So, welcome to our live stream audience watching the Arm Everywhere event. I don't think we've done a live stream event like this.
And to the folks here in the audience, thank you so much for coming to the historic Fort Mason.
And you may not know that Fort Mason here in California was actually an official defence site for the Civil War.
And this is where a very famous battle between Alabama, Georgia, and California took place.
Now you're thinking to yourself, I don't remember that battle.
That's why this area looks so pristine.
That actually was not a battle, but actually was a fort.
So I thought that was actually kind of neat.
I didn't actually know that.
So, thank you again for attending. A big day for us. We have a lot to share with you, so I'm going to jump right into it.

When we thought about how to name this event and how to talk about our company, we thought Arm Everywhere was really appropriate because one of the things that we're very proud of, that we don't always think about in our daily lives at Arm, but is really quite impactful, is just the scale of the company and the magnitude we have.

So we start looking at numbers. 117 billion. What is that number? That's the total humans ever to live on Earth. That is, if you count up, by all of our calculations, how many people have lived on the planet since inception. About 117 billion.

350 billion plus are the number of Arm chips to have ever shipped. That is three times the total number of humans who have ever existed on the planet. So it's not just one for every human, it's three for every human that have ever lived.

Seven times the total number of non-Arm-based CPUs shipped, combined. Let's think about that number.

And 160 Arm chips for every global household. Mine is probably larger than 160, but 160 is about the average. So that just gives you a sense of the scale of what we've done, and it's really important because it feeds into everything that makes us what we are today and, of course, could not be done without our ecosystem partners.

Now, the company's DNA was really born to run off batteries. The company started in the early 1990s. It was a spinout of a British computer company named Acorn. And that company had a mandate to build a chip, and that chip had a couple of requirements. One was it had to run in a plastic package, which back then was really important, and number two, it had to be really low power. The first part was important because of the heat. The second part was important because battery life meant everything, because this was going into the world's first PDA.

So the company, we nailed that. We nailed that objective so solidly that, and this is a true story, when the first Arm development board that had the first ARM1 processor was

powered up, and these were plugged in now into the back of a wall, so we had a development board, lots of logic chips, plugged into an AC outlet, when the AC outlet plug was removed, the chip kept running. And the chip kept running based upon the leakage current that was coming off all the other chips on the board.

So the folks came in the next night and they saw the oscilloscope was still driving a signal. And that is really what, for us, launched the revolution of smartphones. We were designed into the very first GSM phone, for those who remember that Nokia brick on the far edge. But then the BlackBerry, which many of us who had loved still love and wish it came back, all the way to the modern smartphones of Android and iPhones. That is where we started in terms of the battery life. It launched a generation of smartphones.

Now, one of the breaks we got about 10 years ago was when SoftBank bought Arm. Yeah, about 10 years ago is 2016. And when SoftBank bought Arm, Masa gave us an opportunity, now that we were a private company, to invest into areas that we were not able to invest in before. And that gave us the opportunity to expand the platform to a number of other verticals.

We took everything that we knew about smartphones and then expanded that out into the cloud. We launched Neoverse. We got our first design wins in the data center. And then we were also able to invest into autonomous, automotive, physical AI. We could not have done that without that 2016 moment. And this is my thank you to Masa for allowing us to do that. We could not have made that all happen. It's paid significant benefits for the company.

However, as good as our products are, as competitive as the platform is for physical AI, for autonomous, for the cloud, it is really what I like to call the ecosystem of ecosystems that really differentiates us. And this is where the partnership really comes to life because that mobile platform that we built cannot happen without the software. And the software layer, in the case of the mobile area, is iOS, it's Windows, it's Android, MacOS, and then the litany of applications that not only run on the Arm compute platform, but they're highly optimized, highly tuned, and allow the partners in the ecosystem to build great products.

That formula applies to every vertical that we participate in. It applies to what takes place in the cloud, whether it's Linux or OpenAI or Anthropic, and then the platform that runs with it. And this is why we like to call this the ecosystem of ecosystems, because it's not just one vertical. And you can see, when we look at the physical AI platform with automotive, same formula. 22 million-plus software developers that are very unique to a vertical, but they leverage a lot across the ecosystem that allows people to get started in other areas.

So this is the magic. And this is what is uniquely Arm. So it's very, very unique about our compute platform. There's no one on the planet who can serve the edge to the cloud in the way our ecosystem does.

Now, over the past few years, we've been evolving our strategies, largely because we see the demands in the marketplace around the chips are more complex. Cycle times to build these

chips are getting longer. Five nanometer to three nanometer to two nanometer. That means longer fab times, longer packaging times. There's a need to do more and to do it faster.

We've traditionally provided IP, IP in a standalone form: the CPU, the GPU, system IP. And that has served us well for the first 30-plus years of the company. But as I said, we were starting to see huge demand for the need to go faster, make products better, and get time to market sooner. And we introduced something called compute subsystems. We did this about three or four years ago. We invested very heavily in terms of the engineering requirements to do this.

And what this does is it takes all the blocks of IP and pushes them together in a finished way, verified, performant, tested, that the end customer can then take to market. And in some cases it shaves a year, in some cases 18 months, off the time from starting design to getting into production.

It was a very significant investment for us. We put a lot of effort and engineering into it. But we've already seen massive benefits in terms of the customer base. We introduced this three or four years ago. Our business model is a license-plus-royalty. Royalty is the laggard, so royalties start to show up two to three years after we license the product. Already, CSS represents almost 20% of our royalties and growing.

Now that's our evolution.

Of course, we're now in an era where everything is different than we knew it before. And when I think about artificial intelligence, and I get a lot of questions when I talk to analysts or media about, did AI just come up on us by surprise? And I think back to a time when I was in Bletchley Park about a year and a half ago. Bletchley Park is where the original crypto work was done by Alan Turing to help the West against the Germans in World War II.

There is an area there where you can go in the museum and you see papers from Alan Turing about, can machines think? And those papers were written in the 1940s. So the idea of AI is obviously not new. And if you're a sci-fi fiction fan, I certainly was growing up. Arthur C. Clarke was one of my favorite authors. 2001: A Space Odyssey. Now, the people who were even born in 2001 are here. I always looked at this and said, of course this is going to happen. I just didn't think in my lifetime I would see it at the pace that we've seen it.

And for anyone who says, you know, this is a bubble and it's going to pass, it may be a financial bubble in the case of investment, may slow down, and maybe an investment bubble in the sense of the valuations may not be what they are today tomorrow. But if anyone thinks that this is something that is going to go away, it's a little bit ostrich syndrome. This is here with us. And it's really changed how people think about computing.

However, somewhere along the way, people kind of thought CPUs were dead. And there was a thought that the only way you handle AI is through accelerated computing, that the CPU's role in the AI world is no longer relevant.

Now, if we think about the role of the CPU and what happens in the cloud, now this is the cloud before AI, so I'm going to say it's before that last slide that I showed, huge growth in compute cloud. We saw growth from AWS, Microsoft, GCP. And the conventional use of the cloud was: you type in an answer, you do a search. Any seats left the Warriors game? I think there are a lot of seats left for tomorrow's game, by the way. Or tonight's game. You got the prompt back. This is the cloud. Very simple, you do the search. But CPU is very heavy. So when we look at the growth of SaaS 10 plus years ago,10-15 years ago, and all the growth around cloud, the CPUs were doing literally all the work. Now, when you add the AI cloud if you will, and now you are human, you are putting in a prompt, into your device whether its your phone or your PC, of course there are still CPUs involved. The cloud is servicing that request, and that request then gets sent to a token which the accelerator generates, and the CPU in that data center orchestrates and sends the token back. The token being a word or the answer that provides request to the query, so this is all the work that's being done by the AI data center.

CPUs are involved, both in the cloud and obviously they're involved in AI data center. And we estimate that in this data center, there's probably 30 million CPU cores per gigawatt, so there's a lot. Data center here is a combination of what sits right in the AI cluster, whether it's your head node to your accelerator, or what sits next to a dedicated rack. But the math is basically about 30 million CPU cores per gigawatt. OK. And that is the world that we've seen coming up to about the last year or so, or maybe even less.

And what has changed in the last number of months has been this explosion of agents. Agents are essentially tools that act on a request and come back with a full flow of answers. So it's not just a query for an answer, but it's actually work. It's run a payroll task, do a scheduler, go off and write a number of analysis relative to a tool and provide me an answer. And we heard so much about OpenClaw here in the last few weeks, as an example and its not the only example.

Now why is this important? Why am I talking about this? Because as we move to agentic AI, the number of tokens per human go up 15x, if not greater. And if you think about the why of that, it's pretty straightforward. Agents can generate requests far faster than humans, and they don't sleep. They're at it 24/7. So the agents are now pushing these requests into the cloud, into the data center. And what's happening? The data center is choking.

These accelerators, which are very expensive, that generate the tokens, now need to send those tokens back through the cloud. Now, if we think about what an agent is, an agent is a workflow. As I said, it's a payroll task, scheduler task. It's asynchronous. It is a lot of work relative to scheduling. That's what CPUs do. That is what CPUs do. That is not a work that can be done by an accelerator. The way to think about this is the accelerator generates the tokens, but its almost like pushing a dump truck up, and someone's got to move that dirt. The CPUs are the pieces of equipment that move that dirt, and agentic AI only increases that. So what you see is a huge bottleneck now in terms of flow.

So what does that mean? You need more and more CPUs. Lots of them. CPUs near the head node. CPUs next to the accelerator rack. More CPU racks inside the data center. You just need more. And by our calculations, we think this may be a little bit light, it goes up about 4x: 120 million CPU cores for that same gigawatt. OK, so in that same profile we now need 120 million CPU cores.

Now we're trying to put four times the amount of CPU cores in that same power envelope. Power is precious, obviously. The capital required for it is precious. So trying to put all those extra CPUs into a data center that is already stuffed to the brim with accelerators and CPUs doing the core work, that is a problem.

Now every tough problem needs a good solution. And we're announcing our first silicon chip that we are selling to our customers for revenue. The Arm AGI CPU. This is a big, big deal. And I would love to tell you every feed and speed about the product right now, but Mohamed will kill me if I do that. So we will go into a lot of detail about the product and how we conceived it and the why.

But let me be clear. We are now in a new business for Arm. And we are supplying CPUs as chips.

The biggest reason we're doing this is that our partners have asked for it. But we're also really doing this to solve the problem I just described. As agentic AI becomes mainstream, all of the work required to make that happen is CPU-bound, and you need a CPU that has the DNA of being born to run off a battery.

So as I said, reason zero is our partners have asked for it. And one of the partners that we've worked closely with is Meta. And I'm super pleased to have Santosh Janardhan with me today, who's going to do a better job than I can to tell you why Meta has made that choice. Santosh. Please welcome to the stage Meta's Head of Infrastructure, Santosh Janardhan.

Hey folks, welcome.

It's funny, every year I try to run the San Francisco Half Marathon, and they distribute the bibs the day before you run right here. I can tell you it looks very, very different compared to what I guess we're seeing now.

So hi, my name is Santosh Janardhan, and I lead infrastructure at Meta. So what does that mean? Well, it means that we traditionally go and custom-build and design our data centers, run them, we custom-build our hardware, or parts of our CPUs, and we'll get into that quite a bit, the network that connects them, and obviously the software that sort of binds it all together. It's a fancy way to say that if your Instagram is not working, if your WhatsApp service is not working, your message is not arriving, I am the person to blame.

Now, if you think through our family of apps, that amounts to about three, three-and-a-half billion users that use our products daily. Every single day, about half of humanity logs into one of our apps and hammers away at it. As you can imagine, that creates a decent amount

of scale. We run a decent amount of the internet. And we're probably the only hyperscaler that's not a cloud, right?

So if you think about gigawatts of capacity, tens of millions of servers, and increasingly more and more you're seeing bigger and bigger CPU and GPU AI clusters, Rene went through that quite a bit. I think it's interesting to go and look at how this has grown over the last years.

AI clusters are a fairly new thing, really started sort of post-COVID, 2022-23, just after ChatGPT came along. And initial clusters were pretty small. In fact, when I looked back for this, in 2023 our initial clusters were about 128 GPUs. That's it. But as you can see, even in 2023 we started scaling quite a bit, and as you fast-forward, it really started growing. The demand for this has far surpassed what any one of us could imagine it was. We are in the tens of thousands of GPUs stitched together in a single cluster now.

And if I project it forward, and this is the thing I really want to set context on, there is absolutely no sign of this slowing down. In fact, it's almost exponential. Yeah, I only see it accelerating. Right? So the demand is exponential, and as Rene was saying power is constrained.

I want to talk a little bit about some of our clusters. That is Prometheus. Prometheus is one of our bigger clusters. It'll surprise you, valuable gigawatt by the end of this year. There are a lot of GPUs, I can tell you. And we stitched together a bunch of data centers, a bunch of tents. That thing you see, the blue-colored thing, is actually a tent. It's a fancy tent, but still a tent, right? It's weatherproof. It can survive about a Category 2 hurricane.

But we're putting together all of this, stitching it together with a network, and so to our developers, to researchers, what they end up getting is about a gigawatt of AI cluster in a single combined entity, which is pretty powerful as you can imagine.

Like I was saying, the demand is exponential. To put it mildly, that is Hyperion. It is going to go up to five gigawatts in a few years. Most people can't fathom what a gigawatt is. A gigawatt is about 10 Palo Altos. The town of Palo Alto, 10 times what it consumes is one gigawatt. This will be five. That's 50 Palo Altos, right? That's what we're building out. So it is going to go really, really big.

So why do we do this? At Meta, we have this vision of delivering personal superintelligence for every single one of our users. This means creating models that can go and figure out the most relevant experience, the most engaging experience, for every one of you on the platforms. It means creating a personal assistant for every one of you.

Right, now if you have to go and deliver personal superintelligence to billions of people, what kind of systems would that take? You're talking about billions of people, each using an exacting mode of compute over and over. Like I said, over 3 billion users a day, right? This - if this advances, so what does it take? Well, it takes power. It takes land. It takes a decent amount of hardware, software obviously, and most of all, it takes silicon. A lot of silicon.

And this is why I think Arm is such a natural partner for us. What we want is a partner who can match our ambition, who can match our cadence of velocity of innovation. And what we realized when we were sitting down with Arm is that they could develop it. They were as hungry as we were. And most importantly for us, they were as power-conscious and as efficient as we wanted them to be.

This is why Arm is now the primary collaborator and the primary partner. CPU that we are ending up developing, it's pretty foundational. It's not just a Meta CPU. It's not just an Arm CPU. This is something that I think will end up being a foundational CPU for the whole ecosystem.

I think we're at the threshold of something pretty sweet here because you're going to hear more and more about the constraints that data centers are facing. You're going to hear more and more about why the demand for compute is growing while the power is not going in the exact same curve. So this marriage is, I think about it personally, as a win-win situation, right? So it's extremely heartening to see Arm moving on from not just being an IP license provider but actually getting into the game of building something that is production-scale and production-ready. Exciting times. Probably two years, three years in the making, but I think about this as the sweetest things take some time, but we're getting there.

Now, like I said, we're obsessed with efficiency. And if you think about one of the biggest appeals that Arm has had over the years, it's power profile. Rene had this fascinating experience that he was talking about, taking 30 million cores instead of 13 – now making it 120 million and fitting in the same power envelope. But there's one thing you don't want to compromise, and that's performance. Right? This is the thing that I really want to make sure we drive here.

The biggest reason why we sat down with Arm and had this conversation was we want to put in a lot more cores per watt, but we do not want to compromise on the performance space. That marriage is why I really think it's a win-win situation here.

In fact, about two, two-and-a-half years ago, we sat down with Arm. We actually first surveyed the market to see, was there a CPU that could meet the specs that we wanted? If we met the performance, we couldn't get the power. If we got the power, we wouldn't get the performance. And this is why it ended up being Arm such a partner. The ability to scale that Arm gives us when we push lot more cores in. If you think about personal superintelligence, if you think about the orchestration that Rene showed, you don't want to starve your CPUs nor do you want to starve your GPUs. That marriage you end up doing is I think the most people are going to realize pretty soon.

Now, the design point that we chose for this was something to minimize risk for this iteration. We wanted to make sure we get the first CPU right, get it working out of the box. But this is a multi-generational partnership. I just want to emphasize this. When we look at subsequent iterations of things that are already in the hopper of what we're going to build

out, I truly believe that this chip is going to expand the performance on multiple axes. In fact, in this ecosystem it is actually going to be awesome. When you challenge the incumbents, you see innovation across the board, and that, I think, is what all of us will end up achieving.

Now I want to talk about why. I want to take this back, I guess, to why we do this work. At Meta, three, three-and-a-half billion people use our products every single day. This means that your friends are messaging each other. It could be a small or medium business messaging the users on the platform. It could be somebody doing an AI interaction with Meta AI. None of this is possible without infrastructure.

Infrastructure has now gone from being on the backside of technology innovation to being the enabler of technology innovation. AI is built on the backbone of infrastructure. So every interaction, every post, every feed, every call is done on the basis of what we build out on the back end. And at least for us, we custom-build data centers, custom-build hardware, and custom-build silicon. That's why Arm, I think, is such a big partner for us because, for us, we want to squeeze every bit of performance out of what we build out. We think about optimizing things like performance per watt, performance per gigawatt. And Arm allows us to do that. It allows us to increase the efficacy of everything we build out. Why? So that we can serve more users, so that we can hopefully improve every one of your lives in some way, shape, or form.

And that's why Arm has been an awesome partner. So thank you, Rene and team. It has been absolutely a pleasure to work with you, and hopefully we'll do this for years together. Thank you.

Wow, amazing! Santosh, thank you. That was terrific.

I have someone else I'd like to ask to join us, to also talk about how they plan to use our Arm AGI CPU, and that's Kevin Weil from OpenAI. Kevin.

Kevin thank you for joining us.

Thank you for having me.

Welcome to Fort Mason. Have you been here before?

I have! For a few conferences in the past.

So first off, just tell us, and tell me why does this launch today matter to OpenAI?

Well, I thought you did a good job painting this. AI performance these days is system performance. And GPUs kind of get top billing wherever they go, but really the CPU is playing an incredibly important role as an orchestrator. But also, I think as AI becomes more agentic, when you look at a rollout that an agent is doing, it's using tools inside containers. That's CPUs. It's running Python scripts as it performs skills. Those are CPUs. So the CPU

plays an incredibly important role, and it's really the whole system together that makes this all possible.

Now your role in OpenAI is pretty cool, right? You're doing math and science and the stuff that's super compute-heavy. And when you think about compute constraints, and I know when I talk to you or Sam or Mark or anyone at your company, it's, I need more compute.

Yes.

Tell us about that.

That is one of the most common things I hear inside OpenAI, I need more compute. It's kind of the coin of the realm.

I mean, the root of it is, we have more demand from customers. We have more ideas internally that we want to experiment with. We have more things that we want to do than, frankly, the industry can keep up with. And when you get to the bottom of all this, it's certainly about silicon, but it's also about power.

And so if you have a CPU that can draw less power, can be just as performant, but use less power, it means you have more leftover for everything else that you want to do. That means more inference and more compute. That means more intelligence.

And if there's one thing that I've learned in my couple years now at OpenAI, it's that more intelligence leads us to be able to build better products for all of you.

The thing that I keep coming back to, that I try and remind myself of all the time, is as amazing as the models are today, and every year I'm blown away by the amount of progress we make, as amazing as the models are, the model that you use today is the worst AI model that you will ever use for the rest of your life. It's the worst AI model you're going to use for the rest of your life.

And a year from now, you're going to be, you couldn't imagine coming back to the AI models of today because they're getting better at such a rapid pace, which just means there's basically infinite demand for intelligence. So we're not stopping from here.

And in your world, in your new role, where you're looking at verticals that are somewhat untapped today, math and science and things of that nature, when you think about the Arm AGI CPU or, more broadly, what does more compute do for you in that space?

Well, I mean, the more compute you have, the more inference we're able to do, the longer the rollouts you're able to do. AI, as we're progressing from this world of AI as chat to AI solving harder and harder problems, and just like you or me, when you solve harder and harder problems. And just like you or me, when you're solving harder and harder problems, you're going to need to think a little bit longer.

So the more important problems we solve, as we start to think about things like enterprise, AGI, science, you're going to need more compute, which means if you can draw the power that you have, which will always be finite, you can draw that more efficiently, you can do more, we can solve more problems.

And for you personally, what are you most excited about, broadly, in terms of everything we see going on with AI?

Well, I kind of think I have the coolest job in the world. I get to work on accelerating science with AI.

And you've seen sort of a revolution in the past even just three months with GPT-5.2, 5.4, Codex. I mean, it used to be that people said, oh, well, these are just stochastic parrots. You know, they're sampling from a distribution of data that they were trained on, but they can't do novel things.

Now we're seeing every day AI solve open problems in science and mathematics and physics and biology. We're seeing AI help us understand the nature of the universe. We're seeing AI work for weeks on end, using a robotic lab to run 36,000 different experiments to optimize the synthesis of a new protein faster and better than any human could.

So it's an exciting world. I think science is going to move faster than ever, and it's all built on the kind of infrastructure that you're providing here.

We are grateful for your support, Kevin. Thanks.

Thank you so much.

Thank you.

I love the idea that the model we're using today is about as bad as it's going to get. That's crazy.

I want to repeat, in case I wasn't crystal clear on the first go-round: we are now delivering IP, CSS, and chips. IP, CSS, and chips. Contact your local sales representative. Will is here. He can be reached afterwards.

Now seriously, I talked earlier about the ecosystem of ecosystems, and none of this could be done without the ecosystem that we have, particularly around Neoverse. We have many partners that we work with on the supply side, whether it's around memory or connectivity. But we've also got great customers who use our IP today, and they are so supportive of what we're doing.

Santosh talked about the demand. The market is so large. The demand is so significant that no one company can serve it. So what I'd like to do is, rather than me going on and on talking about it, have you hear from some of our partners and friends, who I think you'll probably recognize a few of.

Rene!

Congratulations on launching Arm's first data center chip.

Congratulations to Arm on the launch of the AGI CPU.

Congratulations on the launch of the Arm AGI CPU.

Today's announcement of the launch of the Arm AGI CPU is a significant milestone in AI optimized compute, and for ecosystem.

Congratulations to the Arm team on the launch of an incredible milestone for the ecosystem built on innovation, scale and openness.

The continued growth of the Arm ecosystem with its AGI CPU is a significant milestone

In continuing to bring customers the flexibility to optimize for their specific workloads

And ensuring accessibility of a

New generation of purpose built compute.

AI systems are evolving rapidly.

They're becoming more autonomous and more data intensive.

And that means performance is no longer defined by compute alone,

But by how efficiently compute and memory work together.

Arm's latest platform opens new opportunities for the system

Level innovation across compute, memory and storage.

We look forward to continuing our partnership

With Arm to advance next generation

AI platforms and ecosystem.

We are proud to partner with Arm

In building this open, scalable, power efficient AI future.

Accelerated computing didn't make CPUs irrelevant

It made them essential partners.

Arm architecture has become foundational across

All of our platforms, from Jetson, our robotics system

To Drive, our autonomous vehicle system,

Our data processing units called Bluefield

To Vera, our CPU.

Without the ability to mold and shape and modify

The Arm ecosystem and the Arm platform,

Its impossible for us to build these systems that we build.

And so Arm's adaptability, modifability, if you will,

Really has made it possible for us to integrate

Arm across all our platforms.

This collaboration with Arm has been great for both companies,

And Graviton continues to provide better price performance

For AWS customers.

We see the AWS-Arm partnership contiuining to deliver

Big for customers.

We are excited about the opportunity this creates to expand

Arm AI and data center ecosystem.

Our Azure Cobalt 200 CPU built on Neoverse compute subsystem

Is an important part of how we optimize and accelerate

Every layer of our stack.

Deeply integrating cloud-native capabilities and delivering

The best price-performance and efficiency across our fleet.

Partnership with Arm is a key part of that vision.

We've been a long time and early adopter

of Arm based systems and a strategic partner in advancing the ecosystem.

Having a diverse portfolio of Arm silicon and software,

Gives OCI greater flexibility and differentiation.

Through deep comprehensive collaboration in memory,

Foundry, SoC design and advanced packaging,

I believe Arm and Samsung semiconductor can deliver

Exceptional Arm based AI CPUs worldwide.

Google is proud to partner with Arm to support

Organization’s most demanding cloud-native and AI workloads.

This builds on a long and productive relationship between our two companies,

Working together across standards, technology, and products

To help deliver the infrastructure for the world’s

Leading data centers.

We’re proud to partner with Arm, and we look forward

To building the next generation silicon for AI infrastructure.

We look forward to continuing our partnership together

And to helping customers and users everywhere

Build AI systems that are smarter, faster, and more scalable.

That’s why our long standing partnership with Arm is so important.

One seamless platform from cloud to edge to AI factories.

We look forward to building this future with you.

Congratulations on bringing Arm’s first data center chip

To market.

Charlie, Charlie and Matt and Sanjay and even my old boss did better than I could in terms of talking about this, but this does not happen without fantastic partnership and support from the ecosystem.

Now, I know you are dying to hear about this product, as am I, and I'm now going to turn to Mohammed Awad, who's going to tell you all about the Arm AGI CPU and why it is absolutely amazing.

Mohammed.

Please welcome to the stage, Arm’s Executive Vice President, Cloud AI, Mohammed Awad.

Thank you. Wow. Thank you, Rene. Thank you, Santosh. Thank you, Kevin. Thanks to all of you. Thanks to the entire Arm team that made today possible. We've been looking forward to this, and it is so exciting to be here. It's so exciting to talk to you guys. Thank you, thank you, thank you.

Rene talked about how the world is transitioning from sort of legacy data centers to AGI data centers to agentic data centers, heading down this path, and how the CPU is at the heart of it.

We've designed our AGI CPU around three simple principles. We believe that the heart of what we're doing is the heart of what we focused on, is the heart of how we think about it.

First: performance. Performance, performance. This many threads going on, with this much work to do, with this much orchestration to happen, you can't slow down. Twenty-four hours a day, as Rene said, these agents are going to be running. And if they're not performing fast enough, then the rest of that infrastructure that's relying on it grinds to a halt. So we focused on performance.

Second, we focused on scale. The scale of what we're talking about here is just incredible. You heard Santosh talk about gigawatts. Gigawatts. Scale at the CPU level, scale at the board level, scale at the rack level, scale at the warehouse level, all the way up. We focused on that.

And finally, we focused on efficiency. Maybe most importantly. Because at the end of the day, with this much at stake, with this much compute we're trying to deploy, we're not going to get there unless we provide that performance, we provide that scale, and we do it in an efficient package.

Those are the principles that have guided us. Wait for it. Those are the principles that have guided us, and we refuse to compromise. We've designed on all three.

Play the video now.

I got to tell you. We are so, so proud. Our team has done a fantastic job on this, and it's really been designed from the ground up for this.

Let me tell you a little bit more about what you just saw, because I know there was a lot packed into that video.

Arm AGI CPU starts off with our standard Neoverse V3 compute subsystem. That's the same compute subsystem we make available to the entire ecosystem, and we have other partners building on it. Incredibly proud of that.

We packed in 136 of those cores, which are very high-performant cores, designed to be high-performance. Our V-series is our most performant line, and you've seen it set records across lots of different hyperscale implementations and those of other system providers. We add to that a dedicated 2 MB L2 cache, and we support up to 3.7 GHz in frequency.

But it's not just the CPU core. We thought about the entire system. As part of the design, we went with 96 lanes of PCIe Gen 6, which supports CXL 3, which means you can attach it to any accelerator you like. It also means that you can support things like memory expansion.

On the memory side, DDR5, up to 6 gigabytes per second of memory per core, which can be sustained to each core. That is unique. That level of performance to every single core on both the I/O and the memory is unique to us in this type of a package, at this type of performance point, at this efficiency level.

And it's not just about the bandwidth. It's not just about the I/O. It's about the overall design. You see, we designed the whole thing to be low latency so that you could get to less than 100 nanoseconds of latency from the memory. We did so by sticking with a dual-chiplet design, each chiplet having all of the memory and the I/O directly on it, rather than having to worry about complicated NUMA domains and multiple hops across the silicon.

The result? It wasn't a typo in this slide: 300-watt TDP. Three hundred watts. That is amazing. It's built on a 3 nm TSMC process and allows for that maximum compute density. This is what purpose-built design looks like. This is what we're so proud of.

This CPU is breaking records all over the place for performance, for scale, and for efficiency. You saw some of that in the video.

This is a standard OCP air-cooled rack. Nothing unique about it. Nothing especially exotic about it. Just OCP rack standards, right? That's our Head of OCP right there, clapping, just so everyone’s aware. Thirty-six kilowatts. We pack in over 8,000 of these performance CPU cores. We do so by going to a 2-node 1U server. Thirty of them. You can't do that in other systems because the power consumption is just too high. This is setting records for air-cooled.

But you know what? If you want liquid-cooled, we can do that too. Over 45,000 CPU cores in a 200-kilowatt rack. Again, a standard rack from OCP. Over a petabyte of memory in this thing. And, oh, by the way, fun fact on this one, it's a 200-kilowatt rack. We actually consume about half that much power. We ran out of space. That's why we couldn't put more cores in that.

Yeah, it's pretty wild. The scale of this stuff is crazy. It's just really inspiring. These are standard racks, but there's nothing else like them.

To get to this level of efficiency, we really had to design the Arm AGI CPU from the ground up. And that's why I'm so proud about it.

But before I get to that, I want to just talk about the fact that these are standard racks because it's not only about the fact that we're leveraging OCP and using some of their platforms, we're also giving back. We're in the process of making a bunch of contributions to OCP, things like Arm ServerReady, authenticated access control, and diagnostic tools. And those contributions won't just be for the Arm AGI CPU, they will apply to the entire ecosystem. So what we make available will be beneficial for all Arm-based platforms because it really is an ecosystem that we're building here.

You know, Arm has always been about nurturing and partnering with the ecosystem. That's always been our core identity. And those relationships are paying great dividends now. You saw the video that Rene played, and we're so grateful about all those partnerships.

You know, it's those partnerships, actually, which have allowed us to build the Arm AGI CPU. Some of them are very longstanding partners like TSMC and Samsung and Micron and SK

hynix. These are partners that we've been working with for well over, for decades, literally for decades. And we've also got some new partnerships, which is why we're so proud to say that the Arm AGI CPU is available now.

It says it there, it doesn't say it there. Yeah, so available now.

Arm AGI CPU is available now, and we're so proud of that. It's actually in customers' hands. Customers are actually evaluating it as we speak. We are ready to go, and we're so grateful for our partners, both on the ODM side, on the memory side, on the CPU side, on the manufacturing side, who have helped us get to this point. We'll be in production by the end of the year, and we are excited to share that with you.

We've got firmware ready to go. We've got specifications ready to go. I talked about platforms. I talked about supply. The one thing I haven't talked to you about yet: software.

Let's talk about software. Now the next slide.

So, you know, the reality is that Arm has been investing in the data center software ecosystem for well over 15 years. I don’t know if everyone understands how long we've been investing in the software ecosystem. For the beginning of that time, in the early days, it was just Arm investing in the software ecosystem.

And then something happened in 2019. We launched Arm Neoverse. And what Arm Neoverse did, that compute platform, when we launched it, it allowed our customers to begin to launch products with a much lower barrier to entry. It allowed them to build their own silicon and start to coalesce around a common platform. And that started that software flywheel turning.

You see, when tech leaders started adopting Neoverse, they started optimizing software around it. And the more of those tech leaders that adopted Neoverse, the faster that flywheel started to spin. Today we've got AWS and Google and Meta and Microsoft and Oracle and NVIDIA all investing alongside us in the software ecosystem. And that really was what allowed us to kind of really make some great traction in software.

Together, we've made Arm a first-class citizen on most modern software packages. And for our AI software ecosystems specifically, not only are we a first-class citizen, not only does software run well on Arm, software actually runs best on Arm. And the reason for that is very simple. For AI, the Arm software ecosystem, the Arm architecture, is the primary CPU architecture in support of AI today.

In fact, the work we've done together with technology leaders means that tens of thousands of companies today run their software on Arm in the cloud, on over 1.25 billion Arm Neoverse cores, which we've already shipped into data centers around the world. And that growth is only accelerating. That's actually the curve.

Arm in the data center just works. This is a key point. And I don't know if I'm making it well enough, so I'm going to bring somebody on stage who's got a little experience with software.

Paul Saab has worked on Meta's infrastructure for over 18 years. He's one of the longest-tenured employees at the company. There's a laundry list of things that he's been responsible for, including the adoption of flash storage all the way through the implementation of IPv6. Today, he is specifically focused on making AI more efficient in their infrastructure, and that's how we got to know each other.

Please welcome Paul Saab.

Thank you.

Thank you. Thanks for being here.

So thanks for being here. You know, you've told me the story before, but I really want to hear it. You guys have had a long history with Arm. It goes back longer than just a couple of years ago. Can you maybe give everybody a little bit of a history lesson on how this kind of thing started?

Yeah. You know, I think it was like 2014, 2015, we were looking at Arm. We were really excited about the efficiency gains that we were seeing. We were really back then just targeting our Hack/PHP platform called HHVM. And it was working great. We made it work. It was performant. And then the market kind of went away from us. We didn't really have a platform anymore, and so we just sort of tabled it. We ripped all that code out. Everything in the code base was removed.

Oh geez. OK, so that was 2014 and 2015. Obviously something must have changed or you wouldn't be standing here today, right? So kind of where did we go from there?

Well, the story is kind of funny. It was like post-COVID, we were just coming out of the COVID bubble, and we had a bunch of people over at the house, you know, sitting around socializing and whatever. And I turned to one of my colleagues and I said, hey, I want to boot to Arm again. I kind of had this gut feeling that the ecosystem and the world had changed. And if we didn't start then, we would kind of be playing catch-up when it actually happened.

I didn't even ask my boss here for permission to buy these machines or even start the project.

It's a good thing he approves now.

I don't really ask him permission for much to do.

But yeah, so we started. We found some machines out there. I went to some other colleagues. I said, hey, I want to do Arm. And he actually responded, I was wondering when you were going to ask me. So we got the machines in, started porting, making great

progress, but it was super slow. We only had eight machines, but we had this vast x86 ecosystem.

And I went to the guys and I was like, hey, can we cross-compile? And that's what we ended up doing. We ended up working around the clock. It took us about 90 days, five engineers, and we had a full complete port, full system ready. But then we ran into another problem. We had no silicon to buy.

And this is, you know, Santosh referenced this, that we looked at every partner, and I think this is about the time you and I started talking.

So you'd say the market was a little bit underserved, maybe, for what you guys were looking for?

I think underserved is an understatement.

But let's go back to that. Let's go back to the 90 days, five people. I mean, really, it's OK, so I'm going to take your word for it, it was 90 days, five people. But that's just getting the code working. Now you've got to operationalize and get it performant. How's that going?

It's still a small team. I mean, it's a lot of very devoted people bringing the systems up. From the time we finished that initial port in 2022, it took us about two and a half years to actually get some sort of production-worthy systems in that were TCO-effective, performance-per-watt. And it was still a very small team.

And even today it's really a small team that's focused on hyper-optimizing. It started off with, you know, once those performant systems landed, it was really just one engineer until a few more came in. But that engineer had never written a single line of NEON, never written a single line of SVE, and single-handedly took some of our most precious workloads and made them work on Arm.

And how's it performing now generally, like on typical workloads? How should we think about the performance in the general case?

We're seeing performance that is basically equal to anything you combine that work with today, at massive performance-per-watt improvements.

That's great. That's great.

OK, my light's going to start blinking in a minute here, so I'm not going to keep you on stage too long. But first of all, I want to say thank you. But before I let you go, I guess one question for you. If somebody out there is thinking about, hey, because there are 10,000 companies that are using Arm already, but there are still a few that aren't, what sort of advice or guidance or recommendation would you give to them?

I think, you know, small focused teams doing the port. But if I were starting the port today, I would be using an LLM. I mean, what I'm seeing now, some of the engineers that are now

optimizing even existing Arm-accelerated code, they're using LLMs to even boost those by 10 or 20%.

So the barrier of entry today, porting to Arm, is I would say close to zero, because the LLM is just going to do it for you. I don't even write any handwritten code anymore myself. It's just all LLM, all test cases, all across the board. So there's no excuse to port to Arm today.

Excellent. Thanks Paul.

That was inspring. I mean, you know, Paul and I have obviously known each other for a little while, and the tenacity, what I hear around, is once Paul gets something in his mind, it just kind of happens. So I appreciate all the support, Paul. Thank you.

We're so proud of the partnership that we've had with you and with Meta more broadly. So thank you very much.

You know, what I love about that story is that they had a need. The market was underserved, and together we worked together to go address it.

The reality is the opportunity for the AGI CPU is broad. The software is ready, and we have a great product, and that's why we're seeing such great customer traction. We're seeing it in multiple areas.

If you think about companies like Cerebras and Positron and Rebellions, they're joining Meta and OpenAI by using Arm AGI CPU for things like managing head nodes that they're building, or managing accelerator buildings, a head-node-type use case, or also for agentic orchestration and fan-out. These are specific use cases that they're looking at.

And then in the cloud, we see companies like SAP and SK Telecom and Cloudflare who are actively using or planning on deploying Arm as part of their infrastructure. These are just a few of the customers that are planning on using Arm AGI CPU.

But rather than me tell you about what they're doing, let's listen to them.

Arm's been building IP for hardware for generations.

Over the last decade, built processors for heavy weight compute.

Cloudfare is one of the largest networks in the world,

And Arm is a really important part of our ability

To keep innovating,

Not just as the speed we've always innovated at,

But at the speed at which what s been going on.

Arm has consistently delivered strong price performance,

predictable scalability and excellent power efficiency.

How can we drive more energy efficiency at the end of the day

and more scale to what we are running and deploy?

Also having an eye on from a price performance perspective

on cost reduction, overall.

One of the key criteria for our customers is

the best use of the limited energy available for them.

Getting the technology that gives them the best outcome per

Watt is a critical

Data centers have an obligation in the AI ecosystem.

We use a lot of power.

This is where Arm has been a leader historically,

and Arm AGI really separates itself.

AI has fundamentally turned our business upside down.

It grabbed all the available data center capacity

To the point where they re building many new data centers

Are coming online, which then puts

Pressure on power delivery into those

Data centers

And Arm technology gives them the most outcome per watt.

The AI industry isn't the largest industry in tech.

It might be the largest industry in the history in tech.

There is a big excitement about the Arm AGI CPU.

This is helping us to accelerate the innovation

As we are moving into the future of AI-driven enterprises.

AI is redefining the entire infrastructure from the user,

The consumer, the business, the use cases,

The models, the applications, the infrastructure,

All the way down to the silicon.

Many of the solutions out there are getting power hungry.

Not all customers, their data centers today can handle that.

With Arm silicon coupled with a brand

New line of AI accelerators,

We now have a power efficient solution that we

Can offer our customers.

This performance per watt that we re really going to get out

Of this CPU is really going to help us not only

Save money, but be able to get to places

That were harder for us to get to.

So what makes this partnership compelling

Is their system-level combination.

SK telecom is pairing the Arm AGI CPU

With the Rebellions AI Chip.

This CPU is the perfect fit as we evolve

Into an AI data center developer.

The Arm AGICPU strengthens the orchestration layer of the

system

enabling greater efficiency in the head nodes

that support next generation frontier AI systems.

We are excited to work with the Arm team and

continue building the infrastructure that powers the next wave

of AI

There’s a lot of change happening right now

Its interesting to see how that drives innovations on all levels.

It opens up a whole new world of possibilities

To drive innovation.

That's pretty exciting.

Our mission is smarter technology for all,

And so to be able to be at the front end

Driving platforms and solutions, who better to partner with

Than Arm?

This partnership is not about what's coming next.

This partnership is about the decade to come.

Now, our customers and some partners that are supporting us here today, the support we've gotten has really just been incredible.

We built Arm AGI CPU for you, and we're so pleased with the response.

The Arm AGI CPU has been designed from the ground up to make sure that performance scales and power stays predictable. That's the superpower: performance, scale, and efficiency. And it's resonating with our partners.

It's just a very different approach than is taken by x86. They're burdened with execution overhead, and legacy feature support. They chose to focus on things like modularity, support for lots of different markets and esoteric use cases. We're ruthlessly focused on improving efficiency and reducing latency. Ultimately, this is about architectural philosophy. We're not strapped to the past. We're not strapped to the past.

Listen. We don't support Lotus Notes, OK? We just don't do it. We're focused on exactly and only what the agentic data center needs: performance, scale, and efficiency.

Let me take you through that in a little more detail. It starts with performance. And performance for us is all about doing more work for every clock cycle. This has always been an area. Great IPC has always been an area where Arm has shined. How much work do you get done every single cycle? Our AGI CPU absolutely shines here.

Now what we see is that legacy CPUs, they sometimes try to compete on this vector by doing things like increasing the frequency, going to boost modes. But here's the reality. When you increase the frequency, what else do you increase? Power. That's a problem. These boost modes are not sustainable across long periods of time. They're not sustainable across the chip.

With Arm AGI CPU, what we give you is full performance, sustainably, all the time.

And ultimately that means scale. We linearly scale across cores, and our memory and I/O subsystem is specifically designed to be matched to those cores so that we can continue to feed them 6 gigabytes per second of memory bandwidth to every single core.

In order to scale, what we see some of these legacy architectures do is multithreading, right? What happens when you do multithreading? You throw two jobs at the same core.

That's how they get to a high thread count or try to get to a lot of devices. But the reality on that is your I/O and your bandwidth, that doesn't double, so you just move the bottleneck elsewhere. And, oh, by the way, the CPU needs to be burdened with managing that back and forth, and so your performance degrades. You end up starving your processes.

What we see over and over again is that data center operators have to overprovision their data centers by 30% or more to deal with this lack of nonlinear scaling. This is an actual thing that happens. We take pride in not having to do that.

There's actually a great demo of this out on the show floor. I encourage you all to check it out after the keynote.

And then finally, we have this maniacal focus on efficiency. Obviously that's always been Arm's hallmark. It's always been something that we've been great at. We're leveraging all those techniques and methods and experience that we've built up over the decades around building incredibly efficient processors, incredibly efficient technology. And we're packaging that all in a custom design specifically for this use case. AGI CPU is purpose built without that legacy overhead, because I t all comes back to performance, scale, and efficiency.

At the end of the day: no wasted cycles, no stranded compute, no wasted power, silicon. And we're super proud of that.

Let's look at what it means in practice. I'm going to show you the results and they kind of speak for themselves. First, let's talk about sustained performance. What you see here is the performance you can expect consistently. So this is consistent performance. No performance throttling, because your over power budget, no memory or I/O contentions – this is the sort of performance you're going to see.

You can see with AGI CPU it is world-class, world class performance you can take to the bank.

Next let's talk about scale. How many threads or agents can you run in each rack? How much compute do you actually support with a fixed power budget? With a fixed physical footprint? Remember I showed you those racks I showed you earlier, there you go. That's where we land.

And of course there's efficiency: performance per watt.

What's going on with my screens? Its flipping all over the place? Can you go back, please? Go back one more.

So what you're seeing here, all of these charts are with SMT disabled, so these are single-threaded. Single-threaded cores for us, single-threaded cores for them. So no multithreading whatsoever. OK? I told you what I thought about multithreading, which is why we elected to show it to you this way.

But oftentimes what we hear is that multithreading is going to improve that middle chart. It's going to allow for more scalability. Multithreading is going to improve the performance per watt. Let's take a look at what happens if we turn multithreading on.

OK. So first of all, your performance goes down. That's the chart on the left. And the reason why the performance goes down is because you can't just add more work and expect performance to stay the same. So that's pretty self-explanatory.

And in this particular case, again, we've held it kind of based on the memory and the I/O bandwidth available, kind of where you land. That second one, sustained threads per rack, the reality is that because of the limitations on the device and all the bottlenecks, you end up in a scenario where you can't actually use all of those threads. Many are left idle.

And then finally, performance per watt. Yes, there is a small improvement there, but not enough to change the calculus.

At the end of the day, let's also be clear: this is a killer product. Arm is in a class of its own. Performance, scale, and efficiency. I'll say it one more time. This is what the Arm AGI CPU is built for. And the impact on the AI data center is going to be profound.

Let me turn it back to Rene right now. Thank you.

Thank you Mohammed. Thank you Paul and your LLM agent that is going to do all the conversions for us. So we shared a lot with you today, and I greatfully appreciate your patience and time. If there were just few things that takeaway from this morning, I think it starts here: performance per watt, which translates to performance per rack.

When you look at an x86-equivalent structure, same power delivery, 36 kilowatts, 2x the performance in the same power. That's what you need to remember.

For those of you who are paying for that power, there's another number to remember. For those who are paying for that power, that’s the number to remember. If you think about 1 gigawatt of capacity and you think about the CAPEX associated with that extra power you're spending at the sake of performance, it's up to $10 billion in CAPEX. Obviously, these are serious numbers.

So again, the takeaway from the Arm AGI CPU is 2x performance per watt, probably more than 2x performance per watt.

Now, you heard a number of comments in the videos, including Santosh, that when you embark on the kind of engagement and partnership we're talking about, while a day like this, an event like this, is wonderful and amazing and we're talking about a great product, it's really not about the day, but it's about the future and commitment to a roadmap.

So we are committing to future generations of this product. Arm AGI CPU 2 is coming out soon, as is Arm AGI CPU 3. As you heard in the videos, again, these are multi-generational engagements. We're investing a lot. Our customers are investing a lot. The ecosystem is

investing a lot. We are absolutely committed to a roadmap and a future around this product line.

In addition, we will continue the CSSs around these products. As Mohamed mentioned, one of the big benefits of the CSSs is the speed it allows our customers to get to market. It also enables a lot of benefit for us as well. So the CSS roadmap will continue.

So, I want to close a little bit around what we think the financial opportunity is for Arm.

So before this day, our business has been IP and IP compute subsystems. And we had been doing extremely well in that business, far better than what we had talked to investors about two and a half years ago when we did our roadshow for the IPO. We're actually ahead of that.

When we look at the AI data center business, that represents today about a $3 billion TAM, and now I'm just talking about roughly the royalties. So I mentioned in one of the earnings calls that the cloud AI business will probably be a large business in a few years, and this is really driven by all of the growth that Mohamed talked about, the deployment of 1.25 billion Neoverse cores and forward.

When we think about our business going forward, the Arm AGI CPU, and as Mohamed mentioned, we have committed customers: Meta, OpenAI, Cloudflare, SAP, F5, customers you saw on the video. When we think forward about what is the market opportunity for this business, it is a dramatic sea change for the opportunity.

When we look at what's going on with agentic AI, the growth of CPUs, the benefit that power-efficient CPUs bring in the data center, we think this represents about a $100 billion TAM for us in the future.

So today is all about the Arm AGI CPU. But there will be some tomorrows. And don't ask me about tomorrow today, but there will be some tomorrows. And we think this opportunity to take the work we've done across all of the markets, as you've heard in the videos, from edge to cloud, from milliwatts to gigawatts, we think we have an opportunity to address greater than a $1 trillion TAM by the end of the decade.

So we've got some work to do. But I couldn't be more proud of what our company has changed into, grateful to the ecosystem that helps us achieve it, and the customers that are now committed to buy our product.

I want to close by saying that we stand on the shoulders of our ecosystem. None of this is possible without the ecosystem that we have nurtured for 35-plus years. Many of you who are here today and watching on video, thank you for attending today.

Arm is everywhere. And we appreciate your support.

Over 350 billion chips have been shipped with Arm, or the most pervasive computer architecture ever invented. We touch 100% of the connected population. Anyone with any

digital device is likely using Arm. Arm is one of the great secrets of the technology space. People use Arm technology every day, multiple times. When you look at the Arm ecosystem, it is the de facto platform that everybody is building on top of. We deal with an immense the different range of different companies. In some cases they take our computer subsystem and put their technology alongside that. But increasing number of partners are saying actually we're not experts in buildings to look and. We would like to have an Arm based solution. We chose for many years not to build silicon because when he is all about what our partners want, that gave us the IP business that we've got that will be continuing unimportant part of what we do going forward. So the IP and the CSS Rd maps remain unchanged. We get this tight coupling of lessons that we learn from doing silicon development. That feeds back into the current generations of IP development, which the whole ecosystem, however they engage with Arm will benefit from. And it's all about choice and it depends on where the ecosystem is at an where customers want to engage with us around that and the Arm silicon will combine all of the goodness of all of our technologies. And it's got enhancements to Scalable Vector of extensions that allow you to provide higher quality computing for AI as well. AI is moving incredibly fast. Think everyday you wake up and there's a new technology around AI. Think about where we're going to be an AI 20 years from now, and then think about the fact that everything is growing faster today than it was 20 years ago. The world is evolving to adopt AI and so more and more we're seeing where these optimised systems is optimised platforms. That is the perfect home for a product like the Arm AGI CPU. In the era of agentic AI, they Arm AGI CPU was really purpose built for these agentic AI workloads and specifically for AGI in the future. So that's really how this brand we're really stand up. So as we've learned with AI, we're starting to learn more and more about customers, workloads or applications, what they need. And what we're finding is that the AI solutions today are very, very power hungry. Our reduction is a constant battle. An Arm has an architectural advantage for getting more performance at lower power. Product was designed to run on batteries, which is why the product is so pervasive. But any CPU is only as pervasive as the software that runs on it, and that's really the ecosystem. Compared to the other processors that are out there today, Arm has power efficiency. It gives customers options that they don't currently have today. AGI enables a true dual socket server that we could fit into any existing air cooled data centre and it represents a focus on efficiency that I think we've lost and it also has a focus on performance. I'm excited about these new Arm silicon based systems because it allows customers to have more choices and we do believe that this is going to be a game changer. It's amazing to be a part of bringing this technology and the value in the benefits of this technology to as many people as possible. From to build its own silicon, the first thing we have to understand is what our customers go through to build their own versions. There's a bunch of things that you have to do that we kind of knew about, but now we understand much better. We really focus on. Really robust discipline design practises and the team is really stepped up. The amount of dedication just all in buying for this project has been just outstanding. For me to be a part of this team, to work alongside many of the people that I see, you know, working hard in the labs at night or putting in there all to kind of

solve that late night issue, that's an incredibly rewarding experience. And when the silicon came back and it worked and it's up and running and the software is running and the team is excited and will ship to our customer and our customer has it up and running, the feelings absolutely amazing. I remember the e-mail that I got saying it's alive. That was away moment. It meant that I could ping the CEO and say it's alive. The wire moments won't be just one, it's just going to be a series of them every other week. To build a world class product of Arm AGI CPU, it takes a lot of people, a lot of resources, and we didn't do it overnight, but we did it pretty quick. That's the way you have new things in this market today. Speed matters and I couldn't be more proud of the team to make it happen. I look at this as a chapter in a long novel and we're just getting started. It's going to be groundbreaking. And i feel immensely proud to work for arm and to work with some amazing people to achieve that transformation to the company but in the transformation to the industry under the world