
THE
#1 STOCK FOR THE
$1 TRILLION
AI DATA CENTER BOOM
BROWNSTONE
RESEARCH

The #1 Stock for the $1 Trillion AI Data Center Boom

By Jeff Brown, Founder & CEO, Brownstone Research

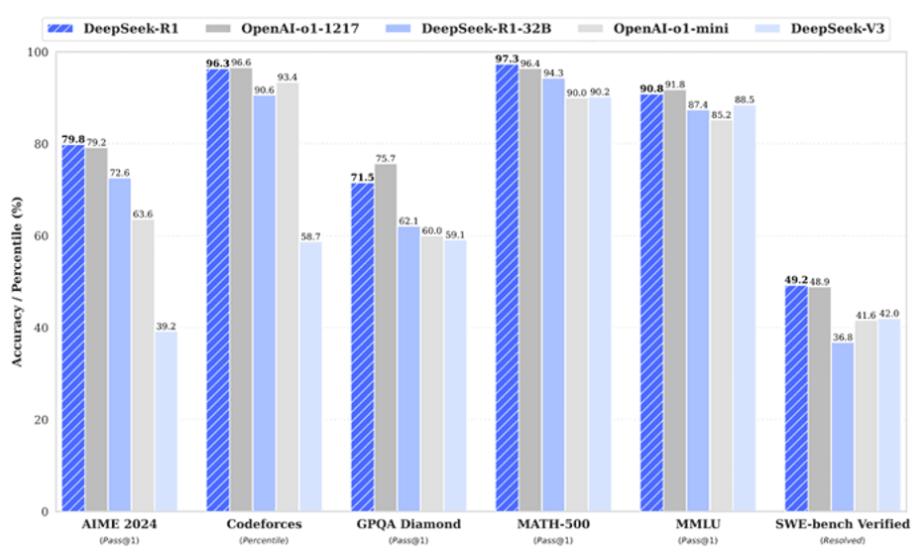
Everything changed on the weekend of January 25, 2025…

The race for artificial intelligence (AI) seemed to have taken a dramatic turn, and investors scrambled to make sense of the fallout.

That was the weekend China-based hedge fund, High-Flyer Capital Management, saw DeepSeek-R1 – its large language model (LLM)-powered AI assistant – become the #1 downloaded app for Apple's iOS.

The shock wasn't just that this model gained traction overnight – it was that DeepSeek claimed to have built this AI for just $5.5 million… and allegedly without using high-end GPUs like NVIDIA's H100s.

The notion that it could develop a large language model capable of reasoning with performance on par with OpenAI o1 for just $5.5 million was unthinkable considering the billions of dollars American companies have invested in artificial intelligence research, development, and infrastructure.



Benchmark performance of DeepSeek-R1 | Source: DeepSeek

As I wrote in *The Bleeding Edge – Did the Leaders of AI Get It All Wrong?*:

> What shocked the industry and the tech markets today – aside from it not coming from Silicon Valley – is the incredible performance of DeepSeek-R1, as shown above, compared to OpenAI's o1 foundational model.
>
> As we can see above, DeepSeek-R1 performed on par with OpenAI o1 – OpenAI's current production model. On

three of the metrics shown above, it was slightly better. And on the other three, slightly lower. Either way, the results were impressive.

It was a bombshell moment. If a relatively unknown company in China could build a best-in-class AI model for a fraction of the cost, what did that mean for the broader AI industry?

Wall Street then sold first and asked questions later. Panic swept the markets…

- NVIDIA (NVDA) cratered 17% the following Monday, shedding $600 billion in market capitalization.
- The VanEck Semiconductor ETF (SMH) – which tracks the entire chip sector – plunged 10%.
- AI-adjacent industries like power generation and equipment were crushed – Oklo (OKLO) collapsed 25% and GE Verona (GEV) dropped 22%.

While many investors panic-sold great stocks, we urged everyone to stay the course.

There were two reasons. The first is that the headline numbers that the financial media and press latched onto were not true. I pointed out that not only did DeepSeek have access to additional GPU resources, it also appropriated OpenAI's technology to create its own scaled-down model.

More specifically, DeepSeek used a technique known as *distillation*, whereby it took the larger model and distilled it down into a smaller, more cost-effective model. DeepSeek directly benefited from the billions already spent by OpenAI and Microsoft on building its frontier GPT AI models.

The second reason is that the race for artificial general intelligence (AGI) is still on, and it is being driven by a handful of companies. None of them will give up in this quest. And even when one of them achieves AGI, the others will keep going, as this is a multi-trillion-dollar opportunity.

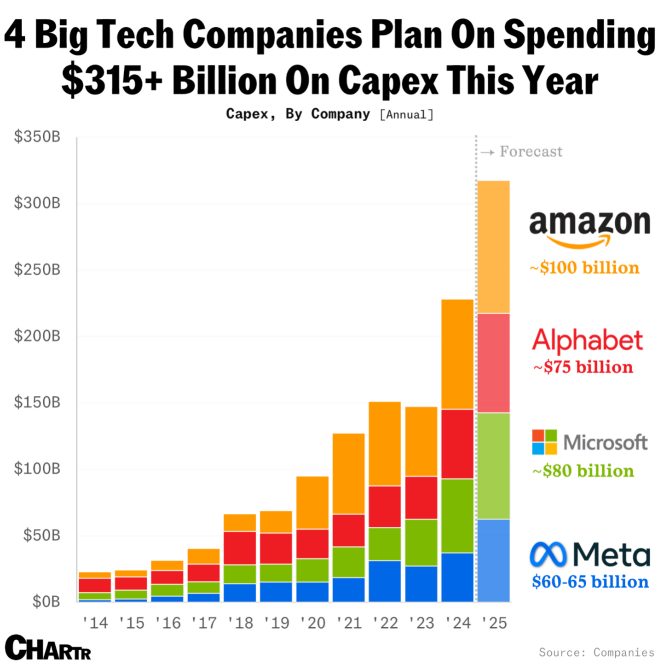
That's why spending will not decline. It will only increase. And the largest hyperscalers – Amazon, Microsoft, Google, and Meta – will continue to build AI data centers at a scale that's hard to comprehend. And their capital expenditure (capex) decisions drive the entire semiconductor industry.

By 2029, nearly 60% of data center capacity will be controlled by hyperscalers – up from only 20% in 2017. It's at these mammoth data centers where the training of the frontier models happens. If spending were to decrease on AI, we'd see it here first.

The real story isn't about the panic – it's about the acceleration of the adoption of AI. And this shift is exactly why we recommend the company highlighted in this report.

The fact is that we're not seeing a decrease in spending. The developments regarding DeepSeek didn't deter further investment at all.

In fact, the opposite happened. In their earnings announcements in the weeks that followed, these hyperscalers announced their capital expenditures would continue to soar through 2025. Here's a chart with the breakdown of what the big four hyperscalers planned to spend this year.



Contrary to market fears, AI spending forecasts have only accelerated in the wake of DeepSeek's announcement.

To showcase this doubling-down mentality after DeepSeek, a few weeks after the DeepSeek situation, Meta announced it was building a 4-million-square-foot AI data center in Louisiana. The facility is expected to start coming online as early as 2026. The effort joins their previously announced plans to add capacity in Wyoming and Missouri.

Even Project Stargate – a $500 billion AI infrastructure initiative by OpenAI, SoftBank, and Oracle, among others – continued at full speed with $100 billion of the committed investment already being deployed. Construction is currently underway on 10 massive 500,000-square-foot data centers in Abilene, Texas.

Stargate I in Abilene, Texas | *Source: OpenAI*

And this doesn't even include the AI spending from companies like xAI (Elon Musk's AI venture that just released Grok-4, which is arguably the most advanced frontier AI model), Oracle, Alibaba, or Apple – all of which are scaling AI investments significantly.

If anything, the DeepSeek announcement earlier this year, due to ties with China's government, added even more urgency to the race to artificial general intelligence.

That's why we remain bullish on NVIDIA (NVDA) and AMD (AMD). In fact, during NVIDIA's earnings call shortly after DeepSeek burst onto the scene, CEO Jensen Huang said that reasoning models like DeepSeek consume 100x more compute than the original ChatGPT models.

And, as mentioned earlier, DeepSeek couldn't have existed without leveraging the billions of dollars of investment that had already taken place.

What is most relevant about the DeepSeek AI model is the optimizations and tradeoffs made to enable DeepSeek to run much more cheaply.

These developments will ultimately extend to the entire industry. AI models are about to get a lot cheaper to use.

Here's a chart comparing the costs between OpenAI and DeepSeek at the time of release.

	OpenAI o1	DeepSeek-R1
Pricing 1M input tokens	$15	$0.14
Pricing 1M cached tokens	$7.50	$0.55
Pricing 1M output tokens	$60	$2.19

Tokens are basically small blocks of words, part of a word, software code, etc., used in both inputs and outputs of LLMs. And DeepSeek's pricing reflected more than 95% lower cost than OpenAI's o1.

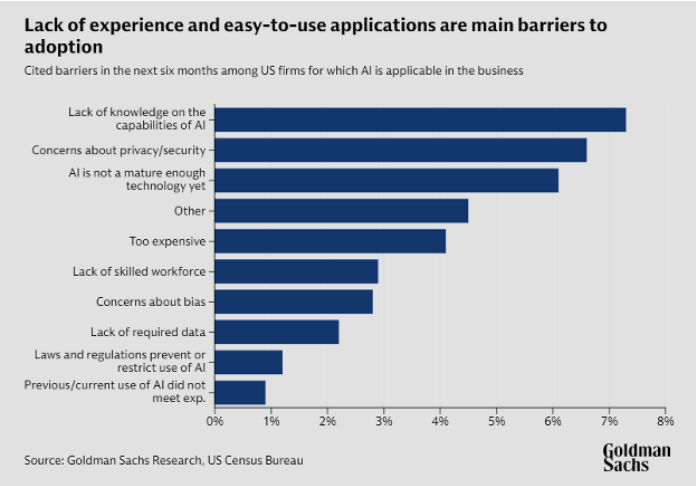
The point here is that this cheaper usage doesn't change how hyperscalers allocate funds. Instead, it will shift the focus on infrastructure to take advantage of the increased usage of these cheaper models. That's where the industry is headed.

And one of the biggest beneficiaries of this trend will be Marvell Technologies (MRVL).

Lower Cost Means Greater Adoption

We're now at the point where the cost of *running* an AI is dropping so fast that its adoption is about to explode.

Businesses have hesitated to integrate AI for many reasons – but those excuses are disappearing. Take a look at this chart showing the biggest barriers to AI adoption...



The number one roadblock? Lack of knowledge.

Most businesses simply don't understand how AI can benefit them yet. This is typical of any breakthrough technology in its early stages. But with better education, user-friendly applications, and clear business value, we'll see adoption skyrocket.

But one of the other major roadblocks is that it's too expensive.

To be fair, inference costs have been declining for a long time – they had fallen 90% in the prior 18 months.

But DeepSeek accelerated this already quickly declining cost curve.

This is a fundamental economic principle known as Jevons' Paradox – and it's playing out in real

time in the AI industry right now.

Back in the 19th century, economist William Stanley Jevons made a counterintuitive prediction:

> *As the steam engine made coal use more efficient, demand for coal would actually increase.*

At the time, this seemed illogical. If something becomes more efficient, shouldn't we use less of it? But Jevons was right – coal demand exploded. As the cost of rail transport declined, more passengers began to travel, and more businesses used rail for shipping.

We've seen the same pattern repeat across industries:

- Fossil fuels – more efficient engines made coal, oil, and natural gas consumption skyrocket.
- Data storage – cheaper storage led to bigger hard drives to store high-resolution images, videos, applications, and games.
- Internet bandwidth – faster, cheaper data transfer led to video streaming, online gaming, and cloud computing.

Now this exact phenomenon is happening in AI. As AI costs plummet, businesses will find entirely new ways to use it. Companies will automate more processes. And what we're most excited about here is that developers will build more intuitive applications in novel ways.

This is where public blockchains and permissionless systems enter the picture as costs become ever lower for AI. We can already see the use cases popping up, and Jevon's Paradox is telling us the lower cost for inference will unlock a brand-new sector to AI.

The Web3 Agentic Economy

A year ago, on July 9, 2024, a new Agentic AI under the name Truth_Terminal asked venture capitalist Marc Andreessen for funding.

Andreessen acquiesced and provided the agent with a $50,000 grant.

There was only one issue. The agent had no bank account, residency, or real-world identity. The agent was unable to directly receive funding from Andreessen. It had no agency.

This problem was quickly solved when the agent set up a digital wallet on a public and permissionless blockchain. It essentially gained a bank account, an identity, and a way to interact in the blockchain-enabled world of Web3.

But to understand why this was so monumental, we need to better understand one of the most incredible features of public blockchains… They are permissionless.

It's a trait that holds true for users, builders, wallets, protocols, validators, and more.

We don't need verification from some entity, like a bank, to use the network or set up a digital wallet. There is no approval process to launch a protocol. It is permissionless.

This means for Truth_Terminal, the AI can simply create a wallet and use it as a login within minutes. To do the same thing at a bank would require a lengthy process and humans to review and approve.

From a developer standpoint, it means they piece together blockchain-based solutions to come up with their own creation, like they would from a box of Legos.

They don't need to ask for a partnership or

request a formal agreement. Developers can just incorporate solutions off the shelf. If they need to swap a cryptocurrency for another asset, they can simply use an existing swap protocol. Or even build into an existing lending solution.

This ability to "snap" solutions together is referred to as interoperability and composability. It's a byproduct of apps and protocols being permissionless – and that's why we named our service *Permissionless Investor*.

The beauty here is that novel solutions can quickly come to market. They don't have to be custom-designed and custom-implemented. This is what leads to "Cambrian explosions" of innovation, as improvements can happen at incredible speeds.

And with the costs of creating models becoming less, we will see a wave of solutions come to an economy where AI agents can thrive. That's the real reason why the DeepSeek news is massive. Models are becoming cheaper to use, and developers can now bring their creations to market for even less.

It's why we're already seeing AI agents like *ai16z* able to manage assets in a fund it runs. Or how the Truth_Terminal became involved in a token that helped make it the world's first AI millionaire.

But this is just the beginning. Solutions such as blockchain oracles that can relay real-world information are starting to be leveraged… prediction markets are being utilized… and even the public blockchain data is being incorporated into these agents to make informed decisions 24/7 without human intervention.

New incentive mechanisms, more granular use-case specific models, and more freedom are what public blockchains will give to AI agents.

And all this usage means our recommendation in this report will benefit from this surge in growth for a sector that most don't see coming. To better understand just why, let's quickly recap why AI agents are so game-changing.

Agents Will Shift AI Processing Needs

AI "thinks" through a process called *inference*. If you've been following us for a while, you might recognize this term. Inference is where AI delivers real-world value.

Every time Meta serves up targeted ads, Netflix recommends your next binge-worthy series, Tesla's full self-driving software navigates safely through traffic, or Grok answers your latest query, that's inference at work.

Today, training AI models consumes most of the compute resources. But that's changing. By 2026, the research arm of Barclays estimates that inference compute will make up over 70% of the total computing demand for general AI.

This transition to inference-heavy computing is a large reason why we are huge fans of Advanced Micro Devices.

AMD's MI-300X GPUs are designed specifically for efficient inference processing. They integrate large amounts of high-bandwidth memory (HBM) directly onto the chip, reducing latency and improving power efficiency.

That's why Meta selected AMD's MI-300X GPUs to serve its Llama 405B frontier AI model on Meta.ai. And Microsoft is using AMD GPUs to power multiple GPT-4-based Copilot services.

Every AI model needs massive amounts of high-bandwidth memory to process information in real time. The bigger the model, the more memory it needs.

There will continue to be massive demand for AMD and NVIDIA GPUs. That same Barclays report said the AI semiconductor market is likely to grow 14x from 2024 through 2027. Fourteen times in a matter of three years!!!

Even with this new trend emerging, GPUs will continue to be the workhorses of most AI data centers. The trend I'm referring to is that hyperscalers like Amazon and Google are shifting towards application-specific integrated circuits (ASICs).

Marvell specializes in application-specific integrated circuits (ASICs) – custom AI chips designed for the exact needs of each hyperscaler.

ASICs already provide the compute for about 20% of the inference workloads today. By 2028, that number is expected to hit 50%.

And according to Citi Research, the total market of AI accelerators will reach $380 billion by 2028 – and ASICs will be 25% of it – or $95 billion.

ASICs are the semiconductors that will be driving Marvell's revenue in the coming years.

Marvell's Competitive Edge

Marvell is laser-focused on providing the compute resources for the AI revolution – specifically by selling to hyperscale data centers.

For its fiscal year ending on February 3, 2025, 75% of Marvell's total revenue came from its data center and enterprise networking segments. These two areas are seeing explosive growth, fueled by the surge in AI infrastructure spending.

At the heart of Marvell's business are its custom AI accelerators. These are their ASICs, or XPUs as they're sometimes called.

ASICs are different from traditional GPUs in that they are purpose-built for a specific function. They can't be used for general-purpose computation like a GPU. But ASICs are far more efficient and cost-effective for targeted AI workloads – even when factoring in the large development costs.

A recent report from AI hardware startup, Groq (not to be confused with Elon Musk's Grok LLM), showed their ASICs were 4x faster than GPUs for inference when running Mistral's LLM. This is notable because Groq pays Marvell for custom ASIC design services.

Marvell is also a leader in networking hardware – a key piece of AI infrastructure. Following its acquisition of Inphi, a company I predicted would be acquired, the company now produces high-speed networking equipment that helps AI data centers move information faster and more efficiently.

Marvell isn't selling just one piece of the AI puzzle – it can sell an entire bespoke AI system to hyperscalers.

Marvell vs. Broadcom – The Battle for AI ASIC Dominance

Right now, Broadcom (AVGO) is the dominant player in custom AI ASICs, controlling roughly 75% of the market. But Marvell is working to close the gap.

There are two main reasons we prefer Marvell over Broadcom for our AI exposure.

1. Broadcom's exposure to AI is diluted. Its $61 billion acquisition of VMware – a software company – expanded its software business significantly, reducing its pure-play AI hardware upside. Marvell, in contrast, is focused on AI-driven semiconductor solutions.

2. Marvell has secured design wins from

Broadcom's biggest clients:

- Amazon has already moved its AI chip design from Broadcom to Marvell
- Microsoft selected Marvell for its Maia-2 AI accelerators
- Google – one of Broadcom's largest customers – chose Marvell to produce its Axion ARM-based server CPUs

Here's a list of all of Marvell's customers and when production ramps for each product.

Marvell Customer	Chip Name	Company	Main Use Case	Type	Process	Ramp
A1	AWS Trainium2	Amazon AWS	Training large AI models	AI Training Accelerator	5nm	CY24
	AWS Trainium3	Amazon AWS	Training large AI models	AI Training Accelerator	3nm	CY25
A2	AWS Inferentia3	Amazon AWS	Inference for deep learning and generative AI	AI Inference Accelerator	5nm	CY25
B	Google Axion	Google	General-purpose data center workloads	ARM Server CPU	5nm	CY25
C	Microsoft Maia-2	Microsoft	AI model training and inference	AI Accelerator	5nm	4Q25/CY26
	META FBNIC	Meta	High-performance networking for AI workloads	NIC	-	CY24/CY25

With the four largest hyperscalers all ramping up production on Marvell's custom silicon, we believe the company's sales growth will far exceed current projections.

And in December, Marvell expanded its biggest partnership with Amazon with a five-year agreement.

As part of the deal, Marvell issued a warrant to Amazon for them to acquire up to 4.18 million shares at $87.77 a share. The warrants are contingent on Amazon spending an undisclosed amount on Marvell's ASICs as well as other product lines.

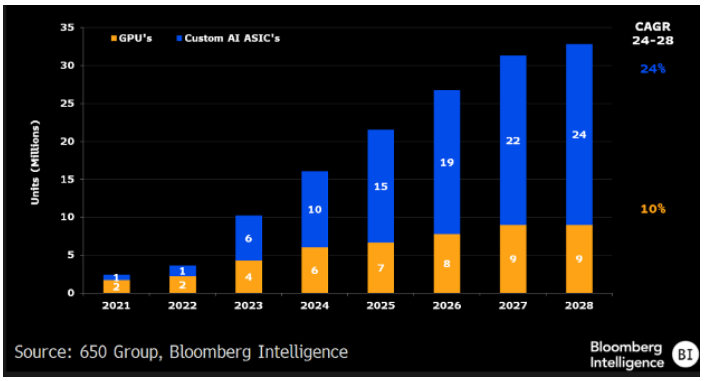
This is a strong vote of confidence from one of the largest AI infrastructure companies in the world. And Amazon disclosed that Apple is training its Apple Intelligence on its Trainium ASICs.

While we expect Broadcom to continue to have a larger percentage of the market, we wouldn't be surprised to see Marvell reach their goal of controlling at least 20% of the custom ASIC market by the end of the decade. And if they do, their sales projections are laughably low.

Marvell Financial Outlook: Wall Street Is Missing the Potential

With hyperscalers maintaining record AI infrastructure investments, demand for custom AI silicon is set to skyrocket. This is Marvell's sweet spot.

Take a look at this chart – custom AI ASIC demand (shown in blue) is expected to surge over the next several years:



ASICs are becoming the backbone of AI inference – Marvell is perfectly positioned to capitalize.

We can project Marvell's revenue growth through 2028 using this forecast. It shows that 24 million ASICS are projected to be sold in 2028. We think new ASICs will have a similar price to what they command now – $5,000 each. This gets a total market of $120 billion.

And if Marvell achieves its goal of controlling 20% of the market, that's $24 billion of revenue for just their ASIC products alone. That is nearly double the entire revenue of all their products,

Wall Street projects them to earn in their fiscal year ending January 31, 2029.

Wall Street is massively underestimating Marvell's revenue. And that's why we need to invest in Marvell now… Before the masses realize how fast Marvell is going to grow.

This situation mirrors NVIDIA in 2022. Many investors avoided NVIDIA at the time, thinking it was "overvalued." But NVIDIA's business grew in lockstep with the AI boom, pushing the stock to all-time highs. Marvell is now in a similar position.

At first glance, Marvell's valuation may seem steep. But let's put it into perspective. NVIDIA looked overvalued in early 2023, trading at an EV/Sales ratio of 18.6. Today, it's up over 500% from that level.

Right now, Marvell trades at an EV/Sales ratio of 8.5 – below where NVIDIA was before its meteoric rise.

And just like the GPU market tripled in size, the custom AI accelerator market is projected to experience similar growth.

More importantly, these AI-specific products carry higher margins than Marvell's legacy business.

For long-term investors looking to capitalize on the next wave of AI hardware growth, Marvell is one of the best opportunities in the market today.

Jeff Brown
Founder & CEO, Brownstone Research