Amazon (AMZN): Trillion Dollar Question

Summary

Amazon trades at roughly $209 today, significantly below what a conservative sum-of-the-parts analysis suggests the business is worth. The business is growing fast, with AWS seeing 24% growth backed by a $244 billion contracted backlog growing 40% year-over-year), advertising growing at 23% annually with essentially no incremental infrastructure cost and retail growing around 10%, but the market is pricing the company as a low growth company losing market share in important divisions and suffering from an increasingly capex, which is misleading. In reality, Amazon is capacity-constrained, not demand-constrained, and is well positioned to grow from the advancement of AI as well as profit from moat-strengthened segments in retail and advertisement.

The thesis is that the market fails to understand each of Amazon's different segments and their respective competitive advantages, pricing power and growth potential. The market seems not to realize that the company has incredible pricing power to improve margins on their retail business, both international and in the US, which would drive predictable FCF over many years. Meanwhile, Amazon Ads has silently grown into one of the world'd largest advertisement platforms and one where the company has strong margins and pricing power given the competitive dynamics of their marketplace. Finally, AWS's backlog is not only one of the largest in the industry but also one of the most reliable, with strong non-AI contracts but also with significant, risk-adjusted exposure to AI megadeals with OpenAI and Anthropic, without much risk of default. Their Trainium and Graviton chips will allow the company to be even stronger in the data center industry and will allow them to serve customers at lower and lower costs, especially as inference AI, not training, gains ground in the next few years.

What Amazon Does

Segment 1: Retail (North America + International + Prime)

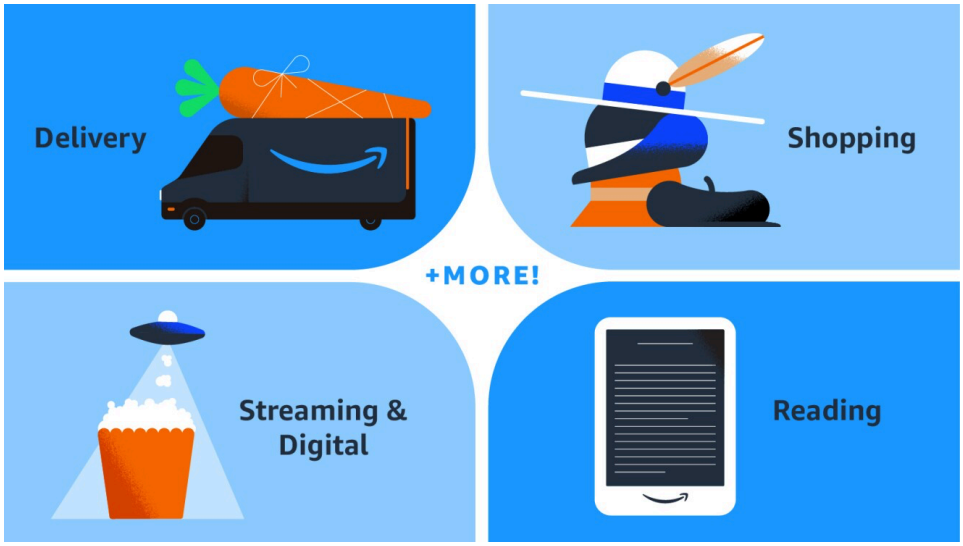
Amazon's retail operation is best understood as two distinct businesses layered on top of each other. The first-party model functions like a traditional merchant where Amazon buys inventory, marks it up, and sells directly to consumers at structurally thin margins in the low single digits. The strategic purpose of 1P is not profitability, it is category leadership and price credibility. By pricing an average of 14% below major competitors, Amazon lowers the barrier for new customers to create an account and make their first purchase, breaking the unfamiliarity barrier and making them more interested in coming back for more products.

The third-party marketplace is where the economics become genuinely compelling. In this case, rather than booking the full gross merchandise value of a sale, Amazon only recognizes the take rate, like listing fees, FBA logistics fees, and referral fees, while bearing none of the

inventory risk. As the mix continues shifting toward 3P, the margin profile of the retail segment improves structurally without any incremental capital investment. The seller does the work; Amazon collects the toll.

Amazon's 3P take rate operates through a base referral fee of 8-15% depending on category, plus FBA logistics fees that add another 10-20% for the 82% of sellers who use Amazon's fulfillment network. Combined with advertising spend, which, as we'll see in the next section, sellers increasingly treat as a mandatory cost of maintaining visibility, the all-in effective rate runs closer to 30-40%, meaning Amazon earns roughly $0.30-0.40 on every dollar of worth of products it never touches, on inventory it never owns.

Prime, their subscription business, offers free shipping, Prime Video, grocery, pharmacy, music, and reading into a single subscription that makes leaving genuinely costly, and this shows up in consumer behavior. Customers who use same-day grocery delivery shop more than twice as often as those who do not, a frequency advantage that compounds across every other category Amazon sells. Subscription services grew 14% in Q4 2025 to $13.1 billion, and Amazon has quietly become a top-tier U.S. grocer with over $150 billion in gross grocery sales.



Analyst estimates put global Prime membership at roughly 260 million, with approximately 185-200 million in the U.S. alone, representing about 80% of American households. Prime members spend an average of $1,400 per year on Amazon versus $600 for non-members, meaning the subscription's primary function is locking in a spending pattern, not necessarily generating $139 in annual fees.

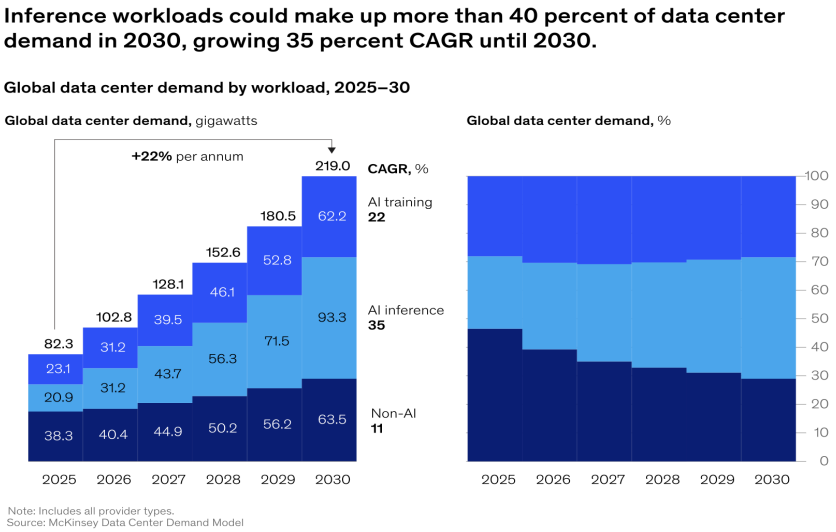
North America retail generated $426.3 billion in revenue, growing 10% YoY, with operating income of $29.6 billion at a 6.9% margin, up from 6.4% in 2024. International contributed $161.9 billion in revenue, growing 13%, with $4.7 billion in operating income marking its first full year of structural profitability after years of deliberate reinvestment.

Segment 2: Amazon Web Services (AWS)

Most companies pay AWS because building and maintaining data center infrastructure is extraordinarily capital-intensive and not their core competency, like a bank or a hospital needs compute and storage to run its operations but has no interest in actually running these programs themselves. AWS offers companies access world-class infrastructure on demand with two pricing structures, usage-based billing by the hour or gigabyte for variable workloads, and committed multi-year contracts at lower rates for predictable ones.

With the rise of AI, demand for compute has split into two distinct categories. Training, the process of building and improving models, makes companies willing to sign the largest contracts in cloud history. Inference, what happens every time a user queries an LLM or an enterprise application calls a model in production, is recurring, high-volume, and growing with every company that moves AI from pilot into deployment. Training gets the headlines, but inference is where the durable volume lives. AWS is well-positioned for both as, in the near term, intensifying competition among AI labs to build better models is driving unprecedented long-term compute commitments that flow directly into the backlog. Over the medium term, as enterprise AI shifts from experimentation into production at scale, inference demand compounds

continuously and becomes the kind of predictable, recurring revenue that places the cloud provider hosting the compute at an extremely advantageous position.



AWS is the profit engine and the growth story simultaneously where in FY2025 it generated $128 billion in revenue, accelerating to 24% growth in Q4, the fastest quarterly rate in years. Operating income reached $45.6 billion at a 35% margin, which is more than half of Amazon's total operating income while representing less than a fifth of revenue.

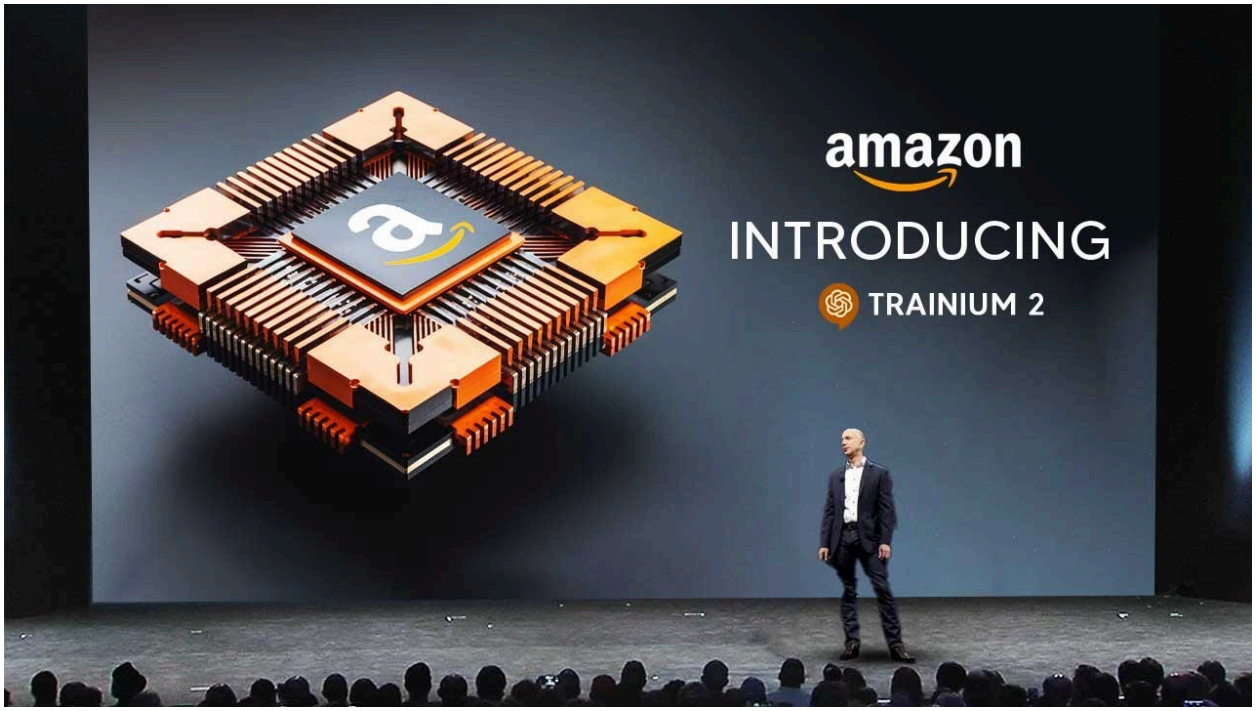
The business model spans over 200 services such as compute, storage, database, analytics, machine learning, and AI-native workloads priced on usage-based or committed-contract terms. The critical feature, as we will explore soon, of the moat here is that switching costs are architectural rather than merely contractual. A company that has built its production infrastructure on AWS doesn't just face a penalty clause if it leaves, it also faces the prospect of rebuilding years of integrated systems from scratch.

The most underappreciated number in the Amazon story is the backlog, and news about it keeps getting better. At $244 billion, up 40% YoY and 22% QoQ, it represents revenue that has

already been contractually committed by enterprise, government, and hyperscaler customers that a simple revenue growth screen entirely misses. On the Q4 earnings call, Andy Jessy said AWS could actually grow faster if supply were available, which means, as we've been seeing on the news of companies needing more and more compute, the constraint is power and data center construction, not customer demand. Amazon responded by adding more data center capacity in 2025 than any other company in the world, 3.99 gigawatts for the full year, including 1.2 gigawatts in Q4 alone.

The custom silicon story adds a compounding dimension. Trainium and Graviton together now carry an annual revenue run rate above $10 billion, growing at triple-digit percentages YoY. Trainium2 is already validated at scale, powering Anthropic's Project Rainier, the world's largest operational AI compute cluster at 500,000+ chips. Jessy said that "Graviton is up to 40% more price-performant than leading x86 processors and is used expansively by over 90% of AWS's top 1,000 customers. Graviton itself is a multi-billion-dollar annualized run rate business, growing more than 50% year-over-year." Trainium3 is in production with nearly all the supply expected to be committed by mid-2026. Each chip generation widens the cost advantage, and as AI workloads grow more compute-intensive, the value of that price-performance gap compounds accordingly.

Long term, when AWS runs workloads on Trainium instead of NVIDIA GPUs, the hardware margin that would otherwise go to NVIDIA flows directly to AWS operating income instead. The near-term reality is a hybrid fleet but as Trainium handles a growing share of the highest-volume workloads, AWS's blended cost per compute unit falls and margins expand without any change in customer pricing. The $10 billion run rate growing at triple digits suggests the substitution is already well underway.



Segment 3: Advertising

Amazon Advertising generated $21.3 billion in Q4 2025, up 23% year-over-year, and roughly $56 billion for the full year, making it the third-largest digital advertising platform in the world behind only Google and Meta. When a seller wants their product to appear at the top of Amazon search results, they bid in a real-time auction against competitors selling similar items and Amazon collects the winning bid every time a customer clicks, regardless of whether a purchase follows. The seller who bids highest for a keyword gets the top placement, and because Amazon's search traffic is full of people who already intend to buy, conversion rates are high enough that sellers bid aggressively to win that visibility.

The core moat is structurally different from either competitor. Google captures what people are searching for and Meta captures who they are and what they look at, meanwhile Amazon captures what people actually want to buy. When a consumer types a product query into Amazon's search bar, they have already cleared the awareness and consideration stages of the purchase funnel and they are now ready to buy. That intent premium drives higher conversion rates, which justify higher auction CPCs, which produce higher ROAS for advertisers, which attract more advertising spend, which makes the auction more competitive. The cycle is self-reinforcing, and it means that for most consumer product categories, not advertising on Amazon is economically equivalent to not appearing in Google search results. For someone selling running shoes, having their running shoes appear at the top of an Amazon search list is an incredibly important input to their sales and worth a significant dollar value in ads.

The margin profile makes this segment particularly compelling, as scalability is easy and capex costs that suppress real margins in AWS are not present here. Amazon does not disclose advertising operating margins, but the economics are straightforward, the marginal cost of serving an additional ad impression on an already-built platform is essentially zero, which means each incremental dollar of advertising revenue falls through to profit at a rate that likely puts the segment at 45-55% operating margins, similar to Google ads or Facebook Ads and higher than AWS on an incremental basis.

Prime Video advertising is still in its early stages, with a 200 million monthly viewer base that has barely begun to be monetized and competition for ad space in Amazon’s website will become increasingly high, driving up ad spend, as the flywheel effect of having your products shown there becomes stronger. The incremental revenue from layering advertising onto it, priced at rates that will rise as targeting precision is demonstrated to buyers, represents a multi-year expansion path that does not yet show up in the current run rate and it is a growth avenue, as well as a multi-billion dollar segment, that most analysts seem to overlook.

Competitive Moat Analysis

AWS: The Switching Costs

AWS's competitive moat operates on three dimensions that have collectively held through a decade of aggressive competition from Microsoft and Google. The first is breadth and

integration depth. AWS offers over 200 services spanning compute, storage, database, analytics, machine learning, and AI-native workloads, and enterprises rarely stay in one corner of that catalog. A company that starts with storage expands into compute, then analytics, then AI services and each layer adopted makes the next migration more painful, because leaving AWS eventually means rearchitecting every integrated dependency simultaneously rather than moving one service at a time. The switching cost is the accumulated complexity of years of build decisions that assumed AWS as the foundation.

The second dimension is the partner ecosystem, as thousands of independent software vendors, systems integrators, and managed service providers have built their entire businesses around AWS certifications and integrations, the human capital and the related AWS services create a high switching cost that forces companies to stay. Every year a company stays on AWS, its internal teams develop deeper AWS-specific expertise, its third-party integrations multiply, and the organizational cost of switching grows, so the cost of retraining developments or reeimplementing related services is just too high to justify a slightly less expensive Azure or Google Cloud.

The third dimension is custom silicon, and it is the one actively widening the moat as its capacity improves and as Amazon decouples their capacity to Nvidia's chips, and the risks of not having optionality, while other cloud providers else can't. Each chip generation compounds the cost advantage, and as AI workloads grow more compute-intensive, the value of that gap grows with them. The infrastructure itself is now proprietary so customers running workloads optimized for Trainium are not just locked into AWS by contracts or convenience, but by hardware-level performance advantages they would forfeit the moment they left. If Nvidia suddenly increases the price of their processors or if another member of the chip supply chain does it, Amazon is way less impacted than competitors and firms, knowing about this optionally, pay a premium for AWS's services.

Consider a mid-size bank that moved its core operations from on-premise hardware to AWS five years ago. It runs its fraud detection on SageMaker, its customer data on RDS, its real-time transaction processing on EC2, and its analytics on Redshift, all deeply integrated, all tuned over years of iteration, all staffed by engineers whose entire cloud expertise is AWS-specific. Its core banking software vendor, its compliance tooling, and its third-party risk platform all integrate natively with AWS. And increasingly, its inference workloads are running on Trainium. To switch to Azure, this bank would need to rebuild every one of those integrations from scratch, retrain its entire technology organization, renegotiate with every vendor, migrate years of sensitive financial data, and accept years of operational risk during the transition, all while its existing AWS setup works perfectly well. The accumulated integration depth, the ecosystem that has grown around it, and the proprietary hardware advantage all point to the same conclusion: switching is theoretically possible but practically irrational, and it gets more irrational with every year the bank stays.

And guess what? If Amazon chooses to increase prices by 10% to boost margins this would still not be enough to get companies to pay the cost of uncertainty and the risk of re-integrating their entire system instantly.

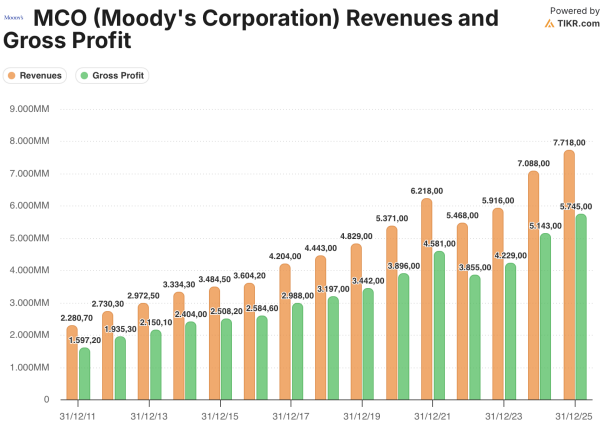
Retail and Prime: The Flywheel Asset

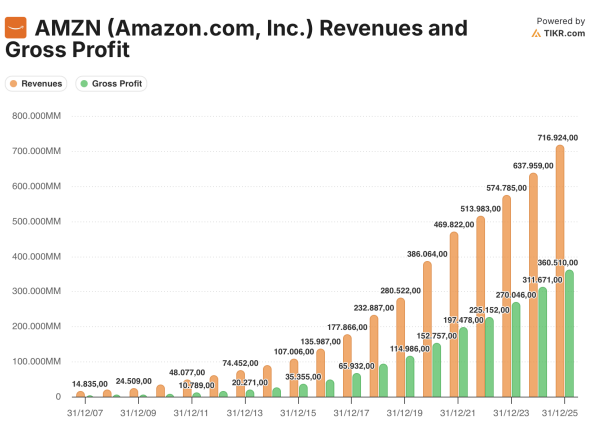
Retail is Amazon's bread and butter and this is where I believe they actually have the strongest competitive advantage coming from extremely strong network effects and an immense cost of opportunity for retailers who think about leaving the platform. On the flipside, its convenience and subsidized pricing scheme makes it the ideal place for consumers to buy their everyday products. As time passes these benefits become stronger and stronger.

Amazon's retail margins are frequently misread as a sign of weak competitive positioning, but they reflect a deliberate choice that no competitor with a different cost structure can replicate. The business operates on thin retail margins by design, using price leadership to drive volume, which in turn feeds the flywheel of marketplace fees, advertising revenue, and Prime subscriptions, as well as consumer familiarity and trust with the platform. The marketplace itself has become effectively non-optional for sellers: if you want to compete in e-commerce, you have to be on Amazon. This creates a dynamic where Amazon takes a cut of every transaction, giving it exposure to the downside of margin pressure from sellers in weak demand environments but significant upside when consumer spending accelerates — the platform captures volume growth without bearing the inventory risk that traditional retailers face.

Amazon's pricing power on its take-rate is substantial and underappreciated. The company can easily increase its take rate by 2 or 3% above inflation annually, and sellers would have no choice but to comply because the alternative, leaving Amazon, is commercially unviable for most businesses and it means losing most of their revenue stream. This is not a fee that sellers can negotiate away, it is closer to a toll on e-commerce itself, and the toll keeps going up and can sustainably keep going up that way for a long time.

Whenever I'm analyzing a moat like this I always remember Chris Hohn's strategy with TCI. Moody's, one of his longest term holdings, did nothing flashy to improve its bottom line, but they managed to pass along pricing increases above inflation for decades which, alongside sustainable demand for their product, reflected an incredible increase in revenue and EPS over the years. I think Amazon, even without the same regulatory moat that MCO has, can achieve the same, and they have been, given their network effect moat.





The introduction of Rufus, Amazon's AI shopping agent, is compounding this dynamic. Used by over 300 million customers in 2025, Rufus drove approximately $12 billion in incremental annualized sales and delivered a 60% lift in purchase likelihood. On top of the boost to retail, each interaction feeds data back into Amazon's advertising targeting engine, making sponsored placements more effective for sellers and more valuable for Amazon.

The logistics infrastructure underneath all of this is the deepest moat. Amazon delivered over 8 billion items same-day or next-day to U.S. Prime members in 2025, a 30% increase year-over-year. Rural same-day delivery monthly customers nearly doubled over the same period. This is not a company harvesting a mature logistics network — it is still actively expanding the moat geographically, pushing same-day delivery into markets that were previously next-day or two-day. Every new fulfillment center, every new delivery route, raises the bar for any competitor attempting to match the service. The capital required to replicate this network is measured in tens of billions and years of execution, and Amazon is still outspending every potential challenger.

This is why sellers are not pushing back harder on take-rate increases: they are making more money even after the fee hike, because the AI-driven conversion improvements more than offset the incremental cost and the massive logistics infrastructure of Amazon makes their lives so much easier, allowing for more volume anyway. Amazon is effectively charging more while simultaneously making the platform more valuable, which is the hallmark of genuine pricing power rather than rent extraction.

Internationally, the picture has shifted from deliberate investment to early-stage harvest. Amazon generated $4.7 billion in operating income from international operations in FY2025 after years of intentional losses. The playbook, invest heavily in logistics infrastructure, subsidize consumer pricing to build volume, earn loyalty through Prime, then monetize with third-party seller fees and advertising, is now proven across enough markets that international looks like an earlier-vintage North America. The losses were not a sign of failure; they were the cost of replicating the domestic flywheel in new geographies. With international now profitable and growing, Amazon has effectively doubled its addressable market for the same toll-road economics that drive the domestic business.

Advertising: The Moat No One Can Replicate

Amazon is in a great position when it comes to their advertising segment. Transaction specific purchase data is the most predictive signal in commercial advertising because it closes the loop between intent and conversion, so advertisers on Amazon are not inferring purchase likelihood from clicks or screen time, they are targeting consumers at the precise moment of buying. Google knows what people search for, Meta knows what people look at, but Amazon knows what people actually buy. This is why Amazon's ad revenue grew 23% in Q4 2025 to $21.3B despite operating in a mature digital ad market, and why its ad margins (estimated at 58-65%) rival or exceed those of the pure-play ad platforms.

Consumers who go on Amazon and type in whatever product they want to buy, say “running shoes”, are one of the highest conversion customers ever, as they are in a marketplace actively looking for a product ready to spend money. Merchants looking to advertise their products there have a strong certainty on their ROI exactly because of the type of customers that are in the platform. Merchants who don’t spend money on advertisement will end up losing market share to the ones that do as they will be shown lower and lower in the search results, which will force them to join the bidding war and actually spend some money to be shown to customers who look on the platform. A new running shoe company should, and is, spending a lot of money on Amazon ads to get more visibility and a higher ROI and the establishment shoe company needs to catch up on advertisement not to lose their Amazon market share.

This dynamic gives Amazon extraordinary pricing power over its ad inventory. Because sellers compete against each other for visibility on a platform where purchase intent is already established, Amazon can raise ad prices without losing advertiser demand. There is no viable substitute, a seller cannot move their Amazon ad budget to Google or Meta and achieve the same closed-loop, bottom-of-funnel conversion. The competitive pressure between sellers effectively sets a floor on ad spend that rises with the number of participants, while Amazon controls the supply of placements, which they can keep deliberately low. Price increases are absorbed because the alternative is losing shelf space to a competitor who will pay the higher rate.

There is a cost of opportunity to not advertise on Amazon, and the company can profit from it.

Trainium and Graviton: Custom Chips leading AWS

Trainium is a chip designed exclusively for AI model training, developed by AWS itself (actually, the company that produces it was acquired by Amazon 10 years ago for $350 million). The key difference between Trainium and NVIDIA chips is that NVIDIA designs a powerful general-purpose chip and then figures out how to connect many of them together, whereas Amazon designed Trainium starting from the question of what a large-scale AI training or inference system needs to do as a whole, and then built the chip to match that system. The result is that the communication between chips, which actually powers a powerful model,

happens in dedicated hardware rather than being managed by software as an afterthought, which leads to some benefits as we will discuss here.

Trainium1 (2023, 7nm) required 16 chips working together to hit 3 PFLOPS (measure of how many calculations per second a chip can make). By Trainium3 (2025, 3nm), a single chip delivers 2.52 PFLOPS, competitive with NVIDIA's B200. At the system level, a full Trainium3 UltraServer (144 chips) pushes 362 PFLOPS with very low latency, per AWS's architecture documentation, which is actually a huge advantage for inference AI models.

On cost, Trainium is the clear winner compared to other top GPUs. According to CloudExpat's cross-platform comparison, instances, the cost to rent a GPU for an hour, run at roughly $4.80/hour versus $9.80/hour for equivalent H100 capacity, translating to 30-50% lower total cost of ownership and the cost of production of Amazon’s GPUs is half as low as NVIDIA’s. Power efficiency is also superior, where Trainium3 achieves 40% better performance-per-watt than Trainium2, and delivers 5x higher output tokens per megawatt. At hyperscale, these economics are actually transformative and really matter for the bottom line, as Anthropic's Project Rainier (500,000 Trainium2 chips across a 1,200-acre Indiana facility) saves millions annually in power costs alone.

Where NVIDIA retains its advantage is raw absolute performance and, critically, the software ecosystem. The B200 offers roughly 2.5x higher raw training performance and up to 4x better inference, making it great to run the cutting edge AI models and train them. More importantly, which is actually a great part of NVIDIA’s economic moat lives in, over 90% of AI developers use CUDA, with 100,000+ third-party applications built on it, which makes switching a whole data center to another system or integrating other chips on it really hard. Trainium requires the Neuron SDK, which struggles with dynamic control flow, and has gaps in framework support, making these GPUs incompatible with most of the current GPU clusters run by major cloud providers based on NVIDIA. Additionally, because of this design and Amazon’s focus on building this GPU only for this one purpose, models with custom operations may require significant refactoring, which makes other uses less friendly. The ecosystem immaturity, with fewer tutorials, smaller community and worse inference capabilities, remains the primary barrier to adoption outside hyperscaler environments.

Customer traction is concentrated but real as AWS disclosed on its Q4 2025 earnings call that Trainium2 is "fully subscribed" and growing 150% quarter-over-quarter as a multi-billion dollar business. In terms of commercial contracts, Anthropic is running 1+ million Trainium2 chips and building Project Rainier as the world's largest non-NVIDIA training cluster and OpenAI has committed to 2 gigawatts of Trainium capacity, small compared to their total demand but significant for a rising GPU model. As explained before, however this adoption is heavily weighted toward hyperscaler partners with engineering teams large enough to absorb the CUDA to AWS system transition costs, which, compared to NVIDIA’s chips, it's not yet a broad market product, nor does it want to be.

The most interesting strategic move is Trainium4 (expected late 2026 / early 2027) which will support mixed clusters of Trainium4 and NVIDIA B200/H100 chips. This is pragmatic and it

actually undermines NVIDIA's competitive advantage by slowly removing their cluster incompatibility, allowing marginal growth of data centers to follow whichever chip has better cost-benefit, letting customers adopt Trainium incrementally without ripping out existing NVIDIA infrastructure.

Amazon has a long way to go if they want to beat NVIDIA in the GPU competition, both in competitive advantages, created by NVIDIA's chip compatibility, set up integration to current data centers and already built in applications, and inference performance. However, this is not the mission of the Trainium program and analyzing it for what it is, a powerful, cheaper and data-center focused GPU, it has the chance to allow AWS to offer a much better competitive position than Azure, Google Cloud or Oracle in the next few years.

Graviton is a family of processors that replace Intel Xeon and AMD EPYC in instances. Unlike Trainium, Graviton doesn't compete with NVIDIA, it competes with x86 incumbents on general cloud workloads, used for general purpose computing like web servers, databases, containers, analytics, and application hosting.

Graviton instances cost up to 20% less than comparable x86 instances, and deliver 40% better price-performance than Intel C5 in most instances. Graviton4 over Graviton3 brought 30% better general compute, 40% faster databases, and 45% faster large Java applications, showing that this CPU is gaining more power while being faster/lower latency than before. The energy angle is significant, as Graviton delivers 60% less energy consumption than comparable x86 instances, which at AWS's scale translates to enormous operational cost savings.

Adoption is arguably Graviton's biggest success story where the company has reported that over 90,000 AWS customers use Graviton as of Q4 2025, representing roughly a third of all new installed workloads. Real-world savings reported include a media company achieving 28% compute cost reduction with 15% faster processing, and retail analytics workloads seeing 32% cost drops with 20% faster completion, per AWS case studies.

The goal of these CPUs isn't to perform any sexy computing for AI training, it is just to do their job as well as possible, as fast as possible while trying not to spend so much energy. And at this Amazon is doing a great job. With these families of chips, there is a very good chance that AWS will be able to offer significantly better prices than their hyperscalers competitors, allowing them to keep and increase their market share over the next few years.

The only real competition comes from Google's TPU, which allows for significant energy and cost reductions compared to NVIDIA chips and positions them as the only other hyperscaler capable of matching AWS's vertical integration on AI training costs and power efficiency. The competition is valid and Google's TPU offers similar power and cost efficiencies, but AWS still has a leverage given their already larger customer base and stronger integration network between GPUs. This competition, however, should be one to watch over the next few years as it could dictate the future of hyperscaler competition.

As inference becomes a bigger thing and AI companies stop spending money in endless training of models, which we explore soon, cost and power efficiency, not performance alone, will be two of the biggest data center preference drivers, and Amazon, as of now, has a credible claim to take the lead.

Q4 2025 and FY2025 Financial Review

Amazon reported Q4 2025 revenue of $213.4B, up 14% YoY and ahead of the $211.3B consensus. AWS was the standout at $35.6B, growing 24% and beating expectations by $700M, advertising came in at $21.3B, up 23% and narrowly above consensus. On the retail front, North America operating income expanded to $11.5B from $9.3B in the prior year, while AWS posted $12.5B at a 35% margin.

The headline EPS of $1.95 missed the $1.97 consensus by two cents, but the miss was entirely attributable to $2.4B in non-recurring charges coming from 3 major sources: an Italy tax and lawsuit settlement of $1.1B, $730M in severance costs, and $610M in asset impairments. Adjusting for these, clean operating income would have been $27.4B versus the $25.0B reported, an almost 10% beat on an underlying basis.

For the full year, Amazon delivered $716.9B in revenue (up 12%), $80.0B in operating income (up 17%), $77.7B in net income (up 31%), and $139.5B in operating cash flow (up 20%). The reported free cash flow figure of $11.2B, down sharply from $38.2B in FY2024, looks alarming in isolation but is driven entirely by a $50B YoY increase in capital expenditure. Operating cash flow grew 20% and remains the correct measure of underlying business cash generation and, as we will discuss soon, the capex increase reflects an investment opportunity guided by genuine demand, not a failure in the company's money generation capabilities.

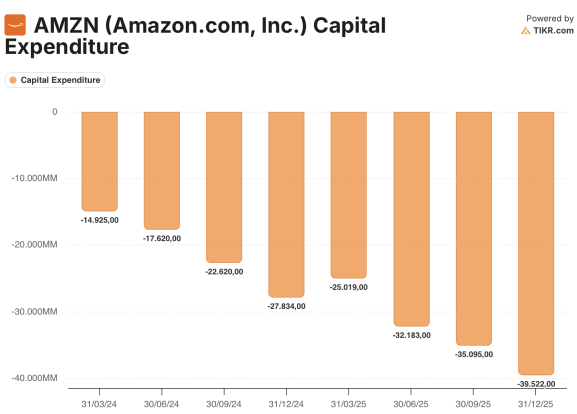
Q1 2026 guidance calls for $173.5–178.5B in revenue and $16.5–21.5B in operating income. The unusually wide operating income range reflects the accounting treatment of Amazon Leo satellite launch costs, with roughly $1B being expensed as incurred before transitioning to capitalized treatment.

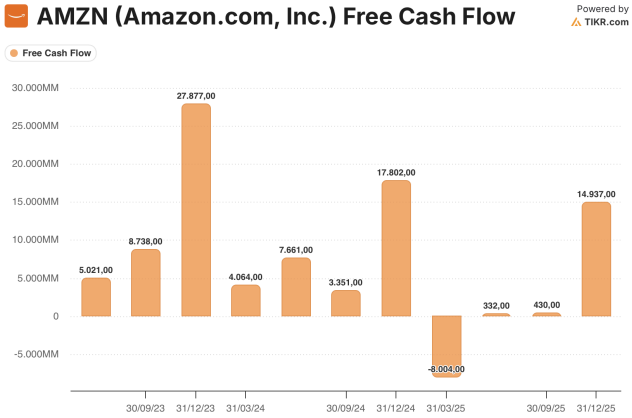
Why the Market Is Pessimistic — And Why It's Overblown

The $200 billion capital expenditure commitment for 2026, a 52% increase over the $131 billion spent in FY2025 and the largest single-year corporate investment commitment in U.S. history was the number that sent the stock down 8% on earnings day.The implicit fear is that Amazon is making a speculative bet on AI demand that may not materialize at the scale required to justify the investment.

Just to be clear, the headline number includes Project Kuiper satellites (20+ launches planned in 2026), robotics/fulfillment, and custom chip development (Trainium, Graviton). Analysts estimate the AWS-specific portion is roughly $130–140B.

Indeed this 2026 capex commitment is already on top of a historic high in capex expenditures for 2025, where Amazon's investment commitments sent Free Cash Flow to only 7.6 billion, down from $32 billion in 2024 even as cash from operations hit a record high in the year. The fear is that Amazon, and other hyperscaler companies such as Microsoft and Google, whose share prices have also plummeted from all time highs, are burning cash that should be returned to investors in a promise to fuel infinite AI growth.



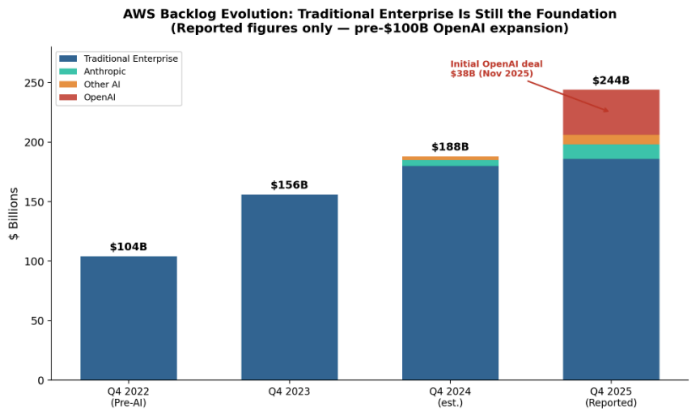


AWS Backlog

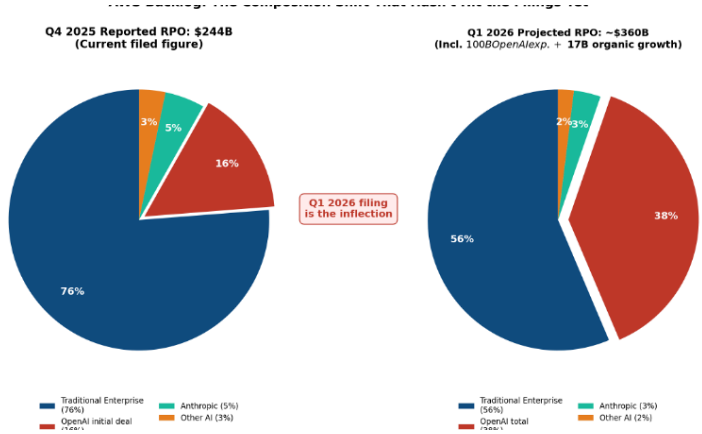
AWS reported $244 billion in remaining performance obligations as of Q4 2025, up 40% YoY and 22% respectively. Compared to an annualized revenue run rate of approximately $142 billion, this represents roughly 1.7 years of forward revenue visibility. However, the $244 billion figure is before the announcement of another $100 billion in backlog coming from a mega OpenAI deal, announced in February 2026.

Amazon does not publicly break out AI versus non-AI backlog, so the composition analysis that follows is reconstructed from known deal announcements, pre-AI baseline growth rates, and public filings.

The initial OpenAI deal, worth $38 billion and announced in November 2025, is captured in the $244 billion figure, accounting for approximately 16% of reported RPO. Anthropic commitments at that point were roughly $12 billion (5%), with other AI customers contributing approximately $8 billion (3%) (estimated). Traditional enterprise therefore constitutes roughly 76% of the reported $244 billion, the vast majority of what is actually on the books.



The next reported RPO, likely disclosed with Q1 2026 earnings, will include the $100 billion OpenAI expansion. But it will not simply be OpenAI added to a static base — traditional enterprise signings continue in the background. AWS was adding roughly $15–20 billion in net new RPO per quarter organically, based on the trajectory from $156 billion to $244 billion over eight quarters minus known AI deal additions. The Q1 2026 reported RPO is therefore likely closer to $355–365 billion, not the naive $344 billion that simple addition would suggest.

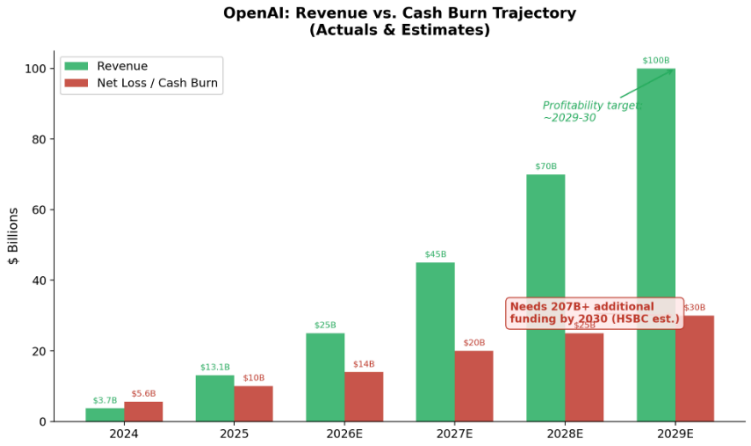


With organic growth factored in, the Q1 2026 composition should be broken down with traditional enterprise respresenting $200–205 billion (approximately 56%), OpenAI at $138 billion (approximately 38%), and Anthropic plus other AI customers at roughly $20 billion (6%). The projected backlog looks heavily skewed toward OpenAI in part because the expansion is a single, publicly disclosed mega-deal, while traditional enterprise growth comes from hundreds of undisclosed renewals and new signings that are never individually announced.

AI vulnerability

OpenAI's annualized revenue was approximately $25 billion as of early 2026, up from $13 billion in FY2025. Despite rapid revenue growth, projected net losses are roughly $14 billion in 2026, with cumulative cash burn estimated at $218 billion through 2029. The company raised $110 billion in February 2026 at an $840 billion valuation, but HSBC estimates they need at least $207 billion more by 2030 just to sustain operations. Profitability is not expected until

2029–2030 at the earliest, meaning OpenAI's cloud commitments are effectively backed by future fundraising, not current cash flows.



The risk is not that OpenAI disappears, it's that they renegotiate. Cloud contracts, despite take-or-pay provisions, have historically been renegotiable when the customer has leverage. OpenAI, representing such an important part of their backlog and burning $25 billion per year in cash, has considerable leverage to demand better terms or reduced minimums.

Anthropic's annualized revenue reached approximately $19 billion by March 2026, roughly doubling from $9 billion at the end of 2025. Cash burn of roughly $3 billion in 2025 is materially lower relative to revenue than OpenAI's, with break-even projected for 2027–2028. They're better capitalized relative to commitments, with an earlier profitability horizon than OpenAI and should have a higher probability of converting their commitments to actual expenditures.

Backlog conversion probability

On the reported $244 billion, conversion probability is high because the backlog is overwhelmingly traditional enterprise. The $186 billion traditional enterprise segment, representing 76% of reported RPO, has an almost 100% chance of converting to revenue on the estimated timeline, as these are established businesses running data, not startups funding uncertain growth. According to 10-K, companies that signed contracts with AWS were "Visa, the NBA, BlackRock, Perplexity, Lyft, United Airlines, DoorDash, Salesforce, U.S. Air Force, Adobe, Thomson Reuters, AT&T, S&P Global, National Bank of Canada, London Stock Exchange Group, Choice Hotels, Accenture, Indeed, HSBC, CrowdStrike, and more".

The initial $38 billion OpenAI tranche, at 16% of the backlog, carries a 60–75% conversion probability as it is smaller, more near-term, and more likely to be honored than the expansion. Anthropic at $12 billion (5%) carries 70–85% conversion probability given that Anthropic has been more conservative on their commitments and given their strong growth over the last few months, and other AI customers at $8 billion (3%) carry 70–80%.

Blended across all segments, approximately 86–91% of the reported $244 billion is likely to convert to actual revenue, or roughly $210–222 billion. This is a solid backlog on filed numbers.

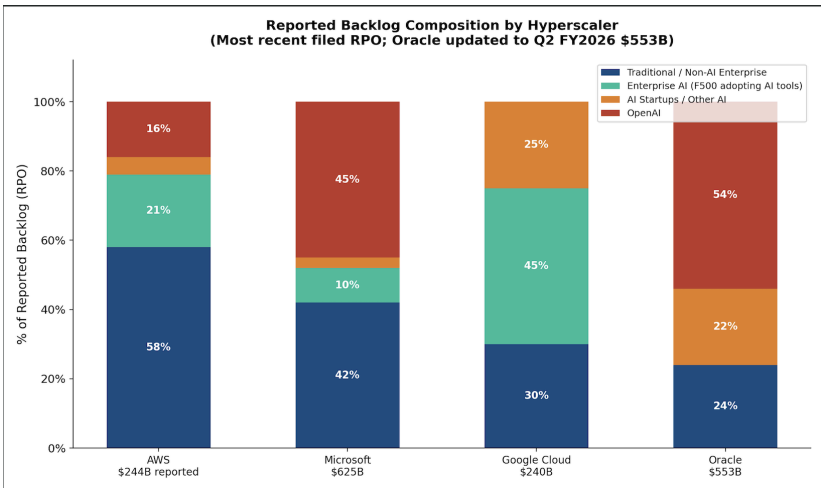
The conversion risk concentrates heavily in future RPO reports once the $100 billion OpenAI expansion is on the books, at which point the blended realizable rate drops to an estimated 72–85%.

Comparison with other hyperscalers

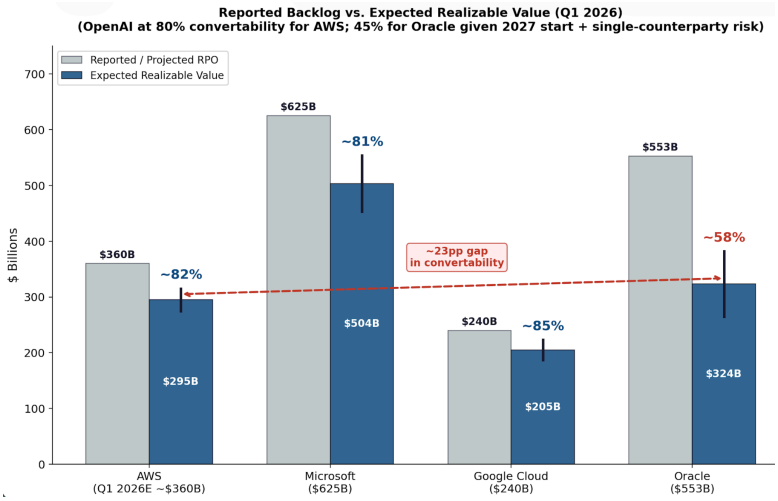
Microsoft's backlog is the largest of the four hyperscalers at approximately $625 billion. OpenAI exposure is roughly $281 billion, or 45% of the backlog, the largest single-customer AI commitment in absolute terms. However, the non-AI enterprise base is also the largest of any hyperscaler, at $250–280 billion anchored by M365, Dynamics, LinkedIn, and core Azure workloads. Microsoft generates over $60 billion in annual free cash flow and operates the stickiest enterprise software stack in the world (try convincing any bank stop using Excel or any consulting firm to stop using Powerpoint). Excluding OpenAI, Microsoft's backlog was growing 28% organically. "Fragile AI" exposure, OpenAI plus AI startups should be running at approximately 48%.

Google Cloud's backlog of $240 billion reached that level after growing 55% YoY. While approximately 70% of the backlog is categorized as "AI-linked," the composition is fundamentally different from Oracle or even AWS. The bulk consists of enterprise customers purchasing Gemini seats (over 8 million sold in four months), Fortune 500 companies that can pay their bills and high volume smaller contracts. Google Workspace, sold to everyday users, or other google suite applications, adds a durable recurring base of roughly $15–20 billion in annual recurring revenue across more than 8 million paying business customers.

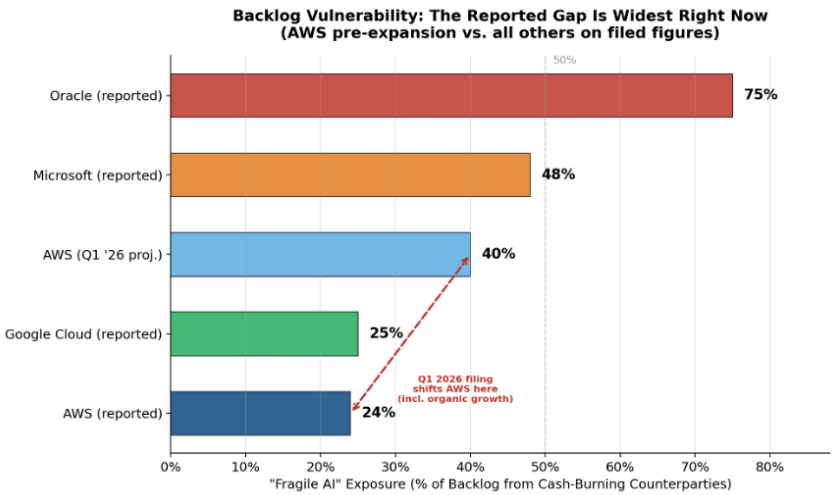
"Fragile AI" exposure should be only approximately 25% of the backlog as their billion-dollar are overwhelmingly enterprise commitments, not single customers startup promises. Alphabet's balance sheet, with over $100 billion in cash and $80 billion in free cash flow, provides near-total insulation from counterparty risk, something that Amazon, apart from their Capex, also has. This is generally a very safe backlog but also doesn't capture much of the leaders of AI's (OpenAI and Anthropic) growth.



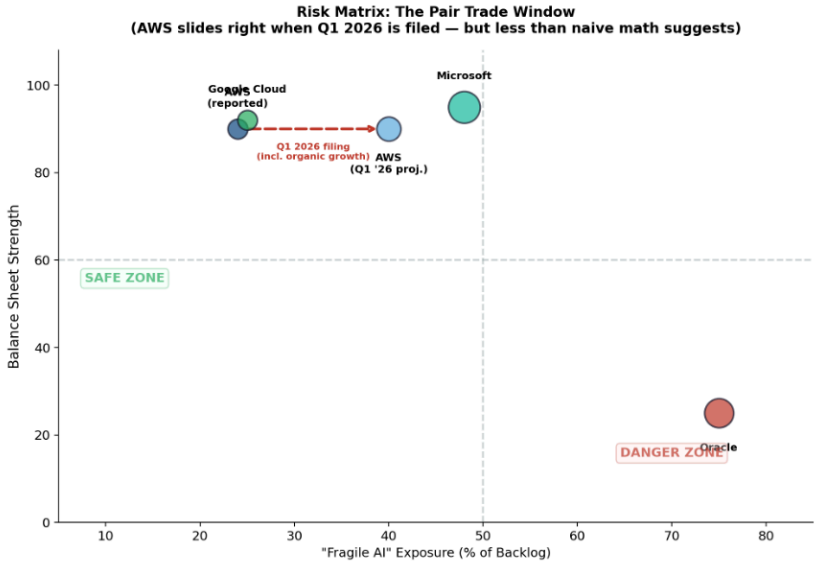
Oracle's backlog reached approximately $553B by Q2 FY2026, up 300%+ year-over-year driven almost entirely by a single counterparty where roughly $300B, or around 54% of the total, comes from OpenAI via a five-year deal beginning in 2027. Another $100-120B sits in other AI infrastructure deals (xAI, GPU cloud customers, other AI firms in the Stargate Project), pushing total AI-linked exposure to roughly 75-80% of reported RPO. The non-AI base of $120-140B in legacy database, ERP, and traditional cloud commitments is the thinnest among the four hyperscalers relative to total backlog. The revenue recognition timeline compounds the risk, and since the OpenAI deal doesn't begin until 2027, giving it some of the longest conversion lag among hyperscaler backlogs, this means Oracle is carrying the capex burden now against revenue that may or may not materialize two years from now.



Unlike the other three hyperscalers, Oracle itself is financially stressed by the AI buildout. Free cash flow has turned negative at roughly negative $10 billion, total debt stands at approximately $124 billion, compared to a $400 billion market cap and Barclays downgraded Oracle’s debt to underweight, warning of a potential slip to BBB-, one notch above junk status and, finally, S&P affirmed BBB but moved the outlook to negative. Oracle needs to raise $45–50 billion in new equity and debt in 2026 just to fund capital expenditures, meaning Oracle faces counterparty risk on both sides where its customer may not pay, and Oracle itself may struggle to finance their capex plans to sustain this demand which may never come.



On the reported $244 billion, AWS's backlog is remarkably resilient. With 76% in traditional enterprise, even a total OpenAI writedown would impair only 16% of the reported figure. Even post-expansion, the traditional enterprise base of approximately $200 billion is growing organically at 15–17% and is effectively recession-proof, underpinned by over 90% of the Fortune 100, the deepest service catalog in cloud computing, and the strongest government and regulated-industry presence. Compared to Oracle, AWS is making the same AI bet from a position of financial strength rather than financial desperation.



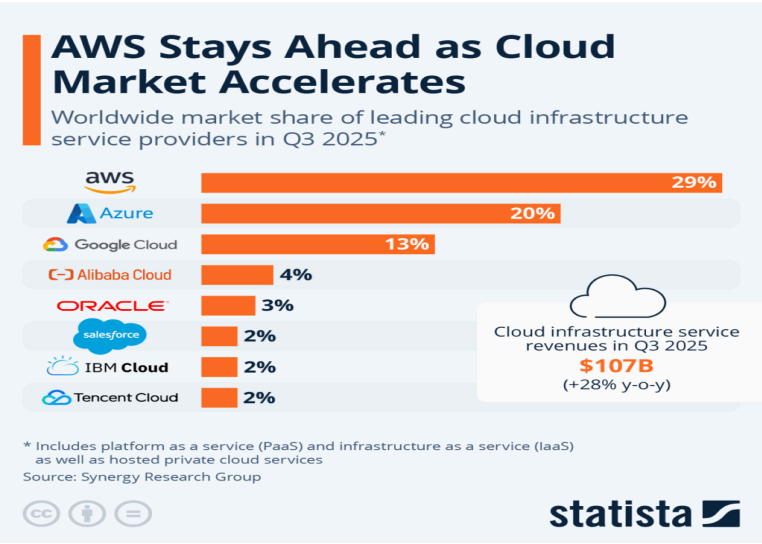
Conclusion on FCF pessimism:

The answer, as we broke down in the previous section, is the $244 billion backlog, which I estimate should grow to around $360B in Q1 given the OpenAI deal and other organic growth. Amazon is not investing speculatively into a technology they're not certain to have demand for,

they are building capacity against revenue that has already been contractually committed by enterprise, government, and hyperscaler customers. Yes, there is a risk that this backlog won't be transformed into actual revenue, but, as modeled and discussed before, I believe that around 80% of it will. Additionally, the 40% YoY growth in the backlog figure itself suggests demand is accelerating faster than the company can deploy infrastructure to fulfill it. Jassy was explicit on the Q4 call, saying that AWS could actually grow faster if supply were available. This is a capacity catch-up program, not a growth bet, and this distinction matters enormously for how you interpret the investment.

However, if revenue recognition is not the risk, margin and ROI on this investment is. AWS currently runs at around 35% operating margins and contributed almost 60% of Amazon's total operating income on just 18% of revenue. The question is whether the incremental $200B earns returns anywhere close to the existing asset base, or whether it gets deployed into capacity that sits underutilized for 2–3 years before filling up. If this is the case, the bears might be right in saying that Amazon is burning cash in order to collect money from it many years down the line. However, Jassy's claim that AI capacity is being monetized "as quickly as it is installed" is critical and, If true, the payback period on the capex is much shorter than bears assume. If demand already outpaces supply, the backlog contracts could turn into revenue as soon as the new data centers are ready and the payback on the capex can start to happen instantly.

Additionally, I think it is worth doing a sanity check in the cloud market. The combined hyperscaler capex is projected at $690B in 2026, coming mostly from Amazon, Oracle, Google and Microsoft. Amazon's $200B is roughly 29% of that total, which compares well with AWS's market share of 29% of the cloud infrastructure and its share of the AI data center market, which they're keeping pace given the OpenAI and Anthropic mega deals.



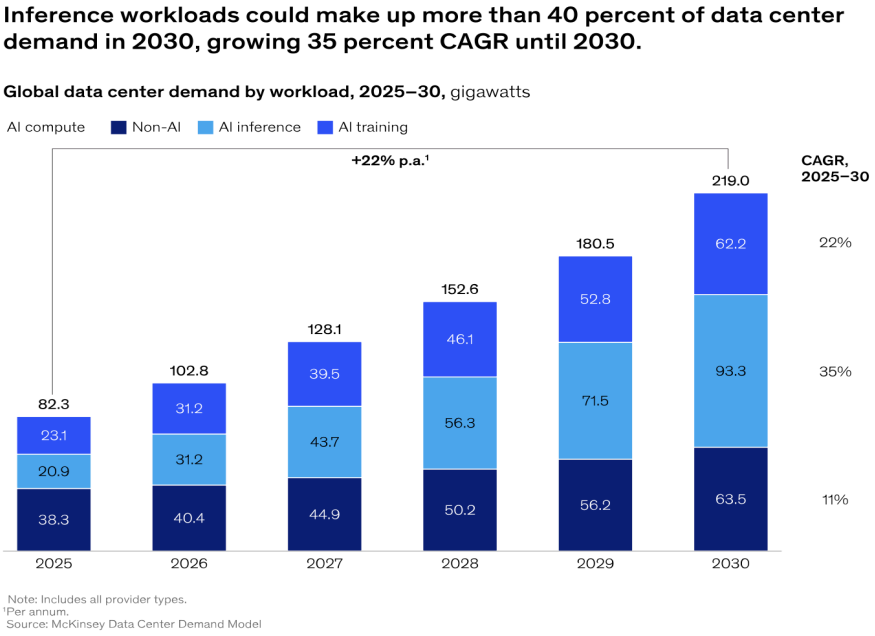
Unlike Oracle or OpenAI, Amazon isn't debt-financing its AI buildout, in fact they have only $30 billion dollars of long term debt, it's funding from operating cash flow. Amazon generated around $115B in operating cash flow in the trailing twelve months so a $200B capex year means FCF goes significantly negative for the short term, but their lack of debt means that they have less

downside risk in the case of an AI race slowdown. The last thing a company would want to do is finance all of its data center buildout with debt, keeping their free cash flow relatively stable, and then having to refinance hundreds of billions worth of loans if a large AI customer fails to go through with their AI compute demands.

Ultimately, as I explained before, most of this capacity will be converted into revenue, with maybe a 10-25% gap as we stretch contracts further down the line and as funding conditions for AI-based companies shift. No doubt, the company's current FCF situation would've been better if their capex wasn't as high but the capacity it will build will allow for Amazon to generate sustainable revenue for the foreseeable future, capitalizing on existing demand/backlog for a trend that is well underway.

Looking beyond the backlog

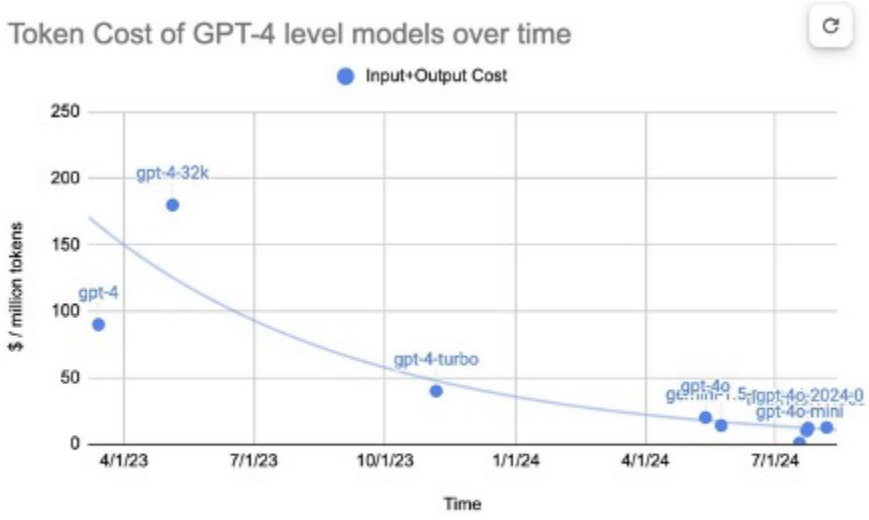
AI Training, which is most of AWS's backlog, is the process of building an AI model, feeding it enormous amounts of data repeatedly until it learns patterns, relationships, and how to respond to questions. Training happens once per model and is extraordinarily expensive. Inference happens billions of times per day across every application that uses AI and is cheap per query but enormous in aggregate volume, basically a person asking an AI model a question. The shift from a world dominated by training spend to one dominated by inference spend is essentially the shift from the construction phase of AI to the operational phase.



Right now, the ratio is roughly 80% training / 20% inference for AI-specific data center capacity but that's already beginning to flip and McKinsey projects that by 2030–2032, the ratio inverts completely to roughly 20% training / 80% inference. Training demand is projected at a 22%

CAGR through 2030 and inference demand grows at 35% CAGR. This presents an interesting and changing landscape for AWS to be in.

There are essentially three ways a company can pay for AI inference. The simplest is pay-per-token, OpenAI calls Amazon's API, uses the model, and pays for exactly what you consume with no commitment, which works well for smaller workloads but is expensive at scale and leaves AWS absorbing the cost of idle hardware between bursts of traffic. The more economically sensible option for serious enterprise deployments is reserved capacity, where a company commits to a guaranteed spend floor in exchange for lower rates, this is essentially what the $244 billion backlog represents, customers locking in future consumption at contracted prices, giving AWS predictable utilization and the customer predictable costs. The third option is renting raw GPU infrastructure and running your own inference stack, which only makes economic sense if you have enough consistent traffic to keep the hardware busy. That dynamic naturally concentrates inference economics in the hands of the largest cloud providers, and AWS, with the broadest enterprise customer base of any hyperscaler, is structurally the best positioned to capture it.



Cost for 2 million tokens (input+output) decreased from $180->$0.75 in 2 years. 240x cheaper

By the early 2030s, the compute purchasing landscape will likely look something like this: a small number of frontier labs (OpenAI, Anthropic, DeepMind, maybe 2–3 others) run massive but infrequent training runs, probably on specialized reserved clusters. The biggest open question is whether inference becomes so cheap and commoditized that it compresses margins for everyone (the "compute becomes electricity" thesis), or whether the sheer volume of agentic demand keeps total spend growing fast enough that providers maintain healthy returns even on lower unit prices. The early data of costs falling 80%+ per year while total enterprise spend increases suggests the volume effect is winning for now.

I believe that this is an incredible opportunity for Amazon to grow more than their competitors, first because of their powerful backlog and secondly because of their Traininum infrastructure. Training requires the absolute best silicon, massive parallelism, tightly synchronized GPU clusters, high energy clusters and this is Nvidia's fortress. But inference workloads are fundamentally different since they are smaller independent computations, lower interdependence between chips, smaller energy requirements, and the key target is cost-per-token rather than raw peak performance.

Amazon claims Trainium delivers 50% lower cost for inference compared to H100-class GPUs, and some benchmarks suggest it's as cheap as a quarter of Nvidia pricing. In an inference-dominant world, customers stop asking "which chip is the most powerful?" and start asking "which chip gives me the lowest cost per million tokens at acceptable latency?" That's exactly the value proposition Trainium and Inferentia are built for.

The migration friction from different coding systems, which will already be solved by Trainium 4, is also lower for inference than training since inference runs standardized models repeatedly with predictable memory access patterns, making the optimization work a one-time cost that compounds with scale. Trainium is already at a multi-billion dollar run rate growing triple digits. If inference becomes 70% of AI data center demand by 2030, Amazon's custom silicon strategy can be a core competitive moat.

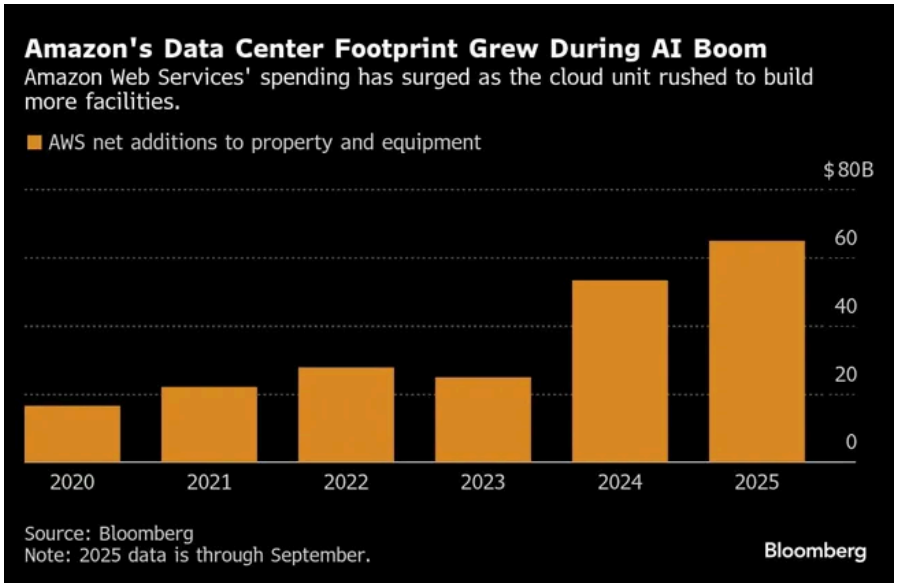
In regards to their backlog and the ability of it being actually realized into revenue, in a world where inference costs drop 90% by 2030 (Gartner's actual forecast), the unit price of compute collapses, but total spend may not. Per-token costs are already falling 80%+ YoY, yet enterprise AI spending is increasing because the volume of inference is exploding. Additionally, companies are going from running a few chatbots to thousands of agentic workflows running 24/7 so the price of a token drops, but the number of tokens consumed grows even faster.

For AWS's backlog specifically, there are a few protective layers, the biggest one is that their large cloud contracts are structured as committed spend floors, not fixed-price-per-unit deals, meaning that even if prices drop they will still can secure most of this revenue promise. If inference gets cheaper per token, the customer still spends the committed amount so they just consume more compute for the same money, which is actually good for AWS utilization metrics and total demand pressure. The risk would be if a customer like OpenAI tried to renegotiate the total commitment down, claiming they need less capacity, but OpenAI's compute appetite is growing, not shrinking and they're scaling to millions of enterprise users and agentic workloads. The more realistic scenario is they consume the committed amount faster and then negotiate additional capacity.

Where backlog could genuinely erode is if a major customer goes bankrupt, pivots away from AI, or gets acquired. That's a credit risk question, not a pricing question. And AWS quietly raised GPU Capacity Block prices by 15% in January 2026, suggesting they still have pricing power despite the cost declines.

Even the bear case where Amazon is spending $200B building massive training clusters, and in 5 years nobody needs training clusters because inference is all that matters is overstating the problem for a few reasons. First, the physical infrastructure, which is actually hard to build and an incredibly powerful asset, is largely workload-agnostic, meaning that a data center built for training can be relatively easily reconfigured for inference. The power density actually decreases significantly as data processing gets lighter, meaning you can fit more inference servers in the same facility and save money in electricity expenditure and cooling costs, even bumping up their total margins.

Second, inference needs to be close to users for latency reasons like metro-adjacent locations, regional deployments and edge facilities. AWS's massive global footprint, that arose from their headstart in capacity building, is a structural advantage in exactly this scenario. AWS operates the largest and most extensive cloud infrastructure footprint in the world, with recent estimates suggesting it manages over 900 data centers, far surpassing competitors. This sprawling network, roughly triple the size of Microsoft's known footprint gives them a very strong competitive advantage in inference.



Third, the chip layer is the most replaceable component and it could play into Amazon’s advantage as they try to expand their silicon capacity simultaneously as companies increase their inference demand. Servers have 3–5 year refresh cycles anyway so if the workload mix shifts to 70% inference by 2030, AWS swaps Nvidia training GPUs for Trainium/Inferentia inference chips during normal refresh cycles and companies will actually appreciate the discounted price given their smaller demand for training. They've been building this optionality deliberately and timing could play into their hands.

The real risk isn't that the capex is "wasted", it's that the returns take longer to materialize than the market expects. If inference becomes so cheap that total cloud spend grows slower than the capacity being built, you get a period of overcapacity and margin compression, though this is

not what we are seeing with current AI trends and expenditure programs. The difference between Amazon earning 25% ROIC on this capex versus 15% is enormous at $200B scale, but still, even if that happens, AWS should be better positioned than most of their competitors given their silicon exposure and cloud backlog composition.

This is both an opportunity and a risk which Amazon has the best risk-adjusted exposure to right now compared to their competitors. I will keep a close eye on this shift from training to inference AI in the next few months and years.

Critical Risks

The $200 billion capital expenditure commitment for 2026 is the central bear case for the AWS thesis, and it deserves honest treatment. If AI demand disappoints, either because enterprise adoption timelines slip or because DeepSeek-style efficiency breakthroughs reduce the compute intensity of production workloads, the return on this capital will be lower than management projects, and the depreciation wave that follows will weigh on reported earnings for years. The $244 billion backlog provides a contractual buffer, but committed contracts can be renegotiated or cancelled with penalties if a customer's circumstances change materially, and as we discussed in the backlog risk section, the composition includes counterparties whose ability to honor multi-year commitments is not beyond question.

That said, the risk is limited as the constraint Amazon faces today is supply, not demand and management has been explicit that AWS could grow faster if capacity were available, and the company is monetizing infrastructure as fast as it can be installed. The scenario where this investment destroys value requires not just a slowdown in new AI commitments but also a simultaneous deterioration of the existing backlog two things happening at once that would need to be quite severe to impair the return on a capital base this large.

The competitive concern is real and worth taking seriously because they started as the main player in the cloud market and now faces pressure from other top companies like Google and Microsoft. Azure grew 39% and Google Cloud grew 48% in Q4 2025 against AWS's 24%, and while the base effect explains much of the gap, it does not explain all of it. If enterprise buyers develop strong preferences for a specific model ecosystem and that ecosystem is natively tied to a competing cloud, AWS could face structural share pressure in the highest-value AI workload category regardless of its infrastructure advantages.

The variable to monitor is where new AI training commitments are flowing. If AWS is holding or gaining share in inference while ceding some ground in frontier training, the thesis remains intact. If it is losing share in both simultaneously, that is a different conversation.

Tariffs represent a genuine second-order risk to the retail and advertising businesses simultaneously since a significant portion of Amazon's third-party seller base sources goods from China, and escalating tariffs compress seller margins directly, when a seller's cost base rises 20-25% overnight, they face a choice between absorbing the hit or raising prices on Amazon. There is a chance demand could slow down with these price hikes and that would

slow down Amazon's revenue from take rate, but if firms choose to eat up most of the costs, which given the recent inflation data and historical background from 2017 tariffs, Amazon's take rate can still remain largely unchanged as they need to be in that marketplace.

The partial mitigant is that Amazon sits in an unusually powerful position relative to its seller base since, as the dominant marketplace, it effectively sets the terms of participation for millions of sellers who have no comparably scaled alternative, and that market power gives Amazon the ability to pass along cost pressures in ways that protect its own economics even when sellers are squeezed. It is not a clean protection, and tariff escalation would still weigh on total revenue and therefore on advertising revenue, but Amazon's structural position as the indispensable retail platform means it is better insulated from tariff pass-through than the sellers who depend on it.

Red Flags That Would Trigger Position Review

- Two consecutive quarters of sub-18% growth while Azure and GCP hold above 30%
- YoY growth falls below 15% for two consecutive quarters without macro explanation
- AWS operating margin falls below 28% on a trailing twelve-month basis as capacity comes online
- NVIDIA regains pricing power and Trainium adoption stalls below $15B annualized run rate in 2026

Sum of the parts exit multiple valuation:

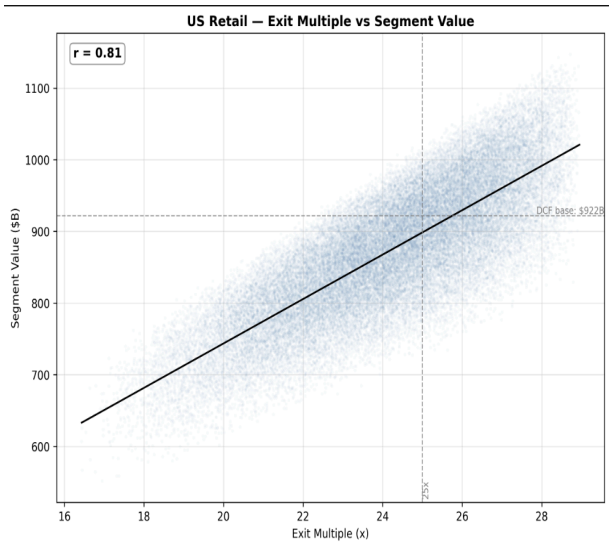
US Retail

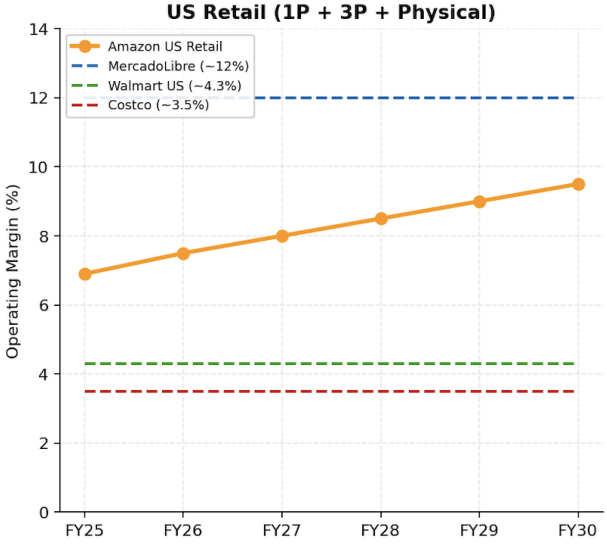
Revenue growth of 9% tapering to 7% reflects the law of large numbers applied to a business already doing $426 billion. The growth isn't coming from Amazon trying harder, it's coming from three structures that the company is already implementing very well: First, Rufus drives a 60% lift in purchase likelihood across 300 million customers, which means the same traffic converts at a higher rate without any incremental marketing spend. Second, same-day delivery expands geographically into markets that were previously next-day, which unlocks basket categories — groceries, consumables, pharmacy that require speed to capture. Not to mention this logistics infrastructure allows Amazon to increase their take rate as necessary and allows for the company to further monetize the marketplace as they see fit. The final mechanism is a very simple one, where Amazon's dominance is one that, now, they can afford to increase their take rates for merchants without causing churn. The 7% terminal is deliberately anchored to Walmart's long-run revenue growth, which is the right discipline for a business at this scale.

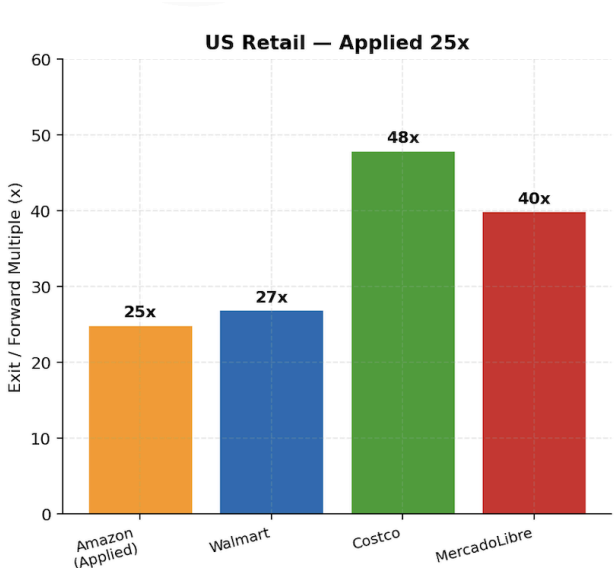
The margin expansion from 6.9% to 9.5% is entirely a mix-shift story. Every dollar of GMV that migrates from 1P to 3P flows through to Amazon as take-rate revenue while Amazon bears none of the inventory risk.The take rate functions as a toll on e-commerce that Amazon can raise 2–3 points above inflation annually without losing sellers, because leaving Amazon is

commercially unviable for most businesses. A seller generating the majority of their revenue through the platform cannot afford to leave even if the toll goes up, and Amazon is simultaneously making the platform more valuable through Rufus and logistics improvements, which means sellers are absorbing fee increases while making more money. Another boost to margins is that now, with their current userbase already at terminal growth levels, the company won't need to spend as much money subsidizing purchases or offering low cost products on their marketplace to attract users. This shift away from low margin 1P, which had the focus of getting more users in the early years, will drive margins up. The 9.5% terminal sits below MercadoLibre's 10–13% and well below pure marketplace economics, which is appropriate because Amazon carries 1P and fulfillment drag that pure platforms don't but also has a structurally better business model compared to retailers like Walmart or Costco.

The 25x exit multiple anchors to Walmart's 25–28x FCF trading range with a modest discount for Amazon's heavier fulfillment investment cycle. It is meaningfully below Costco's 48–52x because Amazon does not have Costco's membership-dollar density in isolation, and above Target's 18–20x because Amazon has a structurally superior moat, better ad integration, and a far more favorable mix trajectory. The Prime membership does generate a strong market value but that will be discussed and expressed, in terms of valuation, in the Prime section of this DCF.







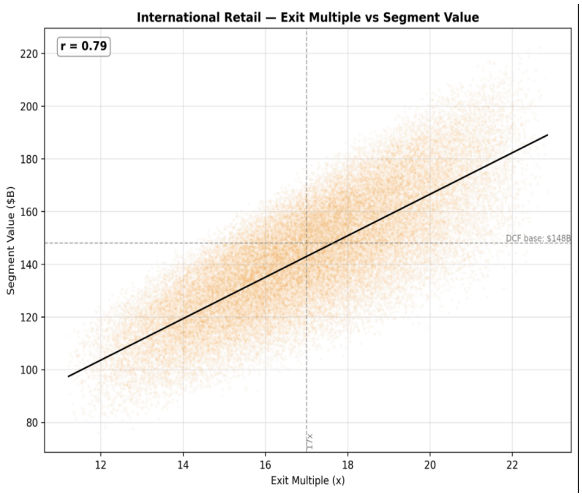
International Retail

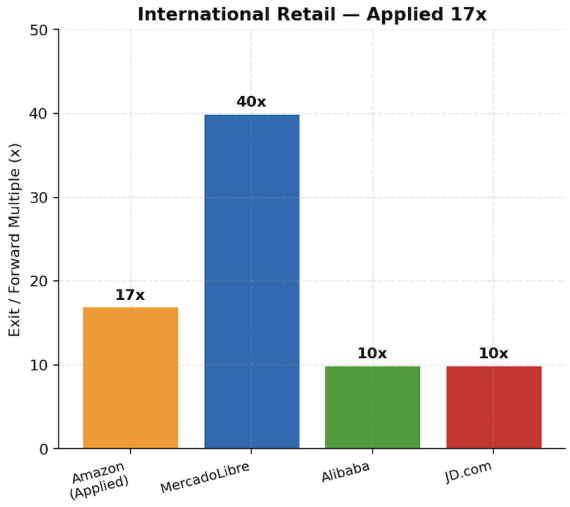
The 13% starting growth rate reflects the thesis that international is best understood as an earlier-vintage North America. The playbook of investing in logistics, subsidizing pricing to build volume, earn loyalty through Prime, then monetize with 3P fees and advertising is now proven across enough markets that international is essentially replicating the domestic flywheel in geographies where e-commerce penetration is still well below US levels. India, Brazil, Mexico, and the Middle East are still in the volume-building phase, which supports above-domestic growth rates even as the UK, Germany, and Japan grow in the mid-single digits. The blend of fast-growing emerging markets and slow-growing mature ones produces 11–13% at the portfolio level, decelerating naturally as the emerging markets scale and the mature markets grow as a larger share of the base.

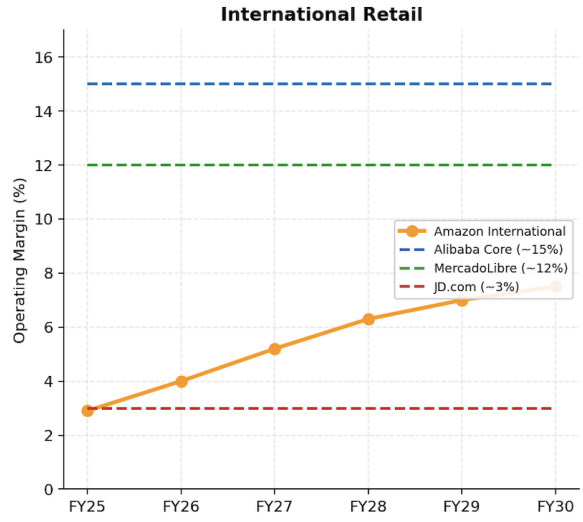
The margin expansion from 2.9% to 7.5% is the natural consequence of the flywheel maturing in markets that have already absorbed the upfront investment. FY2025 was international's first full year of structural profitability, and the trajectory from here follows the same pattern that drove North America from similar starting margins to where it is today: 1P builds the volume, 3P share increases as seller density reaches critical mass, and advertising layers on top as a near-zero-marginal-cost revenue stream. The 7.5% terminal deliberately does not converge to the North America margin because international has more tail markets still in investment mode, more currency and regulatory friction, and a less mature advertising business. As long as their competitive moat abroad is not as strong as it is in the US (which I don’t think they will ever be able to match given each country’s domestic competition) their margins will sit lower.

The 17x exit multiple sits at a deliberate discount to US Retail's 25x, reflecting execution risk across fragmented markets, entrenched local competitors and a 3P mix still maturing. The comp range spans MercadoLibre at 45–55x, which has strong fintech optionality with MercadoPago, down to JD.com at 15–18x, which carries lower margins and the China risk. Amazon International structurally sits between those two, closer to MELI on business model but closer to

JD on geographic complexity, and 17x reflects that positioning while putting a governor on how much value the segment can generate even under optimistic assumptions.





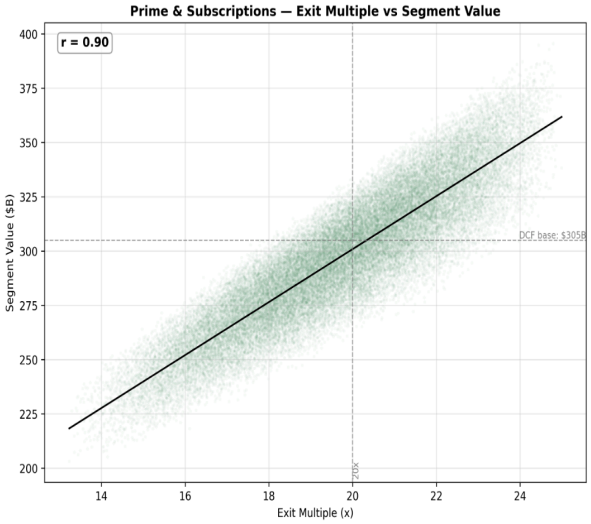


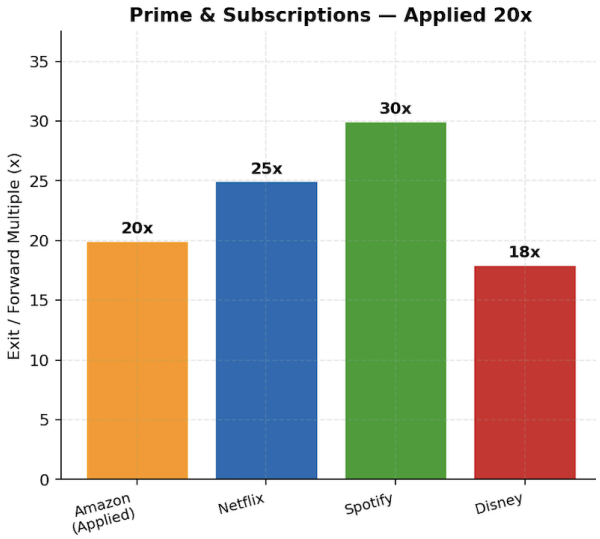
Prime & Subscriptions

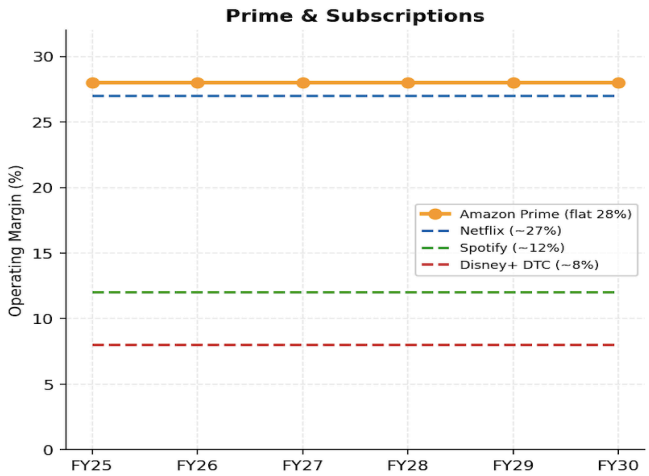
Revenue growth of 10% declining to 8% reflects the tension between two offsetting forces. US penetration is already at roughly 80% of American households, which means the marginal new member is harder to acquire and organic membership growth is largely saturated domestically. But pricing power is real and demonstrably exercised, where Amazon has raised Prime fees twice in recent years without meaningful churn, which is the signal that matters for a subscription business. The 8–10% growth is therefore predominantly a pricing and international expansion story rather than a domestic membership story, and it is the most predictable revenue line in the entire SOTP because the churn dynamics of a bundle this deep are structurally different from any single-service subscription.

The flat 28% margin is the most deliberate assumption in the model. Prime is a near-zero-marginal-cost business, every incremental member costs Amazon almost nothing to serve because the logistics network, video content, and music library are all sunk costs amortized across the existing base. That makes it structurally analogous to Netflix at 27% operating margins rather than to any retail business. There is no realistic path to compression and there is no need to model expansion because 28% already reflects the steady-state economics of a business where the incremental cost of adding a member is essentially zero.

The 20x exit multiple is below Netflix's 38–42x, Spotify's 50–60x, and Disney+'s 25–30x on a pure-play basis. The discount is because leaving Prime means losing free shipping, video, music, and grocery simultaneously rather than just one service, meaning that a lot of its value comes from the connection to retail and other perks that can't be disconnected from the broader business. 20x reflects the strength in their streaming platform and the connections to perks in the retail segment but reflects the downside of separating the businesses.





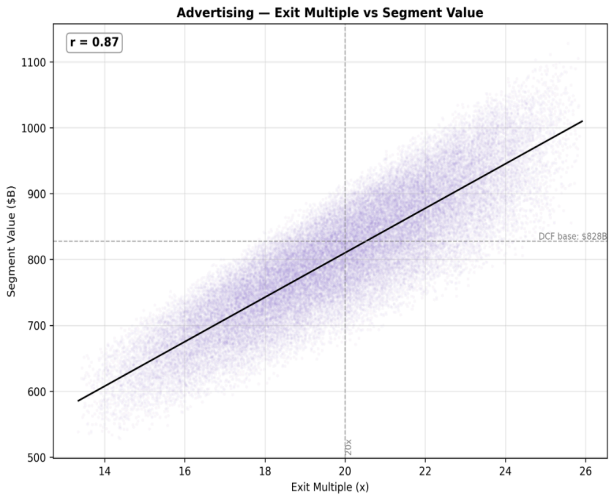


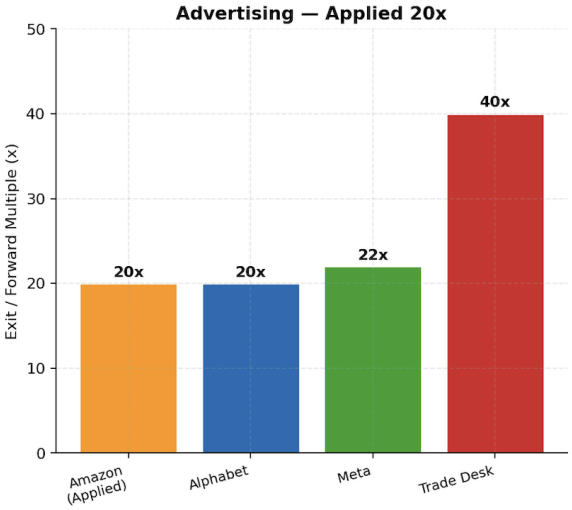
Advertising

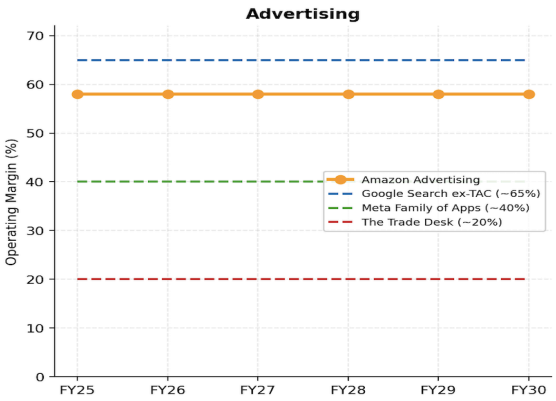
Revenue growth starting at 20% and decelerating to 12% is faster than the overall digital ad market because Amazon is in their growth stage and they are in a prime position to be a major player in the industry given its immense data on consumer behavior and the likelihood of purchase. That intent premium drives higher conversion rates, which justify higher auction CPCs, which attract more advertising spend, which makes the auction more competitive in a self-reinforcing cycle.This will grow to a point where not advertising on Amazon is economically equivalent to not appearing in Google search results for your category, and that non-optional nature of the spend underpins growth that is structurally above the digital ad market. The addition of Prime Video advertising, with a 200 million monthly viewer base that has barely begun to be monetized, adds a second growth vector that does not yet show up in the current run rate and extends the duration of above-market growth.

The flat 58% operating margin reflects the structural reality that the marginal cost of serving an additional ad impression on an already-built platform is essentially zero. The infrastructure is shared with retail and AWS so yes, the margin figures are artificially higher compared to Meta or Google. Holding the margin flat at 58% rather than allowing it to drift toward Google Search reflects this artificiality and prices in a weaker pricing power compared to their competitors. This pricing power dynamics however can easily change in the next few years as Amazon Ads grows and it could be a source of additional FCF for the company in the long term, but for now I believe it is better to be conservative.

The 20x exit multiple is at the low end of the 22–25x range where Google and Meta trade on their advertising segments, and a fraction of The Trade Desk's 55–65x. All of Amazon Ads' competitive advantage and pricing power comes from their retail strength, this segment, similar to Prime, is not strong enough to command higher multiples by itself. The 20x describes the strength of their margins and revenue base but also doesn't double count value added from retail and advertisement.







AWS

Revenue growth, as explained before, is very much backed by current backlog and backlog growth. The $244 billion in remaining performance obligations, not considering the $100B OpenAI deal and the other organic growth, growing 40% year-over-year, represents revenue that has already been contractually committed by enterprise, government, and hyperscaler customers.

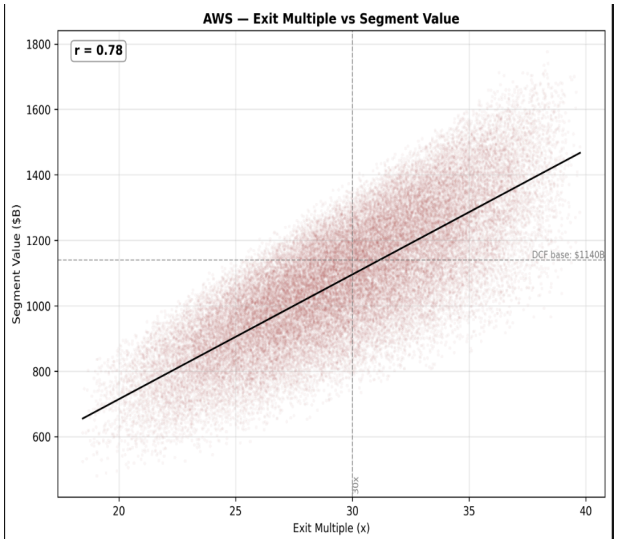
The model converts that backlog into revenue through four buckets: traditional enterprise at 95% conversion (Visa, BlackRock, Salesforce, and the Fortune 100 run production systems on AWS and are not going anywhere), OpenAI at 80% declining to 65% as later tranches depend on uncertain future fundraising, Anthropic at 75–80% given earlier profitability and stronger capitalization relative to commitments, and other AI customers at 75–80% for Bedrock-era enterprise workloads. The deceleration to 12% by the last year reflects the law of large numbers applied to a business approaching $260 billion in revenue and reflects a period where most enterprises will already be on cloud, dampening growth, not a loss of competitive position.

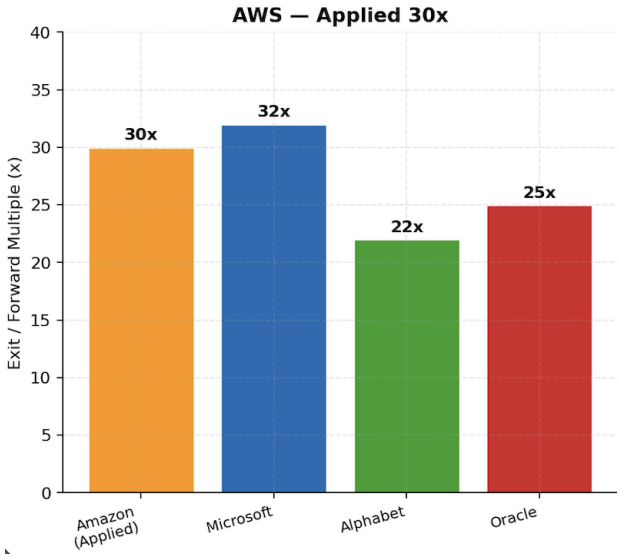
The margin expansion from 35.4% in FY2025 to 43% by FY2030 is driven by two compounding forces. The first is Trainium. When AWS runs workloads on Trainium instead of NVIDIA GPUs, the hardware margin that would otherwise accrue to NVIDIA flows directly into AWS operating income. The $10 billion Trainium run rate growing at triple-digit percentages means this substitution is already well underway, and as the training-to-inference ratio shifts toward inference by 2030, McKinsey projects 80% inference by 2030–2032, Trainium's cost advantage becomes more rather than less valuable, because inference workloads prioritize cost-per-token over raw peak performance, which is exactly the value proposition Trainium is built for. The second force is Graviton, which delivers 40% better price-performance than leading x86 processors and is already used by over 90% of AWS's top 1,000 customers, meaning the CPU cost advantage is pervasive across the workload base. Together these two silicon programs allow AWS to reduce its cost per unit of compute without changing customer pricing, which is pure margin expansion.

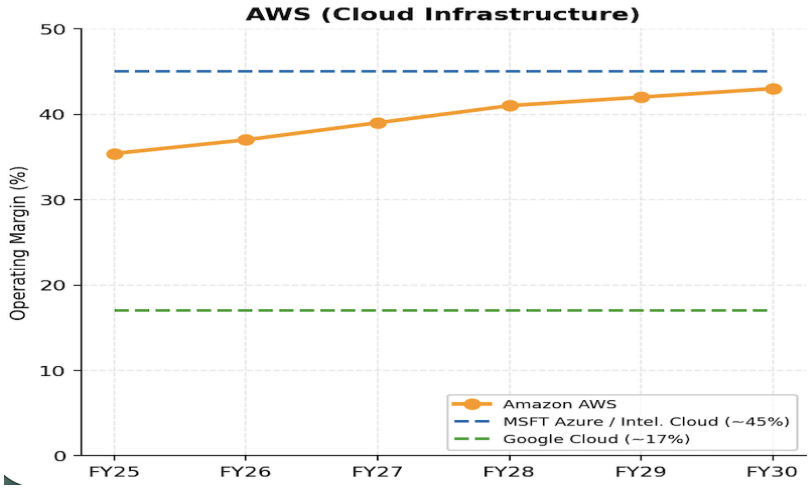
Capex, the big elephant in the room,of $150 billion in Year 1 declining to $80 billion by Year 5 reflects the conviction that peak capex is a 2026–2027 phenomenon driven by a specific AI

buildout window rather than a permanent new baseline. As Trainium's lower unit cost reduces the dollar capex required per unit of compute capacity, and as the data center construction wave crests, the capex intensity normalizes and the FCF profile transforms dramatically. This is the capex-hangover pattern that dominates the bear case, but it is exactly what you would expect from a business building ahead of a backlog it has already sold.

The 30x exit multiple reflects strong capex/buildout fears and the counterparty concentration risk in the OpenAI and Anthropic backlog, which could drag the segment's total value significantly. It is in line with Microsoft Azure's implied 32–35x in sum-of-parts analyses, a premium to Google Cloud's 22–25x reflecting AWS's scale, margin lead, custom silicon and stronger AI exposure, and well above Oracle's distressed positioning with a severely risky backlog of AI demand. This is a multiple that converges current real risks regarding FCF generation in the segment, opportunities from AI training/inference, their custom Trainium and Graviton chips and a calculated upside given their ability to convert backlog into revenue.







	FY2024A	FY2025A	FY2026E	FY2027E	FY2028E	FY2029E	FY2030E	
US Retail (1P + 3P + Physical Stores)								
Revenue ($B)	$388	$426	$465	$502	$542	$580	$621	
YoY Growth (%)		10,0%						
Assumed Growth Rate			9,0%	8,0%	8,0%	7,0%	7,0%	
Operating Margin (%)	6,5%	6,9%	7,5%	8,0%	8,5%	9,0%	9,5%	
Operating Income ($B)	$25,00	$29,60	$34,85	$40,15	$46,07	$52,19	$58,95	
International Retail								
Revenue ($B)	$144	$162	$183	$205	$229	$257	$285	
YoY Growth (%)		12,8%						
Assumed Growth Rate			13,0%	12,0%	12,0%	12,0%	11,0%	
Operating Margin (%)	2,7%	2,9%	3,5%	4,5%	5,5%	6,5%	7,5%	
Operating Income ($B)	$3,80	$4,70	$6,40	$9,22	$12,62	$16,71	$21,40	
Prime & Subscription Services								
Revenue ($B)	$44	$49	$54	$59	$64	$69	$75	
YoY Growth (%)		10,4%						
Assumed Growth Rate			10,0%	9,0%	9,0%	8,0%	8,0%	
Operating Margin (%)	25,0%	27,0%	28,0%	28,0%	28,0%	28,0%	28,0%	
Operating Income ($B)	$11,10	$13,23	$15,09	$16,45	$17,93	$19,37	$20,91	
Advertising Services								
Revenue ($B)	$56	$69	$82	$97	$113	$128	$144	
YoY Growth (%)		22,1%						
Assumed Growth Rate			20,0%	18,0%	16,0%	14,0%	12,0%	R
Operating Margin (%)	55,0%	58,0%	58,0%	58,0%	58,0%	58,0%	58,0%	
Operating Income ($B)	$30,91	$39,79	$47,75	$56,34	$65,35	$74,50	$83,44	
AWS — Non-AI (Traditional Enterprise Cloud)								
Revenue ($B)	$100	$115	$132	$151	$170	$191	$214	E
YoY Growth (%)								
Assumed Growth Rate			15,0%	14,0%	13,0%	12,0%	12,0%	C
Backlog Conversion Rate			95,0%	95,0%	95,0%	95,0%	95,0%	T
Backlog-Adjusted Revenue ($B)			$125,64	$143,23	$161,85	$181,27	$203,02	
AWS — AI (OpenAI, Anthropic, AI Startups)								
OpenAI Contracted Revenue ($B)			$17	$25	$29	$32	$35	$
OpenAI Conversion Rate			80,0%	70,0%	65,0%	65,0%	65,0%	D
OpenAI Realized Revenue ($B)			$14	$18	$19	$21	$23	
Anthropic Revenue ($B)			$3	$5	$6	$7	$7	R
Anthropic Conversion Rate			80,0%	80,0%	78,0%	75,0%	75,0%	B
Anthropic Realized Revenue ($B)			$2	$4	$4	$5	$5	
Other AI Revenue ($B)			$5	$7	$9	$11	$13	B
Other AI Conversion Rate			80,0%	76,0%	75,0%	75,0%	75,0%	
Other AI Realized Revenue ($B)			$4	$5	$7	$8	$10	
Total AWS AI Realized Revenue ($B)			$20	$26	$30	$34	$38	
Additional AWS growth (beyond current backlog)			$10	$13	$15	$17	$19	
Total AWS Revenue ($B)	$108	$129	$156	$183	$207	$232	$260	
AWS Operating Margin (%)	37,0%	35,4%	36,0%	38,0%	41,0%	42,0%	43,0%	M
AWS Operating Income ($B)	$39,80	$45,60	$56,03	$69,49	$84,79	$97,51	$111,65	

	FY2024A	FY2025A	Y1	Y2	Y3	Y4	Y5
Effective Tax Rate			13,0%	13,0%	13,0%	13,0%	13,0%
US Retail — FCF Bridge							
Operating Income ($B)			34,9	40,1	46,1	52,2	58,9
NOPAT ($B)			30,3	34,9	40,1	45,4	51,3
D&A as % of Revenue			4,5%	4,5%	4,4%	4,3%	4,2%
D&A ($B)			20,9	22,6	23,8	24,9	26,1
Capex as % of Revenue			7,0%	6,5%	6,0%	5,5%	5,0%
Capex ($B)			32,5	32,6	32,5	31,9	31,0
ΔWC as % of Revenue			-0,5%	-0,5%	-0,4%	-0,3%	-0,3%
ΔWC ($B)			-2,3	-2,5	-2,2	-1,7	-1,9
Free Cash Flow ($B)			**21,0**	**27,4**	**33,6**	**40,2**	**48,2**
FCF Margin (% of Revenue)			4,5%	5,5%	6,2%	6,9%	7,8%
International Retail — FCF Bridge							
Operating Income ($B)			6,4	9,2	12,6	16,7	21,4
NOPAT ($B)			5,6	8,0	11,0	14,5	18,6
D&A as % of Revenue			5,0%	5,0%	4,8%	4,6%	4,4%
D&A ($B)			9,1	10,2	11,0	11,8	12,6
Capex as % of Revenue			9,0%	8,5%	8,0%	7,0%	6,5%
Capex ($B)			16,5	17,4	18,4	18,0	18,5
ΔWC as % of Revenue			1,0%	0,8%	0,6%	0,5%	0,4%
ΔWC ($B)			1,8	1,6	1,4	1,3	1,1
Free Cash Flow ($B)			**-3,6**	**-0,8**	**2,3**	**7,1**	**11,5**
FCF Margin (% of Revenue)			-2,0%	-0,4%	1,0%	2,8%	4,0%
Prime & Subscriptions — FCF Bridge							
Operating Income ($B)			15,1	16,5	17,9	19,4	20,9
NOPAT ($B)			13,1	14,3	15,6	16,8	18,2
D&A as % of Revenue			2,0%	2,0%	2,0%	1,9%	1,8%
D&A ($B)			1,1	1,2	1,3	1,3	1,3
Capex as % of Revenue			2,5%	2,5%	2,4%	2,3%	2,2%
Capex ($B)			1,3	1,5	1,5	1,6	1,6
ΔWC as % of Revenue			-1,0%	-0,8%	-0,7%	-0,6%	-0,5%
ΔWC ($B)			-0,5	-0,5	-0,4	-0,4	-0,4
Free Cash Flow ($B)			**13,4**	**14,5**	**15,8**	**17,0**	**18,3**
FCF Margin (% of Revenue)			24,9%	24,7%	24,7%	24,6%	24,5%

Prime & Subscriptions — FCF Bridge							
Operating Income ($B)			15,1	16,5	17,9	19,4	20,9
NOPAT ($B)			13,1	14,3	15,6	16,8	18,2
D&A as % of Revenue			2,0%	2,0%	2,0%	1,9%	1,8%
D&A ($B)			1,1	1,2	1,3	1,3	1,3
Capex as % of Revenue			2,5%	2,5%	2,4%	2,3%	2,2%
Capex ($B)			1,3	1,5	1,5	1,6	1,6
ΔWC as % of Revenue			-1,0%	-0,8%	-0,7%	-0,6%	-0,5%
ΔWC ($B)			-0,5	-0,5	-0,4	-0,4	-0,4
Free Cash Flow ($B)			**13,4**	**14,5**	**15,8**	**17,0**	**18,3**
FCF Margin (% of Revenue)			24,9%	24,7%	24,7%	24,6%	24,5%
Advertising — FCF Bridge							
Operating Income ($B)			47,7	56,3	65,4	74,5	83,4
NOPAT ($B)			41,5	49,0	56,9	64,8	72,6
D&A as % of Revenue			1,5%	1,5%	1,4%	1,3%	1,2%
D&A ($B)			1,2	1,5	1,6	1,7	1,7
Capex as % of Revenue			20,0%	19,0%	18,0%	17,0%	16,0%
Capex ($B)			16,5	18,5	20,3	21,8	23,0
ΔWC as % of Revenue			0,5%	0,4%	0,3%	0,3%	0,2%
ΔWC ($B)			0,4	0,4	0,3	0,4	0,3
Free Cash Flow ($B)			**25,9**	**31,6**	**37,8**	**44,3**	**51,0**
FCF Margin (% of Revenue)			31,5%	32,6%	33,6%	34,5%	35,5%
AWS (Total) — FCF Bridge							
Operating Income ($B)			56,0	69,5	84,8	97,5	111,6
NOPAT ($B)			48,7	60,5	73,8	84,8	97,1
D&A as % of Revenue			14,0%	15,0%	15,5%	16,0%	16,0%
D&A ($B)			21,8	27,4	32,1	37,1	41,5
Capex as % of Revenue			96,4%	71,1%	48,4%	34,5%	30,8%
Capex ($B)			150,0	130,0	100,0	80,0	80,0
ΔWC as % of Revenue			1,0%	0,8%	0,7%	0,6%	0,5%
ΔWC ($B)			1,6	1,5	1,4	1,4	1,3
Free Cash Flow ($B)			**-81,0**	**-43,6**	**4,4**	**40,6**	**57,4**
FCF Margin (% of Revenue)			-52,1%	-23,8%	2,1%	17,5%	22,1%

US Retail (1P+3P+Physical)						
	Year 1	**Year 2**	**Year 3**	**Year 4**	**Year 5**	**Year 6**
FCF ($B)	21,0	27,4	33,6	40,2	48,2	52,0
Discount Factor	0,9070	0,8227	0,7462	0,6768	0,6139	0,6139
Present Value ($B)	19,1	22,5	25,1	27,2	29,6	
Sum Discounted CFs ($B)	**123,4**					
Terminal Value						
Year 6 FCF Growth Rate	8,0%					
Year 6 FCF ($B)	52,0					
Exit FCF Multiple	25x					
TV = Year 6 FCF × Multiple ($B)	1.301,0					
PV of Terminal Value ($B)	**798,7**					
Segment Value: US Retail (1P+3P+Physical) ($	**922,1**					
TV as % of Segment Value	86,6%					

International Retail						
	Year 1	**Year 2**	**Year 3**	**Year 4**	**Year 5**	**Year 6**
FCF ($B)	-3,6	-0,8	2,3	7,1	11,5	13,2
Discount Factor	0,9070	0,8227	0,7462	0,6768	0,6139	0,6139
Present Value ($B)	-3,2	-0,6	1,7	4,8	7,0	
Sum Discounted CFs ($B)	**9,6**					
Terminal Value						
Year 6 FCF Growth Rate	15,0%					
Year 6 FCF ($B)	13,2					
Exit FCF Multiple	17x					
TV = Year 6 FCF × Multiple ($B)	224,5					
PV of Terminal Value ($B)	**137,8**					
Segment Value: International Retail ($B)	**147,5**					
TV as % of Segment Value	93,5%					

Prime & Subscriptions						
	Year 1	**Year 2**	**Year 3**	**Year 4**	**Year 5**	**Year 6**
FCF ($B)	13,4	14,5	15,8	17,0	18,3	20,1
Discount Factor	0,9070	0,8227	0,7462	0,6768	0,6139	0,6139
Present Value ($B)	12,2	11,9	11,8	11,5	11,2	
Sum Discounted CFs ($B)	**58,6**					
Terminal Value						
Year 6 FCF Growth Rate	10,0%					
Year 6 FCF ($B)	20,1					
Exit FCF Multiple	20x					
TV = Year 6 FCF × Multiple ($B)	401,9					
PV of Terminal Value ($B)	**246,8**					
Segment Value: Prime & Subscriptions ($B)	**305,3**					
TV as % of Segment Value	80,8%					

Advertising						
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
FCF ($B)	25,9	31,6	37,8	44,3	51,0	56,1
Discount Factor	0,9070	0,8227	0,7462	0,6768	0,6139	0,6139
Present Value ($B)	23,5	26,0	28,2	30,0	31,3	
Sum Discounted CFs ($B)	**139,0**					
Terminal Value						
Year 6 FCF Growth Rate	10,0%					
Year 6 FCF ($B)	56,1					
Exit FCF Multiple	20x					
TV = Year 6 FCF × Multiple ($B)	1.122,4					
PV of Terminal Value ($B)	**689,0**					
Segment Value: Advertising ($B)	**828,0**					
TV as % of Segment Value	83,2%					

AWS — Total						
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
FCF ($B)	-81,0	-43,6	4,4	40,6	57,4	64,3
Discount Factor	0,9070	0,8227	0,7462	0,6768	0,6139	0,6139
Present Value ($B)	-73,5	-35,9	3,3	27,5	35,2	
Sum Discounted CFs ($B)	**-43,4**					
Terminal Value						
Year 6 FCF Growth Rate	12,0%					
Year 6 FCF ($B)	64,3					
Exit FCF Multiple	30x					
TV = Year 6 FCF × Multiple ($B)	1.927,9					
PV of Terminal Value ($B)	**1.183,6**					
Segment Value: AWS — Total ($B)	**1.140,2**					
TV as % of Segment Value	103,8%					

	Value ($B)	% of EV
US Retail (1P+3P+Physical)	922,1	27,6%
International Retail	147,5	4,4%
Prime & Subscriptions	305,3	9,1%
Advertising	828,0	24,8%
AWS — Total	1.140,2	34,1%
Total Enterprise Value ($B)	**3.343,1**	
Less: Net Debt ($B)	-30,0	
Equity Value ($B)	**3.313,1**	
Shares Outstanding (B)	10,7	
Implied Price per Share	**$309,63**	