-export risks, energy bottlenecks, and lack of AI governance capacity.



DataGateway Sdn Bhd (202401021123 – 1566972-V)
UOA Business Park, Tower 8, Level 2, Unit 2-1,
No 1 Jalan Pengatucara U1/51A
40150 Shah Alam, Selangor, Malaysia
T + 603 5590 9883, F + 603 5569 3355
Website: <https://areagroup.my>

- The broader question. Are Tier 2 nations truly prepared to wield Life 3.0 technologies—AI systems that adapt, evolve, and potentially surpass human control?
- The risks Malaysia faces from compliance scrutiny and power grid fragility to the growing trust gap with U.S. partners.
- The tactical options Malaysia can pursue include positioning itself as a neutral inference hub, embracing efficient AI models, investing in export-compliant infrastructure, and doubling down on institutional reform.

Malaysia's Tier 2 status is not just a technical limitation—it reflects a global perception of readiness, trust, and governance. To shift that perception, Malaysia must reinforce its role as a responsible digital partner, not just a hosting market. That includes:

- Strengthening grid resilience and regulatory enforcement,
- Developing a dedicated AI governance authority,
- Building talent pipelines and centers of excellence, and
- Actively contributing to global AI safety standards.

Throughout this document, we introduce a concept that was originally introduced by Max Tegmark and that is Life 3.0 where he explores how AI could reshape the life on Earth and beyond.

In conclusion, while Malaysia may not be the next global hub for frontier AI training or model development, it can carve out a leadership position in inference, infrastructure, and compliance-driven innovation. By focusing on secure deployment, energy-efficient AI services, and component manufacturing, Malaysia can become indispensable to the AI value chain—powering the ecosystem even if it doesn't train the models at its core.

This is not just about adapting to constraints; it's about earning trust, building credibility, and rising—step by step—toward Tier 1 status. With deliberate investment in regulation, governance, and strategic partnerships, Malaysia can transform itself from a Tier 2 implementer into a Tier 1 influencer.

We may not shape the neural networks of tomorrow—but we can shape the frameworks, systems, and safeguards that make them useful, secure, and aligned. That's how Malaysia contributes to the architecture of Life 3.0—and earns its place in the AI future.

Understanding the Diffusion Framework

The U.S. has implemented two complementary mechanisms to shape global AI development: the CHIPS and Science Act and the AI Diffusion Framework. While the CHIPS Act aims to rebuild America’s domestic semiconductor manufacturing capabilities, the AI Diffusion Framework is designed to regulate where and how advanced U.S.-origin AI chips may be used globally.

Chips Act: Economic Repatriation and Guardrails

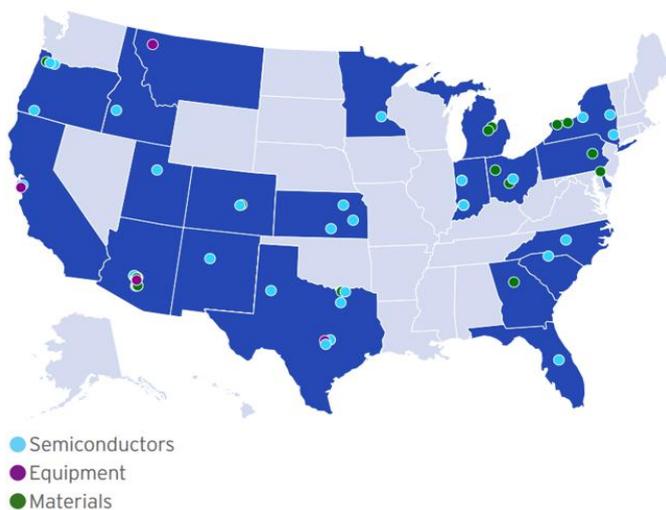
In short, “no American-funded fabs in China,”

Enacted in 2022, the CHIPS Act represents a USD 52 billion industrial policy designed to accelerate semiconductor manufacturing, research, and innovation within the United States. Its core function is economic—to incentivize the reshoring of chip fabrication and reduce U.S. dependence on foreign supply chains.

However, embedded within the funding structure are strategic “guardrails”: recipients of CHIPS subsidies must commit not to expand advanced semiconductor production in Tier 3 countries, including China, Iran, and Russia. These provisions effectively block American taxpayer-funded fabs from contributing to adversarial capabilities abroad.

Notably, the CHIPS Act does not restrict how Tier 1 countries (U.S. allies) may deploy AI infrastructure internationally. Its scope remains limited to the manufacturing footprint, not the deployment of compute resources or cloud-based AI capacity.

NEW SEMICONDUCTOR PROJECTS



50+

new semiconductor ecosystem projects announced across the U.S., including the construction of new semiconductor manufacturing facilities (fabs), expansions of existing sites, and facilities that supply the materials and equipment used in chip manufacturing

\$210B+

in private investments announced across 20 states to increase domestic manufacturing capacity

44,000+

new high-quality jobs announced in the semiconductor ecosystem as part of the new projects, which will support hundreds of thousands of additional jobs throughout the broader U.S. economy

Reference: [2]

Diffusion Framework: Global Compute Governance

At its core, the AI Export Control Framework—introduced by the U.S. Bureau of Industry and Security (BIS) in early 2025—establishes the world’s first formal system for governing the global diffusion of advanced AI compute. Unlike the CHIPS Act, which focuses on domestic fabrication, this framework controls where and how U.S.-origin AI chips may be deployed worldwide, using a tiered geopolitical classification, hardware performance thresholds, and corporate residency rules. *“This policy will help build a trusted technology ecosystem around the world and allow us to protect against the national security risks associated with AI, while ensuring controls do not stifle innovation or US technological leadership,”* said U.S. Secretary of Commerce Gina Raimondo.

In short, the objective is clear:

“Ensure no American top GPUs power adversarial (Tier 3) AI ecosystems.”

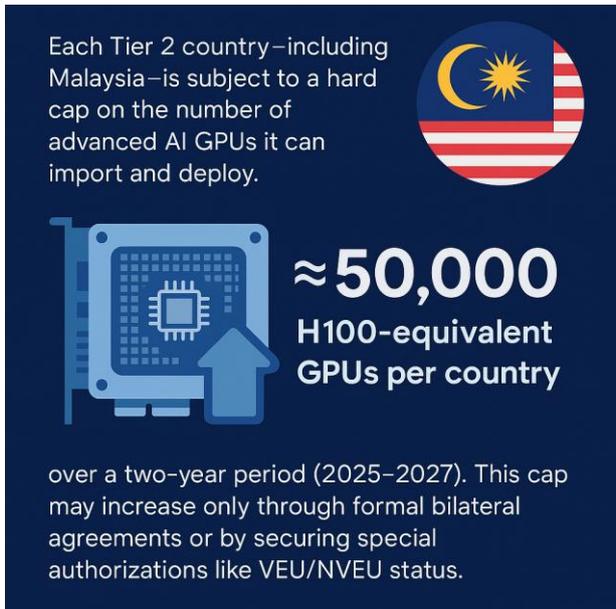
A Geopolitical Framework

Countries are divided into three tiers based on strategic alignment and perceived national security risk.



Tier	Countries	Access Level	Why It Matters
Tier 1	U.S. & key allies (e.g. Japan, UK, Korea, EU core)	<ul style="list-style-type: none"> Full access to U.S. AI chips Must keep ≥75% of compute in Tier 1 Max 7% in any Tier 2 country 	Trusted defense & tech partners. Strong safeguards. Low risk of diversion.
Tier 2	Most countries (e.g. Malaysia, Brazil, UAE)	<ul style="list-style-type: none"> Capped at ~50,000 GPUs (2025–27) Value may increase subject to compliance, usage, and end-user verification. 	Strategic middle ground. Not fully trusted, but not adversaries. Compliance opens doors.
Tier 3	Embargoed states (e.g. China, Russia, Iran)	<ul style="list-style-type: none"> Complete ban on AI chip exports 	High risk of misuse. Seen as adversaries or tech threats. No access permitted.

The framework introduces two interlocking mechanisms of control:



Each Tier 2 country—including Malaysia—is subject to a hard cap on the number of advanced AI GPUs it can import and deploy.

≈ 50,000 H100-equivalent GPUs per country

over a two-year period (2025–2027). This cap may increase only through formal bilateral agreements or by securing special authorizations like VEU/NVEU status.

Absolute Limits

Each Tier 2 country—including Malaysia—is subject to a hard cap on the number of advanced AI GPUs it can import and deploy.

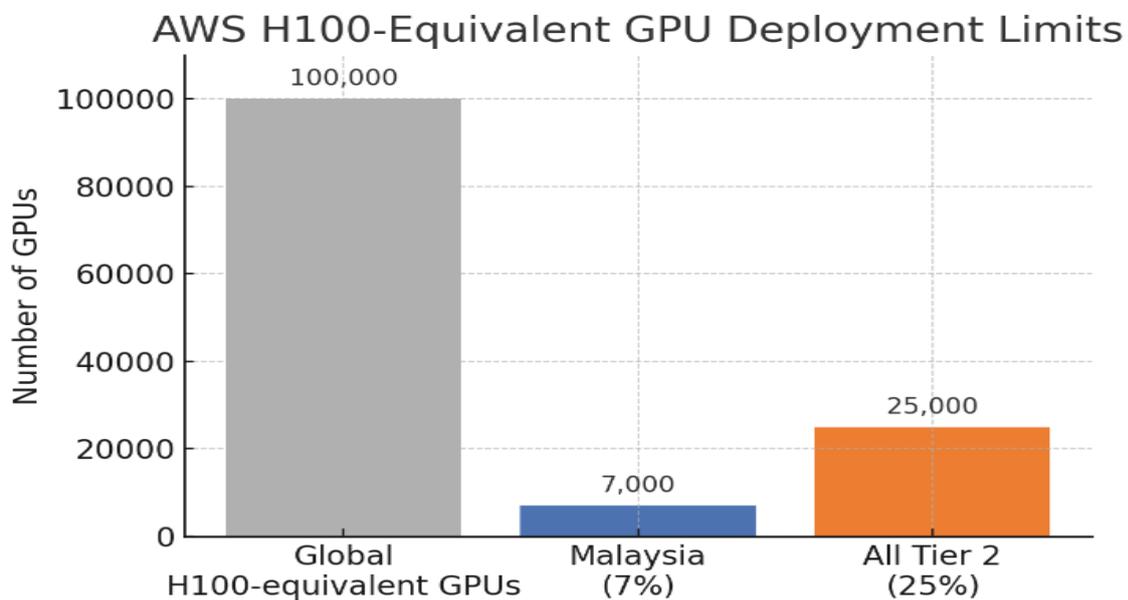
Based on publicly disclosed thresholds and total performance ceilings, this translates to approximately 50,000 H100-equivalent GPUs per country over a two-year period (2025–2027). This cap may increase only through formal bilateral agreements or by securing special authorizations like VEU/NVEU status.

Relative Capacity Allocation Rules

These rules apply to U.S.-headquartered AI cloud providers and dictate how they distribute compute power globally:

- 7% per-country cap for any single Tier 2 nation (e.g., Malaysia).
- 25% ceiling across all Tier 2 countries combined.
- 50% retention requirement: U.S.-based firms must keep at least half their AI compute within the United States.

For example, if AWS has 100,000 H100-equivalent GPUs worldwide, no more than 7,000 can be deployed in Malaysia, and cannot exceed 25,000 GPUS across all Tier 2 markets combined, as explained in the infographic below:



Additional provisions include:

- Unrestricted export rights for Tier 1 allies, who face no hard caps and benefit from streamlined approvals.
- Total prohibition of AI chip exports to Tier 3 countries (e.g., China, Iran, Russia) under ECCN 3A090, due to strategic and national security risks.

Together, these rules form a dual-lock system, limiting both the total volume of AI chips per country and the relative deployment footprint of U.S. firms abroad. This ensures that cutting-edge AI compute remains concentrated among trusted allies while restricting diffusion to potentially adversarial regions.



DataGateway Sdn Bhd (202401021123 – 1566972-V)
 UOA Business Park, Tower 8, Level 2, Unit 2-1,
 No 1 Jalan Pengatucara U1/51A
 40150 Shah Alam, Selangor, Malaysia
 T + 603 5590 9883, F + 603 5569 3355
 Website: <https://areagroup.my>

The table below shows side-by-side comparison table outlining the key differences between the CHIPS Act and the AI Diffusion Framework, with a special emphasis on their respective impacts on Malaysia.

Aspect	CHIPS Act	AI Diffusion Framework
Purpose	Stimulate domestic semiconductor manufacturing and R&D	Regulate export and global deployment of AI compute
Administered By	U.S. Department of Commerce	U.S. Bureau of Industry and Security (BIS)
Target	Chip fabrication and R&D funding	AI chips and compute capacity allocation
Scope	Domestic manufacturing & foreign expansion limits	Global AI infrastructure deployment
Countries Affected	Applies only to U.S. subsidy recipients	All nations (Tier 1, 2, 3 classification)
Restrictions Imposed	No advanced fabs in Tier 3 countries	Caps on chip exports and deployment by geography
Impact on Malaysia	<u>Minimal</u> : Malaysia not restricted for deployment	<u>Significant</u> : Malaysia capped at ~50K GPUs, 7% compute share



DataGateway Sdn Bhd (202401021123 – 1566972-V)
UOA Business Park, Tower 8, Level 2, Unit 2-1,
No 1 Jalan Pengatucara U1/51A
40150 Shah Alam, Selangor, Malaysia
T + 603 5590 9883, F + 603 5569 3355
Website: <https://areagroup.my>

Validated End User (VEU) Pathways – A Loophole for the Trusted

“Entities that comply with U.S. export rules may be eligible for expanded access.”

To soften the impact on industry, the U.S. framework provides a compliance mechanism called Validated End User (VEU) authorization. A VEU is a pre-cleared, trusted entity (company or organization) that can receive controlled items with less red tape. Under the AI chip rules, data center operators can apply to become VEUs to bypass some country caps.

There are two types of validated end users:

Universal VEUs (UVEUs) are global tech giants, often Tier 1-headquartered, like Amazon or Microsoft. They receive blanket approval to deploy AI chips in Tier 2 countries (still within the 7% rule). Practically, this enables U.S. cloud providers to continue expanding into places like Malaysia – but they might have to throttle the size of any one region. (Notably, Nvidia itself and U.S. “hyperscalers” are not counted toward the host country’s cap when they operate the equipment. This means if Nvidia operates an AI cloud in Malaysia, it falls under U.S. entity control and is less restricted.)

National VEUs (NVEUs) are local firms within Tier 2 countries. If a Malaysian data center wants to import more chips—potentially up to 320,000 H100-equivalents—it must earn NVEU status. For example, a Malaysian-owned data center operator or “national champion” could apply to become an NVEU to build a large AI cluster in Malaysia. Getting there isn’t easy. Applicants must meet 19 separate security and policy requirements that include proving ironclad security and a positive track record, showing no ties to Tier 3 actors, and working closely with their governments to demonstrate end-use compliance. Malaysia’s government plays a key role here, often having to guarantee oversight and enforcement as well as regularly reporting their GPU inventory and usage to the BIS.

REQUIREMENTS FOR VALIDATED END USERS

REQUIREMENTS	
General Compliance and Proven Track Record Demonstration of physical, and Intelligence Ties to Tier 3 Nations	Transfer of Chips
No Foreign Military and Intelligence Ties to Tier 3 Nations Cut off supply chain dependencies on advanced semiconductors and network-equipment from Tier 3 countries	Intra-company Transfer Notification
	Geographic Allocations
Transfer of Chips	Advanced AI Training
Intra-company Transfer Notification	Prohibited Uses and Human Rights Safeguards
Geographic Allocations	Reporting of Chip Installations
SECURITY REQUIREMENTS	
Ownership Security of VEU	KEY BENEFITS ✓ Less export red tape ✓ Importing and deployment of more advanced AI chips in Tier 2 locations without being counted against the rational quota ✓ The ability to scale AI workloads faster at near Tier 1 levels
Baseline Security of Chips and Data	
AI-specific Cybersecurity	
Transit Security	

government support to meet required standards.

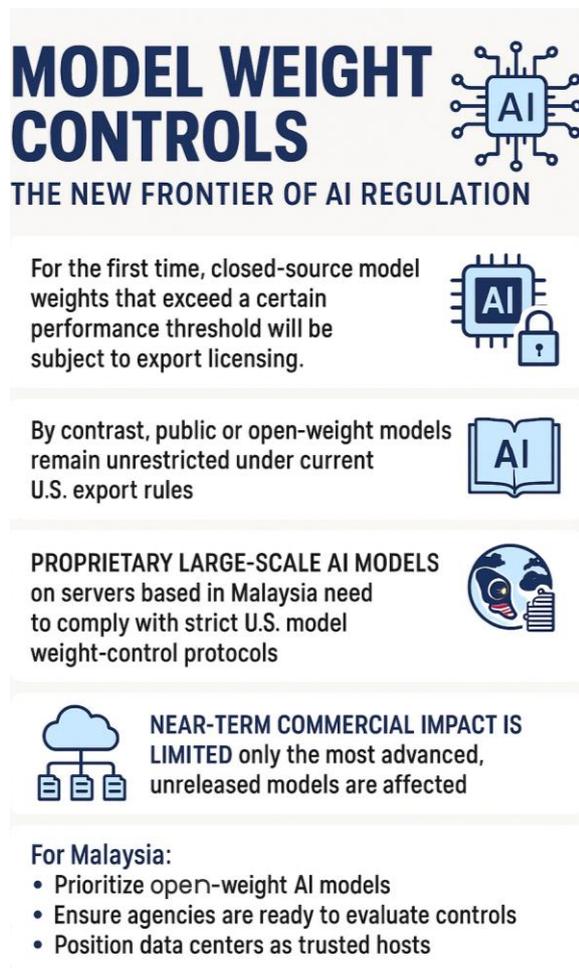
While rigorous, NVEU certification offers near-Tier 1 privileges, critical for Malaysia to scale beyond the 50,000 GPU ceiling.

Major U.S. cloud providers are expected to take advantage of this pathway, effectively serving as de facto “gatekeepers” of global AI deployment. The framework was intentionally designed to streamline exports for trusted entities—such as Microsoft, Google, and Amazon (AWS)—as well as a handful of select international firms likely to secure global authorizations. These authorizations enable continued data center expansion in markets like Malaysia, provided deployments remain under stringent compliance and within the control of trusted parties. The VEU mechanism ensures that the country-level cap does not impede cloud growth—as long as expansion stays in friendly hands.

The VEU program reflects U.S. trust in specific companies rather than entire countries. Malaysian firms aiming to benefit must proactively align with compliance expectations and may need

The New Frontier: Software-Based Controls

In a notable shift, the AI Diffusion Framework introduces export controls not just on hardware, but also on *software-derived intellectual property*—specifically, model weights. For the first time, closed-source model weights that exceed a certain performance threshold will be subject to export licensing. These weights are the critical numerical values that define how a trained AI system functions—essentially, its learned intelligence.



MODEL WEIGHT CONTROLS
 THE NEW FRONTIER OF AI REGULATION

- For the first time, closed-source model weights that exceed a certain performance threshold will be subject to export licensing.
- By contrast, public or open-weight models remain unrestricted under current U.S. export rules
- PROPRIETARY LARGE-SCALE AI MODELS** on servers based in Malaysia need to comply with strict U.S. model weight-control protocols
- NEAR-TERM COMMERCIAL IMPACT IS LIMITED** only the most advanced, unreleased models are affected

For Malaysia:

- Prioritize open-weight AI models
- Ensure agencies are ready to evaluate controls
- Position data centers as trusted hosts

By contrast, public or open-weight models, which are openly shared online and accessible to the global research community, remain unrestricted under current U.S. export rules. This means Malaysia can freely use, host, and even build on these models without triggering compliance concerns.

However, if a company or government agency wishes to store, distribute, or run proprietary large-scale AI models (e.g., GPT-5, Claude Next, Gemini Ultra) on servers based in Tier 2 nations, these nations will need to comply with strict U.S. model weight-control protocols. These include licensing, verification of end-use, and in some cases, restrictions on how the model can be updated or accessed.

While this new control adds another safeguard, its near-term commercial impact is limited—only the most advanced, unreleased models are affected. Over time, this rule may influence where major AI firms choose to host

their most powerful systems, especially in Tier 2 countries like Malaysia.

As the world moves toward Life 3.0, these weight-based controls may become just as important as chip quotas. Malaysia must treat software governance with the same seriousness as hardware control to stay ahead of compliance and ahead in regional competitiveness.



DataGateway Sdn Bhd (202401021123 – 1566972-V)
UOA Business Park, Tower 8, Level 2, Unit 2-1,
No 1 Jalan Pengatucara U1/51A
40150 Shah Alam, Selangor, Malaysia
T + 603 5590 9883, F + 603 5569 3355
Website: <https://areagroup.my>

Foreign Ownership, Control, or Influence (FOCI)

It's also important to note that the Framework for Artificial Intelligence Diffusion is not just about regulating the country destination where the chips are going, but rather the nationality of the end user receiving it. Any company with an ultimate parent in a Tier 3 country (e.g., China, Russia) will fail the “no Tier 3 ties” test and be ineligible for VEU status.

To put it into context, Volvo, which is majority owned by Chinese car manufacturer Geely, would be considered a Chinese company and have their requests for chips limited or deemed ineligible. Although Volvo Cars is headquartered in Sweden (a Tier 1 nation), its ultimate parent is Geely, a Chinese Tier 3 entity. Under the VEU criteria, that Tier 3 ownership stake would likely disqualify Volvo from UVEU status, meaning any H100/A100 request from Volvo Cars would revert to the standard license process, and very likely face a presumption of denial.

White-zone (Inference) vs. Black-zone (Training) vs Grey Zone Chips

The most powerful GPUs (A100/H100, AMD MI300X, etc.) are in the “Black Zone”, heavily restricted. In contrast, inference-optimized chips like Nvidia’s H20 and AMD’s MI300X (with lower TPP/performance density) are in the “White Zone” and remain exportable to non-embargoed countries. Since inference workloads (~70% of AI compute by 2026) are exploding, Malaysia can still freely acquire these “white zone” chips.

However, between these zones lies a “Grey Zone.” This includes:

- Chips or systems that could be used for training or inference, depending on configuration.
- Firms legally registered in Tier 1 or Tier 2 countries but owned or controlled (even partially) by Tier 3 nationals.
- Scenarios where the end use is unclear or unstated, such as reselling, smuggling, or ambiguous cloud deployments.

In this grey area, decisions often rely on case-by-case scrutiny, and even well-intentioned companies can face delays, rejections, or blacklisting if the ownership, usage, or compliance trail isn't crystal clear.

Guide to Updated Chip Export Restrictions (Oct. 2023)

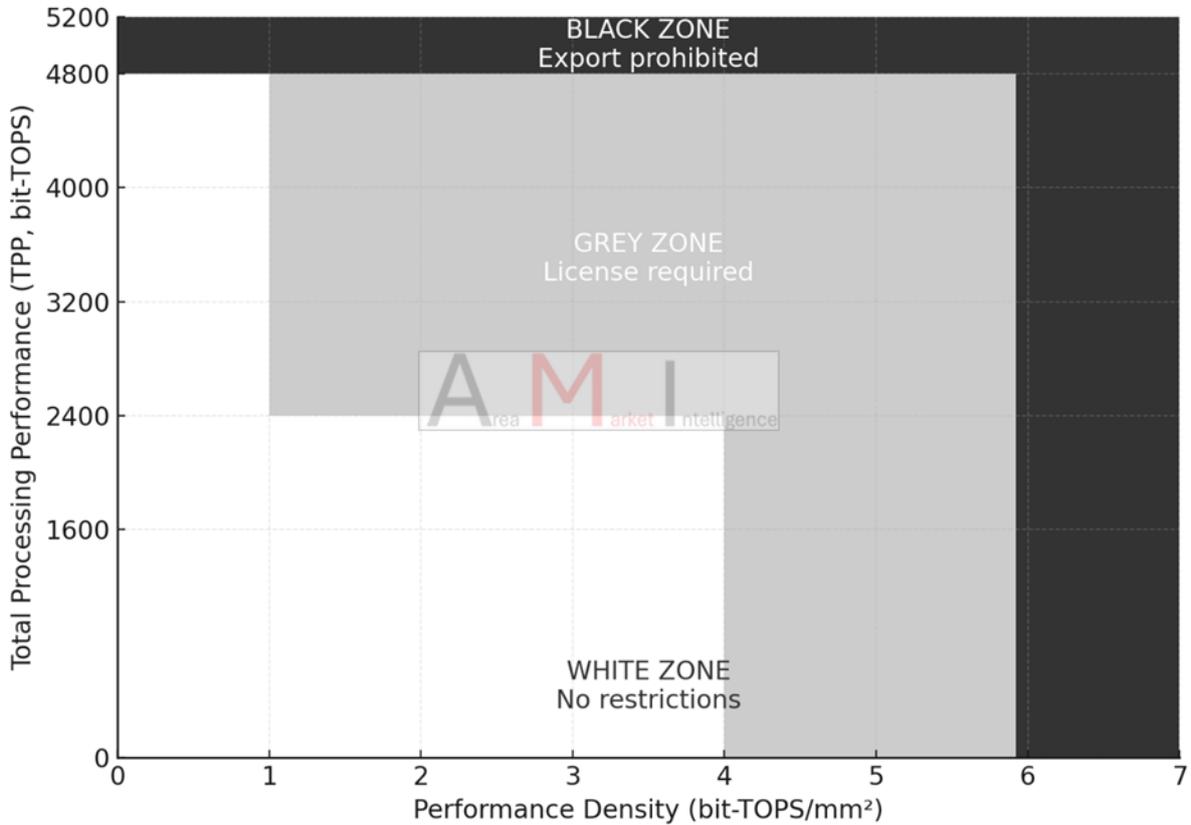


Figure: . Graphic derived from original image by Gregory C.Allen CSIS [17]

This creates an opportunity: Malaysia can emphasize AI services and inference tasks (e.g. cloud AI APIs, edge devices, smart infrastructure) that use allowed hardware, while relying on U.S. or allied Tier 1 partners for the most intensive training.

AI Workload	Typical Chips	Export Status	Malaysia Impact
Training (64-bit)	Nvidia A100/H100, AMD MI300X	Controlled (“Black Zone”) – heavily restricted outside Tier 1	Malaysia can do <i>limited</i> local training (subject to caps) but will mainly rely on Tier 1 partnerships or open models.
Inference (32-bit)	Nvidia H20, AMD MI300X (low precision)	Uncontrolled (“White Zone”) –	Malaysia can host large inference operations and AI services domestically, using available chips without quota.



AI Workload	Typical Chips	Export Status	Malaysia Impact
Grey Zone	Mixed-use chips, unclear ownership, and resale	Case-by-case, subject to scrutiny	Risky territory—requires clean ownership, end-use clarity, and compliance hygiene.

International Reaction

The global response to the U.S. AI Diffusion Rule has been mixed and largely shaped by each country's strategic alignment with the United States. Tier 1 nations—including close allies like the UK, Japan, and EU members—benefited from near-unrestricted access to U.S. AI technologies but raised concerns over regulatory ambiguity and red tape. Meanwhile, Tier 2 countries such as India, Mexico, and Israel expressed frustration, as the imposed limits on AI chip imports were seen as obstacles to their technological growth, prompting diplomatic efforts to renegotiate their positions. Tier 3 countries—including China, Russia, and Iran—were effectively cut off from access, which aligned with U.S. national security goals but deepened geopolitical divides and encouraged the development of alternative, domestic AI ecosystems.

On the industry front, U.S. tech giants like Nvidia and AMD criticized the rule for its overly complex structure, arguing it hindered competitiveness and opened the door for rivals like China to court frustrated partners. Internationally, the rule triggered debate over whether the U.S. was acting unilaterally at the expense of global collaboration and supply chain stability. While some nations viewed it as a necessary safeguard, others feared it could fragment the global AI landscape.

The Trump administration's decision to roll back the rule reflects the ongoing tension between national security priorities and the need to support open innovation in an increasingly interconnected digital economy.

As of Friday, 9th May 2025, details of these new rules are yet to be published.

Why is Malaysia Categorized as Tier 2?

Malaysia is classified as a Tier 2 nation under the U.S. AI export control framework—a system that ranks countries based on their strategic trustworthiness, regulatory maturity, and ability to prevent technology diversion.

While some of this reflects Malaysia’s genuine rise as a data center hub, U.S. agencies have privately expressed concern that Malaysia may also be serving as a re-export proxy—a potential transshipment point for restricted AI GPUs to China. The risk: Chinese entities unable to access NVIDIA’s A100/H100 chips directly may attempt to procure them via Malaysian shell companies or third-party intermediaries. Malaysia has already been flagged for tighter enforcement [11], U.S. authorities have requested enhanced monitoring of end-user recipients and supply chain transparency. And there have been arrests most notably in Singapore where three men, including two Singaporeans and a Chinese national, were charged with fraud related to the movement of AI chips [12].

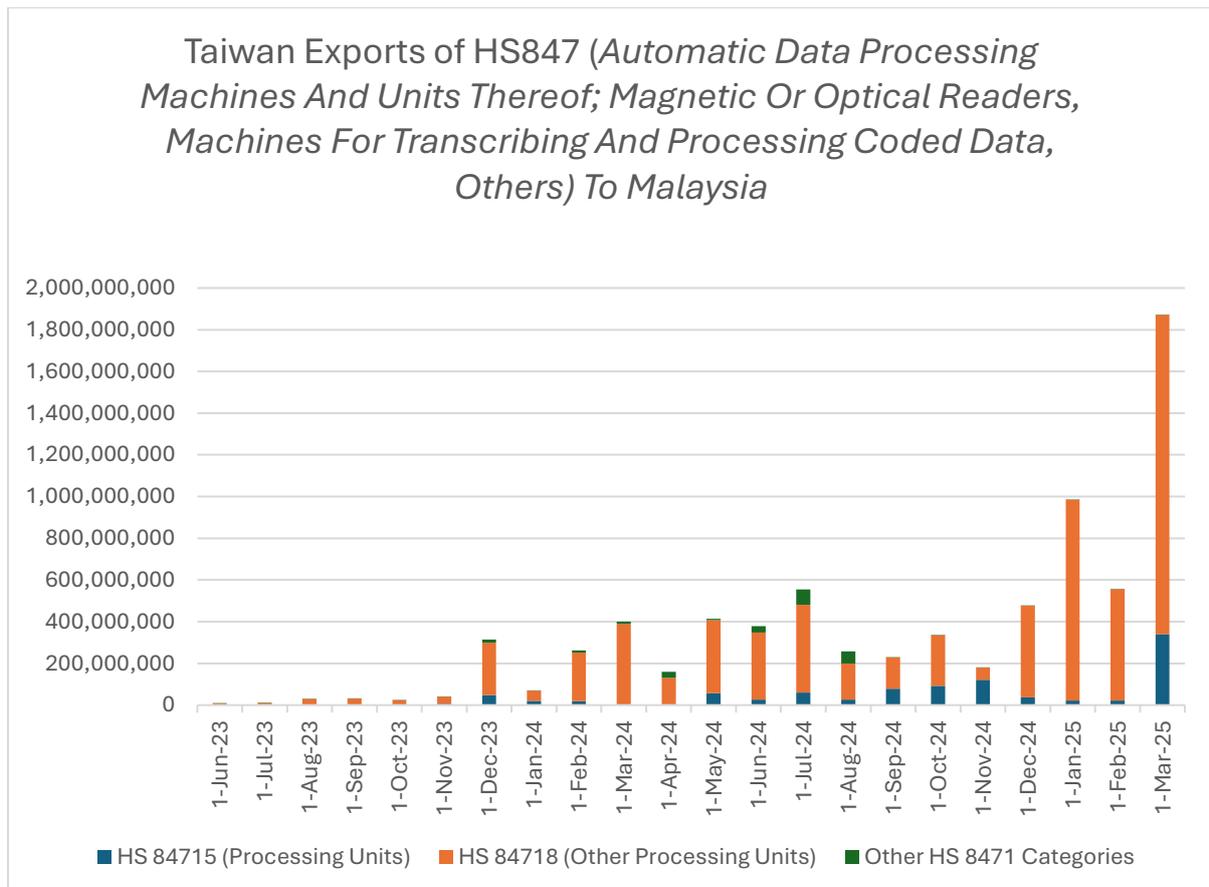


Figure: Taiwan-to-Malaysia Computer Exports, Jan 2023–Mar 2025 [13]



From a regulatory standpoint, Malaysia has yet to implement the kind of institutional oversight expected of Tier 1 countries. The absence of a centralized AI regulatory authority and limited capacity for real-time chip tracking or end-use audits weakens confidence among global technology partners. Cybersecurity laws are developing but have not reached the maturity needed to govern sensitive AI infrastructure.

Infrastructure maturity is another barrier. While Malaysia has attracted major investments from AWS, Microsoft, and Google, these efforts are largely concentrated in inference and general cloud services rather than high-stakes AI training. Moreover, auditing mechanisms and enforceable standards for data center operations remain inconsistent.

Geopolitically, Malaysia walks a fine line. Its neutrality and balanced trade relationships with both the U.S. and China are advantageous for diplomacy but can be viewed as a risk when dealing with sensitive technologies. Furthermore, Malaysia is not a signatory to the Wassenaar Arrangement—a multilateral export control regime that fosters trust among allied countries. This absence further limits Malaysia’s eligibility for streamlined technology exports.

Despite these limitations, Malaysia’s Tier 2 status should not be seen as a permanent ceiling. With targeted institutional reforms—particularly around export control enforcement, AI governance, cybersecurity, and bilateral alignment—Malaysia can strengthen its credibility and make a strong case for reclassification to Tier 1.

The next section will explore how Malaysia can navigate its current constraints and pursue strategic pathways for compliant AI growth.

What does all this mean for Malaysia?

The table below summarizes how key AI tasks are treated under the new U.S. controls.

Workload	Tier-2 Export Rule (including malaysia)
Advanced AI Model Training	Not allowed without license; Tier-1 allied data centers only [1] (beyond 25% fine-tuning exception [1]).
Model Fine-Tuning (≤25%)	Allowed with license and vetting [1].
Model Inference (API/Service)	Allowed (no new export limit); inference “as a service” is permitted [1]

Workload	Tier-2 Export Rule (including malaysia)
Model Weight Transfer/Storage	Restricted to secure Tier-1/VEU sites [1]
Export-Compliant Inference Chips	Freely exportable (no new curbs) – used for inference tasks.

For Malaysia, the AI Diffusion Rule presents a double-edged sword. On one hand, it confirms Malaysia’s placement in Tier 2—a group that is trusted enough to access U.S. inference chips and limited fine-tuning capabilities, but not enough to host full model training operations or store sensitive AI model weights without U.S. vetting. This classification limits the country’s ambitions to scale its AI infrastructure for global clients who require unfettered access to the most advanced computing capabilities. It also raises concerns for investors and hyperscale tenants exploring Malaysia as a regional AI hub.

MALAYSIA’S CONSTRAINTS UNDER THE DIFFUSION FRAMEWORK

As a Tier 2 nation under the U.S. AI Diffusion Framework, Malaysia faces significant structural limitations in scaling advanced AI infrastructure.

Absolute GPU Cap



Limited to 50,000 advanced GPUs through 2027, with scrutiny for large shipments

Relative Capacity Rules



U.S. firms bound by 7% per-country and 25% Tier 2 share limits

Asymmetry of Control



Subject to U.S. inspections and audits, but no influence over quota setting

Opportunity in Inference



Full use of export-compliant chips to power AI-as-a-Service solutions

However, this restriction can also be reframed as an opportunity. Malaysia could position itself as a specialized AI inference hub, offering "as-a-service" models for multinational clients focused on deployment and application rather than training. By leaning into inference-based workloads and compliance-driven innovation, Malaysian data centers and

cloud providers can attract clients seeking safe, regulated, and affordable environments. Moreover, this constraint may catalyze strategic partnerships with Tier-1 nations to bridge capability gaps and ensure Malaysia remains integrated in global AI value chains.

To end this section on a positive note, one area where Malaysia retains full operational freedom is in the use of export-compliant inference chips. These GPUs and accelerators—designed specifically for running AI models rather than training them—are not subject to the same licensing or quota restrictions. This opens a practical and scalable pathway for Malaysia to develop a robust AI-as-a-Service industry. By doubling down on inference infrastructure, local cloud providers and data center operators can offer commercial AI capabilities (like chatbot hosting, vision APIs, and industrial automation solutions) without breaching U.S. export controls. In effect, while Malaysia may face constraints in frontier model development, it is uniquely positioned to lead in downstream deployment, providing a launchpad for economic growth and regional digital leadership.

In short, Malaysia’s ability to grow its AI economy under the current framework depends on how well it can:

- Align with U.S. compliance expectations,
- Invest in trusted infrastructure,
- And pivot toward inference, efficiency, and strategic partnerships.

The following sections explore the tactical steps Malaysia can take to unlock growth despite these restrictions.

Are Tier 2 nations ready for Life 3.0?

We have talked about all the restrictions and potential strategies for working with Tier 2 nations are facing real limitations, especially in building large-scale AI training clusters due to GPU export caps and licensing barriers. These restrictions undoubtedly slow the pace at which these countries can catch up to Tier 1 powers in frontier AI development.

Internationally, the rule sparked debates about the balance between national security and global technological collaboration. Some nations viewed the policy as a unilateral move that could disrupt global supply chains and technological ecosystems. Others saw it as a necessary step to prevent the misuse of advanced technologies.

But there is a lingering issue that no one is really talking about? Are Tier 2 nations truly prepared to responsibly wield Life 3.0 technologies—systems that can learn, adapt, and potentially outpace human capabilities? *We are not going to delve into what AI is or what its main proponents are. That is a topic for another article.*

Are Tier 2 Nations Ready for Life 3.0?

The coming age of self-evolving AI poses deep questions, beyond just hardware



The Power Illusion

- Strategic Reserves: Malaysia's strategic reserves of (as of 2020) stood at 40%; today it's closing in on 30%.
- Grid Reliability: 28% - 30% to affectively remain stable
- Infrastructure Strain: New power plants will have to be planned, and grids will need to be upgraded.



Talent Pipeline Challenges

- Nascent R&D: Malaysia's frontier AI R&D is still developing.
- Brain Drain: 1.86 million Malaysians have left over the past 50 years, representing 5.6% of the population, higher than the global average of 3.6%.
- Lack of Excellence Centers: Without homegrown centers of excellence, we risk becoming AI users, not creators.



Governance Still Catching Up

- Regulatory Gaps: No dedicated AI regulatory body.
- Under-addressed Areas: Critical areas like algorithmic bias, safety, and accountability remain under-addressed.
- Need for Leadership: In the age of Life 3.0, compliance isn't leadership—trusted governance is.



Strategic Influence

- Reactive Stance: Malaysia follows global rules—but rarely shapes them.
- Credibility Through Contribution: True credibility comes from contributing, not just complying.
- Neutrality vs. Influence: Our neutrality has value—but influence demands a voice.



DataGateway Sdn Bhd (202401021123 – 1566972-V)
UOA Business Park, Tower 8, Level 2, Unit 2-1,
No 1 Jalan Pengatucara U1/51A
40150 Shah Alam, Selangor, Malaysia
T + 603 5590 9883, F + 603 5569 3355
Website: <https://areagroup.my>

The Power Illusion

Malaysia's strategic power reserves have declined from 40% (in 2020) to approximately 30% today, barely above the reliability threshold. Grid stability in the range of 28–30% is required to support stable high-compute infrastructure. This creates a fragile foundation for AI scalability. Any future growth in AI clusters will require urgent upgrades to national grids and planning for new power generation capacity. Without energy resilience, ambitions around hosting AI inference or training platforms may falter. It will also explain why the Malaysian government is investing billions in upgrading the national grid in the southern state of Johor. Johor's challenge isn't generation capacity—it's grid bottlenecks. By focusing on upgrading substations and transmission infrastructure (as TNB is doing with its RM90 billion investment), Malaysia can address data center demand without disrupting the national energy strategy. Malaysia is already investing billions into upgrading its grid to incorporate more green energy from Solar. But this takes time, so the illusion that Malaysia is flush with unlimited power is truly an illusion.

Governance Still Catching Up

Malaysia lacks a dedicated AI regulatory authority. Key areas—such as algorithmic safety, ethical review boards, and data accountability—remain underdeveloped. In the era of Life 3.0, compliance is no longer sufficient; leadership in governance is essential. Without a clear institutional home for AI oversight, the country cannot guarantee that deployed systems will align with democratic norms, security standards, or human rights obligations.

Talent Pipeline Challenges

Malaysia's research ecosystem for frontier AI remains nascent. The country suffers from brain drain, with 1.86 million Malaysians having left over the past 50 years, equivalent to 5.6% of the population, a figure notably higher than the global average. More critically, the country lacks globally recognized centers of excellence for advanced technologies such as Artificial Intelligence. Without such institutions, Malaysia risks becoming a perpetual user of imported AI systems, rather than a creator or co-architect of foundational technologies.

Malaysia Strategic Influence

Malaysia continues to operate reactively, aligning with global rules but not shaping them. True credibility comes not from compliance, but from meaningful contribution. While neutrality offers geopolitical flexibility, in AI governance, neutrality without influence results in marginalization. To shape its own future, Malaysia must earn a voice, not just through economic participation, but through leadership in standards-setting, safety, and innovation policy.



DataGateway Sdn Bhd (202401021123 – 1566972-V)
UOA Business Park, Tower 8, Level 2, Unit 2-1,
No 1 Jalan Pengatucara U1/51A
40150 Shah Alam, Selangor, Malaysia
T + 603 5590 9883, F + 603 5569 3355
Website: <https://areagroup.my>

National Security and Strategic Caution

There's also a deeply rooted national security logic behind restraining Tier 2 nations from unrestricted AI development. Frontier AI systems—particularly those capable of self-improvement, autonomy, or advanced deception—pose existential risks, not just technical challenges. The U.S. and Tier 1 allies are still grappling with the implications of Life 3.0 technologies. If even the pioneers don't fully understand the systems they're building, what happens when countries without established guardrails, research maturity, or institutional oversight start developing their own?

From a Western perspective, the emphasis lies on Tier 2 nations' lack of a long-standing contribution to the global scientific and technological commons. Many have minimal involvement in the ethical, safety, or theoretical development of AI, and few possess credible institutions capable of safely governing systems that could rival state-level capabilities.

From a Western perspective, it's not just about capability—it's about trust, maturity, and track record.

If AI is indeed the next nuclear moment—capable of transforming economies, militaries, and societies—then it is reasonable, from a national security standpoint, to argue that access must be conditional, not universal. Especially when:

- The creators (Tier 1 nations) don't yet know how it works fully,
- The potential misuse scenarios include mass surveillance, autonomous weapons, or destabilization,
- And the recipient nations often lack the institutional capacity to safely manage such power.

This doesn't mean Tier 2 countries should be permanently sidelined—but access must come with oversight, transparency, and shared responsibility.



Pathways, Innovation and Strategic Positioning - Malaysia

Pathway to Compliance and Growth in line with the Framework

Despite the hardware restrictions imposed under the U.S. AI Diffusion Framework, Malaysia retains several viable pathways to expand its AI compute capacity while remaining fully compliant with export control laws. These mechanisms, if effectively leveraged and institutionalized, can support Malaysia’s evolution from a Tier 2 deployment zone into a trusted, rules-based digital partner, potentially paving the way toward Tier 1 recognition in the future.

This section outlines the practical strategies Malaysia can adopt to scale its AI infrastructure under current regulations, while reinforcing its commitment to transparency, reliability, and responsible innovation.

Small-Shipment Exemption

Malaysia can import shipments of up to ~1,700 GPU equivalents without a license. These exemptions typically support academic institutions, healthcare providers, and small-scale research labs, enabling early-stage AI experimentation without regulatory burden [5].

Advantages:

- Enables experimental/R&D readiness across education and healthcare sectors.

Universal VEU (UVEU)

Tier 1-headquartered U.S. companies—such as Amazon Web Services, Google, and Microsoft—can operate under “Universal Validated End User” status. This allows them to deploy AI clusters in Tier 2 countries like Malaysia, provided they comply with the global 50% retention and 7% per-country limits [1]. These deployments do not count against Malaysia’s national GPU cap if the equipment remains under U.S. entity control, offering a low-risk channel for cloud growth.

Advantages:

- Offers a low-risk channel for Malaysia to attract cloud investments without breaching its national computing cap.
- Supports the expansion of inference services tied to global platforms.

National VEU (NVEU)

Malaysian-based data center operators can apply for “National VEU” status to unlock



DataGateway Sdn Bhd (202401021123 – 1566972-V)
UOA Business Park, Tower 8, Level 2, Unit 2-1,
No 1 Jalan Pengatucara U1/51A
40150 Shah Alam, Selangor, Malaysia
T + 603 5590 9883, F + 603 5569 3355
Website: <https://areagroup.my>

significantly higher import quotas—up to ~320,000 GPUs over two years[5]. However, this route demands rigorous vetting, including cybersecurity controls, ownership transparency, and active oversight by Malaysian regulators.

Advantages:

- Provides a framework for long-term, sovereign AI capacity building.
- Encourages institutional maturity and national regulatory alignment.

Legacy and Export-Compliant Hardware

Malaysia can accelerate its AI development by pivoting toward export-compliant chips—such as older-generation GPUs or specialized inference accelerators—that fall outside ECCN 3A090 restrictions. These components are freely exportable under existing rules and are ideal for inference-heavy workloads, enabling Malaysia to build commercial AI services without violating U.S. export controls.

Advantages:

- Supports rapid deployment of AI applications without breaching licensing rules.
- Aligns with Malaysia’s positioning as a regional hub for inference services.

Government-to-Government (G2G) Alignment

Malaysia can negotiate higher chip quotas or expedited licensing by enhancing its export-control safeguards. This includes adopting U.S.-aligned end-user screening, real-time chip tracking, and cybersecurity audits. Countries that demonstrate robust compliance may be rewarded with expanded access to advanced AI hardware [4].

Advantages:

- Enhances Malaysia’s case for higher national quotas and strategic exemptions.
- Builds long-term trust with U.S. and multilateral regulatory agencies.

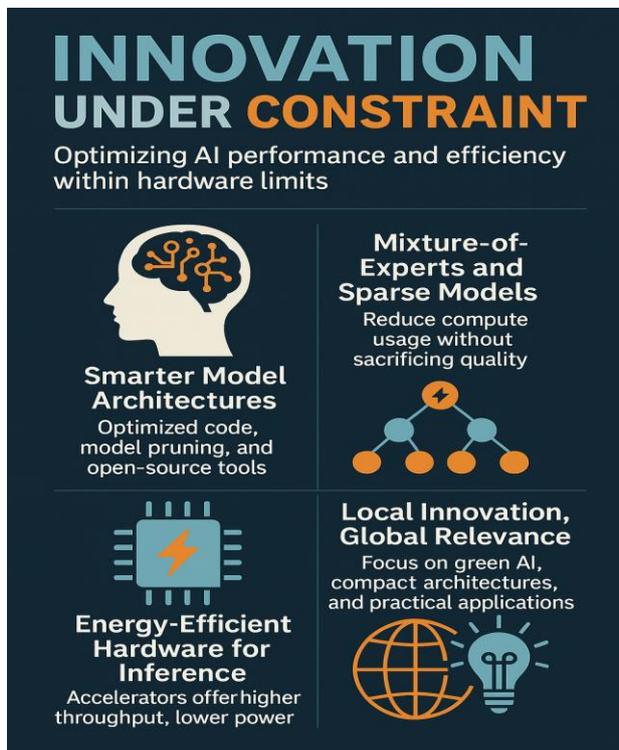
In Practice:



Malaysian firms are already navigating these channels. For instance, YTL Power’s AI Cloud partnered with NVIDIA to host DGX Cloud systems locally, leveraging NVIDIA’s UVEU status to legally operate beyond the national GPU cap. Going forward, Malaysia is likely to pursue a hybrid strategy: blending small institutional imports, selective VEU participation, and investment in inference-based cloud services.

Innovation Under Constraint

Malaysia’s chip import cap may seem like a barrier, but in reality, it’s fuelling a wave of global innovation aimed at doing more with less. For investors and policymakers, this moment offers a chance to back technologies that prioritize efficiency, creativity, and compliance over brute computing force.



INNOVATION UNDER CONSTRAINT
 Optimizing AI performance and efficiency within hardware limits

- Smarter Model Architectures**
 Optimized code, model pruning, and open-source tools
- Mixture-of-Experts and Sparse Models**
 Reduce compute usage without sacrificing quality
- Energy-Efficient Hardware for Inference**
 Accelerators offer higher throughput, lower power
- Local Innovation, Global Relevance**
 Focus on green AI, compact architectures, and practical applications

Smarter Models, Lower Costs
 Startups like *DeepSeek* in China have proven that you don’t need the latest top-tier chips to develop competitive AI. By training their language model using mid-range, export-compliant chips (like the NVIDIA H800), DeepSeek kept costs below USD 6 million, just a fraction of what major U.S. players spend.

This shows that with the right software and architecture, Tier 2 nations like Malaysia can remain competitive without breaching export rules.

Next-Generation AI Architectures
 One standout approach is called Mixture-of-Experts (MoE). Rather than activating all



DataGateway Sdn Bhd (202401021123 – 1566972-V)
UOA Business Park, Tower 8, Level 2, Unit 2-1,
No 1 Jalan Pengatucara U1/51A
40150 Shah Alam, Selangor, Malaysia
T + 603 5590 9883, F + 603 5569 3355
Website: <https://areagroup.my>

parts of an AI model for every task, MoE selectively uses only the most relevant “experts,” dramatically reducing the computing needed.

This kind of architecture can:

- Train faster than traditional dense models
- Deliver strong results with fewer GPUs
- Optimize both costs and energy usage

For Malaysia, adopting MoE models offers a way to stretch chip quotas further and build scalable, compliant AI services.

Energy-Efficient Compute for AI Services

Malaysia can also gain a cost and ESG advantage by embracing hardware that delivers more AI output per watt. Compared to traditional CPU clusters, GPU-based systems consume far less energy for the same task. In fact, hosting AI inference services—like chatbots, search engines, or vision applications—on export-compliant GPUs can:

- Lower energy and cooling bills
- Deliver faster service to end-users
- Qualify for sustainability-linked financing

What This Means for Malaysia

Rather than viewing the export rules as a blockade, Malaysia can position itself as a global leader in efficient, regulation-aligned AI innovation. This will appeal not only to U.S. partners looking for trusted deployment zones, but also to investors searching for scalable, high-return AI platforms that don’t rely on top-tier GPUs.

Key Takeaways for Policymakers and Investors:

- Support R&D into efficient AI models like MoE and open-source architectures
- Channel incentives toward energy-efficient data centers and compliance-ready infrastructure
- Treat export constraints not as limitations, but as a push toward smarter, more sustainable innovation

By focusing on efficiency, compliance, and commercial AI services, Malaysia can thrive under constraint—and carve out a leadership position in the next phase of global AI development.



DataGateway Sdn Bhd (202401021123 – 1566972-V)
UOA Business Park, Tower 8, Level 2, Unit 2-1,
No 1 Jalan Pengatucara U1/51A
40150 Shah Alam, Selangor, Malaysia
T + 603 5590 9883, F + 603 5569 3355
Website: <https://areagroup.my>

Building the Backbone

While global discourse on artificial intelligence remains centered on compute capacity and GPU availability, the true backbone of AI scalability lies in physical infrastructure. The construction of hyperscale data centers depends on vast quantities of highly specialized hardware—often overlooked but fundamentally essential to ensure reliability, uptime, energy efficiency, and future scalability.

As Southeast Asia accelerates its digital transformation, Malaysia is uniquely positioned to emerge as the region’s principal manufacturing and export hub for data center hardware. This opportunity is not merely a matter of capacity—it is a function of strategic neutrality, industrial depth, and regional integration.

Core Industrial Opportunities

Malaysia’s manufacturing ecosystem can supply a wide range of critical data center components, including:

- **Backup Power Systems:** Diesel and natural gas generators, uninterruptible power supplies (UPS), and high-voltage switchgear compliant with Tier III and Tier IV reliability standards.
- **Mechanical and Structural Infrastructure:** Fire-rated pipework, HVAC ducting, containment systems, modular cable trays, server racks, and hot/cold aisle frameworks.
- **Power Distribution and Safety Equipment:** Precision power distribution units (PDUs), grounding and EMI shielding solutions, fire suppression systems, and biometric access control technologies.
- **Prefabricated Modular Assemblies:** Factory-assembled skids, power rooms, and portable data center enclosures designed for rapid on-site deployment.
- **Competitive Advantage Through Strategic Positioning**
- **Malaysia offers a distinctive combination of cost efficiency, geopolitical neutrality, and export-readiness:**
- **Neutral Supply Chain Partner:** Unlike Chinese manufacturing bases, Malaysia is not subject to global export sanctions or restrictions, making it a trusted origin for critical infrastructure goods.
- **Cost-Competitive Alternative:** Compared to U.S. and European production centers, Malaysia offers significantly lower fabrication and labor costs without compromising on quality.

- **Industrial Maturity:** With deep local supply chains in steel, electronics, and electrical components, Malaysia is well-equipped to support end-to-end manufacturing of complex assemblies.
- **Trade and Logistics Integration:** Benefitting from multiple free trade agreements and advanced port infrastructure, Malaysia offers seamless access to regional and global markets.

Strategic Benefits of Domestic Component Manufacturing

Investing in this segment offers long-term value for the nation and for international partners:

- Reduces Dependency on Single-Source Imports from jurisdictions facing export restrictions
- Builds Resilient, Compliance-Ready Supply Chains aligned with global hyperscaler standards
- Generates High-Value Industrial Employment in engineering, precision manufacturing, and logistics
- Catalyzes Regional Growth by anchoring new industrial corridors adjacent to digital infrastructure zones such as Johor, Selangor, and Penang

As the global data center model shifts toward modularity, speed-to-deploy, and local customization, the need for scalable, region-specific component production will only increase. By aligning domestic standards to global procurement frameworks and instituting robust quality control systems, Malaysia can expand its role in the AI ecosystem—from a host of digital infrastructure to a builder of its physical foundation.

To conclude, Malaysia does not need to dominate in AI model training or GPU design to lead in the AI economy. Its strategic opportunity lies in becoming indispensable to the physical layer of AI infrastructure. By manufacturing the hardware that powers, cools, and protects AI compute environments, Malaysia can secure a durable leadership position within the global digital supply chain.

From Market of Convenience to Partner of Choice

Yet Malaysia's path to becoming a true AI hub cannot rest on its geopolitical position alone. While hyperscaler investments and diplomatic neutrality provide a strong foundation, these advantages must be reinforced through deliberate planning, infrastructure readiness, and strategic governance.



DataGateway Sdn Bhd (202401021123 – 1566972-V)
UOA Business Park, Tower 8, Level 2, Unit 2-1,
No 1 Jalan Pengatucara U1/51A
40150 Shah Alam, Selangor, Malaysia
T + 603 5590 9883, F + 603 5569 3355
Website: <https://areagroup.my>

Johor’s rise as a data center hotspot is instructive. Its success stemmed less from central planning and more from geographic fortune—proximity to Singapore, ample land, and early commercial interest. But this organic growth has outpaced existing infrastructure. Today, Johor faces acute strain on water and energy systems, prompting urgent multi-billion-ringgit investments to upgrade basic utilities in support of its Special Border Economic Zone (SBEZ). This underscores a critical point: success without systems is not sustainable.

To avoid repeating this strain and to credibly position Malaysia as a future Tier 1 AI nation, a national roadmap is required—one that outlines not just ambition, but capability. The following ingredients are essential to build a resilient, attractive AI hub:

- Infrastructure Readiness: Develop pre-qualified AI zones with robust access to power, water, and fiber; ensure redundancy and ESG-aligned energy sourcing.
- Site Selection Intelligence: Its important to identify alternative areas with the right infrastructure in place already in place instead of pure green field locations in areas no one has ever heard about costing multiples more.
- Regulatory Trust: Enforce U.S.-aligned export controls, provide clear VEU/NVEU pathways, and establish a dedicated AI governance body.
- Industrial Integration: Support local manufacturing of AI data center components—generators, racks, PDUs, pipework—through incentives and tech parks.
- Talent & Compliance: Build workforce pipelines for AI and data center operations; certify sites for third-party compliance and monitoring.
- Geopolitical Strategy: Maintain commercial flexibility with China while aligning national standards with U.S. security frameworks.

In doing so, Malaysia can move from being a beneficiary of tech decoupling to becoming a strategic enabler of inference-scale AI—especially in areas such as multilingual models, edge inference, and cost-efficient hosting for global platforms. Kuala Lumpur, in particular, could evolve into a neutral AI operations capital for Southeast Asia—a role that balances innovation, compliance, and diplomacy.

Malaysia’s story must now shift from “a market of convenience” to “a partner of choice.” With the right national strategy, it can anchor regional AI deployment, earn greater U.S. trust, and build a resilient digital economy that transcends the current Tier 2 label.

Strategic Risks and Tactical Openings

Managing these changes entails both challenges and strategic openings. The table below summarizes key risks and complementary opportunities for Malaysia under the AI export framework:

Risks / Challenges	Opportunities / Strategies
Capped AI Training Capacity: The 50,000-GPU limit (without VEU) restricts raw computing power for local training.	Surging FDI in Cloud Infrastructure: U.S. tech giants are investing heavily in Malaysia’s cloud (AWS \$6.2B, Google/MS ~\$2B each) [9] to service ASEAN demand. Malaysia can capture this by offering secure hosting and talent.
Smuggling and Compliance Pressure: U.S. may expect Malaysia to crack down on illicit chip exports (recently it urged tighter oversight) [3].	Trusted Digital Hub: By enforcing export controls and vetting, Malaysia can market itself as a secure AI center [10]. This reinforces U.S. confidence and can lead to expanded access (e.g. joint US-Malaysia R&D projects).
Balancing US–China Ties: Aligning with U.S. could cool some Chinese collaborations.	Enhanced Global Partnerships: Leveraging U.S. and allied support (e.g. tech standards, training) can improve Malaysian AI capabilities. Middle-power diplomacy allows participation in multiple ecosystems.
Infrastructure Strain: Meeting the GPU cap (~35 MW IT load for 50k GPUs) demands power/cooling investment.	Focus on Inference & Niche AI: Emphasizing AI inference services (lower power per chip [8]) or specialized domains (edge AI, industrial) can deliver value without massive training rigs.

In summary, Malaysia’s task is to minimize the impact of restrictions (risks) while maximizing new advantages (opportunities). Strict export compliance is burdensome, but doing so unlocks U.S. trust and investment. The 50k cap limits local training, but it also forces a shift toward efficient AI models, turnkey AI services, and deep integration with global cloud platforms. By championing secure, innovative AI deployment, Malaysia can turn a regulatory challenge into a launching pad for its AI ecosystem [4] [10].

Conclusions: From Compliance to a Competitive Edge

Malaysia's Tier 2 designation under the U.S. AI Diffusion Framework places it in a position of both limitation and latent opportunity. While restricted from engaging in frontier model training without explicit licenses, Malaysia is not locked out of the AI race. Instead, it holds a distinct advantage in hosting AI inference operations, deploying export-compliant hardware, and acting as a neutral bridge in the bifurcated tech landscape.

The path forward is clear: Malaysia must lead through trust, infrastructure, and specialization.

By investing in export-compliant data center infrastructure, enforcing robust compliance measures, and leveraging Validated End User (VEU) pathways, Malaysia can signal to global stakeholders, especially the U.S., that it is a responsible digital partner. The rise of inference-based AI workloads, which are forecasted to dominate global compute demand, aligns perfectly with Malaysia's strengths: access to skilled labour, cost-effective energy, and geopolitical neutrality.

At the same time, Malaysia must address critical readiness gaps: governance structures must mature, grid stability must be enhanced, and its talent pipeline must deepen to support sustained AI service deployment. If these foundational issues are tackled head-on, the country can shift from being seen as a "market of convenience" to a partner of choice.

Malaysia does not need to train the next GPT-6 to lead in AI. Instead, it can host the infrastructure, manufacture the physical backbone, and power the edge applications that make global AI run. In doing so, Malaysia will not merely adapt to the AI future—it will help define its operational reality.

With strategic foresight, institutional discipline, and diplomatic balance, Malaysia can transform AI export constraints into a launchpad for long-term digital leadership in a Life 3.0 world.

References

Number	Reference
[1]	Bureau of Industry and Security 15 CFR Parts 732, 734, 740, 742, 744, 748, 750, 762, 772, and 774 [Docket No. 250107–0007] RIN 0694–AJ90
[2]	CHIPS-Science-Act-Fact-Sheet – Semiconductor Industry Association - https://www.semiconductors.org/wp-content/uploads/2023/08/CHIPS-Science-Act-Fact-Sheet.pdf
[3]	https://www.tomshardware.com/tech-industry/artificial-intelligence/massive-366-percent-chip-shipment-surge-to-malaysia-amid-increased-nvidia-ai-gpu-smuggling-curbs-ahead-of-looming-sectoral-tariffs#:~:text=on%20HS%20codes%2C%20it%20is,are%20being%20funneled%20to%20China
[4]	https://www.justsecurity.org/106545/the-future-of-the-ai-diffusion-framework/#:~:text=The%20real%20innovation%20lies%20in,incloding%20cybersecurity%20standards%2C%20personnel%20vetting
[5]	https://bidenwhitehouse.archives.gov/briefing-room/statements-releases/2025/01/13/fact-sheet-ensuring-u-s-security-and-economic-strength-in-the-age-of-artificial-intelligence/#:~:text=,technology%20around%20the%20world
[6]	https://huggingface.co/blog/moe#:~:text=Mixture%20of%20Experts%20enable%20models,counterpart%20much%20faster%20during%20pretraining
[7]	https://www.weforum.org/stories/2025/02/open-source-ai-innovation-deepseek/#:~:text=Another%20key%20differentiator%20of%20DeepSeek,of%20January%20as%20a%20result
[8]	https://blogs.nvidia.com/blog/energy-efficient-ai-industries/#:~:text=ran%2C%20as%20well%20as%20how,of%20the%20world%E2%80%99s%20largest%20supercomputers
[9]	https://www.ciodive.com/news/aws-asia-pacific-hyperscale-cloud-region-malaysia-microsoft-google/725017/#:~:text=,2%20billion%2C%20respectively%2C%20in%20May
[10]	https://www.mida.gov.my/mida-news/malaysia-can-be-middle-power-in-global-supply-chain-amid-us-china-tensions/#:~:text=Malaysia%2C%20as%20a%20middle%20power%2C,RE%29%20sectors
[11]	https://www.reuters.com/technology/malaysia-tighten-semiconductor-regulations-under-us-pressure-ft-reports-2025-03-24/
[12]	https://www.cnbc.com/2025/03/03/nvidia-unofficial-exports-to-china-face-scrutiny-after-singapore-arrests.html#:~:text=Singapore%20has%20been%20working%20to,Nvidia's%20highly%20sought%20Dafter%20chips.
[13]	https://publicinfo.trade.gov.tw/cuswebo/FSCE3000C?table=FSCE3010F
[14]	https://theedgemalaysia.com/node/744625



DataGateway Sdn Bhd (202401021123 – 1566972-V)
UOA Business Park, Tower 8, Level 2, Unit 2-1,
No 1 Jalan Pengatucara U1/51A
40150 Shah Alam, Selangor, Malaysia
T + 603 5590 9883, F + 603 5569 3355
Website: <https://areagroup.my>

Number	Reference
[16]	<i>Press Release 13 Jan 2025: Biden-Harris Administration Announces Regulatory Framework for the Responsible Diffusion of Advanced Artificial Intelligence Technology New Framework Advances AI Innovation While Protecting U.S. National Security</i>
[17]	<i>Center for Strategic and International Studies. Understanding the Biden Administration's Updated Export Controls by Gregory C Allen December 2024. Image from: https://x.com/Gregory_C_Allen/status/1714662722338504829</i>