

PitchBook Data, Inc.

Nizar Tarhuni Executive Vice President of Research and Market Intelligence

Paul Condra Global Head of Private Markets Research

James Ulan Director of Emerging Technology Research

Institutional Research Group

Analysis



Dimitri Zabelin
Senior Research Analyst,
AI and Cybersecurity
dimitri.zabelin@pitchbook.com

Data

Oscar Allaway
Data analyst

Owen Walker
Intern

pbinstitutionalresearch@pitchbook.com

Publishing

Designed by **Jenna O'Malley**

Published on October 27, 2025

Contents

Executive summary	1
Sovereign AI	2
The infrastructure super-cycle	4
The geopolitical divide	9
Regional strategies	17
Industry intelligence	29
Macroeconomic and market forces	32
Investor playbook and outlook	33
Glossary	37

EMERGING TECH RESEARCH

Sovereign AI: The Trillion-Dollar Frontier

PitchBook is a Morningstar company providing the most comprehensive, most accurate, and hard-to-find data for professionals doing business in the private markets.

Executive summary

This report provides a comprehensive analysis of the sovereign AI market, a defining technological and geopolitical trend of the mid-2020s. We assess that the pursuit of national autonomy in artificial intelligence is no longer a niche policy objective but the central pillar of economic and national security strategy for developed and emerging nations alike. This shift is catalyzing a historic, multitrillion-dollar redirection of state-backed capital into the AI technology stack, creating a durable, long-term investment super-cycle. Our analysis indicates that this market is not monolithic; it is bifurcating into two distinct, competing ecosystems led by the US and China, compelling other nations to make strategic alignment choices that will dictate capital flows and technological standards for the next decade.

The primary driver of this trend is an intensifying global rivalry, compelling governments to secure their own computational infrastructure, data, and AI models to avoid dependency on nondomestic powers. This imperative is creating unprecedented demand guarantees for companies across the AI value chain. We have identified four critical layers for investment:

1. Semiconductors, where US export controls are creating a protected market for domestic Chinese players and a premium for Western-aligned suppliers.
2. Infrastructure, where sovereign wealth funds (SWFs) in the Middle East and state-backed consortia are underwriting hyperscale datacenter build-outs, de-risking growth for the entire supply chain from power and cooling to interconnects.
3. Foundation models, where capital is concentrating in a few US-based leaders who are becoming quasi-sovereign entities through deep partnerships with allied nations.
4. Applications, a nascent but high-potential area where sovereign-specific use cases in defense, healthcare, and public services will emerge.

For investors, the sovereign AI landscape offers generational opportunities but is fraught with geopolitical risk. The US ecosystem, anchored by legislative support like the CHIPS and Science Act and corporate champions like NVIDIA and OpenAI, remains the dominant destination for private capital due to its transparency and

innovation leadership. Conversely, China's state-directed model offers a deep but increasingly isolated market, with significant opportunities in domestic substitution. The most dynamic near-term growth is emanating from the Middle East, where SWFs from the United Arab Emirates to Saudi Arabia are deploying tens of billions of dollars to rapidly build domestic capacity in partnership with US firms.

Our key recommendation is to adopt a geopolitically aware investment thesis. Portfolios should seek exposure to the foundational infrastructure layers of the AI stack, as these are the immediate, primary beneficiaries of state spending. Specifically, companies providing enabling technologies for datacenters such as advanced packaging, optical interconnects, and next-generation cooling, are poised for outsized returns. As the market matures, the focus will shift from training to inference, creating a second wave of opportunities in specialized hardware and efficient software. Navigating this complex environment requires a deep understanding of the policy levers, national champions, and geopolitical fault lines shaping this new world order in technology.

Sovereign AI

The geopolitics of AI

The concept of sovereign AI has rapidly evolved from a theoretical construct into the single most powerful force shaping the global technology landscape. At its core, sovereign AI is a nation's capacity to independently develop, deploy, and govern artificial intelligence systems. This entails controlling the entire technology stack, from the silicon upon which models are trained to the data that gives them intelligence and the applications that deliver economic and strategic value. What was once a concern for only the largest geopolitical powers has now become a strategic imperative for nearly every nation with aspirations of economic relevance and national security in the 21st century. The race for AI supremacy is, fundamentally, a race for sovereignty itself.

The primary impetus for this global movement is the recognition that dependence on nondomestic technology, particularly in a domain as foundational as AI, creates unacceptable strategic vulnerabilities. The intensifying geopolitical rivalry between the US and China has acted as a powerful accelerant. Smaller and middle-power nations, observing the weaponization of supply chains and the imposition of technology export controls, have concluded that technological autonomy is synonymous with national security.

This realization has triggered a wave of state-directed industrial policy and capital allocation on a scale not seen since the Cold War space race. Governments are no longer passive observers of technological progress; they are now the largest and most motivated customers for the core components of the AI revolution.

This shift has profound implications for private markets. Sovereign AI initiatives function as a massive, state-underwritten demand guarantee for the entire technology ecosystem. Private markets have benefited from two types of government support. First, they receive direct state-backed funding. Second, they benefit indirectly from infrastructure strategies, where public companies partner

with government to build the core AI infrastructure. Private market players are then able to create on top of this foundation, thereby strengthening the ecosystem of innovation, funding, and deployment.

North America, Europe, Asia, and the Middle East are the primary regional players in sovereign AI, with each government leveraging a unique set of policy tools. The three capital vectors are sovereign wealth funds, legislative funding, and bureaucratic allocation. Each lever is optimized—or constrained—by the political system of each country in their respective region.

This capital is not being deployed speculatively; it is targeted, long-duration investment aimed at constructing national champion companies, onshoring critical supply chains, and building fortress-like data infrastructure. For investors, this creates a durable, multidecade tailwind, transforming the risk calculus for capital-intensive sectors like semiconductors, datacenters, and foundational model research.

The impact of this trend ripples through the entire market structure, affecting a wide range of stakeholders. For end customers, particularly enterprises in regulated industries like finance, healthcare, and defense, sovereign AI means the emergence of national or regional cloud platforms that comply with stringent data residency and security requirements. This simplifies compliance but may limit access to the most cutting-edge global AI innovations.

For technology vendors, the market is bifurcating. Companies must now navigate a complex patchwork of national procurement standards, security protocols, and geopolitical alignments. Success is no longer determined solely by technological merit but by a company's ability to position itself as a trusted partner to a sovereign government and its chosen geopolitical bloc.

The industry dynamics are being fundamentally reshaped. The traditional dominance of US hyperscalers is being challenged by a new class of sovereign-backed cloud providers and specialized AI cloud companies. In the semiconductor industry, US export controls are simultaneously hobbling China's advance and creating a protected, high-growth market for domestic Chinese firms while reinforcing the market power of Western-aligned players like NVIDIA and TSMC. The foundation model layer is consolidating around a few key US players (OpenAI, Anthropic, and others) who are acting as quasi-sovereign entities, forging deep, exclusive partnerships with allied nations to become the standardized platform for their ecosystems.

Sovereign AI constraints

However, the path to sovereign AI is fraught with significant risks and challenges that threaten to derail national ambitions. The most acute bottleneck is the global shortage of elite AI talent. The concentration of top researchers and engineers in a few US technology hubs creates an intense competition for human capital that many nations cannot win on salary alone. Attracting and retaining talent requires building vibrant domestic research ecosystems, which is a multigenerational challenge.

Second, the costs associated with building a competitive AI stack are astronomical. A state-of-the-art semiconductor fabrication plant costs upward of \$20 billion, while training a frontier AI model can cost over \$1 billion in compute time alone. Building the supporting datacenter infrastructure requires tens of billions more. Only the wealthiest nations and sovereign wealth funds can afford to compete at the highest level, creating a new global divide between the “AI-haves” and “AI-have-nots.”

The energy requirements for large-scale AI datacenters are staggering. According to the International Energy Agency’s Electricity 2024 report, global electricity consumption from datacenters, artificial intelligence, and cryptocurrency operations could double by 2026.¹ In many regions, electrical grid infrastructure is insufficient to meet these demands, making energy availability a critical constraint on sovereign AI deployment. Companies and nations that can overcome this bottleneck, whether through alternative energy sources such as Crusoe’s gas-flare-powered datacenters or through more efficient hardware, will hold a decisive competitive advantage in scaling AI infrastructure.

Finally, the rapid pace of technological innovation creates a significant risk of obsolescence. A multibillion-dollar investment in today’s leading-edge GPU technology could be rendered uncompetitive by a breakthrough in a new architecture (for example, optical or analog computing) within a few years. This forces nations into a perpetual, high-stakes technology race where a single misstep can result in billions of dollars in stranded assets.

For investors, this complex interplay of geopolitical ambition, massive capital flows, and significant execution risk creates a unique opportunity set. The winners will be those who can identify the critical, enduring bottlenecks in the sovereign AI value chain and invest in the companies providing the “picks and shovels” for this global technological build-out. The key is to understand that this is not merely a technology cycle; it is the construction of a new, geopolitically defined world order in technology, and the foundational layers of that order are being laid today.

The infrastructure super-cycle

The inference imperative: Capturing value in the next phase of sovereign AI

The first wave of sovereign AI investment has been defined by a singular focus: securing massive amounts of computation power, or “compute,” for training foundational models. This “training-first” approach, exemplified by the global rush to procure NVIDIA H100 and B100 GPUs, is a necessary first step. However, it represents only a fraction of the long-term economic opportunity. The enduring value and the majority of future infrastructure spending will be dedicated to inference—the process of using a trained model to generate predictions, answers, and content in real-world applications. For investors, understanding this imminent shift from training to inference is critical to positioning capital for the next decade of growth.

¹: “Electricity 2024 Executive Summary,” International Energy Agency, January 24, 2024.

Training a frontier model is a massive but infrequent event, costing hundreds of millions of dollars. In contrast, inference happens billions or trillions of times a day as users interact with AI services. The technical and economic requirements for inference are fundamentally different from those for training, creating opportunities for a new class of specialized companies to challenge incumbent dominance.

- **The total cost of ownership (TCO) over performance:** Training prioritizes raw computational power (measured in flops) to minimize the time it takes to train a model, with cost as a secondary concern. Inference prioritizes a model's performance per watt and performance per dollar to minimize the operational cost of running it at scale. For a sovereign state running nationwide AI services, a 20% improvement in TCO can translate to billions of dollars in annual savings.
- **Latency over throughput:** Training is a batch process where throughput is key. Inference, particularly for real-time applications like chatbots or autonomous systems, is latency-critical. A 100-millisecond delay can render an application unusable. This requires hardware and software stacks optimized for immediate response times.
- **Hardware diversity over GPU monoculture:** While GPUs are effective for both training and inference, their architecture is not always optimal for the latter. This has created a massive market opportunity for specialized inference accelerators. Startups like Groq (with its Language Processing Units), d-Matrix (with its chiplet-based architecture), and numerous others are designing chips from the ground up to maximize efficiency and minimize latency for inference workloads. These companies are gaining traction with sovereign and enterprise customers seeking a more cost-effective alternative to a pure-GPU strategy.

As sovereign nations complete the initial training of their national models, their focus will pivot to deploying these models to their populations and industries. This will trigger a massive procurement cycle for inference-optimized infrastructure. This includes not only specialized chips but also the surrounding ecosystem: high-speed networking to connect them, efficient power delivery and cooling, and sophisticated software to manage and orchestrate workloads.

For investors, the inference imperative signals a clear action plan. While exposure to NVIDIA remains crucial, portfolios must be diversified to include the emerging leaders in the inference space. The private companies developing novel accelerator architectures and the public companies providing the essential networking and infrastructure components for this build-out are poised to become the next big winners in the sovereign AI market.

Investment criteria for sovereign AI

Given the rapidly shifting geopolitical and regulatory landscape, traditional due diligence must now account for factors such as state alignment, policy exposure, and long-term infrastructure dependencies. The following priorities outline a framework for evaluating opportunities and risks in this evolving environment.

- **Map geopolitical alignments:** For every target investment, evaluate key customers and supply chain partners to determine whether the company operates within the US or Chinese ecosystem. Understanding this alignment helps assess the likelihood of exposure to future sanctions or export restrictions.
- **Evaluate technical moats in enabling technologies:** In the competitive hardware segment, focus on companies with defensible intellectual property in interconnect fabrics, thermal management, and advanced packaging. These areas offer more durable differentiation than model architectures and are less vulnerable to rapid displacement.
- **Assess customer concentration risk:** Large sovereign contracts can signal strong demand but also heighten dependency on a single buyer. Companies with a balanced mix of sovereign and enterprise clients demonstrate greater revenue stability and policy resilience.
- **Diligence the inference roadmap:** Review hardware product strategies to determine preparedness for the fast-growing inference market. Firms concentrating solely on training hardware may lack long-term positioning, while those transitioning toward inference are aligned with future compute demand.
- **Monitor sovereign policy shifts:** Establish a process for tracking national AI strategies, subsidy allocations, and export control updates from the US Department of Commerce. These developments serve as early indicators of capital flows, procurement trends, and changing market access.

Together, these priorities provide a structured approach for navigating the sovereign AI market with greater precision and foresight. As national interests continue to shape capital allocation, disciplined diligence will determine who captures durable value in this new era of state-backed innovation.

Sizing the sovereign AI market

The total addressable market (TAM) for sovereign AI is a composite of multiple established and emerging technology sectors, supercharged by state-directed spending. While a single, universally agreed-upon TAM is difficult to define, leading market intelligence firms have begun to quantify the opportunity, consistently pointing to a multitrillion-dollar market by the end of the decade. The data from Q3 2025 underscores the sheer scale of this state-catalyzed build-out.

Industry forecasts indicate that sovereign and AI-linked infrastructure spending is accelerating at an unprecedented pace. McKinsey projects global datacenter capital expenditures to reach \$6.7 trillion by 2030, with approximately \$5.2 trillion allocated to AI-optimized facilities designed to support large-scale model training and inference.² Oppenheimer estimates the sovereign AI market opportunity at roughly \$1.5 trillion, reflecting a surge in national investments to localize compute capacity

2: "The Cost of Compute: A \$7 Trillion Race to Scale Datacenters," McKinsey & Company, Jesse Noffsinger, Mark Patel and Pankaj Sachdeva, et al., April 28, 2025.

and strengthen chip and software supply chains.³ Fortune Business Insights expects the sovereign cloud market to grow from \$155 billion in 2025 to \$824 billion by 2032, representing a compound annual growth rate of about 27%.⁴ Collectively, these figures highlight the rapid institutionalization of digital sovereignty as a central pillar of the global AI economy.

IDC and Dell'Oro data point to a structural shift in AI infrastructure toward specialized compute. Dell'Oro projects global datacenter CAPEX will rise at a 21% CAGR through 2029 to roughly \$1.2 trillion, with GPUs and custom AI accelerators already representing about one third of total spend.^{5, 6} IDC forecasts that by 2028, accelerated servers containing GPUs or custom chips will exceed 75% of AI server infrastructure spending, growing at a five year CAGR of 42%.⁷ These trends show how nations and hyperscalers are investing in sovereign-grade infrastructure by expanding domestic datacenters, deploying accelerator clusters, and funding local chip manufacturing to secure localized compute capacity and maintain greater control over AI resources.

Industry research suggests the financial burden of sovereign AI development extends well beyond compute infrastructure. Brookfield estimates more than \$500 billion in capital may be required over the next decade to expand power generation and transmission capacity for AI demand.⁸ Goldman Sachs projects that datacenter power requirements will rise roughly 160% by 2030, with as much as \$5 trillion in combined digital infrastructure and energy investment needed to sustain growth.^{9, 10, 11} Meketa's June 2025 whitepaper highlights the necessity of coordinated buildouts spanning chip fabrication, datacenters, and utility systems, while Surocap forecasts total investment across AI datacenters and supporting energy infrastructure could reach \$6 trillion to \$8 trillion by the end of the decade. Collectively, these projections illustrate how energy and grid modernization are becoming integral components of national AI and digital sovereignty strategies.

PitchBook research highlights the growing influence of sovereign capital in private markets. State-backed investors and sovereign wealth funds are playing a larger role in late-stage AI and infrastructure deals, helping stabilize demand and accelerate capital formation.

Collectively, these forecasts paint a clear picture: the sovereign AI trend is not an incremental market opportunity but a foundational, state-driven economic shift. The alignment of national security interests with massive capital deployment is creating one of the largest and most durable technology investment cycles in history.

3: "Why Oppenheimer Is Excited About Nvidia's Potential Gains From Sovereign AI Demand," Investopedia, Andrew Kessel, June 16, 2025.

4: "Sovereign Cloud Market Size," Fortune Business Insights, September 29, 2025.

5: "Datacenter Capex to Grow at 21 Percent CAGR Through 2029, According to Dell'Oro Group," Dell'Oro Group, August 6, 2025.

6: "Market for AI Accelerators to Reach \$382 Billion by 2029, According to Dell'Oro Group," Dell'Oro Group, February 6, 2025.

7: "Artificial Intelligence Infrastructure Spending to Surpass the \$200Bn USD Mark in the Next 5 years, According to IDC," IDC, February 18, 2025.

8: "Building the Backbone of AI," Brookfield, August 2025.

9: "The Building Blocks Behind AI's Next Wave," Goldman Sachs, n.d., accessed October 20, 2025.

10: "How AI Is Transforming Datacenters and Ramping Up Power Demand," Goldman Sachs, August 29, 2025.

11: "Powering the AI Era," Goldman Sachs, 2025.

Mapping the sovereign AI value chain

To effectively allocate capital within the sovereign AI landscape, investors must move beyond a monolithic view of AI and adopt a structured framework that dissects the technology stack into its core layers. Each layer presents distinct investment opportunities, competitive dynamics, and geopolitical sensitivities. We analyze the sovereign AI value chain as four interdependent layers, each of which is a primary target for state-directed capital.

Layer 1: semiconductors and core components

At the base of the sovereign AI stack lies the semiconductor foundation, where intelligence is forged in silicon. It includes the designers and manufacturers of GPUs, ASICs, high-bandwidth memory (HBM), and advanced interconnects. This layer is the most geopolitically sensitive, as demonstrated by stringent US export controls. Sovereign investment here is primarily defensive, focused on onshoring manufacturing and fostering domestic champions to ensure security of supply. The US CHIPS and Science Act and China's Big Fund III, a massive \$47.5 billion vehicle confirmed in May 2024, are textbook examples.

Serving as the system's critical chokepoint, this tier concentrates technological and geopolitical leverage. Dominant incumbents in the Western bloc (NVIDIA, TSMC, and SK Hynix) command immense pricing power and are direct beneficiaries of sovereign build-outs. In the private markets, the most compelling opportunities lie in "pick and shovel" plays that are difficult to replicate: advanced packaging firms that bond chiplets together and developers of novel interconnect technologies like optical I/O that are critical for scaling AI systems.

Layer 2: Cloud and datacenter infrastructure

Above the silicon sits the infrastructure layer, where compute is delivered as a service through vast networks of datacenters and hyperscale facilities. It encompasses the entire supply chain for building and operating hyperscale datacenters: from power and cooling systems to servers, networking gear, and the physical facilities themselves. Sovereign capital is flowing aggressively into this layer to ensure data residency and control over national compute resources. The multibillion-dollar datacenter campuses being built in the UAE (via MGX Fund Management) and Saudi Arabia (via the Public Investment Fund) are prime examples of state-underwritten infrastructure projects.

This stage of the stack is where sovereign spending is the most direct and predictable. Private equity firms specializing in digital infrastructure (for example, Blue Owl and Blackstone) are scaling rapidly by partnering with sovereign funds. For venture investors, the opportunity is in next-generation enabling technologies that solve the key bottlenecks of power consumption and heat dissipation, such as liquid cooling solutions and AI-driven datacenter management software.

Layer 3: Foundation models and platforms

Here, raw compute is translated into general-purpose intelligence through large language and multimodal models that power the broader ecosystem. It is dominated by a small number of LLMs and multimodal model developers who provide the core platforms upon which applications are built. Nations are pursuing two strategies here: either partnering deeply with a leading US provider (for example, the UAE's partnership with Microsoft/OpenAI) to standardize on a Western platform, or investing heavily to build an indigenous "national model" (for example, France's support for Mistral AI and India's for Sarvam AI).

The platform layer exhibits powerful network effects, leading to a likely consolidation around a few trusted, geopolitically aligned winners. Direct investment in the top US players (OpenAI and Anthropic) offers exposure to the heart of the Western AI ecosystem as they become quasi-sovereign partners. For VCs, the opportunity lies in backing emerging regional champions in Europe and Asia that can capture the "home market" through linguistic, cultural, and regulatory advantages.

Layer 4: Vertical applications and services

At the top of the value chain, foundational models are fine-tuned and integrated to solve domain-specific, high-value problems across defense, public administration, and healthcare. In the sovereign context, these are applications built for core government functions: national security and defense, public health analytics, tax fraud detection, and critical infrastructure management. This layer is the ultimate goal of all sovereign investment: to translate technological capacity into tangible economic and societal benefits.

The vertical applications layer is currently the most nascent but holds immense long-term potential. The key is to identify startups building applications for regulated industries where sovereign procurement is the primary go-to-market channel. Companies with deep domain expertise in defense, intelligence, and public administration, and that can obtain the necessary security clearances, are positioned to win large, multi-year government contracts as sovereign platforms come online.

The geopolitical divide

The US-China technological fault line

The global sovereign AI landscape is not a unified, open market; it is a geopolitical battleground defined by the intensifying technological rivalry between the US and China. These two powers are actively constructing parallel, often incompatible, technology ecosystems. Every other nation is increasingly forced to make a strategic choice: align with the US-led bloc, characterized by a market-driven approach with strong corporate champions, or the China-led bloc, defined by state-directed industrial policy and a focus on self-sufficiency. This bifurcation is the single most important factor for investors to understand, as it dictates market access, supply chain stability, and the flow of capital.

The US-led ecosystem: The “hub-and-spoke” model

The American ecosystem operates on a hub-and-spoke model, with US corporations and technology standards positioned at the center. This bloc is defined by deep public-private partnerships and strategic coordination with other market-based economies.

Corporate champions sit at the core of the model. NVIDIA leads in hardware, Microsoft and OpenAI dominate the platform layer, Google spans both platforms and research, Oracle provides sovereign cloud infrastructure, and TSMC serves as the manufacturing ally that anchors the semiconductor supply chain.

Strategic levers underpin the system’s global reach. The US leverages capital market depth, venture capital dynamism, and soft power through technological leadership. These strengths are reinforced by hard power policies such as the CHIPS and Science Act, which subsidizes domestic and allied manufacturing, and export controls managed by the US Bureau of Industry and Security (BIS) that restrict China’s access to chokepoint technologies.

Capital flows define the hub-and-spoke structure. Funding originates from US venture capital firms, corporate venture arms, and allied sovereign wealth funds such as the UAE’s MGX and Singapore’s Temasek. These flows concentrate around dominant US platforms and their supply chains. The implicit understanding is that allied nations gain access to leading technology in exchange for alignment with the US stack and continued support for its key players. The UAE’s substantial NVIDIA chip allocation demonstrates how this arrangement reinforces demand for US firms and sustains a self-reinforcing cycle of investment and technological adoption.

The China-led ecosystem: The “walled garden” model

China’s ecosystem functions as a state-designed walled garden built for self-reliance and protection from US sanctions. The overarching goal is to establish a fully domestic, vertically integrated technology stack that operates independently of nondomestic suppliers and is insulated from external political pressure.

Corporate champions define each layer of this system. Huawei and its chip design arm HiSilicon lead in hardware, SMIC anchors domestic semiconductor manufacturing, and Baidu, Alibaba, and Tencent dominate the platform and cloud layers. Together, they form the backbone of China’s sovereign technology network and enable the state to maintain control over critical digital infrastructure.

Strategic levers are driven primarily by industrial policy. The government channels significant capital through state-directed mechanisms such as the Big Fund III, targeting strategic sectors like semiconductors, AI, and advanced manufacturing. These efforts are reinforced by procurement mandates that prioritize domestic suppliers and by data security regulations that limit nondomestic platform access. China’s dual circulation strategy underpins this approach, emphasizing domestic consumption and technological self-reliance while selectively engaging global markets to reduce external vulnerability and sustain long-term growth.

Capital flow remains predominantly domestic. Funding is sourced from state-backed guidance funds, policy banks, and the corporate capital expenditures of major technology firms that operate under state influence. The objective is import substitution across the entire value chain, from semiconductor manufacturing equipment to large language models, reducing reliance on Western suppliers and ensuring long-term national resilience.

Caught in the middle

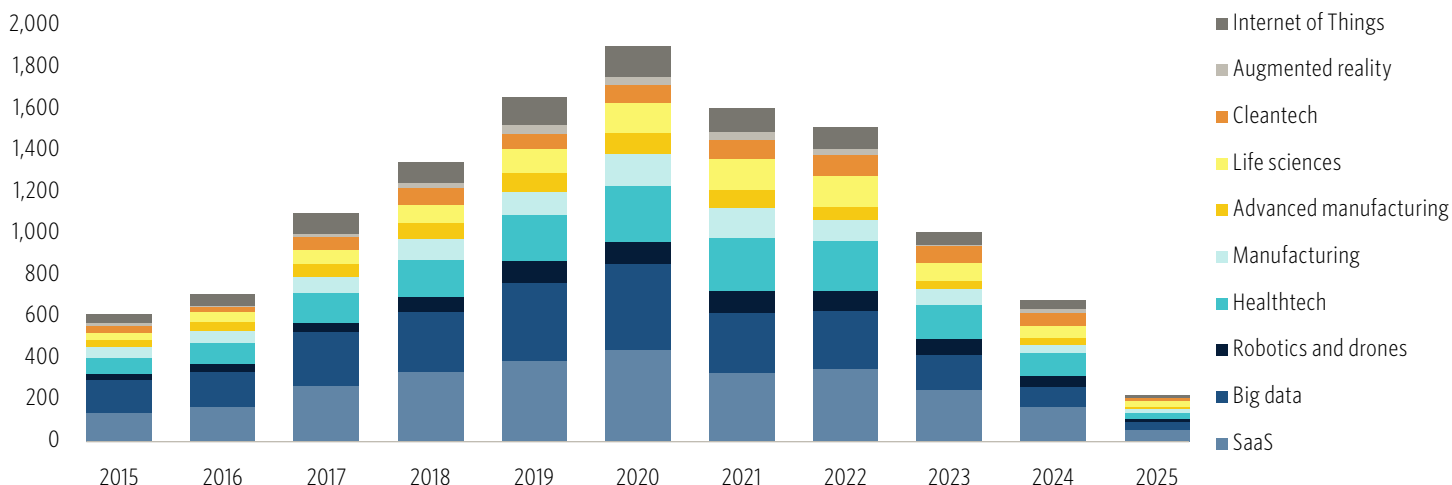
Nations in the Middle East, Southeast Asia, and South America face a critical strategic dilemma. Full alignment with the US provides access to superior technology but risks economic blowback from China, often a larger trading partner. Conversely, adopting Chinese technology offers a lower-cost alternative but risks future sanctions from the US and potential cybersecurity vulnerabilities.

The emerging strategy for capital-rich nations like Saudi Arabia and the UAE is a pragmatic “barbell” approach. They are partnering with US leaders for their core, mission-critical sovereign AI infrastructure (for example, the Stargate Project) to secure access to state-of-the-art advancements. Simultaneously, they continue to procure Chinese technology for less sensitive applications like smart city infrastructure and consumer electronics. For investors, this means the most valuable, high-margin opportunities will remain concentrated in the US ecosystem, as this is where nations are placing their most strategic bets.

Public sector capital and grant activity

Grant allocation serves as a proxy for state conviction, signaling where public capital is building the infrastructure likely to anchor future private investment. Global grant activity in AI & machine learning (ML) has concentrated heavily in a handful of verticals over the past decade. From 2015 through 2025, software as a service (SaaS) and Big Data emerged as the clear leaders by deal count, together attracting nearly 5,500 grant-backed projects. SaaS consistently held the top spot in nearly every year of the period, finishing with 2,920 deals in total, while Big Data followed closely at 2,541. The dominance of these subsegments reflects how governments and institutions prioritized broad, horizontal platforms that could enable enterprise adoption at scale. This data provides an essential leading indicator for investors, revealing where governments see the most utility and potential for widespread economic impact, thereby de-risking early-stage investment in these categories.

AI & ML grant count by top 10 VC verticals

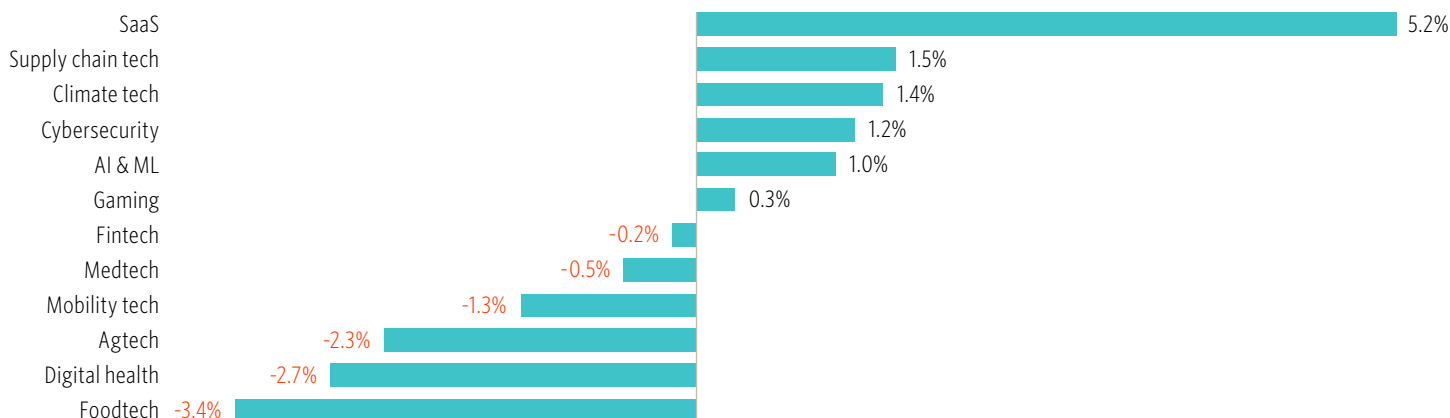


Source: PitchBook • Geography: Global • As of August 12, 2025

A second tier of verticals trails by a wide margin over the same ten-year window. Healthtech is the clear third-place finisher with 1,803 total deals, underscoring persistent policy interest in applying AI to diagnostics, drug discovery, and digital health platforms. Beyond healthtech, life sciences, manufacturing, and the Internet of Things each attracted fewer than 1,000 total deals. Meanwhile, robotics & drones, cleantech, and advanced manufacturing hovered in the mid-hundreds. Augmented reality (AR) ranks last with just 240 deals, highlighting its niche position in grantmaking. This tiered structure suggests that while verticals like healthtech receive consistent support, they lack the broad, cross-industry applicability that governments prioritize in their largest funding initiatives.

The concentration of sovereign grants in SaaS and Big Data signals to investors that governments are prioritizing the broadest, most scalable layers of AI. PitchBook Senior Research Analyst Derek Hernandez provides further context on this trend. He states that over the past 10 years, SaaS and Big Data together have encompassed the promise of proprietary data as the new digital oil, requiring collection, processing, and refinement to become valuable and proprietary assets. The ML and deep learning booms in the 2000s and 2010s were both reflections of this promise and the massive investments that followed each new technological breakthrough.

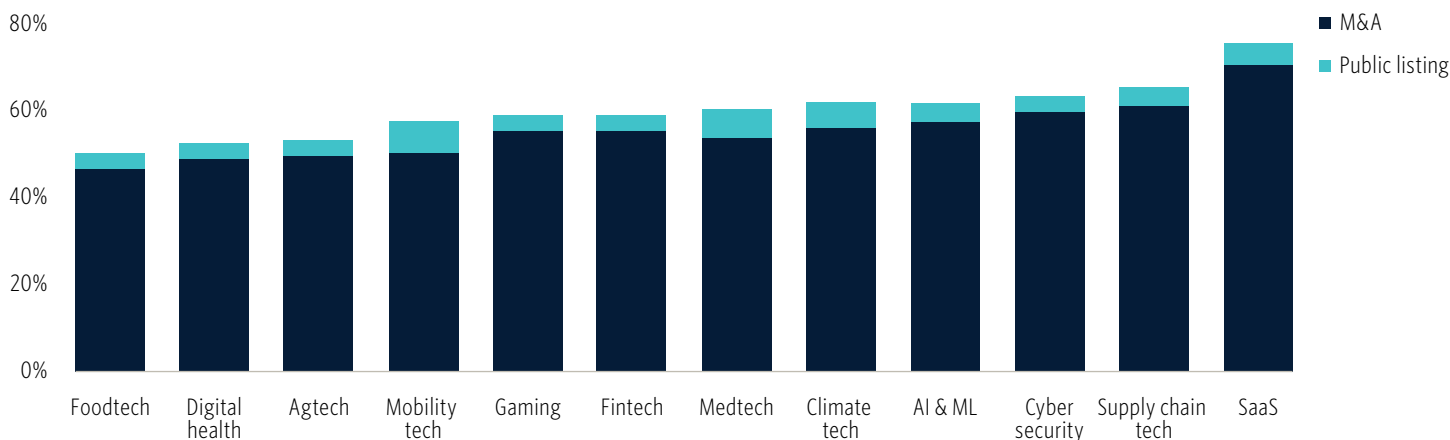
Annualized expected returns relative to cross-vertical average by VC vertical



Source: PitchBook • Geography: Global • As of December 31, 2024

Another major driver has been disproportionate returns and successful exits. In this area, SaaS, on a relative basis, has outperformed all the other verticals covered by PitchBook's Industry and Technology team. It shows a 5.5% higher annualized expected return relative to the cross-vertical average. Startups in the SaaS vertical have an elevated expectation of a successful exit compared with every other vertical, at 78.2% (75% M&A and 3.2% IPO), and a combined successful exit probability between approximately 54% and 65%, per the [PitchBook Exit Predictor methodology](#). This track record of liquidity and returns creates a powerful feedback loop, attracting further private capital into sectors already validated by government grants.

Expected share of exits for early-stage companies by successful exit type



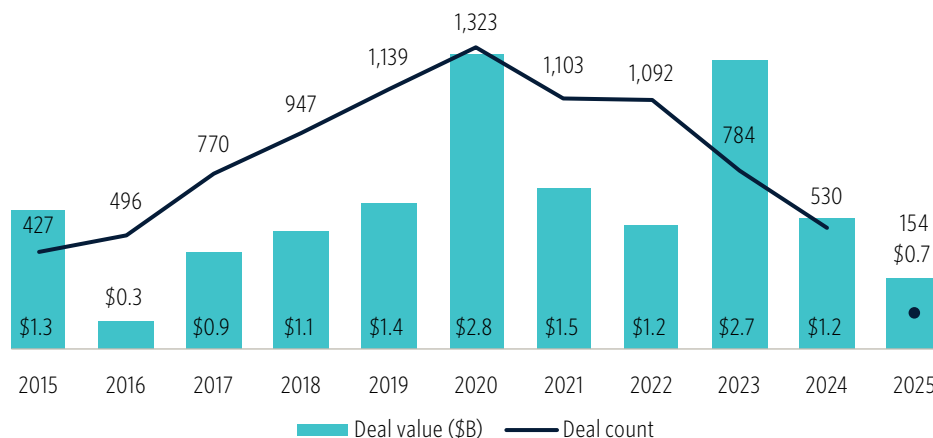
Source: PitchBook • Geography: Global • As of December 31, 2024

VCs often interpret this grant activity as a durable source of demand because these categories enable adoption across multiple industries rather than being confined to niche use cases. Categories with limited grant activity, such as AR or robotics, lack the same state-backed demand anchors in the form of procurement contracts, regulatory alignment, or public research & development (R&D) programs.

That tends to make investors more cautious and can limit the depth of late-stage financing compared with sectors where sovereign support is stronger. Grant activity can act as an early signal of sovereign conviction, and many investors incorporate it when weighing risk and allocating capital, even if it is only one factor among several.

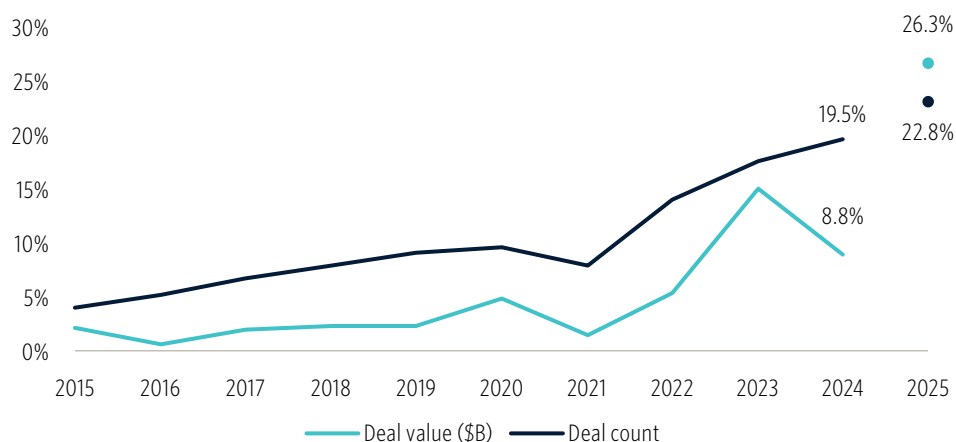
Globally, AI grant deal activity from the public sector has been falling. However, AI's share of all deal counts has been steadily rising since 2021 and surged from 7.9% in 2021 to 22.8% in 2025. AI's share of grant deal value saw an even bigger swing, jumping from 8.8% in 2024 to 26.3% in 2025. The fact that AI's share is rising despite lower absolute levels means funders are concentrating their bets in AI. They may be trimming exposure in peripheral areas but keeping (or reallocating) a relatively larger slice of limited resources to AI, signaling conviction in the sector's resilience. This trend indicates a maturing of public funding strategies, shifting from broad experimentation to focused investment in proven, high-impact areas of AI.

AI & ML grant activity



Source: PitchBook • Geography: Global • As of August 12, 2025

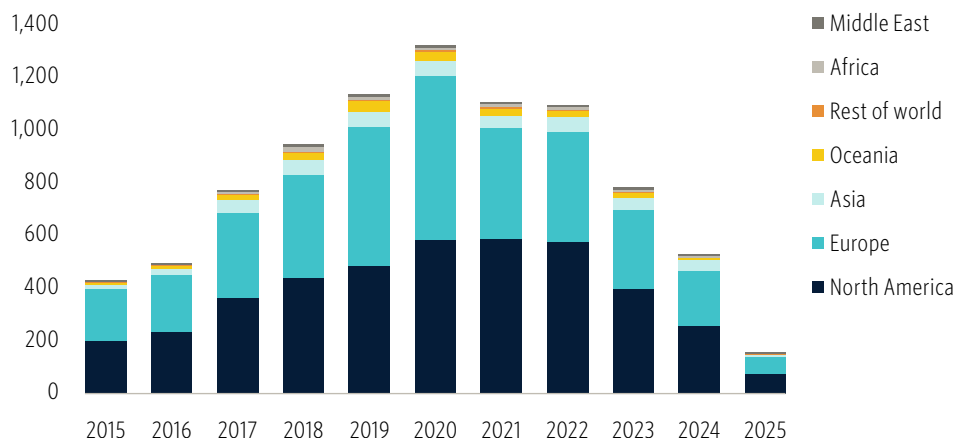
AI & ML grant share of all grant activity



Source: PitchBook • Geography: Global • As of August 12, 2025

Grant deal counts by region overwhelmingly show the US and Europe in the lead by several magnitudes, with Asia, Oceania, and the rest of the world far behind. This geographic concentration reflects the established innovation ecosystems and deep pools of research talent in these regions, making them the most efficient allocators of public R&D capital. For investors, this reinforces the view that despite the global nature of the AI race, the core innovation and the most attractive early-stage investment opportunities remain heavily concentrated in Western markets.

AI & ML grant count by region

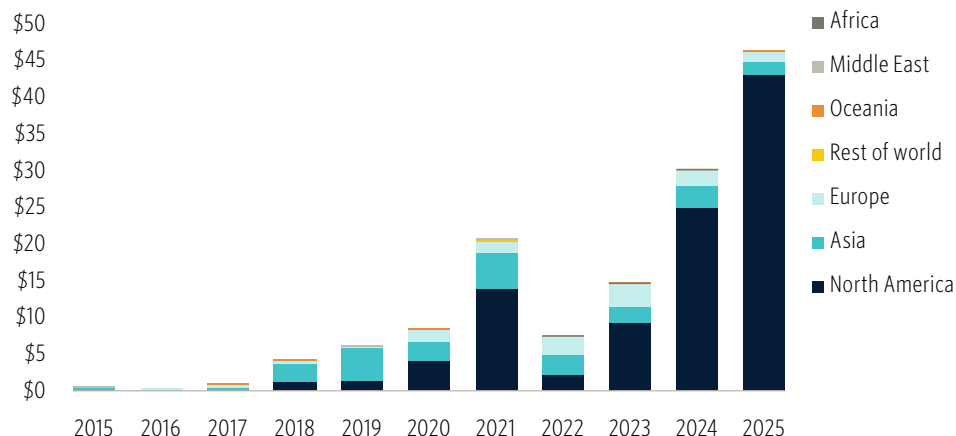


Source: PitchBook • Geography: Global • As of August 12, 2025

Governments and SWFs have consistently accounted for only about 2% of global AI & ML deal count over the past decade, underscoring their limited role in deal volume. Their influence instead shows up in deal value, which has climbed from 3% in 2015 to 31% in 2025. That share has risen steadily since 2022, highlighting their growing involvement in AI-related venture activity through larger checks rather than broad participation. This “whale” investing strategy means that sovereign capital is primarily a late-stage accelerant, providing the massive growth rounds that allow platform winners to achieve global scale.

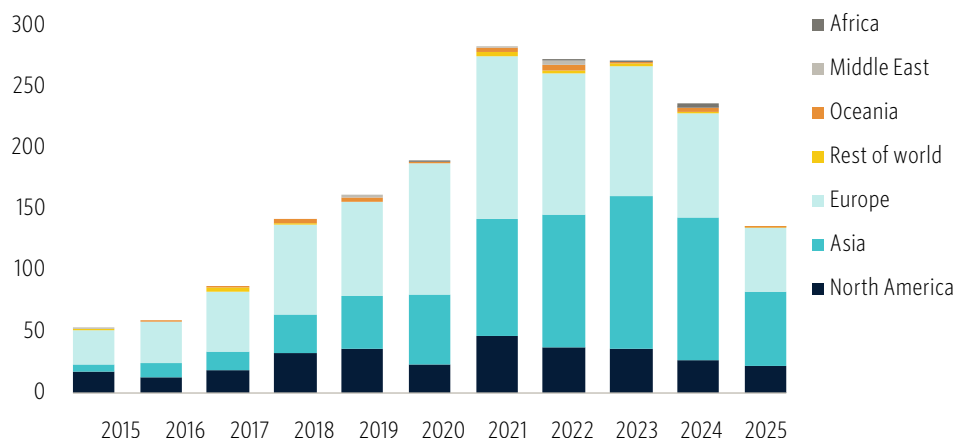
For deal value involving government or SWF participation, the US has dominated, especially over the past two years. From 2015 to 2025, North America shifted from a marginal participant to the clear center of AI & ML venture activity involving government or SWF participation. In 2015, the region accounted for just \$0.3 billion in AI & ML VC deal value with government or SWF participation, compared with \$0.1 billion in Asia and virtually none in Europe. By 2025, North America reached \$43.3 billion, dwarfing Asia’s \$1.7 billion and Europe’s \$1.3 billion. However, with deal counts in 2025 involving government participation or SWF, the pyramid is flipped on its head. Asia led with 61, followed by Europe at 52, and the US at 22.

AI & ML VC deal value (\$B) with government or SWF participation by region



Source: PitchBook • Geography: Global • As of August 12, 2025

AI & ML VC deal count with government or SWF participation by region



Source: PitchBook • Geography: Global • As of August 12, 2025

The data reveals a decade-long reordering of global AI capital flows, with sovereign-backed investment consolidating decisively in the US. This consolidation is a direct result of the US' lead in foundational model development and its deep, liquid capital markets. International sovereign investors have concluded that to gain meaningful exposure to the most advanced AI, they must invest in the US ecosystem. Consequently, it is appropriate to start any regional analysis with the nation magnetizing the most global capital.

The global pursuit of sovereign AI is not a uniform phenomenon; it is a collection of distinct regional and national strategies shaped by unique political systems, economic structures, and geopolitical objectives. From the legislative, market-driven approach of the United States to the state-directed industrial policy of China and the SWF-fueled ambition of the Middle East, each region offers a different set of opportunities and risks for investors. A granular understanding of these regional dynamics is essential for effective capital allocation in this new global landscape.

Regional strategies

North America: Industrial policy as a catalyst

The gravitational pull of the AI ecosystem is anchored in the US, which has solidified its position as the undisputed leader of the Western AI bloc through a combination of public investment, private sector innovation, and deep capital markets. A pivotal moment came in 2022 with the CHIPS and Science Act, which authorized \$280 billion to reinforce the country's AI and semiconductor base. This legislation represents the legislative vector for capital distribution in action. As of early 2025, the Department of Commerce's CHIPS Program Office had made 20 awards totaling \$33.7 billion in direct funding and \$5.5 billion in loans,¹² helping catalyze over \$540 billion in announced private-sector semiconductor investments nationwide.¹³

Subsidies are underwriting TSMC's advanced node fabrication facilities (fabs) in Arizona; Samsung's logic fabs in Texas; Intel's logic and advanced packaging builds in Ohio, Arizona, New Mexico, and Oregon; and Micron's DRAM projects in New York and Idaho. This onshoring reduces nondomestic dependence and expands domestic compute capacity. Most of the new capacity will arrive later this decade, but it signals to investors that the future supply of advanced logic, memory, and packaging is being de-risked. For frontier AI developers like OpenAI, Anthropic, and xAI, whose growth depends more on access to compute than on algorithms, this sovereign push acts as an indirect accelerator that shapes how markets price their ability to train and deploy models.

The US sovereign AI strategy now spans both chips and compute. The Stargate Project in Texas, a reported \$500 billion datacenter campus initiative by Microsoft and OpenAI, is an anchor for sovereign-scale infrastructure. Affiliated deployments, such as Oracle's sovereign cloud regions in Norway and the UAE, extend US capacity abroad, reinforcing America's central role in the global AI stack. At the same time, capital commitments between Oracle, NVIDIA, and OpenAI operate like sovereign guarantees, securing long-term compute supply and de-risking upstream revenue forecasts. For private markets, these initiatives expand the startup pipeline in power, cooling, interconnects, and packaging.

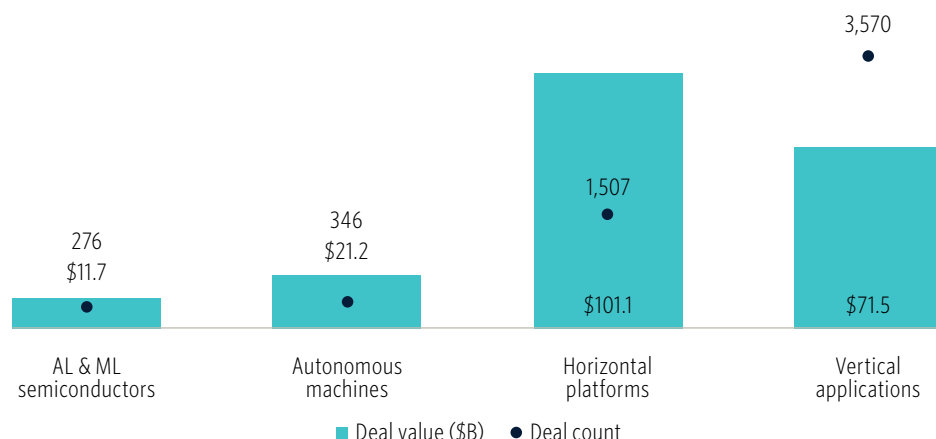
Long-duration contracts from corporate and sovereign partners reduce uncertainty around demand. In other infrastructure sectors, such as renewables and datacenters, similar agreements have historically supported higher valuations and smoother exit paths by anchoring investor confidence in predictable cash flows. AI infrastructure is beginning to follow the same pattern, with sovereign-scale contracts functioning as demand guarantees that ripple through the supply chain. This dynamic significantly lowers the risk profile for private equity and growth-stage investments in AI infrastructure.

¹²: "Commerce CHIPS Act Programs: Status Report," U.S. Department of Commerce Office of the Inspector General, June 2, 2025.

¹³: "Innovation Lightbulb: Tracking Chips Act Incentives," Center for Strategic and International Studies, Chris Borges, Yutong Deng, and Julie Heng, April 25, 2025.

When looking at the top 30 global AI VC deals involving a government or SWF participation, 60% of those companies—including OpenAI, Databricks, xAI, and Crusoe—were based out of the US. PitchBook Q1 and Q2 2025 deal flow data underscores the same pattern. Specifically, horizontal platforms have consistently captured the largest share of capital, outpacing vertical applications, semiconductors, and autonomous machines. This indicates a "winner-take-all" dynamic at the platform layer, where both private and sovereign capital are consolidating bets on a few key players.

AI & ML trailing 12-month VC deal activity by segment



Source: PitchBook • Geography: Global • As of June 30, 2025

Sovereign-linked capital is reinforcing this imbalance, channeling funds disproportionately toward US-based horizontal platforms that anchor the global AI stack. PitchBook also spoke with corporate VCs who noted—on condition of anonymity—that investment strategies remain weighted toward horizontal platforms. The application layer is still unsettled, but the prevailing view is that eventual winners will be built on top of foundation model providers such as OpenAI, Anthropic, and xAI. Capital flows reflect this conviction, with both venture firms and government-backed investors concentrating resources at the platform level. The heavy concentration of sovereign grants in SaaS and Big Data foreshadowed this pattern, highlighting how both public and private capital have consistently clustered around general-purpose layers of the AI stack that enable adoption across multiple industries.

Europe: In regulation we trust

Europe's sovereign AI push has established regulatory guardrails and limited industrial support, but it has not catalyzed US-level private market depth. The European Chips Act, adopted in 2023, commits €43 billion (\$46 billion) to double Europe's semiconductor share to 20% by 2030. For private markets, that signals state co-funding for fabs and pilot lines that lowers execution risk for suppliers and startups tied into the supply chain. Biodrone, Edgybees, and Stegra were among the largest recipients of grants at \$1.7 billion, \$958.3 million, and \$387.3 million, respectively, but most companies raised less than \$50 million, underscoring Europe's fragmented sovereign approach.

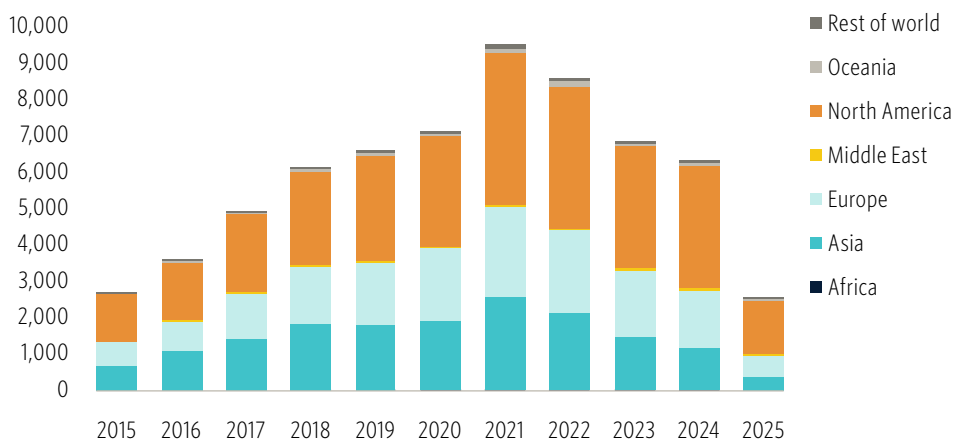
European AI & ML companies receiving government grants or funding from SWFs (2015-2020)

Company	2015-2025 grant value (\$M)	Investors
Biodrone	\$1,739.4	European Commission
Edgybees	\$958.3	United States Air Force, AFWERX, United States Department of Defense
Stegra	\$387.3	The Government Offices Of Sweden, EU-Egypt Innovation Fund, Energimyndigheten
Zebra Medical Vision	\$198.0	NHSX
VTT	\$151.8	Horizon 2020
Multiverse Computing	\$70.0	Perte Chip, Government of Spain, The Spanish Society for Technological Transformation, Innovate UK, European Innovation Council
Pathlake	\$60.21	UK Research and Innovation, UK Department of Health
ev.energy	\$42.0	California Energy Commission, Innovate UK, EDF Energy
Lightpoint Medical	\$31.8	National Institutes of Health, National Cancer Institute, European Commission, Horizon 2020, Biomedical Catalyst, Innovate UK, Wellcome Trust
Agrimetrics	\$29.2	Innovate UK

Source: PitchBook • Geography: Global • As of September 30, 2025

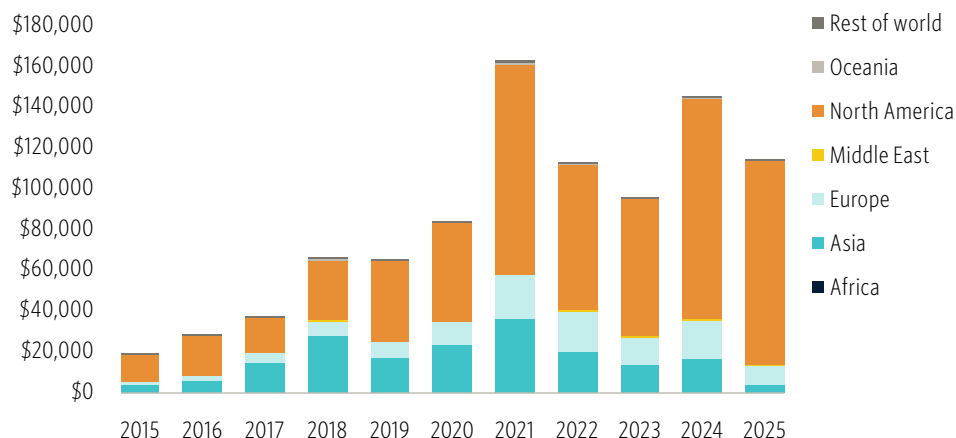
By the numbers, Europe continues to trail North America in AI & ML VC activity. In 2024, Europe logged 1,585 deals worth \$18.8 billion, while North America registered more than 2x at 3,369, with a value north of \$107 billion. Asia recorded 1,160 deals valued at \$16.7 billion, while the Middle East added 60 deals worth \$573.7 million. In H1 2025, Europe stood at just over 581 deals worth \$9 billion, while North America registered 1,493 deals worth \$100.1 billion. Asia posted 403 deals worth \$4.5 billion, and the Middle East registered 27 deals worth \$215.8 million. The gravitational center of sovereign-linked capital remains anchored in the US.

AI & ML VC deal count by region



Source: PitchBook • Geography: Global • As of June 30, 2025

AI & ML VC deal value (\$B) by region



Source: PitchBook • Geography: Global • As of June 30, 2025

The EU's AI Act, which entered into force in August 2024, phases in obligations through 2027 and carries penalties of up to €35 million or 7% of global revenue. In H1 2025, AI startups accounted for 24.1% of all European down rounds, compared with 13.8% in fintech and 19% in cleantech. Overall down rounds rose from 12.1% of deals in Q1 to 20.7% in Q2, underscoring that sovereign initiatives have not insulated AI from repricing pressures. The regulatory burden, while intended to create trust, is perceived by many investors as a tax on innovation that slows down development and increases compliance costs, making European startups less competitive than their US counterparts.

Late-stage dynamics reinforce the shallow depth of capital formation. Median Series E or later pre-money valuations more than doubled past €1.2 billion (\$1.28 billion) in 2025, yet the median deal size fell 13% to €83.5 million. The result is headline valuation growth without corresponding late-stage capacity. Outliers such as Mistral AI, the Paris-based foundation model developer that raised more than \$1 billion in its latest round in September 2025, highlight Europe's technical talent but also its reliance on US capital, with major rounds led by Andreessen Horowitz and General Catalyst.

Europe's regulatory-first stance is unlikely to attract growth capital on its own. However, it could create investable niches in compliance software, audit tooling, and privacy tech that can monetize mandatory requirements. This is not without precedent. For example, the rollout of the EU's General Data Protection Regulation in 2018 catalyzed a surge in privacy and compliance tooling. Startups such as OneTrust, TrustArc, and BigID raised billions as enterprises were compelled to adopt consent management and data protection software to meet regulatory mandates. A similar regulation tech ecosystem could form around the AI Act.

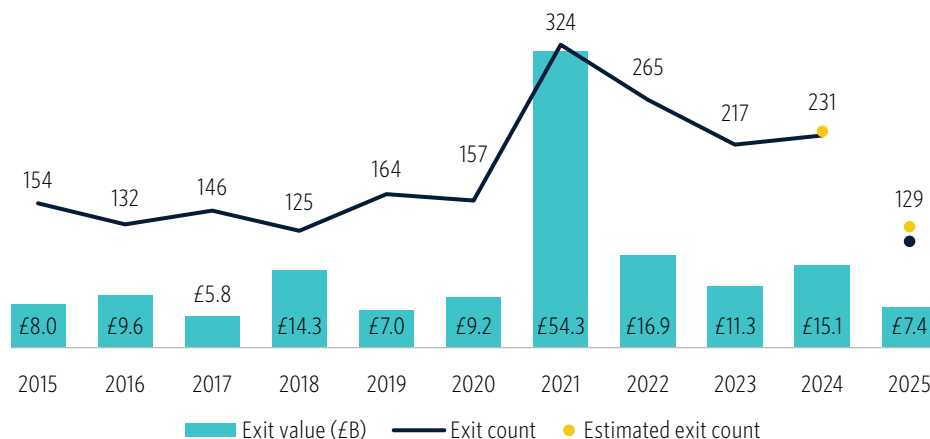
Taken together, Europe's sovereign initiatives provide regulatory guardrails—or what many view as innovation-inhibiting restraints—and selective industrial support, but they have not translated into the scale of capital formation or depth of private market investment seen in the US. Investors interpret the European Chips Act as a modest de-risking tool for hardware. Meanwhile, the AI Act is priced in as a cost center that shapes valuations and curtails late-stage appetite in sensitive sectors. For VCs, this means European AI remains a higher-risk allocation, where valuations and scaling hinge on cross-border capital rather than sovereign backstops.

The UK: Mind the exits

As for the UK, the government earmarked £500 million (\$625 million) through 2025 under the Department for Science, Innovation and Technology to expand sovereign AI capacity. The country's new sovereign AI Unit is directing funds into the AI Research Resource, with flagship supercomputers Isambard-AI in Bristol and Dawn in Cambridge forming the backbone of government-controlled compute. Execution relies on public-private partnerships. NVIDIA and University College London are building sovereign AI platforms to run on this infrastructure, while additional partners are expanding the high-performance compute base. The initiative is designed to provide the government with allocatable compute for security, health, and science, while reducing dependence on nondomestic providers.

Sovereign backing has not prevented fundraising from cooling. In H1 2025, UK AI startups raised £2.3 billion (\$2.9 billion), putting the market on pace for a 10.8% YoY decline from 2024's £5.2 billion (\$6.5 billion) peak. The area where the initiatives are showing impact is in exits and liquidity. AI ranked second only to SaaS by exit count in H1, with 14 acquisitions and buyouts, including Papercup and Convergence.

UK VC exit activity



Source: PitchBook • Geography: Global • As of June 30, 2025

For corporates, sovereign-funded compute hardware, like Isambard-AI and Dawn, lowers execution risk in regulated verticals such as health and finance, making acquisitions more attractive. Liquidity infrastructure is also widening. The launch of PISCES, a government-regulated secondary trading venue in June 2025, enables institutional investors and founders to transact private shares pre-IPO under Financial Conduct Authority oversight. Together, this points to sovereign AI initiatives shaping the UK ecosystem less by boosting inflows and more by reinforcing credible exit pathways and alternative liquidity channels.

The Middle East: A desert bloom

Gulf sovereign wealth funds are accelerating regional AI ambitions by underwriting both domestic infrastructure and cross-border partnerships. Capital is focusing on training clusters, datacenters, and sovereign platforms, tightening deployment timelines and opening new pathways for private capital. The Gulf has emerged as one of the world's most active new markets for AI infrastructure, with projects like the UAE's Stargate AI campus and multibillion-dollar sovereign tech initiatives confirming rising demand.

In the United Arab Emirates, Abu Dhabi has become the anchor for sovereign AI in the region. Mubadala Investment Company and MGX, the state-owned AI investor formed with G42, are deploying capital at scale. MGX was launched in 2024 with several billions of dollars in seed funding and has announced plans to raise tens of billions more through sovereign and private co-investors. MGX is a core equity partner in Stargate, OpenAI's multi-gigawatt compute program. The Stargate UAE site, announced in 2025, will operate as a one-gigawatt cluster within a five-gigawatt AI campus in Abu Dhabi. It is being developed with G42, Oracle, OpenAI, and global suppliers including NVIDIA, Cisco, and SoftBank. Operations are expected to begin in 2026. For OpenAI, Stargate's UAE node extends US capacity abroad and embeds the company in sovereign ecosystems. This gives it entrenchment power and ensures that future vertical applications are built on top of its models.

Hardware flows reflect the scale of this bet. A US agreement in 2025 cleared the way for the UAE to import up to 500,000 NVIDIA chips annually, worth an estimated \$15 billion to \$20 billion at prevailing H100 prices of \$30,000 to \$40,000 per unit. From this, G42 was allocated around 100,000 units valued at \$3 billion to \$4 billion. This secure, large-scale supply of cutting-edge hardware is a key differentiator for the UAE's strategy.

Saudi Arabia is pursuing a parallel path. In September 2025, the Public Investment Fund (PIF) announced a new \$40 billion fund, AI-Sadaf, dedicated to AI. The fund has launched HUMAIN, a national AI platform investing across chips, infrastructure, and models. PIF also signed a memorandum with Qualcomm to establish AI datacenters and edge-to-cloud services. In parallel, DataVOLT announced a \$5 billion AI campus at NEOM, structured around net-zero energy. These projects pull forward demand for semiconductors, cooling, and interconnect suppliers while opening new commercial opportunities for model developers and hyperscalers.

Qatar is scaling digital infrastructure through its sovereign fund as well. The Qatar Investment Authority partnered with Blue Owl to launch a \$3 billion platform targeting datacenters. This ties directly into private market momentum. Blue Owl recently closed \$7 billion for its Digital Infrastructure Fund III, and through joint ventures has already committed \$750 million to build AI datacenters for CoreWeave and \$3.4 billion for a 1.2 gigawatt campus with Crusoe in Texas. The sovereign-backed platform guarantees scale for these investors and ensures hyperscaler demand in the Middle East is captured.

Kuwait is participating through the Kuwait Investment Authority, which joined the AI Infrastructure Partnership alongside MGX, BlackRock, Global Infrastructure Partners, and Microsoft. The partnership is capitalized at roughly \$30 billion to finance next-generation AI datacenters. This signals that Gulf SWFs are not only financing domestic capacity but also embedding themselves in global infrastructure consortia. Oman is aligning its updated national AI strategy with targeted equity stakes in chip and compute firms such as Saliency Labs. This reflects an effort to localize parts of the supply chain while laying groundwork for domestic datacenter development.

While state-backed entities are leading the construction of national AI infrastructure, private companies in key industries are moving quickly to align with these sovereign priorities. Tabby, a fast-growing fintech firm active across the Gulf Cooperation Council region, announced a major investment in NVIDIA HGX systems to localize its AI and data operations. The deployment strengthens the firm's ability to manage compute workloads securely within national borders. Chief Operating Officer and Co-Founder Daniil Barkalov said the investment "sets the foundation for sovereign AI in financial services" by combining faster processing capabilities with greater control over data sovereignty.¹⁴

SWF capital is acting as a demand guarantee for global AI suppliers. NVIDIA benefits from secured orders across sovereign campuses, with its partnership with OpenAI tied to as much as \$100 billion in incremental investment over the coming years. OpenAI has leveraged Stargate to entrench its models in national infrastructures, shaping downstream enterprise adoption. CoreWeave's ability to tap sovereign-backed demand has strengthened its growth narrative, contributing to its successful IPO in July 2025, while Crusoe has secured billions in joint venture financing to scale its own sovereign-aligned campuses. Startups in cooling, interconnect, and chip design are also positioned to benefit, with sovereign projects reducing execution risk for scaling. Private equity firms like Blue Owl are tapping sovereign partnerships to scale portfolios, while venture investors see clearer exit pathways through sovereign-anchored AI infrastructure.

Collectively, the Middle East's sovereign AI push is reshaping how private markets price growth in both foundational models and the enabling supply chain. This dynamic reinforces US leadership in AI, with NVIDIA securing sovereign demand for infrastructure and OpenAI entrenching its models abroad. This ultimately cements US control over both the hardware base and the horizontal platform layer of the global AI stack.

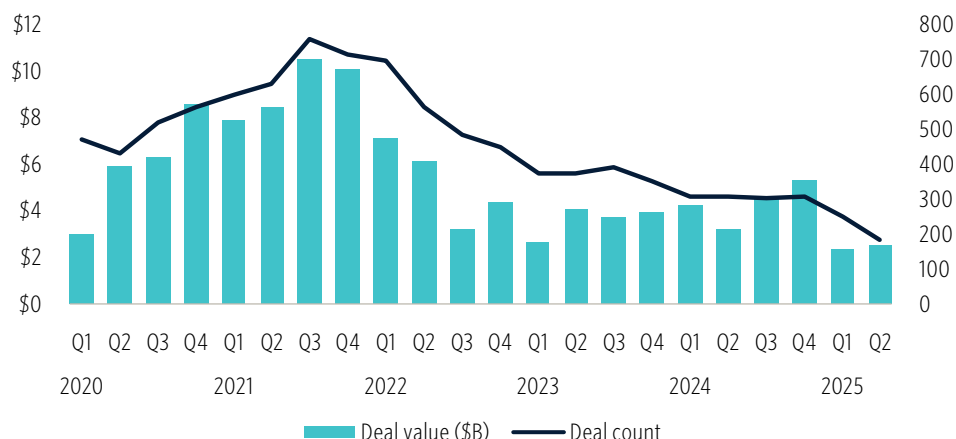
¹⁴: "Tabby Invests in NVIDIA HGX Systems to Power Advanced AI Infrastructure," Tabby, October 12, 2025.

Asia: Fragmentation and consolidation

Asia presents the most complex and fragmented sovereign AI landscape, characterized by the state-driven, self-sufficiency model of China and the more open, partnership-oriented strategies of industrial democracies like Japan, South Korea, and Singapore.

As far as funding, AI VC deal activity in the Asia-Pacific (APAC) region hit its lowest count on a quarterly basis going back to Q1 2020, from a peak of 761 in Q3 2021 to 186 in Q2 2025. China, historically the largest VC market in the region, saw its VC investment levels and deal count fall to a multi-year low. This substantial contraction in one of the region's main drivers is the largest quantitative contributor to the overall APAC decline.

APAC AI & ML VC deal activity by quarter



Source: PitchBook • Geography: Global • As of June 30, 2025

The sharp decline in deal volume was not matched by a pullback in total capital deployment, with combined APAC AI VC deal value reaching \$4.9 billion in H1 2025. The divergence underscores a clear flight-to-quality trend as investors navigate macroeconomic uncertainty and a constrained exit environment. Capital is concentrating in a smaller set of later-stage, high-conviction AI deals, while early-stage startups face growing challenges in securing follow-on funding. Geopolitical tensions and tariff risks are reinforcing this pattern, driving capital away from China toward more stable regional growth hubs such as Japan and India, where investors see stronger policy support and clearer exit visibility.

China: The self-reliance loop

China's sovereign AI strategy is driven by industrial policy and bureaucratic allocation that push capital into a domestically controlled stack spanning chips, compute, models, and applications. The 2017 New Generation AI Development Plan set the top-level blueprint to 2030, while Made in China 2025 aligns subsidies and procurement to strategic technologies. Since 2014, Beijing has operationalized this with government guidance funds that channel public and bank capital into "hard tech" and AI.

Capital scale has accelerated. The China Integrated Circuit Industry Investment Fund, known as the Big Fund, raised 138.7 billion yuan (\$19 billion) in Phase I and 204 billion yuan (\$28 billion) in Phase II. Phase III was formed in May 2024 with 344 billion yuan (\$47.5 billion), the largest investment to date, backed by the finance ministry and major state banks to prioritize equipment, manufacturing, and AI-relevant semiconductors such as HBM chips. In January 2025, China added a 60 billion yuan (\$8.3 billion) National AI Industry Investment Fund to target early-stage AI projects. A planned National Venture Capital Guidance Fund is designed to mobilize roughly 1 trillion yuan (\$138 billion) in “social capital” for seed and startup investments in AI, robotics, and embodied intelligence. Local governments mirror this at the city level. For example, Shanghai launched 100 billion yuan (\$13.8 billion) of funds in 2024 for integrated circuits, biomedicine, and AI, with explicit targets across intelligent chips, software, autonomous driving, and robotics.

This policy architecture underwrites a dual-economy approach. External controls and supply risk are addressed by import substitution and domestic scale, while internal demand is driven by cloud, e-commerce, and industrial automation customers. Huawei exemplifies the chip leg of this strategy. The company is reportedly expanding production of its Ascend 910C AI processors, with plans to roughly double output in 2026 to about 600,000 units. US officials, however, have assessed that 2025 output could remain limited to around 200,000 advanced AI chips due to export-control constraints.^{15, 16, 17} The trajectory is nonetheless evident: growing domestic chip output reduces exposure to US sanctions and strengthens local supply options for Chinese clouds. On the platform side, Baidu’s ERNIE continues to anchor a state-aligned foundation model track. Additionally, SenseTime’s generative AI business has pivoted its revenue mix toward model-driven products, with H1 2025 revenue at roughly 2.4 billion yuan (\$330 million) and losses narrowing as scale improves.

The hyperscale cloud cohort is matching state capital with corporate capital expenditure. Alibaba has disclosed a 380 billion yuan (\$52.5 billion) three-year program for AI infrastructure and datacenters, positioning its Qwen models and AI services as a core pillar. Tencent increased its AI capital expenditure after spending \$10.7 billion in 2024 and is expanding support for domestic AI chips across its cloud stack. Baidu reports rising AI cloud costs tied to model training and inference ramps, alongside buybacks that stabilize equity while it reallocates toward AI services. These corporate allocations translate policy into deployed compute, new sovereign clouds, and domestic inference at scale.

Private market implications are tangible, though they flow differently than in open markets. First, the guidance-fund complex and the new AI fund are direct sources of domestic growth equity for chipmakers, tool vendors, and model companies. Big Fund III’s 344 billion yuan (\$47.5 billion) pool prioritizes equipment, materials, manufacturing, and memory that feed AI workloads, benefiting listed names such as Semiconductor Manufacturing International Corporation, Hua Hong Semiconductor, and private suppliers in lithography subsystems, deposition, and testing. Second, the

15: “China’s Huawei Hypes Up Chip Computing Power Plans in Fresh Challenge to Nvidia,” Reuters, Brenda Goh and Che Pan, September 18, 2025.

16: “US Says China’s Huawei Can’t Make More Than 200,000 AI Chips in 2025,” Reuters, June 12, 2025.

17: “Huawei Chips Are One Generation Behind US but Firm Finding Workarounds, CEO Says” Reuters, Brenda Goh, June 9, 2025.

60 billion yuan (\$8.3 billion) National AI Industry Fund supplies early capital to AI startups at a time when global venture capital access is limited by data and security rules. Third, corporate capital expenditure from Alibaba, Tencent, and Baidu acts as a demand guarantee for domestic GPU substitutes, interconnects, power systems, cooling, and datacenter developers. SenseTime's shift toward generative AI revenue shows how model vendors monetize under this protected demand.

Nondomestic investor access is narrower than in the US because AI and semiconductors are treated as strategic. Nondomestic direct investment and control limits restrict direct stakes in sensitive assets, and variable-interest-entity structures are constrained for data-intensive businesses. Capital formation remains strong but is largely internal, with valuations and liquidity events clustering in Hong Kong and mainland exchanges for model companies and equipment makers. Private opportunities are concentrated in domestic rounds anchored by guidance funds and state-bank capital. For global allocators, exposure is more likely to come via supply chain beneficiaries outside China, power and datacenter vendors serving demand outside China, or selective Hong Kong listings of Chinese AI firms.

Policy reinforces this separation. The nation's dual-circulation strategy prioritizes self-reliance in technology and demand while preserving export channels where possible. Export controls constrain NVIDIA's addressable market in China and raise the relative value of domestic accelerators. At the same time, large domestic capital expenditure programs ensure that compute supply grows onshore, which supports local model ecosystems and sector applications from industrial vision to autonomous systems and fintech risk. The result is a durable, state-backed capital stack for sovereign AI that reduces execution risk for domestic suppliers, shapes cash-flow visibility for listed Chinese semis and AI software names, and limits cross-border arbitrage. For private markets, China represents a deep but increasingly closed loop. The opportunity set is substantial, the funding vectors are clear, and the exit venues are increasingly domestic, with policy continuity pointing to sustained allocations across chips, compute, and models through the rest of the decade.

South Korea: Legislative coordination and chaebol alignment

South Korea is deploying sovereign AI strategies primarily through legislative funding and bureaucratic coordination. In 2024, the government announced a 9.4 trillion won (\$7 billion) package through 2027 to support AI chips, national compute, and model development. The National Artificial Intelligence Committee, created in 2024, oversees implementation across ministries and chaebol partners. Funds target next-generation AI semiconductors, sovereign datacenters, and algorithm development, with SK Hynix, Samsung Electronics, and Naver positioned as anchors. SK Hynix received state backing to accelerate HBM development, leveraging its global lead in HBM3E. Samsung, already investing tens of trillions of won annually in semiconductors, benefits from matching state incentives that prioritize AI workloads and sovereign compute clusters. Naver has partnered with LG AI Research to advance HyperCLOVA, Korea's large language model, which gained government contracts for domestic inference in 2025.

Private market implications are visible in venture funding and corporate exits. AI chip startups such as Rebellions and FuriosaAI raised capital with participation from state-aligned funds and chaebols, stabilizing valuations despite US and Chinese competition. Sovereign initiatives function as a demand guarantee, as government contracts for defense, finance, and language localization pull early revenues. Exit pathways remain tied to KOSDAQ listings, where state-backed firms can expect favorable liquidity, though global allocators face constraints on direct participation. For investors, South Korea's sovereign AI program creates predictable growth for memory, model, and accelerator firms, with valuations supported by state demand and chaebol alignment.

Japan: Industrial policy

Japan is advancing sovereign AI through industrial policy and regulation rather than sovereign funds. In 2023, Tokyo launched the AI Japan Strategy with an initial 150 billion yen (\$1 billion) allocation to AI research and compute. In 2024, the government approved a 2 trillion yen (\$13 billion) package aimed at semiconductors and AI, tied to the Ministry of Economy, Trade and Industry (METI). This program supports domestic fabs like Rapidus, which is building advanced logic plants with AI workloads in mind, and sovereign compute initiatives such as the AI Bridging Cloud Infrastructure supercomputer at the National Institute of Advanced Industrial Science and Technology, one of the world's fastest AI training systems.

Private market players are benefiting through preferential access to state contracts and R&D co-funding. Preferred bidders include SoftBank, which announced a \$65 billion AI-focused investment fund in partnership with METI in August 2025, aiming to expand sovereign-scale AI compute in Japan. Other beneficiaries include startups such as Preferred Networks, which raised significant private rounds at valuations above \$2 billion with state research contracts underpinning demand. Japan's AI strategy also emphasizes regulation, with METI drafting frameworks that will shape compliance and open opportunities for startups in privacy, audit, and compliance tooling.

For private markets, Japan's sovereign AI push supports growth for semiconductor firms like Rapidus, model developers like Preferred Networks, and cloud/compute providers such as SoftBank. Valuations are being supported by government co-funding and procurement guarantees, though the scale remains smaller than in South Korea or China. Together, South Korea and Japan illustrate how advanced industrial economies deploy sovereign AI without SWF-style vehicles, instead relying on legislative funding, chaebol or corporate alignment, and targeted subsidies. For investors, this creates a landscape where domestic champions receive valuation support and exit visibility, while global allocators gain exposure primarily through public markets and selective corporate partnerships.

Singapore: Small state, big compute

Singapore's sovereign AI strategy is anchored by Temasek Holdings, its \$300 billion sovereign wealth fund, which has positioned itself at the center of global AI infrastructure. In June 2025, Temasek joined the AI Infrastructure Partnership (AIP) alongside Microsoft, BlackRock, and Abu Dhabi's MGX, committing several

billion dollars in initial capital to build sovereign-scale datacenters and AI campuses across Asia. This program mirrors the Gulf's sovereign AI initiatives, with Temasek functioning as both an equity provider and strategic convener. The city-state supplements this sovereign capital with strong regulatory leadership, issuing one of the most advanced AI governance frameworks in Asia and creating a policy environment favorable for scaling private solutions.

Temasek has been an active investor in global AI players, backing Anthropic, Databricks, and emerging model developers. Through the AIP, Singapore provides guaranteed demand for infrastructure vendors like Equinix and ST Telemedia Global Data Centres, as well as chip suppliers aligned with NVIDIA's global expansion. Domestic firms such as Grab and SEA.AI are incorporating sovereign-backed AI infrastructure into their platforms, expanding vertical applications in fintech, logistics, and gaming. Local startups including ADVANCE.AI and TAIGER have benefited from government procurement pilots and co-investment vehicles, stabilizing valuations in a tighter venture environment.

For investors, Singapore represents the most open sovereign AI program in Asia. Temasek's participation in AIP creates co-investment opportunities for private equity in datacenters and connectivity, while sovereign-backed demand reduces execution risk for early-stage AI companies. Valuations for domestic startups remain supported by government and sovereign procurement, while global allocators gain exposure through Temasek's syndicates and selective growth rounds. Singapore's sovereign AI push highlights how a small state can leverage its sovereign wealth fund to build regional infrastructure, secure strategic partnerships, and channel private market capital into both horizontal platforms and enabling infrastructure.

India: Code and bureaucracy

India's sovereign AI strategy is structured around legislative funding and bureaucratic allocation rather than sovereign wealth capital. In March 2024, the government approved the IndiaAI Mission with a budget of 103 billion rupees (\$1.3 billion) over five years. The program is divided across three verticals: building compute infrastructure, developing indigenous foundation models, and supporting AI innovation through grants and accelerators. Implementation is coordinated by the Ministry of Electronics and Information Technology, which is setting up IndiaAI compute clusters and data platforms to provide domestic access to training resources.

The mission has direct implications for private markets. Sarvam AI, selected in 2024 to build India's first sovereign foundation model, has raised more than \$50 million in venture funding led by Lightspeed Capital and Peak XV Partners. State support for Sarvam acts as a demand guarantee, ensuring domestic enterprise adoption of its models. Other startups, including Ola Krutrim (founded by Ola co-founder Bhavish Aggarwal) and AI4Bharat, are also benefiting from government co-funding and procurement pilots. Valuations for these firms have remained buoyant despite weaker global conditions, supported by sovereign contracts that underpin early revenues.

India's sovereign AI policy also extends to infrastructure. Public procurement of GPUs and compute clusters is expected to generate billions in contracts for datacenter operators and cloud providers. Reliance Jio and Bharti Airtel, India's leading telecoms, are scaling AI cloud capacity to align with IndiaAI clusters, while nondomestic hyperscalers such as AWS and Microsoft are partnering with Indian ministries to co-develop sovereign-aligned infrastructure. For investors, India's program creates clear entry points at the early and growth stages. Startups aligned with the IndiaAI Mission can secure higher valuations and predictable exit pathways through IPOs on domestic exchanges. Infrastructure vendors and telecom operators gain sovereign-backed demand visibility, reducing execution risk. India's sovereign AI approach reflects its ambition to localize models and infrastructure while keeping private markets engaged through co-funding and public-private partnerships.

Industry intelligence

To bring an industry perspective into the analysis, PitchBook interviewed Andy Hock, senior VP of product and strategy at Cerebras, on the role of sovereign policy, infrastructure, and geopolitics in driving demand for advanced compute. Cerebras is a US-based company that developed the world's largest chip- and wafer-scale systems, giving it a distinctive role in powering national AI infrastructure projects across the US, the Middle East, and beyond. The company's responses have been edited for clarity.

PitchBook: To what degree do legislative frameworks like the US CHIPS and Science Act directly expand your addressable market? Do they create reliable demand signals that VCs can use to benchmark growth opportunities?

Hock: The CHIPS & Science Act is an indirect accelerator of Cerebras' addressable market, as it eases supply bottlenecks (advanced logic and memory), strengthens the US R&D / national-lab market for big compute, and helps shift corporate capital expenditure toward domestic supply chains. Cerebras already handles manufacturing and packaging in the US, and it improves our ability to scale and sell our wafer-scale systems. It's also worth noting that the role of the US government has always been that of a partner. In April 2025, Cerebras won a hardware development contract worth \$45 million from the Defense Advanced Research Projects Agency to build a high-performance AI system using co-packaged optics. This investment in innovation by the US government helps accelerate the AI market and continues to position the US as a leading innovator. I think VCs can look at the CHIPS Act as a useful macro tailwind and screening filter—not as a single, fully reliable demand guarantee, but as one of many data points.

Which regions are showing the strongest near-term momentum in sovereign AI, and what signals tell you that?

We're seeing exciting developments and ambitious visions for AI initiatives and infrastructure around the world. Sovereign AI is a very valuable market with massive implications for society and the global economy. The UAE was one of the first to show a strong commitment in creating a vision for global and sovereign AI, and our strategic partner G42 in particular has been a national champion leading

the charge. G42 has a strong national charter and vision for AI, enabling them to shape and own their destiny in the future of AI. They have invested and invented accordingly, including massive infrastructure build-outs with our Condor Galaxy supercomputers, made in partnership with Cerebras; introducing leading language models in Arabic (Jais), Hindi (NANDA) and reasoning (K2 Think), and strong partnerships with leading AI players, like OpenAI, Microsoft, and Cerebras. We're seeing similarly ambitious and visionary initiatives coming from the UK, Europe, India, and the Middle East, including Qatar, Oman, Saudi Arabia, and Central Asia (Kazakhstan, Uzbekistan, Tajikistan), as well as from Southeast Asia and South America. And, last but not least, the US is not only a leader of industry in AI and technology, but the US government is positioning itself through programs like the AI Research Resource, the National Science Foundation, and the Department of Energy supercomputing programs to lead in sovereign and public initiatives as well.

What challenges do you foresee in the global rollout of sovereign AI?

There is a massive global shortage of qualified AI talent—ML researchers, systems engineers, infrastructure architects, and more. Many of these are concentrated in the US and a handful of other tech hubs. With difficulty in attracting or retaining talent, sovereign AI initiatives must match what private or nondomestic labs can offer. Sovereign AI also requires high-performance compute, storage, high-bandwidth memory, fast interconnects, cooling, and power. Building or procuring this infrastructure is expensive. Some countries face cost, logistics, or supply chain issues. Lastly, export controls, trade restrictions, geopolitical risks can impact the rollout of sovereign AI.

How concerned are you about modern geopolitical tensions affecting Cerebras' sovereign AI deployment strategies?

We carefully consider geopolitical factors every day of our business, across our domestic and international work. Cerebras is on a mission to put the right high-performance AI compute into the hands of more users, more developers, and more organizations across our friend and ally nations around the world. We believe there will be fundamentally stronger AI innovation and better outcomes for society with democratized access to these resources. We have a long ongoing beneficial relationship with the US government, including the Department of Commerce and the Bureau of Industry and Security. We work closely with them to understand where we can export and where, and we follow the letter of the law in complete coordination with them. We're watching the shifting geopolitical landscape and, at the end of the day, we only ship where we're allowed to ship. Democratization of AI compute will lead to better outcomes and we want to enable our friends and allies.

When assessing sovereign AI partners, what are things that you look for?

First, we work closely with our partners in the US government to define a commercial strategy to deliver AI compute to our friends and allies. The common thread that we have seen in the strongest candidate prospects is an ambitious AI vision for both public and private/industry benefit across science, industry, and society. The strongest sovereign AI partners have a well-defined government

department or commercial champion organization—such as a ministry of AI or a public/private champion organization like G42 in the UAE, for example—so the rest of the ecosystem knows who to partner with to help scale development of AI infrastructure. This means that, as a compute and model development partner, Cerebras can immediately contribute and deliver value when there's a well-defined need or set of projects around model development and the public/private compute infrastructure is in place—supported by datacenters, power, and access to datasets.

To add another industry perspective, PitchBook interviewed d-Matrix on how sovereign policy, infrastructure, and geopolitical dynamics are shaping demand for inference-centric compute. d-Matrix is a US-based company that develops inference accelerator chips with an integrated software stack, enabling efficient deployment of transformer and agentic AI workloads. Its architecture is designed to lower latency and reduce power requirements, which positions the company well for sovereign AI programs where inference at scale is a critical bottleneck. With engagements across the US, the Middle East, and Asia, d-Matrix provides insight into how countries are approaching sovereign AI after the initial wave of model training. The company's responses have been edited for clarity.

PitchBook: Are sovereign AI buyers asking for modular, plug-and-play systems that interoperate with the global AI stack, or are they demanding bespoke, closed stacks that only run inside their own walls?

d-Matrix: To supply sovereign AI programs, providers need to deliver the full stack. d-Matrix works with partners to integrate hardware and software, enabling deployment at scale. While we engage with hyperscalers globally, our current focus is the US. We anticipate demand for inference-focused datacenters abroad will accelerate once countries with sovereign AI initiatives complete model training. Timelines will vary by region. For agentic AI, particularly in verticals such as consumer, healthcare, and finance, we see US adoption of inference within the next two years, with the Middle East following in roughly two to four years.

How do regional differences shape the way sovereign AI infrastructure is being built?

Not all sovereign AI build-outs look the same. Regional constraints and enablers create divergent strategies. In the Middle East, countries such as Saudi Arabia, Oman, Kuwait, Qatar, and the UAE are pushing AI as part of broader digital-economy agendas. Energy costs are higher due to the climate, but abundant supply means it is not viewed as a headline risk. Conversations with sovereign AI stakeholders confirm that energy rarely surfaces as a barrier. Instead, the challenge lies in driving adoption and ensuring populations integrate AI into domestic economies.

This contrasts with Singapore and Europe, where energy availability is a more pressing concern, but digitally indigenous populations make adoption and scaling far easier. For markets facing power constraints and relying on GPU-heavy infrastructure, inference-optimized systems from companies like d-Matrix offer a 5x to 10x efficiency gain, creating a differentiated path to sovereign AI.

We also spoke with Multiverse Computing, a European company specializing in AI model compression and quantum software. Its CompactifAI platform uses quantum-inspired techniques to reduce the cost and energy consumption of large AI models without major performance loss. Headquartered in Donostia-San Sebastián, the company has offices across North America and Europe. Clients include Iberdrola, Moody's Analytics, the Bank of Canada, and BASF. Multiverse Computing has received industry recognition, including CB Insights' "100 Most Promising Companies in AI" list and Digital Europe's "Future Unicorn 2024 Award."

PitchBook: What is the funding environment like for startups in Europe? Do you see a big difference in capital accessibility between early-stage and late-stage/venture-growth investing?

Multiverse Computing: Generally, early-stage startups in Europe can access capital from venture capital firms and government funds with relative ease, though the key word is "relative." Once they reach the scaling phase, such as later-stage VC or venture growth, funding options become more limited. Because Europe's startup ecosystem is less mature than that of the US, some companies may turn to US investors for larger capital injections, even though the situation in Europe has been improving overall in the last years.

How are governments playing a role in funding?

There is a growing sense of urgency among European policymakers who fear falling behind China and the US in the AI race. Europe already missed out on the social media boom, which was dominated by Silicon Valley. For instance, half of the capital from our last round came from the Spanish government, and we have additional investments from European and regional governments. Policymakers now recognize the need to accelerate European innovation and ecosystem development, and there have been gradual improvements in how capital and policy are being deployed to support that goal, which in turn attracts more VCs.

Macroeconomic and market forces

The Federal Reserve and monetary policy

Broader macroeconomic trends and geopolitical forces are acting as powerful accelerators and shapers of the sovereign AI market. The era of frictionless globalization has definitively ended, replaced by a new paradigm of geostrategic competition where technological leadership is paramount. This environment provides a durable, noncyclical foundation for sovereign tech spending that is largely insulated from conventional economic cycles. Two macroeconomic factors are particularly impactful: the evolution of monetary policy and the deepening of the US-China technology rivalry.

The accommodative monetary policy that has re-emerged in the second half of 2025 is creating favorable conditions for capital-intensive AI projects. In September 2025, the Federal Reserve cut rates by 25 basis points, its first reduction since December 2024. The move showed a recalibration of its dual mandate, tilting away from a singular focus on inflation and toward stabilizing a softening labor market.

This pivot, occurring while core inflation remains above target, signals a higher tolerance for inflation in the service of economic growth. For private markets, a lower cost of capital reduces the discount rate for long-duration projects, making venture-style bets on foundational AI technologies more attractive. Should labor markets weaken further, additional cuts appear likely, potentially amplifying an AI market already delivering record highs in deal value and volume.

More profoundly, the US-China rivalry is the central organizing principle of the sovereign AI landscape. Security-first policies, including US export controls on advanced semiconductors, are politicizing supply chains and steering technology build-outs away from economic maximization toward geopolitical optimization. On the US side, corporate champions like OpenAI and NVIDIA are capturing sovereign-backed demand through programs like Stargate, which guarantee long-term capital flows into both horizontal platforms and infrastructure. These partnerships effectively extend the US technology ecosystem to allied nations, cementing US standards and market share.

In China, a mirror image is emerging. Huawei and Baidu are leading domestic efforts to build an AI stack independent of US technology, with Huawei developing indigenous chips and Baidu advancing foundation models such as ERNIE. This is fueled by a massive, state-directed capital stack designed to achieve self-sufficiency. As de-risking and decoupling gain momentum, countries and companies are being forced to align their investment and procurement with either the US or Chinese ecosystems. The result is a digital dichotomy where capital flows mirror geopolitical blocs, shaping which platforms and suppliers become global standards. For investors, this bifurcation creates both risks and opportunities. While it fragments the global market, it also creates protected, high-growth ecosystems for the dominant players within each bloc.

Investor playbook and outlook

Portfolio construction and allocation strategy

Fund managers will likely capture asymmetric upside in the AI race if they allocate capital strategically along an optimized risk-reward ladder. The following 10 points are key considerations to take into account when investing in sovereign AI.

1. **Prioritize infrastructure enablers:** Focus investment on the “picks and shovels” of the sovereign AI build-out. Target private companies in advanced cooling, optical interconnects, and power management systems, as these are critical, high-margin bottlenecks for all datacenter construction.
2. **Gain exposure to advanced packaging:** The onshoring of semiconductor supply chains via US and EU public investment creates a durable tailwind for advanced packaging startups. These companies are essential for integrating heterogeneous chips (chipllets) required for AI workloads and represent a less crowded segment than GPU design.

- 3. Invest in the inference hardware ecosystem:** While NVIDIA dominates training, the inference market is more fragmented. Allocate capital to emerging chiplet-based accelerator companies and software startups optimizing inference workloads, as this is where the majority of long-term compute demand will reside.
- 4. Adopt a “US-plus-allies” strategy:** Concentrate capital in US-domiciled AI leaders and their key suppliers who are the primary beneficiaries of sovereign contracts from allied nations in the Middle East, Europe, and Asia. This strategy captures global growth while mitigating direct exposure to geopolitical hotspots.
- 5. Identify niche opportunities in European regulatory tech:** The EU’s AI Act, while a headwind for some, creates a captive market for startups specializing in AI compliance, auditing, and explainability. These regulatory tech firms offer noncorrelated growth as adoption becomes mandatory.
- 6. Leverage SWF partnerships in the Middle East:** For growth equity and private equity funds, seek co-investment opportunities alongside Gulf SWFs (for example, the UAE’s MGX and Saudi Arabia’s PIF). Their large-scale capital acts as a demand guarantee, de-risking infrastructure projects and providing clear exit pathways.
- 7. Avoid direct investment in China’s core AI stack:** Due to US sanctions and increasing market isolation, direct investment in Chinese semiconductor and foundation model companies carries substantial geopolitical risk. Funds may consider reducing exposure to publicly listed supply chain beneficiaries on the Hong Kong exchange.
- 8. Monitor the emergence of sovereign-specific applications:** While the application layer is nascent, begin due diligence on early-stage companies developing AI for regulated, sovereign-critical industries like defense, national security, and public health, as these will be the next frontier for procurement.
- 9. Increase allocation to datacenter private equity:** The unprecedented demand for power and physical space makes datacenter operators and developers a prime target. Invest in specialized PE funds like Blue Owl that are partnering with sovereign capital to build next-generation AI campuses.
- 10. Stress-test portfolios for supply chain bifurcation:** Actively model the impact of a deepening US-China technological decoupling on portfolio companies. Favor firms with resilient, multi-sourced supply chains and those aligned with the dominant Western technology stack.

Taken together, these strategies position fund managers to participate in the sovereign AI build-out while managing geopolitical and technological risk. The ability to align capital with secure supply chains, strategic partners, and long-term infrastructure demand may define the next generation of outperforming funds.

The next decade of sovereign AI

The rise of sovereign AI signals a structural realignment in the global technology landscape, driven by intensifying geostrategic competition. Governments are deploying capital at unprecedented scale to secure national control over compute, data, and models. This shift is creating a durable, state-backed investment cycle that spans the entire AI value chain. The market is no longer unified but is splitting into two distinct ecosystems led by the US and China. Within this bifurcated environment, the most accessible and investable opportunities for private capital remain concentrated in the US-aligned bloc, which continues to draw significant funding from sovereign wealth funds in the Middle East and Asia.

For investors, the focus should remain on the foundational layers of the AI stack. The strongest near-term opportunities are in enabling infrastructure such as semiconductors, interconnects, power systems, and cooling technologies that underpin sovereign compute expansion. As these assets scale and infrastructure matures, value creation will begin to shift from training to inference. This evolution will open new opportunities for specialized, energy-efficient hardware and optimized software platforms capable of delivering AI performance at scale. Navigating the sovereign AI landscape requires a new playbook that blends deep technology diligence with sophisticated geopolitical analysis. A holistic approach to these mega-trends is as necessary as is prudent. Upon this rock, sovereigns will build their datacenters, and investors will patiently wait to see who will wear the crown of circuits.

Additional resources

The following tables highlight public and private companies positioned to capture outsized gains from the global acceleration of AI adoption, amplified by the rise of sovereign investment programs. PitchBook's Q2 2025 AI comp set supports this view, with performance concentrated among publicly traded hyperscalers and semiconductor firms that sit at the core of the AI infrastructure build-out.

Select private AI & ML companies

Company	Category	Description
Crusoe Energy	Datacenter infrastructure	Leverages stranded natural gas to power modular, low-cost datacenters for AI workloads, directly benefiting from sovereign demand for sustainable compute.
Cerebras Systems	Semiconductors	Developer of wafer-scale chips for AI training, securing contracts with national labs and international sovereign partners.
Groq	Semiconductors	Produces high-speed inference chips (LPUs) that offer a compelling alternative to GPUs for latency-sensitive sovereign applications.
d-Matrix	Semiconductors	Develops chiplet-based inference compute cards for generative AI, positioned to capture demand as nations shift focus from training to deployment.
CoreWeave	Cloud infrastructure	A specialized cloud provider for large-scale AI, partnering directly with sovereign-backed entities to build out dedicated compute clusters.
Lambda Labs	Cloud infrastructure	Provides AI-focused cloud compute and hardware, attracting sovereign clients seeking alternatives to hyperscalers.
Celestial AI	Optical interconnects	Designs optical fabrics for datacenters, a critical enabling technology for the massive compute clusters funded by sovereign states.
Astera Labs	Interconnects	A newly public leader in PCIe and CXL interconnects that remove performance bottlenecks in AI datacenters.
Mistral AI	Foundation models	The leading European foundation model developer, positioned as a regional champion and beneficiary of EU sovereign initiatives.
Sarvam AI	Foundation models	India's leading foundation model startup, directly supported by the IndiaAI Mission to create sovereign-specific models.

Source: PitchBook • Geography: Global • As of August 12, 2025

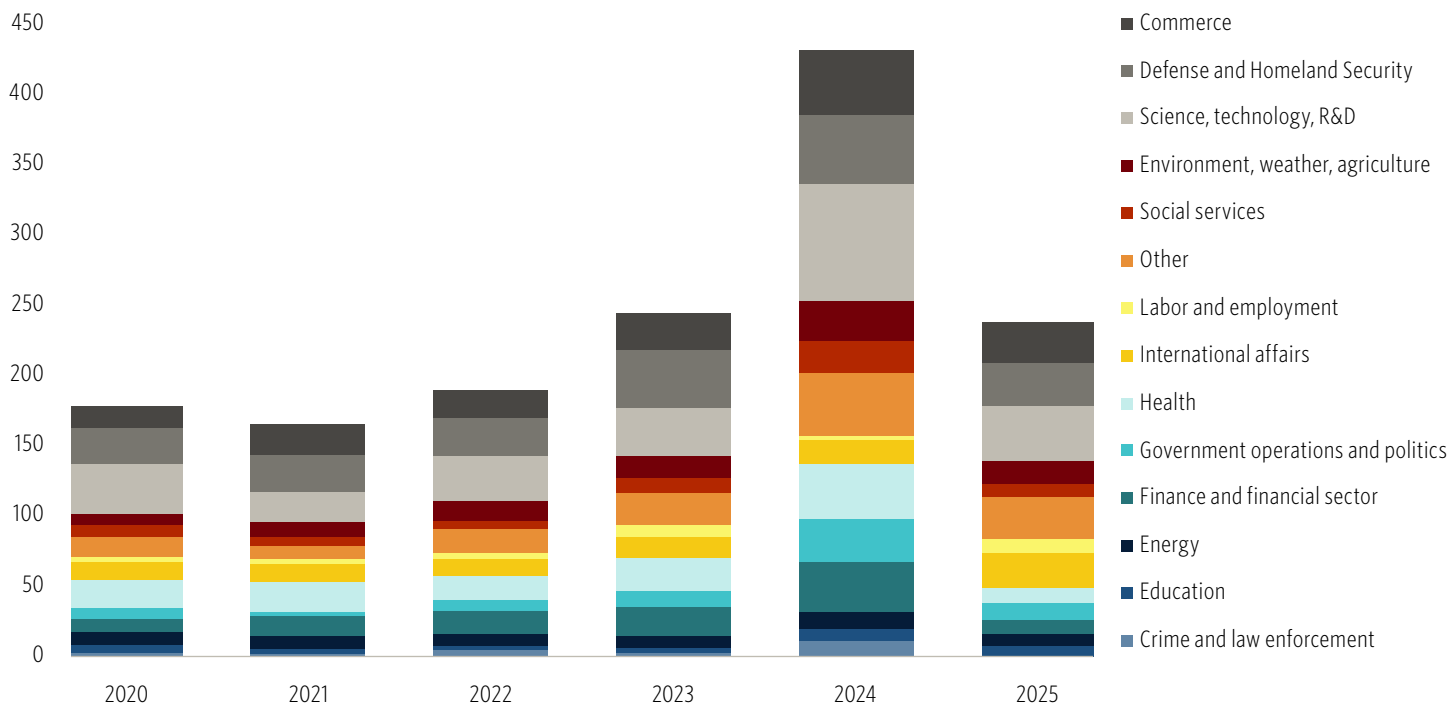
Select public AI & ML companies

Company	Category	Description
NVIDIA	Semiconductors	The dominant provider of AI training GPUs, serving as the foundational hardware layer for nearly every global sovereign AI initiative.
Microsoft	Cloud/models	A primary partner for sovereign nations through its Azure cloud and strategic investment in OpenAI, providing the core infrastructure and platform layer.
Oracle	Cloud infrastructure	Aggressively building out sovereign-focused cloud regions and partnering with NVIDIA and Middle Eastern SWFs to deliver national AI infrastructure.
TSMC	Semiconductors	The leading manufacturer of advanced AI chips, benefiting directly from CHIPS Act subsidies to build fabs in the US that will supply sovereign projects.
Vertiv	Datacenter infrastructure	Provides critical power and thermal management solutions, seeing massive order growth driven by sovereign-scale AI datacenter construction.
Eaton	Power management	A key supplier of electrical equipment for datacenters, benefiting from the immense power requirements of AI infrastructure.
Marvell Technology	Interconnects/networking	Producer of optical modules and custom silicon essential for networking within AI datacenters.
SK Hynix	Memory	The market leader in high-bandwidth memory (HBM), a critical component for AI accelerators, with demand directly tied to sovereign GPU build-outs.
CrowdStrike	Cybersecurity	As nations build sovereign AI clouds, securing these critical assets becomes paramount, driving demand for advanced cybersecurity platforms.
Palo Alto Networks	Cybersecurity	Provides network and cloud security solutions essential for protecting sovereign data and AI models from state-sponsored threats.

Source: PitchBook • Geography: Global • As of August 12, 2025

Legislative momentum for AI is building as evidenced by the number of times “AI” or “artificial intelligence” was used in legislation by US policymakers. A pattern of AI mentions also appeared in European legislative trends.

US government document count with mentions of AI (2020-2025)



Sources: [Congress.Gov](#), [Federal Register](#), and [GovInfo](#) • Geography: Global • As of July 18, 2025

Glossary

- **Sovereign AI:** A nation's capacity to independently develop, deploy, and govern artificial intelligence systems using its own infrastructure, data, and talent.
- **Sovereign wealth fund (SWF):** A state-owned investment fund that invests in real and financial assets on behalf of a national government.
- **CHIPS and Science Act:** A 2022 US law authorizing roughly \$280 billion in new funding to boost the domestic research and manufacturing of semiconductors.
- **Foundation model:** A large AI model trained on a vast quantity of data at scale, designed to be adapted to a wide range of downstream tasks (for example, GPT-4).
- **Inference:** The process of using a trained AI model to make a prediction or generate an output based on new input data.
- **Training:** The process of teaching an AI model to learn patterns from a large dataset. It is computationally intensive.

- **Graphics processing unit (GPU):** A specialized electronic circuit designed to rapidly manipulate memory to accelerate the creation of images. Currently, GPUs are the primary hardware for AI training.
- **High-bandwidth memory (HBM):** A high-performance RAM interface for 3D-stacked memory. HBM hardware is essential for feeding data to powerful AI accelerators.
- **Advanced packaging:** A set of semiconductor manufacturing techniques that enable the integration of multiple chips (chipselets) into a single, more powerful device.
- **Chipselet:** A small, specialized integrated circuit designed to be combined with other chipselets in an advanced package to form a larger, more complex system.
- **Optical interconnect:** The use of light (photons) to transmit data between chips, servers, and racks, offering higher bandwidth and lower latency than traditional electrical interconnects.
- **Compute Express Link (CXL):** An open standard interconnect for high-speed communication between CPUs, GPUs, and other accelerators.
- **Geopolitical bifurcation:** The splitting of the global technology landscape into two distinct, often incompatible, ecosystems, primarily one led by the US and one led by China.
- **Dual-circulation strategy:** China's economic strategy to reduce its reliance on overseas markets and technology while fostering domestic consumption and innovation.
- **EuroHPC JU:** The European High-Performance Computing Joint Undertaking, a legal and funding entity for developing a pan-European supercomputing ecosystem.
- **Application-specific integrated circuit (ASIC):** A chip customized for a particular use rather than for general-purpose use, often used for high-efficiency AI inference.
- **Data residency:** The requirement that data collected from citizens of a country must be stored and processed on servers physically located within that country's borders.
- **Bureau of Industry and Security (BIS):** An agency of the US Department of Commerce that deals with issues involving national security and high technology, responsible for administering export controls.

- **Total cost of ownership (TCO):** A financial estimate intended to help buyers and owners determine the direct and indirect costs of a product or system. For AI, it includes hardware, energy, and operational costs.
- **Latency:** The delay between a user's action and the time a system responds. It is a critical metric for real-time AI applications.
- **Throughput:** The amount of work a system can perform in a given period. It is a critical metric for batch-processing tasks like AI training.
- **Onshoring:** The practice of transferring a business operation that was moved overseas back to the country from which it was originally relocated.
- **Electronic design automation (EDA):** A category of software tools for designing electronic systems such as integrated circuits and printed circuit boards.
- **Extreme ultraviolet (EUV) lithography:** A next-generation semiconductor manufacturing technology that uses light of an extremely short wavelength to print complex patterns on silicon wafers.
- **Vertical application:** A software application designed to meet the specific needs of a particular industry or function, as opposed to a horizontal application that can be used across many industries.
- **Regulatory technology:** A class of software and technology solutions that help businesses comply with regulations efficiently and effectively. Also known as regtech.