80 YEARS

JAG CAPITAL MANAGEMENT LLC

Forward. Together.

JAG Growth Equity Thematic Insights: Q4 2025

Total reading time = 4 minutes

Need For Speed: Silicon Edition

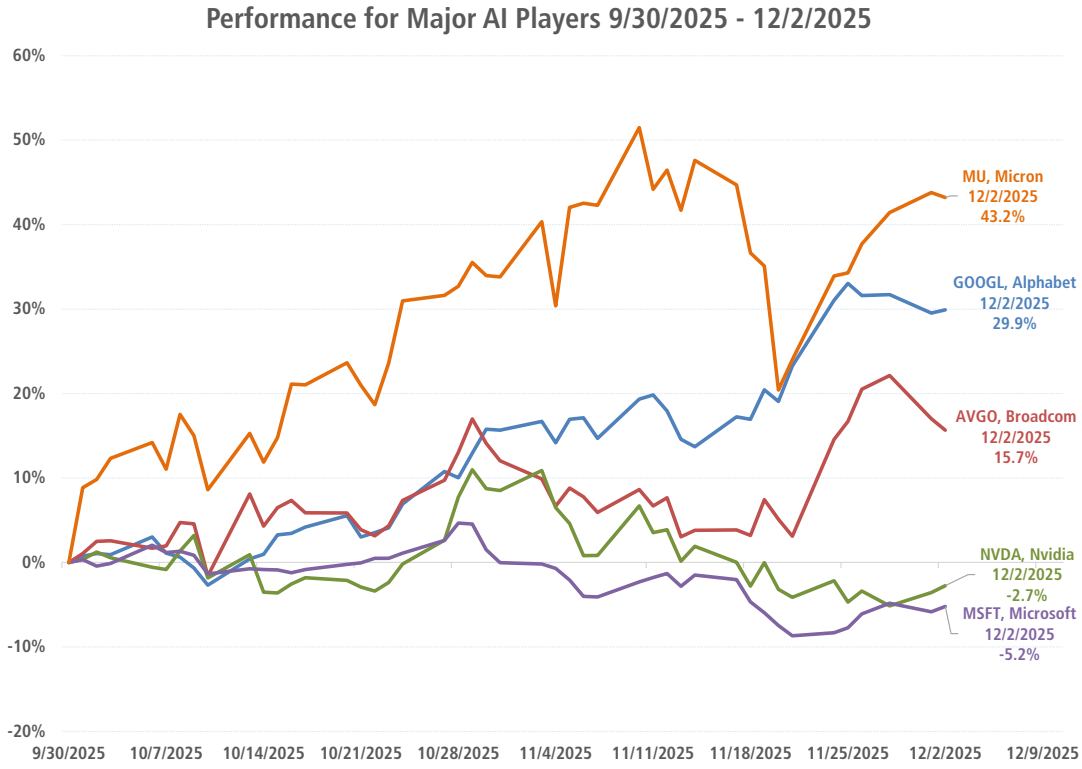
The AI landscape continues to evolve rapidly, with recent headlines highlighting a growing competitive threat to Nvidia (NVDA) as Google (GOOGL) is reportedly considering external sales of its custom AI accelerator chip, the tensor processing unit (TPU). This note will outline why Nvidia is likely to remain the dominant supplier of AI chips, and why memory manufacturers – often overlooked in the AI narrative – will continue to benefit from the ongoing rise in AI workloads.

GPUs vs TPUs: A Comparative Look

Nvidia developed the first graphics processing units (GPUs) to enable more advanced graphics rendering, with the technology used for many years primarily in video gaming applications. Because GPUs excel at parallel computing, they later became the go-to solution for training neural networks that power today's large language models (LLMs).

Google began developing TPUs after recognizing the need for a faster and more cost-efficient way to meet rising computing demand across its services, particularly following the rollout of its voice search feature. While Nvidia has always sold GPUs to external customers, Google has historically used TPUs exclusively for internal applications such as Search, YouTube recommendations, and more recently to train its AI models. The recent reports that Google could offer TPUs to outside customers have therefore understandably raised concerns about potential pressure on Nvidia's market position.

Figure 1: Recent Stock Performance of Select AI Players

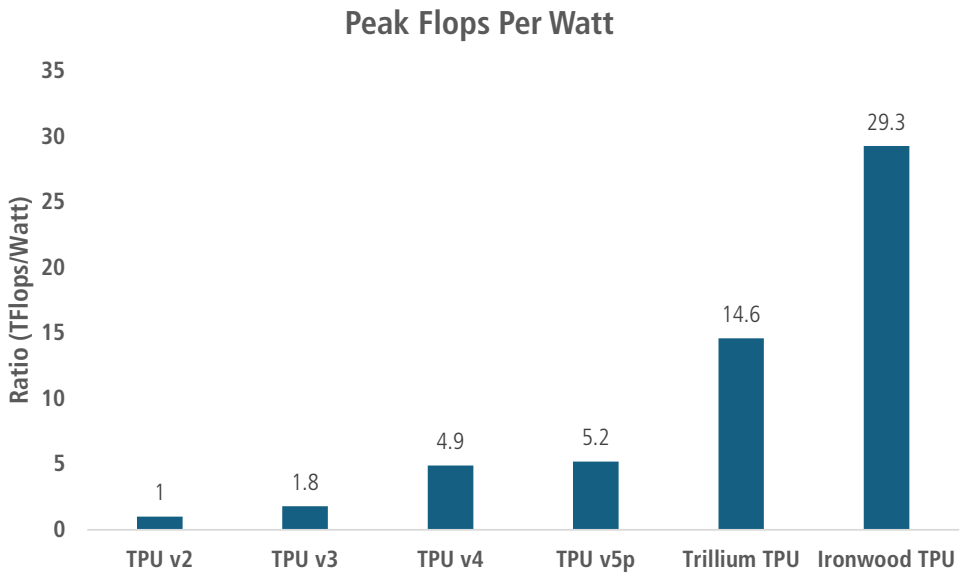


Source: FactSet, JAG Research

TPUs are ASICs (application-specific integrated circuits) that Google designed in collaboration with Broadcom (AVGO), a leading global semiconductor provider. JP Morgan estimated the high-end ASIC chip market at roughly $30B in 2024 and expects it to grow at 30% CAGR in the coming years. This is much smaller compared to the overall AI accelerator market, which AMD expects to exceed $500B by 2028. There are two primary reasons why GPUs have seen far wider adoption than ASICs: greater flexibility and a more mature software ecosystem.

While ASICs such as TPUs could offer significant cost and performance improvements versus GPUs for specific workloads, they are narrow in scope and difficult to adapt when workloads or architectures change. For example, most enterprises run more than just AI workloads (data analytics, recommendation engines, etc.) and GPUs can support all of them – while ASICs cannot. On the ecosystem front, Nvidia's extensive CUDA (Compute Unified Device Architecture) libraries allow engineers to write software that optimizes GPU performance, while TPUs are designed to work with a JAX/TensorFlow ecosystem that is less developed.

Figure 2: Google TPU Performance Improvement



Source: Google blog post

So given ASIC limitations, why has there been such optimism surrounding Google's TPU program recently? The primary reason is the successful rollout of Gemini 3.0 – Google's latest LLM that was reportedly exclusively trained on TPUs. Given rapidly changing training requirements that demand flexible hardware, LLM training is dominated by GPUs. However, Google's expertise across the entire stack – from model architecture to chip design – combined with its partnership with Broadcom for networking and supply chain expertise, has given TPUs a notable advantage over other ASIC efforts. Google's newest generation of TPU, Ironwood, shows strong improvements (Figure 2) that according to Google, rival or exceed Nvidia's Blackwell chips performance in specific workloads. This could lead to a material reduction in total cost of ownership (TCO) for specific, hyperscale training/inference workloads, and potential market share gains vs GPUs.

While Google is a formidable competitor, we do not believe TPUs will have a negative impact on Nvidia's growth or margin profile in the next few years. First, as already noted, GPUs are flexible and have a more developed ecosystem that makes them more suitable for most enterprises. Second, Ironwood TPU compares well vs Blackwell chips on specific workloads, but Blackwell chips became available roughly one year earlier. As a result, Nvidia is ahead on its roadmap and expects to launch its next generation of AI architecture, Vera Rubin, in the second half of 2026. The Rubin platform features significant upgrades, notably on the memory side (more on this later) that will likely materially exceed TPU's performance. Third, most enterprises use multiple cloud providers and moving the data from one to another can be very costly. TPUs are currently only available through Google Cloud Platform (GCP), but Nvidia's GPUs are available on every major cloud platform. This provides enterprises with greater flexibility and avoids a situation where they are forced to rely on a single cloud vendor (Google). Finally, we believe GPU vs TPU

is not a zero-sum game, and rapidly growing AI demand creates room for multiple winners. Google itself is spending billions of dollars to purchase Nvidia's AI accelerator chips. We expect capacity to stay tight as computing demand continues to outpace supply, with high-end GPUs virtually sold out.

Memory Chips – Underappreciated AI Enabler

Regardless of what side you take on the GPU vs TPU debate, one thing is clear – AI workloads need substantially higher memory capacity. For example, according to Google's blog post, Ironwood TPU has an HBM (high-bandwidth memory) capacity of 192 gigabytes, compared to "only" 32 gigabytes for TPU v4 that became available to GCP customers in 2022. Nvidia's standard Rubin chip is expected to feature 288 gigabytes of HBM capacity.

GPUs have become increasingly more powerful over the years, but they can only do processing as fast as the memory throughput allows. This limitation is known in AI industry parlance as the "memory wall." To address this limitation, chip makers have turned to HBM technology. HBM was first developed over a decade ago but has seen relatively little adoption until recently. This was primarily due to significantly higher manufacturing costs vs conventional memory chips due to HBM's 3D stacking technology. However, while conventional memory was generally sufficient for PC/gaming, its relatively limited bandwidth meant powerful AI accelerators sat idle while waiting on more data. Unlike conventional memory chips, HBM sits near or on the same package as the GPU and has a wider bus that transfers very large amounts of data simultaneously.

The memory industry, including the HBM segment, is an oligopoly with the top three players – Samsung, SK Hynix, and Micron – holding approximately 90% market share. Historically, the memory industry has been very cyclical and prone to boom-bust cycles. Like other commodity markets, memory producers generally lacked supply discipline and would overbuild amid surging demand – leading to the subsequent oversupply and plummeting memory prices. While supply discipline has improved as the industry has consolidated, memory markets continue to experience significant cyclicality. For example, during the most recent downturn that occurred in 2022, weak demand and high supply weighed on memory pricing. Eventually pricing started to recover as memory manufacturers significantly reduced supply. At the same time, the AI buildout has introduced significant tailwinds on the demand side of the equation. Collectively, these dynamics have led to an unprecedented spike in memory pricing.

We expect the memory industry to remain highly cyclical, but the current robust demand for HBM solutions should drive better pricing and improved profitability. All available HBM capacity is already sold out for next year, with memory producers moving to more frequent contract negotiations to achieve higher pricing.

JAG Portfolio Positioning

JAG's Large Cap Growth (LCG) strategy is broadly positioned to benefit from these dynamics. Nvidia, Google, and Broadcom are among the strategy's five largest holdings as of the date of this report. LCG also maintains sizeable positions in AMD (AMD) and Micron (MU), which should see tailwinds from rising data center GPU demand and improving memory pricing, respectively. In addition, Taiwan Semiconductor Manufacturing (TSM) remains well positioned as a key foundry supplier for leading-edge chips, including Nvidia's GPUs and Google's TPUs.

Outside of public cloud providers and semiconductor companies, LCG also has exposure to businesses that stand to benefit from AI adoption across software, services, and enterprise infrastructure. These include leading AI platform provider Palantir Technologies (PLTR), networking leader Cisco Systems (CSCO), and data center power and cooling systems supplier Vertiv (VRT). We believe this broad-based and balanced exposure positions our portfolio well to benefit from accelerating demand for AI-driven solutions.

– JAG's Growth Equity Research Team

Disclosures

This report was prepared by the staff of JAG Capital Management, LLC, an SEC-registered investment adviser. The information herein was obtained from various sources including but not limited to FactSet, Bloomberg, Reuters, Standard & Poor's, ChatGPT, and the United States Bureau of Labor Statistics, and believed to be reliable; however, we do not guarantee its accuracy or completeness. The information in this report is given as of the date indicated. We assume no obligation to update this information, or to advise on further developments relating to securities discussed in this report. The opinions expressed are those of the adviser listed above as of the date of this report and are subject to change without notice. The opinions of individual representatives may not be those of the Firm. Additional information is available upon request.

The information contained in this document is prepared and circulated for general information only. It does not address specific investment objectives, or the financial situation and the particular needs of any recipient. Investors should not attempt to make investment decisions solely based on the information contained in this communication as it does not offer enough information to make such decisions and may not be suitable for your personal financial circumstances. You should consult with your financial professional prior to making such decisions. For institutional investors: JAG Capital Management, LLC, has a reasonable basis to believe that you are capable of evaluating investment risks independently, both in general and with regard to particular transactions or strategies. For institutions who disagree with this statement, please contact us immediately. **Past performance should not be considered indicative of future performance. Any investment contains risk including the risk of total loss.**

This document does not constitute an offer, or an invitation to make an offer, to buy or sell any securities discussed herein. J.A. Glynn & Co., JAG Capital Management, LLC, and its affiliates, directors, officers, employees, employee benefit programs and discretionary client accounts may have a position in any securities listed herein.

A Veteran Owned Business (VOB) is a business owned, (51% ownership or greater) by a Veteran who has met the definition of a veteran. The following represents the criteria that the National Veteran Business Development Council (NVBDC) uses in determining ownership:

Ownership: Fifty-one percent ownership by a Veteran or Veterans. The applicant must share in all risk and profits commensurate with their ownership interest.

Control and Management: Proof of active management of the business. Veteran must possess the power to direct or cause to direct the management and policies of the business.

Contribution of Expertise and Capital: Contribution of capital and/or expertise by Veteran owner(s) to acquire their ownership interest shall be real and substantial and be in proportion of the interest acquired.

Independence: The Veteran owner(s) shall have the ability to perform in their area of specialty/expertise without substantial reliance on non-Veteran-owned businesses.

About JAG

JAG Capital Management (JAG) actively invests for institutions and individuals in highly selective, customizable, and nimble equity and fixed income strategies. JAG is a boutique, independent, employee-owned investment management firm in St. Louis.

JAG CAPITAL MANAGEMENT LLC

1610 Des Peres Road, Suite 120
St. Louis, MO 63131