NVDA Earnings Call – FY2026 Q2

Generated by discountingcashflows.com

Date: August 27, 2025

Sarah

Good afternoon. My name is Sarah, and I will be your conference operator today. At this time, I would like to welcome everyone to NVIDIA Corporation's Second Quarter Fiscal 2026 Financial Results Conference Call. All lines have been placed on mute to prevent any background noise. After the speakers' remarks, there will be a question and answer session. If you would like to ask a question during this time, simply press star followed by the number one on your telephone keypad. If you would like to withdraw your question, press star 1 again. Thank you. Toshiya Hari, you may begin your conference. Thank you.

Toshiya Hari

Good afternoon, everyone, and welcome to NVIDIA Corporation's conference call for 2026. With me today from NVIDIA Corporation are Jensen Huang, president and chief executive officer, and Colette Kress, executive vice president and chief financial officer. I would like to remind you that our call is being webcast live on NVIDIA Corporation's 2026. The content of today's call is NVIDIA Corporation's property. It cannot be reproduced or transcribed without our prior written consent. During this call, we may make forward-looking statements based on current expectations. These are subject to a number of significant risks and uncertainties, and our actual results may differ materially. For a discussion of factors that could affect our future financial results and business, please refer to the disclosure in today's earnings release, our most recent forms 10-K and 10-Q, and the reports that we may file on Form 8-K with the Securities and Exchange Commission. All our statements are made as of today, 08/27/2025, based on information currently available to us. Except as required by law, we assume no obligation to update any such statements. During this call, we will discuss non-GAAP financial measures. You can find a reconciliation of these non-GAAP financial measures to GAAP financial measures in our CFO commentary, which is posted on our website. With that, let me turn the call over to Colette.

Colette Kress

We delivered another record quarter while navigating what continues to be a dynamic external environment. Total revenue was $46.7 billion, exceeding our outlook as we grew sequentially across all market platforms. Data center revenue grew 56% year over year. Data center revenue also grew sequentially despite the $4 billion decline in H20 revenue. NVIDIA Corporation's Blackwell platform reached record levels, growing sequentially by 17%. We began production shipments of GB300 in Q2. Our full stack AI solutions for cloud service providers, Neo Clouds, enterprises, and sovereigns are all contributing to our growth. We are at the beginning of an industrial revolution that will transform every industry. We see $3 to $4 trillion in AI infrastructure spend by the end of the decade. The scale and scope of these build-outs present significant long-term growth opportunities for NVIDIA Corporation. The GB200 NBL system is seeing widespread adoption with deployments at CSPs and consumer Internet companies. Lighthouse model builders, including OpenAI, Meta, and Mastral, are using the GB200 NBL72 at data center scale for both training next-generation models and serving inference models in production. The new Blackwell Ultra platform has also had a strong quarter, generating tens of billions in revenue. The transition to the GB300 has been seamless for major cloud service providers due to its shared architecture, software, and physical footprint. The GB200 enables them to build and deploy GB300 racks with ease. The transition to the new GB300 rack-based architecture has been seamless. Factory builds in late July and early August were successfully converted to support the GV300 ramp. Today, full production is underway. The current run rate is back at full speed, producing approximately 1,000 racks per week. This output is expected to accelerate even further throughout the third quarter as additional capacity comes online. We expect widespread market availability in the second half of the year as CorWeave prepares to bring their GV300 instance to market as they are already seeing 10x more inference performance on reasoning models compared to H100. Compared to the previous Hopper generation, GV300 NDL72 AI factories promise a 10x improvement in token per watt energy efficiency, which translates to revenues as data centers are power limited. The chips of the Rubin platform are in fab. The Vera CPU, Rubin GPU, CX9 Supernic, NBLink 144 scale-up switch, SpectrumX scale-out and scale-across switch, and the silicon photonics processor. Rubin remains on schedule for volume production next year. Rubin will be our third-generation NVLink Rack Scale AI supercomputer with a full-scale supply chain. This keeps us on track with our pace of an annual product cadence and continuous innovation across compute, networking, systems, and software. In late July, the US government began reviewing licenses for sales of H20 to China customers. While a select number of our China-based customers have received licenses over the past few weeks, we have not shipped any H20 based on those licenses. USG officials have expressed an expectation that the US will receive 15% of the revenue generated from licensed H20 sales, but to date, the USG has not published a regulation codifying such a requirement. We have not included H20 in our Q3 outlook as we continue to work through geopolitical issues. If geopolitical issues reside, we should ship $2 to $5 billion in H20 revenue in Q3. And if we had more orders, we can bill more. We continue to advocate for the US government to approve Blackwell for China. Our products are designed and sold for beneficial commercial use, and every license sale we make will benefit the US economy, the US leadership, in highly competitive markets. We want to win the hearts of every developer. America's AI technology stack can be the world's standard if we race and compete globally. Notably in the quarter was an increase in Hopper 100 and H200 shipments. We also sold approximately $650 million of H20 in Q2 to an unrestricted customer outside of China. The sequential increase in Hopper demand indicates the breadth of data center workloads that run on accelerated computing and the power of CUDA libraries and full stack optimizations, which continuously enhance the performance and economic value of our platform. As we continue to deliver both Hopper and Blackwell GPUs, we are focusing on meeting the soaring global demand. This growth is fueled by capital expenditures from the cloud to enterprises, which are on track to invest $600 billion in data center infrastructure and compute this calendar year alone, nearly doubling in two years. We expect annual AI infrastructure investments to continue growing, driven by several factors: reasoning agentic AI requiring orders of magnitude more training and inference compute, global build-outs for Sovereign AI, enterprise AI adoption, and the arrival of physical AI and robotics. Blackwell has set the benchmark as it is the new standard for AI inference performance. The market for AI inference is expanding rapidly with reasoning and agentic AI gaining traction across industries. Blackwell's RackScale NVLink and CUDA full stack architecture addresses this by redefining the economics of inference. New NVFP44 bit precision and NVLink 72 on the GB300 platform delivers a 50x increase in energy efficiency per token compared to Hopper, enabling companies to monetize their compute at unprecedented scale. For instance, a $3 million investment in GV200 can generate $30 million in token revenue, a 10x return. NVIDIA Corporation's software innovation combined with the strength of our developer ecosystem has already improved Blackwell's performance by more than 2x since its launch. Advances in CUDA, TensorRT LLM, and Dynamo are unlocking maximum efficiency. CUDA library contributions from the open-source community along with NVIDIA Corporation's open libraries and frameworks are now integrated into millions of workflows. This powerful flywheel of collaborative innovation between NVIDIA Corporation and global community contribution strengthens NVIDIA Corporation's performance leadership. NVIDIA Corporation is a top contributor to OpenAI models, data, and software. Blackwell has introduced a groundbreaking numerical approach to large language model pre-training. Using NBFP4 computations on the GB300 can now achieve 7x faster training than the H100, which uses FP8. This innovation delivers the accuracy of 16-bit precision with the speed and efficiency of 4-bit, setting a new standard for AI factor efficiency and scalability. The AI industry is quickly adopting this revolutionary technology, with major players such as AWS, Google Cloud, Microsoft Azure, and OpenAI, as well as Cohere, Mistral, Kimi AI, Perplexity, Reflection, and Runway, already embracing it.

NVIDIA Corporation's performance leadership was further validated in the latest MLPerth training benchmarks, where the GB200 delivered a clean sweep. Be on the lookout for the upcoming MLPerf inference results in September, which will include benchmarks based on the Blackwell Ultra. NVIDIA RTX Pro servers are in full production for the world system makers. These are air-cooled PCIe-based systems integrated seamlessly into standard IT environments and run traditional enterprise 90 companies, including many global leaders, are already adopting RTX Pro servers. Hitachi uses them for real-time simulation and digital twins, Lily for drug discovery, Hyundai for factory design and AV validation, and Disney for immersive storytelling. As enterprises modernize data centers, RTX Pro servers are poised to become a multibillion-dollar product line. Sovereign AI is on the rise as the nation's ability to develop its own AI using domestic infrastructure, data, and talent presents a significant opportunity for NVIDIA Corporation. NVIDIA Corporation is at the forefront of landmark initiatives across the UK and Europe. The European Union plans to invest €20 billion to establish 20 AI factories across France, Germany, Italy, and Spain, including five gigafactories to increase its AI compute infrastructure tenfold. In the UK, the Isambard AI supercomputer powered by NVIDIA Corporation was unveiled as the country's most powerful AI system, delivering 21 exaflops of AI performance to accelerate breakthroughs in fields of drug discovery and climate modeling. We are on track to achieve over $20 billion in Sovereign AI revenue this year, more than double that of last year. Networking delivered record revenue of $7.3 billion, and escalating demands of AI compute clusters necessitate high if and low latency networking. This represents a 46% sequential and 98% year-on-year increase with strong demand across Spectrum X Ethernet, InfiniBand, and NVLink. Our Spectrum X enhanced Ethernet solutions provide the highest throughput and lowest latency network for Ethernet AI workloads. Spectrum X Ethernet delivered double-digit sequential and year-over-year growth with annualized revenue exceeding $10 billion. At hot chips, we introduced SPECTRUM, XGS, Ethernet, a technology designed to unify disparate data centers into gigascale AI super factories. Farweave is an initial adopter of the solution, which is projected to double GPU to GPU communication speed. InfiniBand revenue nearly doubled sequentially, fueled by the adoption of XDR technology, which provides double the bandwidth improvement over its predecessor, especially valuable for the model builders. The world's fastest switch NVLink, with 14x the bandwidth of PCIe Gen five, delivered strong growth as customers deployed Brace Blackwell and D Link Rack Scale systems. The positive reception to NVLink fusion, which allows semicustom AI infrastructure, has been widespread. Japan's upcoming Fugaku NEXT will integrate Fujitsu's CPUs with our architecture via NVLink Fusion. It will run a range of workloads, including AI, supercomputing, and quantum computing. Fugaku NEXT joins a rapidly expanding list of leading quantum supercomputing and research centers running on NVIDIA Corporation's CUDA Q quantum platform, including ULEC, AIST, NNF, and NERSC. Supported by over 300 ecosystem partners, including AWS, Google Quantum, AI, Quantinuum, QEra, and SciQuantum. Justin Thor, our new robotics computing platform, is now available. Thor delivers an order of magnitude greater AI performance and energy efficiency than NVIDIA AGX Orin. It runs the latest generative and reasoning AI models at the edge in real-time, enabling state-of-the-art robotics. Adoption of NVIDIA Corporation's robotics full stack platform is growing at a rapid rate. Over 2 million developers and 1,000 plus hardware, software applications, and sensor partners are taking our platform to market. Leading enterprises across industries have adopted Thor, including Agility Robotics, Amazon Robotics, Boston Dynamics, Caterpillar, Figure, Hexagon, Medtronic, and Meta. Robotic applications require exponentially more compute on the device and in infrastructure, representing a significant long-term demand driver for our data center platform. NVIDIA Omniverse with Cosmos is our data center physical AI digital platform built for the development of robot and robotic systems. This quarter, we announced a major expansion of our partnership with Siemens to enable AI automatic factories. Leading European robotics companies, including Agile Robots, neurorobotics, and universal robots, are building their latest innovations with the Omniverse platform. Transitioning to a quick summary of our revenue by geography. China declined on a sequential basis to low single digits percentage of data center revenue. Note, our Q3 outlook does not include H20 shipments to China customers. Singapore revenue represented 22% of second quarter's billed revenue as customers have centralized their invoicing in Singapore. Over 99% of data center compute revenue billed to Singapore was for US-based customers. Our gaming revenue was a record $4.3 billion, a 14% sequential increase and a 49% jump year on year. This was driven by the ramp of Blackwell GeForce GPUs as strong sales continued as we increase supply availability. This quarter, we shipped GeForce RTX 5060 desktop GPU. It brings double the performance along with advanced ray tracing, neural rendering, and AI-powered DLSS four. Gameplay to millions of gamers worldwide. Blackwell is coming to GeForce NOW in September. This is GeForce NOW's most significant upgrade, offering RTX 5080 class performance, minimal latency, and 5K resolution at 120 frames per second. We are also doubling the GeForce NOW catalog to over 4,500 titles, the largest library of any cloud gaming service. For AI enthusiasts, on-device AI performs the best RTX GPUs. We partnered with OpenAI to optimize their open-source GPT models for high-quality, fast, and efficient inference on millions of RTX-enabled window devices. With the RTX platform stack, window developers can create AI applications designed to run on the world's largest AI PC user base. Professional visualization revenue reached $601 million, a 32% year-on-year increase. Growth was driven by an adoption of the high-end RTX workstation GPUs and AI-powered workloads like design, simulation, and prototyping. Key customers are leveraging our solutions to accelerate their operations. Activision Blizzard uses RTX workstations to enhance creative workflows, while robotics innovator Figure AI powers its humanoid robots with RTX embedded GPUs. Automotive revenue, which includes only in-car compute revenue, was $586 million, up 69% year on year, primarily driven by self-driving solutions. We have begun shipments of NVIDIA Thor SoC, the successor to Orin. Thor's arrival coincides with the industry's accelerating shift to vision, language, model architecture, generative AI, and higher levels of autonomy. Thor is the most successful robotics and AV computer we have ever created or will power. Our full stack drive AV software platform is now opening up billions to new revenue opportunities for NVIDIA Corporation while improving vehicle safety and autonomy. Now moving to the rest of our P&L. GAAP gross margin was 72.4%, and non-GAAP gross margin was 72.7%. These figures include a $180 million or 40 basis point benefit from releasing previously reserved H20 inventory. Excluding this benefit, non-GAAP gross margins would have been 72.3%, still exceeding our outlook. GAAP operating expenses rose 86% on a non-GAAP basis sequentially. This increase was driven by higher compute and infrastructure costs as well as higher compensation and benefit costs. To support the ramp of Blackwell and Blackwell Ultra, inventory increased sequentially from $11 billion to $15 billion in Q2. While we prioritize funding, our growth and strategic initiatives in Q2, we returned $10 billion to shareholders through share repurchases and cash dividends. Our board of directors recently approved a $60 billion share repurchase authorization to add to our remaining $14.7 billion of authorization at the end of Q2. Okay. Let me turn it to the outlook for the third quarter. Total revenue is expected to be $54 billion plus or minus 2%. This represents over $7 billion in sequential growth. Again, we do not assume any H20 shipments to China customers in our outlook. GAAP and non-GAAP gross margins are expected to be 73.3% and 73.5%, respectively, plus or minus 50 basis points. We continue to expect to exit the year with non-GAAP gross margins in the mid-seventies. GAAP and non-GAAP operating expenses are expected to be approximately $5.9 billion and $4.2 billion, respectively. For the full year, we expect operating expenses to grow in the high thirties range year over year, up from our prior expectations of the mid-thirties. We are accelerating investments in the business to address the magnitude of growth opportunities that lie ahead. GAAP and non-GAAP other income and expenses are expected to be an income of approximately $500 million, excluding gains and losses from non-marketable and publicly held equity securities. GAAP and non-GAAP tax rates are expected to be 16.5%, plus or minus 1%, excluding any discrete items. Further financial data are included in the CFO commentary and other information available on our website. In closing, let me highlight upcoming events for the financial community. We will be at the Goldman Sachs Technology Conference on September 8 in San Francisco. Our annual NDR will commence the October GTC data center begins on October 27 with Jensen's keynote scheduled for the twenty-eighth. We look forward to seeing you at these events. Our earnings call to discuss the results of our 2026 is scheduled for November 19. We will now open the call for questions. Operator, would you please poll for questions?

Sarah

Thank you. At this time, I would like to remind everyone in order to ask a question, press star, then the number one on your telephone keypad. Thank you. Your first question comes from C.J. Muse with Cantor Fitzgerald. Your line is open.

C.J. Muse

I guess with wafer-in to rack-out lead times of twelve months, you confirmed on the call today that Rubin is on track for ramp in the second half. And, obviously, many of these investments are multiyear projects contingent upon power, cooling, etcetera. I was hoping perhaps you could you take a high-level view and speak to, you know, your vision for growth into 2026. And as part of that, if you could kinda comment between network and data center, would be very helpful. Thank you.

Toshiya Hari

Yeah. Thanks, C.J. At the highest level of growth, drivers would be the evolution, the introduction, if you will, of reasoning agentic AI. In a word, chatbots used to be one shot. You give it a prompt, and it would generate the answer. Now the AI does research, it thinks and does a plan. And it might use tools. And so it's called long thinking. And the longer it thinks, oftentimes, it produces better answers. And the amount of computation necessary for one shot versus reasoning agentic AI models could be a 100 times, a thousand times, and you potentially even more as the amount of research and basically reading and comprehension that it goes off to do. So the amount of computation that has resulted in agentic AI has grown tremendously. And, of course, the effectiveness has also grown tremendously. Because of agentic AI, the amount of hallucination has dropped significantly. You can now use it, you can now use tools and perform tasks. Enterprises have been opened up. As a result of agentic AI and vision language models, we now are seeing a breakthrough in physical AI in robotics, autonomous systems. So the last year, AI has made tremendous progress. And agentic systems, reasoning systems, is completely revolutionary. Now we built the Blackwell MVLink 72 system, a rack scale computing system for this moment. We have been working on it for several years. This last year, we transitioned from MVLink eight, which is a node scale computing. Each node is a computer. To now NVLink 72 where each rack is a computer. That disaggregation of NVLink 72 into a rack scale system was extremely hard to do. But the results are extraordinary. We are seeing orders of magnitude speed up and, therefore, energy efficiency and, therefore, cost-effectiveness of token generation because of MBLink 72. And so over the next over the next couple of years, you are gonna over well, you asked about longer term. Over the next five years, we are gonna scale into with Blackwell, with Rubin, and follow-ons to scale into a effectively a $3 to $4 trillion AI infrastructure opportunity. The last couple of years, you have seen that CapEx has grown in just the top four CSPs by has doubled and grown to about $600 billion. So we are in the beginning of this build-out, and the AI technology advances have really enabled AI to be able to adopt and solve problems to many different industries.

Sarah

Your next question comes from Vivek Arya with Bank of America Securities. Your line is open.

Vivek Arya

Colette, just wanted to clarify the $2 billion to $5 billion in China. What needs to happen and what is the sustainable pace of that China business as you get into Q4? And then Jensen, for you on the competitive landscape, several of your large customers already have or are planning many ASIC projects. I think one of your you know, ASIC competitors, Broadcom, signaled that they could grow their AI business almost, you know, 55, 60% next year. Any scenario in which you see the market moving more towards ASICs and away from NVIDIA Corporation GPU? Just what are you hearing from your customers? How are they managing this split between their use of merchant silicon and ASICs. Thank you. Thanks for that. So let me first answer your question regarding what will it take for the H20s to be shipped. There is interest in our H20s. There is the initial set of licenses that we received. And then additionally, we do have supply that we are ready, and that is why we communicated that somewhere in the range of about $2 to $5 billion this quarter we could potentially ship. We are still waiting on several of the geopolitical issues going back and forth between the governments and the companies trying to determine their purchases and what they want to do. So it is still open at this time, and we are not exactly sure what that full amount will be that this quarter. However, if more interest arrives, more licenses arrive, again, we can also still build additional H20 and ship more as well.

Jensen Huang

NVIDIA Corporation builds very different things than ASICs. But let us talk about ASICs first. A lot of projects are started. Many startup companies are created. Very few products go into production. And the reason for that is it is really hard. Accelerated computing is unlike general-purpose computing. You do not write software and just compile it into a processor. Accelerated computing is a full stack co-design problem. And AI factories in the last several years have become so much more complex. Because the scale of the problems has grown so significantly. It is really the ultimate, the most extreme computer science problem the world has ever seen, obviously. And so the stack is complicated. The models are changing incredibly fast. From generative based on autoregressive to generative based on diffusion to mixed models to multimodality, the number of different models that are coming out that are either derivatives of transformers or evolutions of transformers is just daunting. One of the advantages that we have is that NVIDIA Corporation is available in every cloud. We are available from every computer company. We are available from the cloud to on-prem to edge to robotics. On the same programming model. And so it is sensible that every framework in the world supports NVIDIA Corporation. When you are building a new model architecture, releasing it on NVIDIA Corporation's most sensible. And so the diversity of our platform both in the ability to evolve into any architecture, the fact that we are everywhere, and, also, we accelerate the entire pipeline. You know, everything from data processing to pretraining to post-training with reinforcement learning all the way out to inference. And so when you build a data center with NVIDIA Corporation, platform in it, the utility of it is best. The lifetime usefulness is much, much longer. And then I would just say that in addition to all of that, it is just a really extremely complex systems problem anymore. You know, people talk about the chip itself. There is one ASIC. The GPU that many people talk about. But in order to build Blackwell the platform, and Reuben, the platform, we had to build CPUs, that connect fast memory, low extremely energy-efficient memory. For large KB caching necessary for agentic AI. To the GPU to a super NIC to a scale-up switch we call NVLink, completely revolutionary when we are in our fifth generation now. To a scale-out switch, whether it is quantum or Spectrum X Ethernet, to now scale across switches so that we could prepare for these AI super factories. With multiple gigawatts of computing all connected together. We call that Spectrum XGS. We just announced that at Hotchips this week. And so the complications, the complexity of everything that we do is really quite extraordinary. It is just done at a really, really extreme scale now. And then lastly, if I could just say one more thing. You know, we are in every cloud for a good reason. Not only do are we the most energy-efficient, our perf per watt is the best of any computing platform. And in a world of power-limited data centers, perf per watt drives directly to revenues. And, you know, you have heard me say before, that in a lot of ways, the more you buy, the more you grow. And because our perf per dollar, the performance per dollar is so incredible, you also have extremely great margins. So the growth opportunity with NVIDIA Corporation's architecture and the gross margins opportunity with NVIDIA Corporation's architecture is absolutely the best. And so there are a lot of reasons why NVIDIA Corporation has chosen by every cloud and every startup and every computer company. We are, you know, really a holistic full stack solution for AI factories.

Sarah

Your next question comes from Ben Reitzes with Melius. Your line is open.

Ben Reitzes

Hey, thanks a lot. Jensen, I wanted to ask you about your $3 to $4 trillion in data center infrastructure spend by the end of the decade. Previously, you talked about something in the $1 trillion range, which I believe was just for compute. By 2028. If you take past comments, you know, $3 to $4 trillion would imply maybe $2 trillion plus in compute spend. And just wanted to know if that was right and that is what you are seeing by the end of the decade. And wondering what you think your share will be of that, your share right now of total infrastructure compute-wise is very high. So, wanted to see. And also if there are any bottlenecks you are concerned about, like power. To get to the $3 to $4 trillion. Thanks a lot.

Jensen Huang

Yeah. Thanks. As you know, the CapEx of just the top four hyperscalers has doubled in two years. As the AI revolution went into full steam, as the AI race is now on, the CapEx spend has doubled to $600 billion per year. There are five years between now and the end of the decade. And $600 billion only represents the top four hyperscalers. We still have the rest of the enterprise companies building on-prem. You have enter you have cloud service providers building around the world. United States represents about 60% of the world's compute. And over time, you would think that artificial intelligence would reflect GDP scale and growth. And so and would be, of course, accelerating GDP growth. And so our contribution to that is a large part of the AI infrastructure out of a gigawatt AI factory, which can go anywhere from 50 to, you know, plus or minus 10%, let us say, 50 to $60 billion, we represent about 35 plus or minus of that. And 35 out of fifty or so billion dollars for a gigawatt data center. And of course, what you get for that is not a GPU. Think people, you know, were famous for building the GPU and inventing the GPU. But as you know, over the last decade, we have really transitioned to become an AI infrastructure company. It takes six chips just to build six different types of chips just to build an AI, a Rubin AI supercomputer. And just to scale that out, you know, to a gigawatt, you have hundreds of thousands of GPU compute nodes. And a whole bunch of racks. And so we are really an AI infrastructure company and we are hoping to continue to contribute to growing this industry, making AI more useful, and then very importantly, driving the performance per watt because the world, as you mentioned, limiters it will always likely be power limitations or AI or AI building limitations. And so we need to squeeze as much out of that factory as possible. NVIDIA Corporation's performance per unit of energy used drives the revenue growth of that factory. It directly translates. If you have a 100-megawatt factory, perf per 100 megawatt drives your revenues. It is tokens per 100 megawatts of factory. In our case, also, the performance per dollar spent is so high that your gross margins are also the best. But anyhow, these are the limiters going forward. And $3 to $4 trillion is fairly sensible for the next five years.

Sarah

Next question comes from Joe Moore of Morgan Stanley. Your line is open.

Joe Moore

Great. Thank you. Congratulations on reopening the China opportunity. Can you talk about the long-term prospects there? You have talked about, I think, half of AI software world being there. You know, how much can NVIDIA Corporation grow in that business? And you know, how important is it that you get the Blackwall architecture ultimately licensed there?

Jensen Huang

The China market, I have estimated to be above $50 billion of opportunity for us this year. If we were able to address it, with competitive products and if it is $50 billion this year, you would expect it to grow say, 50% per year. As the rest of the world's AI market is growing as well. It is the second-largest computing market in the world, and it is also the home of AI researchers. About 50% of the world's AI researchers are in China. The vast majority of the leading open-source models are created in China. And so it is fairly important, I think, for the American technology companies to be able to address that market. And open source, as you know, is created in one country, but it is used all over the world. The open-source models that have come out of China are really excellent. DeepSeek, of course, gained global notoriety. QN is excellent. Kimi is excellent. There is a whole bunch of new models that are coming out. They are multimodal. They are great language models. And it is really fueled the adoption of AI in enterprises around the world because enterprises want to build their own custom proprietary software stacks. And so open-source model is really important for enterprise. It is really important for SaaS. Who also would like to build proprietary systems. It has been really incredible for robotics around the world. And so open source is really important. And it is important that the American companies are able to address it. This is going to be a very large market. We are talking to the administration about the importance of American companies to be able to address the Chinese market. And, as you know, H20 has been approved for companies that are not on the entities list. And many licenses have been approved. And so I think the opportunity for us to bring Blackwell to the China market is a real possibility. And so we just have to keep advocating the sensibility of and the importance of American tech companies to be able to lead and win the AI race. And help make the American tech stack the global standard.

Sarah

Your next question comes from the line of Aaron Rakers with Wells Fargo. Your line is open.

Aaron Rakers

Yes. Thank you for the question. I want to go back to the Spectrum XGS announcement this week and, you know, thinking about the Ethernet product now pushing over $10 billion of annualized revenue. You know, Jess, what is the opportunity set that you see for Spectrum XGS? Do you think about this as kind of the data center interconnect layer? Any thoughts on the sizing of this opportunity? You know, within that Ethernet portfolio? Thank you.

Jensen Huang

We now offer three networking technologies. One is for scale-up, one is for scale-out, and one for scale across. Scale-up is so that we could build the largest possible virtual GPU, the virtual compute node. NVLink is revolutionary. NVLink 72 is what made it possible for Blackwell to deliver such an extraordinary generational jump over Hopper's MBLink eight. At a time when we have long-thinking models, agentic AI reasoning systems, the NVLink basically amplifies the memory bandwidth, which is really critical for reasoning systems. And so NVLink 72 is fantastic. We then scale out with networking, which we have two. We have InfiniBand, which is unquestionably the lowest latency, the lowest jitter, the best scale-out network. It does require more expertise in managing those networks, and for supercomputing, for the leading model makers, InfiniBand, quantum InfiniBand is the unambiguous choice. If you were to benchmark an AI factory, ones

with InfiniBand are the best performance. For those who would like to use Ethernet because their whole data center is built with Ethernet, we have a new type of Ethernet called Spectrum Ethernet. Spectrum Ethernet is not off the shelf. It has a whole bunch of new technologies designed for low latency and low jitter and congestion control, and it has the ability to come closer, much, much closer to InfiniBand than anything that is out there. And that is we call that Spectrum X Ethernet. And then finally, we have Spectrum XGS, a gigascale for connecting multiple data centers, multiple AI factories into a super factory. A gigantic system. And you are going to see that networking obviously is very important in AI factories. In fact, choosing the right networking, the performance, the throughput, improvement. Going from, you know, 65% to 85% or 90%. That kind of step up because of your networking capability effectively makes networking free. You know? Choosing the right networking, you are basically paying, you know, you will get a return on it like you cannot believe because the AI factory, a gigawatt, as I mentioned before, could be $50 billion. And so the ability to improve the efficiency of that factory by tens of percents results in $1 to $20 billion worth of effective benefit. So, you know, the networking is a very important part of it. It is the reason why NVIDIA Corporation dedicates so much in networking. That is the reason why we purchased Mellanox five and a half years ago. And Spectrum X, as we mentioned earlier, is now quite a sizable business, and it is only about a year and a half old. So Spectrum X is a home run. All three of them are going to be fantastic. NVLink scale-up, SpectrumX, and InfiniBand, scale-out. And then Spectrum XGS for scale across.

Sarah

Your next question comes from Stacy Rasgon with Bernstein Research. Your line is open.

Stacy Rasgon

Hi, guys. Thanks for taking my question. Have a more tactical question for Colette. So on the guide, we are up, you know, over $7 billion. The vast bulk of that is going to be from data center. How do I think about apportioning that $7 billion out across Blackwell versus Hopper versus networking? I mean, it looks like Blackwell was probably $27 billion in the quarter up from maybe $23 billion last quarter. You know, Hopper is still $6 or $7 billion. Post the H20. Like, do you think the Hopper strength continues? Because how do I think about parsing that $7 billion out across all the three those three different components?

Colette Kress

Thanks, Stacy, for the question. First part of it, looking at our growth between Q2 and Q3, Blackwell is still going to be the lion's share of what we have in terms of data center. But keep in mind, that helps both our compute side as well as it helps our networking side. Because we are selling those significant systems that are incorporating the NVLink that Jensen just spoke about. Selling Hopper, we are still selling it. H100, H200s, we are. But, again, they are HCX systems, and I still believe our Blackwell will be the lion's share of what we are doing. On there. So we will continue. We do not have any more specific details, in terms of how we will finish our quarter but you should expect Blackwell again to be the driver of the growth.

Sarah

Your next question comes from Jim Schneider of Goldman Sachs. Your line is open.

Jim Schneider

Good afternoon. Thanks for taking my question. You have been very clear about the reasoning model opportunity that you see and you have also been relatively clear about the technical specs for Rubin, but maybe you could provide a little bit of context about how you view the Rubin product transition going forward. What incremental capability does that offer to customers? And would you say that Rubin is a bigger, smaller, or similar step up in terms of performance for capability perspective relative to what we saw with Blackwell? Thank you.

Jensen Huang

Yeah. Thanks. Ruben. Ruben, we are on an annual cycle. And the reason why we are on an annual cycle is because we can do so to accelerate the cost reduction and maximize the revenue generation for our customers. When we increase the perf per watt, the token generation per amount of usage of energy. We are effectively driving the revenues of our customers. The perf per watt of Blackwell will be for reasoning systems an order of magnitude higher than Hopper. And so for the same amount of energy and everybody's data center energy limited by definition, for any data center, that we using Blackwell, you will be able to maximize your revenues. Compared to anything we have done in the past compared to anything in the world today. And because the perf per dollar the performance is so good, that the perf per dollar invested in the capital would also allow you to improve your gross margins. To the extent that we have great ideas for every single generation, we could improve their revenue generation, improve the AI capability, improve the margins, of our customers, by releasing new architectures. And so we advise our partners, our customers to pace themselves and to build these data centers on an annual rhythm. Ruben is going to have a whole bunch of new ideas. I paused for a second because, you know, I have got plenty of time between now and a year from now to tell you about all the breakthroughs that Rubens are going to bring. But Ruben has a lot of great ideas. I am anxious to tell you but I cannot right now. And I will save it for a GTC to tell you more and more about it. But, nonetheless, for the next year, we are ramping really hard into now Grace Blackwell, GB200, and then now 300, we are ramping really hard into data centers. This year is obviously a record-breaking year. I expect next year to be a record-breaking year. And while we continue to increase the performance of AI capabilities as we race towards artificial superintelligence on the one hand, and continue to increase the revenue generation capabilities of our hyperscalers on the other hand.

Sarah

Your final question comes from Timothy Arcuri with UBS. Your line is open.

Timothy Arcuri

Thanks a lot. Jensen, I wanted to ask you just answer the question. You threw out a number. You said 50% CAGR for the AI market. So I am wondering how much visibility that you have into next year. Is that kind of a reasonable bogey in terms of how much your data center revenue should grow next year? I would think you will grow at least in line with that CAGR. And maybe are there any puts and takes to that? Thanks.

Jensen Huang

Well, I think the best way to look at it is we have reasonable forecasts from our large customers for next year. A very, very significant forecast. And we still have a lot of businesses that we are still winning. And a lot of startups that are still being created. Do not forget that the number of startups for AI native startups was $100 billion was funded last year. This year, the year is not even over yet. It is $180 billion funded. If you look at AI native, the top AI native startups that are generating revenues, last year was $2 billion. This year is $20 billion. Next year, being 10 times higher than this year, is not inconceivable. And the open-source models are now opening up large enterprises, SaaS companies, industrial companies, robotics companies, to now join the AI revolution another source of growth and you know, whether it is AI natives or enterprise SaaS or industrial AI, or startups. We are just seeing just enormous amount of interest in AI and demand for AI. Right now, the buzz is I am sure all of you know about the buzz out there. The buzz is everything sold out. H1 Hers sold out. H2 hundreds are sold out. Large CSPs are coming out renting capacity from other CSPs. And so the AI native startups really scrambling to get capacity. So that they could train their reasoning models. And so the demand is really, really high. But the long-term outlook between where we are today, CapEx has doubled in two years. It is now running about $600 billion a year just in the large hyperscalers. For us to grow into that $600 billion a year representing a significant part of that CapEx is not unreasonable. And so I think the next several years, surely through the decade, we see just a really fast-growing, really significant growth opportunities ahead. Let me conclude with this. Blackwell is the next-generation AI platform the world has been waiting for. Delivers an exceptional generational leap. NVIDIA Corporation's NVLink 72 rack scale computing is revolutionary. Arriving just in time as reasoning AI models drive order of magnitude increases in training and inference performance requirement. Blackwell Ultra is ramping at full speed, and the demand is extraordinary. Our next platform, Rubin, is already in fab. We have six new chips that represent the Rubin platform. They have all taped out the TSMC. Rubin will be our third-generation MB Link Rack Scale AI supercomputer. And so we expect to have a much more mature and fully scaled-up supply chain. Blackwell and Rubin AI factory platforms will be scaling into the $3 to $4 trillion global AI factory build-out through the end of the decade. Customers are building ever-greater scale AI factories. From thousands of Hopper GPUs in tens of megawatt data centers to now hundreds of thousands of Blackwells in 100-megawatt facilities and soon, we will be building millions of Rubin GPU platforms powering multi-gigawatt multisite AI super factories. With each generation, demand only grows. One-shot chatbots have evolved into reasoning agentic AI, that research plan, and use tools. Driving orders of magnitude jump and compute for both training and inference, agentic AI is reaching maturity, and has opened the enterprise market to build domain and company-specific AI agents. For enterprise workflows, products, services. The age of physical AI has arrived unlocking entirely new industries in robotics, industrial automation, every industry and every industrial company will need to build two factories. One to build the machines, and another to build their robotic AI. This quarter, NVIDIA Corporation reached record revenues and an extraordinary milestone in our journey. The opportunity ahead is immense. A new industrial revolution has started. The AI race is on. Thanks for joining us today, and I look forward to addressing you next week. Next earnings call. Thank you.

Sarah

This concludes today's conference call. You may now disconnect.