Bank of America Securities 2022 Global Technology Conference

Company Participants

- Jensen Huang, Founder, President and Chief Executive Officer

Other Participants

- Vivek Arya, BofA Securities

Presentation

Vivek Arya {BIO 6781604 <GO>}

So while everyone is settling in, good morning, everyone. I'm Vivek Arya. I lead the semiconductor equipment research team here at Bank of America Securities. And before we get started with the session, NVIDIA just asked me to say that today's discussions contain some forward-looking statements, and investors are advised to read NVIDIA's filed SEC reports for risks and uncertainties facing the business.

So with that, really delighted and honored to have Jensen Huang, the CEO and Co-Founder of NVIDIA, with us. Jensen needs very little introduction, but suffice to say that NVIDIA under Jensen's leadership has been an industry pioneer in driving the boundaries of artificial intelligence, gaming, cloud computing, robotics, I can go on and on. But really delighted to have Jensen with us sharing his perspective. So a warm welcome.

Jensen Huang {BIO 1782546 <GO>}

Thank you, Vivek. It's great to be here. Great to see all of you.

Vivek Arya {BIO 6781604 <GO>}

Yes. So this is actually our first in-person conference after three years. So it's --

Jensen Huang {BIO 1782546 <GO>}

It's my first conference in three years. That's a lot of you.

Questions And Answers

Q - Vivek Arya {BIO 6781604 <GO>}

(Question And Answer)

Yes. So maybe, Jensen, let's start with the high level. So in the field of AI, why is it an important problem to solve? Why is it a hard problem to solve? And why do you think the GPU is the right way to solve it? Because we see a number of other approaches, whether it is custom silicon, right, whether it's FPGAs, right, there are other companies making GPUs also. So why is it an important problem? Why is it a hard problem? And why do you think your approach is the best one?

A - Jensen Huang {BIO 1782546 <GO>}

Is that the one and only question? Because that's going to take 40 minutes. That's an excellent question. It's a very important question. First of all, machine learning, artificial intelligence is a computer that writes software itself. It writes software that no humans can write. And it looks for patterns and relationships from data. And it creates a model that can then infer and make predictions from data that it sees in the future that it hasn't seen before.

And so, if you think about what I just said, the characteristics of machine learning that -- it's a computer that writes by itself, it writes software normal humans can and that can make predictions about the future, you've got to ask yourself how important is that, and how would you apply that, and what kind of problems can you now finally solve the very first time that because we can't write these type of software, there are no principled mathematics, no principled equations to make that prediction, like, for example, there's no Maxwell's Laws, there's no Newton's Laws, none of those laws exist, no Schrodinger's equations. None of those equations exist or can't solve problems at the scale that we can do with machine learning. Maybe it's because it's multiple physics. Maybe thermodynamics and fluid dynamics have to come together to solve a problem. Maybe it's multi-physics, and it's impossible for us to have a simple numerical answer for it. And so for the very first time, machine learning gives us the opportunity to solve those problems.

Now the implications to -- three large domains, I'll just characterize it as three large domains, is utterly profound and you're seeing the benefits of it. One is, of course, information automation, what we call IT. For the vast majority of the industry, the time that we've all been in the industry, and I've been in the industry since I was 20 years old, information technology has been about retrieval of information. If you look at your data center, the vast majority of your data centers, a bunch of computers connected to storage, it's about retrieval. It's about storing files, retrieving files, sharing files, modifying files. It's about retrieval.

For the very first time, we can now take that data and infer some model from it, create some model from it, predictive model from it. And this AI has the ability to go into that data and figure out what are the predictive features, a predictive feature so you guys know, no different than an equation of -- a force and you know that mass and acceleration are predictive features. It's no different than Maxwell's equation has predictive features.

Inside your company, there's a whole bunch of data, and there's no way for you to figure out what the predictive features are. Finally, machine learning comes along and it figures out what the predictive features are all by itself. And it could create a model that could then make predictions from. So information automation.

The second field, which is very important, is science automation. Some people have now made a claim and it's too far that because of AI, because of machine learning, it's the end of the scientific method. I think the scientific method is sound and will continue to exist, but it will be augmented by machine learning. And the reason for that is you're going to observe the world and use AI to go figure out how to predict the future. In the past, scientists used thought experiments. And the thought experiments, Einstein, for example, didn't observe anything. It was all completely based on mathematics. It's based on thought experiments. Sometimes, it's through observation.

And so you could say that when a scientist makes observations, intuitive observations, and then figures out what the numerical-first principal methods are building upon previous science, that method is now going to be augmented at scale. And so, I don't think it's the end of scientific method, but that scientific method is going to get boosted. It's called physics ML, physics machine learning.

The third area of great impact, and you could almost look at the world and break it down into these three areas, and you'll see some really profound work being done. The third is, for the very first time, we can -- we have the ability to write software, have machines write software, that can control really, really complicated things, like, for example, through sensors infer or perceive the environment, reason about the context, reason about the environment and its mission and its goals and then come up with a set of plans. What I just described of perception, reasoning and planning is the foundation of intelligence, is the foundation of robotics.

And so for the very first time, we're going to see, because we have this new form of software, we are going to be able to automate industries, not just information but automate industries. And we can talk about these three areas. So first of all, the profound impact of machine learning is fundamentally that. The second, why is it hard? Well, first of all, the computational method of machine learning, ultimately, the model has -- there are two computers that has to be built. There are two basic processes that has to be built for artificial intelligence.

One is a computer that takes a whole bunch of data and -- it takes a whole bunch of data and find the predictive features and find the predictive patterns and relationships. The relationships could be over space. For example, one pixel to another pixel, computer vision. It could be over time, one sound versus another sound, speech recognition. Or it could be over time and space, video. And so the finding relationship across -- and it could be in the frequency domain, for example, FFTs. It could be in physical domains. It could be information spatial domains. And so, there's a lot of different dimensions that you could learn from.

So I think the first part is you have to create a computer that can sufficiently learn from the data that it's presented, all these different types of data that's presented a predictive model. That piece of software, so that it can learn from the data because it could be -- the number of predictive features could be in the hundreds of millions, whereas F equals MA has two variables in the case of many artificial intelligence. The work that you do for example, if you just think about how do you predict certain things.

The amount of modalities of information that you have to bring in, the dimensionality of the information that you bring in is utterly gigantic. And to figure out how to take all of that information in, create an architectural model that could learn and predict from that data set without being overfit, meaning it only saw -- you gave it an example of a fruit and that fruit is orange, and it thinks that orange is the only fruit. So it can't be overfit on the one hand. And the size of the data implies something about the size of the model, which implies something about the computer.

And ImageNet went from a few million images to now companies are training on hundreds of terabytes, moving to tens of petabytes of data. And so the size of the data, the modality of the data, the dimensionality of the data and the model that you wanted create to predict that data is proportional. So that's the thing -- that's why it's hard. And it's the ultimate high-performance computing problem.

Then the second problem is how do you now deploy that model into a world to make the predictions. In the case of mobile devices, that model is running in the cloud. And almost everything that all of you do every day, whether it's doing search or shopping or video or playing TikTok or whatever it is, short videos, long videos, whatever it is, everything has a recommender system behind it. It's the single-most important economic engine in the world today is unquestionably the most valuable piece of software the world's ever discovered. It's worth trillions of dollars, as all of you know. And that recommender system is running in the cloud.

On the other hand, there's another AI model that is developed in the cloud in the same way that I just described earlier. However, it has to be deployed at the edge. So a perfect example of that would be a self-driving car that has an artificial intelligence model that takes all of the sensor input. It has to make sense of what it sees through a LIDAR or radars or surround cameras, create a world model from it, localize everything around it, localize itself to it and then reason about what it should do based on the mission that it has. And so now that, that computer is at the edge.

Notice the difference between the cloud computer that is doing inference or prediction, making a recommendation for you. Every time you click -- and you might be clicking once every second. In computer time, that's a very long time. On the other hand, you have an edge device like a self-driving car that has to make predictions in completely real time. If it doesn't make a prediction in real time all the time, something terrible could happen. And so I've just described the two ends and how the computing model is radically different.

And then lastly, why GPUs? There's several reasons why NVIDIA GPUs. First of all, NVIDIA's GPU is not a graphics chip. About 20 years ago, we started the journey of making it a general-purpose parallel accelerator. And the parallel accelerator created -- we created a new programming model called CUDA, which is the most successful programming language the world's ever seen. And you could argue it is the only parallel programming model the world's ever seen. It took us about some 20 years to make it happen. And today it's used in just by every field of science. It's in every -- it's offered by every computer maker, every cloud maker. It's used in computer graphics, image processing to quantum physics to quantum chemistry to machine learning to robotics. It's the most popular robotics platform. And so anyways, why us?

The first part is the technology. The technology is obviously very hard. The ability to not just run one or two threads of execution on a CPU but to be able to run, orchestrate, manage tens of thousands of threads in one GPU. And in the case of a data center, one application is running across an entire data center. There are hundreds of millions of threads being orchestrated by this one scheduler. You just got to imagine what kind of a scheduler, what kind of a programming model that can take a problem and break it down into hundreds of millions of little tasks and then orchestrating all of that. And so the technology is hard.

The second is machine learning is a complicated computing problem. It's complicated at the algorithm level. It's complicated at the compiler level. Remember, if you look at a neural network, it looks like a compute graph. Well, it is a compute graph. It's a giant compute graph. It's a compute graph with, well, we're coming up on 530 billion nodes. There are only 7 billion people. So our compute graph called Megatron has 530 billion nodes, and those nodes, parameters in it. And the size of the computer, that size of software doesn't fit in any normal computer. It needs a DGX computer to fit.

And so how do you, number one, write that software? And then number two, how do you run that software? And so the entire computing stack is hard. The computing architecture is hard. Just imagine, we're trying to write a piece of software that is -- has all the characteristics that I just described, has data from all these different modalities. You have to ingest the data. You have to bring it all into system memory. You have to operate on it to find relationships across everything. And so the computation model of the system, if you look at our systems, it's a complicated CPU problem, GPU problem, memory problem, networking problem, system bus problem, distributed computing problem, storage problem. It's a problem everywhere.

And so the invention -- we reinvented the entire stack from the chip to the system to the system software to the compilers, graph analyzers and graph optimizers, all the way up to the algorithms itself. And so I think the answer to your incredibly hard question is that machine learning is the most impactful computing science problem that we've ever challenged. It has tremendous and profound impact on all the industries that we've mentioned. But you can't do that by just designing a chip. You have to do -- you can only do this well by reinventing the whole computing science at the computer stack. And that's what we've been working on the last 10 years.

Q - Vivek Arya {BIO 6781604 <GO>}

Jensen, where do you think we are in that adoption curve? Because other parts of the market that you think are starting to get mature where you might face more competitors. So if you could help us kind of think through where we are in the adoption curve. For example, what we do is we look at supercomputers, right? Every six months, there at that top 500 list, and the number of accelerators there now is close to 1/3 of all the machines. Is that the right way to think about where we are in the adoption curve, looking at what supercomputers are doing? Then where are we in that similar curve for, let's say, hyperscalers or enterprises?

A - Jensen Huang {BIO 1782546 <GO>}

Yes. That's an indicator, but here's the way I would look at it. It took -- there are four major data centers that -- data center classes that all of us know, and there are two new ones that are coming out. The data centers -- can you guys hear me? The gentlemen back there told me to do this. Am I okay?

Q - Vivek Arya {BIO 6781604 <GO>}

Yes.

A - Jensen Huang {BIO 1782546 <GO>}

All right. Thank you. I follow instructions really well. The four data centers -- and they emerged and they came into the world in this way. The first data center was a supercomputing center, right, Amdahl, Cray, so on, supercomputing centers. The second is the enterprise computing data center, IBM. The third, hyperscalers, the invention of Hadoop, in-storage computing, Yahoo!, okay? And then the next one is cloud computing, which is Amazon. Does that make sense? So I just described the early days of each one of the four data centers that exist today. They're all quite large. Each one of them added to the previous data center because it has a different need. It serves a different purpose.

There are two new data centers that are coming in, you can tell that are different than all of the other four. The new one that's coming out is what I call an AI factory. An AI factory does one thing, just like a factory does one thing, it could be manufacturing cars or it can be refining oil or whatever you want to, making chemicals or whatever it is, making plastic, whatever it is. And so that factory does one thing. Data, in this case, data comes in, it gets refined and what comes out as a model. Data is coming in continuously. It's being refined continuously, 24/7, and models are being updated continuously. It does one thing.

In fact, if you look at one of the most popular applications in the world and potentially the most disruptive new Internet application in the world, TikTok. There is a factory that is refining the AI model continuously. It's gigantic, utterly gigantic, potentially one of the largest data centers in the world. We're building many, many of those all over the world. In my opinion, there are 115,000 large factories of traditional industrial revolution times. Now you're going to see 150,000 giant factories, and their job is just to refine data, create models, AI factories. We're in the beginning part of that.

If you look across all the companies that are doing things, and you think to yourself, is this a service -- is this an application that has a continuous ingestion of data and a continuous output of model? Well, we have one. NVIDIA have run some of the largest industrial supercomputers in the world and their AI factories and it completely revolutionized NVIDIA. We ingest data from a fleet of cars. We're processing it continuously, petabytes and petabytes of data, and what comes out is an AI model for self-driving cars. We're doing that in a whole lot of different fields. And so that's AI factories.

And then the last data center, a new data center, is what I described earlier at the edge. Every single factory, every single warehouse, retail stores, cities, public places, cars, robots, shuttles, they're all going to have little data centers inside. They're all going to be orchestrated by Kubernetes. They're all going to be orchestrated from a panel from afar. They'll all run containers. You're going to OTA new containers to it. All of them are going to work together as a fleet that's going to generate the fleet's memory. The memory is going to be constructed into some virtual world model. That virtual world model will be updated continuously. And that loop will just sit there and just run continuously. That's a factory.

Okay. And so you got the AI factory and then you have the edge data center. These two data centers are brand-new. We are in the early phases of the growth of that. In the case of the hyperscalers -- let me come back to your supercomputing question. 30% of the world's supercomputers are now accelerated.

Q - Vivek Arya {BIO 6781604 <GO>}

Top 500.

A - Jensen Huang {BIO 1782546 <GO>}

And the Top 500 are now accelerated, that's the installed base. 80% of the new systems are NVIDIA-accelerated. And so if you look at the installed base number, it is 30%. If you look at the brand-new systems that are created this year, about 80%. So over time, that 30% will get larger and larger.

In the case of supercomputer, it's actually fairly hard. It's easy to move to get into the early parts of it. It's hard to move the rest of it. And the reason for that is if you look at the world's supercomputing centers and the applications it's running, it goes from quantum physics to quantum chemistry to, right, astrophysics to molecular dynamics to healthcare or life sciences, physical sciences, astrophysics to weather simulation. And so the tail of algorithms is really, really long, thousands of applications. And that's why for certain supercomputing centers, you can move fast because they're pioneering ones. For the vast majority of the Top 500, it takes a long time. And we've been at it for 15 years, right? And so now after 15 years, we're at 80%.

Now in the case of hyperscalers, that's a little faster. I think that every single hyperscaler will be GPU-accelerated or will have some kind of accelerator in their computers. And the reason for that is because the vast majority of the hyperscalers write their own software. And we contribute a lot of software to open source. We contribute a lot of software to them. And then they have large IT teams that -- computer science teams that can develop the software.

And so they -- when something new comes along, for example, the two major drivers, data center -- NVIDIA's Data Center business is largely focused on acceleration and largely focused on machine learning. Our Data Center business is strong, And it was strong last quarter. It's going to be strong next quarter. And the reason for that is because we're in an early-adoption phase of machine learning across all of these data centers.

They could absolutely measure their earnings growth by investing in NVIDIA's GPUs. And the reason for that is because all of them have the ability to do this thing called A/B testing. They have a digital twin. There's a digital twin of all of us in every single cloud service. And whenever they train a new model, they could predict would you have clicked on something more with this new model than you did with the last model. And because they know the click-through rate and they can predict from the click-through rate purchasing rate or engagement rate or ad payments, because that algorithm is so clear-cut and so well understood. They have the ability literally to model the impact of certain new enhancements to the AI on future growth.

And so they are very focused on it. They want to enhance the quality of the service. They want to enhance the engagement rate. And at the core of this is this new model called deep learning recommender system, DLRM. It is the single-most important AI model framework in the world. It takes a whole bunch of deep learning models to create the predictive features, and then it creates -- and then -- just giant, giant factory. This is the most valuable computer science project in the world today.

The second one that's most important is natural language understanding, and what is known as a large language model. If you ever get a chance, look up in your Internet, LLM. LLM is a very, very important thing, and it's probably one of the great breakthroughs in computer science in the last three years. Last three years, the greatest breakthrough in computer science ever and very, very important implications to the future of computer -- of machine learning and AI and how machines learn. Imagine a machine that doesn't have to learn at all but has common sense, and it's called zero-shot learning.

There's a whole bunch of new AI technologies. Those two areas are driving just enormous investments in cloud service providers. And so we -- you just -- each one of the data center has its own adoption rate. It's hard to generalize across all of it, Vivek. But here's my prediction, I predict that you'll have data retrieval systems inside your company, and they'll continue to be enterprise data centers like today. But every single company, ours has and every other company will too, will want to invest in an intelligence generation data center. And that intelligence data center will be 100% GPU-accelerated.

Q - Vivek Arya {BIO 6781604 <GO>}

Jensen, what do you see as the next phase for NVIDIA? Because at the last GTC, right, and other events, you have described the move into software and subscriptions. How do you see that evolving for NVIDIA over the next few years? Is it additive to your business? Is it something you already do but you would just be putting it into a formal umbrella? Or do you think it's a brand-new growth opportunity for NVIDIA?

A - Jensen Huang {BIO 1782546 <GO>}

You can characterize everything that we're doing. You can go on that -- you can characterize our strategy in this way. The first thing that we're doing is we have to reimagine the computer -- computing system from top to bottom. We call that full-stack innovation. Chips, systems, system software, algorithms and libraries, and so that's full stack.

The second thing is we're the only platform in the world of any accelerators at all, aside from the CPU. So CPU just does it very, very slowly. In the case of acceleration, we're the only platform in the world that does end-to-end machine learning from ingestion of the data from the actual query of the data from a database to the processing of it called ETL. Anybody here in the field of data science? ETL is fully accelerated with NVIDIA, whether it's RAPIDS or Spark. And then we go into the training part of it to inference.

We are end-to-end. We're the only platform that can train and inference any model that's created. There's a very important contest that's called MLPerf. We're the only company that has ever -- well, we're the only company that finishes the test. We finished the test every time on training. And the number of tests, it's like -- it's harder than the SAT. But the number of tests is quite enormous. We finished all -- we're the only one that submits a result for everyone in the test for data center, for edge, for training and inference. Every single model, every single time. We're the only one that does it -- even come close. And so we're end-to-end, we're full stack, we're in any model. And so that's our first mission, is to reinvent computer science.

The second thing is to put our computing architecture in anywhere that people want to do computing. So I just mentioned there are six different classes of data centers. And they all have different stacks and they all have different needs and different networking, different storage, different provisioning. We're the only company that has the architecture across all six, okay? So that's the second mission.

The third mission is to invest in the reinvention of today's information technology automation, but also build a foundation for the largest of all the opportunities, which I think is going to be industrial automation and putting AI literally everywhere. And so everything that moves will be automated. There's no question about it. It will be safer. It will be easier to manage. And so the thing that we're working on now which is related to Omniverse is that.

We also want to make it possible for companies, whether they have internal computer science organizations and IT organization, large engineering organizations like clouds or enterprise, to be able to adopt AI inside the company the way we describe it is to democratize AI. We want to put AI in the hands of enterprises, healthcare companies, financial services companies, you name it, okay, and retail companies, large logistics companies, automotive companies, transportation companies. We want to put AI in everybody's hands.

The only way to do that is to take software that we otherwise open source or put into GitHub or provide a source to CSPs to package that up into a licensable product. Because most companies don't have the ability to go and cobble all of that complexity together from the algorithm level to the system software level to the networking and storage level. It's just too much. And it needs to be multi-tenant, it needs to scale out, and it needs to be secure. And so it's just too much software to do.

And so we package all that up into NVIDIA AI. We package all that up into NVIDIA Omniverse. And we have an enterprise license. The enterprise license is $1,000 per node. And for us, there are 25,000 enterprises around the world that's already using NVIDIA's technology for AI. We now can offer them an enterprise software license that they could get direct support from us, access to our expertise to learn how to use it and get maintenance and support. And so that licensable products is a new product of ours. It's off to a great start. I think it's going to be probably one of the largest -- my estimation, one of the largest enterprise software products in the world. It has the opportunity, though, to ride NVIDIA's go-to-market. So we have the opportunity not to have to build up a large enterprise sales force because we will go to market on the computing platforms that we already sell. And so I'm very excited about NVIDIA AI.

Q - Vivek Arya {BIO 6781604 <GO>}

Got it. Since we have a lot of investors in the audience, I can't resist but ask a little bit shorter-term question, which is there seems to be this conflict where the semiconductor industry, right, sounds very strong, that demand is not the problem. Supply is the only challenge, right, because of lockdowns or other issues. Whereas the broader market, right, seems to be implying, we are headed into a big slowdown. So what's on your dashboard? How do you perceive the demand environment? And what kind of risks do you foresee over the next four to six quarters?

A - Jensen Huang {BIO 1782546 <GO>}

If the slowdown results in loosening of supply chain, that's good. Our strategy is to grow faster than the economy could be impacted. Of course, China and Russia has an impact on our consumer product -- consumer gaming business. It's impacted our Q2 by about $400 million. China is a significant market. Russia is a meaningful market for our gaming business. However, gaming remains solid even in the face of China and Russia. Q1 sell-through was -- grew year-over-year over last year, which was a really fantastic year. And so gaming sell-through remains solid.

We are working on the single-greatest opportunity in the history of computing. Artificial intelligence has been the holy grail of computer scientists for a very long time. For the very first time, we know how to in specific areas, specific areas of skills and things to automate, achieve superhuman levels, not to mention that, achieve global scale because of cloud computing. The combination of machine learning and cloud computing is really quite tremendous.

We've now succeeded in creating AIs for information automation, recommender systems and language understanding, as I mentioned. In physical sciences, we invented a new physics molecule, FNO, that can be used for weather prediction, one that's going to be used for quantum chemistry. And so we're making tremendous breakthroughs in physics ML and revolutionizing science. And then of course, all of the work that we're doing with Omniverse and robotics for industrial automation. And so those -- that work, I believe, and the foundation of our end-to-end, full-stack capability, we can bring and democratize AI for all of the industries that I mentioned. And with a company of our scale and our footprint and the reputation for being very good at this, I think we have an opportunity for years of growth ahead.

Q - Vivek Arya {BIO 6781604 <GO>}

Got it. I know I won't be able to go through the other 20 questions in the next --

A - Jensen Huang {BIO 1782546 <GO>}

Well, that's -- I will teach you to ask the ultimate first question. Whose first question is why, why and why? What is the meaning of life and what is the --

Q - Vivek Arya {BIO 6781604 <GO>}

So maybe just in the last few minutes, Jensen, so you've recently launched the Grace CPU. If I'm an x86 server CPU vendor, should I be scared? Or should I think of that as well, they're only going after a niche market. What are your ambitions and plans over the longer term? Because I remember meeting the first ARM server company, and they are no longer there. So many people have tried. Why do you think you will be successful this time? And what does the ambition look like?

A - Jensen Huang {BIO 1782546 <GO>}

First, you can only make a computing architecture succeed if you have software ecosystem. Period. It's all about the software. It's not about the chip. There's a really important question about why we do what we do. So let me explain.

If we are a component maker of CPUs, memories, networking chips and storage chips, Wi-Fi chips, USB chips, if we're a component maker, a chip maker, it doesn't matter at all that we have software. It doesn't matter at all that we are full stack. And the reason for that is because it's industry standard, and Wi-Fi is Wi-Fi, 802.11, right? Video chip, AV1. And so you name it, okay, USB 3. So industry standard is industry standard. X86 is an industry standard.

If you build an ARM SoC for mobile devices and embedded systems, you're good to go. The ecosystem is there. However, if you want to build a new chip or a new market that, that architecture has never been, you have no hope without an ecosystem. And so that's number one. If you build an ARM chip for data centers, you have no hope unless you have the data -- you have the infrastructure and the ecosystem.

Number two, why are we building a CPU? We're -- we buy a lot of x86s. We have great partnerships with Intel and AMD. For the Hopper generation, I've selected Sapphire Rapids to be the CPU for NVIDIA Hopper. And Sapphire Rapids has excellent single-threaded performance. And we're qualifying it for hyperscalers all over the world. We're qualifying it for data centers all over the world. We are qualifying it for our own server, our own DGX. We're qualifying it for own supercomputers.

And so we partner with the ecosystem. And I buy everything that I can. As a practice, I buy everything I can. And the reason for that is because I've got smart engineers who wants to invent the future. The last thing I'm going to go do is squander their time and squander their life's work on repeating somebody else's work. And so that is just one of the core values of our company. The reason why the world's great computer scientists want to come work at NVIDIA is because we choose work for them that is singular. We choose work for them that has never been done before.

So Grace is going to be an amazing CPU, and it's not like anything that's ever been built for. It has the benefit of two things. One, it's designed for a new class of applications that emerge in the last couple two, three years, that has proven to be really, really impactful. I mentioned two of them: recommender systems and large language models. These two applications have such giant data spaces that wants to be accelerated, that unless you do something new, you're just going to have lots and lots of bottlenecks or just cost too much. And so Grace is going to help solve that.

Grace has the advantage that in every single application domain that we go into, we have the full stack. We have all of the ecosystem all lined up. Whether it's data analytics or machine learning or cloud gaming or Omniverse, digital twin simulations or in all of the spaces that we're going to take Grace into, we own the whole stack. So we have the opportunity to create the market for it.

And so I think the -- what NVIDIA does for a living, as you know, is not build components that are chips that are industry standard. I think those are all terrific. What we try to do is build platforms that open new markets. Whether it's the market we recently opened up on operational research or the work that we're doing in quantum chemistry, quantum physics, of course, robotics, of course, industrial automation and AI and things like that, in each one these, we create the whole stack so that we can open markets. And those markets are important. We're really passionate about it. We're very good at it, and that's what really makes our company special.

Q - Vivek Arya {BIO 6781604 <GO>}

Great. Terrific. Thank you, Jensen. Really appreciate your time. Really appreciate your insights. Thanks, everyone, for coming.