



AI capex boom intensifies cash flow focus

Nimish Mathur, head of tech coverage for North America, and **Ugur Bitiren**, global head of sectors and advisory, International Trade and Transaction Banking at Crédit Agricole CIB, share their views on how the large-scale capital expenditure required to fund the artificial intelligence boom is pressuring players to optimise cash flows worldwide.

Artificial intelligence investments have boomed over the past couple of years. The graphics processing units (GPUs) in AI servers, ordered in the hundreds of thousands to build “AI factories”, can push each new project into the billions. At GTC 2025, Nvidia CEO Jensen Huang noted that a 1GW AI factory could cost as much as US\$40bn.

An “AI factory” refers to compute and networking equipment. In this article, we focus on financing the AI factory itself and the related cash flow challenges, contrasting this with the financing of real estate and power supply for data centres, which banks have typically funded through project finance structures.

The boom in AI factory capex

For perspective, AI Invest estimates that the top 11 cloud providers will invest more than US\$390bn globally in AI infrastructure capex in 2025.¹ It is important to note that around 75% of this spend is directed toward the AI factory

itself, with compute and networking equipment representing the most expensive component.

As a result, firms in the sector have been investing heavily in AI factory capex. Our estimates show that, on average, companies are allocating around 60% of their cash flow from operations (CFO) to capex (see Figure 1).

This spending is expected to be recurring, with advances in semiconductor design driving firms to purchase the latest generation as soon as it becomes available. While GPUs are typically depreciated over five to six years, we find they continue to be used beyond full depreciation as AI use cases expand. For the foreseeable future, we expect demand will be driven by the need for fresh capacity rather than simple replacement, meaning mid-term financing needs are likely to persist.

Despite the large cash balances of some companies sponsoring new AI factories, the sheer scale of capex spend has pushed them to optimise cash flows

as a financing source, complementing traditional debt, equity and, more recently, private credit.

At the same time, the number of players is growing with the emergence of “neoclouds” and telecom operators offering GPU-as-a-service, competing with established hyperscalers. This rise of new AI cloud providers is further supported by Europe’s data sovereignty concerns.

Financing bottlenecks

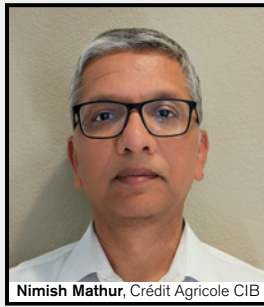
Purchases of GPUs from Nvidia and AMD, or application-specific integrated circuits from Broadcom, usually require deposits or binding purchase orders months in advance due to the strong demand for the latest generation of products. These purchases are costly and represent the first point of stress in the value chain. The GPUs are then assembled by Original Equipment Manufacturers and Original Design Manufacturers – such as Foxconn, Wistron, Quanta, Inventec and Supermicro – in a process that can take

30–45 days. Finally, buyers such as hyperscale cloud providers and emerging “neoclouds” are often able to negotiate extended payment terms with these manufacturers.

Companies in the middle of this value chain face a squeeze from both sides. This is just one of several bottlenecks involving high-value goods that could benefit from innovative financing structures. The strong demand for GPUs creates a liquid, high-value asset that can underpin a wide variety of financing solutions.

Operators of AI factories, which purchase these AI servers, then provide compute capacity to leading AI companies through long-term take-or-pay offtake agreements. While hyperscalers and large technology firms with ample cash reserves can absorb the capital expenditure with little difficulty, emerging neocloud providers and telecom operators must find alternative ways to fund such large investments. This creates the next bottleneck in the value chain. In addition to equity and debt (where accessible, such as for later-stage startups or listed companies), more creative structures – including working capital facilities and export credit-supported solutions that leverage the underlying asset (AI servers) – are gaining traction.

For AI factory operators, complexity is often compounded by delays in real estate development or power supply construction. In these cases, expensive equipment may be procured but cannot yet be fully deployed to generate revenue. To manage this, operators aim to delay server purchases to a just-in-time basis – without compromising access to the required computing and networking capacity. The high value and



The strong demand for graphics processing units creates a liquid, high-value asset that can underpin a wide variety of financing solutions.



marketability of these servers make such arrangements feasible.

At the end of the chain are the AI foundation model builders and startups. Companies such as OpenAI, Anthropic, Cohere and Mistral, as well as application developers like Perplexity and Harvey, typically follow an asset-light model. Instead of owning servers, they ‘rent’ the

compute from AI factory operators, often through multi-year GPU-as-a-service contracts. Given the scale, duration and take-or-pay structure of these agreements, they can be ring-fenced and monetised, creating yet another funding channel for capital-intensive AI factory operators.

Data sovereignty and domestic AI ecosystem developments

Data sovereignty concerns in Europe have driven providers to establish multiple AI factory sites across different jurisdictions and currency regimes. This added operating complexity has created a need for multinational AI players – both large and small – to implement efficient cash management systems tailored to their day-to-day operations.

In parallel, certain export credit agencies (ECAs) are exploring ways to finance local AI factories as strategic domestic investments, with banks lending directly to operators under ECA cover. This form of support can complement the other financing solutions already being deployed in the sector.

With respect to China, the AI ecosystem has developed rapidly. Domestic cloud providers such as Alibaba Cloud have invested heavily in AI factory deployments and pledged significant future commitments. While their initial focus was on the domestic market, these companies are now expanding into regions such as Southeast Asia, Emea and Latin America. In response to uncertainty over US chip export regulations – and with a desire to build in-house capabilities – Chinese players have also begun developing their own AI chips. These trends are driving greater financing needs, both for domestic and international AI factory investments.

Conclusion

Crédit Agricole CIB has been at the forefront in providing market-leading solutions across the artificial intelligence value chain. The scale of investment required means that many players, despite strong balance sheets, are seeking to optimise free cash flow. With its global presence, product expertise, and relationships across the AI ecosystem, Crédit Agricole CIB is well positioned to remain a strong partner to corporates in this space.

Reference

<https://www.ainvest.com/news/ai-infrastructure-boom-era-capital-spending-long-term-growth-2507/>

Figure 1 Capex/CFO evolution of AI firms

