JEFF'S BULLETPROOF AI PICK

This "Bulletproof" AI Player Will Be Essential in the Data Center Buildout

By Jeff Brown

Neither company will say it out loud... But in many ways, they're partners tied at the hip.

In fact, nine of the world's 30 fastest supercomputers use products from these two companies to most efficiently process large amounts of data.

Most people believe these two companies are competing with each other. After all, they do have similar products. But they each have different strengths... And because both companies are enabling the future of AI, they both can provide huge returns for investors.

This kind of setup happens from time to time in different industries. I experienced it myself when I worked in the semiconductor industry as well as in the IT industry when I worked at Juniper Networks – a pioneer of the networking and communications industry.

The late 1990s marked a major turning point for the internet...

Global internet traffic exploded from about 44 million users in 1995 to nearly 281 million by 1999. It was a massive surge – like nothing the world had ever seen before. And internet infrastructure wasn't anything like what we have today. It wasn't built to handle the flood of traffic.

So, we saw a rapid build-out of fiber optic cables connecting telecommunications hubs to data centers, which allowed all of us to access the internet. And all that new traffic needed to be directed to the right places – whether it was to America Online, AltaVista search, eBay auctions, or Napster to download music.

The devices that handled this job were called routers. For nearly 40 years now, one company has maintained the majority market share in the industry: Cisco Systems (CSCO).

At the time of the internet surge, Cisco owned about 80% of the market share for various kinds of routers. But the industry was hungry for alternatives. They needed competition... not just to drive innovation but also to keep prices in check.

That's where Juniper Networks (JNPR) came into play.

Its founder, Pradeep Sindhu, knew he was coming into the game late. The idea for Juniper didn't take shape until 1996, well after the internet boom had started. Cisco – founded in 1984 – had a 12-year head start.

But Sindhu found his niche: high-end routers – the kind that telecom companies used as the backbone of the internet.

This was a genius move for two reasons.

First, by focusing on high-end routers, he could attract early-stage funding from big-money potential clients like Ericsson, Lucent Technologies, and Nortel Networks.

Second, the market for big internet protocol (IP) routers – the ones directing traffic for large communications companies – was growing at an astonishing 150% annual rate.

When a market grows that fast, there's plenty of room for multiple companies to succeed.

Juniper didn't need to knock Cisco off its perch. It just needed to show it could produce routers as good as, or better than, Cisco. The sales would follow.

The reality is that the market was expanding so rapidly and would keep expanding for many years to come... there was plenty of room for both players.

Juniper's strategy worked right out of the gates. It grew from $3.8 million in revenue in 1998 when it launched its first product to $887 million by 2001.

Investors in both companies made out like bandits during that time. Most of us have heard Cisco's success story... It skyrocketed more than 100,000% from 1990 to its peak in April 2000. It was even the largest company by market cap for a short while.

Juniper investors did quite well too. Kleiner Perkins, the storied venture capital firm, was an early investor in Juniper... and reportedly saw a 230,000% gain on that investment. Public shareholders did well too – Juniper's IPO in June 1999 to its peak in October 2000 saw gains of 4,188%.

And here's the kicker: Juniper never had to displace Cisco.

Its products became just as big of a part of the internet's backbone as Cisco's, and their products worked side by side routing bits of data to the right locations all over the world.

This is important to remember because, in this report, we're highlighting another semiconductor leader that's been essential to the artificial intelligence (AI) boom... That's NVIDIA's silent partner: Advanced Micro Devices (AMD).

This may sound hard to believe, but the reality is that NVIDIA's GPUs are designed to work hand-in-hand with AMD CPUs. Neither company will say it out loud, but in many ways, they're partners.

Companies that manufacture AI servers offer a powerful combination of AMD EPYC CPUs and NVIDIA H100 GPUs. This pairs what many consider the best CPU with the fastest GPU.

The AI data center market is growing fast enough for both NVIDIA and AMD to profit immensely... And just as Juniper made a name for itself in the rise of the internet, AMD is carving out a not-so-little niche for itself in the data center buildout.

The Data Center Buildout Is Gaining Speed

AI is attracting more investment capital than any other trend. And it's not just investors pouring money into AI. Companies are also positioning themselves to capitalize on the massive adoption of this technology.

The total capital expenditures (capex) of just the big four hyperscalers – Microsoft (MSFT), Meta (META), Google (GOOG), and Amazon (AMZN) – exceeds $50 billion each quarter. And as we can see from the chart on the next page, capex is still on the rise.

These are just the four largest hyperscale companies. It doesn't include others like Elon Musk's xAI, Oracle, Apple, etc...

The scale of these data centers being built right now is staggering.

And these footprints are only getting larger. Take Microsoft's Project Stargate, for instance – a $100 billion AI data center slated for completion in 2028. That's nearly three times what Microsoft spent on all its capital expenditures in 2023.

I recently visited Microsoft's $1 billion supercomputer data center, currently under construction in Mt. Pleasant, Wisconsin.
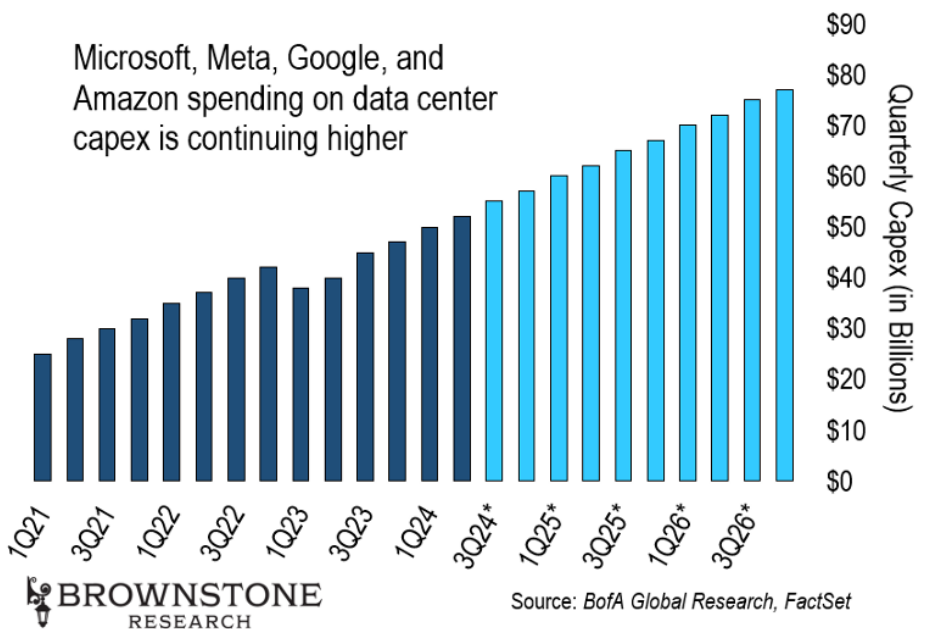
When finished, it's expected to span 1,030 acres – the equivalent of 780 football fields or 1.6 square miles. Now just imagine a $100 billion data center.

The scale of the facility seen behind me under construction was massive. It's hard to imagine something that costs 100 times more to construct. It will be more than a magnitude larger in scale.

Looking ahead a few years, it's not out of the question that we could see a $1 trillion data center built in the 2030s to fully harness the economic power of true artificial intelligence.

Hyperscaler Quarterly Capital Expenditures

Microsoft, Meta, Google, and Amazon spending on data center capex is continuing higher

Source: BofA Global Research, FactSet

This will likely be required to develop an artificial superintelligence (ASI).

The key point is that these massive AI-specific data centers are multibillion-dollar, multi-year projects. And they're going to continue no matter what happens with the economy, interest rates, or the U.S. election.

If we're concerned about a potential recession or the upcoming election, investing in sectors where money is flowing no matter what is a smart move.

Nothing short of a catastrophic event would delay this spending. The prize is just too big.

Artificial General Intelligence (AGI) Is the Ultimate Prize

In the short term, all the major hyperscalers are in a race to achieve artificial general intelligence (AGI).

AGI is the next monumental leap in artificial intelligence. Unlike the AI systems we're used to today – those designed for specific tasks like image recognition or natural language processing – AGI is a form of intelligence that mirrors human cognitive abilities.

It's an AI that can learn, reason, and apply its knowledge across a broad range of tasks, adapting to new situations just as a human would.

Today's AI can excel at particular tasks like beating a world champion at chess or translating languages. AGI would have the capability to understand and perform any intellectual task that a human can. And it would do so with greater speed and accuracy.

But this isn't just about processing more data. It's about creating an AI that truly understands the world around it and can make decisions in real time, in any context, without needing specific programming for each scenario.

AGI is the ultimate prize of AI research. A machine that can think, learn, and reason like a human, but with the ability to improve itself continuously. And AGI will be capable of performing research and development autonomously around the clock.

This kind of intelligence will revolutionize industries, solve complex global challenges, and transform the way we live and work.

And it's not going to be just one company building an AGI. It will be multiple companies with competing AGIs.

Currently, we have seven major players (ex-China) with powerful foundational models. Five of them have well-established user interfaces – a technical term I use to describe their existing products that will act as immediate distribution for their AI technology. (See table below.)

The performance of these models is tracked in real-time and measured against industry standard benchmarks. We have yet to see a "perfect" model. Some are better than others at certain tasks.

Foundation Model Major Players

COMPANY	FOUNDATION MODEL INTERFACE	FOUNDATION MODEL
Microsoft	Microsoft Office, Bing Search, LinkedIn, Xbox, Skype	Has chosen sides with its ~13 billion investment and exclusivity with OpenAI.
Meta	Facebook, Instagram, WhatsApp, Meta Smart Glasses	Has been developing its own foundation model - Llama 2 and beyond.
Alphabet (Google)	Google Search, G-Suite Productivity Products (Competes With Microsoft), YouTube	Has developed its own foundation model, Gemini.
xAI	X (Twitter)	Has developed its own foundation model, Grok, available on X.
Apple	iPhone, iPad, Apple Watch, Vision Pro	Is still woefully behind in AI - has built its own proprietary model and is partnering with OpenAI where it is weak.
Anthropic	Access Through APIs	U.S.-based private company with stated focus on building reliable and interpretable AI models.
Mistral AI	Access Through APIs	Paris-based private company focused on building open-sourced artificial intelligence.

But I'm predicting these companies will all have their own AGI. They all have the capital resources, human resources, and technology to achieve AGI.

The economic benefits of running an AGI will be astronomical. This will justify increased spending on AI by every company in the world... which will justify even more spending on AI data centers. I have no doubt this will be the largest productivity breakthrough in history.

The other thing for us to keep in mind is that some of the most advanced research on artificial intelligence is published in the open. As are open-source foundational models like Meta's Llama 3.1 – which is performing almost on par with the most advanced model from Anthropic, Claude 3.5 Sonnet.

This allows anyone knowledgeable enough to utilize top-quality AI models on their own servers, with their own hardware, and with their own private data.

Providing server equipment for companies running private instances of these large language models (LLMs) or smaller LLMs is already a multibillion-dollar industry.

It's Not AMD vs. NVIDIA... It's AMD and NVIDIA

Too often, financial pundits pit these two companies against each other. But that's the wrong perspective. In reality, the offerings from AMD and NVIDIA often complement each other in significant ways.

We've talked about how the NVIDIA H100 is the gold standard for AI GPUs. But when it comes to server CPUs, AMD's EPYC processors have set the benchmark.

As mentioned earlier, nine of the world's 30 fastest supercomputers pair AMD CPUs with NVIDIA GPUs. These two components were designed to work together seamlessly. Many data centers purchase their servers from other companies – like SuperMicro (SMCI), Dell (DELL), HP Enterprises (HPE), Lenovo, Asus, and Gigabyte. All these major server providers offer configurations that combine AMD CPUs with NVIDIA GPUs.

AMD's server CPU revenue stands at $1.6 billion in the second quarter of 2024. That's larger than their GPU revenue of $1 billion. But I expect that to shift next year. GPUs are more expensive, and in most server setups, one CPU can manage the central processing needs of four GPUs.

Neither company is standing still. NVIDIA has recently developed its Grace CPUs – its breakthrough CPU superchip that is, "designed to meet the performance and efficiency needs of today's AI data centers." And they are directly integrating them with their Hopper H100 GPUs in their Grace Hopper offering.

Once again, this isn't a direct threat to AMD.

Both AMD and NVIDIA are likely to expand their share of the server CPU market. Right now, AMD controls 24.1% of that market, while NVIDIA's share is next to nothing. The other 75% is held by Intel (INTC) which continues to cede market share to AMD.

The fact that Intel still has that much of the CPU server market is quite astonishing. They have bungled every release of server CPUs in the past decade. It started with the two-year delay of their 14nm (nanometer) Broadwell processors in 2014. Every release since has been nearly two years delayed past their planned release date.

And Intel's latest chips, the 4th-generation Xeon series called Sapphire Rapids are built on a 10nm node. That's two generations behind AMD's latest Genoa EPYC CPU which is built on 5nm technology.

Having a smaller node makes the semiconductor faster and more energy efficient. AMD's EPYC outperforms Intel on most general-purpose workloads.

The reality is that AMD absolutely dominates the high end of the server CPU market as Intel has no offering at all.

The only reason Intel has been able to hold onto this market share is because it's a painful process to switch. And many enterprise applications and systems are optimized to use Intel's architecture just because they were there first. And for non-time-sensitive applications, Intel's CPUs still work fine.

But the future of data centers is changing. Here's a recent comment from NVIDIA CEO Jensen Huang…

> *Every single data center… 100% of the trillion dollars' worth of infrastructure… will be completely changed. And there will be new infrastructure built on top of that.*

If data centers don't upgrade their network and computing infrastructure, they won't be able to handle the coming AI workload.

Since these companies will be rebuilding new data center infrastructure from the ground up, they'll continue to switch to AMD for dramatically better performance. And that means that AMD will continue to eat into Intel's market share.

As I said above, the AMD EPYC CPUs have the best benchmark performance. NVIDIA's Grace CPUs are comparable to Intel's Xeon… But that's just in the benchmarking test.

NVIDIA is likely to have optimized compatibility between its Grace CPU and its GPUs. That means, in the real world, their CPU will outperform Intel's as well.

AMD's rise has been impressive. In 2017, it only had 0.8% of the server CPU market, now that number is approaching 25%. With what's happening in the market right now, I can see a path for AMD to attain more than 50% market share of the server CPU market over the next several years.

There has never been a sharper catalyst to upgrade to higher performance than what we're seeing right now with these AI factories.

When it comes to the server compute market, competition is essential. Just like the industry welcomed Juniper's competition to Cisco in the 90s, today's market wants more than one CPU and GPU vendor.

The industry needs and wants competition between at least two companies to keep prices in check and to drive forward the capabilities of GPUs.

This year, both NVIDIA and AMD increased the cadence of their upgrade cycles from 18–24 months down to just 12 months. Now, each company will release a new GPU every year, pushing each other to new heights.

Currently, AMD's flagship server GPU is the MI300X, their answer to NVIDIA's H100. But AMD isn't trying to overtake NVIDIA for training large language models (LLMs)…

Training vs Inferencing

Where AMD is already carving out a lead for itself is in inferencing.

When deploying a model like ChatGPT, OpenAI first must train it using massive computational resources, and then they release the model for consumer and/or business use just like any other software application.

The work that the AI model does to respond to our queries is called inference. We can think of inference as the model "thinking." It makes predictions about what we want based on what was learned from training.

A simpler way to think about inference is as the running of AI applications (as opposed to the training of an AI).

Since inferencing happens after training, AMD made a smart move focusing on this area. Its GPU technology is just about as good as NVIDIA's for training, but inference is like a second wave of computation that needs to be optimized to run AI applications.

The growth in inference compute is expected to outpace training compute in the coming years. One firm predicts demand for inference compute will grow 90% annually through the rest of the decade.

The reason for that is twofold. First, as adoption increases it'll require more compute to handle the increased requests. Second, these generative AI models are creating more complex outputs.

AMD has designed its latest MI300X GPU specifically to handle these inferencing tasks efficiently. The MI300X boasts 2.4 times the memory capacity and 1.6 times the memory bandwidth. (See table below.)

Having more memory is crucial because it allows larger models to fit entirely on the memory of a single GPU. For instance, models like Mistral AI's Mixtral 7x8B LLM, which is popular in production workflows, won't fit on a single H100 GPU but can fit on one MI300X GPU.

This makes the MI300X far more efficient for inference tasks. That's why AMD is seeing increased adoption from major players like Microsoft, Amazon, and Google. In fact, AMD noted that over "100 enterprises and AI customers are actively developing or deploying the MI300X."

In Microsoft's July earnings release, CEO Satya Nadella said:

> *I am really excited to share that we are the first cloud to deliver general availability of VMs (virtual machines) based on AMD's MI300X AI accelerator. It's a big milestone for both AMD and Microsoft. We've been working on it for a while, and it's great to see that today, as we speak, it offers the best price-performance on GPT-4 inference.*

Performance like this is what will drive AMD's business forward and push its stock price higher. AMD CEO Lisa Su mentioned that the MI300X AI GPUs were the "fastest-ramping product in AMD history." They sold over $1 billion worth of these chips in not even two quarters.

Now, this isn't to say that AMD is lagging in training – it's actually running at comparable speeds to the H100 on many benchmark tests.

And AMD powers the world's fastest supercomputer. The Oak Ridge National Laboratory's Frontier supercomputer uses AMD CPUs and GPUs to process data...

AMD Instinct™ MI300X GPU vs. Competition

Hardware Specifications		MI300X (Up to)	H100 SXM	AMD Instinct™ Advantage (Up to)
	TBP	750W	700W	-
	Memory Capacity	192 GB HBM3	80GB HBM3	2.4x
	Memory Bandwidth (Peak Theoretical)	~5.3 TB/s	3.3TB/s	1.6x

AMD is prepared for competition from NVIDIA. Estimates suggest that NVIDIA's new Blackwell GPUs will show a 30x improvement in inference... something AMD will likely match with its MI350X release next year. This is a very healthy industry dynamic that is expanding the entire market and accelerating technological advancement.

Below is a chart of the product roadmap for both AMD and NVIDIA AI GPUs.

We should expect this technological leapfrogging to continue for years to come. And that's going to be great for both companies as AI data centers scramble to equip themselves with the latest and greatest GPUs for training models and running inference.

AMD's New $10 Billion Niche

GPUs are undoubtedly going to be a massive business for AMD in the years ahead. But they aren't the only chips that handle inference tasks.
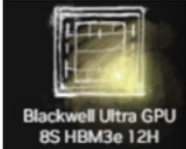
AMD is a leader in field programmable array (FPGA) chips. Long-time readers may remember former portfolio holding Xilinx (XLNX).

We owned Xilinx for its FPGA portfolio and held the shares when AMD announced its intention to buy the company.

Xilinx, now under AMD, controls about 50% of the FPGA market, making it the clear leader in this space. FPGAs are incredibly versatile chips. As the name suggests, these chips are programmable, allowing them to efficiently run a wide range of tasks.

This versatility will be crucial for non-time-sensitive tasks in data centers, like inference. But it also positions AMD as a major player in AI at the "edge" of networks. The "edge" is an industry term to describe when technology is working in devices out in the field, like AI technology embedded in our smartphones, autonomous vehicles, smart homes, robotics, and more.

The FPGA market alone is projected to grow at a rate of 16.4% annually, reaching $25.8 billion by 2029. If AMD can maintain its 50% market share, this implies that its embedded unit could grow from $3.6 billion this year to over $12 billion by 2029.

	2024	2025	2026	2027
NVDA	Blackwell (delayed 1H25)	Blackwell Ultra	Rubin	Rubin Ultra
SKUs	B100/200/102/200A	B210/102 Ultra	...	...
HBM	8H HBM3e	12H HBM3e	HBM4	HBM4
Node	TSMC 4nm	TSMC 4nm	TSMC 3nm	TSMC 3nm
AMD	MI325X	MI350	MI400	...
HBM	HBM3e	HBM3e	HBM4	...
Node	TSMC 5nm (CDNA 3)	TSMC 4nm (CDNA 4)	(CDNA "Next")	
INTC	Gaudi 3	Falcon Shores		

Source: company reports, MSUSA

That's growth no one's really predicting… but I believe it's likely as AI compute moves out to the edge.

And there's one more thing that I predict will result in an even more competitive advantage for AMD.

Intel is in trouble right now, and it is looking to sell its Altera division, which Intel acquired in 2015. Altera was Xilinx's #1 competitor.

Assuming Intel sells Altera, it will be caught up in the typical distractions and restructuring that come with normal M&A activity. AMD's Xilinx product lines will capture even more market share as a result.

Valuation

To get to our valuation, let's first start by looking at revenue growth projections. As the chart above shows, current projections are for AMD's revenue to continue growing 20% a year through the end of the decade.

And even more impressive is the projected EBITDA growth over the same period, 37% a year. (See second chart above.)

This kind of growth allows us to pay a high multiple today but still make above-average returns. But AMD isn't that expensive today when we look at its EV/EBITDA ratio.

It currently trades 36.1x 2024 estimated EBITDA. That's reasonable for a company projecting to grow EBITDA by 37% annually.

AMD (AMD) Annual Sales

AMD's sales are projected to grow over 20% a year going forward — 20% CAGR

2019: $6.7, 2020: $9.8, 2021: $16.4, 2022: $23.6, 2023: $22.7, 2024*: $25.6, 2025*: $32.8, 2026*: $39.3, 2027*: $47.1, 2028*: $57.1

Source: Bloomberg

AMD (AMD) EBITDA

EBITDA is growing nearly twice as quick as sales — 37% CAGR

2020: $0.9, 2020: $1.8, 2021: $4.2, 2022: $6.0, 2023: $4.2, 2024*: $6.5, 2025*: $9.9, 2026*: $12.5, 2027*: $17.1, 2028*: $20.5

Source: Bloomberg
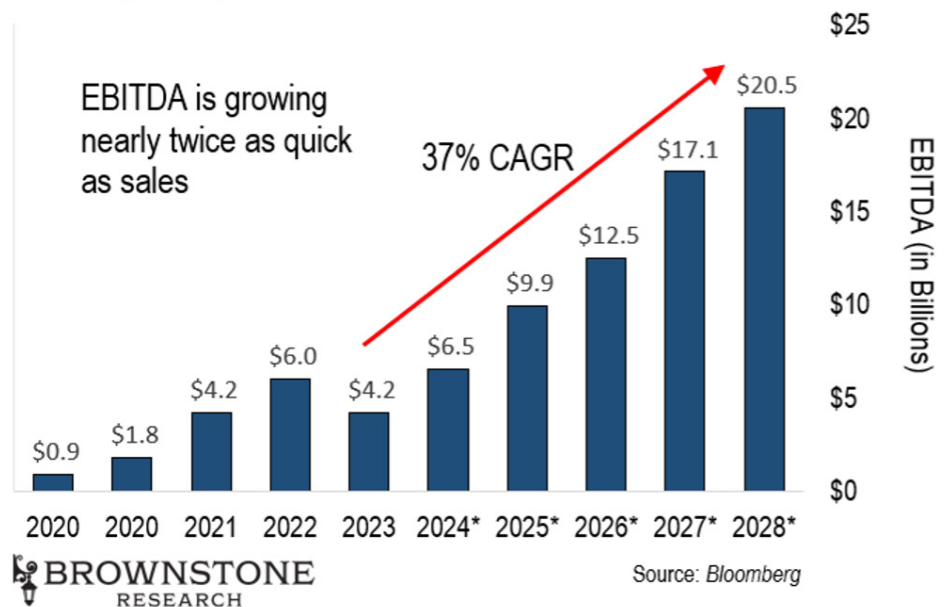
I'd even venture to say it's undervalued.

But in the spirit of remaining conservative with our expectations, we'll keep the valuation multiple the same going forward.

By the end of 2026, the consensus estimate for AMD's EBITDA is $12.5 billion. Using a 36x multiple we get an enterprise value of $450 billion – a 92% gain from today's levels in just 16 months.

<u>Action to Take:</u> Buy **AMD (AMD)** up to a price of $180 a share.

Risk Management: At the time of publication, **AMD (AMD)** is not in the portfolio, so we will not provide regular updates unless we officially add it. If anyone purchases AMD, make sure to use a volatility-adjusted trailing stop.

The percentage stop may change as market conditions change. The current stop loss is 43% for anyone wishing to enter that into their risk management calculations. The exact dollar amount of everyone's stop loss will vary depending on their entry price. I will rarely use a set dollar amount for a stop because, in the past, hedge funds have pushed prices down to our stop, forcing us to sell, so they could establish a position at a lower price.