



### Investment Process

Our investment approach is based on idea generation, a systematic framework for analyzing companies and proactive risk management. Utilizing this approach, we seek to construct a focused portfolio designed to maximize alpha while limiting downside risk over the long term.

#### Idea Generation

We believe a key element in alpha generation is finding areas where our views on industry fundamentals differ from consensus estimates. In this pursuit, we seek to identify inflections in multi-year trends which may be caused by changes in supply/demand dynamics, societal behavior, market conditions, technology, laws/regulations and business models, among other variables. We believe these inflections are often misunderstood by market participants, and can lead to powerful re-ratings of industries and companies. Identifying themes helps us develop a focused universe of companies to analyze more thoroughly.

#### Systematic Analytical Framework

We apply a systematic framework for analyzing companies across sectors and themes, creating a repeatable and methodical decision-making process. Our proprietary company models focus on multi-year earnings power differentiation, expected outcome scenario analysis, return on invested capital and discounted cash flow valuations. Visual outputs are then produced through our internally developed technology solutions, allowing us to consistently evaluate positions across the portfolio.

#### Proactive Risk Management

We incorporate risk management into all stages of our investment process. Metrics evaluated include crowding, correlation, volatility, stress tests, liquidity, factor analysis and macro drivers, all of which inform portfolio construction and position sizing. We also use various instruments, such as options, in an effort to magnify alpha and minimize downside.

### Team Leadership

The investment team applies the same approach to idea generation and fundamental company analysis that Portfolio Manager Chris Smith has honed throughout his career. Research analysts are sector specialists with deep knowledge of their coverage areas. Our process blends a collaborative team mentality with individual accountability.

### Portfolio Management



Christopher Smith  
Portfolio Manager

### Investment Results (% USD)

As of 30 June 2025	Average Annual Total Returns						
	QTD	YTD	1 Yr	3 Yr	5 Yr	10 Yr	Inception <sup>1</sup>
Composite — Gross	20.25	18.28	30.00	22.12	16.87	—	19.83
Composite — Net	19.97	17.71	28.74	20.92	15.72	—	18.66
S&P 500® Index	10.94	6.20	15.16	19.69	16.63	—	14.35

### Calendar Year Returns (% USD)

	2020	2021	2022	2023	2024
Composite — Net	29.53	23.95	-25.67	15.93	31.39

Source: Artisan Partners/S&P. Returns for periods less than one year are not annualized. <sup>1</sup>Composite inception: 1 May 2017.

Past performance does not guarantee and is not a reliable indicator of future results. Current performance may be lower or higher than the performance shown. Composite performance has been presented in both gross and net of investment management fees. The Strategy's investments in initial public offerings (IPOs) made a material contribution to performance. IPO investments may contribute significantly to a small portfolio's return, an effect that will generally decrease as assets grow. IPO investments may be unavailable in the future.

**Investment Risks:** Investments will rise and fall with market fluctuations and investor capital is at risk. Investors investing in strategies denominated in non-local currency should be aware of the risk of currency exchange fluctuations that may cause a loss of principal. These risks, among others, are further described near the back of this document, which should be read in conjunction with this material.



“I don't want an organization of order-takers because order-takers don't take responsibility for results... Innovation and progress are achieved only by those who venture beyond standard operating procedure.”  
 —D. Michael Abrashoff, “It’s Your Ship”

**Exhibit 1: Investment Results**

As of 30 Jun 2025	Q2 2025	YTD 2025	Cumulative (since inception)	Annualized (since inception)
Antero Peak Strategy (Net)	19.97%	17.71%	304.78%	18.66%
S&P 500® Index	10.94%	6.20%	199.09%	14.35%
Strategy Alpha	902bps	1,150bps	10,570bps	431bps
S&P 500® Equal Weight Index	5.46%	4.82%	133.65%	10.95%
Strategy Alpha	1,451bps	1,288bps	17,114bps	771bps

Source: Antero Peak Group/S&P. As of 30 Jun 2025. **Past performance does not guarantee and is not a reliable indicator of future results.** Returns greater than one year are annualized unless otherwise noted. Strategy inception: 1 May 2017.

In Q2, the Antero Peak Strategy (net) outperformed the S&P 500® Index by 902bps. This follows a strong 2024 and Q1 2025, in which the Strategy outperformed the benchmark by 637bps and 239bps, respectively. We are pleased with our continued strong performance in the first half of 2025, having achieved YTD relative outperformance versus the S&P 500® Index of 1,150bps.

Despite an ever-swinging pendulum of macro conditions—tariffs/no tariffs, inflation/no inflation, rate cuts/no cuts, risk-on/risk-off, global unrest/world peace—we believe our portfolio and risk management execution have shown great consistency in 2025.

**Performance Review**

Once again, our results were tied to the fundamental performance of our underlying investments. Top contributors in Q2 included Microsoft, General Electric, Netflix, Broadcom, Rolls Royce and Siemens Energy. During the quarter, these companies experienced sharp positive earnings revisions across the board. This stood in stark contrast to the S&P 500® Index, which saw 2025 estimates revised 3.6% lower. Our portfolio, on a blended average basis, experienced +4.6% upward revisions, with the majority of our portfolio companies faring better than the benchmark.

**Exhibit 2: Q2 Execution Was Excellent Across Sectors and Geographies**  
 Selected Portfolio Holdings Fiscal Year Earnings Estimate Revisions During Q2

Industry	Q2 Changes in 2025 FY Estimates
EU Industrial	+5.6%
Aerospace	+2.5%
EU Aerospace	+8.0%
Internet	+6.7%
Semiconductor	+4.8%
Software	+2.0%
Capital Goods	+1.7%
Electrical Equipment	+5.9%
Utility	+4.2%
Financial Services	+0.4%

Source: Artisan Partners/Bloomberg/FactSet. As of 30 Jun 2025. **Past performance does not guarantee and is not a reliable indicator of future results.** Based on a representative account within the Antero Peak Strategy Composite. Industry classifications are determined solely at the discretion of the team. EU Industrial (Siemens Energy), Aerospace (General Electric), EU Aerospace (Rolls Royce), Internet (Netflix), Semiconductor (Broadcom), Software (Microsoft), Capital Goods (AXON), Electrical Equipment (GE Vernova), Utility (Vistra), Financial Services (Mastercard). Selected portfolio holdings are provided for illustrative purposes only and are not representative of all portfolio securities classified by industry.

Bottom contributors for Q2 included Equinix, Crown Castle, Apollo Global Management, Primo Brands, Safran SA and VanEck Semiconductor ETF. The absolute impact of our “losers” was mild, with only one stock impacting us by more than 50bps (VanEck Semiconductor ETF). This led to a very strong slugging ratio of 2.5 (the ratio of dollars gained on “winners” versus dollars lost on “losers”) and a Sharpe ratio at 4X that of the benchmark. Good risk management was also on display in Q2—we had just 76% downside into the market lows on April 8, and as markets recovered, we experienced 120% upside capture.

The outlook for the portfolio remains remarkably strong, in our view. Based on consensus estimates, our holdings are projected to outgrow the S&P 500® Index by approximately 4% in the second half of 2025. Our research also supports a blended 5% upward revision to consensus earnings estimates for our portfolio companies over the same period. At the same time, we see downside risk to overall S&P 500® Index EPS, which has continued to trend lower as of this writing—particularly for 2026 and 2027. This dynamic sets the stage for a widening relative growth gap that could approach double digits in the second half of 2025. Meanwhile, our portfolio has continued to generate twice the return on invested capital (ROIC) of the index, with a slightly lower cost of capital.

Overall, we're very pleased with our execution across both 2024 and the first half of 2025. Our process has shown strong performance across a variety of elements, including modeling accuracy, sector allocation and risk management.

### A Culture of Accountability

At just 36 years old, Michael Abrashoff was appointed commander of the USS Benfold, a US Navy destroyer equipped for anti-aircraft warfare and among the first vessels fitted with the Aegis Ballistic Missile Defense System. By 2010, the Benfold became the first ship capable of simultaneously defending against both ballistic and cruise missiles.

When Abrashoff took command, the Benfold was one of the worst performing ships in the Navy. Within a year, it became the best. He accomplished this remarkable feat by transforming the ship's culture. He listened intently to his crew, eliminated bureaucratic friction and empowered every sailor to take ownership of the mission.

The title of his book, "It's Your Ship," captures this ethos. Abrashoff encouraged each sailor to act as though they were the captain—to make decisions, speak up and be accountable. He removed the traditional command-and-control mindset and replaced it with decentralized leadership rooted in trust, humility and clear expectations. That mindset resonates deeply with our team. In high-performance environments—whether on a destroyer or in an investment team—culture drives outcomes. Ownership, humility and a shared sense of purpose are not just ideals; they're strategic edges just as much as good models are.

"Show me an organization in which employees take ownership, and I will show you one that beats its competitors."  
— D. Michael Abrashoff, "It's Your Ship."

When I founded the Antero Peak Group in 2016, we started with just four senior investment team members. Today, we are all still here. The continued presence of these core team members speaks not only to their individual talent and commitment, but also to the culture we've built—one rooted in trust, mutual respect and shared ownership.

We've created an environment where people feel empowered, accountable and invested in the outcome—not just their own, but the team's and the firm's. That sense of alignment is intentional. We don't separate performance from culture; we believe the two are inseparable. People who thrive here think like owners, take initiative and hold themselves to the highest standards.

The longevity of our core team isn't accidental—and it's not in spite of the high bar we set. It's because of it. High expectations attract those who want to be challenged and who take pride in building something lasting. That's the foundation we started with, and it remains central to who we, the Antero Peak Group, are today.

### Revisiting Batting and Slugging

In the [past](#), we have discussed the interplay of batting average and slugging percentage. Batting average represents the percentage of stocks active managers select that beat their benchmark. Slugging percentage, on the other hand, represents the amount gained on "winners," divided by the amount lost on "losers." For the three-year period ending March 31, 2025, the median batting average for large-cap active managers was 45%, with the top quartile around 51%.

The commonly held belief that success in investing stems from being right more than 50% of the time oversimplifies a much more nuanced reality. In our view, and consistent with insights from leading investors like Stanley Druckenmiller, the true driver of long-term alpha is not how often you're right, but how much you make when you are.

Our investment process prioritizes slugging over batting average. By concentrating capital in high-conviction ideas identified through rigorous diligence and ongoing scrutiny, we seek to maximize the asymmetry between "winners" and "losers." While this approach demands more upfront research and greater accountability, we believe it allows us to capture outsized returns when we are right and cut losses quickly when we're not.

This high-conviction focused strategy has delivered tangible results. Just as baseball evolved beyond batting average to more predictive stats like on-base plus slugging (OPS), we believe investing should evolve to focus on the true drivers of outperformance: payoff ratios, position sizing and disciplined risk management. For example, for an equally weighted portfolio with ~10% single-name volatility, annual returns are surprisingly

insensitive to batting average. Raising the batting average of this theoretical portfolio from 50% to 55% could add a mere 100bps of alpha, while improving the slugging has five times the impact.

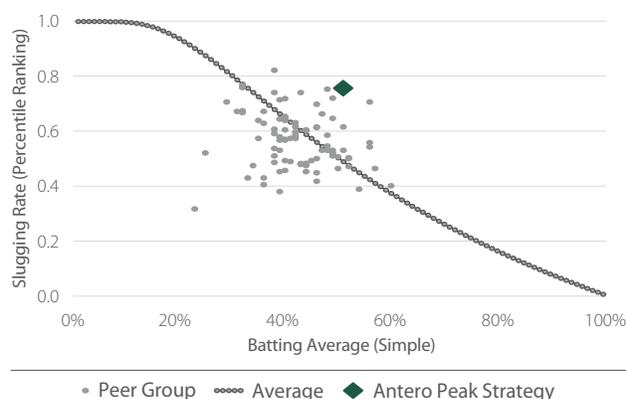
We follow a disciplined approach grounded in two core principles: First, it is likely we will often be wrong; second, it is likely we will often be right. By fully embracing these truths, we focus on cutting losing positions quickly and sizing winning positions aggressively—an approach that has been key to driving our historical returns.

Culture is essential to executing this framework effectively. Team members must set aside personal bias and the need to be “right,” instead approaching decisions with humility and an acceptance that being wrong is part of the process. This mindset is not just tactical—it’s character-driven. As Mike Abrashoff taught the sailors on his battleship, the team must prioritize the mission—in our case, the portfolio—over ego or personal ambition.

Industry norms—like low turnover and balanced sector exposure—can also hinder this dynamic, suppressing the payoff ratio and diluting the potential for outperformance. In fact, even with strong research and a decent hit rate (in the top quartile), these constraints often result in returns just above fees.

Since late 2017, our portfolio has delivered a 52% hit rate with a 300% absolute slugging ratio (240% relative to the S&P 500® Index). This validates our approach, which emphasizes rigorous time allocation to both our largest positions and those that are underperforming. Our process has helped us identify negative fundamental revisions early, which can lead a dramatically reduced hit rate on a go-forward basis. This enables us to objectively reduce risk fast. This focus on maximizing payoffs while minimizing losses is at the heart of our success.

**Exhibit 3:** Essentia Analytics Percentile Scatterplot of Comparable Portfolios



Source: Antero Peak Group/Essentia Analytics. As of 16 May 2025. **Past performance does not guarantee and is not a reliable indicator of future results.** Based on a representative account within the Antero Peak Strategy Composite. Chart is provided for illustrative purposes only as a visual representation of the batting average and percentile slugging rate of the peer group with a best-fit line showing the average. The line is bent due to the use of percentile rankings for slugging. Data measurement period is the most current available (1 Jan 2024 through 16 May 2025). Peer group includes 134 actively managed portfolios within the Essentia Analytics database. Portfolios are inclusive of separately managed accounts, mutual funds and ETFs with varying styles (e.g., value and growth), market caps, regions and benchmarks.

### A Focus on Alpha Over Beta

We believe there’s a tendency to categorize managers as value, growth, core and so on. To be clear, we are an alpha-focused portfolio. Our goal is to outperform the S&P 500® Index every year—regardless of style factors or market conditions. This objective is supported by the quantitative analysis of our performance. We’ve outperformed the S&P 500® Index in six of the last nine calendar years, including 1H 2025.

It’s particularly instructive to examine the sources of our tracking error over this period. Since inception, our asset-specific (idiosyncratic) risk has averaged approximately 55%. When combined with industry risk, nearly three-quarters of our relative tracking error can be attributed to stock and sector selection. In contrast, style factors such as value, growth, momentum and overall market direction—elements beyond our control—have had a far smaller impact. In our view, this is exactly what one should want from a true active manager: returns driven by security and sector selection, not by style drift or macro conditions.

#### Exhibit 4: Deconstructing Our Tracking Error

Antero Peak Strategy (Net) Metrics		
Alpha ITD, Total/Annualized	+10,570bps/+431bps	
Beta	0.93	← Lower Beta
Portfolio Volatility	16.7%	← Similar Volatility
Benchmark Volatility	16.2%	
Tracking Error Deconstruction		
<i>Idiosyncratic (Idio)</i>	54.8%	Most of our tracking error was driven by idiosyncratic risk and industry selection
<i>Industry Selection</i>	17.7%	
<b>Total Idio + Industry Selection</b>	<b>72.4%</b>	←
<i>Value and Growth</i>	2.3%	Our relative alpha was not tightly tied to changes in factors like value, growth, momentum and broader market direction
<i>Momentum</i>	6.9%	
<i>Market Direction</i>	<1.0%	

Source: Antero Peak Group/Bloomberg/FactSet/S&P. As of 30 Jun 2025. **Past performance does not guarantee and is not a reliable indicator of future results.** Based on a representative account in the Antero Peak Strategy Composite. Risk model provided for illustrative purposes only, is based on the team's analysis and is subject to material revisions. Composite Inception: 1 May 2017. Volatility measured by standard deviation. Idio (idiosyncratic or asset-specific risk) represents the risk attributable to a particular portfolio company, exclusive of overall market factors. Beta, volatility and tracking error measures based on monthly returns beginning May 2017, the first full month since inception. Tracking error statistic is comprised of idio and factor risk (industry, value, growth, momentum, market direction and other factors not outlined above). The figures shown for these risk factors represent their average respective contribution to tracking error over the measurement period.

#### Conclusion

Given recent client discussions and the increasing relevance and importance of artificial intelligence (AI), we thought we would share our perspective with you on the AI infrastructure landscape and why we think this tectonic shift in computing architecture will continue to provide a fertile investment period. We expect the underlying investment in infrastructure—silicon, energy and data pipelines—to remain very robust through 2030, to support the next wave of generative-AI growth, agentic AI. We hope you find our thoughts in the pages that follow useful and welcome the opportunity to discuss it further with you.

As you all know, as part of our personal and professional development, we keep a book club of sorts. This quarter, we read "It's Your Ship." An extended list can be viewed [here](#).

Thank you for your continued trust and partnership.

---

#### ARTISAN CANVAS

Timely insights and updates from our investment teams and firm leadership

Visit [www.artisancanvas.com](http://www.artisancanvas.com)

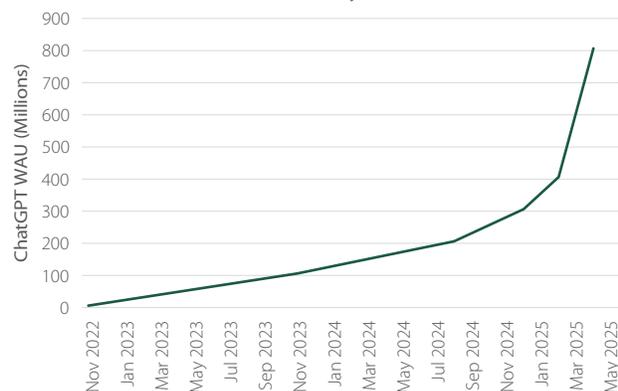
---

#### AI Update

As we previously wrote about in our [Q1 2023](#) investment letter, the Cambrian moment for AI decisively arrived in November 2022, with ChatGPT bursting onto the scene. ChatGPT has now scaled to nearly one billion monthly active users (MAUs) in under three years. This explosive expansion is reshaping capital markets, industrial supply chains and corporate strategy. For the investing community, it has also introduced a lot of volatility in AI stocks driven by 1) scepticism around the sustainability of infrastructure investment, 2) the return on investment (ROI) and use cases of AI, and 3) misconceptions surrounding the implications of new model architectures (e.g., the DeepSeek moment, which caused NVIDIA's stock to decline nearly 17% in one day, the largest one-day market value decline for any stock in history).

These fears seem misplaced in hindsight, as each of the worries have been addressed in the past six months: 1) hyperscale capital expenditure has materially stepped up YTD, with Amazon, Google and Meta explicitly raising forecasts, and OpenAI beginning to ramp up the initial \$100 billion phase of Stargate; 2) new agentic AI use cases have emerged, and revenue growth at Azure has materially accelerated; and 3) ChatGPT's weekly active users (WAUs) have more than doubled YTD as reasoning models have driven a material inflection in tokens (usage).

**Exhibit 5: Growth of ChatGPT's Weekly Active Users**



Source: OpenAI. As of 11 Apr 2025.

### AI Progress and Scaling

We do not envision a peak in AI spending on either training or inferencing through the end of the decade. The continued acceleration of generative AI (GenAI) is placing stress on the existing infrastructure, whether it's the supply chains of the physical chips, the data centers or the power supply. We hypothesize that capital spending on AI infrastructure will remain frantic until we see the capabilities of AI driven by the largest deployed data center projects.

Hyperscalers, the neoclouds (OpenAI, xAI, Anthropic, etc.) and now large sovereign AI projects (HUMAIN) are investing in larger training clusters (to advance model capabilities) and in densifying their accelerated compute architectures to reduce token per unit costs of inferencing as agentic AI has driven a material acceleration in usage patterns.

Over the past 24 months, the industry has moved from a "train-at-any-cost" mindset toward an operating model where inference drives the majority of incremental demand. Yet, the two workloads remain interdependent and place markedly different stresses on silicon, networking and data center economics.

Our data center tracker has highlighted significant projects being planned for the next 4–5 years. The Stargate development alone is projected to potentially reach ~\$500 billion of capital deployed and require 10 GWs of power. If we haven't reached artificial general intelligence (AGI) following this scale of investment, we may see a slowdown thereafter. However, given the recent improvements in model capabilities, we think AI investment will remain robust through the end of the decade.

In the following section, we address the investment pipeline for large-scale training clusters and how expanding agentic use cases and positive ROI on spend will require more inferencing infrastructure.

### Training Clusters Are Getting Bigger and Mega-Scale Infrastructure Projects Runways Are Long

Scaling laws, the continued growth in parameter counts and more compute-intensive post-training enhancements are all driving increasing spend on training large language models (LLMs). We continue to see larger training clusters being built, as we think this higher initial training cost can create incremental model capabilities that importantly reduce future inferencing costs. The superior compute capacity effectively unlocks new model architectures that can reduce per-token flops and materially reduce memory requirements.

The computational demands of LLMs scale with parameter count and data volume. The pre-training phase often involves petabyte-scale datasets. For example, Baidu's (the Chinese search engine) foundational model Ernie 4.0 is widely reported as having multiple trillion parameters, while GPT-4 and DeepSeek-V3 are both reported to utilize 1 trillion parameters. Meta has delayed its Behemoth LLM multiple times, which is expected to be the largest version of the Llama 4 family, with a parameter count 400 billion.

Innovation has been apparent in post-training techniques, with the larger models supporting Mixture of Expert models (MoE), key-value cache compression and new distillation techniques. GPT-4 is an MoE model and has almost 10X the parameter count of GPT-3 (1.7 trillion versus 175 billion parameters). Effective new post-training techniques, such as reinforcement learning and chain of thought reasoning models, explore multiple solution paths that drive higher computational intensity. For companies that are trying to drive scale, investing in bigger training runs can help minimize lifetime model costs, which can compound over billions of queries.

The largest training cluster to date is the Colossus supercomputing cluster built out by xAI to power Grok3 and combines 200,000 GPUs. However, this will likely seem small in the coming years, given the number of hyperscalers developing plans and dedicating resources to develop

greater than one million GPU clusters. For example, the initial \$100 billion phase of the Stargate project will require 1.2 GWs of power (this alone is 20% of the entire power consumption of the NOVA area, which is the largest data center market globally and has been built up over decades).

Planned data center capacity expansions in the US have increased 150% over the past year—with almost 40 GWs of expansion currently in the planning and development phase across the country. Similarly in Europe and Asia, expansion plans have more than doubled in the last year. Some of the largest data center developments are yet to come online, with numerous 1 GW deployments scheduled to come online in the next few years. This is a huge amount of data center capacity that is still to be built, before any compute architecture is deployed. Sovereign AI factories are emerging, with multiple deals announced in the Middle East over several months. NVIDIA on its most recent earnings call confirmed this level of demand, noting nearly 100 AI factories built in the quarter and noting it sees visibility to projects requiring “tens of GW” of power.

**Exhibit 6: Upcoming Data Center Deployments**

Project	Location	Tenants	Scale	Expected
Stargate	USA	OpenAI/OCI	>10 GW	Phase from 2025+
Sines	Portugal	Hyperscalers	1.2 GW	2025+
Amazon Project Rainier	USA	Amazon	2.2 GW	2026+
Quantum Loophole	USA	Hyperscalers	>2 GW	2026
Wonder Valley	Canada	Hyperscalers	>7 GW	1.4 GW by 2027–2028
Mistral	France	Mistral	1.4 GW	2028
Meta AI Cluster	USA	Meta	>2 GW	2030
HUMAIN	Saudi Arabia	PIF	1.9 GW	2030

Source: Artisan Partners.

### Agentic AI Era Could Drive Material Spend in Inferencing Infrastructure With Addition of Billions of New Computing Users

While blockbuster training cycles could continue to set narrative peaks, sustained value should accrue to vendors and utilities that can deliver low-latency, low-cost inference at scale. The transition from predictive models to autonomous agents is underway, unlocking new revenue pools but demanding exponential compute resources. NVIDIA’s Blackwell family—particularly GB200 and GB300 rack-scale systems—should provide the performance per watt and per token that make large-scale agentic AI economically feasible. We think enterprises that align their data center strategies with this roadmap could be positioned to capitalise on the next wave of AI-driven productivity.

Agentic AI systems can scale out by working together and can automate and execute tasks with minimal human oversight. When an agentic AI receives a task (e.g., “plan a marketing campaign,” “write a Python script,” “summarize or edit this document”), the input text is broken down into tokens (converting human text into numerical format that the model can process). This token generation is compounding.

In its Q3 FY2026 earnings call, Microsoft noted that it had processed 100 trillion tokens in the quarter, up ~5X YoY within Azure OpenAI services. At this scale and growth, Microsoft is already driving multi-exabytes of traffic. While inference often operates at lower precision (FP16, FP8) to reduce memory usage, this does not mitigate the cost, as agentic AI systems simply generate more tokens per query. For popular models like ChatGPT, serving billions of requests daily means the total compute cost can surpass the initial training expense.

To put it simply, the rise in agentic AI has effectively added billions of new users to existing compute infrastructure in less than two years, overloading the system and necessitating scaled accelerated compute infrastructure investment.

### Large-Scale Investment Requires Positive ROI

Large-scale infrastructure investment needs to be met with positive ROI for it to be sustainable. Two proof points of this positive ROI include 1) better profitability among the hyperscalers; and 2) expansion of use cases of agentic AI that are driving tangible productivity gains. As an investor, this creates an interesting opportunity around the specialized hardware that is driving the inference cost per token lower, including GPUs and custom ASICs.

Hyperscalers have reported positive gross margin developments from their AI services investments. Azure and Amazon Web Services (AWS) have both called out how AI services gross margins are tracking above cloud gross margins at commensurate scale and the companies expect terminal margins to be similar. This was surprising to us, as initial cloud deployments had very strong total cost of ownership (TCO) arguments for lift and shift from existing on-premise workloads, while AI services have typically been net new workloads. This margin profile has allowed Microsoft to materially beat its internal margin expectations over the past 12 months, now guiding operating margins to be up YoY, versus an initial expectation for ~100bps dilution from its AI investments. We believe this tangible ROI realization has been a key driver of accelerated investment in the underlying compute architecture—we have seen large capital expenditure increases this year at Amazon, Google, Meta and Oracle. Meanwhile, Microsoft has shifted its capital expenditure plans in the next 12 months to favor compute chips (GPUs and CPUs) over data center infrastructure.

## Expansion of Agentic AI Use Cases Are Helping Drive Productivity Enhancements

Industries continue to face pressure to improve efficiency and reduce costs, driving the adoption of automation and agentic AI use cases. Multiple Agent System (MAS) frameworks facilitate collaboration among agents, each specializing in tasks like natural language processing, computer vision or data analysis. MAS is crucial for applications requiring diverse expertise, such as software development or architectural design.

Customer service has emerged as a leading use case for agentic AI due to its ability to enhance efficiency and personalization. For example, Salesforce's Agentforce has enabled AI agents to handle customer queries autonomously, escalating only complex issues to human agents. Adoption is not just siloed to the technology industry. Wells Fargo CIO, Chintan Mehta, revealed in an interview with VentureBeat that its AI assistant exceeded 245 million customer interactions in 2024, a more than 10-fold increase from 2023. Another interesting anecdote was that roughly 80% of 2024's usage was in Spanish.

Amazon CEO, Andy Jassy, has been vocal of the productivity gains of GenAI in recent shareholder letters: "As we roll out more Generative AI and agents, it should change the way our work is done. We will need fewer people doing some of the jobs that are being done today, and more people doing other types of jobs...but in the next few years, we expect that this will reduce our total corporate workforce as we get efficiency gains from using AI extensively across the company."

Rapidly growing new companies that leverage AI are also emerging and have the potential to disrupt existing industries, which has the potential to shape corporate strategies even among the mega-cap companies.

Within the software ecosystem, Cursor, developed by Anysphere, is growing rapidly in the AI coding assistant market, with reports that its annual recurring revenue (ARR) was \$500 million and that figure has been doubling every two months. This growth is notably faster than many established software companies. We have seen M&A among software companies accelerate, with ServiceNow's acquisition of agentic AI solution, Moveworks; Salesforce's acquisition of data integration tool, Informatica; and OpenAI's acquisition of AI coding tool, Windsurf.

### Exhibit 7: Benefits of Agentic AI for Selected Applications

Use Case	Key Benefit	Application
Customer Service	<ul style="list-style-type: none"><li>• Faster Response Times</li><li>• Handle More Peak Volume</li><li>• Personalized Support</li></ul>	<ul style="list-style-type: none"><li>• Salesforce Agentforce</li><li>• Glean</li><li>• Moveworks</li></ul>
Software Development	<ul style="list-style-type: none"><li>• Code Editing and Automation</li><li>• Code Reviews</li></ul>	<ul style="list-style-type: none"><li>• Github Copilot</li><li>• Anysphere (Cursor)</li><li>• Windsurf</li></ul>
Content Creation	<ul style="list-style-type: none"><li>• Accelerated Creative</li><li>• Personalized Content</li></ul>	<ul style="list-style-type: none"><li>• Midjourney</li><li>• Canva AI</li><li>• Jasper</li></ul>

Source: Artisan Partners. Key benefits based on the team's analysis and subject to material revision.

Control over compute, energy, data and talent is becoming the new competitive moat for the mega-cap tech companies, and we have seen recent evidence of the mega-caps evaluating their corporate strategies in light of the shifting AI landscape. To us, it signals a critical inflection point, where AI capabilities can determine future market leadership across the technology sector. Recent reports of large potential investments at both Apple and Meta are indicative of this.

Apple has reportedly been in discussions to acquire Perplexity AI for \$14 billion, which comes as a surprise given how nonacquisitive the company culture has been. The deal to acquire Beats for \$3 billion in 2014 remains the largest deal in the company's 49-year history. Apple has delayed Siri integration, and its Apple Intelligence features have not resonated well with users, compared to the AI adoption at ChatGPT. Integrating AI search capabilities into Safari and Siri to create a seamless, AI-enhanced user experience is likely a critical avenue that Apple would like to pursue, whether by organic or inorganic means (i.e., Perplexity).

An acquisition could help accelerate and augment its internal R&D effort, and a deal for Perplexity would be a defensive hedge for two reasons. Firstly, against a potential adverse ruling in the DOJ review of the Google search agreement (worth a reported \$25 billion per year). Secondly, to position the company to compete against OpenAI, which has been actively working on hardware AI systems and successfully recruited Jony Ive from Apple. Ive was integral to the design philosophy of many of the iconic Apple devices, including the iMac, iPod, iPhone and iPad.

Meanwhile, Meta has seemingly shifted its Metaverse aspirations to achieving AGI. It has entered a deal to acquire a 49% non-voting stake in ScaleAI for \$14.3 billion, with the ScaleAI co-founder, Alexandr Wang, joining Meta to lead a new "Superintelligence" lab. Meta is also in discussions to acquire venture capital fund NFDG, run by Nat Friedman and Daniel Gross, and reportedly is offering \$100 million sign-on bonuses to key talent at OpenAI and in the field of GenAI research.

Meta’s ambitions seem more tied to acquiring high-quality training data through ScaleAI’s data labelling expertise and establishing a leadership position in advanced AI research. This may also be defensive in nature, given the challenges Meta has encountered with delayed model launches (Behemoth) and talent leakage to the frontier LLMs. However, we certainly see Meta as well placed to leverage this AI expertise in its advertising algorithms to drive conversion and new ad formats, as well as to accelerate innovation in its virtual reality/augmented reality (VR/AR) initiatives.

These large-scale strategic investments are likely to further catalyse the fundamental shift toward agentic AI systems over the next few years. As investors, we think the silicon and power providers are the biggest near-term beneficiaries, given they are the critical enablers and the primary bottlenecks in supply.

### NVIDIA's Computing Architecture Is Helping Enable the Scale Up of Agentic AI

NVIDIA’s new accelerated compute architectures have enabled the scale up of agentic AI by reducing the inferencing costs per token. The B200 and B300 introduce higher density high-bandwidth memory (HBM) to remove bottlenecks for large context windows. For context, the GB300 (which is currently in testing and in volume shipments to hyperscalers in 2H 2025) is almost 700X more efficient at inferencing than the Ampere chips that were in production in 2020–2021.

NVIDIA’s management is emphasizing “lowest cost per inference token” as the design goal of new product launches. NVIDIA has demonstrated expertise in advancing the inferencing efficiency of its compute architecture, supported by an ambitious roadmap and accelerated product cadence that creates significant hurdles for competitors attempting to keep up. Following the debut of its Blackwell architecture in late 2024, the B300 (shipping in 2H 2025) is expected to deliver a 50% boost in inferencing efficiency with only a 10%–15% increase in system cost (ASP) versus the B200. This enhancement offers substantial performance-per-dollar advantages for hyperscalers building out large clusters over the next few years. The upcoming Vera-Rubin architecture in 2026 and Feynman in 2028 are both strategically designed to help optimize inferencing workloads, which we expect will further cement NVIDIA’s dominance.

On top of the rapid hardware progress, NVIDIA’s CUDA and TensorRT stack is also helping to boost inference throughput.

“Rapid advancements in NVIDIA software algorithms boosted Hopper inference throughput by an incredible 5X in one year and cut time to first token by 5X,” stated NVIDIA CFO Colette Kress on the Q3 FY2025 earnings call.

#### Exhibit 8: Estimated Inference Speed Growth Since A100 Launch

Model	Launch Year	Inference Speed Versus A100
A100 (Ampere)	2020	1X
H100 (Hopper)	2022	30X
B200 (Blackwell)	2024	450X
GB300 (Blackwell)	2025	675X
Vera-Rubin (Rubin)	2026	1,125X (estimated at 2.5X B200)

Source: NVIDIA/Artisan Partners. Estimates based on the team’s analysis of reported specs and subject to material revision.

### Key AI Investments

Historically, during tectonic computing shifts, the best investments throughout the cycle have typically been the technology leaders—whether it was IBM during the mainframe era, Intel and Microsoft during the PC era, Apple and Qualcomm during the smartphone era, or Amazon and Microsoft during the cloud computing era. We don’t see much reason to stray from the playbook for the era of accelerated compute. The highest margin businesses in semiconductors are NVIDIA and Broadcom, which are the technology leaders for AI infrastructure and the compute and the networking layers. NVIDIA leads in system computing, while Broadcom leads in ethernet networking and custom ASIC design (helping to make hyperscaler capital expenditure more efficient and conversely freeing up more budget for NVIDIA). Both are considered pivotal to the AI infrastructure build-out. We see a catalyst-rich path for both, NVIDIA as it accelerates its product roadmap and scales out its rack-level ecosystem, and for Broadcom as it begins to ramp its new Tomahawk 6 switch and as it hits a projected inflection point in new customer ASIC ramps from 2026–2028.

### NVIDIA's Supply Improvements Provide Confidence on Near-Term Estimate Revisions

Demand for NVIDIA AI compute architecture is still outstripping supply, but the pain points in the supply chain have been shifting. We find it most accurate to build up our near-term estimates to solve for maximum capacity in the bottlenecks. In 2024, the primary constraint was chip-on-wafer-on-substrate (CoWoS) capacity, so we rearchitected our revenue model to map CoWoS capacity upgrades at Taiwan Semiconductor (TSM).

Since late 2024, the bottleneck in supply has shifted to rack-scale deployments of the GB200 system, so we have started mapping our next 24-month estimates to supply capacity. NVIDIA recognizes revenue when it ships its GPUs to its original design manufacturer (ODM) partners (HonHai, Quanta and Wistron), but the ramp has been delayed over the past six months as the ODMs encountered multiple scaling issues of the

rack-scale Blackwell deployments due to the complex system integration and high-power density format. For context, the NVL72 form factor has power consumption of 120 kW per rack, far exceeding typical data center rack power (12 kW–20 kW) and even the high-density H100/H200 racks. Issues were also encountered surrounding the ramping of the liquid cooling systems and the backplane connectors.

For context around the complexity of the compute architecture, within a GB200 NVL 72, there are 108 chips (72 Blackwell GPUs, 36 Grace CPUs), 18 NVSwitch5 ASICs, 72 Infiniband NICs, 18 Bluefield-3 DPUs, 36 ConnectX-8 Super NICs, 162 OSFP cages, 36 PSUs, 5 BBUs, 300 supercapacitors, and ~5,000 copper cables, liquid cooling plates and connectors. The GB300 systems have even more discrete. The GB300 systems will have even more components and a larger memory attach footprint.

As a result, rack yields from December 2024 to May 2025 were sub-50%, resulting in smaller upward revenue revisions at NVIDIA. As we look forward, we expect bigger revenue revisions as supply ramps up. Our conversations with ODMs and suppliers suggest there have been rapid yield improvements in recent months, with rack-level yields likely to reach 80% in Q3. More recent channel checks from the sell side have suggested the ODM ecosystem has recently reached production rates close to 1,000 racks per week. This is a huge improvement from the 200–300 racks per month in February and March of this year.

As rack yields improve, we expect the next shortage in the supply chain to move to component supply, where we would expect to further rework and redefine our forecasting methodology to better align with the supply chain bottlenecks.

Exhibits 9–13 illustrate a breakdown of our compute model build for NVIDIA for 2025, which highlights the complexity and scale of the business.

Hopper revenues have started to decline from ~\$80 billion in 2024 to ~\$27 billion in 2025 as the Blackwell systems ramp up.

"When Blackwell starts shipping in volume, you couldn't give Hoppers away," NVIDIA CEO Jensen Huang stated at GTC 2025.

NVIDIA also has had to cease supply of its H20 chip, which was optimized to comply with prior export restrictions, which cost the company ~\$15 billion of lost sales for the year and the opportunity to compete in the ~\$50 billion China accelerated compute total addressable market (TAM).

**Exhibit 9: Hopper GPU and Average Selling Price (ASP) Estimates**

H200	SKU	Q1 FY26	Q2 FY26	Q3 FY26	Q4 FY26	FY 2026	
GPUs	H200	250,000	200,000	150,000	50,000	650,000	← Volumes taper off through Blackwell ramp
ASP (\$)	H200	24,473	23,984	23,264	22,334	23,879	
<b>Revenue (\$MM)</b>	<b>H200</b>	<b>6,118</b>	<b>4,797</b>	<b>3,490</b>	<b>1,117</b>	<b>15,522</b>	
H20	SKU	Q1 FY26	Q2 FY26	Q3 FY26	Q4 FY26	FY 2026	
<b>Pre-ban</b>							
GPUs	H20	350,000	700,000	400,000	250,000	1,700,000	← China restrictions caused a \$4.5 billion write-off, and an initial estimate ~\$15 billion of lost sales in FY26
ASP (\$)	H20	11,525	11,410	11,295	11,183	6,665	
<b>Post-ban getting revoked July 10</b>							
GPUs	H20	350,000	—	250,000	400,000	1,000,000	← Sales should re-ramp from inventory in Q3, with new wafer starts being ramped at TSM
ASP (\$)	H20	11,525	11,410	11,295	11,183	11,331	
<b>Revenue (\$MM)</b>	<b>H20</b>	<b>4,034</b>	<b>—</b>	<b>2,824</b>	<b>4,473</b>	<b>11,331</b>	
		<b>Q1 FY26</b>	<b>Q2 FY26</b>	<b>Q3 FY26</b>	<b>Q4 FY26</b>	<b>FY 2026</b>	
GPUs		600,000	200,000	150,000	50,000	1,000,000	
ASP (\$)		16,920	23,984	42,090	111,794	26,852	
<b>Revenue (\$MM)</b>		<b>10,152</b>	<b>4,797</b>	<b>6,314</b>	<b>5,590</b>	<b>26,852</b>	← We forecast Hopper revenue at ~\$27 billion for FY26, down 66% YoY (from ~\$80 billion)

Source: NVIDIA/Artisan Partners. As of 28 Jul 2025. Estimates based on the team's analysis and subject to material revision.

B200 chips have been shipping since late in Q4 FY25 for NVIDIA, with \$30 billion of revenues recognized in the first two quarters. As ODMs experienced yield issues with the rack-scale GB200 format, hyperscalers were aggressively ordering discrete chips and the prior generation HGX form factor.

**Exhibit 10: Blackwell B200 Rack Yields and Revenue Estimates**

B200	SKU	Q1 FY26	Q2 FY26	Q3 FY26	Q4 FY26	FY 2026
Rack Yields	B200	>85%	>85%	>85%	>85%	
Racks		8,000	8,000	10,000	10,000	
<b>Mix</b>						
Liquid		50%	55%	60%	60%	
Air		50%	45%	40%	40%	
<b>Racks</b>						
Liquid		4,000	4,400	6,000	6,000	
Air		4,000	3,600	4,000	4,000	
<b>GPUs per system</b>						
Liquid		64	64	64	64	
Air		32	32	32	32	
<b>Total GPUs</b>						
Liquid		256,000	281,600	384,000	384,000	
Air		128,000	115,200	128,000	128,000	
Total GPUs		384,000	396,800	512,000	512,000	1,804,800
ASP (\$)	B200	33,000	32,670	32,343	32,020	
<b>Revenue (\$MM)</b>	<b>B200</b>	<b>12,672</b>	<b>12,963</b>	<b>16,560</b>	<b>16,394</b>	<b>58,589</b>

← No yield issues on HGX form factor

← Liquid cooling allows higher GPU density/more GPUs per rack

← Supply chain capacity >2MM units

Source: NVIDIA/Artisan Partners. As of 28 Jul 2025. Estimates based on the team's analysis and subject to material revision.

We see a material ramp in rack-scale deployments in 2H of the year as rack yields improve. The GB200 is expected to be the biggest driver of revenue growth at NVIDIA this year.

**Exhibit 11: Blackwell GB200 Rack Yields and Revenue Estimates**

GB200 - NVL-72	SKU	Q1 FY26	Q2 FY26	Q3 FY26	Q4 FY26	FY 2026
Rack Yields	NVL-72	35%	60%	85%	>85%	
Racks		1,500	5,500	11,000	12,500	30,500
GPUs Per System		72	72	72	72	
CPUs Per System		36	36	36	36	
Total GPUs	NVL-72	108,000	396,000	792,000	900,000	2,196,000
Total CPUs	NVL-72	54,000	198,000	396,000	450,000	1,098,000
ASP Per System (\$)	NVL-72	3,100,000	3,069,000	3,038,310	3,007,927	
<b>Revenue (\$MM)</b>	<b>NVL-72</b>	<b>4,650</b>	<b>16,880</b>	<b>33,421</b>	<b>37,599</b>	<b>92,550</b>
<b>Compute Revenue (\$MM)</b>	<b>NVL-72</b>	<b>3,813</b>	<b>13,841</b>	<b>27,406</b>	<b>30,831</b>	<b>75,891</b>
<b>GB200 - NVL-36</b>						
Racks		300	2,000	4,000	4,000	10,300
GPUs Per System		36	36	36	36	
CPUs Per System		18	18	18	18	
Total GPUs	NVL-36	10,800	72,000	144,000	144,000	370,800
ASP Per System (\$)	NVL-36	1,800,000	1,782,000	1,764,180	1,746,538	
<b>Revenue (\$MM)</b>	<b>NVL-36</b>	<b>540</b>	<b>3,564</b>	<b>7,057</b>	<b>6,986</b>	<b>18,147</b>
<b>Compute Revenue (\$MM)</b>	<b>NVL-36</b>	<b>443</b>	<b>2,922</b>	<b>5,787</b>	<b>5,729</b>	<b>14,880</b>
<b>GB200 Total</b>						
Racks		1,800	7,500	15,000	16,500	40,800
GPUs Per System		66	62	62	63	
Total GPUs	GB200	118,800	468,000	936,000	1,044,000	2,566,800
ASP Per System (\$)	GB200	2,883,333	2,725,800	2,698,542	2,702,136	
<b>Revenue (\$MM)</b>	<b>GB200</b>	<b>5,190</b>	<b>20,444</b>	<b>40,478</b>	<b>44,585</b>	<b>110,697</b>
<b>Compute Revenue (\$MM)</b>	<b>GB200</b>	<b>4,256</b>	<b>16,764</b>	<b>33,192</b>	<b>36,560</b>	<b>90,771</b>

← Early expectations of 20,000–25,000 are biased higher as yields improve

← GB200 is sold in system with >1,000 components, including Grace CPUs and ConnectX-Super NICs

← Isolating Compute Revenue (GPU and CPU)

← GB200 is sold in system with >1,000 components, including Grace CPUs and ConnectX-Super NICs

← Isolating Compute Revenue (GPU and CPU)

← Matches with our channel check on total GB200 unit supply

Source: NVIDIA/Artisan Partners. As of 28 Jul 2025. Estimates based on the team's analysis and subject to material revision.

We expect GB300 to ramp up in 2H as well, with rack-scale production ramping up in late Q4 2025. This system has been optimized for inferencing, with more memory attach driving a 50% improvement in inferencing efficiency. We expect the GB300 to drive revenue growth through 1H of calendar year 2026, when Vera-Rubin arrives.

### Exhibit 12: Blackwell GB300 Rack Yields and Revenue Estimates

GB300	SKU	Q1 FY26	Q2 FY26	Q3 FY26	Q4 FY26	FY 2026	
GPUs	GB300			50,000	250,000	300,000	
ASP (\$)	GB300			35,000	34,300	34,417	
<b>Revenue (\$MM)</b>	<b>GB300</b>	<b>—</b>	<b>—</b>	<b>1,750</b>	<b>8,575</b>	<b>10,325</b>	
GB300 - NVL-72	SKU	Q1 FY26	Q2 FY26	Q3 FY26	Q4 FY26	FY 2026	
Rack Yields	NVL-72				>35%		
Racks					750		← Expect GB300 to ramp faster than GB200 in Q1
GPU per System					72		
Total GPUs	NVL-72				54,000		
ASP per System (\$)	NVL-72				3,350,000		← 10% – 15% ASP premium to GB200
<b>Revenue (\$MM)</b>	<b>NVL-72</b>				<b>2,513</b>	<b>2,513</b>	
<b>Compute Revenue (\$MM)</b>	<b>NVL-72</b>				<b>2,060</b>	<b>2,060</b>	← Isolating Compute Revenue (GPU and CPU)
GB300 - NVL-36	SKU	Q1 FY26	Q2 FY26	Q3 FY26	Q4 FY26	FY 2026	
Racks					1,200		
GPU per System					36		← Expect NVL-36 form factor to ramp faster given earlier availability and before yield rates improve on NVL-72
Total GPUs	NVL-36				43,200		
ASP per System (\$)	NVL-36				2,050,000		← 10% – 15% ASP premium
<b>Revenue (\$MM)</b>	<b>NVL-36</b>				<b>2,460</b>	<b>2,460</b>	
<b>Compute Revenue (\$MM)</b>	<b>NVL-36</b>				<b>2,017</b>	<b>2,017</b>	← Isolating Compute Revenue (GPU and CPU)
GB300 Total	SKU	Q1 FY26	Q2 FY26	Q3 FY26	Q4 FY26	FY 2026	
Racks					1,950	1,950	
Total GPUs	GB300			50,000	347,200	397,200	
ASP per System (\$)	GB300			35,000	39,019		
<b>Revenue (\$MM)</b>	<b>GB300</b>			<b>1,750</b>	<b>13,548</b>	<b>15,298</b>	
<b>Compute Revenue (\$MM)</b>	<b>GB300</b>			<b>1,750</b>	<b>12,652</b>	<b>14,402</b>	← GB300 is in high demand given 50% better inferencing performance, at a highly attractive price point versus GB200. Strong demand from hyperscalers

Source: NVIDIA/Artisan Partners. As of 28 Jul 2025. Estimates based on the team's analysis and subject to material revision.

In aggregate, we see scope for material positive revenue revisions as supply improves through the rest of the year, and we would expect fewer yield issues with the GB300 form factor, given the experience gained from the supply chain over the past 12 months.

### Exhibit 13: Blackwell and Total Compute GPU and ASP Estimates

Blackwell	Q1 FY26	Q2 FY26	Q3 FY26	Q4 FY26	FY 2026	
GPUs	502,800	864,800	1,498,000	1,903,200	4,768,800	
ASP (\$)	33,667	34,375	34,380	34,472	34,341	
<b>Revenue (\$MM)</b>	<b>16,928</b>	<b>29,727</b>	<b>51,502</b>	<b>65,607</b>	<b>163,763</b>	← We forecast Blackwell compute revenue at ~\$165 billion for FY26 (up >100% versus Hopper)
Total Compute	Q1 FY26	Q2 FY26	Q3 FY26	Q4 FY26	FY 2026	
GPUs	1,102,800	1,064,800	1,648,000	1,953,200	5,768,800	← Matches our channel checks on total GPU supply of ~6 million
ASP (\$)	24,556	32,423	35,082	36,451	33,042	
<b>Revenue (\$MM) - Supply</b>	<b>27,080</b>	<b>34,524</b>	<b>57,815</b>	<b>71,196</b>	<b>190,615</b>	← We forecast total compute supply will reach ~\$190 billion in FY26 versus Street estimates of \$160 billion
Street Compute Revenue (\$MM)	34,155	35,660	41,657	47,001	159,114	
<b>Delta vs. Street</b>					<b>20%</b>	

Source: NVIDIA/Artisan Partners. As of 28 Jul 2025. Estimates based on the team's analysis and subject to material revision.

### Broadcom

One of the things that we believe is underappreciated about both NVIDIA and Broadcom is their reliability with execution. Since its initial Tesla data center chip, NVIDIA has innovated consistently on a two-year cadence, which has accelerated recently. Broadcom too has consistently introduced its upgraded version of its high-performance ethernet family of switches, labeled Tomahawk. These chips are essential to driving down compute as they deliver higher bandwidth and lower latency for efficient networking in scaled environments.

Broadcom has two major catalysts for its AI business over the coming years: 1) the ramp of its high-performance Tomahawk 6 Ethernet switch; and 2) the ramp of multiple new ASIC customers, to augment its scaled out TPU footprint with Google.

On the switch side, the new Tomahawk 6 chip can now support both scale-up and scale-out environments, which will be essential for some of the large-scale training clusters coming online over the next few years, with Hock Tan, Broadcom CEO, noting there are multiple 1 million + XPU clusters in development. Due to the inherent cost savings the chip will provide, Broadcom has been able to charge more than its typical 2X ASP uplift per generation with the TPU 6 family.

**Exhibit 14: Broadcom Tomahawk Switch Comparison**

Version	Process	ASP	Capacity	2017– 2018	2019– 2021	2022– 2024	2025– 2027
3	12 NM	\$2,000	12.8 TB/s	█			
4	7 NM	\$4,000	25.6 TB/s		█		
5	5 NM	\$8,000	51.2 TB/s			█	
6	3 NM	\$20,000	102.4 TB/s				█

Source: Broadcom/Artisan Partners. As of 30 Jun 2025. Estimates based on the team’s analysis and subject to material revision.

On the ASIC side, demand for ASICs is increasing, as hyperscalers want to lower the cost of ownership versus NVIDIA’s high-margin GPUs. ASICs typically solve for lower cost in internal-facing workloads that have consistent features. Broadcom has the broadest portfolio of networking IP and can be fastest to market at scale for its customer base. We think this experience and consistency of manufacturing at scale positions Broadcom to win multiple new customers over the coming years.

All of the large hyperscalers currently have ASICs in production at varying levels of success (Amazon and Microsoft utilize Marvell, Alchip and GUC). For Broadcom specifically, the company has announced three hyperscale customers for its custom ASICs, including Google, Meta and ByteDance, and additionally the company has four prospects in advanced development for its own next-generation AI accelerators, though these are not yet classified as customers. Broadcom highlights a software asset management just for its three announced customers in the range of \$60 billion to \$90 billion by 2027. We think the other customers include OpenAI, Apple, xAI and Softbank, which we expect to ramp from 2026–2028 and support 40% AI revenue CAGR through 2029.

**Exhibit 15: Broadcom’s Announced and Prospective ASICs Customers**

Customer	Product	Process	2024	2025	2026	2027	2028
Google	V1-V5	> 5 NM	█				
Google	V6	3 NM		█			
Google	V7	3 NM			█		
Google	V8	2 NM				█	
Meta	MTIA V1	7 NM		█			
Meta	MTIA V2	5 NM		█			
Meta	MTIA V3	3 NM			█		
OpenAI	Titan 1	3 NM			█		
OpenAI	Titan 2	A16				█	
Bytedance		5 NM			█		
Bytedance		3 NM				█	
Softbank	Izanagi V1	3 NM			█		
Softbank	Izanagi V2	2 NM					█
xAI	V1	3 NM			█		
Apple	V1	3 NM					█

Source: Broadcom/Artisan Partners. As of 30 Jun 2025. Estimates based on the team’s analysis and subject to material revision.

---

For more information: Visit [www.artisanpartners.com](http://www.artisanpartners.com)

---

**Investment Risks:** Non-diversified portfolios may invest larger portions of assets in securities of a smaller number of issuers and performance of a single issuer may have a greater impact to the portfolio's returns. Use of derivatives may create investment leverage and increase the likelihood of volatility and risk of loss in excess of the amount invested. High portfolio turnover may adversely affect returns due to increased transaction costs and creation of additional tax consequences. Securities of small- and medium-sized companies tend to have a shorter history of operations, be more volatile and less liquid and may have underperformed securities of large companies during some periods. International investments involve special risks, including currency fluctuation, lower liquidity, different accounting methods and economic and political systems, and higher transaction costs. These risks typically are greater in emerging and less developed markets, including frontier markets. These risks, among others, are further described in Artisan Partners Form ADV, which is available upon request. This is a marketing communication.

Unless otherwise indicated, the Artisan Strategy characteristics relate to that of an investment composite or a representative account managed within a composite. It is intended to provide a general illustration of the investment strategy and considerations used by Artisan Partners in managing that strategy. Individual accounts may differ, at times significantly, from the reference data shown due to varying account restrictions, fees and expenses, and since-inception time periods, among others. Where applicable, this information is supplemental to, and not to be construed with, a current or prospective client's investment account information. References to individual security performance relate to a representative account in the composite. Individual holding periods may differ.

For the purpose of determining the portfolio's holdings, exposures are delta-adjusted at the issuer level and may include multiple securities of the same issuer. The holdings mentioned above comprised the following percentages of a representative account within the Antero Peak Strategy Composite's total net assets as of 30 Jun 2025: Apollo Global Management Inc 1.7%, Axon Enterprise Inc 2.8%, Broadcom Inc 3.5%, Equinix Inc 0.4%, GE Vernova Inc 0.9%, General Electric Co 6.9%, Mastercard Inc 1.0%, Microsoft Corp 10.6%, Netflix Inc 4.2%, NVIDIA Corp 9.3%, Rolls-Royce Holdings PLC 5.8%, ServiceNow Inc 2.0%, Siemens Energy AG 4.9%, Taiwan Semiconductor Manufacturing Co Ltd 3.3%, Vistra Corp 4.3%, VanEck Semiconductor ETF -4.5%. Securities named in the Commentary, but not listed here are not held in the portfolio as of the date of this report.

This commentary represents the views of the manager as of 30 Jun 2025 and do not necessarily represent those of Artisan Partners. The views and opinions expressed are based on current market conditions, which will fluctuate and those views are subject to change without notice. While the information contained herein is believed to be reliable, there is no guarantee to the accuracy or completeness of any statement in the discussion. Any forecasts contained herein are for illustrative purposes only and are not to be relied upon as advice or interpreted as a recommendation.

**Contribution to Return (CTR)** represents the contribution of an investment to the total portfolio return. CTR is calculated by multiplying a security's portfolio weight by its in-portfolio return daily for the period and has been derived from a transaction-based methodology, which includes the impact of transactions. **Attribution** quantifies the relationship between a portfolio's relative returns and the active management decisions differentiating the portfolio from the benchmark. **CTR and Attribution** are not exact, nor representative of actual investor returns due to several variables (e.g., fees, expenses, transactions) and therefore should be considered an approximation and examined in conjunction with the performance of the portfolio during the period. If applicable, securities of the same issuer are aggregated and may be delta-adjusted to determine the weight in the portfolio, and aggregation of corporate affiliates is subject to the determination of Artisan Partners. Further information on the methodology used is available upon request.

Theme classifications are at the sole discretion of the team. Themes and constituents are as of the date indicated and are subject to change. Certain holdings have been reclassified subsequent to initial investment, which has impacted theme performance during the period. Portfolio sector classifications are defined by the investment team based on GICS.

The S&P 500 and S&P UBS Leveraged Loan Indices are products of S&P Dow Jones Indices LLC ("S&P DJI") and/or its affiliates and has been licensed for use. Copyright © 2025 S&P Dow Jones Indices LLC, a division of S&P Global, Inc. All rights reserved. Redistribution or reproduction in whole or in part are prohibited without written permission of S&P Dow Jones Indices LLC. S&P<sup>®</sup> is a registered trademark of S&P Global and Dow Jones<sup>®</sup> is a registered trademark of Dow Jones Trademark Holdings LLC ("Dow Jones"). None of S&P DJI, Dow Jones, their affiliates or third party licensors makes any representation or warranty, express or implied, as to the ability of any index to accurately represent the asset class or market sector that it purports to represent and none shall have any liability for any errors, omissions, or interruptions of any index or the data included therein.

Portfolio statistics are obtained from various data sources and intended to provide a general view of the portfolio, or Index, at a point in time. Artisan Partners excludes outliers when calculating portfolio characteristics and may use data from a related security to calculate statistics if information is unavailable for a particular security. Percent of net assets represents the portfolio's exposures based on the economic value of investments (including delta-adjusting options exposures). Delta-adjusted options exposure is a measure of the market exposure created by the options and accounts for the sensitivity of options to changes in price of the underlying security. In comparison, measuring the exposure of an option at the market value of the option or notional value can understate or overstate, respectively, the economic exposure and risk. This estimate of portfolio exposure is only an approximation of the portfolio at a point in time.

Securities referenced may not be representative of all portfolio holdings. Securities of the same issuer are aggregated to determine a holding's portfolio weight. Portfolio statistics calculations exclude outlier data and certain securities which lack applicable attributes, such as private securities. Artisan Partners may substitute information from a related security if unavailable for a particular security. This material is as of the date indicated and is subject to change without notice. Totals may not sum due to rounding.

Net-of-fees composite returns were calculated using the highest model investment advisory fees applicable to portfolios within the composite. Fees may be higher for certain pooled vehicles and the composite may include accounts with performance-based fees. All performance results are net of commissions and transaction costs, and have been presented gross and net of investment advisory fees. Dividend income is recorded net of foreign withholding taxes on ex-dividend date or as soon after the ex-dividend date as the information becomes available to Artisan Partners. Interest income is recorded on the accrual basis. Performance results for the Index include reinvested dividends and are presented net of foreign withholding taxes but, unlike the portfolio's returns, do not reflect the payment of sales commissions or other expenses incurred in the purchase or sale of the securities included in the indices.

S&P 500<sup>®</sup> Index measures the performance of 500 US companies focused on the large-cap sector of the market. S&P 500<sup>®</sup> Equal Weighted Index gives each constituent the same weight in the index, versus the market weighted index where bigger companies hold a larger share of the index. The index(es) are unmanaged; include net reinvested dividends; do not reflect fees or expenses; and are not available for direct investment.

The Global Industry Classification Standard (GICS<sup>®</sup>) is the exclusive intellectual property of MSCI Inc. (MSCI) and Standard & Poor's Financial Services, LLC (S&P). Neither MSCI, S&P, their affiliates, nor any of their third party providers ("GICS Parties") makes any representations or warranties, express or implied, with respect to GICS or the results to be obtained by the use thereof, and expressly disclaim all warranties, including warranties of accuracy, completeness, merchantability and fitness for a particular purpose. The GICS Parties shall not have any liability for any direct, indirect, special, punitive, consequential or any other damages (including lost profits) even if notified of such damages.

**Discounted cash flow (DCF)** is a valuation method used to estimate the value of an investment based on its expected future cash flows. **Return on Invested Capital (ROIC)** is a measure of how well a company generates cash flow relative to capital invested in the business. **Alpha** is a quantitative measure of the volatility of the portfolio relative to a designated index. A positive alpha of 1.0 means the fund has outperformed its designated index by 1%. Correspondingly, a similar negative alpha would indicate an underperformance of 1%. **Earnings per Share (EPS)** is the portion of a company's profit allocated to each outstanding share of common stock. **Beta** is a measure of the volatility of a security or a portfolio in comparison to the market as a whole. **Earnings Before Interest & Tax (EBIT)** is an indicator of a company's profitability, calculated as revenue minus expenses, excluding tax and interest. **Sharpe Ratio** is a measure of risk-adjusted return—it is the average return earned in excess of the risk-free rate per unit of volatility or total risk. **Compound Annual Growth Rate (CAGR)** is the year-over-year average growth rate of an investment over a period of time. It is calculated by taking the nth root of the total percentage growth rate, where n is the number of years in the period being considered. **Hit Rate** is a measure of an investment manager's success rate within a portfolio, calculated as the percentage of profitable issuers in a portfolio relative to the total issuers in the portfolio. **Consensus Earnings Estimates** represent the average financial forecast for a company's future performance, typically its earnings per share (EPS) and revenue, as predicted by a group of financial analysts. **Return on Investment (ROI)** measures the amount of return on an investment relative to the investment's cost. **Slugging Ratio** measures the ratio of average gains on profitable investments versus the average losses on non-profitable investments. **Standard Deviation** defines how widely returns varied from an average over a given period of time. Higher deviation represents higher volatility. **Tracking error** is the difference between the price behavior of a position or a portfolio and the price behavior of a benchmark. **Upward Revisions** represent positive changes to previous earnings estimates or earnings forecast and indicate a more positive or increased outlook for the company, portfolio or index.

**Risk Factors: Idiosyncratic Risk (Asset-Specific Risk)** refers to the inherent risk due to specific qualities of an asset or group of assets. The **Momentum** factor refers to the tendency of "winning" stocks to continue performing well in the near term. **Momentum** is categorized as a "persistence" factor, i.e., it tends to benefit from continued trends in markets. The **Growth** factor refers to a characteristic or set of metrics used to identify companies or stocks that are expected to grow at an above-average rate compared to their industry or the overall market. The **Value** factor is based on the idea that stocks with lower prices relative to their fundamental metrics (such as earnings, book value, or cash flow) are likely to be undervalued. The **Industry** factor refers to the specific characteristics of an industry that can affect the performance of companies within that industry. The **Market Direction** factor refers to the overall trend or direction of the broad market.

This material is provided for informational purposes without regard to your particular investment needs and shall not be construed as investment or tax advice on which you may rely for your investment decisions. Investors should consult their financial and tax adviser before making investments in order to determine the appropriateness of any investment product discussed herein.

Artisan Partners Limited Partnership (APLP) is an investment adviser registered with the U.S. Securities and Exchange Commission (SEC). Artisan Partners UK LLP (APUK) is authorized and regulated by the Financial Conduct Authority and is a registered investment adviser with the SEC. APEL Financial Distribution Services Limited (AP Europe) is regulated by the Central Bank of Ireland. APLP, APUK and AP Europe are collectively, with their parent company and affiliates, referred to as Artisan Partners herein. Artisan Partners is not registered, authorised or eligible for an exemption from registration in all jurisdictions. Therefore, services described herein may not be available in certain jurisdictions. This material does not constitute an offer or solicitation where such actions are not authorised or lawful, and in some cases may only be provided at the initiative of the prospect. Further limitations on the availability of products or services described herein may be imposed.

This material is only intended for investors which meet qualifications as institutional investors as defined in the applicable jurisdiction where this material is received, which includes only *Professional Clients* or *Eligible Counterparties* as defined by the Markets in Financial Instruments Directive (MiFID) where this material is issued by APUK or AP Europe. This material is not for use by retail investors and may not be reproduced or distributed without Artisan Partners' permission.

In the United Kingdom, issued by Artisan Partners UK LLP, 25 St. James's St., Floor 10, London SW1A 1HA, registered in England and Wales (LLP No. OC351201). Registered office: Phoenix House, Floor 4, Station Hill, Reading Berkshire RG1 1NB. In Ireland, issued by Artisan Partners Europe, Fitzwilliam Hall, Fitzwilliam Pl, Ste. 202, Dublin 2, D02 T292. Registered office: 70 Sir John Rogerson's Quay, Dublin 2, D02 R296 (Company No. 637966).

**Australia:** This material is directed at wholesale clients only and is not intended for, or to be relied upon by, private individuals or retail investors. Artisan Partners Australia Pty Ltd is a representative of APLP (ARBN 153 777 292) and APUK (ARBN 603 522 649). APLP and APUK are respectively regulated under US and UK laws which differ from Australian laws and are exempt from the requirement to hold an Australian financial services license under the Australian Corporations Act 2001 in respect to financial services provided in Australia.

**Canada:** This material is distributed in Canada by APLP and/or Artisan Partners Distributors LLC, which conduct activities in Canada under exemptions from the dealer, portfolio manager and investment fund manager registration requirements of applicable Canadian securities laws. This material does not constitute an offer of services in circumstances where such exemptions are not available. APLP advisory services are available only to investors that qualify as "permitted clients" under applicable Canadian securities laws.

© 2025 Artisan Partners. All rights reserved.

**For Institutional Investors – Not for Onward Distribution**

