

NVIDIA: In Conversation

In a League of Its Own, With More Room to Run

Our core framework at Aurelion is to identify investment themes at inflection points. We believe conversations with management teams are one of the best ways to identify them early.

Our recent post-earnings discussion with NVIDIA gave us a closer look at where the company is heading and what will drive the next phase of the AI cycle.

Because NVIDIA touches so many parts of the AI ecosystem, the conversation also gave us useful insights into several areas of the industry:

- **Memory.** AI chips rely on a special type of high-speed memory that is in short supply. Prices have risen sharply, and we expect them to stay high into 2027-28, which is good news for the companies that make it.
- **Inference.** Most AI spending so far went into building models. The next wave comes from people and businesses using them, which will require more servers and power than expected.
- **Robotics.** Robots are starting to move from demos into real factories. The companies supplying the computing power and automation tools can benefit no matter which robot maker comes out ahead.

We believe inference will drive NVIDIA's next growth phase. Growing agentic workloads create a significant new source of compute demand.

Leo-Pierre Trudel

Lead Equity Analyst

ltrudel@aurelionresearch.com

Jérémie Boyer

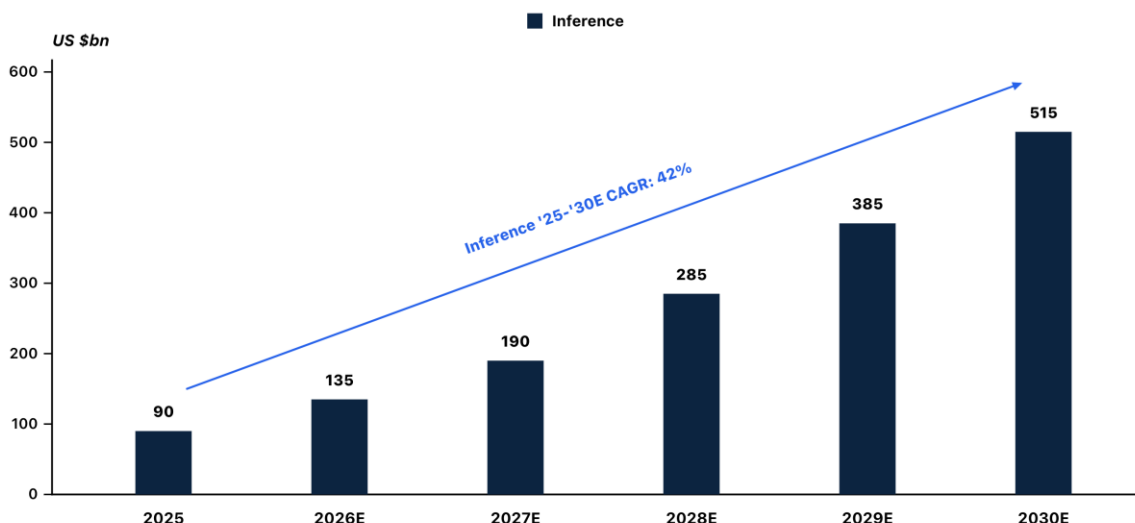
Lead Research Analyst

jboyer@aurelionresearch.com

One-Year Price Performance Chart



Inference: The Next Major Driver of AI Compute



Source: Bloomberg Intelligence, Aurelion Research.

NVIDIA has lagged because of concerns that AI spending may not generate sufficient returns, but we do not think the stock deserves to remain a laggard.

In this piece, we will explore why NVIDIA sits at the heart of the AI ecosystem, the exponential growth opportunities ahead, and what is driving hyperscaler spending. We will also look at the memory bottleneck and the next areas of growth across inference, agentic AI, robotics, and physical AI.

Methodology: *Regarding our conversation with NVIDIA's Investor Relations team, we were not granted permission to expressly quote or identify the individuals with whom we spoke. Our analysis below was informed by this discussion, but all views and analysis expressed are those of Aurelion Research and are based on publicly disclosed information.*

Nothing in this report should be construed as a direct quotation from, endorsement by, or representation of NVIDIA or its views. Aurelion Research does not receive compensation from the companies it covers. More information at aurelionresearch.com.

Table of Contents

1. Why We Are Bullish on NVIDIA.....4

 1.1 NVIDIA Is the Heart of the AI Ecosystem.....6

 1.2 Exponential Growth Opportunities.....9

 1.3 Hyperscaler Spending Is Durable.....12

2. AI Computing & Infrastructure.....15

 2.1 The Race for Compute.....15

 2.2 How Long Can the Memory Bottleneck Last?.....17

3. The Rise of Inference.....19

4. The Agentic AI Opportunity.....22

5. Physical AI: Where Are the Robots?.....24

6. Key Risks.....28

7. Our Final Thoughts.....30

AI Terms Glossary

Training: Teaching an AI model using large amounts of data.

Inference: Using a trained AI model to produce an answer or complete a task.
Every time someone asks a chatbot a question, that is inference.

Tokens: Units of text AI processes or generates, used to measure AI usage.

Agentic AI: AI that completes tasks and uses tools (agents).

Physical AI: AI operating in the real world through robots, vehicles, and machines.

Accelerated computing: Using GPUs alongside or instead of traditional CPUs to run AI faster.

AI factories: Data centers built to train and run AI models.

1) Why We Are Bullish on NVIDIA

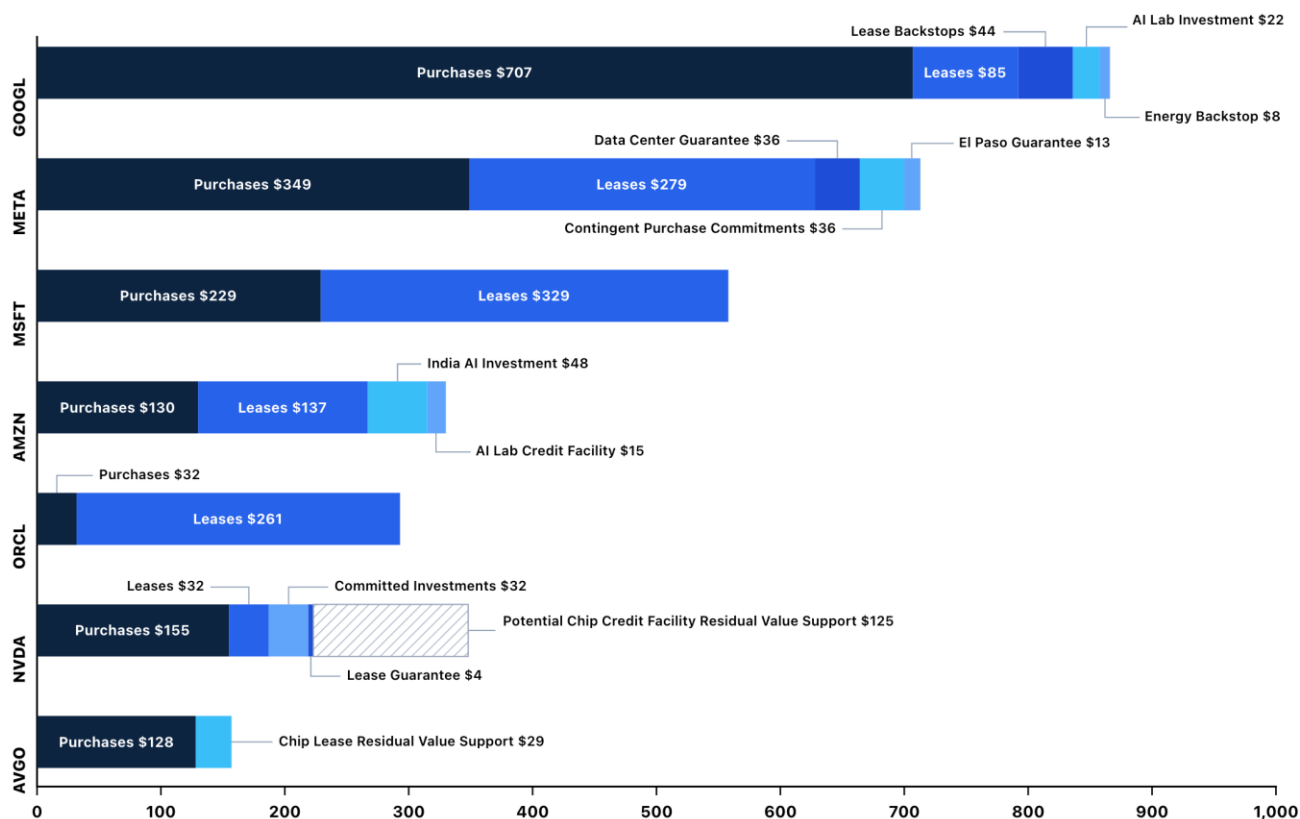
Our discussion with the company reinforced our view that the next phase of the AI chip race will be about making intelligence cheaper to use.

As inference becomes a larger part of AI workloads, the companies that can deliver more intelligence per dollar of compute will have the biggest opportunity. We believe NVIDIA is best positioned to lead this shift.

The AI infrastructure opportunity is expanding in two important ways: there are more applications for AI, and there are more customers spending on it.

The first phase of the AI cycle was largely driven by a handful of hyperscalers building massive amounts of infrastructure, with AI-related infrastructure commitments expected to exceed \$3.1 trillion. That is still happening and continues to grow, but the next phase will be much broader, with spending coming from enterprises, AI-native companies, and cloud providers.

Hyperscalers' Off-Balance-Sheet Commitments Exceed \$3.1 Trillion

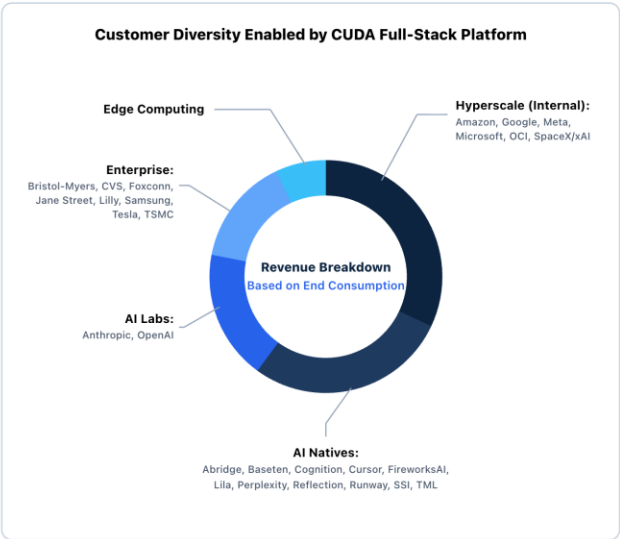
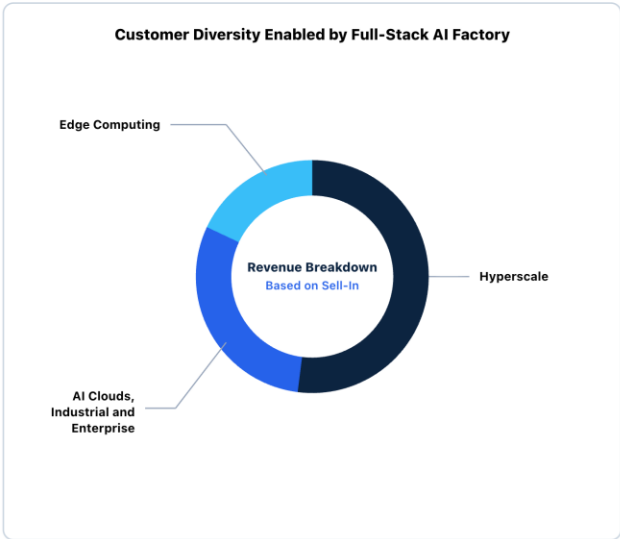


Source: Company Data & Filings, Bloomberg, LSEG, Aurelion Research.

In the past, NVIDIA was obviously an attractive company, but we felt the story was still heavily driven by its chips and a relatively concentrated group of hyperscaler customers. **Today, we see a much wider opportunity.**

Demand is still being driven by hyperscalers, but our recent discussion with NVIDIA’s Investor Relations team reinforced our view that the customer base is expanding well beyond the largest cloud providers. **We see this broader demand as an important source of growth for NVIDIA in the years ahead.**

NVIDIA's Demand is Increasingly Diverse

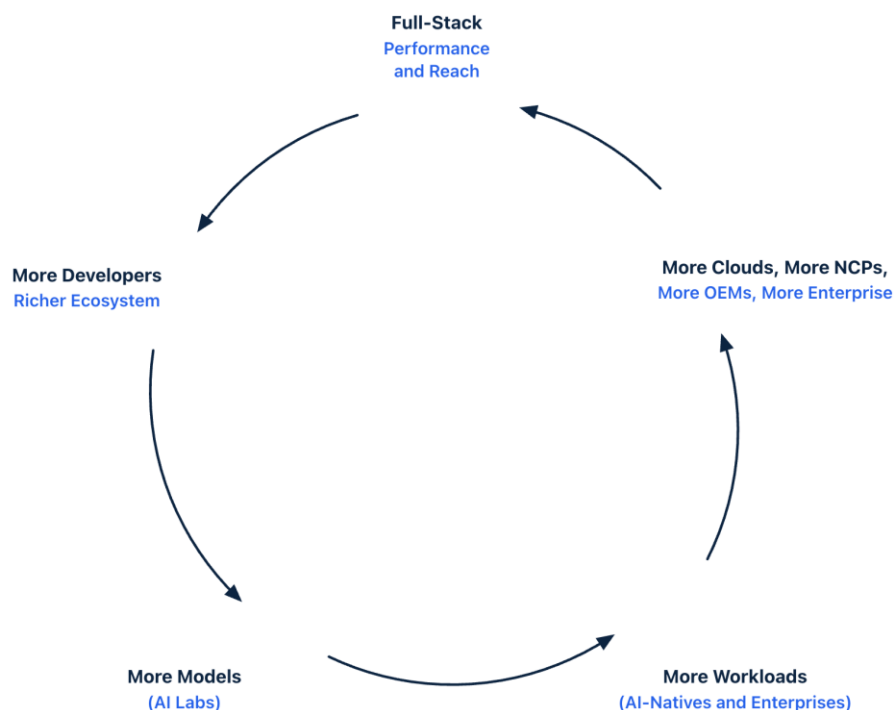


Source: Nvidia Corporation Presentation, Aurelion Research.

1.1 NVIDIA Is the Heart of the AI Ecosystem

The next AI phase will be less concentrated, with spending spreading across a much wider range of companies and applications. This should create new sources of compute demand beyond the infrastructure buildout we have seen so far.

NVIDIA Creates Value for Every Part of the AI Ecosystem

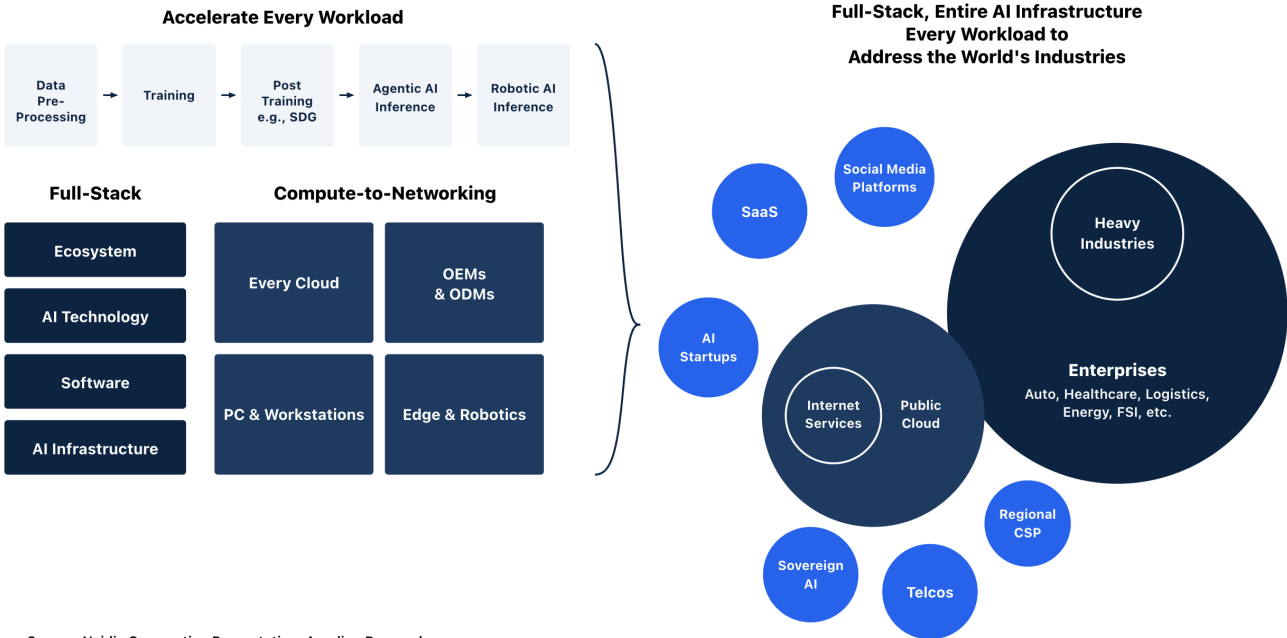


Source: Nvidia Corporation Presentation, Aurelion Research.

The key with NVIDIA is that the market often seems to be questioning whether the change happening right now is sustainable. If it isn't, the concern is that the large investments being made in AI infrastructure may not generate the returns expected, or that some of that capacity could ultimately go unused.

What stands out from our conversation with NVIDIA is how broad the platform has become. Its exposure across so many parts of the value chain means that even smaller opportunities can add meaningfully to growth.

NVIDIA AI Platform & Ecosystem Reaches Every Market



Source: Nvidia Corporation Presentation, Aurelion Research.

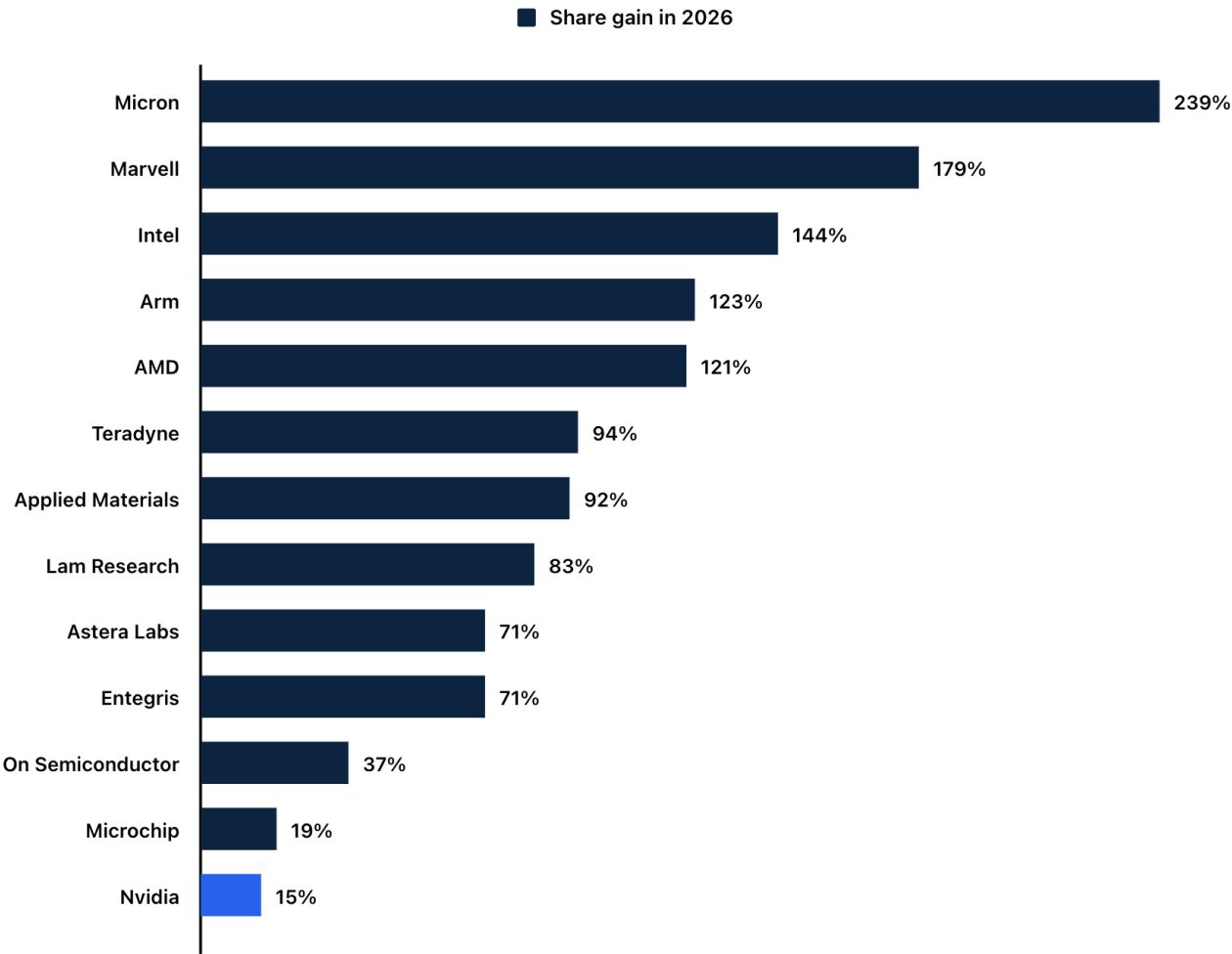
The other side of that is execution risk. It is difficult to be the best at everything, and NVIDIA is expanding into more areas where it will face strong competition. If it cannot maintain a leading position across the different workloads and markets it is targeting, that broad exposure could become a problem.

We think this is one of the reasons NVIDIA has lagged in 2026.

It has been one of the biggest under-performers among U.S. chip stocks, while Micron, Marvell, Intel, and AMD have all outperformed. **The performance gap is substantial, but the market is already pricing in many of these concerns.**

Nvidia Has Lost Some Luster With Investors This Year

Among US chip stocks, the company has been a laggard

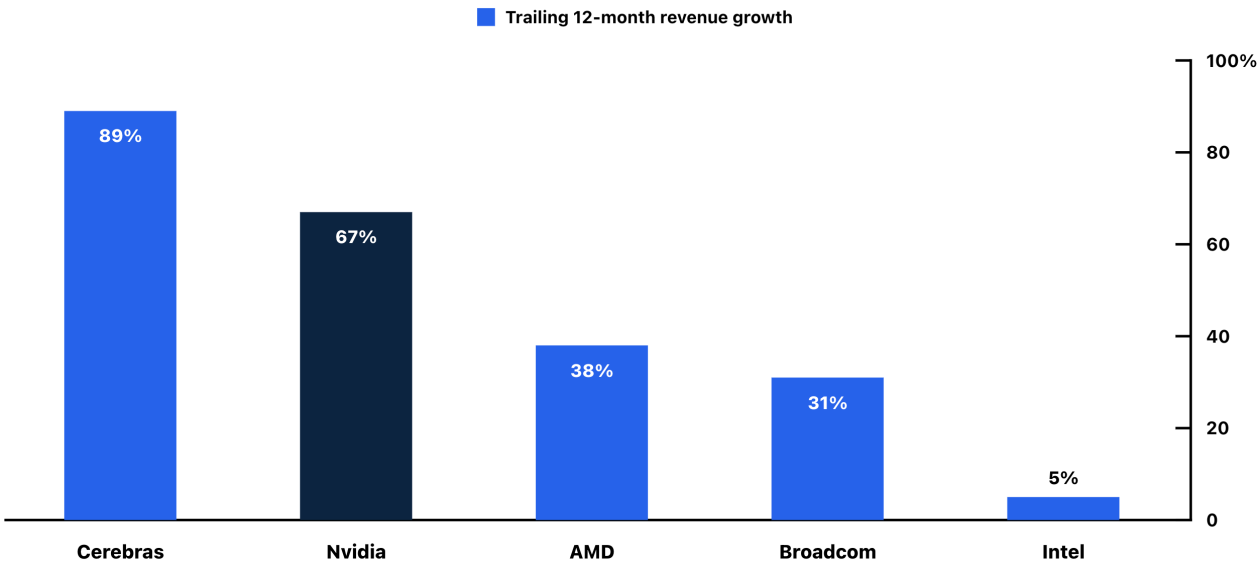


Source: Bloomberg, Aurelion Research.

As we said earlier, we think those concerns are overdone. The disconnect between NVIDIA’s fundamentals and its stock performance is where we see the opportunity. The company continues to grow at an exceptional pace, generate significant free cash flow, and expand into new areas of AI computing.

Nvidia’s Growth Is Still the Envy of Most Major Rivals

Only the much smaller Cerebras is seeing a bigger surge in sales



Source: Company data, Bloomberg, Aurelion Research.

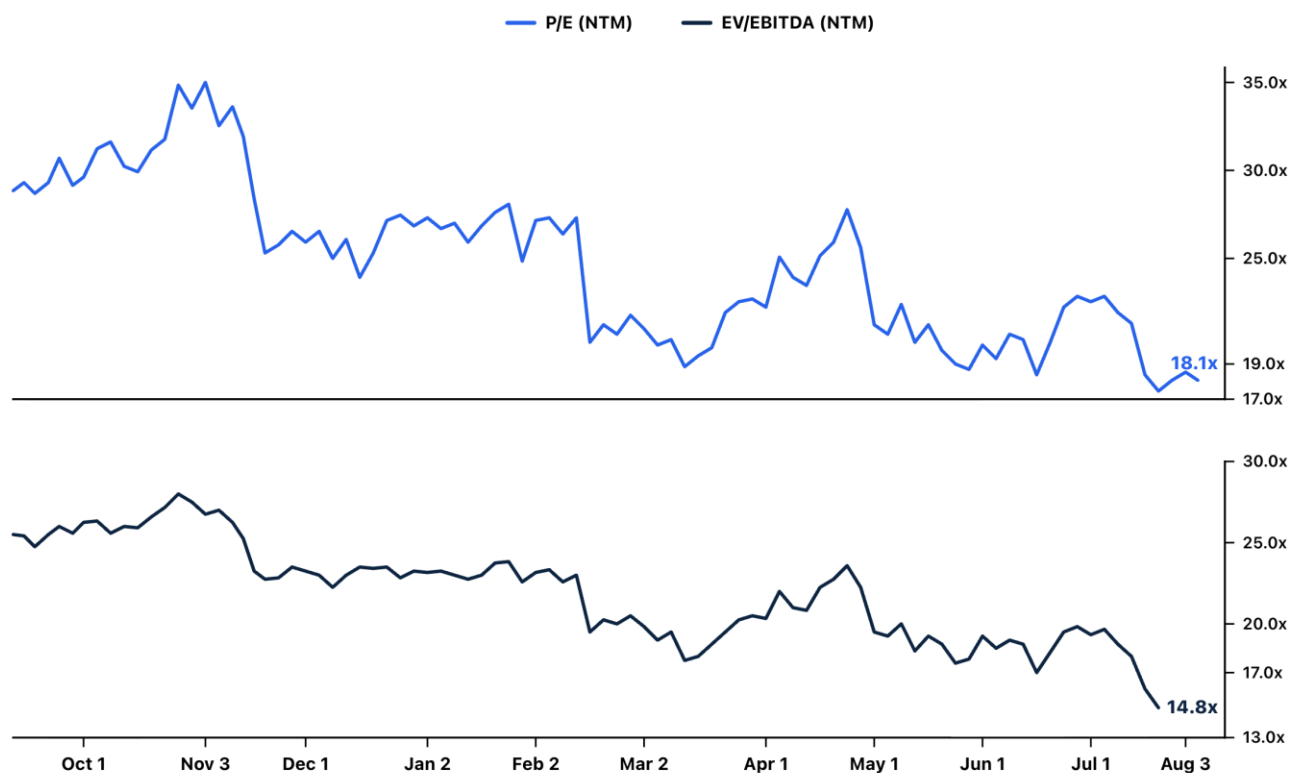
The key for us is that NVIDIA has several years of growth ahead, while the market is already debating when that growth will slow. If AI infrastructure spending continues to rise, we see a clear path to much higher earnings and cash flow from here. That is what makes the current setup so attractive to us.

1.2 Exponential Growth Opportunities

NVIDIA needs to sustain a very high rate of growth, and we think it can. Valuation is one of the biggest challenges in AI today, with some companies trading at 50x or even 100x earnings despite generating little or no free cash flow. We find those valuations difficult to justify. **NVIDIA looks very different.**

At 18.1x NTM P/E and 14.8x NTM EV/EBITDA, the valuation looks reasonable given its growth, profitability, and cash generation. Management will rarely say outright that its shares are cheap, but NVIDIA's aggressive buyback activity suggests that it sees its stock as an attractive use of capital at current levels.

NVIDIA Valuation Multiples



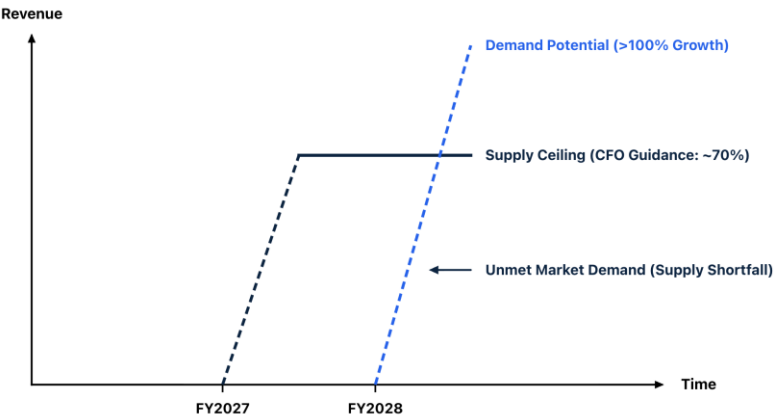
Source: FactSet, Aurelion Research.

But where will the growth come from? That is the key debate around NVIDIA today. Some investors believe it can continue growing at the pace needed to support a higher valuation, while others think growth will slow much sooner.

In a recent earnings call, NVIDIA's CFO said the company expects revenue to grow by 70% in FY2028, adding that this is a supply-constrained outlook.

She also highlighted the pressure from memory costs, describing the current environment as "extreme pricing conditions in memory."

NVIDIA Demand Potential vs. Supply Constraints (FY27–FY28)



Source: Nvidia Data, Aurelion Research.

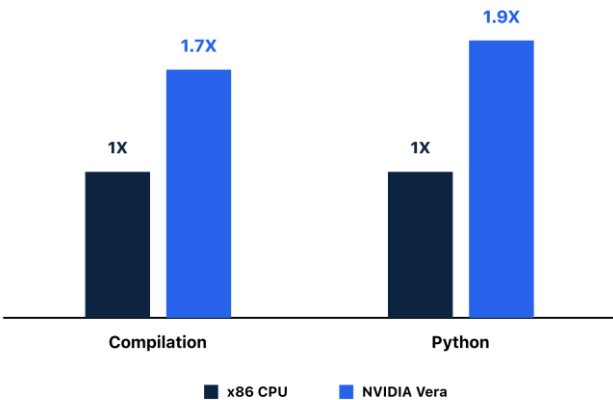
Both comments point to the same thing: demand for NVIDIA’s products remains ahead of available supply. NVIDIA expects these supply constraints to persist through at least the end of FY2028, which should continue to support strong demand for its computing products as more capacity comes online.

What makes NVIDIA particularly interesting to us is the number of different areas that can drive growth from here. Training, inference, enterprise AI, agentic AI, physical AI, and accelerated computing are all at different stages of adoption, giving NVIDIA multiple ways to grow as the AI market develops.

Some of these areas are still in their early stages, while others are already scaling quickly. CPUs are a good example. The CPU market is often seen as mature or slowing, but NVIDIA’s Vera CPU could open up another meaningful source of growth as more data centers adopt accelerated computing.

Vera CPU: A New Pillar of Growth

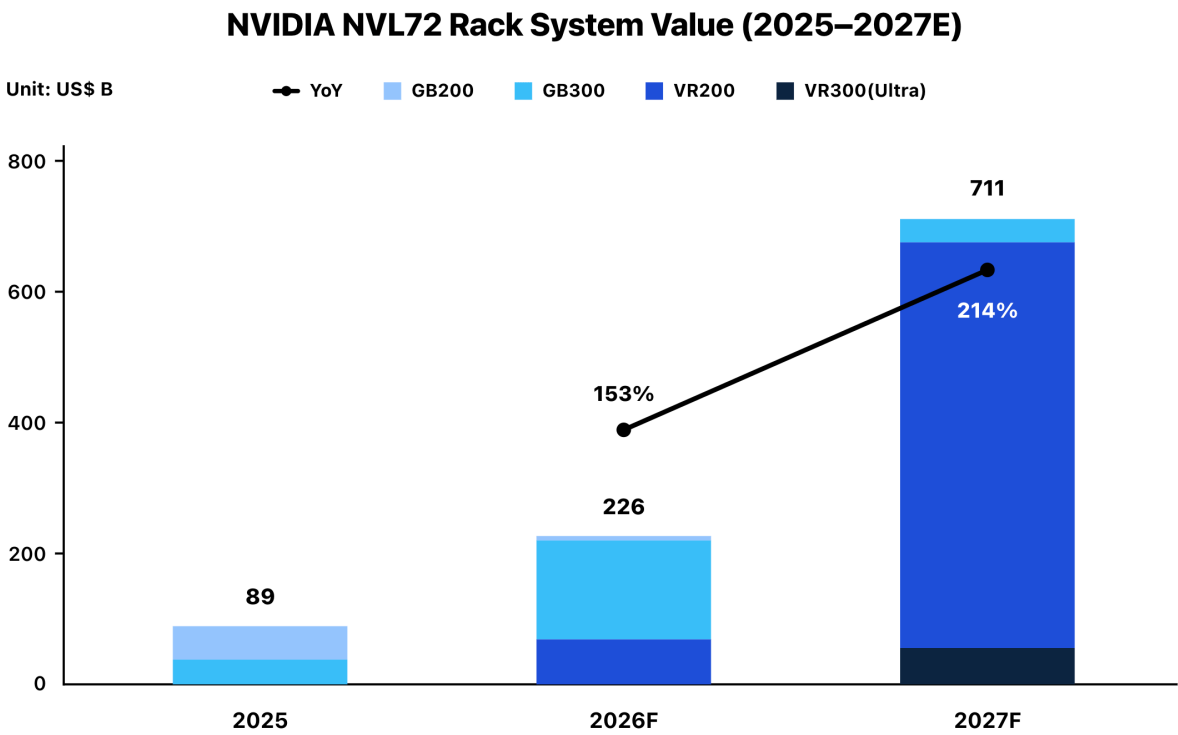
1.8X Agentic Sandbox Performance vs. x86



Source: Nvidia Corporation Presentation, Aurelion Research.

Another opportunity comes from NVIDIA’s complete computing systems.

TrendForce expects the output value of NVIDIA NVL72-class rack systems to grow from \$226B in 2026 to \$711B in 2027.



Source: TrendForce (Aug. 2026), Aurelion Research.

As AI systems become more complex, customers are buying complete computing systems instead of individual GPUs. This will give NVIDIA more opportunities to grow revenue within its existing computing business and capture more value from each deployment.

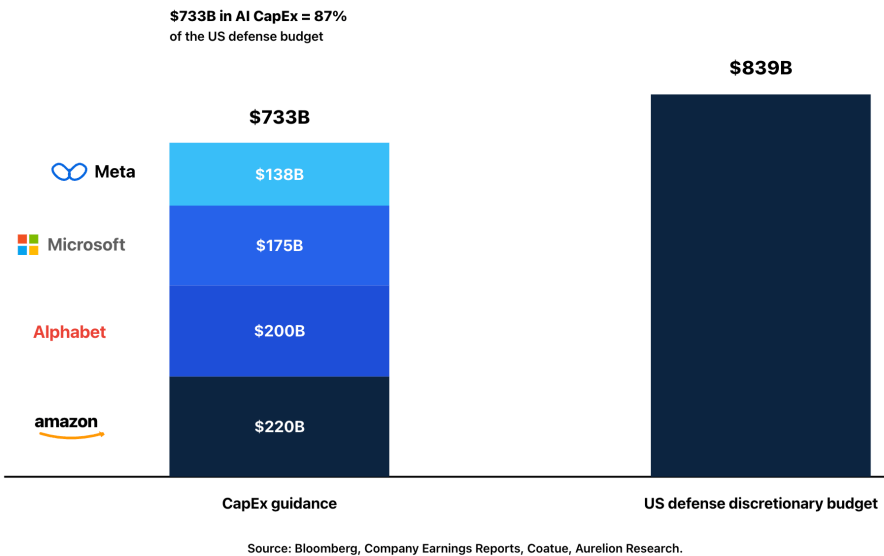
1.3 Hyperscaler Spending Is Durable

One of the biggest concerns around NVIDIA is what happens if hyperscalers slow down their spending. If capex growth fades, it would clearly be a material headwind for NVIDIA given how much of its business is tied to AI infrastructure.

Could that happen? Yes. Do we expect it to? Not really. The hyperscalers have been clear that continued investment in AI infrastructure is necessary to support their growth. Amazon, Microsoft, Google, Meta, and Oracle are all committing significant amounts of capital, and their earnings calls and industry conferences continue to reinforce that message. If these companies were to suddenly cut capex, it would be negative for both their own growth and NVIDIA.

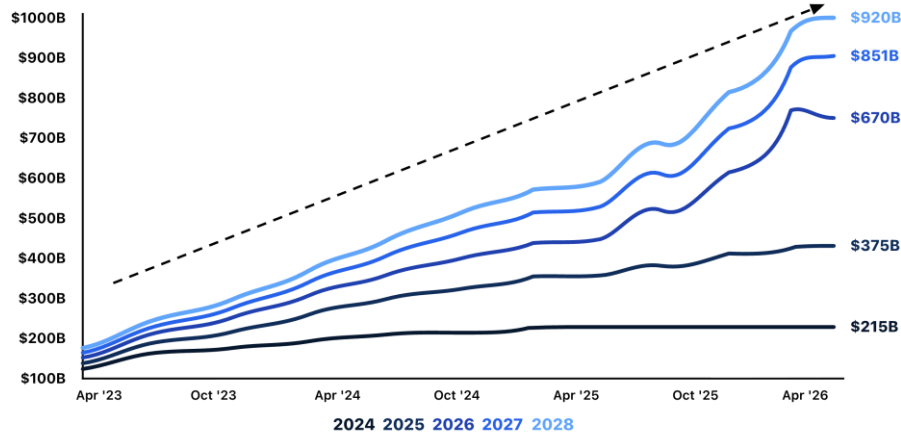
For context, projected hyperscaler capex in 2026 is equivalent to around 87% of the annual U.S. defense discretionary budget.

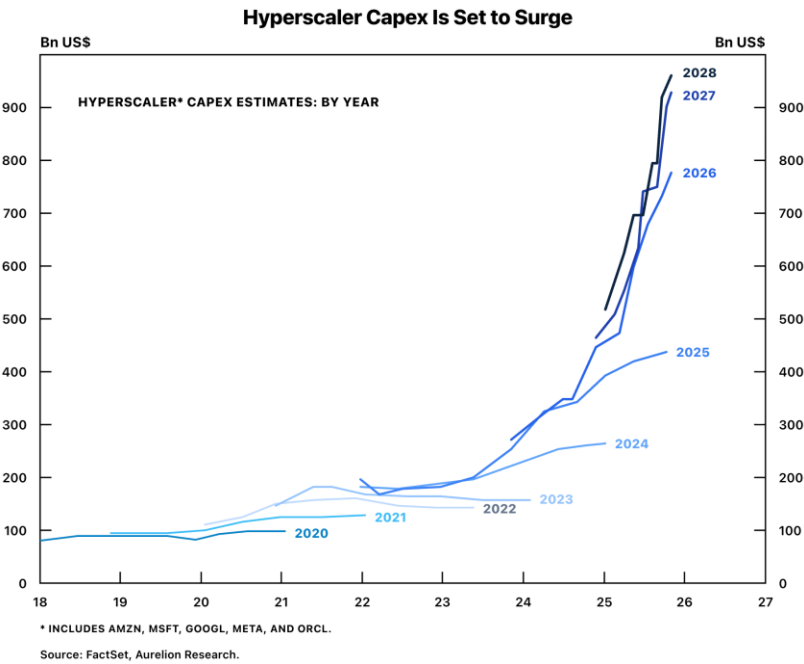
Hyperscaler 2026E Capex vs. US FY2026 Defense Budget (\$B)



The spending estimates also point in the same direction, with plenty of room for hyperscaler capex to continue growing from current levels. And even if we are partly wrong about the pace of spending, NVIDIA still has significant room to grow. It’s generating massive amounts of cash, expanding its business, and benefiting from a market where computing demand continues to increase.

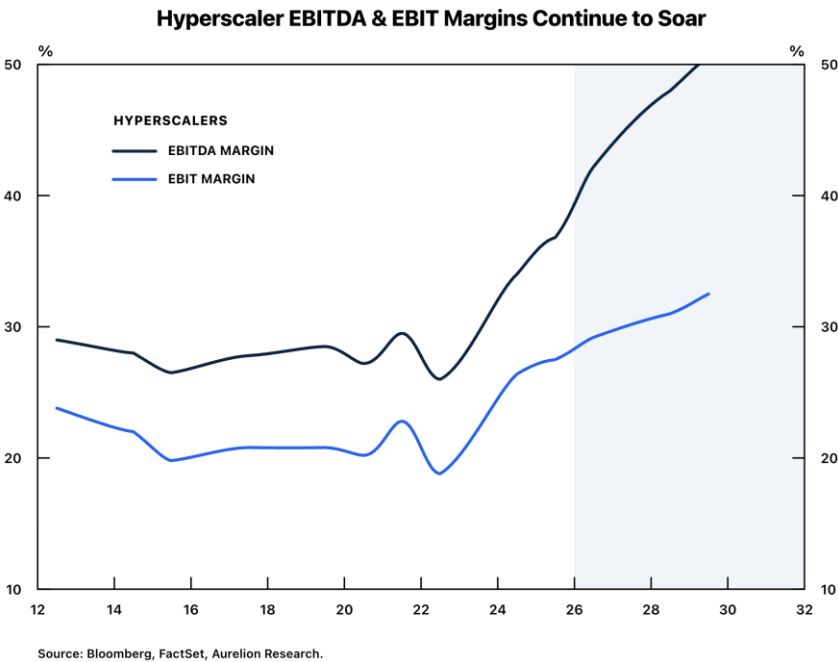
U.S. Big 4 Hyperscalers: Consensus Capital Spending





One thing that also gives us confidence that hyperscalers can continue spending heavily is that their EBITDA and EBIT margins are still rising.

They are generating strong profits and cash flow, giving them plenty of capacity to keep investing in AI infrastructure while continuing to grow their businesses.



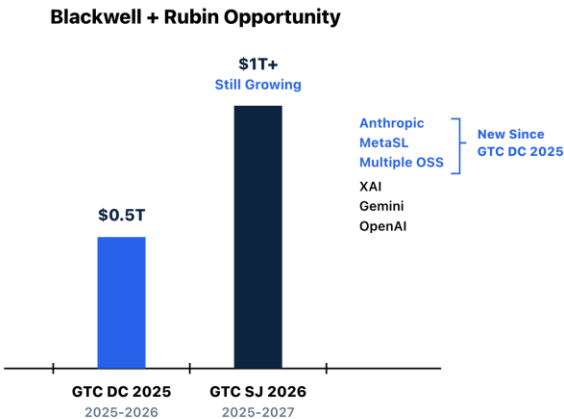
2. AI Computing & Infrastructure

NVIDIA is far more than a GPU company. The company has shown this clearly through its recent events and product launches, but we think the market still underestimates how broad its business has become.

NVIDIA is No Longer Just a GPU Story

	FY 2024	FY 2025	FY 2026	2027 (Q1)
Datacenter	\$48	\$115	\$194	\$75
Hyperscale	25	54	106	38
AI Clouds, Industrial Enterprise	23	61	88	37
Edge Computing	13	15	22	6
TOTAL	\$61	\$130	\$216	\$82

(\$ in Billions)



Source: Nvidia Corporation Presentation, Aurelion Research.

The company is present across edge computing, data centers, hyperscalers, AI clouds, industrials, and enterprises, while its platform spans GPUs, CPUs, networking, memory, software, and complete computing systems. **Very few companies are positioned across as many parts of the AI ecosystem.**

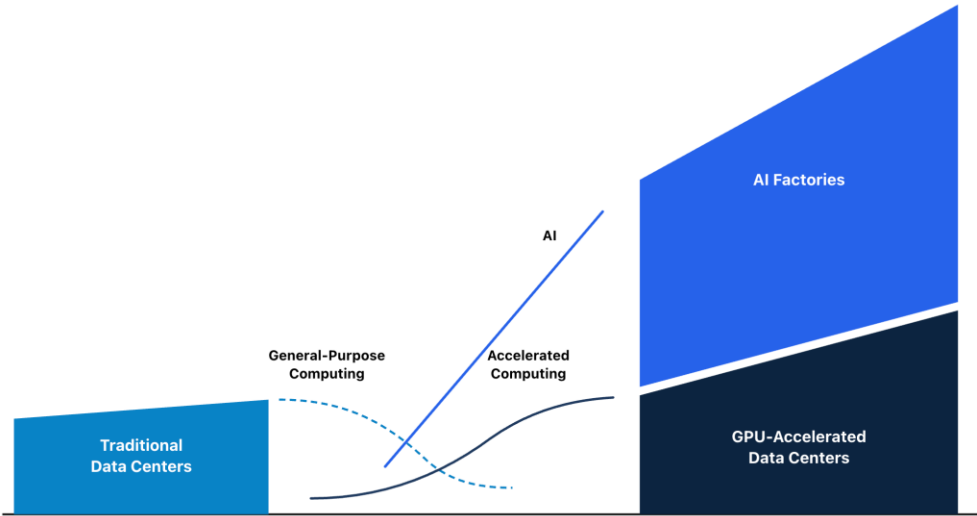
2.1 The Race for Compute

The \$1 trillion installed base of general-purpose CPU data center infrastructure is being upgraded to accelerated computing. NVIDIA sees this as a much broader shift in computing, with GPUs, networking, software, and complete systems all becoming part of the new AI infrastructure stack.

This is also expanding the size of the data center opportunity.

NVIDIA estimates that AI factories could push the data center market to \$2 trillion and beyond in the coming years, as companies across industries begin building infrastructure to train and run AI models.

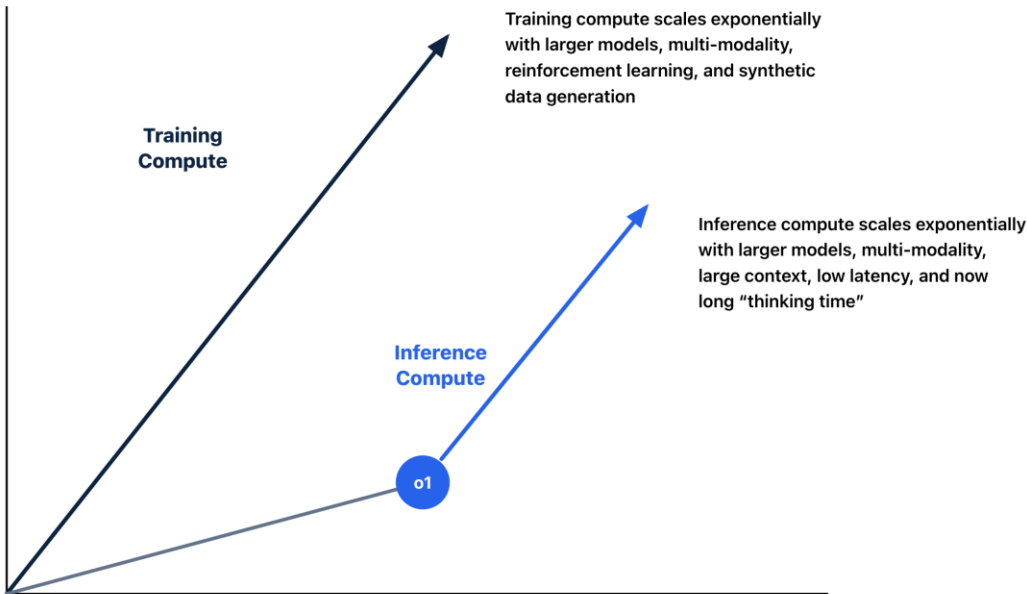
Accelerated Computing & Generative AI Create Trillion-dollar Opportunities



Source: Nvidia Corporation Presentation, Aurelion Research.

AI Clouds, Industrial, and Enterprise customers now account for roughly 50% of NVIDIA’s Data Center revenue, and the company expects this group to outgrow hyperscalers over the long term. It also has visibility into more than \$1 trillion of cumulative Blackwell & Rubin revenue from 2025 through 2027, before factoring in additional opportunities across CPUs, networking, and other products.

AI Scaling Laws Drive Exponential Demand for Compute

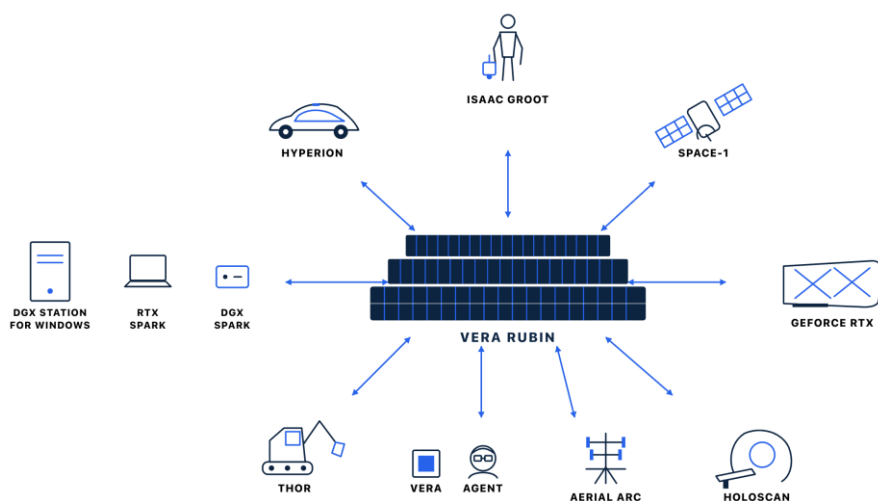


Source: Nvidia Corporation Presentation, Aurelion Research.

The customer base is expanding as well, with companies such as Anthropic, Meta, and OpenAI, as well as models such as Gemini and multiple open-source models, driving additional demand. This is why NVIDIA should be seen as more than a semiconductor company. That said, the broader platform also creates risks.

Bears could argue that NVIDIA is expanding into too many areas, increasing execution risk and giving competitors more opportunities to challenge its position.

NVIDIA Powers Global AI Economy



Source: Nvidia Corporate Presentation, Aurelion Research.

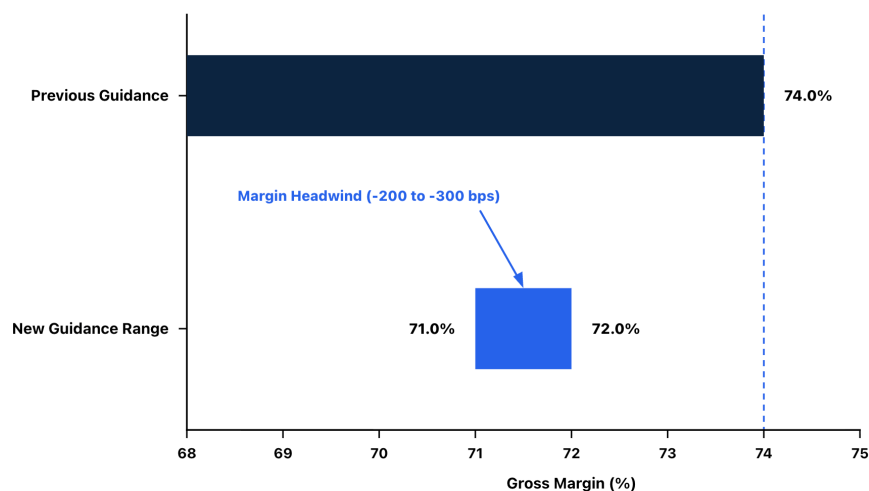
The company's advantage comes from bringing together the hardware, software, and developer ecosystem required to build AI systems. As customers become more deeply integrated into its ecosystem, moving to an alternative becomes more difficult, helping it defend its position as AI adoption expands.

2.2 How Long Can the Memory Bottleneck Last?

One of the most important comments from NVIDIA's earnings call was around memory. CFO Colette Kress said **the company is experiencing "extreme pricing conditions in memory," with price increases exceeding previous expectations and likely to move even higher into next year.**

This is putting pressure on NVIDIA's margins, with Q4 gross margins now expected at around 71% to 72%, compared with the previous expectation of roughly 74%. We view this as a headwind that could become more significant over time, but for now, it remains manageable given NVIDIA's overall profitability.

NVIDIA Q4 Gross Margin Guidance Compression



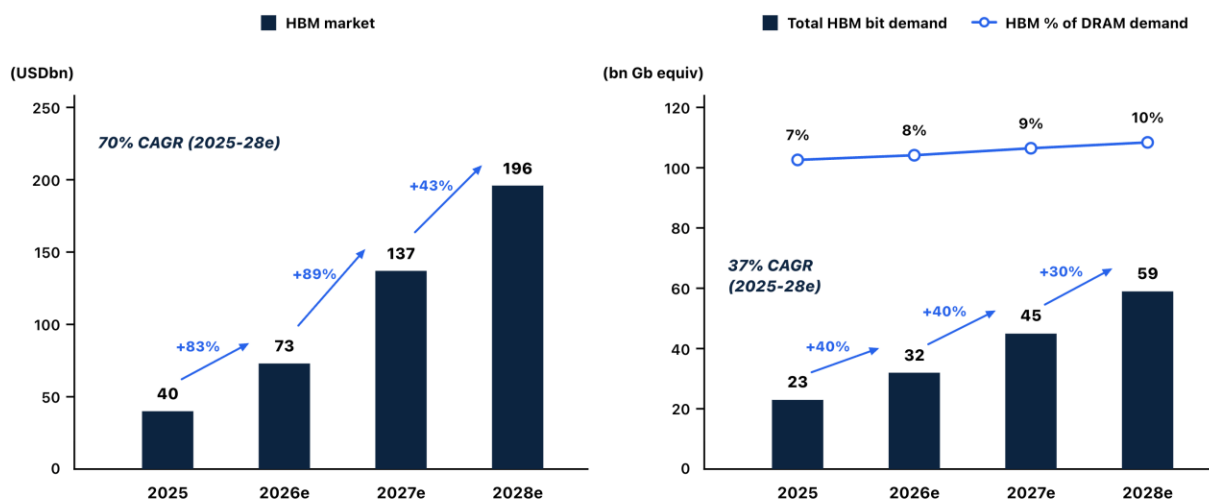
Source: NVIDIA Data, Aurelion Research.

But how long will these memory pricing conditions last?

NVIDIA believes the current memory shortage is being driven largely by the AI buildout, as AI systems require large amounts of high-bandwidth memory (HBM). In other words, higher memory costs are partly a consequence of the same infrastructure spending that is driving demand for NVIDIA's products.

The HBM market is expected to keep expanding rapidly through 2028. Total HBM bit demand is forecast to grow at a 37% CAGR, while HBM's share of total DRAM demand is expected to rise from around 7% in 2025 to 10% by 2028.

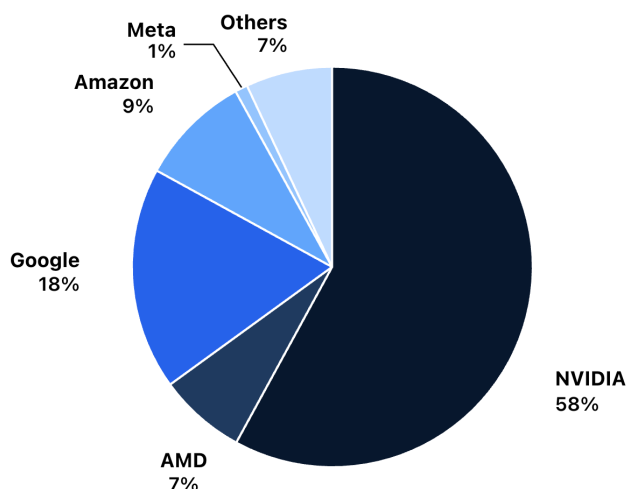
HBM Demand: 2026–2027 Outlook



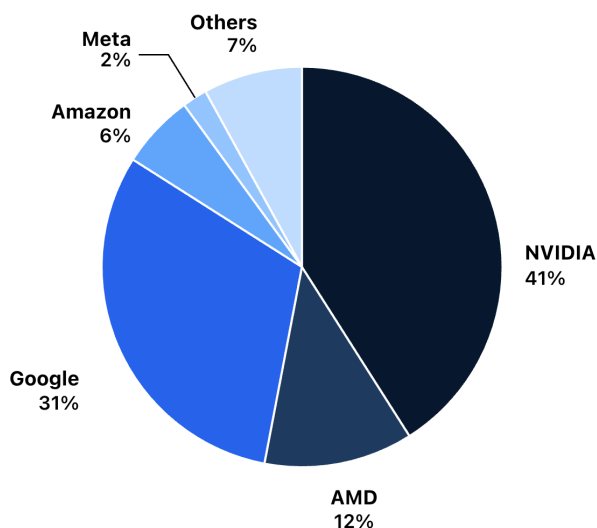
Source: WSTS, TrendForce, Company Data, HSBC Estimates, Aurelion Research.

NVIDIA's share of HBM demand is expected to fall from roughly 58% this year to 41% next year as competing AI accelerators gain ground. While the decline in share is worth monitoring, we do not view it as a material concern on its own.

HBM demand by customer (2026E)



HBM demand by customer (2027E)



Source: Bloomberg, Company Data, Aurelion Research.

The key is the size of the overall market: NVIDIA can lose share while still increasing its HBM consumption as total demand continues to grow.

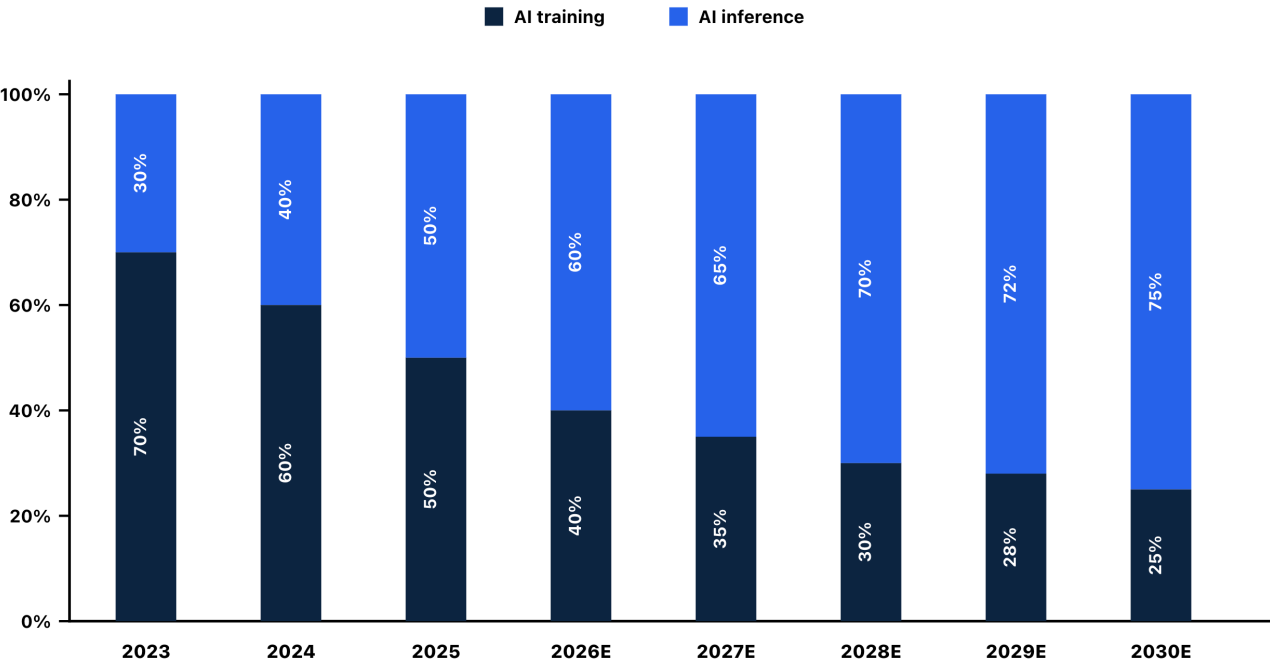
3. The Rise of Inference

An interesting takeaway from our conversation is that, for now, inference gets much less attention than training, but this is about to change. Training gets the headlines because building capable models requires enormous amounts of compute. But once a model is trained, it has to run every time someone uses it.

As AI moves from occasional use to everyday software, business workflows, and increasingly autonomous agents, a shift we are already seeing, inference demand should grow rapidly.

We think inference will become one of the biggest drivers of NVIDIA's next phase of growth. NVIDIA said on its latest earnings call that inference is now larger than training, after the two were roughly balanced only 18 months ago.

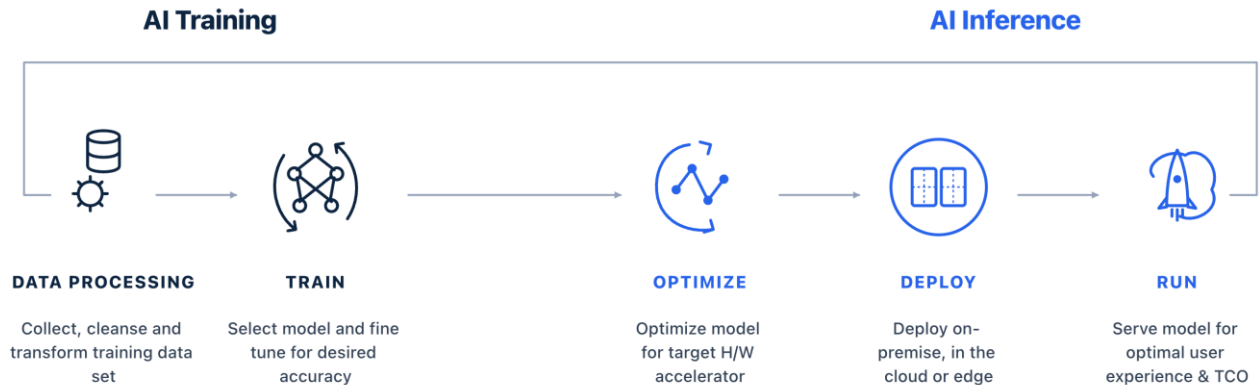
Datacenter TAM: AI Training vs. Inference



Source: IDC, Gartner, CLSA, Aurelion Research.

It also said its revenue could grow by more than 100% YoY on an unconstrained basis, but ultimately guided to nearly 70% growth because supply is limiting how much it can deliver. To us, that is an important distinction: the constraint is increasingly how much NVIDIA can supply, rather than a lack of demand.

Training Workloads vs. Inference Workloads



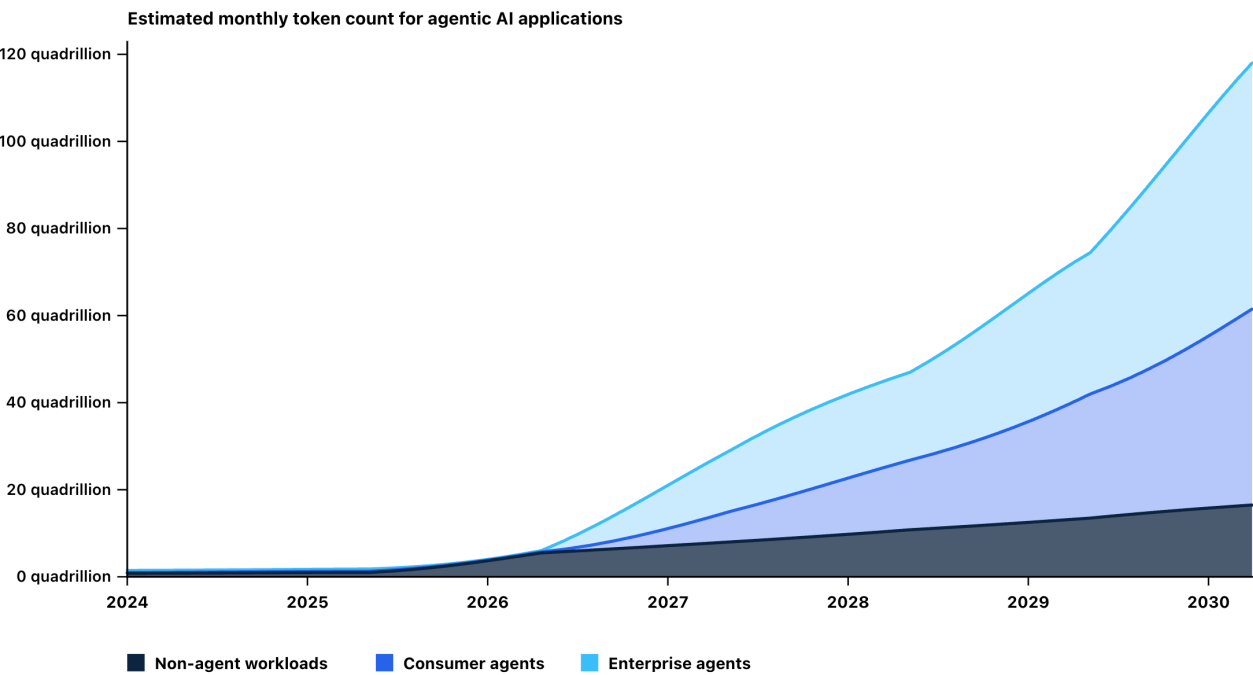
Source: Nvidia Insights, Aurelion Research.

In that environment, paying more for compute or accepting lower efficiency in exchange for faster deployment can make economic sense. The value of getting inference capacity online quickly is becoming extremely high.

The bigger opportunity, in our view, is how quickly AI usage is growing.

Dell estimated late last year that AI inference would reach 1 quadrillion tokens per month by the end of 2028. Its latest estimate is now 57 quadrillion tokens per month for the same period. **That means Dell’s estimate increased by 57x in less than a year.** Goldman Sachs has since published an estimate that is even higher, at roughly 10x Dell’s revised forecast. That is a massive change in expectations, and it shows just how quickly the amount of AI being used is growing.

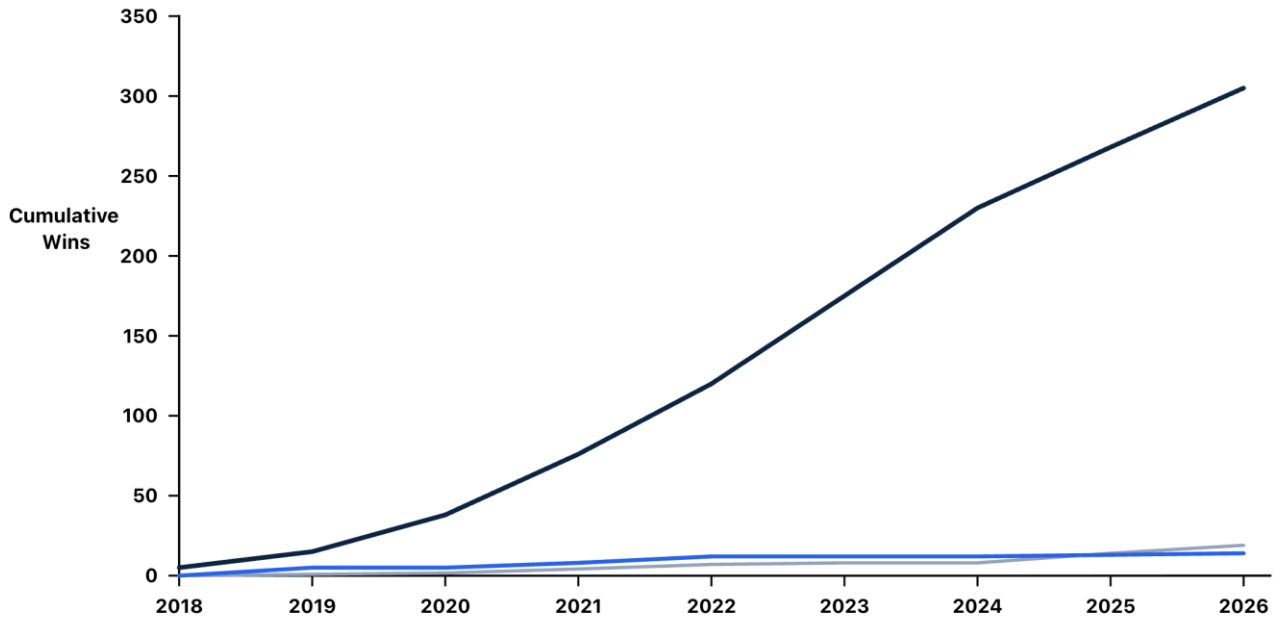
AI Agent Token Usage Could Rise 24x by 2030



Source: Bloomberg, Dell Data, Goldman Sachs Estimates, Aurelion Research.

If token generation continues to grow at anything close to these rates, the amount of compute required to serve AI workloads could expand far beyond what current infrastructure spending suggests. For NVIDIA, this creates an important opportunity: inference does not end when a model is trained.

MLPerf Training and Inference Wins



Source: Nvidia Corporate Presentation, Aurelion Research.

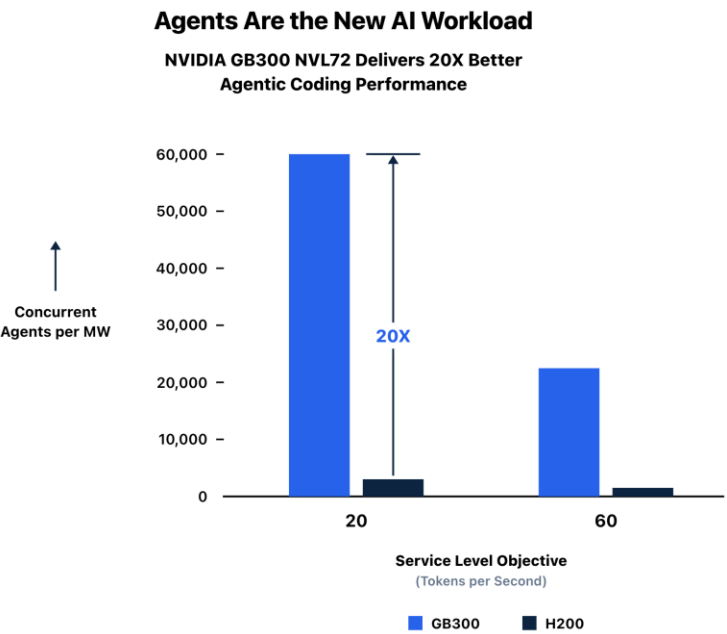
— NVIDIA GPU — Google TPU — All Others

For NVIDIA, the key point is that if AI usage grows anywhere close to these forecasts, the amount of compute required to serve that usage could be far higher than today's infrastructure spending suggests.

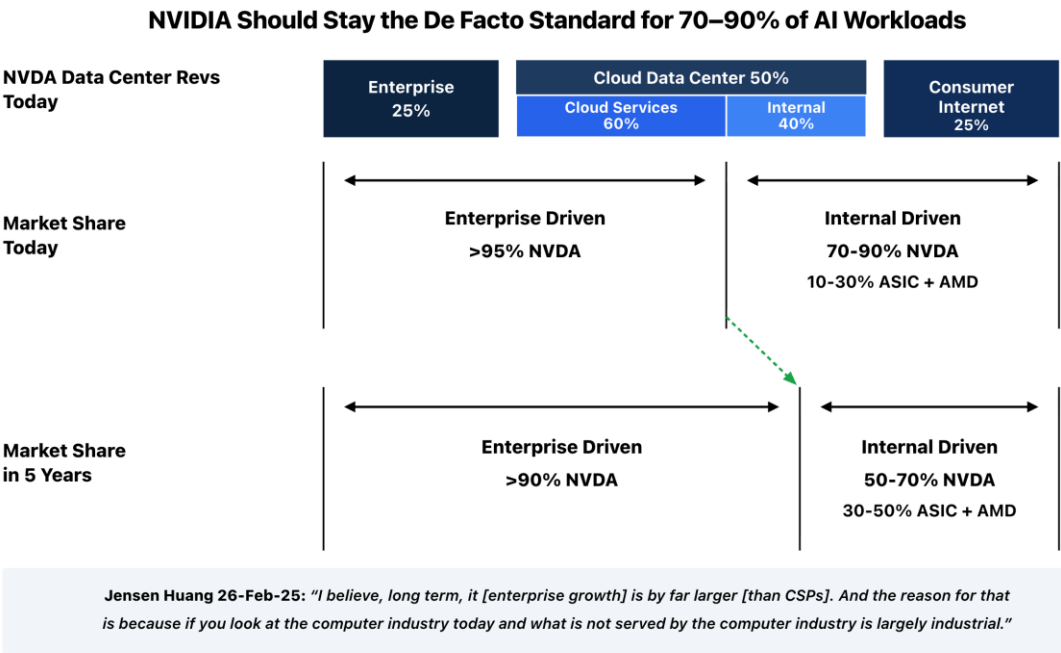
4. The Agentic AI Opportunity

Another standout from our conversation and research is how agentic AI could become one of the most important catalyst for the next phase of AI adoption.

AI is moving beyond answering questions toward completing tasks, using tools, and interacting with existing workflows. As these systems become more capable, companies should have more incentive to deploy them at scale, creating another source of demand for AI compute.



Enterprise AI is already an area where NVIDIA has a very strong position, with more than 90% market share today and a similar share expected over the next five years. The competitive picture is different in hyperscaler and consumer internet workloads, where NVIDIA currently holds roughly 70–90% share but is expected to fall toward 50–70% as custom chips and alternative accelerators gain ground.



Source: Company Data, Evercore ISI, Aurelion Research.

NVIDIA may lose share in hyperscaler and consumer internet workloads while maintaining a strong position in enterprise AI and continuing to grow. The market should keep this in perspective when assessing the competitive risks.

5. Physical AI: Where Are the Robots?

Beyond digital AI, physical AI and robotics could become another meaningful growth driver for NVIDIA over the next decade. **We are bullish on the long-term potential, but we also think adoption will vary significantly across industries.** Autonomous vehicles, and humanoids are moving toward broader commercial use, while many traditional industries are likely to adopt robotics more slowly.

NVIDIA is positioning itself as the computing and software platform behind the robotics industry, providing the technology needed to train, simulate, and operate these machines. CEO Jensen Huang has been vocal about the opportunity.

At CES 2026, he said: “Physical AI has arrived. Every industrial company will become a robotics company.”

While we are very bullish on robotics, we disagree with Jensen on one point. We have a hard time seeing a world where every of companies become robotics companies. Some industries do not change that quickly, and many traditional businesses will continue to operate in ways that have worked for decades.

Take an aluminum plant or an oil refinery. Both can benefit from robotics and automation, but that does not mean they will suddenly become robotics companies. **The important point for the AI community is that some industries are inherently slower to change, and many of these businesses will continue to have a strong human element even as they adopt more AI and automation.**

An oil refinery may use more robots, AI, and autonomous systems over time, but we doubt the people running these traditional industries will suddenly become AI evangelists. Many of these companies are run by operators who have spent decades building businesses around processes they know and trust.

They tend to value control, reliability, and proven technology, and they may have little interest in replacing people with robots simply because the technology exists.

We can already imagine a landman in the Texas Permian Basin saying: “So they really think we’re going to let robots replace us? They’ve probably never been to Texas. This is an oil field, not a robotic daycare.”

Technology adoption is very different across industries, and some of the world’s oldest businesses will remain much more resistant to change than Silicon Valley expects. But that doesn’t mean we don’t see robotics growing massively.

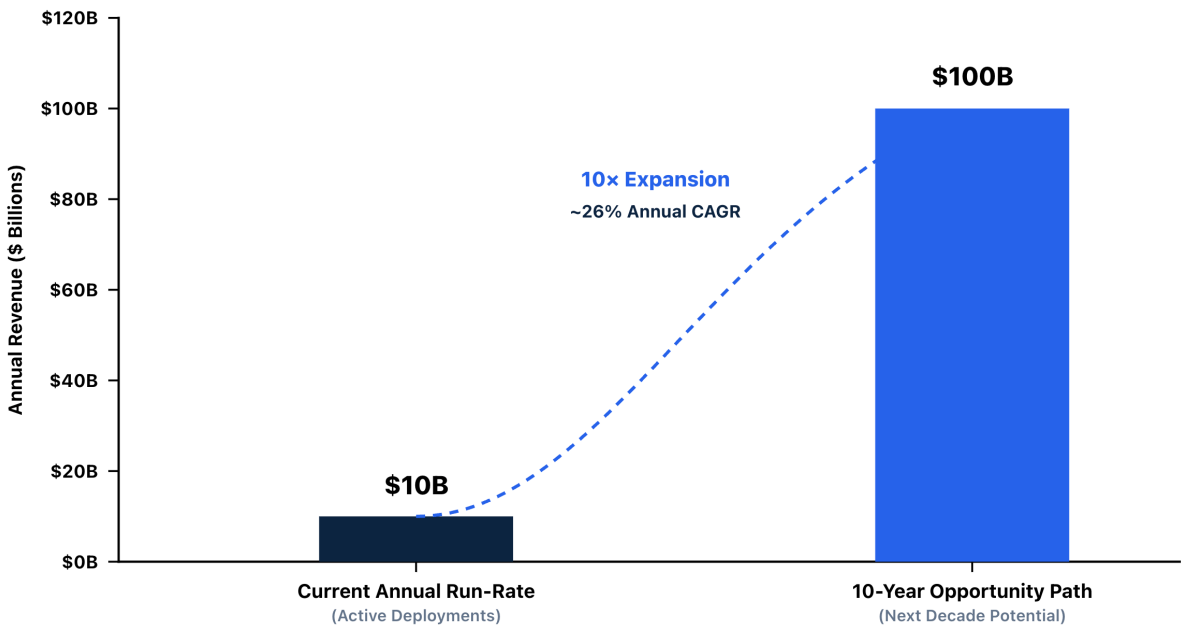
Our take: After inference, we see robotics as NVIDIA’s second-largest growth opportunity in the years ahead, with potential to generate substantial revenue.

The opportunity is enormous. The computing giant estimates that physical AI could eventually be 10x greater than digital AI, given that roughly 90% of the world’s actions happen in the physical world.

NVIDIA estimates that its physical AI business is already running at \$10B in annual revenue, with Huang seeing a path toward \$100B within the next decade.

NVIDIA Physical AI Revenue Expansion Trajectory

Robotics, Autonomous Vehicles & Physical AI Systems



Source: Nvidia Data, Aurelion Research.

What makes NVIDIA particularly interesting is that it can participate across several stages of the robotics ecosystem.

First comes training. Robots need to learn how to see, understand their surroundings, and perform tasks in the real world. NVIDIA’s GPUs provide the computing power needed to train the models that make this possible.

Then comes simulation.

Before putting a robot into a factory or warehouse, developers can train and test it in virtual environments. NVIDIA's Isaac platform, Omniverse and other simulation tools allow developers to generate synthetic data and test robots in different environments without having to collect all of that data in the real world.

Finally comes deployment.

Once the robot is operating in the real world, it needs enough computing power to process information and make decisions in real time. NVIDIA's Jetson Thor is designed for this, bringing powerful AI computing directly into the robot and enabling it to understand its surroundings and respond with very low latency.



Source: NVIDIA website

“We have built three key computers for robotic systems: training computers, synthetic data generation and simulation computers, and robotic body embedded computers.” — Jensen Huang, GTC 2026

This is an important part of the investment case. It can benefit from the growth of robotics without having to predict which company will build the winning robot. Its technology can be used across different types of robots and applications.

NVIDIA is also building the software ecosystem around this hardware.

Its Isaac GROOT platform provides foundation models and tools for humanoid robots, while the company is working with robotics companies including Figure, Agility, Boston Dynamics, FANUC, KUKA, ABB and Yaskawa.

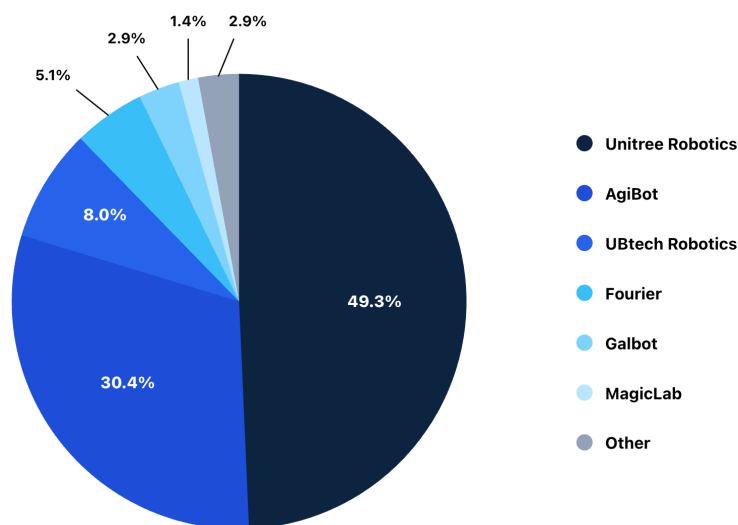
“NVIDIA’s full-stack platform — spanning computing, open models and software frameworks — is the foundation for the robotics industry.” - Jensen Huang, GTC 2026

The opportunity also goes well beyond humanoids.

NVIDIA is targeting industrial automation, warehouses, logistics, autonomous vehicles and other physical AI applications. This gives the company several ways to benefit even if humanoid adoption takes longer than expected.

We are already seeing early signs of commercialization. TrendForce expects China’s humanoid robot output to increase by up to 94% in 2026, with Unitree Robotics & AgiBot together accounting for 80% of shipments.

Market Share of Humanoid Robot Production by Chinese Developers, 2026E



Source: TrendForce (April 2026), Aurelion Research.

The market is still small compared with AI infrastructure today, but the pace of development is what stands out to us. Robotics is moving from research and demonstrations toward real deployments across manufacturing, logistics and other industries. **If that adoption continues, we think robotics could become a meaningful new source of AI compute demand over the next decade.**

“Robotics is a once-in-a-generation opportunity for the European nations.”
— Jensen Huang, World Economic Forum, Davos 2026

The opportunity is still very early, and there are clear questions around the cost, reliability, and economics of humanoid robots. For us, the key thing to remember is that robotics could open another large market for NVIDIA’s existing technology as AI moves further into the physical world.

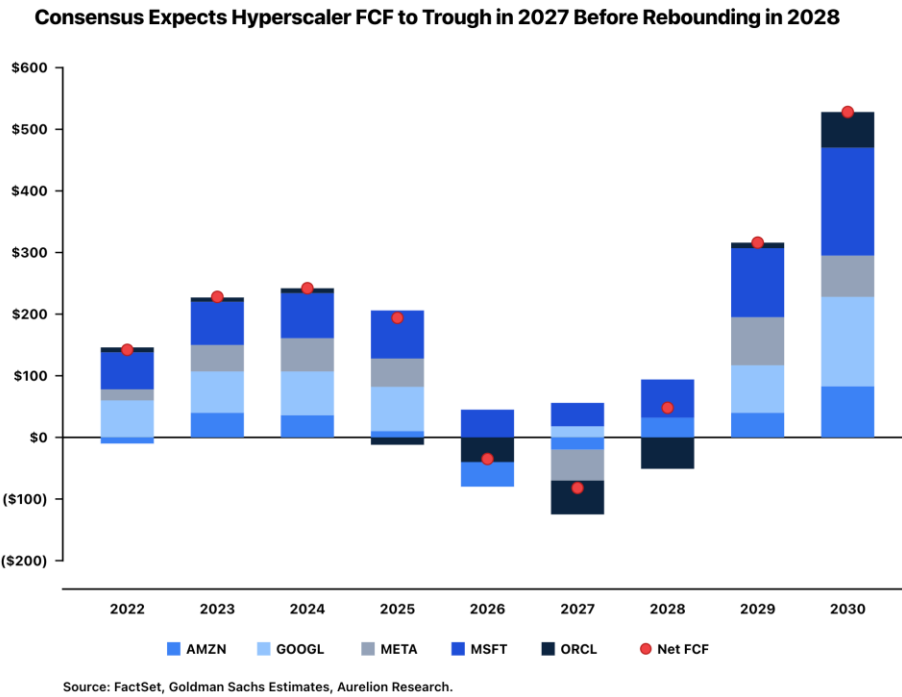
6. Key Risks

Many risks exist around anything related to AI, and NVIDIA is no different.

We see four main risks for the company:

1. AI Spending & Returns

NVIDIA’s growth depends heavily on continued AI infrastructure spending, particularly from the hyperscalers. The key risk is that these investments fail to generate the expected returns, eventually causing customers to slow or reconsider their spending. That said, this does not mean hyperscalers are unprofitable or unable to keep investing. Most of the largest players continue to generate strong profits and operating cash flow, giving them capacity to fund AI infrastructure. The risk is whether the returns are high enough to justify the pace of spending.

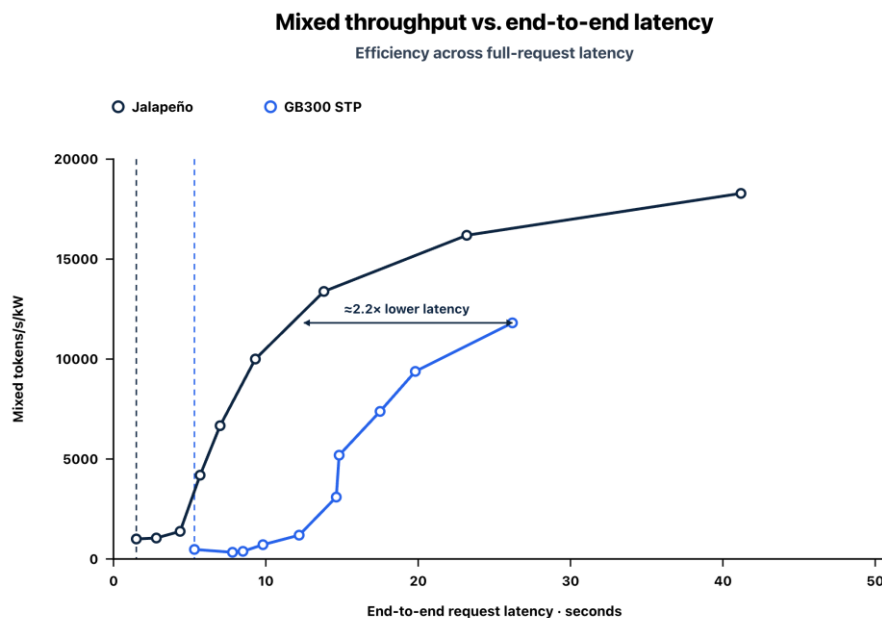


2. Custom Silicon & Competition

There is significant attention today on custom silicon and new AI accelerators. Another company may be able to build a competitive chip for a specific workload, but replicating NVIDIA’s broader platform is much harder.

This includes its software, developer ecosystem, networking, and complete systems. For us, the real competitive risk is therefore whether another company can replicate the full NVIDIA ecosystem, instead of simply matching its chips.

For example, OpenAI is developing its own custom AI chip with Broadcom, designed specifically for inference. Early testing suggests it can outperform NVIDIA and AMD solutions on certain inference workloads



Source: OpenAI, SemiAnalysis, Aurelion Research.

OpenAI plans to begin using the chips internally in late 2026, with broader production expected in 2027. This shows that some of NVIDIA's largest customers are increasingly looking to develop their own alternatives.

3. Supply Chain Constraints

The AI buildout requires the entire supply chain to expand at the same time. Advanced memory, semiconductor manufacturing, packaging, power generation, and data-center capacity all need to increase alongside demand.

This creates the risk of not securing enough components or having to pay more to secure supply, something we have already seen in the past. At the same time, we would not interpret supply constraints as a sign that AI demand is weakening. The industry may simply be growing faster than the supply chain can keep up.

4. Pace of Technological Change

AI is moving extremely fast, with new architectures, accelerators, and models emerging constantly. It will need to keep innovating across its platform while maintaining strong profitability and defending its position against competitors.

7. Our Final Thoughts

After spending time with NVIDIA and looking at the different areas of the AI ecosystem, we came away more confident in the long-term opportunity.

The market is increasingly broadening beyond training and hyperscalers, with inference, agentic AI, robotics, and physical AI creating new sources of demand. This gives the company several ways to continue growing as AI adoption expands.

At the same time, we understand why the market remains cautious.

NVIDIA needs to maintain a very high rate of growth, AI infrastructure requires enormous amounts of capital, competition is increasing, and supply constraints remain a real challenge. We also cannot assume that every area of AI will develop as quickly as the most optimistic forecasts suggest. These are important risks, but they are already well priced in by the market.

Our view is that NVIDIA is one of the best-positioned companies to benefit from the next phase of AI. It has multiple growth opportunities, a strong ecosystem, and exposure to most areas of AI computing. We think the market still underestimates how much room the company has to grow from here.

Important Legal Information, Analyst Certification, and Disclosures

Purpose and intended audience

This report (the “Report”) has been prepared by Aurelion Research (“Aurelion”) solely for informational and educational purposes. Aurelion is an independent equity research provider and is not currently registered as an investment adviser, broker dealer, securities dealer, or fiduciary with any securities regulatory authority in any jurisdiction.

This Report is intended exclusively for institutional investors and other financially sophisticated persons who are capable of independently evaluating the information presented and the risks associated with it.

Nothing in this Report constitutes, or should be construed as, investment advice, a personal recommendation, an offer, or a solicitation to buy, sell, or hold any security, financial product, or instrument, or to engage in any investment strategy. No advisory, fiduciary, agency, or client relationship is created by the receipt or use of this material.

Aurelion does not provide personalized investment, tax, accounting, or legal advice, and this Report has been prepared without regard to any recipient’s specific investment objectives, financial situation, or particular needs. Recipients are solely responsible for their own investment decisions, should not rely solely on this Report, and are strongly encouraged to conduct independent analysis and consult their own professional advisers before making any investment decision.

Analyst certification

The analyst(s) responsible for the preparation of this Report certify that the views expressed herein accurately reflect their personal views regarding the subject securities and issuers. No part of the analyst(s)’ compensation was, is, or will be directly or indirectly related to the specific recommendations, ratings, or views expressed in this Report.

Ratings, time horizon, and terminology

Any rating, opinion, estimate, or forecast in this Report represents Aurelion’s judgment as of the date of publication and is subject to change without notice. Ratings are opinions, not statements of fact, and do not imply suitability for any particular investor. Unless otherwise stated, references to expected performance, price targets, and catalysts should be interpreted over a 12 month horizon from the publication date.

Price targets and valuation

Where provided, Aurelion price targets represent Aurelion’s estimate of the security’s potential value at a specified future date (typically 12 months), based on Aurelion’s analytical methods and assumptions. Price targets are inherently imprecise and depend on the analyst’s judgment. Price targets may be derived from one or more valuation approaches, including discounted cash flow, multiple based valuation, precedent transactions, sum of the parts, net asset value, or other methods deemed appropriate.

There can be no assurance that any price target will be achieved. Actual outcomes may differ materially due to changes in company fundamentals, competitive dynamics, regulation, macro conditions, capital markets, execution, corporate actions, or unforeseen events.

Risks

Investing in equities involves risk, including the possible loss of principal. Any risk discussion is illustrative and not exhaustive. Recipients should independently assess risks and determine whether any investment is appropriate for their circumstances.

Sources of information, verification, and third party data

Information in this Report has been obtained from sources believed to be reliable, including publicly available sources, company disclosures, discussions with company management, and third party data providers (including market data and alternative data sources). Such information has not necessarily been independently verified by Aurelion. No representation or warranty, express or implied, is made as to accuracy, completeness, or timeliness. All information is provided “as is” and may change without notice.

Aurelion assumes no responsibility or liability for errors, omissions, inaccuracies, or interruptions in any third party data, calculations, or model outputs.

Forward looking statements

This Report may contain forward looking statements, including expectations, projections, estimates, or assumptions regarding future events or performance. Such statements are inherently uncertain and subject to a wide range of known and unknown risks and uncertainties, and actual outcomes may differ materially from those expressed or implied.

No duty to update

This Report speaks only as of the date of publication. Aurelion undertakes no obligation to update, revise, or reaffirm any information, opinions, estimates, price targets, ratings, or forward looking statements contained herein, whether as a result of new information, future events, changed circumstances, market conditions, or otherwise. Circumstances, market conditions, and issuer fundamentals may change rapidly after publication without a subsequent note being issued.

Conflicts of interest, independence, and holdings

Aurelion operates independently and does not engage in investment banking, underwriting, mergers and acquisitions advisory, brokerage, or commission based activities. Aurelion's analysts, principals, employees, or affiliated parties may, from time to time, hold positions in, transact in, or refrain from holding the securities of companies discussed. Any such positions or transactions may be inconsistent with the views expressed and may change without notice. Aurelion may highlight alignment of interests (for example, that founders' personal capital is invested in Buy-Rated companies). Such alignment may also constitute a conflict because Aurelion personnel may benefit from price movements in securities discussed.

Hypothetical performance and the Aurelion Index

Any performance figures referenced by Aurelion may relate to the "Aurelion Index," a hypothetical and theoretical model portfolio constructed for illustrative purposes only. The Index does not represent an actual investment product, is not actively managed on behalf of any client, and is not available for direct investment. Past performance is not indicative of future results, and no representation is made that any strategy, process, or model portfolio will achieve similar results.

Performance figures may be presented on a gross basis and do not reflect taxes, transaction costs, management fees, financing costs, or market slippage, all of which would materially reduce actual returns. Total return figures may include an approximation of dividend income based on forward dividend yield estimates that may differ materially from dividends actually received. The Index may employ an equally weighted methodology and may be rebalanced immediately upon additions or removals. Aurelion reserves the right to modify its research process, analytical framework, assumptions, or index methodology at any time without notice.

Third party platforms, redistribution, and display errors

This Report may be distributed, hosted, reformatted, excerpted, or displayed through third party platforms or data aggregators (for example Bloomberg, LSEG Refinitiv, FactSet, and similar services). Aurelion is not responsible for any redistribution by third parties or for any errors, omissions, truncation, formatting changes, or display issues introduced during ingestion, processing, transmission, or presentation by such third parties. Distribution of this material by any third party does not constitute an endorsement of such party or its services by Aurelion, and does not confer any ownership, license, or rights to Aurelion's intellectual property. Recipients should refer to the original Aurelion version as delivered by Aurelion as the authoritative record.

Confidentiality, jurisdictional restrictions, and permitted recipients

This material is confidential and intended solely for the recipient to whom it has been delivered by Aurelion or its authorized partners. It is not intended for distribution to, or use by, any person or entity in any jurisdiction where such distribution or use would be contrary to applicable laws or regulations. Any reproduction, redistribution, or transmission of this material, in whole or in part, without Aurelion's prior written consent is strictly prohibited.

Intellectual property and prohibition on AI or machine learning use

The Report, including its content, structure, analyses, models, charts, and presentation, is the intellectual property of Aurelion and is protected by applicable intellectual property laws.

All rights are reserved. Without Aurelion’s prior written consent, the contents of this Report may not be used, in whole or in part, in connection with the training, development, testing, fine tuning, benchmarking, calibration, or operation of any machine learning or artificial intelligence system, tool, or application, including by ingestion into datasets or prompts intended to improve such systems. This prohibition is intended to mirror the modern protections used by institutional research publishers.

Limitation of liability

To the fullest extent permitted by law, Aurelion shall not be liable for any direct, indirect, incidental, special, or consequential losses arising from the use of, or reliance upon, this material.

Contact and version control

This Report is current only as of its publication date and may be superseded by subsequent Aurelion publications. Aurelion may change its research coverage, frequency of updates, methodology, and analytical framework at its sole discretion.