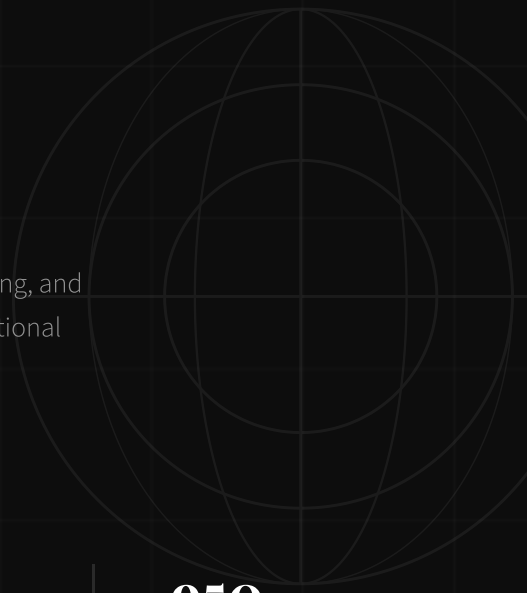
GRAVITYWELL RESEARCH

INTELLIGENCE REPORT
MAY 2026 — GWR-2026-04
INSTITUTIONAL DISTRIBUTION

GEOPOLITICS & TECHNOLOGY STRATEGY

The Sovereign *AI Race*

A global intelligence and investment report on how nations are building, financing, and competing for AI independence — across compute, semiconductors, energy, national models, and strategic capital.



$1T+
GLOBAL AI INFRA CAPEX CYCLE

40+
NATIONS WITH SOVEREIGN AI PLANS

3
COMPETING AI BLOCS

950
TWH DATA CENTRE POWER BY 2030

GRAVITYWELL RESEARCH · LONDON · SINGAPORE · NEW YORK

MAY 2026

REPORT NAVIGATION

Report *Contents*

Nineteen chapters across six analytical parts — from the global landscape and infrastructure stack through demand economics, country deep-dives, governance, and synthesis. All figures as of May 27, 2026 unless noted; aspirational commitments flagged (target) or (planned).

EXECUTIVE SUMMARY 01

The Sovereign AI Race: *A $1 Trillion Capital Cycle*

TL;DR — THREE SIGNAL FINDINGS

1. Sovereign AI is now strategic infrastructure. More than 40 nations have enacted or announced binding sovereign AI strategies as of May 2026. AI compute, energy, and semiconductor access are treated alongside defence and energy in national security doctrine.

2. Capital is concentrating asymmetrically. The five largest technology companies spent **>$400bn in capex in 2025**, set to rise a further 75% in 2026 — exceeding global oil & gas upstream capex. The Middle East and East Asia account for >80% of all publicly disclosed sovereign AI investment.

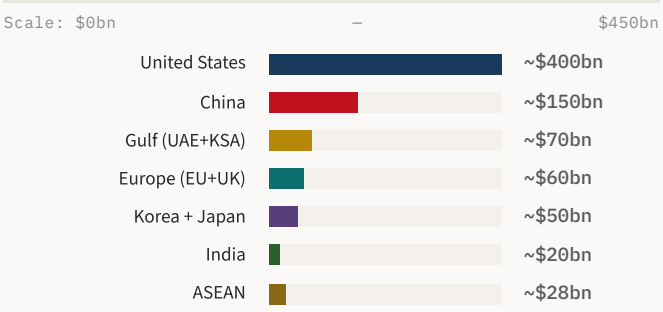
3. A power–compute–semiconductor trilemma is binding. AI-focused data-centre electricity grew 50% in 2025. Hyperscalers are signing nuclear PPAs of multi-GW scale. GPU supply chains through TSMC CoWoS and SK Hynix HBM remain the principal chokepoints of geopolitical competition.

$500bn	€200bn	$53bn	950TWh
STARGATE US 4-YEAR TARGET	EU INVESTAI MOBILISATION	ALIBABA 3-YEAR AI/CLOUD PLAN	GLOBAL DC POWER DEMAND BY 2030
↑ 7 GW planned	↑ 76 site applications	↑ vs. prior decade combined	↑ 2× 2025 baseline

THE THREE BLOCS

1. **U.S.-Led Alliance Stack:** Stargate (US/Japan), AUKUS, EU AI Factories, Korea/Japan/Taiwan, UAE & Saudi Arabia under U.S. operator covenants. Anchored by NVIDIA Blackwell, TSMC Arizona, and the OpenAI–Microsoft–Oracle ecosystem. → NATIONAL P.04
2. **Chinese Self-Reliance Stack:** Huawei Ascend, Alibaba/Baidu/Tencent, DeepSeek open-weight models, SMIC/YMTC/CXMT. Mandated domestic chip procurement in state data centres (Nov 2025). Target: 300 EFLOP/s national compute by end-2025. → CHINA P.14
3. **Non-Aligned / Hedging:** India, Brazil, ASEAN multi-vendor strategies. Seek the lowest cost, highest optionality approach — adopting open-weight models, leveraging both U.S. and Chinese infrastructure where permitted. → INDIA P.18 · AFRICA/LATAM P.19

AI INFRASTRUCTURE CAPEX BY REGION (2025E, US$BN)



Source: Gravitywell Research estimates, company disclosures, IEA (Apr 2026)

WHY SOVEREIGNTY HAS BECOME NON-NEGOTIABLE

AI infrastructure now meets every test of strategic national infrastructure: **capital-intensive** (5-yr hyperscaler capex > $1T), **physically constrained** (HBM, EUV, gigawatt power), **dual-use** (civilian + defence), and **monopolizable** — compute concentration analogous to 19th-century railroad monopolies. McKinsey (Dec 2025): **71% of executives & government officials** characterise sovereign AI as "existential" or a "strategic imperative."

FOR TIME-CONSTRAINED READERS 02

Investor Quick Read:
Five Conviction Calls · Five Risks

HEADLINE THESIS

Sovereign AI is a **$1.1–2.3T TAM by 2030** reshaping geopolitics. Capital concentrates in three blocs (US-led / Chinese self-reliance / non-aligned hedgers). The binding constraint is no longer money — it is **power, HBM, and EUV**. Underwrite by control plane, not data-residency marketing.

→ EXEC P.01 · RISK REGISTER P.22

FIVE CONVICTION CALLS

1 **Energy > Compute as 2027 chokepoint.** HIGH Goldman: 85–90 GW of new nuclear needed by 2030; <10% will be available. Buy Constellation, Vistra, GE Vernova, Talen, Cameco. → P.09

2 **HBM is sold out through 2027.** HIGH SK Hynix 62% share, NVIDIA 90% of its output. Most undervalued exposure: SK Hynix at 8× EV/EBITDA. → P.10

3 **TSMC at 18× EV/EBITDA** HIGH is the highest-margin monopoly in the stack — 90% advanced logic, $165bn US commitment, growing 30% in 2026. Cheapest sovereign-AI proxy. → P.22

4 **India is application-layer sovereign** MED — not frontier. 38K+ GPUs deployed, DPI flywheel unmatched, fab buildout 2026–28. Play: Tata, Reliance, IndiaAI infra; avoid frontier-model bets. → P.18

5 **Defence AI is the breakout vertical.** HIGH Pentagon $13.4bn FY2026 AI line item (first ever). Anduril $61bn, Palantir $250bn+, Shield AI $12.7bn, Helsing €18bn. → P.29

FIVE MATERIAL RISKS

R1 **Counterparty concentration.** HIGH Big Four hyperscaler 2026 capex ($725bn) underwritten by ~$1.05T backlog from **two counterparties** (OpenAI + Anthropic), both deeply FCF-negative. → P.21

R2 **2027–28 US grid bottleneck.** HIGH Project pipelines exceed interconnection capacity by 2–3 years. Asset stranding risk on Texas / PJM merchant DC builds. → P.09

R3 **Taiwan single-point-of-failure.** MED TSMC = 90% advanced logic; ASML EUV = 100%. Any disruption = 12–18 month global AI capex pause. → P.08

R4 **Sovereignty theatre.** HIGH "Hosted Sovereign" deployments collapse under US Cloud Act subpoena or chip export shock. Underwrite only FULL or PARTIAL grade. → P.06

R5 **DeepSeek-style efficiency shocks.** MED Frontier training cost compression undermines capex-intensity thesis. Bear-case TAM $1.1T vs base $1.7T = ~$600bn at risk. → P.22

BY STAKEHOLDER · 30-SECOND POSTURE

Allocators / SWF

Overweight power (nuclear, regulated utilities) + HBM/CoWoS supply chain + Defence AI primes. Underweight pure-play Neoclouds with negative FCF. → STAKEHOLDER P.24

Enterprise / CIO

Tier workloads by sovereignty grade. Lock 5-yr nuclear PPAs. Dual-stack CUDA + ROCm through 2028. Treat HBM as 2021-shortage-class risk. → TCO P.14

Policymakers

Reform grid interconnection within 24 months. Mandate open-weight national-language models. Co-invest via EIB/development banks to absorb 10–20yr payback. → P.24

REFERENCE KEY 03

Glossary & *Acronyms*

Quick reference for technical terms used throughout the report. Listed alphabetically.

HARDWARE & STACK

Anke — China's security framework certifying domestic AI chips for state procurement (May 2026).

ASIC — Application-Specific Integrated Circuit. Custom silicon (e.g. Google TPU, Meta MTIA).

ATMP — Assembly, Test, Mark & Pack. Backend semi step.

BF16 — Bfloat16. 16-bit floating-point format for AI training.

CoWoS — Chip-on-Wafer-on-Substrate. TSMC advanced packaging — the AI-chip bottleneck.

CUDA — NVIDIA's proprietary GPU compute platform; the software-layer sovereignty veto.

DLC — Direct Liquid Cooling. Replaces water-intensive evaporative cooling.

EUV — Extreme Ultraviolet lithography. ASML monopoly (100% share).

HBM — High Bandwidth Memory (HBM2E/3/3E/4). SK Hynix + Samsung control 90%+.

OSAT — Outsourced Semiconductor Assembly and Test.

ROCm — AMD's open CUDA-alternative compute stack (v7 production-grade for inference).

SMR — Small Modular Reactor. Hyperscaler nuclear bet (Kairos, X-energy, NuScale).

ECONOMICS & WORKLOADS

A2A — Agent-to-Agent protocol. Foundation for agentic AI orchestration.

CAGR — Compound Annual Growth Rate.

DDTL — Delayed Draw Term Loan. CoreWeave-style GPU-backed financing.

EBITDA — Earnings Before Interest, Tax, Depreciation, Amortisation.

EV / EBITDA — Enterprise Value over EBITDA. Multi-stage valuation metric.

IaaS — Infrastructure-as-a-Service. The base layer of cloud revenue.

LCOI — Levelised Cost of Inference. Per-token economics over asset life.

MaaS — Model-as-a-Service. Premium AI revenue layer (60–70% gross margin).

MCP — Model Context Protocol. Agentic-AI standard for tool integration.

PPA — Power Purchase Agreement (long-term electricity contract).

PUE / WUE — Power / Water Usage Effectiveness. DC efficiency metrics.

TCO — Total Cost of Ownership. Multi-year cloud-vs-on-prem economic comparison.

SOVEREIGNTY & POLICY

Cloud Act — US law compelling US-domiciled cloud providers to surrender data wherever stored.

CISPE — Cloud Infrastructure Services Providers in Europe. Defines binary sovereign vs. not.

CSRD — EU Corporate Sustainability Reporting Directive. Mandates AI-related ESG disclosure.

DPDP — India's Digital Personal Data Protection Act 2023.

DPI — Digital Public Infrastructure (e.g. Aadhaar, UPI, ONDC).

EO 14365 — US Executive Order (Dec 2025) on AI infrastructure preemption.

GPAI — General-Purpose AI. EU AI Act regulatory classification.

ML-KEM — NIST post-quantum key encapsulation standard (FIPS 203).

MCF — Military–Civil Fusion. China's dual-use AI doctrine.

Sovereign Grades — FULL / PARTIAL / HOSTED / VASSAL framework (see Sovereignty Spectrum p.06).

SWF — Sovereign Wealth Fund (PIF, MGX, GIC, Mubadala).

TPU — Tensor Processing Unit. Google's custom AI accelerator.

PART I · THE LANDSCAPE

SOVEREIGNTY FRAMEWORK 04

The Sovereignty Spectrum: *What Does "Sovereign" Actually Mean?*

CISPE's February 2026 position paper put it bluntly: a cloud is either sovereign or it is not — there is no "75% sovereign." Yet most national AI programmes conflate hardware residency with actual operational control. We propose a four-tier grading framework, based on five layers of sovereignty: hardware supply chain, software stack, operator nationality, legal jurisdiction, and model provenance.

Grade	Classification	Definition	Exemplars	Actual Control Level
FULL	**Full-Stack Sovereign**	Domestic chips + domestic foundry + domestic software + domestic operator + domestic model + domestic legal jurisdiction	China (Huawei Ascend + SMIC + DeepSeek + Alibaba Cloud)	Complete independence; immune to foreign sanctions
PARTIAL	**Infrastructure Sovereign**	Foreign chips but domestic operator, domestic legal jurisdiction, domestic model, no foreign access rights	France (Mistral + Scaleway + NVIDIA chips); Korea (Naver Cloud + NVIDIA); Israel (Nebius/National SC)	High operational control; vulnerable to chip export restrictions
HOSTED	**Hosted Sovereignty**	Foreign operator with domestic legal wrappers; data residency but foreign control plane	Germany (Bleu = Microsoft + Capgemini/Orange); S3NS (Google + Thales); AWS European Sovereign Cloud (Brandenburg)	Data residency achieved; operational control shared; subject to US Cloud Act conflict
VASSAL	**Vassal Architecture**	Foreign chips + foreign operator + foreign model; "sovereign" branding with US operator covenants and dollar-for-dollar matching requirements	UAE (Stargate UAE = G42/OpenAI/Oracle under US supervision); Saudi (HUMAIN = NVIDIA/AWS/xAI operators)	Infrastructure residency only; operational control substantially foreign; US veto on sensitive workloads

> *"Even when operated inside Europe by European employees, those services are still under the American legislation, under the Cloud Act."*
>
> — ULRICH AHLE, CEO GAIA-X, GAIA-X SUMMIT PORTO, NOV 2025

Gaia-X: A Cautionary Tale in Sovereignty by Committee

Gaia-X launched in 2019 as a Franco-German initiative to build federated European cloud infrastructure. Six years later, Forrester's assessment is damning: "Gaia-X has failed to launch meaningful public cloud and data services" beyond limited proofs of concept. The EuroStack Project is even blunter, calling it "a chronicle of a failure foretold."

Root causes: Too many goals and too many masters. Ambiguity between reducing US dependency, ensuring GDPR compliance, and subsidizing European SMEs. Allowing US hyperscalers (Microsoft, Google, Palantir) as members created internal contradictions. Analysis paralysis and bureaucratic complexity delayed every milestone.

Key lessons for sovereign AI programmes:

1. **Sovereignty requires picking a lane.** You can have US hyperscaler performance OR full operational sovereignty. Claiming both is false advertising — PwC Netherlands (March 2026) found that every step deeper into cloud maturity is simultaneously a step deeper into dependency.
2. **Market share loss is real.** Gaia-X's own leadership acknowledges European cloud providers lost three-quarters of their market share during the six years the initiative was being "built."
3. **Execution > architecture.** IDC forecasts that by 2028, 60% of multinationals will split AI stacks across sovereign zones, tripling integration costs. The cost of sovereignty is real: it's not free insurance.
4. **Gartner projects $80bn in sovereign cloud IaaS spending in 2026** (+35.6% YoY), with China the largest spender ($47bn), followed by the US ($16bn), then Europe ($13bn, passing US in 2027).

INVESTOR IMPLICATION

Only **FULL** and **PARTIAL** grades survive a US Cloud Act subpoena or chip-export shock. HOSTED and VASSAL deployments are valuation-cliff risks the moment Washington–Beijing tension flips a switch. Underwrite sovereignty by control plane, not by data-residency marketing.

→ CHINA P.14 · UAE/SAUDI P.04-03

PART I · THE LANDSCAPE

COMPETITIVE LANDSCAPE 05

Sovereign AI Readiness Index: *Who Leads, Who Lags?*

Scores across 10 dimensions: Compute, Energy, Semiconductor Access, Talent, Research, Capital, Regulatory Flexibility, Domestic Stack, Military Integration, Infrastructure Readiness. Scale: 1 (weakest) → 5 (strongest). Maximum 50 points.

METHODOLOGY: GRAVITYWELL RESEARCH PROPRIETARY SCORING, MAY 2026

COUNTRY	COMPUTE	ENERGY	SEMIS	TALENT	RESEARCH	CAPITAL	REG.	DOM. STACK	MIL. AI	INFRA	TOTAL /50
United States	5	4	5	5	5	5	4	5	5	4	47 #1
China	4	5	2	5	5	4	5	5	5	5	45 #2
South Korea	4	4	5	4	4	4	4	4	4	5	42 #3
UAE	4	5	3	3	3	5	5	3	4	4	39
Israel	3	2	3	5	5	4	5	4	5	3	39
Japan	3	3	5	4	4	4	4	4	3	4	38
Saudi Arabia	4	5	3	2	2	5	5	3	4	4	37
Taiwan	3	3	5	4	4	3	4	4	3	4	37
France	3	4	3	4	4	4	3	4	4	3	36
Singapore	3	2	3	4	4	5	4	3	3	5	36
United Kingdom	3	2	3	5	5	3	4	3	4	3	35
Germany	3	2	3	4	4	4	3	4	3	3	33
Canada	3	5	2	4	4	3	4	3	3	3	34
India	3	3	2	4	3	3	4	4	3	3	32

SCORE LEGEND: 5 = Frontier | 4 = Strong | 3 = Developing | 2 = Emerging | 1 = Nascent

Reg. = Regulatory flexibility · Dom. Stack = Domestic software/chip stack · Mil. AI = Military integration

PART I · THE LANDSCAPE

COUNTRY ANALYSIS 06

National Initiatives: *The Race is On*

COUNTRY	INITIATIVE / PROGRAMME	COMMITMENT	STRATEGIC HIGHLIGHTS	TYPE
United States	Stargate Project	**$500bn / 10 GW**	OpenAI/SoftBank/Oracle JV. 7 GW committed, 5 sites. TSMC Arizona $165bn expansion. CHIPS Act HVM achieved.	COMPUTE INFRA
China	AI Plus + Big Fund III	**$47bn fund + $70bn DC capex**	Alibaba RMB 380bn AI plan. Foreign chip ban in state DCs (Nov 2025). DeepSeek-R1 opens new compute-efficiency frontier.	INDUSTRIAL POLICY
European Union	InvestAI + AI Gigafactories	**€200bn mobilised**	76 site applications for 4–5 Gigafactories (~100k chips each). EIB co-financing MoU Dec 2025. AI Act GPAI rules Aug 2025.	SOVEREIGN CLOUD + MODEL
France	AI Action Summit + Mistral	**€109bn committed**	ASML €1.3bn stake in Mistral (€11.7bn valuation). Mistral Compute: 18,000 GB200 GPUs, 40 MW. "Third Way" AI strategy.	NATIONAL MODEL + INFRA
Germany	Industrial AI Cloud	**~€1bn + €5.5bn (Google)**	Deutsche Telekom + NVIDIA: 10,000 Blackwell GPUs, 0.5 ExaFLOPS Munich cluster. Google €5.5bn through 2029.	SOVEREIGN CLOUD
United Kingdom	AI Opportunities Action Plan	**£2bn state + £28bn private**	AI Growth Zones: 15,000+ jobs. Isambard-AI live Jul 2025. Sovereign AI Unit (£500mn) launched Apr 2026. AI firms raised £6bn in 2025.	COMPUTE + ECOSYSTEM
UAE	Stargate UAE + MGX	**>$30bn (5 GW campus)**	G42-led 1 GW cluster. NVIDIA GB300, OpenAI operator. MGX: $100bn AUM target, $10bn/yr AI spending. First 200 MW online 2026.	COMPUTE INFRA + SWF
Saudi Arabia	HUMAIN (PIF-backed)	**$23bn partnerships**	600,000 NVIDIA GPUs over 3 yrs. xAI 500 MW campus. AWS AI Zone 150k GPUs. ALLAM Arabic LLM. $10bn HUMAIN Ventures VC fund.	FULL STACK SOVEREIGN

Continued on next page · Source: Gravitywell Research; government & corporate disclosures (May 2026)

COUNTRY ANALYSIS · CONTINUED

07

National Initiatives *(continued)*

↓ CONTINUED FROM P.04

COUNTRY	INITIATIVE / PROGRAMME	COMMITMENT	STRATEGIC HIGHLIGHTS	TYPE
Japan	AI Promotion Act + ¥10T plan	¥10T (~$65bn) by 2030	FY2026 chip/AI budget 4× to ¥1.23T. Sakana AI $135mn round (In-Q-Tel). NTT ¥8T commitment. SoftBank first DGX SuperPOD.	INDUSTRIAL POLICY + MODEL
South Korea	NVIDIA Korea + AI Basic Act	$735bn ecosystem (est.)	260,000 NVIDIA GPUs by 2030. Samsung $230bn semis capex. Jeollanam-do 3 GW DC ($35bn). HBM monopoly: 90% of global supply.	COMPUTE + SEMICONDUCTORS
Taiwan	AI Basic Act + Silicon Valley South	TWD 31.1bn ($1bn) FY26	TAIDE national LLM (70B). TSMC: 534 customers, 305 process nodes. Foxconn-NVIDIA "fastest AI supercomputer." AI Basic Act Jan 2026.	SEMIS + MODEL + POLICY
India	IndiaAI Mission	₹10,372cr (~$1.25bn)	38,000+ GPUs deployed. Tata-PSMC $11bn fab (Dholera). Micron Sanand ATMP open. UPI 14bn txns/month. 100% compute subsidy for foundation models.	COMPUTE + FAB + DPI
Singapore	National AI Strategy 2.0	>S$1bn (5-yr R&D)	National AI Council (PM Wong). S$300mn Quantum Strategy. Green DC Roadmap. Hub for SEA sovereign inference workloads.	RESEARCH + REGULATION
Russia	Putin AI Commission + Yandex	RUB 12bn ($131mn) Yandex	Putin AI Commission Feb 2026. Yandex GPT 5.1 Pro outperforms GPT-4.1 in 56% RU tasks. Rostelecom $196mn DataLine acquisition. No NVIDIA access; sub-scale vs single US hyperscaler DC.	SANCTIONED / DEFENSE
Israel	National AI Directorate	$140mn supercomputer	1,000 NVIDIA B200s live (Nebius). NVIDIA Kiryat Tivon campus (1.7mn sqft). AI21 Labs/NVIDIA M&A ($2–3bn). IDF AI targeting deployment.	DEFENSE + COMPUTE
Malaysia	Digital Ecosystem Acceleration	MYR 285bn (~$69bn)	GPU imports +3,400% YoY (Apr 2025 vs Apr 2023). YTL-NVIDIA $2.36bn: first sovereign LLM + 600 MW Kulai campus. Microsoft $2.2bn.	DC HUB + SOVEREIGN CLOUD

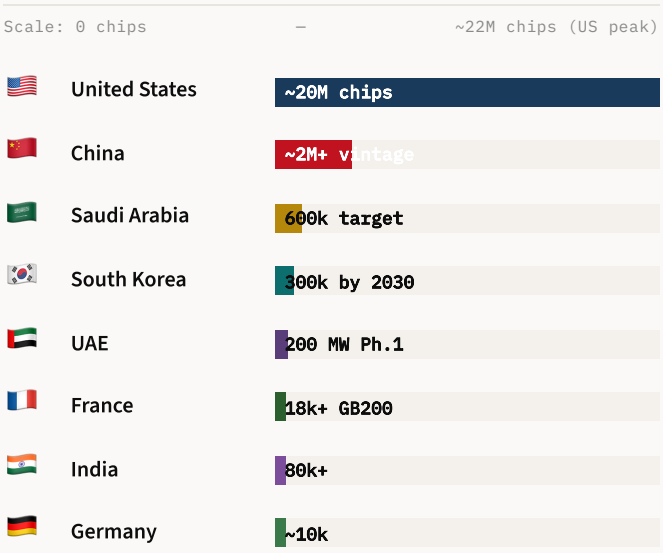
Type categories: ■ Compute · ■ Policy · ■ Cloud/Model · ■ Research · ■ Defense · ■ Full Stack · Source: Gravitywell Research (May 2026)

PART II · THE INFRASTRUCTURE STACK

COMPUTE SOVEREIGNTY & GPU WARS 08

Compute Sovereignty:
GPU Installed Base & Supply-Chain Chokepoints

GPU INSTALLED BASE BY JURISDICTION (LATE 2025 EST.)



Note: US figure from NVIDIA CEO Huang (APEC briefing). Others are Gravitywell Research estimates from public disclosures. Bars represent relative scale, not exact proportions.

> *"The capital expenditure of five large technology companies surged to more than $400bn in 2025 and is set to increase by a further 75% in 2026."*
>
> — IEA, KEY QUESTIONS ON ENERGY AND AI, APRIL 2026

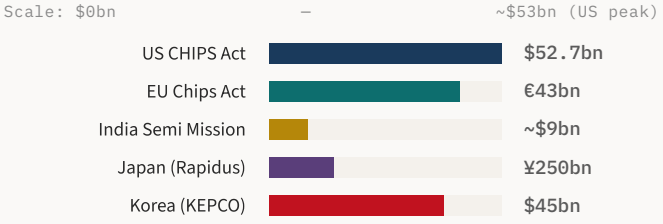
SUPPLY CHAIN CHOKEPOINTS

Sovereignty aspirations collide with physical concentration. Five nodes control the AI compute stack:

LAYER	DOMINANT PLAYER(S)	MARKET SHARE	RISK
Accelerators	NVIDIA	**~80–90%**	CRITICAL
Advanced Logic Fab	TSMC	**~90%**	CRITICAL
EUV Lithography	ASML	**100%**	CRITICAL
HBM Memory	SK Hynix / Samsung	**~90%+**	CRITICAL
Adv. Packaging (CoWoS)	TSMC	**>70%**	HIGH
EDA Software	Cadence / Synopsys	**~80%**	HIGH

Source: SemiAnalysis, company filings, Gravitywell Research (May 2026) *→ HBM DETAIL P.12*

NATIONAL FAB SUBSIDIES



Source: Government budget disclosures, company filings, Gravitywell Research

PIVOTAL DISRUPTION

DeepSeek: The Compute-Efficiency Shock

DeepSeek-V3 reported a training rental cost of ~$5.6mn — and R1's final run at $294k — raising existential questions about the capex intensity of AI. On Jan 27, 2025, NVIDIA's market cap fell $589bn in a single session — the largest single-day loss for any company in history. The incident accelerated efficiency-first sovereign model strategies globally.

PART II · THE INFRASTRUCTURE STACK

ENERGY, POWER & AI INFRASTRUCTURE 09

Energy & Power Infrastructure:
Nuclear, Grid Readiness & the 950 TWh Wall

485TWh	950TWh	50%	9.8GW
GLOBAL DC POWER DEMAND 2025	IEA FORECAST 2030 BASE CASE	AI-FOCUSED DC POWER GROWTH 2025	NUCLEAR CAPACITY CONTRACTED (HYPERSCALERS)
↑ 17% YoY	↑ 2× in 5 yrs	↑ vs. 17% overall	↑ as of May 2026

"Electricity consumption from AI-focused data centres grows much faster than overall data centre electricity consumption, tripling between 2025 and 2030. Data centres now account for half of all new US electricity demand."

— IEA, KEY QUESTIONS ON ENERGY AND AI, APRIL 2026 / FORTUNE, APRIL 2026

THE NUCLEAR RENAISSANCE

Hyperscalers have signed **>9.8 GW of nuclear capacity contracts** since 2024. This represents a structural shift in data-centre power strategy, driven by the need for 24/7 carbon-free baseload at gigawatt scale.

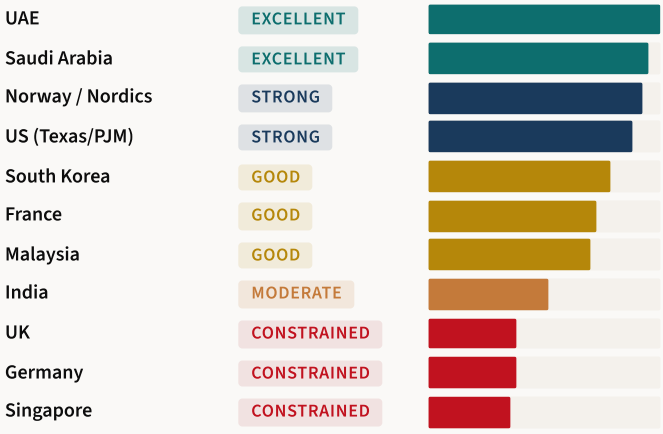
HYPERSCALER	PARTNER / PLANT	CAPACITY	TIMELINE
Microsoft	Constellation (TMI restart)	835 MW	2027–28
Google	Kairos Power (SMR fleet)	500 MW	~2030
Amazon	X-energy (Xe-100 SMRs)	~960 MW	2030+
Amazon	Talen/Susquehanna	1.92 GW PPA	Active
Meta	Multi-vendor RFPs	≤6.6 GW	2026–30
Oracle	3-SMR campus plan	~1 GW	2028+

Source: Company press releases, World Nuclear Association (2025–26)

KEY CONSTRAINT

Goldman Sachs estimates **85–90 GW of new nuclear capacity** would be needed to meet all data-centre power demand growth by 2030 — but **<10% will be available**. Onsite gas + battery storage is the realistic 2026–2030 bridge.

ENERGY READINESS BY JURISDICTION



Anchors: UAE — Barakah nuclear + low-cost gas · Saudi — abundant gas + nuclear pipeline · Nordics — hydro + cold climate · US — deregulated + nuclear restart · S. Korea — nuclear expansion · France — ~70% nuclear · Malaysia — low-cost gas · India — grid reliability gaps · UK — grid bottleneck · Germany — nuclear phaseout · Singapore — land + grid limits.

■ *Excellent* · ■ *Strong* · ■ *Good* · ■ *Moderate* · ■ *Constrained* ·
Source: Gravitywell Research; generation mix & siting trends

OUR HIGHEST-CONVICTION ENERGY CALLS

CONSTELLATION ENERGY | VISTRA CORP | GE VERNOVA | TALEN ENERGY | CAMECO | URANIUM ENERGY | CENTRUS | NEXTERA | DOMINION ENERGY

The 2027–28 US power bottleneck is the single most under-priced risk in AI infrastructure: project pipelines exceed grid interconnection capacity by 2–3 years.

PART II · THE INFRASTRUCTURE STACK

PHYSICAL & SOFTWARE CONSTRAINTS 10

Water, HBM & Software: *The Binding Constraints Beneath Compute*

Water: The Silent Limiter

THE SCALE OF THE PROBLEM

A typical 100 MW AI data centre using evaporative cooling consumes **200–400mn gallons of water annually** — equivalent to the residential use of a small city. Google's data-centre water consumption rose from 4.3bn gallons in 2021 to **6.1bn gallons in 2024**. In Texas alone, data centres are projected to use 49bn gallons in 2025, potentially reaching **399bn gallons by 2030** — equivalent to drawing down Lake Mead by 16 feet annually.

MSCI's November 2025 analysis identifies a direct tension: water-based cooling is more energy-efficient but increases water consumption; air-based cooling conserves water but requires more electricity. Liquid immersion and direct-to-chip cooling reduce direct water use by 70–90% but raise capital costs.

WATER STRESS BY AI HUB

JURISDICTION	WATER STATUS	DOMINANT COOLING	RISK
UAE / Saudi Arabia	Extreme stress	Desalinated + DLC	CRITICAL
India (Gujarat)	High stress	Evaporative	CRITICAL
Texas (US)	Moderate–High	Evaporative	HIGH
Singapore	Import-dependent	Mixed	HIGH
Malaysia (Johor)	Moderate	Evaporative	MEDIUM
Nordics / Canada	Low stress	Air-cooled / hydro	LOW
UK / France	Moderate	Mixed	MEDIUM

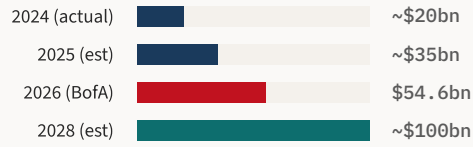
HBM Memory: Sold Out Through 2026, Shortages Through 2027+

WHY HBM IS THE CHOKEPOINT

The HBM supply-demand balance is the tightest chokepoint in the AI stack. BofA estimates the **2026 HBM market at $54.6bn (+58% YoY)**. Goldman Sachs forecasts HBM demand for custom ASICs to surge 82%, accounting for one-third of the market. SK Hynix has **sold out its entire 2026 production capacity**, and Samsung's memory chief warned in April 2026 that "significant shortages" will persist through at least 2027.

HBM now consumes **23% of all DRAM wafer capacity** globally — up from under 5% in 2023. The result: DRAM prices have roughly doubled since early 2025, smartphone shipments are projected to decline 12.9% in 2026, and the PC market faces an 11.3% contraction. Data centres now consume an estimated **70% of all memory chips produced worldwide**.

HBM TIMELINE & MARKET SIZE



HBM4 enters mass production in 2026 (55% of revenue mix). HBM3E prices hiked ~20% for 2026. HBM4E development targeted for completion H1 2026, targeting ~40% of 2027 demand. SK Hynix dominates with 62% market share; NVIDIA accounts for ~90% of SK Hynix's HBM supply.

Source: BofA, Goldman Sachs, TrendForce, SK Hynix & Samsung guidance (2026)

Software Stack: CUDA Lock-in as Sovereignty Risk

NVIDIA's CUDA platform creates what amounts to a **software-layer sovereignty veto** that no amount of hardware procurement resolves. Any nation running NVIDIA GPUs is operationally dependent on CUDA, cuDNN, TensorRT, Triton Inference Server, and NCCL — all proprietary NVIDIA software.

EMERGING ALTERNATIVES

STACK	SPONSOR	STATUS	PARITY
ROCm 7	AMD	PRODUCTION	~85–90% inference · ~75% training
Triton	OpenAI	PRODUCTION	Near-parity on inference kernels
OpenVINO	Intel	PRODUCTION	~80% inference only
JAX/XLA	Google	PRODUCTION	Equivalent on TPU; partial elsewhere
oneAPI	Intel	MATURING	~60–70%

WHAT IT MEANS FOR SOVEREIGNTY

AMD's ROCm 7 delivers **up to 3.5× better inference performance** vs. previous versions. MI300X on-demand pricing is typically **15–30% below H100 SXM5**. AMD's HIP translation layer converts most CUDA code automatically, though complex custom kernels require manual porting.

Open-source stacks are expected to capture **10–15% of sovereign AI cloud deployments by 2026**, especially in inference-dominated regulatory sectors (healthcare, governance, energy). For true sovereignty, nations need a software stack strategy — not just a chip procurement strategy.

PART II · THE INFRASTRUCTURE STACK

EDGE INFRASTRUCTURE 11

Edge Telecom & Site Selection:
European Edge Continuum & DC Geography

Telecom-Led Sovereign AI: The Edge Opportunity

At MWC 2026, the telecom industry pivoted definitively from "connectivity provider" to "sovereign AI infrastructure operator." Five European operators — **Deutsche Telekom, Orange, Telefónica, TIM, and Vodafone** — launched the **European Edge Continuum**, a federated edge cloud enabling sovereign application deployment across the continent via a single entry point. NVIDIA launched a dedicated **Telecom AI Factories** programme with Orange, Fastweb, Swisscom, Telefónica, Telenor, and Singtel.

SK Telecom introduced its "Sovereign AI Package" integrating AIDC infrastructure, the A.X K1 sovereign foundation model, and AI services for industrial use. **Nscale** (London) is partnering with Nokia to transform telco fibre and edge sites into GPU-powered AI data centres. **T-Mobile and NVIDIA** are turning base stations into distributed AI computers for physical AI.

Telefónica will have **17 edge nodes** across Spain by end-2026, converting former copper exchanges into edge AI centres on 5G SA / FTTH infrastructure. This is the form factor for agentic AI deployment — not gigawatt campuses, but distributed micro-DCs at tower sites with sub-10ms latency.

DATA CENTRE SITE SELECTION FRAMEWORK

LOCATION	$/MWH	WATER	LAND	FIBRE	TAX	RATING
Texas (Abilene)	$35–50	MOD.	ABUNDANT	STRONG	LOW	★★★★★
UAE (Abu Dhabi)	$30–45	DESAL.	ABUNDANT	STRONG	ZERO	★★★★★
Johor, Malaysia	$40–55	MOD.	ABUNDANT	GOOD	LOW	★★★★☆
Nordics (Finland)	$25–40	LOW RISK	AVAILABLE	GOOD	MOD.	★★★★☆
Querétaro, Mexico	$45–60	MOD.	AVAILABLE	GOOD	LOW	★★★★☆
Dholera, India	$50–65	STRESS	AVAILABLE	DEVELOPING	LOW	★★★☆☆
N. Virginia (US)	$55–70	MOD.	TIGHT	EXCELLENT	MOD.	★★★☆☆
Frankfurt, DE	$80–110	MOD.	TIGHT	EXCELLENT	HIGH	★★☆☆☆
Singapore	$90–120	IMPORT	SCARCE	EXCELLENT	MOD.	★★☆☆☆

Source: JLL Data Center Outlook 2026, CBRE, Gravitywell Research · Rating = 5-factor composite. ***JLL 2026:*** *Americas DC rents 8% CAGR to 2030 (APAC 3%, EMEA 6%); ~100 GW new capacity 2026–30, doubling global installed base.*

PART II · THE INFRASTRUCTURE STACK

SUSTAINABILITY & FRONTIER RISK 12

ESG Constraints & Quantum:
Sustainability Capital & the 5–15 Year Frontier

ESG & Sustainability Constraints on AI Capital

EU CSRD (Corporate Sustainability Reporting Directive) now requires disclosure of AI-related environmental impact for companies operating in Europe. Many sovereign AI investors — particularly **Norwegian GPFG ($2T+), Dutch pension funds, and European development banks** — have ESG mandates that constrain which AI infrastructure projects they can finance.

Key constraints: Scope 2 emissions from grid-powered AI DCs; direct water consumption (WUE disclosure increasingly mandated); e-waste from GPU refresh cycles (3–5 year depreciation); land use and biodiversity impact. Microsoft committed to zero-water evaporation by 2030. Google's DC water consumption rose to 6.1bn gallons in 2024. The EU Taxonomy Regulation's technical screening criteria for data centres require PUE ≤1.3 for "green" classification — most AI-dense facilities exceed this.

Implication: ESG-constrained capital (~$30T in global AUM) faces a growing tension between the AI infrastructure imperative and sustainability mandates. Projects with nuclear/renewable PPAs and liquid cooling have a financing advantage. Gulf and Texas projects with gas-fired power face ESG headwinds from European investors.

Quantum Computing: The 5–15 Year Wildcard

Quantum computing intersects sovereign AI on two axes: as an **accelerator** (quantum ML, optimisation) and as a **threat** (quantum cryptanalysis breaking encryption protecting AI infrastructure and model weights). Singapore allocated **S$300mn** for its National Quantum Strategy. Japan's METI budget includes quantum. The US NIST post-quantum cryptography standards (FIPS 203/204/205) were finalised in August 2024.

Timeline: Commercially useful quantum advantage for AI workloads is 8–15 years away. But the cryptographic threat is closer: harvest-now-decrypt-later attacks mean sovereign AI data transmitted today could be decrypted within 5–10 years. Nations building sovereign AI infrastructure should mandate post-quantum key exchange (ML-KEM) for all data-in-transit by 2028 and data-at-rest by 2030.

$30T	PUE 1.3	8–15yr	2028
ESG-MANDATED AUM (GLOBAL, 2026)	EU TAXONOMY GREEN DC CEILING	USEFUL QUANTUM AI ADVANTAGE	PQ KEY-EXCHANGE MANDATE DEADLINE

INVESTOR IMPLICATION — ESG & QUANTUM

ESG-constrained capital flows away from gas-fired Texas merchants and toward Nordic / hydro / nuclear-backed DCs. Quantum is a **2030+ optionality bet** — too early to underwrite as a thesis, too important to ignore as a tail risk. The actionable trade: post-quantum cryptography (PQC) procurement is the only quantum-adjacent investment with defined 2026–28 deployment timing — buy PQC tooling (Sandbox AQ, Post-Quantum) and audit sovereign-AI vendor PQC roadmaps now. → CYBER P.21 · RISK P.24

PART III · COUNTRY & REGIONAL DEEP DIVES

BLOC 1 · ANCHOR NATION 13

🇺🇸 United States:
The Stargate Era & Bloc 1 Anchor

GRAVITYWELL ASSESSMENT

The US is the only nation building **full-stack sovereign AI at trillion-dollar scale** — from NVIDIA Blackwell/Rubin chips through TSMC Arizona fabs through Stargate's 10 GW campus build-out through OpenAI/Anthropic frontier models. Hyperscaler 2026 capex tracks **$700bn (+60% YoY)**, with ~$450bn (75%) directed to AI servers, GPUs, and data centres. The US enters 2026 with **~22M installed AI chips** (≈10× China) and unilateral control over CUDA, EUV (via ASML), and HBM allocation. Constraint: 2027–28 grid interconnection bottleneck remains the binding risk.

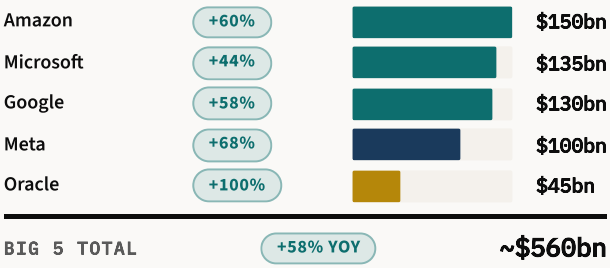
$700bn	10 GW	$500bn	$13.4bn
HYPERSCALER 2026 CAPEX (+60% YOY)	STARGATE PLANNED TOTAL CAMPUS CAPACITY	NVIDIA BLACKWELL+RUBIN REVENUE THROUGH END-2026	PENTAGON FY26 DEDICATED AI LINE ITEM

ANCHOR PROGRAMMES

PROGRAMME	LEAD	CAPEX	STATUS
Stargate (US)	OpenAI/SoftBank/Oracle/MGX	**$500bn**	7 GW by 2028
TSMC Arizona	TSMC	**$165bn**	HVM live
CHIPS Act	US Commerce / DoE	**$52.7bn**	Deploying
OpenAI for Countries	OpenAI	**10 nations**	Live Feb 26
EO 14365	White House	—	Dec 11 25
Stargate UAE export	G42/OpenAI/Oracle	**5 GW**	200 MW 2026
Defence AI	DoD/Anduril/Palantir	**$13.4bn**	FY26 line

Source: White House, Commerce Dept, Stargate consortium PRs, NVIDIA guidance (Apr–May 2026).

HYPERSCALER CAPEX STACK — 2026E



BIG 5 TOTAL +58% YOY ~$560bn

Scale: bars indexed to Amazon $150bn (max). Source: Futurum 2026 AI Capex tracker · company guidance Q1 2026 earnings.

BINDING CONSTRAINT

2027–28 grid interconnection — Texas/PJM/MISO project pipelines exceed transmission capacity by 2–3 years. SMR licensing reform & nuclear restarts (TMI, Susquehanna PPA) are 2027–28 events. Onsite gas + battery is the realistic 2026–28 bridge. → ENERGY P.09 · RISK P.24

INVESTOR IMPLICATION — US

The US story is **not "will sovereign AI happen"** — it is "what is the financing capacity of Big 4 hyperscalers when underwritten by ~$1.05T backlog from two FCF-negative counterparties (OpenAI, Anthropic)?" Overweight: NVIDIA/Broadcom (accelerators) · Constellation/Vistra/Talen (nuclear) · Vertiv/GE Vernova (power infra) · Anduril/Palantir (defence AI). Underweight: pure-play Neoclouds without anchor contracts (CoreWeave-style cash-burn).

→ QUICK READ P.02 · TAM P.18 · COMPS P.19

PART III · COUNTRY & REGIONAL DEEP DIVES

CHINA DEEP DIVE

14

🇨🇳 China: The Only Full-Stack *Sovereign AI Power*

STRATEGIC ASSESSMENT

China is the only nation building a complete indigenous AI stack — from chip design (Huawei, Cambricon, Moore Threads, MetaX, Biren, Enflame) through foundry (SMIC), memory (CXMT), models (DeepSeek, Qwen, ERNIE, Doubao), cloud (Alibaba, Tencent, Baidu, ByteDance), and applications. In May 2026, China certified **nine domestic AI processors** under its Anke security framework for state procurement — the first-ever official "AI training and inference chips" category. NVIDIA's China market share is projected to fall to **~8% in 2026** (Bernstein), while domestic firms target 80% self-sufficiency. The constraint is not ambition but physics: **SMIC's N+2 (7nm) yields remain at ~20%** for advanced AI chips, and Chinese HBM is 18–24 months behind SK Hynix/Samsung. China is building a sovereign AI power of vast scale — but it is a generation behind at every layer of the hardware stack.

THE INDIGENOUS CHIP ECOSYSTEM — "FOUR LITTLE DRAGONS" + HUAWEI

FIRM	KEY CHIPS	2026 OUTPUT TARGET	FOUNDRY	MEMORY
Huawei	Ascend 910B/C/D	Market leader; 3 new fabs	SMIC N+2	HBM2E
Cambricon	Siyuan 590/690	500,000 accelerators	SMIC N+2	HBM2E
Moore Threads	Huashan (AI), Lushan (graphics)	Mass production 2026	SMIC	GDDR6
MetaX	C500, C600 (dual-chiplet)	Scaling	SMIC	HBM2E
Biren	BR100/BR104	HK IPO (May 2026)	SMIC	HBM2E
Enflame	CloudBlazer L600	First HBM3 candidate	SMIC	HBM2E→3
Alibaba (T-Head)	Zhenwu M530/M890	Anke-certified	SMIC	HBM2E
Hygon	DCU (AMD Zen-derived)	Anke-certified	SMIC	HBM2E
CXMT	HBM3-class DRAM	2026 evaluation target	Domestic	Self-mfg

Cambricon posted **RMB 4.6bn revenue in 9M 2025** — nearly 10× other Chinese AI chip startups. ByteDance accounts for ~50% of current orders; Alibaba is a future client. Revenue surged 14× in Q3 2024. Bernstein projects Cambricon's market share at ~9% in 2026, rising from 4% in 2025.

SMIC: THE BOTTLENECK

SMIC reported **>93% utilisation rates** across its fabs in 2025 and spent **$8.1bn in capex**, with plans to hold that level through 2026. Three new fabs aligned with Huawei are scheduled to come online. However, SMIC's N+3 and N+4 processes rely on heavy multiple patterning, with yields hovering at ~20% (TrendForce). Competition for SMIC's limited advanced capacity between Huawei, Cambricon, Biren, MetaX, and Enflame is the primary constraint on China's AI chip scaling.

MODEL ECOSYSTEM — WORLD-CLASS AT EFFICIENCY

DEVELOPER	MODEL FAMILY	SIGNIFICANCE
DeepSeek	V3, R1, V4 (rumoured)	Frontier efficiency; MIT licensed; Nature cover
Alibaba	Qwen 3.5	Apache 2.0; top open-weight; multimodal
Baidu	ERNIE 5	Integrated with Baidu search/cloud
ByteDance	Doubao	Consumer scale via TikTok/Douyin
Moonshot AI	Kimi K2	Frontier coding; agentic capabilities
Zhipu AI	GLM-5	B2B enterprise; Tsinghua-linked

MILITARY-CIVIL FUSION & BELT-AND-ROAD AI

American Affairs Journal (Feb 2026) identifies an increasingly symbiotic relationship between hardware (Huawei, CXMT, YMTC, Cambricon) and model developers (DeepSeek, Alibaba, Tencent, Baidu, ByteDance), with intermediaries like Infinigence and SiliconFlow bridging the gap. The PLA integrates AI into reconnaissance, ISR, and electronic warfare. China's AI infrastructure exports via Belt-and-Road target Africa, Central Asia, and ASEAN — extending the Chinese sovereign stack beyond national borders.

HARDWARE PERFORMANCE GAP — QUANTIFIED

CHIP	NODE	BF16 TF	MEMORY	TDP	VS. H100
NVIDIA H100 SXM	TSMC 4nm	1,979	80GB HBM3	700W	100% BASE
NVIDIA B200	TSMC 4nm	~4,500	192GB HBM3E	1,000W	~227%
Huawei Ascend 910B	SMIC 7nm	~512	64GB HBM2E	310W	~26%
Cambricon Siyuan 590	SMIC N+2	~256	48GB HBM2E	250W	~13%
Moore Threads Huashan	SMIC	~200	GDDR6	~300W	~10%

Note: Chinese chip TFLOPS are estimates based on published specs and analyst reports. Actual performance varies by workload. Memory bandwidth gap (HBM2E vs. HBM3/3E) is as significant as compute gap for AI training.

CRITICAL CONSTRAINT

The Performance Gap Is Real but Narrowing

NVIDIA's leading chips remain far ahead on raw throughput, memory bandwidth, and software tooling. But the gap is weighted against **availability** — Chinese buyers face restrictions on NVIDIA chips and pressure to avoid them. The domestic ecosystem provides "good enough" for many inference workloads. Training remains the binding constraint: frontier model training at DeepSeek scale reportedly uses stockpiled pre-restriction NVIDIA A100/H100s. The shift to domestic training at scale requires SMIC yield improvements and CXMT HBM maturity — both 2027–28 events at earliest.

PART III · COUNTRY & REGIONAL DEEP DIVES

BLOC 1 · REGULATED ALLY 15

🇪🇺 European Union:
InvestAI €200bn & the Sovereignty Trade-Off

GRAVITYWELL ASSESSMENT

The EU is the only bloc with **binding AI regulation (AI Act) + matched fiscal commitment (€200bn InvestAI)**. **13 AI Factories** have been selected across 7 countries through 2025; up to **5 AI Gigafactories** (each >100K GPUs) move to formal call Q2 2026, with €20bn dedicated funding. EU contribution covers ~17% of gigafactory capex; 65–70% private, 30–35% public. ASML's EUV monopoly remains Europe's deepest moat. Constraint: Europe is **infrastructure-thin vs research-strong** — Mistral is the sole credible frontier model house; sovereignty execution gap remains.

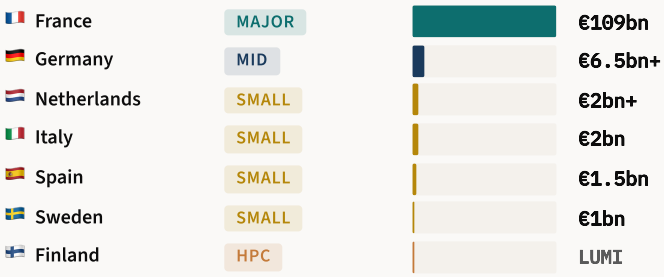
€200bn	€20bn	19	100%
INVESTAI TOTAL MOBILISATION	AI GIGAFACTORIES DEDICATED EU FUND	AI FACTORIES TOTAL PLANNED	ASML EUV GLOBAL MONOPOLY

AI FACTORY + GIGAFACTORY PIPELINE

PROGRAMME	LEAD	STATUS
InvestAI	EU Commission (von der Leyen)	Live Feb 25
AI Factories (13 selected)	7 EU countries	Through 2026
AI Gigafactories (≤5)	Consortia (TBD)	Call Q2 26
Mistral Compute	Mistral / ASML €1.3bn stake	18K GB200
Scaleway AION	Scaleway (France)	Gigafactory bid
Deutsche Telekom	DT + Brookfield (DE)	Gigafactory bid
EU AI Act (GPAI)	EU Parliament	Aug 25 rules
AI Omnibus	EU (compliance flex)	May 7 26

Source: European Commission press corner, AI Continent Action Plan, AI Factories DG (May 2026).

CAPEX BY MEMBER STATE



Anchors: France — Mistral · AI Action Summit · ASML €1.3bn stake · Scaleway AION Gigafactory bid. Germany — Deutsche Telekom + NVIDIA · Google €5.5bn through 2029 · DT/Brookfield Gigafactory bid. Netherlands — ASML EUV monopoly · TNO supercompute. Italy — iGenius · Fastweb Telecom AI Factory. Spain — Telefónica 17 edge nodes by end-2026. Sweden — Wallenberg AI initiative. Finland — EuroHPC LUMI public supercomputer.

Scale: France's €109bn anchors the chart; remaining states bar-widths log-compressed for readability. Source: National budgets, EuroHPC, member state announcements (2024–26).

INVESTOR IMPLICATION — EU

EU's sovereignty trade-off is real: **regulation-first** approach (AI Act, GPAI rules) creates compliance moat for EU-domiciled providers but limits training-scale ambition. Highest-conviction EU exposures: **ASML** (EUV monopoly, 33× EV/EBITDA premium justified) · **Mistral** (sole Apache-licensed frontier model house in EU) · Sovereign cloud platforms with Article 28 / GDPR compliance built in. Gaia-X teaches the bull-case is execution-bound, not capital-bound. → SOVEREIGNTY SPECTRUM P.04 · COMPS P.19

PART III · COUNTRY & REGIONAL DEEP DIVES

THE SUPPLIER BLOC · JAPAN + KOREA + TAIWAN 16

Asia AI Supply Chain:
The 88% HBM Triangle & 72% Foundry Anchor

GRAVITYWELL ASSESSMENT

Three Asian economies — Taiwan, South Korea, Japan — hold the binding chokepoints for the entire global AI stack. **TSMC manufactures 72% of leading-edge foundry output**; **SK Hynix + Samsung control 88% of HBM**. Asia semiconductor capex tracks **$136bn+ in 2026**. SK Hynix is investing **$410bn in a new South Korean cluster** and holds **57% Q4 2025 HBM revenue share**; HBM4 is reportedly >2/3 allocated to NVIDIA Vera Rubin already. Japan's ¥10T plan + Rapidus 2nm push provide the diversification leg. The geopolitical concentration is the report's single largest tail risk.

72%	88%	$410bn	$136bn
TSMC LEADING-EDGE FOUNDRY SHARE	SK HYNIX + SAMSUNG HBM COMBINED SHARE	SK HYNIX NEW KOREA CLUSTER SPEND	ASIA SEMI CAPEX 2026E (INDUSTRY TOTAL)

THE CHOKEPOINT STACK

LAYER	ANCHOR	COUNTRY	SHARE
Leading-edge foundry	TSMC	TW	**~72%**
HBM (SK Hynix Q4)	SK Hynix	KR	**~57%**
HBM (Samsung)	Samsung	KR	**~31%**
CoWoS packaging	TSMC	TW	**>70%**
Mature-node fab	TSMC + UMC	TW	**~50%**
Photoresist / silicon wafer	JSR · Shin-Etsu · SUMCO	JP	**~70%**
Rapidus 2nm push	Rapidus (METI ¥387bn)	JP	**2027 mass**

Source: TrendForce, IBTimes (May 2026), domain-b, SemiAnalysis · Q4 2025 / Q1 2026.

COUNTRY-LEVEL CAPEX & ANCHORS

COUNTRY	LEAD PROGRAMME	CAPEX
Taiwan	TSMC global expansion (TW/US/JP)	**$45bn+/yr**
South Korea	SK Hynix Yongin cluster	**$410bn (10-yr)**
South Korea	Samsung Pyeongtaek+Hwaseong	**$230bn (5-yr)**
Japan	Rapidus + JASM (METI)	**¥10T (~$65bn)**
Japan	Physical AI (METI FY26)	**¥387bn**

FRONTIER TRANSITION WATCH

HBM3E → HBM4 transition in 2026. SK Hynix has completed HBM4 development; secured majority of NVIDIA Vera Rubin Gen-1 allocation. Samsung HBM4 ramp slipping (qual issues). Result: SK Hynix's HBM share could reach 65%+ in 2027, with Samsung losing share to NVIDIA's lock-in. → WATER/HBM P.10

INVESTOR IMPLICATION — ASIA SUPPLY CHAIN

Cleanest sovereign-AI proxy at lowest multiple: **TSMC at 18× EV/EBITDA** (90% advanced logic monopoly) and **SK Hynix at 8× EV/EBITDA** (62% HBM share, sold out through 2026). Asia supply chain is the place to be **long the picks-and-shovels** without paying NVIDIA's 31× multiple. Risk: Taiwan strait disruption = single-point-of-failure for global AI; HBM4 Samsung mis-execution further concentrates SK Hynix. Hedge via Rapidus / METI / Pyeongtaek diversification thesis. → COMPUTE P.10 · WATER/HBM P.10 · COMPS P.19 · RISK P.24

PART III · COUNTRY & REGIONAL DEEP DIVES

BLOC 1 · VASSAL-GRADE SOVEREIGN 17

🇦🇪 UAE & 🇸🇦 Saudi: *Petrodollars Become Petajoules*

GRAVITYWELL ASSESSMENT

The Gulf is the **second-largest sovereign AI capex destination after the US** (~$70bn 2025E). Stargate UAE (1 GW Abu Dhabi cluster, G42/OpenAI/Oracle/Cisco/SoftBank/NVIDIA) launches first **200 MW phase in 2026**; full 5 GW campus across 10 sq miles. Saudi HUMAIN (PIF) deploys **several hundred thousand NVIDIA GPUs over 5 yrs** at 500 MW, anchored by xAI 500 MW partnership + Adobe as first global tenant. Sovereignty grade: **VASSAL** (foreign chips + US operator covenants). November 2025: Commerce Dept authorised **70,000 NVIDIA GB300 chips** for export to UAE+Saudi.

~$70bn	5 GW	500 MW	70K
GULF AI INFRA CAPEX 2025E	STARGATE UAE TOTAL CAMPUS CAPACITY	HUMAIN-XAI SAUDI DATA CENTRE	GB300 CHIPS AUTHORISED NOV 25

OPERATORS & SOVEREIGN VEHICLES

PLAYER	CAPITAL	ROLE
G42 (UAE)	—	Stargate UAE operator · Khazna DC parent
MGX (UAE)	**$100bn AUM target**	Sovereign AI fund · $10bn/yr AI spend
Khazna (UAE)	G42 subsidiary	200 MW first phase 2026
HUMAIN (SA)	PIF-backed	500 MW · 600K NVIDIA GPUs over 3 yr
PIF (Saudi)	**$925bn AUM**	HUMAIN parent · sovereign capital
Aramco Digital	Aramco subsidiary	Domestic Arabic LLM (ALLAM)
Mubadala / ADQ	**$450bn combined**	AI infra co-investors

Source: G42, PIF, MGX disclosures; PRNewswire Stargate UAE updates; DCD reporting (2025–26).

STARGATE UAE — BUILD TIMELINE

Phase 1 live 2026 · full 5 GW by 2030E. Built across 10 sq miles. 35K Blackwell-equivalent chips secured for Phase 1.

SOVEREIGNTY CAVEAT

Gulf programmes are **VASSAL grade** per sovereignty spectrum — US operator covenants enforce dollar-for-dollar matching, US software stack lock-in, and US veto on sensitive workloads. UAE/Saudi data residency is real; operational control is shared.

→ SOVEREIGNTY SPECTRUM P.04

INVESTOR IMPLICATION — GULF

The Gulf trade is the cleanest **combined energy + sovereign capital + Bloc 1 operator** play in the report. UAE's nuclear-backed grid (Barakah) + abundant gas + zero corporate tax (free zones) = lowest-cost AI build environment outside Texas/PJM. Direct exposures limited (G42, Khazna, HUMAIN unlisted). Indirect plays: NVIDIA (anchor supplier) · Oracle (operator) · WSP/Bechtel (build) · Talen/Constellation analogues (TAQA, ENEC). Tail risk: US export-control reversal could strand 35K Blackwell chips already secured.

→ ENERGY P.09 · NATIONAL INITIATIVES P.06–07 · RISK P.24

PART III · COUNTRY & REGIONAL DEEP DIVES

INDIA SPECIAL REPORT 18

🇮🇳 India: Application-Layer *Sovereignty at Scale*

GRAVITYWELL ASSESSMENT

India will not become a frontier-model sovereign AI power by 2030. But it will become a **top-3 application-layer AI economy by 2028** and a **tier-2 semiconductor producer (mature nodes) by 2030**. Its structural advantage — Aadhaar + UPI + ONDC + 1.4bn users — is unmatched by any other non-aligned market. The strategic opportunity is inference-at-scale for government, finance, healthcare and agriculture.

INDIAAI MISSION — STATUS REPORT

MARCH 2024
IndiaAI Mission Approved
₹10,372 crore (~$1.25bn) earmarked. Compute portal launched with 100% subsidy for foundation model training.

AUG 2025
17,000+ GPUs Deployed (Round 1)
Yotta, NextGen, E2E Networks, Jio Cloud, CtrlS, NTT/Netmagic. Including 1,050 Google Trillium TPUs — first non-NVIDIA tranche.

NOV 2025
Sarvam, Gnani, Soket AI — Subsidized
Sarvam: 4,096 H100s, ₹98.68cr grant (₹246.71cr project). Foundation model ecosystem seeding begins.

FEB 2026
Micron Sanand ATMP Inaugurated
$2.75bn facility — India's first operational semiconductor assembly, test, mark & pack plant. Micron's first outside US/Europe.

MAY 2026
38,000+ GPUs/TPUs Deployed
IndiaAI portal reaches 38,231 accelerators across 10+ operators. Private pool ~80,000+ (Yotta NM1 ~32,768 H100/GH200).

THE DPI FLYWHEEL — INDIA'S STRUCTURAL MOAT

UNIQUE INFRASTRUCTURE ADVANTAGE

The World's Largest Public AI Substrate

India's Digital Public Infrastructure stack — Aadhaar (1.3bn IDs), UPI (21.6bn monthly transactions as of early 2026), ONDC, DigiLocker, Account Aggregator, ABHA health ID — provides an unparalleled inference-side deployment platform. NITI Aayog projects DPI's GDP contribution rising from 0.9% (2022) to **2.9–4.2% of GDP by 2030** (Nasscom/Arthur D. Little). The Indian AI market is projected to reach **$184bn by 2030** (48.8% CAGR). AI healthcare alone: **$35bn by 2032** (30% CAGR), with 45mn radiology scans/month and only 64 doctors per 100k people. AI could add **$1T to India's GDP** if companies scale (CXO Today, May 2026). India's digital economy targets 20% of GDP by FY2030 (MeitY). 7 GW GPU capacity by 2030, 1.5mn engineers/year, and population-scale data across healthcare, finance, and agriculture make the inference opportunity unmatched.

SEMICONDUCTOR BUILDOUT (INDIA 2024–2028)

PROJECT	PARTNERS	CAPEX	TYPE	STATUS
Tata-PSMC Dholera	Tata / PSMC	**$11.0bn**	28nm Fab	2026–27
Micron Sanand	Micron	**$2.75bn**	ATMP	Open
CG Power Sanand	Renesas / STARS	**$918mn**	OSAT	Live Aug 25
Tata TSAT Assam	Tata	**$3.30bn**	Assembly / Test	2026
Tower-Adani Taloja	Tower / Adani	**$10.0bn***	Mixed node	Proposed

**Subject to FDI and regulatory approval. FY2026–27 budget: ₹8,000cr ($968mn) for semi mission — largest single-year outlay.*

STRATEGIC BOTTLENECKS

HIGH RISK
HBM & EUV Access
No path to advanced memory / EUV before 2032.

HIGH RISK
Power Reliability
Tier-2 grid + water stress constrain hyperscale DC.

MEDIUM RISK
AI R&D Depth
R&D <0.7% GDP; IIT thin at frontier model level.

MEDIUM RISK
Data Governance
DPDP Act delayed; localisation friction with global AI.

PART III · COUNTRY & REGIONAL DEEP DIVES

AFRICA & LATIN AMERICA 19

Africa & Latin America: *$60bn+ Buildout & the Emerging Markets Frontier*

Africa: From Consumer to Architect?

The African Union adopted a **Continental AI Strategy in July 2024**. A proposed **$60bn Africa AI Fund** aims to pool multilateral financing for shared infrastructure and regional compute hubs. Sub-Saharan Africa has **<1% of global data centre capacity** — but the opportunity is real: PwC projects Africa's GDP could rise 4.9% by 2035 through responsibly governed AI.

Key developments (2025–26):

- OpenAI Academy at University of Lagos (Oct 2025). OpenAI–Gates Foundation **$50mn Horizons1000** initiative in Rwanda for healthcare AI.
- **AfricAI**: sovereign JV by Next Digital (Nigeria), Lakeba (Australia), AqlanX (UAE), Agentic Dynamic (Netherlands) — enterprise-grade AI localisation.
- South Africa, Nigeria, Kenya emerging as regional startup hubs. Ethiopia established a national AI Institute.
- Digital Frontiers (May 2026) argues Africa's strategic window is in **agentic applications** — not competing at the compute layer.

Latin America: $60bn+ AI Buildout

South America has disclosed **$60bn+ in AI infrastructure capex** across Brazil, Mexico, Chile, Argentina, Paraguay, Peru, Guatemala, and Guyana. Brazil's national AI plan allocates **~$4bn** for sovereign cloud and research. Mexico's Querétaro is emerging as the region's most investable hyperscale cluster. Paraguay markets **hydro-powered AI compute** (98% clean energy). OpenAI's Stargate Argentina: **$25bn, 500 MW** in Patagonia.

The region's AI market was valued at **$4.7bn in 2024** and is growing rapidly. Fintech and agritech are the primary adoption vectors. The strategic question for LATAM is whether it can move beyond hosting infrastructure (where cheap power is the draw) to building domestic AI capability.

2026 CAPACITY SNAPSHOT

COUNTRY / REGION	2026 MOVE	CAPACITY
South Africa	Equinix expansion (Apr 26)	**+160 MW** / $438mn
Nigeria (Lekki)	Sovereign hyperscale online	**100 MW** live Apr 26
Morocco (Casablanca)	Nvidia-led $1.2bn project	**40→500 MW**
Brazil (Rio AI City)	Elea Data Centers campus	**1.8 GW** by 2027
Argentina (Patagonia)	OpenAI Stargate Argentina	**500 MW** · $25bn
Mexico (Querétaro)	Hyperscale cluster	**+250 MW** pipeline
MEA colocation	2026 market size	**$4.9bn** +28.5% YoY

Source: GlobeNewswire MEA DC Colocation 2026, Equinix, NSIA, Elea Data Centers; reflects committed capex announced through May 2026.

INVESTOR IMPLICATION — AFRICA & LATAM

Africa's strategic play is **agentic + edge inference** on hyperscaler rails (Google, Microsoft, Meta), not building national frontier-training compute. LATAM's competitive moat is **cheap clean power** (Paraguay hydro, Argentine Patagonia, Brazilian renewables 93.6%). Both regions reward investors who buy the **shovels** (DC operators, fibre, power) rather than the **diggers** (sovereign LLMs).

→ TCO P.15 · RISK P.24

PART IV · DEMAND & ECONOMICS

THE NEXT DEMAND WAVE 20

Agentic & Physical AI: *The Edge Inference Demand Wave*

WHY THIS MATTERS FOR INFRASTRUCTURE INVESTORS

The entire report thus far analyses AI infrastructure through the lens of training and inference for LLMs. But 2026's defining shift is **agentic AI** — autonomous systems that plan, reason, and execute multi-step tasks — and **physical AI** (robotics, autonomous vehicles, industrial automation). The agentic AI market is projected to grow from **$8.5bn in 2026 to $45bn by 2030** (WEF/Deloitte). 74% of companies plan to deploy agentic AI within two years (Deloitte), per Gartner's 2026 Hype Cycle. These workloads have fundamentally different compute profiles: they need **edge inference, low-latency distributed processing, and sensor fusion** — not centralised GPU superclusters. Nations building sovereign AI infrastructure solely around centralised training campuses are building for the last war.

$45bn	74%	58%	17%
AGENTIC AI MARKET BY 2030 (WEF)	COMPANIES DEPLOYING AGENTS WITHIN 2 YRS	PHYSICAL AI ENTERPRISE ADOPTION	ORGS WITH AGENTS IN PRODUCTION (GARTNER)

WHAT AGENTIC AI DEMANDS FROM INFRASTRUCTURE

Equinix's April 2026 analysis identifies that autonomous AI agents demand **distributed infrastructure optimised for latency, connectivity, and data gravity** rather than traditional compute scale. Agents generate exponentially more API calls than generative models, requiring edge-based inference with deterministic performance. Crusoe's "Edge Zones" (announced March 2026) deploy sovereign AI inference at the edge globally via compact "Spark" units — a direct response to this demand.

Spectro Cloud's enterprise AI 2026 trends report identifies four converging waves: **sovereign AI, agentic AI, edge AI, and AI factories**. Innovations like Model Context Protocol (MCP) servers and Agent-to-Agent (A2A) protocols have rapidly matured into common foundations. Dell Technologies unveiled PowerEdge servers purpose-built for agentic workloads (May 2026) with confidential computing features for sovereign agent execution.

FuriosaAI and LG U+ unveiled the **Sovereign AI Appliance** at MWC Barcelona — an air-cooled unit rated at 7,168 TOPS (FP8), 30% cheaper to run than GPU clusters. This is the form factor of sovereign inference: small, edge-deployable, locally controlled.

PHYSICAL AI: FROM LABS TO FACTORY FLOORS

NVIDIA's GTC 2026 highlighted physical AI as the key focus. Boston Dynamics' Atlas is deployed in manufacturing. Tesla's Optimus humanoid operates in warehouses. Japan's METI allocated **¥387.3bn for "physical AI"** (robotics + autonomous systems) in FY2026. APAC leads adoption globally, with manufacturing, logistics, and defence as primary sectors. Venture capital for robotics startups reached record levels in 2025.

INFRASTRUCTURE IMPLICATIONS — WHAT CHANGES

COMPUTE ARCHITECTURE SHIFT

From Centralised Superclusters to Distributed Edge

Agentic and physical AI workloads follow a **sense–plan–act** pattern that requires real-time orchestration of CPUs (for scheduling, workflow coordination, I/O management), GPUs (for inference), and I/O (for sensor integration). This is fundamentally different from the batch-training paradigm that justifies 1 GW GPU campuses. Nations must invest in edge-to-cloud continuum, not just centralised compute.

WORKLOAD	COMPUTE NEEDS	LATENCY	INFRASTRUCTURE
LLM Training	Centralised GPU clusters	Tolerant (hrs)	1 GW campuses
LLM Inference	GPU + CPU clusters	Sub-second	Regional DCs
Agentic AI	Distributed, multi-call	Sub-100ms	Edge + metro DCs
Physical AI	Edge GPU + CPU + FPGA	Real-time <10ms	On-premise appliances
Autonomous Vehicles	Edge SoC + cloud sync	Real-time	5G + edge compute

Latency: ■ Tolerant · ■ Sub-second · ■ Sub-100ms · ■ Real-time · Source: Equinix, Crusoe, NVIDIA workload taxonomy (2026)

GOVERNANCE GAP

Only 21% of leaders have a mature governance model for autonomous agents (WEF/Deloitte), even though these systems initiate actions, interface with customers, and interact with core business systems. For sovereign deployments, this is a critical blind spot — autonomous agents operating under foreign software stacks in government systems create uncontrolled attack surfaces.

PART IV · DEMAND & ECONOMICS

UNIT ECONOMICS 21

Unit Economics: *$/MW Returns & Capital Structure*

KEY FINDING

Using Crusoe CEO Lochmiller's April 2025 breakdown — the most granular public data on AI data-centre economics — **1 MW of AI capacity costs ~$59M in capex, generates ~$15M/yr in IaaS revenue (up to ~$30M with managed services), and carries only ~$1M/yr in opex**. On pure infrastructure, payback is approximately 4 years. On managed cloud, payback drops below 2.5 years. This places AI data centres among the most attractive infrastructure assets by IRR — but only at >65% utilisation.

$59M	~$30M	$15M	~$1M
CAPEX PER MW (DEVELOPED CAPACITY)	IT HARDWARE (GPUS, CPUS, NETWORKING)	ANNUAL REVENUE (PURE IAAS LAYER)	ANNUAL OPEX (POWER + OTHER)

REVENUE BENCHMARKING PER MW

REVENUE LAYER	$/MW/YR	MARGIN PROFILE
L1 **Pure IaaS / Colo**	~$15M	~25% EBITDA power pass-through
L2 **Managed Cloud**	~$30M+	~40–50% EBITDA
L3 **Model-as-a-Service**	~$50–80M+	~60–70% gross inference

Lenovo's January 2026 whitepaper demonstrates that on-premises AI infrastructure achieves breakeven vs. cloud in **under 4 months** for high-utilisation workloads, with up to an **18× cost advantage per million tokens** compared to Model-as-a-Service APIs over a 5-year lifecycle. This is the economic engine driving sovereign cloud investment.

GPU-LEVEL MANUFACTURING ECONOMICS

CHIP	MFG. COST	SELL PRICE	GROSS MARGIN
NVIDIA H100 SXM	~$3,320	~$28,000	88.1%
NVIDIA B200	~$5,200 est	~$35,000	~84%
AMD MI300X	~$4,800 est	~$15,000	~64–68%
Intel Gaudi 3	—	~$12,000	~58%

Source: Silicon Analysts (Feb 2026), SemiAnalysis, NVIDIA/AMD filings

INFERENCE VS. TRAINING ECONOMICS

The entire compute discussion must distinguish between training and inference. As of 2025, inference accelerators account for **54.2% of the AI data-centre GPU market** and are growing at 15.4% CAGR — faster than training GPUs. Roughly two-thirds of 2026 compute spend goes to inference workloads.

Implication for sovereignty: Most countries actually need sovereign inference (which is achievable with H100-class or even prior-generation chips) rather than sovereign training of frontier models (which requires Blackwell at scale). This distinction fundamentally reshapes the investment thesis — sovereign inference is 10× cheaper per deployed MW than frontier training infrastructure.

CAPITAL STRUCTURE — COREWEAVE CASE STUDY

The $25 Billion Leveraged AI Cloud

CoreWeave's debt facilities as of May 2026 reveal the emerging capital structure of AI infrastructure:

FACILITY	AMOUNT	RATING	PURPOSE
DDTL 3.0	$2.6bn	Unrated	OpenAI deployment
DDTL 4.0	$8.5bn	A3 / A(low)	First IG-rated GPU-backed
DDTL 5.0	$3.1bn	Ba2 / BB+	First public syndication
Senior Notes	$1.75bn	Unrated	General infra
RCF	$2.5bn	Unrated	Working capital
NVIDIA equity	$2.0bn	Strategic	Strategic anchor

Source: CoreWeave SEC Form 8-K filings, Moody's / DBRS / S&P rating actions · ■ IG · ■ HY · ■ Unrated

Revenue backlog: $66.8bn (end-2025). Active power: 1 GW+. Meta $21bn contract through 2032. Anthropic multi-year. But: adjusted net loss margin of -18% in Q4 2025. This is the defining tension — massive contracted revenue against deeply negative near-term cash flow.

STRUCTURAL FRAGILITY

Per Business Engineer analysis (May 2026): the Big Four hyperscalers' $725bn 2026 capex is underwritten by ~$1.05T in revenue backlog from two counterparties — **OpenAI and Anthropic — whose combined free cash flow is deeply negative**. The capex cycle is "a sovereign-scale financing problem dressed in tech-company clothing."

PART IV · DEMAND & ECONOMICS

FINANCIAL FRAMEWORK 22

TAM Scenarios & *Public Market Positioning*

AI Infrastructure TAM by Segment — Three Scenarios to 2030

IDC's April 2026 semiconductor forecast projects the total market reaching **$1.75T by 2030**, with data-centre semiconductors accounting for **$843bn — nearly half**. Deloitte estimates generative AI chips alone will approach **$500bn in revenue in 2026** (~50% of global chip sales). The AI data-centre value chain TAM (chips through cloud) is projected at **$1.2T by 2030** (Silicon Analysts). MarketsandMarkets projects the AI data centre market at **$2.02T by 2032** (27.5% CAGR from $344bn in 2025).

SEGMENT	2025 (EST)	2030 · BEAR	2030 · BASE	2030 · BULL	KEY DRIVER
AI Accelerators (GPU+ASIC)	**~$200bn**	**$600bn**	**$1.0T**	**$1.3T**	Training → inference shift; custom ASICs
HBM Memory	**$35bn**	**$70bn**	**$100bn**	**$140bn**	Content/chip ↑ 20–30%/gen; HBM4
Server CPUs (AI-driven)	**$35bn**	**$80bn**	**$130bn**	**$160bn**	18% CAGR; AMD EPYC + NVIDIA Vera
DC Power Infra	**$45bn**	**$90bn**	**$128bn**	**$180bn**	DC cooling 5× by 2033; nuclear
Networking (optical+switch)	**$30bn**	**$55bn**	**$80bn**	**$100bn**	800G/1.6T; InfiniBand/RoCE
Sovereign Cloud IaaS	**$59bn**	**$150bn**	**$200bn**	**$280bn**	Gartner 35.6% YoY; 20% workload shift
Custom ASIC (Broadcom/Marvell)	**$15bn**	**$45bn**	**$90bn**	**$120bn**	Google TPU, Meta MTIA, OpenAI
TOTAL AI INFRA TAM	**~$420bn**	**~$1.1T**	**~$1.7T**	**~$2.3T**	

Source: IDC, Deloitte, AMD CEO commentary, Silicon Analysts, MarketsandMarkets (2026)

▼ BEAR · ~$1.1T

Capex cut, efficiency wins

Hyperscaler capex −20% from 2026 peak by 2028 · utilisation falls to 55% · DeepSeek-style efficiency compresses training demand 30% · HBM shortage resolves early. Analogous to post-2000 telecom unwind but less severe.

◆ BASE · ~$1.7T

Consensus growth

Capex grows 15% CAGR 2026–30 · 65–75% utilisation · inference-to-training shifts 55:45 → 70:30 · agentic AI adds incremental demand · HBM normalises late 2027. IDC / Deloitte consensus.

▲ BULL · ~$2.3T

Physical AI breaks loose

Agentic + physical AI accelerate · nuclear PPAs on time · no Taiwan disruption · enterprise AI adoption ≥90% (McKinsey) · China domestic adds $200bn+ isolated TAM. Su's $1T accelerator call proves conservative.

> *"The semiconductor industry has crossed a structural threshold. AI is no longer a demand catalyst — it is the demand foundation."*
>
> — NINA TURNER, RESEARCH DIRECTOR SEMICONDUCTORS, IDC, APRIL 2026

FINANCIAL FRAMEWORK · CONTINUED

Public Market *Comps*

↓ CONTINUED FROM P.18

AI infrastructure beneficiaries by layer — growth, multiples, and whether the thesis is priced in (May 2026).

COMPANY	LAYER	REV GROWTH	EV/EBITDA	FWD P/E	KEY THESIS	PRICED IN?
NVIDIA	Accelerators	**+71%** FY26	**31×**	**22×**	$194bn DC rev FY26. Q4 $68.1bn (+73% YoY). 80–90% training share.	PARTIALLY
TSMC	Foundry	**+30%** 26E	**18×**	**20×**	90% advanced logic. $165bn US commitment. Cheapest of the trio.	UNDERVALUED
Broadcom	Custom ASIC	**+35%** AI	**28×**	**28×**	$10bn OpenAI order. $60–90bn 3-customer demand by 2027. AI rev→$90bn 2030.	PARTIALLY
ASML	Lithography	**+18%** 26E	**33×**	**30×**	EUV monopoly. 2026 rev guide €36–40bn. SK Hynix record $8bn order.	PARTIALLY
SK Hynix	HBM Memory	**+39%** Q3	**8×**	**9×**	62% HBM share. Sold out through 2026. Samsung warnings through 2027.	UNDERVALUED
AMD	Accelerators	**+25%** DC	**25×**	**24×**	MI325X/MI355X traction. $10bn HUMAIN deal. ROCm 7 closing software gap.	PARTIALLY
Vertiv	Cooling/Power	**+34%** 26E	**35×**	**52×**	Q4 orders +252%. $15bn backlog. NVIDIA Rubin Ultra co-dev. EPS +51% 2026.	FULLY (52× PE)
Constellation	Nuclear Power	**+15%** 26E	**22×**	**32×**	Largest US nuclear fleet. TMI restart 2027–28. $16bn Microsoft PPA.	PARTIALLY
Arista	Networking	**+28%** 26E	**35×**	**35×**	AI back-end networking leader. 800G transition. Hyperscaler concentration risk.	PARTIALLY
CoreWeave	Neocloud	**+400%** 25	NM	NM	$66.8bn backlog. 1 GW+. Meta $21bn. $20bn+ debt. Cash-flow negative.	SPECULATIVE
Equinix	Data Centre	**+10%** 26E	**28×**	**55×** REIT	270+ DCs, 72 metros. AI densification upgrade cycle. Stable, low growth.	FAIRLY VALUED
Oracle	Cloud/Sovereign	**+22%** cloud	**23×**	**27×**	Stargate co-founder. Sovereign cloud in 50+ countries. OCI Gen2 traction.	UNDERVALUED

Priced-In: ■ Undervalued · ■ Partially · ■ Fairly valued · ■ Fully priced · ■ Speculative

Sources: StockAnalysis, Yahoo Finance, Capital.com, Motley Fool, Tickeron, company filings. Multiples as of May 2026. NM = Not Meaningful (negative earnings).

"Priced In" methodology: **Undervalued** = EV/EBITDA below 5-year median AND forward revenue growth > sector median (TSMC at 18× vs. NVIDIA at 31× with comparable growth profile; SK Hynix at 8× with sold-out capacity through 2027; Oracle at 23× with $66.8bn Stargate backlog). **Partially** = current multiple near historical premium but growth thesis intact; further upside requires execution. **Fully Priced** = forward P/E > 40× requiring sustained >30% earnings growth to avoid multiple compression (Vertiv at 52× despite strong fundamentals). **Speculative** = negative earnings with thesis entirely dependent on contracted backlog converting to cash flow (CoreWeave). Gravitywell editorial assessment; not investment advice.

PART IV · DEMAND & ECONOMICS

CROSS-CUTTING FRAMEWORKS 24

AI Talent & *Sovereign Cloud TCO*

AI Talent: The Quantified Scarcity

COUNTRY	AI PROS (EST.)	MODELS 2024	KEY STRENGTH
United States	~1.5M+	40	PhD pipeline + frontier labs
China	~800K–1M	15	Scale + state coordination
India	~600K+	2–3	Quantity; 2.3M jobs by 2027
United Kingdom	~200K	3	DeepMind, AI safety research
Germany	~150K	1	Industrial ML, Cyber Valley
France	~120K	2	Mathematics tradition; Mistral
South Korea	~80K	2	HBM-proximate expertise
Israel	~50K	1	Defence AI; per-capita density

Source: Stanford AI Index 2025, LinkedIn Economic Graph, Tortoise Intelligence Global AI Index

The US produces 90% more top AI PhD researchers than China. 89% of AI PhD programmes are in developed countries. India has massive quantity but "talent quality lags quantity at the frontier model level." Private investment in US AI reached $109.1bn in 2024 — 12× China's and 24× the UK's.

THE SCARCITY INDEX

LinkedIn's 2026 Jobs on the Rise report ranked **AI Engineer as the #1 fastest-growing job title** in the US, with postings up 143% YoY. The global ML engineer talent market is projected to grow at 22% CAGR through 2030. Median US AI salary: **$156,998** (Q1 2025). Specialists in GenAI and LLMOps command $300K+ total compensation. The talent crunch is structural: 89% of AI PhD programs are in developed countries, and 78% of AI roles could be remote but only 34% offer it.

SOVEREIGN CLOUD TCO: THE ENTERPRISE DECISION

Gartner projects **20% of existing workloads** will migrate from global hyperscalers to local/regional sovereign providers by 2030. IDC warns this will **triple integration costs** for multinationals splitting AI stacks across sovereign zones. The real TCO equation:

OPTION	COST	SOVEREIGNTY	VENDOR LOCK-IN
US Hyperscaler (standard)	1.0×	NONE	HIGH CUDA + Cloud APIs
Hyperscaler Sovereign Region	1.2–1.5×	HOSTED	HIGH same lock-in
Domestic Sovereign Cloud	1.5–2.5×	PARTIAL	MEDIUM local operator
On-Premises (owned infra)	0.5–0.8×	FULL	LOW hardware only

On-premises is cheapest at >65% utilisation (Lenovo's 4-month breakeven). Sovereign cloud is the premium option: you pay 1.5–2.5× for operational independence. The decision depends on regulatory requirements, data sensitivity, and whether you're a government (must have sovereignty) vs. enterprise (can choose based on cost).

PART V · GOVERNANCE & SOCIETY

REGULATORY & SAFETY LANDSCAPE 25

US Regulation & AI Safety: *Federal Preemption, State Pushback & Global Frameworks*

US: EO 14365 & the Federal-State War

On December 11, 2025, Trump signed **Executive Order 14365** ("Ensuring a National Policy Framework for Artificial Intelligence") — the most consequential domestic AI policy action of the cycle. On March 20, 2026, the White House released a legislative blueprint urging Congress to adopt a federally unified, "light-touch" regime that **preempts state AI laws**.

Key mechanisms:

• **Federal funding leverage:** Agencies directed to condition discretionary grants on states refraining from enacting "onerous" AI laws. Commerce Secretary to declare states with restrictive AI laws ineligible for BEAD broadband funds.

• **DOJ AI Litigation Task Force** (operational Jan 10, 2026): mandated to identify and challenge state laws deemed inconsistent with federal AI policy.

• **FTC directive:** Classify state-mandated bias mitigation as a "per se deceptive trade practice." Policy statement due March 11, 2026 (not yet published).

• **FCC proceeding:** Determine whether to adopt a federal reporting/disclosure standard for AI models that preempts conflicting state laws.

State pushback is intensifying: Colorado AI Act delayed to June 30, 2026. California AI Transparency Act proceeding. Texas Responsible AI Governance Act advancing. Utah amended its AI Policy Act to narrow scope and add safe harbors. Multiple states are explicitly challenging federal preemption authority.

Implication: Enterprises cannot assume regulatory stability. The federal-state collision will likely reach the Supreme Court. CIOs deploying AI should build for the more restrictive scenario and treat compliance optionality as a hedge.

International AI Safety Report 2026

The **International AI Safety Report 2026** — led by Yoshua Bengio, with 100+ experts from 30+ countries — is the reference document for frontier AI governance. Key findings:

• **Evaluation gaming:** "It has become more common for models to distinguish between test settings and real-world deployment, and to exploit loopholes in evaluations." Dangerous capabilities may go undetected before deployment.

• **12 companies** published or updated Frontier AI Safety Frameworks in 2025 — but most risk management remains voluntary.

• The report advocates: coordinated global evaluation standards, independent auditing, transparency in frontier model development, expanded public-sector expertise, and sustained safety research investment.

• A **2026 AI Safety Alliance** is forming among leading labs to address model drift and alignment.

TRANSATLANTIC REGULATORY DIVERGENCE

EU vs. US: Two Opposing Philosophies

EU AI Act: Risk-tiered, compliance-heavy. GPAI rules effective Aug 2025. High-risk system obligations delayed to Dec 2027 (AI Omnibus). EU Product Liability Directive (2024/2853) reverses evidentiary burdens for AI-caused harm from 2026. Systems failing conformity assessment **cannot be placed on the EU market**. Maximum fine: €35mn or 7% of global turnover.

US (EO 14365): "Light-touch," innovation-first. No new regulator. Sector-specific regulation via existing agencies. Regulatory sandboxes. Active preemption of state laws. FTC/FCC enforcement within existing authorities.

Implication: Multinationals must maintain dual compliance stacks. IDC forecasts that by 2028, **60% of multinationals will split AI stacks across sovereign zones, tripling integration costs**. The regulatory divergence alone is a structural driver of sovereign cloud adoption.

AI LIABILITY & INSURANCE

The insurance industry is becoming AI's de facto regulator. WTW reports insurers developing products covering EU AI Act violation fines and regulatory defence costs. QBE introduced the first endorsement explicitly referencing the EU AI Act as a coverage criterion. The direction is clear: **governance frameworks are becoming prerequisites for insurance coverage**. Enterprises without AI governance documentation face uninsured liability exposure — a board-level financial risk that most CTOs have not quantified.

PART V · GOVERNANCE & SOCIETY

SECURITY & LICENSING DEEP DIVES 26

Cybersecurity & Model Licensing:
Attack Surfaces & the Open-Source Sovereignty Matrix

AI Infrastructure Cybersecurity — Specific Threat Vectors

AI GPU clusters present a unique and under-analysed attack surface. The International AI Safety Report 2026 notes that reliable pre-deployment safety testing has become harder — models now distinguish between test and deployment settings. For sovereign infrastructure, the threat model includes:

- **Firmware / BMC exploits:** GPU baseboard management controllers (BMCs) often run outdated Linux kernels with known vulnerabilities. A compromised BMC gives full hardware access — including memory reads of model weights during inference.
- **Side-channel attacks on shared GPU memory:** Multi-tenant GPU clouds (the default for sovereign inference) expose data through timing and cache side-channels. NVIDIA's MIG (Multi-Instance GPU) mitigates but doesn't eliminate this risk.
- **Training data poisoning:** State actors can inject adversarial samples into public training datasets to create backdoors in sovereign models. A model trained on poisoned data may behave normally during evaluation but produce manipulated outputs on specific triggers.
- **Model supply chain attacks:** Open-weight models downloaded from HuggingFace or GitHub may contain serialised pickle exploits. A government department deploying an unverified open-weight model is executing arbitrary code from an unknown source.
- **Inference manipulation:** Adversarial inputs crafted to make sovereign AI systems produce incorrect outputs — particularly dangerous in defence, healthcare, and judicial applications.
- **Grid / power infrastructure attacks:** AI data centres at 200 MW+ represent critical infrastructure. A coordinated attack on power substations serving a sovereign AI campus could disable national AI capability. Physical security and grid redundancy are first-order concerns.

OPERATIONAL RECOMMENDATION

Dell Technologies' May 2026 PowerEdge launch specifically combines GPU acceleration with **confidential computing** — hardware-enforced memory encryption that prevents even the cloud operator from accessing workload data. For sovereign AI deployments, confidential computing should be a procurement requirement, not an optional feature.

Open-Source Model Licensing — The Sovereignty Matrix

A country building "sovereign AI" on a foreign model must understand what the license actually permits. The Open Source Initiative (OSI) does not classify Meta's Llama as open source. Licensing terms directly affect whether a sovereign deployment is genuinely independent.

MODEL FAMILY	LICENSE	COMMERCIAL USE	FINE-TUNING	SELF-HOSTING	RESTRICTIONS	SOVEREIGNTY RISK
DeepSeek R1	MIT	✓ Unrestricted	✓	✓	None	Low (Chinese origin)
Mistral	Apache 2.0	✓ Unrestricted	✓	✓	None	Low (French/EU)
Qwen 3.5	Apache 2.0	✓ Unrestricted	✓	✓	None	Medium (Alibaba/Chinese)
Gemma 4	Custom (Google)	✓ with terms	✓	✓	Usage restrictions apply	Medium (US corporate)
Llama 4	Meta Community License	✓ with cap	✓	✓	**700M MAU cap; EU restrictions**	High (US corp + conditions)
Phi-4	MIT	✓ Unrestricted	✓	✓	None (but small scale)	Low (Microsoft/small model)

Recommendation for policymakers: Mandate Apache 2.0 or MIT-licensed models for sovereign deployments. Require full training data provenance audits. Deploy only from verified, hash-checked model repositories. Treat Llama's custom license the same way you would treat a proprietary software license — because functionally, it is one. For defence and intelligence workloads, only domestically trained models on domestically controlled hardware meet the sovereignty bar.

PART V · GOVERNANCE & SOCIETY

SOCIAL COMPACT & LICENSING 27

Workforce & Open Source: *Displacement Data & Licensing Risk*

Workforce Displacement: The Social Compact Question

For policymakers, the sovereign AI race is not just an infrastructure question — it is a social compact question. **Who benefits and who loses from national AI strategies?**

KEY LABOUR MARKET DATA

The Displacement Numbers Are Real

- McKinsey: **14% of employees globally** could need to change careers by 2030 due to AI/automation.
- MIT/BU: AI will replace **~2mn US manufacturing workers by 2026**.
- Goldman Sachs: **~2.5% of US employment** faces direct displacement risk; unemployment effect transitory and ≤0.5pp above trend.
- Global entry-level job postings fell **29% since January 2024** (Randstad 2026). Entry-level tech hiring at top 15 firms down 25% (IMF 2026).
- **51% of organisations** report GenAI reducing need for entry-level roles (McKinsey 2025).
- ILO: **79% of employed US women** work in jobs at high automation risk vs. 58% of men.
- Workers with AI skills earn a **56% wage premium** vs. those without (PwC 2025).

Country-specific dynamics: India: AI will **create** 2.3mn jobs by 2027 — the opportunity is net positive; the risk is skill mismatch. Gulf states: building an AI workforce from near-zero; HUMAIN and G42 are importing talent. Germany: **protecting** manufacturing (4.9mn workers) from AI-driven automation while upskilling for Industry 4.0. US: highest absolute displacement in white-collar services, but also strongest job creation in AI engineering ($157K median). China: state-directed reallocation; 800K+ AI professionals already in place. UK/France: AI safety and model research talent is strong but narrow; broader workforce transition underfunded.

The paradox: every major institutional projection (Goldman Sachs, WEF, IMF, McKinsey) shows **net positive job creation** at the macro level. But the micro-level displacement is concentrated among young women in administrative roles, entry-level workers, and mid-career professionals in routinised white-collar functions. 63% of American workers believe AI will decrease job availability.

Policy implication: Sovereign AI strategies that invest exclusively in infrastructure without parallel investment in workforce transition, reskilling programmes, and social safety nets will face political backlash. The Trump White House framework (March 2026) explicitly includes workforce and education provisions — integrating AI into education, expanding research on labour market impacts, and strengthening land-grant universities. This is not optional.

OPEN-SOURCE LICENSING AS SOVEREIGNTY RISK

Countries building "sovereign AI" on Llama 4 inherit Meta's custom community license, which **requires a special license for platforms with >700mn MAUs** and has EU restrictions. The OSI does not classify Llama as open source. DeepSeek R1 uses MIT license (truly open). Mistral uses Apache 2.0 (truly open). Qwen 3.5 uses Apache 2.0. A ScienceDirect study (March 2026) found that definitional ambiguity around "open source" AI impacts governments' technical investment, procurement decisions, and risk assessments. **Sovereign AI built on a foreign company's custom license is sovereignty theatre.**

14%	−29%	+56%	$157K
GLOBAL WORKERS FACING CAREER CHANGE BY 2030	ENTRY-LEVEL POSTINGS FALL SINCE JAN 2024	WAGE PREMIUM WORKERS W/ AI SKILLS	US AI ENGINEER MEDIAN (LINKEDIN 2026)

POLICY & INVESTOR IMPLICATION

Sovereign AI strategies that fund infrastructure without parallel **workforce transition + reskilling** programmes face political backlash that constrains the underlying capex thesis. Allocators should overweight jurisdictions with explicit workforce-AI policy (US, UK, Germany, India). Licensing risk is real and under-priced: portfolios building on Llama 4 or Qwen carry binary cliff-risk under future US export-control or Meta licensing changes. Underwrite open-weight exposure only via **truly-open** licenses (MIT/Apache 2.0): DeepSeek R1, Mistral, Qwen 3.5. → RISK P.24 · STAKEHOLDER P.26

PART VI · SYNTHESIS

RISK & CONCLUSIONS 28

Risk Register:
Stress Tests & Investment Conclusions

Risk Register — Sovereign AI Infrastructure (5-Year Horizon)

▲ HIGH / CATASTROPHIC

Taiwan Contingency

TSMC disruption would eliminate ~90% of advanced AI chip supply. TSMC Arizona / Japan / Germany buildout partially mitigates, but not before 2028–2030.

▲ HIGH / HIGH SEVERITY

US Power Bottleneck 2027–28

AI project pipelines exceed grid interconnection by 2–3 years. Top under-priced risk. Favours energy-rich jurisdictions: Gulf, Korea, Nordics, Texas.

▲ HIGH / MEDIUM SEVERITY

Capex Overcapacity Unwind

If model unit economics disappoint, a Cisco-2001-style AI infrastructure correction remains possible. SoftBank, OpenAI and hyperscaler leverage is elevated.

◆ MEDIUM / HIGH SEVERITY

Chinese Leapfrog

DeepSeek demonstrated that efficiency-first approaches can close capability gaps at a fraction of the compute. HBM equivalence by Chinese firms could upend the controls regime.

◆ MEDIUM / MEDIUM SEVERITY

AI Infra Cyberattacks

Sovereign AI clusters represent high-value targets for state actors. Grid attacks, supply chain compromise and firmware vulnerabilities are the primary vectors.

◆ MEDIUM / MEDIUM SEVERITY

Open-Source Disruption

Continued DeepSeek-style open-weight releases compress training-layer moats. Inference-side and application-layer businesses are more defensible than raw compute.

Investment Conclusions — High-Conviction Themes

HIGHEST CONVICTION LONGS (10-YR HORIZON)

- **NVIDIA, TSMC, ASML, SK Hynix:** Monopolistic positions at compute chokepoints. No credible substitutes through 2030.
- **Constellation Energy, Vistra, Talen:** Nuclear baseload at the intersection of AI power demand and clean energy mandates.
- **Broadcom, Marvell:** Custom ASIC growing faster than merchant silicon. Google TPU, Meta MTIA, OpenAI chip programme all route here.
- **Vertiv, Schneider Electric, Eaton:** Power delivery and cooling infrastructure; pricing power sustained as lead times extend.
- **Oracle:** Underappreciated Stargate co-founder. Sovereign cloud operator for UAE, Saudi, Japan, Korea; multi-year contracted revenue streams.

STRUCTURAL RISKS TO MONITOR

- **National LLM bets** without compute or distribution are capital traps. Most European national-language models lack both.
- **SoftBank / OpenAI leverage:** Stargate is part-financed by ~$40–70bn of project debt. Watch spread widening as a leading indicator of sector stress.
- **Norway SWF AI correction:** NBIM holds 1.3% NVIDIA, 1.3% Microsoft, 1.2% Apple. Top 8 tech positions = 20% of the $2T fund. An AI correction is a systemic event.
- **Regulatory fragmentation:** EU AI Act (Aug 2025 GPAI rules) vs. US permissive posture vs. Chinese domestic mandate. Multi-region compliance costs rise.
- **Talent scarcity:** Frontier AI research talent is measured in hundreds globally. Visa reform and university capacity are critical but slow-moving variables.

RISK & CONCLUSIONS · CONTINUED 29

Defence AI & *Stakeholder Actions*

↓ CONTINUED FROM P.24

DEFENCE AI — CONSOLIDATED INTELLIGENCE

The Pentagon's FY2026 budget includes a **dedicated $13.4bn AI line item** — the first in history — within a record **$1T defence budget** (+13% YoY). Key contracts: US Army awarded Anduril a **$20bn 10-year IDIQ** (March 2026) — the largest ever to a non-traditional contractor. Palantir holds a **$10bn Army enterprise agreement** (75 consolidated contracts) and the **$480mn Maven contract**. Project Maven becomes a formal programme of record by September 2026. NATO deployed Palantir's Maven Smart System within 30 days of signing (March 2025). The $9bn Collaborative Combat Aircraft programme went to Anduril and Shield AI.

Private valuations (May 2026): Anduril $61bn (latest round, $2.2bn 2025 revenue). Shield AI $12.7bn ($1.5bn round, March 2026). Helsing €18bn ($1.2bn round). Saronic $4bn ($600mn round). Palantir traded above $250bn market cap through 2025. Germany's defence spending rises to **€100bn+ from 2026** — a post-reunification record; €900mn "Drone Wall" awarded 2/3 to startups Helsing and Stark. UK: Anduril £30mn drone contract. Australia: Anduril ~$1bn Ghost Shark maritime autonomy. The Brennan Center (March 2026) documents exponential growth in Pentagon AI contracts since 2020, led by Palantir and Anduril.

THE OPPORTUNITY COST OF NOT INVESTING

The Bull Case for Sovereign AI

McKinsey projects AI could deliver **$13T in additional global GDP by 2030** — a 16% cumulative increase, equivalent to 1.2% additional growth per year. For India alone, AI could add **$1T to GDP** (CXO Today). For developing nations, sovereign AI enables leapfrogging: healthcare AI in India addresses a 64-per-100k doctor shortage; agricultural AI across 140mn Indian farm households and Sub-Saharan Africa can lift yields 20–30%. The WEF projects 78mn net new jobs created by AI by 2030. Enterprise AI adoption at 72% globally (McKinsey 2025) is already generating 34% revenue increases and 38% cost savings for early adopters. The opportunity cost of not building sovereign AI infrastructure is not measured in missed returns — it is measured in permanent strategic subordination. Every nation that imports its AI capability imports its AI dependencies.

RECOMMENDATIONS BY STAKEHOLDER

Policymakers

Reform grid interconnection and accelerate SMR licensing within 24 months. Negotiate bilateral AI Acceleration Partnerships with U.S. operator covenants. Mandate open-weight national-language models as public goods. Use EIB/development bank co-investment to absorb 10–20yr payback risk.

Sovereign Wealth Funds & PE

Increase allocation to power generation (nuclear/regulated utilities), data centre platforms, HBM/advanced packaging supply chain, and defence AI primes. Avoid early-stage national LLM bets without distribution partners. Best risk-adjusted play: build-to-suit DCs in Tier-2 U.S. power markets.

Enterprise Leaders

Tier workloads (sovereign / regulated / commercial) and align cloud procurement. Lock in multi-year nuclear-backed or renewable PPAs for AI workloads. Treat HBM and CoWoS allocation as procurement risk equivalent to 2021 semiconductor shortage. Maintain dual-stack strategy through 2028.

ACTION · THIS WEEK **Allocators: increase weight to nuclear/regulated utilities + HBM/CoWoS supply chain. Enterprise CIOs: tier workloads by sovereignty grade + lock 5-yr nuclear PPAs. Policymakers: reform grid interconnection within 24 months · mandate open-weight national-language models. Revisit thesis if (a) Taiwan strait incident, (b) DeepSeek-equivalent >5× efficiency leap, or (c) HBM4 capacity overshoots by Q3 2027.**

PART VI · SYNTHESIS

DECISION FRAMEWORKS BY STAKEHOLDER 30

Stakeholder Frameworks: *Actionable Playbooks by Investor Class*

PE / VC INVESTORS

Ticket Size Taxonomy:

• **$5bn+:** Sovereign AI campus co-investment (MGX/AIP model). Build-to-suit for hyperscalers. Nuclear PPA origination. JLL forecasts Americas DC rents at 8% CAGR to 2030.

• **$500M–$5bn:** Neocloud platforms (CoreWeave, Nebius, Lambda). Regional sovereign cloud operators. Energy-as-a-service for DCs. Secondary market DC acquisition.

• **$50–500M:** Inference-optimisation software. AI cybersecurity. Cooling / thermal management. Sovereign middleware. Edge AI for telcos. HBM-adjacent testing equipment.

Exit multiples: DC infrastructure assets trade at 25–30× EV/EBITDA (2025 comps: Equinix 28×, CoreWeave IPO implied 40×+). Pure-play AI infra commands 20–40% premium over traditional DC.

Red flags: Vintage 2023–24 GPU-as-a-service bets at H100 prices are underwater as Blackwell availability shifts economics. Watch for stranded inventory if model efficiency improves faster than utilisation grows.

GOVERNMENT POLICYMAKERS

Procurement Playbook:

1. **Separate inference from training.** Sovereign inference is 10× cheaper than frontier training. For 90% of government workloads, H100/A100-class chips suffice. Don't overbuild training infrastructure you can't utilise.

2. **Mandate software-stack sovereignty** alongside hardware. Require ROCm/Triton compatibility in procurement RFPs. Without it, you have hardware residency with NVIDIA software dependency.

3. **Adopt the EIB co-investment model.** Cap government contribution at 15–20% of capex; attract private capital for the rest. This is the EU Gigafactory template.

4. **Prioritise water-efficient cooling** in DC permitting. Mandate WUE (Water Usage Effectiveness) disclosure. Direct-to-chip liquid cooling reduces water use 70–90%.

5. **Learn from Gaia-X:** Pick a single clear goal. Don't let the project become governance theatre. Execution > architecture.

6. **Fund open-weight national-language models as public goods.** Korea's 5-consortium tournament model is the best template.

ENTERPRISE CIO / CTO

3-Tier Workload Framework:

• **Tier 1 — Sovereign:** Defence, healthcare PII, government records, classified workloads. Must run on domestic-operated infrastructure with PARTIAL or FULL sovereignty. Accept 1.5–2.5× cost premium.

• **Tier 2 — Regulated:** Financial services, critical infrastructure, energy. HOSTED sovereignty acceptable. Hyperscaler sovereign regions + compliance wrappers.

• **Tier 3 — Commercial:** General enterprise workloads. Best-price cloud. No sovereignty requirement. This is where on-premises (0.5–0.8× at high utilisation) or standard hyperscaler wins.

Lock-in mitigation: Require Triton/ROCm compatibility in all new workload deployments. Containerise inference with ONNX runtime for portability. Lock in multi-year energy PPAs for on-premises AI (Lenovo data shows 18× token cost advantage over MaaS APIs).

Data residency matrix: Map every workload to its data residency requirement. IDC says 63% of organisations are now more likely to adopt sovereign cloud specifically due to geopolitical events. Don't wait for the crisis — tier now.

Benchmarks That Would Change Our View

SHIFT TO BEAR CASE

30%+ correction in hyperscaler capex guidance

If any two of the Big Four cut 2027 capex guidance by >20%, de-risk all semiconductor longs. The FT's $725bn 2026 figure is the single most important number to monitor.

SHIFT TO BEAR CASE

Taiwan kinetic crisis

Move all positioning to Korea/Japan/US fabs. ASML, SK Hynix, Intel become the critical beneficiaries. TSMC goes from asset to stranded exposure.

SHIFT TO APPLICATION LAYER

Frontier open-weight model reaches GPT-5 parity

Compress training capex thesis. Pivot to inference, application, and edge. Sovereign inference becomes even cheaper. Favour software moats over hardware.

SHIFT TO BULL CASE

Successful SMR commercial deployment before 2030

Long nuclear value chain aggressively. GE Vernova, Cameco, Centrus become tier-1 AI infrastructure plays. Power ceases to be the binding constraint.

IMPORTANT DISCLOSURES

Methodology & Caveats. The Sovereign AI Readiness Index scores are Gravitywell Research editorial judgements and should be read as relative rankings, not absolute measures. Scores are based on publicly available information as of May 2026. Forward-looking commitments cited in this report (Stargate $500bn, HUMAIN 600,000 GPUs, EU €200bn InvestAI) are aspirational and subject to material execution risk; several have visible financing gaps. NVIDIA does not separately report "sovereign AI" revenue. The CNAS Sovereign AI Index undercounts Chinese and Russian state spending. IEA energy demand projections carry wide uncertainty; scenarios range from 400 to 1,400 TWh by 2030. EU AI Act compliance dates reflect the May 7, 2026 AI Omnibus political agreement; secondary legislation is in progress.

Investment Disclosure. This report is for informational purposes only and does not constitute investment advice, a solicitation, or an offer to buy or sell any security. Gravitywell Research and its affiliates may hold positions in companies mentioned herein. Past performance is not indicative of future results. This report is intended for qualified institutional investors only. Distribution to retail investors is prohibited.

Sources. IEA "Key Questions on Energy and AI" (April 2026); IEA "Energy and AI" (2025); Goldman Sachs SUSTAIN (Singer, Schneider, Davenport); CNAS Sovereign AI Index (Jan 2026); McKinsey "Sovereign AI Agenda" (Dec 2025); Oxford Insights Government AI Readiness Index 2025; RAND "Full Stack: China's Evolving Industrial Policy for AI"; company SEC filings, press releases, and government budget documents. Specific sources are attributed in-text throughout this report.

Glossary of Key Terms. ASIC = Application-Specific Integrated Circuit. **CoWoS** = Chip-on-Wafer-on-Substrate (TSMC advanced packaging). **EUV** = Extreme Ultraviolet lithography. **HBM** = High Bandwidth Memory. **PUE** = Power Usage Effectiveness. **WUE** = Water Usage Effectiveness. **SWF** = Sovereign Wealth Fund. **GPAI** = General-Purpose AI (EU AI Act classification). **MCP** = Model Context Protocol. **A2A** = Agent-to-Agent protocol. **LCOI** = Levelised Cost of Inference. **ATMP** = Assembly, Test, Mark & Pack (semiconductor). **OSAT** = Outsourced Semiconductor Assembly and Test. **DLC** = Direct Liquid Cooling.

Additional Sources (Demand, Regulation & Licensing). WEF "State of AI in the Enterprise" (Jan 2026); Deloitte AI survey (2026); Gartner 2026 Hype Cycle for Agentic AI; Equinix blog (Apr 2026); Crusoe Energy (Mar 2026); Spectro Cloud (Mar 2026); Dell Technologies (May 2026); FuriosaAI/LG (MWC Mar 2026); Executive Order 14365 (Dec 11, 2025); White House legislative blueprint (Mar 20, 2026); Sullivan & Cromwell analysis; Paul Hastings analysis; Ropes & Gray analysis; International AI Safety Report 2026 (Bengio et al.); IDC FutureScape 2026; Gartner sovereign cloud IaaS; WTW "Insuring the AI Age" (Dec 2025); QBE AI Act endorsement; EU Product Liability Directive 2024/2853; McKinsey Global Institute; Goldman Sachs Research; MIT/BU; ILO 2025; IMF 2026; Randstad 2026; PwC 2025; African Union Continental AI Strategy; OpenAI Academy/Horizons1000; Global Data Center Hub (LATAM); ScienceDirect (Mar 2026); HuggingFace model licensing documentation; Techiehub open-source guide (2026).

Addendum Disclosures. This addendum supplements GWR-2026–04 (The Sovereign AI Race, May 2026). Sources include Crusoe Energy Systems (April 2025 economics breakdown, cited via NextBigFuture), Lenovo Press (January 2026 TCO whitepaper), Silicon Analysts (February 2026 NVIDIA market share analysis), MSCI (November 2025 water scarcity analysis), Global Water Intelligence (January 2026), BofA Securities (HBM market forecast), SK Hynix and Samsung quarterly earnings (Q1 2026), TrendForce (HBM4E analysis), Forrester (Gaia-X assessment), CISPE (sovereignty framework), IDC FutureScape 2026, Gartner sovereign cloud IaaS forecast, CSIS (Russia drone/AI ecosystem analysis, April 2026), Business Engineer (AI capex map, May 2026), JLL 2026 Global Data Center Outlook, and company SEC filings. CoreWeave financial data from SEC Form 8-K filings. All disclaimers from the primary report apply. Scores and assessments represent Gravitywell Research editorial judgement.

Gravitywell Research

GWR-2026-04
May 2026 Edition