

Date: October 30, 2025

Operator

Thank you for standing by. Good day, everyone, and welcome to the Amazon.com Third Quarter 2025 Financial Results Teleconference. [Operator Instructions] Today's call is being recorded. And for opening remarks, I will be turning the call over to the Vice President of Investor Relations, Mr. Dave Fildes. Thank you, sir. Please go ahead.

Dave Fildes

Hello, and welcome to our Q3 2025 financial results conference call. Joining us today to answer your questions is Andy Jassy, our CEO; and Brian Olsavsky, our CFO. As you listen to today's conference call, we encourage you to have our press release in front of you, which includes our financial results as well as metrics and commentary on the quarter. Please note, unless otherwise stated. All comparisons in this call will be against our results for the comparable period of 2024. Our comments and responses to your questions reflect management's views as of today, October 30, 2025 only, and will include forward-looking statements. Actual results may differ materially. Additional information about factors that could potentially impact our financial results is included in today's press release and our filings with the SEC, including our most recent annual report on Form 10-K and subsequent filings. During this call, we may discuss certain non-GAAP financial measures. In our press release, slides accompanying this webcast and our filings with the SEC, each of which is posted on our IR website. You will find additional disclosures regarding these non-GAAP measures, including reconciliations of these measures with comparable GAAP measures. Our guidance incorporates the order trends that we've seen to date and what we believe today to be appropriate assumptions. Our results are inherently unpredictable and may be materially affected by many factors, including fluctuations in foreign exchange rates, changes in global economic and geopolitical conditions, tariff and trade policies and customer demand and spending, including the impact of recessionary fears; inflation, interest rates, regional labor market constraints, world events, the rate of growth of the Internet, online commerce cloud services and new and emerging technologies and the various factors detailed in our filings with the SEC. Our guidance assumes, among other things, that we don't conclude any additional business acquisitions, restructurings or legal settlements. It's not possible to accurately predict demand for our goods and services, and therefore, our actual results could differ materially from our guidance. And now I'll turn the call over to Andy.

Andrew Jassy

Thanks, Dave. We saw strong growth across our business in Q3, and we're reporting \$180.2 billion in revenue, up 12% year-over-year, excluding the impact from foreign exchange rates. Operating income was \$17.4 billion, but would have been over \$21 billion, if not for 2 special Q3 expenses, \$2.5 billion for an FTC settlement and \$1.8 billion for estimated severance costs. Trailing 12-month free cash flow was \$14.8 billion. I'll start with AWS. AWS is growing at a pace we haven't seen since 2022, reaccelerating to 20.2% year-over-year, our largest growth rate in 11 quarters. It's worth remembering that year-over-year percentage growth is a relative term. It's very different having 20% year-over-year growth on a \$132 billion annualized run rate and to have a higher percentage growth rate on a meaningfully smaller annual revenue, which is the case with our competitors. Backlog grew to \$200 billion by Q3 quarter end and doesn't include several unannounced new deals in October, which together or more than our total deal volume for all of Q3. AWS is gaining momentum. Customers want to be running their core and AI workloads in AWS given its stronger functionality, security and operational performance and the scale I see in front of us gives me significant confidence in what lies ahead. I'll share a little more detail on why. It starts with AWS having much broader infrastructure functionality. Start-ups, enterprises and governments want to move their production workloads to the place that has the broadest and deepest array of capabilities. AWS has more services and deeper features within those services than anybody else and continues to innovate at a rapid clip. These are key building blocks for anything that customers want to create, and they're a big part of why Gartner has named AWS leader in its strategic cloud platform services Magic Quadrant for 15 consecutive years. We're bringing the same building block approach to AI. SageMaker makes it much simpler for companies to build and deploy their own foundation models. Bedrock gives customers leading selection of foundation models and superior price performance to deploy inference into their next-generation applications. A lot of the future value companies will get from AI will be in the form of agents. AWS is heavily investing in this area and well positioned to be a leader. Companies will both create their own agents and use agents from other companies. For those building their own, it's been harder to build than it should be. It's why we launched strands to make it much easier to create agents from any foundation model that builders desire. For companies who successfully built agents, they've hesitated putting them into production because they lack secure scalable runtime services or memory or observability built specifically for agents. It's why we launched AgentCore, a set of infrastructure building blocks that allow builders to deploy secure, scalable agents. Ericsson used AgentCore to deliver AI agents across their workforce, Sony used it to build a agentic AI platform with enterprise-level security, observability and scalability. And Cohere Health is using AgentCore to deploy agents that will reduce medical review times by up to 30% to 40%. AgentCore's SDK has already been downloaded over 1 million times, and our builders are excited about it. It's an enabler. Companies will also use other agents, and AWS continues to build many of the agents we believe builders will use in the future. For coding, we've recently opened up our agentic coding IDE called Kiro. More than 100,000 developers jumped into Kiro in just the first few days of preview and that number has more than doubled since. It's processed trillions of tokens thus far, weekly actives are growing fast, and developers love its unique spec and tool call and capabilities. For migration and transformation, we offer an agent called Transform. Year-to-date, customers have already used it to save 700,000 hours of manual effort. The equivalent of 335 developer years of work. For example, Thomson Reuters used it to transform 1.5 million lines of code per month, moving from Windows to open source alternatives and completing tasks a times faster than with other migration tools. Customers have also already used Transform to analyze nearly 1 billion lines of mainframe code as they move mainframe applications to the cloud. For business customers, we've recently launched QuickSleep to bring a consumer AI-like experience to work, making it easy to find insights, conduct deep research, automate tasks, visualize data and take actions. We've already seen users churn months long projects in today's get 80% plus time savings on complex tasks and realize 90% plus cost savings. And for contact centers, we offer Amazon Connect which creates a more personalized and efficient experience for contact center agents, managers and their customers. Connect has recently crested \$1 billion annualized revenue run rate with 12 billion minutes of customer interactions being handled by AI in the last year and is being used by large enterprises like Capital One, Toyota, American Airlines and Ryanair. These are real practical results for customers, and there are many more examples like them. Because of its advantaged capabilities, security, operational performance and customer focus, AWS continues to earn most of the big enterprise and government transformations to the cloud. As a result, AWS is where the preponderance of company's data and workloads reside and part of why most companies want to run AI and AWS. To enable customers to do so, we need to have the requisite capacity, and we've been focused on accelerating capacity the last several months, adding more than 3.8 gigawatts of power in the past 12 months, more than any other cloud provider. To put that into perspective, we're now double the power capacity that AWS was in 2022, and we're on track to double again by 2027. In the last quarter of this year alone, we expect to add at least another 1 gigawatt of power. This capacity consists of power, data center and chips, primarily our custom silicon, Trainium and NVIDIA. We've recently brought Project

Rainier Online, our massive AI compute cluster spanning multiple U.S. data centers and containing nearly 500,000 of our Trainium2 chips. Anthropic is using it now to build and deploy its industry-leading AI model, Claude, which we expect to be on more than 1 million Trainium2 chips by year-end. Trainium2 continues to see strong adoption, is fully subscribed and is now a multibillion-dollar business that grew 150% quarter-over-quarter. Today, Trainium is being used by a small number of very large customers but we expect to accommodate more customers starting with Trainium3. We're building Bedrock to be the biggest inference engine in the world and in the long run, believe Bedrock could be as big a business for AWS as EC2, and the majority of token usage in Amazon Bedrock is already running on Trainium. We're also continuing to work closely with chip partners like NVIDIA, with whom we continue to order very significant amounts as well as with AMD and Intel. These are very important partners with whom we expect to keep growing our relationships over time. You're going to see us continue to be very aggressive in investing in capacity because we see the demand. As fast as we're adding capacity right now, we're monetizing it. It's still quite early and represents an unusual opportunity for customers in AWS. I'll now turn to stores. Where the team continues to deliver and innovate for customers across our key priorities, selection, low prices and convenience, particularly fast delivery, we're offering 14% more selection since last quarter from popular brands like The North Face and Charlotte Tilbury, and we've added hundreds of thousands of items from popular brands this year. Everyday Essentials continues to grow quickly, and year-to-date is growing nearly twice as fast as the rest of the business. We continue to make it easier for customers to order low-priced perishable groceries from Amazon, and customers in more than 1,000 cities and towns now can shop fresh groceries alongside millions of Amazon.com products with free same-day delivery. This is a game changer for customers who can now order milk alongside electronics, check out with one cart and have everything delivered to their doorstep within hours. The team also invented a new add to delivery button that lets customers add items to previously scheduled orders and it's been used more than 80 million times since launch, and it's just launch. It's an example of one of those seemingly simple but powerful innovations that make customers' lives easier. We remain committed to staying sharp on price and meeting or beating prices of other major retailers. In July, we had our biggest Prime Day event ever, with customers saving billions of dollars across more than 35 categories. We continue to break records on speed. We're on track to deliver at our fastest speeds ever for Prime members globally once again this year, and we've started rolling out 3-hour delivery in select U.S. cities. We're also continuing to invest in infrastructure to speed up rural deliveries and serve more customers in more communities. That includes committing over \$4 billion to expand our rural delivery network across the U.S. These are small towns where people want fast delivery, but where other companies have been backing out and reducing service. In contrast, we've already increased the number of rural communities with access to our same-day and next-day delivery by 60%, reaching roughly half of the total communities we plan to expand to by the end of the year. The stores team is also innovating rapidly with AI. For example, Rufus, our AI-powered shopping assistant has had 250 million active customers this year with monthly users up 140% year-over-year, interactions up 210% year-over-year and customers using Rufus during a shopping trip being 60% more likely to complete a purchase. Rufus is on track to deliver over \$10 billion in incremental annualized sales. Here are the highlights. Our generative AI-powered audio feature that combines product summaries and reviews to make shopping easier has expanded from hundreds of products at launch to millions of products and millions of customers have used it streaming almost 3 million minutes. In Amazon Lens, an AI-powered visual search tool that lets customers find products with their phones camera, a screenshot or a bar code, now includes Lens Live, which instantly scans products and shows real-time matches in a swipeable carousel. Tens and millions of customers are using Amazon Lens each month. Moving on to Amazon ads. We're pleased with the continued strong growth, generating \$17.6 billion of revenue in the quarter and growing 22% year-over-year. We see strength across our broad portfolio of full photo advertising offerings that helps advertisers reach an average ad-supported audience of more than \$300 million in the U.S. alone. We also continue to be excited about our demand side platform, Amazon DSP, which lets advertisers plan, activate and measure full funnel investments. Last quarter, I mentioned our partnership with Roku and we've built on that with a partnership with Netflix, providing advertisers using Amazon DSP with direct access to Netflix's premium ad inventory. We announced integrations with Spotify and SiriusXM. With Spotify, we provide advertisers with direct programmatic access to a global audience of more than 400 million monthly ad-supported listeners. And with SiriusXM, brands can reach 160 million monthly digital listeners across services like Pandora and SoundCloud and we're excited about the advertising opportunity around prime video live sports. Live sports got a lot of interest from advertisers in upfront negotiations for 2025, '26, and we exceeded our own expectations for upfront commitments with significant growth across the board. Finally, we're continuing to invade for advertisers with AI. For example, in September, we announced an agentic AI tool and creative studio that plans and executes the entire creative process in a matter of hours instead of weeks. We're also inventing and seeing strong momentum in several other areas, and I'll mention just a few. In Prime Video live sports, NBA on Prime tipped off last week and our opening night doubleheader averaged 1.25 million viewers in the U.S., a double-digit increase over last season on cable. You'll see us bring the same constant innovation here that we brought to our NFL broadcast. We're adding golf with The Masters in 2026 and new skins competition with the PGA Tour on Black Friday this year. And we've added Peacock and FOX One to Prime Videos add-on subscription offering of over 100 channels in the U.S. We continue to be energized by the response to Alexa+ compared to what we call the classic Alexa experience, Alexa+ customers are talking to Alexa 2x more. Those interactions are much longer, and they're covering a broader range of topics. So using Alexa+ and Fire TV at 2.5x the rate of classic using natural conversation to discover audio content 4x more, engaging with photos 4x more and customers are completing 4x more shopping conversations that end in a purchase. We've expanded the number of project hyper satellites and space to more than 150 and delivered over 1 gigabit per second speeds and test with our enterprise-grade customer terminal, the first commercial phased array we know of to clear that threshold. Finally, Zoox robotaxis are available to riders in Las Vegas, and we've announced Washington, D.C. as the eighth testing location. We're excited for these to continue rolling out to more riders. Q4 is one of our busiest and most energizing times of the year, and we're excited about the continued demand for AWS. The innovations will announce the reinvent in December, the positive customer response to our AI-powered experiences, all the guests will be delivering throughout the holiday season and a lot more. Thanks in advance to our teammates around the world who are gearing up to deliver for customers once again. With that, I'll turn it over to Brian for a financial update.

Brian Olsavsky

Thanks, Andy. Starting with our top line financial results. Worldwide revenue was \$180.2 billion, a 12% increase year-over-year, excluding a 90 basis point favorable impact of foreign exchange. In Q3, we reported worldwide operating income of \$17.4 billion. This operating income includes 2 special charges, which reduced operating income by \$4.3 billion. The first charge of \$2.5 billion is related to a legal settlement with the Federal Trade Commission, which impacts the North America segment and is recorded in the other operating expense line. The second charge of \$1.8 billion relates to severance costs for roll eliminations and impacts all 3 of our segments. The severance charge is recorded primarily in the technology and infrastructure, sales and marketing and general and administrative expense line items. Excluding these 2 charges, worldwide operating income would have been \$21.7 billion or \$1.2 billion above the high end of our guidance range. Moving to our segment results. We remain encouraged by the innovation our teams are delivering for customers across all 3 segments. In the North America segment, third quarter revenue was \$106.3 billion, an increase of 11% year-over-year. International segment revenue was \$40.9 billion, an increase of 10% year-over-year, excluding the impact of foreign exchange. Worldwide paid units grew 11% year-over-year. We continue to prioritize the inputs that matter most to our customers. In the third quarter, our sharp pricing, broad selection and fast delivery speeds continue to resonate with customers. Customers appreciate the ability to quickly receive items essential for their daily needs, including perishable groceries and have them delivered in the same day. Our millions of global third-party sellers continue to be important contributors to our vast selection, which helps customers find the items they need at competitive prices. We're committed to building innovative services and features for our sellers, including our ongoing advancements in generative AI. Today, more than 1.3 million sellers have used our generative AI capabilities to more quickly launch high-quality listings. Better listings translate into better traction with customers. And

in Q3, worldwide third-party seller unit mix was 62%, up 200 basis points from Q3 of last year. Shifting to profitability. North America segment operating income was \$4.8 billion, with an operating margin of 4.5%. Excluding the \$2.5 billion charge related to the legal settlement with the FTC, North America segment operating income would have been \$7.3 billion with an operating margin of 6.9%. North America segment operating margin also includes a portion of the severance charge. International segment operating income was \$1.2 billion, with an operating margin of 2.9%. Excluding the impact of the severance charge International segment operating margins expanded year-over-year. Globally, our progress on key inputs is delivering a better customer experience while driving a more efficient cost structure. For example, we're making notable strides in improving inventory placement to speed up delivery to customers. And as a result, for the third year in a row, we are on track to deliver our fastest speeds ever for Prime members in 2025. We continue to tune and improve our fulfillment operations and our regionalized network is operating at scale. We see many benefits of our inbound process improvements, including a reduction of U.S. inbound lead time by nearly 4 days compared to last year. This allows us to be more efficient with our inventory purchasing, which benefits working capital. We're also placing inventory more strategically throughout the network. And by leveraging our existing infrastructure, we're now offering U.S. customers the ability to order perishable groceries and receive them the same day and as little as 5 hours. We're seeing positive early results since launching in January, when customers start shopping groceries on Amazon, they are visiting the site more often and returning twice as often as nonperishable shoppers. Looking ahead, we see further opportunity to improve our activity in our global fulfillment and transportation network. We continue to improve inventory placement to drive down distance travel and touches for package. We will also build on the gains from our regionalized network through algorithmic improvements as well as launching robotics and automation. Operating margin may fluctuate quarter-to-quarter, we have a delivered approach to achieve sustained progress over the long term. Shifting to advertising. Advertising revenue was \$17.7 billion and growth accelerated for the third consecutive quarter. We continue to see strong growth on an increasingly large base, as our full funnel advertising approach of connecting brands with customers is resonating. Moving next to our AWS segment. Revenue was \$33 billion, up 20.2% year-over-year. This is an acceleration of 270 basis points compared to last quarter, driven by strong growth across both our AI and core services and more capacity, which has come online to support customer demand. AWS revenue increased \$2.1 billion quarter-over-quarter and now has an annualized revenue run rate of \$132 billion. AWS operating income was \$11.4 billion, and reflects our continued growth, coupled with our focus on driving efficiencies across the business. We are expanding our data center footprint, largely to accommodate Gen AI. And to the extent those assets were placed into service related to depreciation does impact our margins. As we've long said, we expect AWS operating margins to fluctuate over time, driven in part by the level of investments we're making at any point in time. Now turning to our cash CapEx, which was \$34.2 billion in Q3. We've now spent \$89.9 billion so far this year. This primarily relates to AWS as we invest to support demand for our AI and core services and in custom silicon, like Trainium as well as tech infrastructure to support our North America and international segments. We'll continue to make significant investments, especially in AI, as we believe it to be a massive opportunity with the potential for strong returns on invested capital over the long term. Additionally, we continue to invest in our fulfillment and transportation network to support the growth of the business, improve delivery speeds and lower our cost to serve. These investments will support growth for many years to come. Looking ahead, we expect our full year cash CapEx to be approximately \$125 billion in 2025, and we expect that amount will increase in 2026. I'll finish up my remarks with net income. While we primarily focus our comments on operating income, our third quarter net income of \$21.2 billion includes a pretax gain of \$9.5 billion related to our investment in Anthropic. This investment activity is not related to Amazon's ongoing operations and is included in nonoperating income. We're encouraged by the start of the peak season and we are ready to serve customers in the coming months. I want to thank our teams across Amazon for their hard work as we get ready to delight customers during the holiday season. Our commitment to elevating the customer experience is the only reliable way to drive sustainable value for our shareholders. With that, let's move on to your questions.

Operator

[Operator Instructions] And our first question comes from the line of Justin Post with Bank of America.

Justin Post

I'll ask on AWS. Can you just kind of go through how you're feeling about your capacity levels and how capacity constrained you are right now? And then in your prepared remarks, you mentioned Trainium3 demand and maybe broadening out your customer base. Can you talk about the demand you're seeing outside of your major customers for Trainium?

Andrew Jassy

Yes. On the capacity side, we brought in quite a bit of capacity, as I mentioned in my opening comments, 3.8 gigawatts of capacity in the last year with another gigawatt plus coming in the fourth quarter and we expect to double our overall capacity by the end of 2027. So we're bringing in quite a bit of capacity today, overall in the industry, maybe the bottleneck is power. I think at some point, it may move to chips, but we're bringing in quite a bit of capacity. And as fast as we're bringing in right now, we are monetizing it. And then on the Trainium demand, outside of our major customers. So first of all, as I mentioned on Trainium2, it's really doing well. It's fully subscribed on Trainium2. We have -- it's a multibillion-dollar business at this point. It grew 150% quarter-over-quarter in revenue. And you see really big projects at scale now, like our Project Rainier that we're doing with Anthropic, where they're running their next version of -- they're training the next version of Claude on top of Trainium2 on 500,000 Trainium2 chips going to 1 million Trainium2 chips by the end of the year. As I mentioned, we have -- today, with Trainium2, we have a small number of very large customers on it. But because Trainium is 30% to 40% more price performance than other options out there, and because as customers, as they start to contemplate broader scale of their production workloads, moving to being AI-focused and using inference, they badly care about price performance. And so we have a lot of demand for Trainium. Trainium3 should preview at the end of this year with much fuller volumes coming in the beginning of '26, we have a lot of customers, both very large, and I'll call it, medium-sized who're quite interested in Trainium3.

Operator

And the next question comes from the line of Brian Nowak with Morgan Stanley.

Brian Nowak

Congrats on the quarter, guys. So maybe 2. One, Andy, sort of a philosophical chip question. There's a lot of questions in the market about Trainium and sort of its positioning versus other third-party chips. So how do you think about the key hurdles of Trainium3 need to overcome to really make Trainium adoption broader, to your point on the last question and continue to drive Trainium as opposed to satisfying what could be broader demand with third-party chips in the near term?

Andrew Jassy

Yes. Well, first of all, we're always going to have multiple chip options for our customers. It's been true in every major technology building block or component that we've had in AWS. Really in the history of AWS, it's never just one player that over a long period of time has the entire market segment and then can satisfy everybody's needs on every dimension. And so we have a very deep relationship with NVIDIA. We have for a very long time. And we will for as long as I can foresee the future. We buy a lot of NVIDIA. We are not constrained in any way in buying NVIDIA, and I expect that we'll continue to buy more NVIDIA both next year and in the future. But we're different from most technology companies in that we have our own very strong chip team, and this is our Annapurna team. And you saw it first on the CPU side with what we built with Graviton which is about 40% better price performance than the other x86 processors, and you're seeing it again on the custom silicon on the AI side with Trainium, which is about the same amount of price performance benefit for customers relative to other GPU options. And our customers to be able to use AI as expansively as they want. And remember, it's still relatively early days at this point. They're going to need better price performance and they care about it deeply. And so I mentioned earlier the momentum that Trainium2 has. And I think that for us, as we think about Trainium3, I expect Trainium3 will be about 40% better than Trainium2 and Trainium2 is already very advantaged on price performance. So we have to, of course, deliver the chip. We have to deliver it in volumes and deliver it quickly. And we have to continue to work on the software ecosystem, which gets better all the time. And as we have more proof points like we have with Project Rainier with what Anthropic's doing on Trainium2, it builds increasing credibility for Trainium. And I think customers are very bullish about it. I'm bullish about it as well.

Operator

And our next question comes from the line of Doug Anmuth with JPMorgan.

Douglas Anmuth

I'll stick with basically the same topic, Andy. But can you just talk a little bit about the architecture of Project Rainier and how it's differentiated and what that means for customers and for AWS? And do you expect Rainier to expand beyond Anthropic? And how do you replicate Rainier with Trainium3 chips?

Andrew Jassy

Yes. I think what is compelling for entropic around Project Rainier is really is the Trainium2 chip, which we built a very -- first of all, we built a very large cluster that they can use in a very expansive way. And it's not simple to be able to build a cluster that has 500,000 plus chips going to 1 million. That's an infrastructure feat that's hard to do at scale. And so some piece of it is the infrastructure capabilities that we've built over a long period of time in AWS that is unusual in the industry. But it's just also the performance of the chip and the price performance, both of which matter. And I think that Project Rainier is something that is specific for anthropic, but we have a lot of other customers who are interested in employing large clusters of Trainium chips that we're going to hopefully give them a chance to do so with Trainium3.

Operator

The next question comes from the line of Mark Mahaney with Evercore ISI.

Mark Stephen Mahaney

I want to ask about 2 topics, groceries and then how to think about head count in the future. And on groceries, I want to -- the perishables, I think last quarter, you talked about 70% or something of users had never purchased from perishables from Amazon before. Just talk about whether you -- I think you used the term, Game Changer, before. Does this mean that maybe we don't -- you no longer need to Amazon Fresh stores. You always had this DVD delivery van density advantage. And have you kind of reached a point you think of scale and speed that you really can change people habit and really have them consider Amazon as one of their first grocery options? Do you really feel like you're at that point? And then secondly, just on the head count, some of the recent news. Just talk to us about how you think about head count going forward? Are you seeing -- is the level of efficiencies that you're getting from AI such that you can keep head count relatively flattish for the foreseeable future? Just talk about the pros and cons or the wins and losses in terms of that head count going forward?

Andrew Jassy

Yes. So I'll start with grocery, Mark. We have a very large grocery business. If you look at our entire grocery business, if I don't even count Whole Foods Market and Fresh, in the last 12 months to over \$100 billion of gross merchandising sales, which would make us a top 3 grocery in the U.S. A good chunk of it is a lot of the items that you'd find in the middle aisle so consumables and canned goods and pet food and health and beauty, very significant and continues to grow at a very good clip. But then we also have Whole Foods Market, which is the pioneer in organic foods, which is also growing at a faster clip than most grocery companies with an attractive trajectory on profitability, and we'll expand our Whole Foods physical presence over the coming years here. And I'm also very excited about this new concept, daily shop that we have, which is a smaller version of Whole Foods in urban settings, which we have 3 that we've launched that are off to very good starts that you should expect to see more of as well. And we have always been -- as you referenced, we've talked a lot about having a larger mass physical presence. And we continue to experiment with various formats. But the one that we are most excited about is what you referenced, which is the ability to provide perishable groceries with same-day deliveries. And if you think about how many of our customers are buying from us multiple times a week and who are buying things like shampoo or detergent or paper cups or water, where the ability to add milk and eggs and yogurt and other perishables to their order and have it live in the same shop in cart and then show up a few hours later, is very compelling. And we started with a few markets about a year ago, and we were really taken aback at the adoption, not just the number of people that started buying perishables from us very quickly but how often they came back downstream to buy perishables and groceries from us in the future. And so we've now expanded that to 1,000 cities around the U.S. and will be in 2,300 by the end of the year. And it's really changing the trajectory and the size of our grocery business. And I also believe that this many years tradition of the weekly stock up grocery stock up is changing. And I think we're a big part of that. And I think there's a lot of potential there for the grocery side. It doesn't mean that we won't continue to experiment with other physical formats, but we're on to something very significant with what we're doing with perish both from our same-day facilities. And then on your head count question, what I would tell you is the announcement that we made a few days ago was not really financially driven and it's not even really AI-driven, not right now, at least. It really -- its culture. And if you grow as fast as we did for several years, the size of businesses, the number of people, the number of locations, the types of businesses you're in, you end up with a lot more people than what you had before, and you end up with a lot more layers. And when that happens, sometimes without realizing that you can weaken the ownership of the people that you have who are doing the actual work and who own most of the 2-way door decisions, the ones that should be made quickly and right at the front line, and it can lead to slowing you down. And as a leadership team, we are committed to operating like the world's largest start-up. And that means removing layers. It means increasing the amount of ownership that people have, and it means inventing and moving quickly. And I don't know if there's ever been a time in the history of Amazon or maybe business in general with the

technology transformation happening right now, where it's important to be lean, it's important to be flat, and it's important to move fast, and that's what we're going to do.

Operator

And the next question comes from the line of Eric Sheridan with Goldman Sachs.

Eric Sheridan

Wanted to know, Andy, if you could reflect on the opportunity that's continuing to present itself in terms of rolling out more robotics and automation and the broader theme of physical AI across your operations? And how should we be thinking about that as a driver of potential efficiencies, but also as a driver of the ability to possibly reinvest back in the business over the long term?

Andrew Jassy

Robotics is a very substantial area of investment for us. We have over 1 million robots in our fulfillment network at this point. And I would say that while that's significant, we have a lot of invention in flight. So I expect that we'll have more over a period of time. Robotics are very important for us and for our customers and for our teammates because they improve safety, they boost productivity, they increased speed, and they let our human teammates focused on problem solving and what they do best. And we expect that our people remain at the heart in the center of our fulfillment network as they have from when we first started working the robotics. And we expect that over time, we will have a fulfillment network where robots and humans complement each other and work together. But I think you're going to continue to see us advance invest very significantly in robotics. It's going to help on the safety, the productivity, the speed and ultimately some of the cost pieces, which will allow us to continue to improve the customer experience.

Operator

And the next question comes from the line of John Blackledge with TD Cowen.

John Blackledge

How does Amazon think about agentic commerce going forward? And how do you think Amazon will serve customers using agents to purchase goods on Amazon in the future?

Andrew Jassy

I'm very excited about -- and as a business, we're very excited about in the long term the prospect of agentic commerce. And it has a chance to be good for customers has a chance to be really good for e-commerce. And I think if you're -- if you know what you want to buy, there are a few experiences that are better than coming to Amazon. But if you don't know what you want, it's a physical store with a physical salesperson still has some advantages. Obviously, lots of people do it on Amazon all the time. But you very often want to ask questions and help get help narrowing what you're going to look for. And as you keep asking new questions, having a whole bunch of different options presented to you. And I think AI and agentic commerce are going to change the experience online where that experience where you're narrowing what you want when you don't know is going to get better online than it even is in physical environments. Now we obviously have our own efforts here in agentic commerce. We have Rufus, which I talked about in my opening comments, which is continuing to get better and better and used more broadly. And we have features like Buy for Me where we will surface on Amazon, even items that we don't stock that other merchants have. And then if customers want us to go and buy it for them on those merchants websites, we will do that. And both of those have been successful for us. But we're also having conversations with and expect over time to partner with third-party agents. And I think that it reminds me in some ways of the beginning of search engines many years ago being sources of discovery for commerce. And you had to kind of figure out the right way to work together. And today, search engines are a very small part of our referral traffic and third-party agents are a very small subset of that. But I do think that we will find ways to partner. We have to find a way, though, that makes the customer experience good. Right now, I would say the customer experience is not -- there's no personalization. There's no shopping history. The delivery estimates are frequently wrong. The prices are often wrong. So we've got to find a way to make the customer experience better and have the right exchange value. But I do think that the exciting part of this and the promise is that AI and agentic commerce solutions are going to expand the amount of shopping that happens online. And I think that's really good for customers, and I think it's really good for Amazon because at the end of the day, you're going to buy from the outfit that allows you to have the broadest selection, great value and continues to deliver for you very quickly and reliably. And I think that bodes well for us.

Operator

And our final question comes from the line of Colin Sebastian with Baird.

Colin Sebastian

I guess first on AWS, following up there. How much of this acceleration is driven by core infrastructure versus AI workload monetization? And I think part of it is trying to understand how important newer services like AgentCore are becoming and bringing enterprises to AWS to build agents? And then I guess, secondly, regarding the acceleration in advertising, if you could potentially disaggregate the core advertising contribution versus DSP and Prime video. That would be helpful as well.

Andrew Jassy

I'll start on the AWS side, we are seeing -- we're really pleased with the results from this quarter, 20% year-over-year on an annualized run rate of \$132 billion is unusual. And we have momentum. You can see it. And we see the growth in both our AI area, where we see it in inference. We see it in training. We see it in the use of our Trainium custom silicon. Bedrock continues to grow really quickly. SageMaker continues to grow quickly. And I think that the number of companies who are working on building agents is very significant. I do believe that a lot of the value that companies will realize over time and AI will come from agents. And I think that building agents today is still harder than it should be. You need tools to make it easier, which is why we built strands, which is an open source capability that lets people build agents from any model that they can imagine. But even more so, when you talk to enterprises or companies that care a lot about security and

scale. They're starting to build agents, and they don't really feel like they've got -- they've had building blocks that allow them to have the type of secure scalable agents that they need to bet their businesses and their customer experience and their data. And that's why -- that was really the inspiration behind AgentCore was to build another set of primitive building blocks like we built in the early days of AWS, where it was compute and storage and database. We defined a set of building blocks that you needed to be able to deploy agents securely and scalably that we provide in AgentCore. And then when we talk to our customers, it really resonates. There is not anything else like it, it's changing their time frame and their receptivity to building agents, and it's very compelling for them. So I do think the combination of what we're doing to enable agents to be built and run securely and scalably as well as some of the agents that we're building ourselves that our customers are excited about are compelling for them. And I think the other place we see a lot of growth in AWS also is just the number of enterprises who are -- who have gotten back to moving from on-premises infrastructure to the cloud. And we continue to earn the lion's share of those transformations. And I look at the momentum we have right now, and I believe that we can continue to grow at a clip like this for a while. I think on the advertising side, that is also an area where I think collectively, we feel very pleased about the progress. Every single one of our advertising offerings this quarter grew in a meaningful way. I think there's a few things going on for us. We have what I think of as a pretty unusual full funnel offering. And if you look at the top of the funnel, which typically tends to be awareness building and broad scale to be able to use our own Prime Video and our live sports capabilities as well as going all the way down to the bottom of the funnel at point of sale being able to use sponsored products, that's -- most people don't have a full funnel offering as robust as that. And then when you layer on top of it, the combination of the audience curation and development we can do, along with the advantage measurement, it just all leads to a return on advertising spend is very unusual. And I think there are multiple places where we can expect to continue to grow. One is in our stores business. I still think if you look at the worldwide market segment share of retail, still 80% to 85% of it lives in physical stores. And that equation is going to flip over time. And I think AI is going to only accelerate that. So I think we have an opportunity -- a significant opportunity still in our existing stores. And then I think video, we've only been at this for a little bit of time, but it's already a very large amount of advertising revenue, and we're still relatively early stage. I think that will continue to be a big area of growth. And then as you referenced, the amount -- their demand-side platform or Amazon DSP, that is growing really quickly as well. And some of it had to do with the fact that we had some features. We always had a number of the core components people wanted around some of our properties, the measurement capabilities, Amazon Marketing Cloud, but we lack some features for a while as we were building out our DSP that customers told us mattered and the team over the last 20 months have closed those gaps in a very significant way so that now people feel like our DSP is fully featured. And then you look at some of the partnerships that we've done, the Roku partnership gives us the largest connected TV footprint in the U.S. And you layer on top of that, what we've recently done in providing our DSP customers, the opportunity to integrate with the ad inventory in Netflix and Spotify and SiriusXM, it's powerful. And so we are growing very quickly on the demand side platform. So very optimistic about what we're doing there. We've continued work to do, obviously, but I don't think we're close to being able to grow there.

Operator

Thanks for joining us on the call today and for your questions. A replay will be available on our Investor Relations website for at least 3 months. We appreciate your interest in Amazon and look forward to speaking with you again next quarter.