RESEARCH STUDY — FEBRUARY 2026

AI FOR AMERICANS FIRST

AI, Energy & Semiconductor Protectionism:
US/Europe Divergence Trajectories 2024–2030

Integrated Geostrategic and Economic Analysis

Chapter III

Fabrice Pizzi
Sorbonne University
Master in Economic Intelligence — Intelligence Warfare

75% global AI compute = USA | $675B US capex 2026 | 7–12× US/EU ratio

Paris — February 2026

7 chapters • 4 prospective scenarios • 3 geographical zones

Keywords: artificial intelligence, technology protectionism, semiconductors, export controls, sovereign compute, AI geopolitics, France, United States, China

CHAPTER III

Empirical Diagnosis 2020–2026

This chapter establishes the factual foundation of the analysis. It covers four interdependent dimensions: the energy trajectory of data centers, the evolution of the semiconductor market, the geographical distribution of installed AI compute, and the timeline of American regulatory measures. The data presented here constitute the "predetermined elements" (in the Schwartz sense) that structure the prospective scenarios in Chapter V. All time series are sourced, and where data diverge between sources, the discrepancy is made explicit.

3.1 Data Center Energy Trajectory: Demand Doubling in Six Years

3.1.1 Global Consumption 2020–2024

The International Energy Agency (IEA, April 2025) estimates global data center electricity consumption at approximately 415 TWh in 2024, representing 1.5% of global electricity consumption.[1] This consumption has grown at an average of 12% per year since 2017, a rate four times higher than total electricity consumption growth. In 2020, it stood at approximately 270 TWh; the increase thus represents +54% over four years.

The geographical distribution is highly concentrated. The United States absorbs 45% of global data center consumption (approximately 180 TWh in 2024), followed by China (25%, ~102 TWh) and Europe (15%, ~70 TWh for the EU according to the European Commission).[2] The four main European countries (Germany, France, United Kingdom, Netherlands) total approximately 41 TWh, or 10% of the global total. Total installed global data center capacity reached nearly 100 GW in 2024. Nearly half of US capacity is concentrated in just five regional clusters.

3.1.2 Projections 2024–2030: The IEA Central Scenario

The IEA's central scenario ("Base Case") projects global consumption of 945 TWh by 2030, representing a doubling compared to 2024 and equivalent to Japan's current electricity consumption. AI is identified as "the most important driver of this growth." AI's share of data center consumption, estimated at 5–15% in recent years, could reach 35–50% by 2030.[3]

Growth is geographically uneven. The United States adds approximately +240 TWh (+130% compared to 2024), China +175 TWh (+170%), and Europe only +45 TWh (+70%). The IEA notes a 20% risk of project delays in Europe, linked to grid connection constraints and permitting timelines.[4] In the United States, data centers will account for nearly half of electricity demand growth by 2030; the country will consume more electricity for its data centers than for aluminum, steel, cement, and chemical production combined.

Table 4. Data center electricity consumption by region, 2020–2030. Source: IEA (2025), Energy and AI.

Region	2020	2024	2030 (IEA)	Δ 24-30	Global Share 2024
United States	~120 TWh	~180 TWh	~420 TWh	+240 TWh	45%
China	~60 TWh	~102 TWh	~280 TWh	+175 TWh	25%
Europe (EU)	~45 TWh	~70 TWh	~115 TWh	+45 TWh	15%
World	~270 TWh	~415 TWh	~945 TWh	+530 TWh	100%

3.1.3 The Energy Factor as Competitive Advantage

The energy asymmetry constitutes a structural advantage for the United States. The IEA notes that natural gas currently provides 40% of US data center electricity, renewables 24%, nuclear 20%, and coal 15%. In Europe, the energy mix is more constrained by climate objectives and higher energy costs. Industrial electricity prices are typically 2 to 3 times higher in Europe than in the United States, before Power Purchase Agreement negotiations. As documented by the Federal Reserve Board (October 2025), there is a significant negative correlation between energy costs and AI adoption at the European enterprise level.[5]

France, however, has a specific asset in this context: its nuclear electricity mix (approximately 65–70% of production), which offers relatively inexpensive, decarbonized, and baseload-available electricity. RTE estimates an additional need of +10 GW for data centers in France by 2030, which remains compatible with installed capacity subject to investments in the transmission network.

3.1.4 Amplifying Effect: AI Robotics and Industrial Energy Demand

A factor still poorly quantified in the literature is the impact of AI robotics on energy demand. Autonomous robots require both embedded compute (edge computing) and centralized compute (cloud AI for training and model updates). The widespread deployment of AI automation in industry could add 20 to 30% additional energy demand in affected sectors, on top of data center growth. This factor is integrated as a sensitivity variable in our scenarios (Chapter V), as no precise and reliable estimate yet exists. The IEA nonetheless devotes a section of its report to the "Jevons rebound effect": even with significant efficiency gains (illustrated by the DeepSeek case), demand growth absorbs the gains.[6]

3.2 Semiconductor Market: AI-Driven Growth

3.2.1 Global Sales Trajectory 2020–2025

The global semiconductor market has followed a remarkable trajectory since 2020. After a record of $555.9 billion in 2022, the sector contracted by 8.2% in 2023 ($526.8B), before a spectacular rebound in 2024: global sales reached $630.5 billion (revised SIA/WSTS data, March 2025), a 19.1% year-on-year increase and a new all-time record.[7]

The year 2025 confirmed the acceleration. The SIA announced in February 2026 global sales of $791.7 billion in 2025, a 25.6% increase over 2024.[8] The fourth quarter of 2025 ($236.6B) was 37.1% higher than the same quarter in 2024. The SIA now projects sales close to $1 trillion in 2026, a symbolic threshold that seemed out of reach two years ago.

Two segments are driving this growth. Logic chips (processors, GPUs, ASICs) reached $301.9B in 2025 (+39.9%), becoming the top category by sales. Memory (DRAM, NAND) reached $223.1B (+34.8%). Together, these two segments represent over 66% of global sales and directly reflect AI infrastructure demand: servers, data centers, and HPC (High Performance Computing).

Table 5. Global semiconductor sales, 2020–2026. Sources: SIA/WSTS (Feb. 2025, Feb. 2026). The 2026 projection is based on the WSTS autumn 2025 forecast ($975.4B) and the SIA February 2026 statement.

Year	2020	2022	2023	2024	2025	2026 (proj.)
Sales ($B, SIA)	440	556	527	631	792	~975–1,000
Growth (%)	+6.8%	+3.3%	-8.2%	+19.1%	+25.6%	+23–26%

3.2.2 The Scope Question: SIA vs McKinsey

As noted in Chapter II, a notable methodological gap exists between SIA/WSTS data and McKinsey estimates. The SIA records semiconductor sales in the strict sense (components sold by manufacturers). McKinsey (January 2026) adopts an expanded scope including the value of captive designers (Apple, Amazon, Tesla, Google), i.e., the economic value of chips designed internally but manufactured by contract foundries. Under this expanded scope, McKinsey values the market at approximately $775B in 2024 and projects $1.5 to $1.8 trillion by 2030.[9] The expanded scope is more relevant to our analysis because it captures the full value of the AI chain, including hyperscaler design investments.

3.2.3 Regional Concentration of Manufacturing Capacity

The geographical distribution of foundry capacity constitutes a key vulnerability factor. Taiwan (TSMC) concentrates most of the leading-edge production (sub-7nm nodes). The United States represents approximately 10% of global installed capacity but is growing rapidly (the Chips Act has generated nearly $500B in announced private investment, with a goal of tripling US capacity by 2032).[10] The EU represents approximately 8% of global capacity, a figure the European Chips Act (€43B) aims to raise to 20% by 2030, a target deemed unrealistic by many observers. China, despite restrictions, is progressing from 21% toward a 30% target of installed capacity by 2030, primarily on mature nodes (28nm and above).

3.3 Geographical Distribution of Installed AI Compute

3.3.1 American Dominance: 75% of Global Performance

The most significant data point for our analysis is the geographical distribution of AI-dedicated computing capacity. The work of Epoch AI (Pilz et al., April 2025) provides the best available estimate, based on a census of 728 GPU clusters, covering 10 to 20% of estimated global capacity.[11]

The result is unequivocal: as of May 2025, the United States concentrates approximately 75% of global GPU cluster performance, China holds approximately 15%, and all other countries, including all of Europe, share the remaining 10%. The GeoCoded/Sanchez estimate (August 2025), which cross-references Epoch AI and Georgetown University, refines this figure: US 74.5%, China 14.1%, EU 4.8%.[12] In other words, the US/EU ratio in installed AI compute is approximately 15:1 — a gap of considerable magnitude.

Two dynamics explain this concentration. First, the private sector's share of global AI compute has grown from 40% in 2019 to 80% in 2025, and the technology companies investing massively in AI are almost exclusively American (Microsoft, Meta, Google, Amazon, xAI). Second, cluster sizes have exploded: systems with more than 10,000 chips were rare in 2019; in 2024, xAI deployed a 200,000-GPU Colossus cluster. The performance of leading AI supercomputers doubles every 9 months.[13]

3.3.2 Total Electrical Power of the AI Chip Fleet

Epoch AI (January 2026) estimates the total electrical capacity of the global AI chip fleet at approximately 30 GW at end-2025, comparable to the peak power consumption of New York State.[14] This estimate is based on quarterly AI chip sales from major manufacturers (Nvidia, AMD, Google TPU), multiplied by their nominal power (TDP) and a 2.5× factor for data center infrastructure. Global AI chip production doubles every 7 months, a pace that exceeds all prior forecasts. The five largest US AI server investors (Microsoft, Google, Meta, Amazon, xAI) announced $320B in investments in 2025, compared to $230B in 2024.

Table 6. Indicators of US dominance in AI compute (2024–2025). Sources: Epoch AI (2025), IEA (2025), GeoCoded/Sanchez (2025).

Indicator	United States	China	EU	US/EU Ratio
GPU cluster performance share	~75%	~15%	~5%	~15:1
Private sector share (2025)	~65% of global	~12%	~3%	~22:1
Data center consumption 2024	180 TWh	102 TWh	70 TWh	2.6:1
AI investment 2025	$320B (5 Big Tech)	n.a.	~€20B (EU total)	>15:1

3.3.3 Implications for the CACI Metric

These data allow an initial calibration of the CACI (defined in Chapter II). With an installed US/EU compute ratio of approximately 15:1, an EU/US energy cost of approximately 2–3×, and comparable GDP (US ~$28T vs EU ~$18T), the CACI(US)/CACI(EU) ratio falls in a range of 7 to 12 depending on assumptions for the human capital factor L(r). In other words, at comparable GDP and human capital units, American players have 7 to 12 times more effective compute than European players. Chapter V will project the evolution of this ratio under each scenario.

3.4 Timeline of American Regulatory Measures (2022–2026)

The sequence of American semiconductor and AI control measures constitutes the guiding thread of our hypothesis. Four phases are identifiable, marking a progressive escalation in the scope and intensity of restrictions.

Phase 1 — The Initial Shock (October 7, 2022)

The BIS (Bureau of Industry and Security) of the Department of Commerce publishes a final interim rule that radically transforms American export controls on semiconductors. The measures cover three components: (i) controls on advanced computing chips (GPUs above performance thresholds defined by TTP — Total Processing Performance), (ii) controls on semiconductor manufacturing equipment (SME, including EUV and DUV lithography), and (iii) restrictions on US person activities supporting advanced chip manufacturing in China.[15] The target is explicitly China: the stated objective is to prevent Chinese military modernization through AI compute access. The restrictions include three new Foreign Direct Product (FDP) rules that extend American jurisdiction to products manufactured outside the United States if US technologies are used in their production.

Phase 2 — Closing the Gaps (October 2023)

The BIS publishes two new interim rules that strengthen and broaden the October 2022 controls. Technical thresholds are adjusted to capture Nvidia chips specifically designed to circumvent restrictions (A800, H800). The geographic scope is extended to approximately 40 additional countries (Country Groups D:1, D:4, and D:5), with a differentiated licensing regime by country category.[16] Controls on manufacturing equipment are deepened. It is during this phase that European actors begin to perceive the indirect effects of restrictions, even though the EU is not the primary target.

Phase 3 — The AI Diffusion Rule (January 2025, Biden)

The Biden administration publishes in January 2025 the AI Diffusion Rule, which represents a qualitative shift. For the first time, restrictions apply not only to physical chips but also to AI model weights and cloud compute access. The rule classifies 120 countries into three categories: (i) trusted allies (broad access), (ii) intermediate countries (quotas), (iii) embargoed countries.[17] European reactions are strong: the European Parliament raises alarm about restrictions threatening the EU's ability to train models on its AI Factories. France and Germany are classified as trusted partners, but other member states face caps on importable GPU volumes.

Phase 4 — The Trump Break: Section 232 and Explicit Protectionism (January 2026)

On January 14, 2026, President Trump signs Proclamation 11002, invoking Section 232 of the Trade Expansion Act of 1962.[18] This action marks a qualitative break with previous phases, for three reasons.

First, the legal instrument changes. Phases 1 to 3 fell under export controls (export regulation, defensive national security logic). Section 232 is a tariff instrument (import duties), whose logic is protectionist: it aims to protect domestic production, not merely restrict an adversary's access.

Second, the tariff creates an explicit competitive advantage for American companies. The 25% tariff hits advanced GPUs (H200, MI325X) imported unless destined for American domestic uses: US data centers, R&D, startups, public sector, industrial and non-data-center consumer applications. Concretely, an American company using these chips on US soil does not pay the tariff; a foreign company importing the same chips for re-export to China pays 25%.[19]

Third, the proclamation announces an imminent expansion. The text provides that the Secretary of Commerce and USTR negotiate within 90 days with semiconductor-producing countries, and that broader tariffs, accompanied by a tariff offset program for companies investing in US production, could be imposed. The Secretary of Commerce must provide by July 2026 a report on the semiconductor market used in American data centers.[20]

Table 7. Timeline of American regulatory measures on semiconductors and AI (2022–2026). Sources: BIS, White House, Pillsbury Law (2026), Gibson Dunn (2026).

Date	Admin.	Measure	Scope / Target
Oct. 2022	Biden	BIS export controls: advanced GPUs, SME, US persons	China (military)
Oct. 2023	Biden	Threshold reinforcement + 40-country extension + HBM	China + 40 countries (D:1/D:4/D:5)
Dec. 2024	Biden	Wave 3: 24 SME types, HBM, 140 entities, ECAD	China (full chain)
Jan. 2025	Biden	AI Diffusion Rule: model weights, cloud, 3-tier countries	120+ countries (incl. EU effects)
Jul. 2025	Trump	America's AI Action Plan: deregulation, US compute	US (domestic strategy)
Sep. 2025	Trump	Announcement: China sales authorized against 25% revenue	China (monetization)
Jan. 2026	Trump	Section 232: 25% tariff advanced GPUs + BIS China license	Global (US domestic exemption)

3.4.5 Interpretation: From Defensive Control to Offensive Protectionism

The 2022–2026 sequence reveals a qualitative transformation of American policy. The Biden phases (2022–2025) follow a denial strategy logic:

preventing a designated adversary (China) from accessing key technologies, within a multilateral framework (coordination with Japan, Netherlands). The Trump phase (2025–2026) adds a capture strategy logic: generating revenue (25% tariff), prioritizing American companies (domestic exemptions), and using compute access as a negotiation lever with third countries.

This is precisely the transformation that validates the central hypothesis of our study: American technology protectionism is no longer limited to denying access to an adversary; it actively constructs a structural competitive advantage for American companies. The Section 232 domestic exemptions mean that, for the first time, the cost of accessing cutting-edge compute is legally differentiated by the user's nationality. Even though direct effects on the EU remain limited in January 2026 (the tariff primarily targets re-exportation to China), the proclamation opens the way for an expansion that could directly affect Europe — precisely scenarios B, C, and D of our analysis.

3.5 Diagnostic Synthesis: The "Predetermined Elements" of 2026

The four dimensions of the diagnosis converge toward a coherent picture that structures the scenarios in Chapter V:

(1) AI compute demand is growing exponentially (chip sales doubling in two years, energy consumption doubling projected over six years), and this growth shows no signs of slowing.

(2) This demand is structurally concentrated in the United States (75% of installed compute, 45% of data center energy consumption, >80% of private investment), a concentration that is increasing over time.

(3) Europe starts from a structurally deficit position (~5% of compute, ~15% of consumption, energy costs 2–3× higher, 72–80% dependency on US hyperscalers for AI cloud), with investment plans (Chips Act, AI Factories, SNIA) whose scale remains at least 10× lower than American private investments.

(4) The American regulatory framework crossed a qualitative threshold in January 2026 with the shift from export controls to Section 232 tariffs, creating a legal mechanism for differentiating compute access costs by nationality. The proclamation explicitly signals the possibility of expansion, whose scope and timeline will depend on political variables — precisely the critical uncertainties that our scenarios explore.

The following chapter (Chapter IV) analyzes the mechanisms by which this compute asymmetry translates into measurable competitive advantage for American companies.

Notes

[1] IEA (2025), Energy and AI, Paris, IEA. https://www.iea.org/reports/energy-and-ai. The 415 TWh figure includes all data centers (cloud, enterprise, colocation), not just AI-dedicated data centers. The IEA notes that consumption has grown by 12% per year since 2017.

[2] IEA (2025), op. cit., chapter "Energy demand from AI." European Commission (November 2025), "In Focus: Data Centres — An Energy-Hungry Challenge," energy.ec.europa.eu. The 70 TWh figure for the EU is an IEA estimate in the absence of precise consumption data.

[3] IEA (2025), op. cit.; Carbon Brief (September 2025), "AI: Five Charts that Put Data-Centre Energy Use — and Emissions — into Context." The 35–50% range comes from a report prepared for the IEA (Kamiya & Coroamă, 2025, IEA-4E).

[4] IEA (2025), op. cit. The IEA further notes that in its "Headwinds" scenario, global consumption would only reach 790 TWh, 40% less than the central scenario, illustrating the scale of uncertainty.

[5] Federal Reserve Board (October 2025), State of AI Competition in Advanced Economies. The report documents a significant correlation between industrial energy costs and AI adoption rates, particularly for large European companies.

[6] IEA (2025), op. cit., chapter "Case study: DeepSeek and efficiency gains." The IEA concludes that algorithmic efficiency gains, while significant, tend to be absorbed by demand growth ("Jevons paradox").

[7] SIA (February 2025, revised March 2025), "Global Semiconductor Sales Increase 19.1% in 2024." The initial figure of $627.6B was revised to $630.5B by the WSTS.

[8] SIA (February 2026), "Global Annual Semiconductor Sales Increase 25.6% to $791.7 Billion in 2025." Q4 2025 sales ($236.6B) were 37.1% higher than Q4 2024, marking a notable acceleration.

[9] McKinsey (January 2026), "Hiding in Plain Sight: The Semiconductor Industry's Expanding Perimeter." The gap with the SIA is explained by the inclusion of captive designers and vertically integrated fabless operators.

[10] SIA (July 2025), State of the U.S. Semiconductor Industry Report. The SIA notes that the Chips & Science Act has generated nearly $500B in announced private investment and is expected to create or support more than 500,000 jobs.

[11] Pilz, K.F., Rahman, R., Sanders, J. & Heim, L. (2025), "Trends in AI Supercomputers," arXiv:2504.16026. The dataset covers 728 clusters, of which 501 have been operational since 2019, representing 10–20% of estimated global capacity.

[12] Sanchez, C. (2025), "GeoCoded Special Report: State of Global AI Compute (2025 Edition)," Sanchez.vc. The estimate cross-references Epoch AI and Georgetown University data. The author notes that coverage varies and actual shares could differ by 5 percentage points.

[13] Pilz et al. (2025), op. cit. Leading AI supercomputer performance doubles every 9 months, driven by a 1.6×/year increase in chip count and a 1.6×/year increase in per-chip performance.

[14] Epoch AI (January 2026), "Global AI Power Capacity Is Now Comparable to Peak Power Usage of New York State." The 30 GW estimate is based on quarterly AI chip sales × TDP × 2.5× infrastructure factor.

[15] BIS (October 7, 2022), "Commerce Implements New Export Controls on Advanced Computing and Semiconductor Manufacturing Items to the People's Republic of China," Federal Register. See also GAO (2025), "Export Controls: Commerce Implemented Advanced Semiconductor Rules," GAO-25-107386.

[16] Skadden (October 2023), "BIS Updates October 2022 Semiconductor Export Control Rules." The new rules notably capture the A800 and H800, Nvidia chips specifically designed to fall just below the October 2022 thresholds.

[17] Carnegie Endowment (May 2025), Winter-Levy, H. & Phillips-Robins, A., "What Is the AI Diffusion Rule?" The AI Diffusion Rule remained in effect only briefly before being partially replaced by the Trump approach, but it established the precedent of control over models and cloud, not just physical chips.

[18] White House (January 14, 2026), Presidential Proclamation 11002, "Adjusting Imports of Semiconductors, Semiconductor Manufacturing Equipment, and Their Derivative Products into the United States." The proclamation invokes Section 232 of the Trade Expansion Act of 1962 (19 U.S.C. 1862).

[19] Pillsbury Law (January 2026), "Trump Admin Targets Advanced AI Semiconductors, Defers Broader Tariffs." The legal analysis notes that the combination of Section 232 and the BIS rule of January 15, 2026 operationalizes the September 2025 announcement that the US government would collect 25% of AI chip sales to China.

[20] White House (January 14, 2026), Fact Sheet: "President Donald J. Trump Takes Action on Certain Advanced Computing Chips to Protect America's Economic and National Security." Gibson Dunn

(January 2026), "The Trump Administration's New Tariffs on and Export Licensing Requirements for Advanced Semiconductors."