11192024 Ignite KEY01 Satya Nadella

# Ignite Keynote
**Satya Nadella – CEO, Microsoft**
**Tuesday, November 19, 2024**

(Applause.)

**SATYA NADELLA:** Hello, and welcome to Ignite! It's really fantastic to be back here in Chicago with all of you and everyone joining from around the world. It's always fun to be in conferences like this, at times like this, when major platform shifts are in the air. They're exciting. The buzz, in some sense, I live for these.

In fact, I was — this morning, I was reflecting, what is it? Like 32 years ago, right here in this very conference center, at the spring COMDEX is when we launched Windows 3.1, and that was like a big deal. It was just actually a few months before I joined Microsoft, and it made Windows what it finally became. It was just the most major release of it.

And in fact, in 2015, Ignite itself started in Chicago, and that was right in the middle innings of the cloud. And so here, to be back again during the middle innings of AI is just fantastic. And so, I'm really excited about it. And you know, given I see this room, it's breadth first. So I'm going to give a keynote that's breadth first. So I'm going to try and cover everything that we have for you throughout the show.

With every platform shift, it's always perhaps good to build a deep context and understanding of the underlying forces. And today we talk about them as scaling laws. Just like Moore's Law, we saw the doubling in performance every 18 months. With AI, we've now started to see that doubling every six months or so.

Now, in fact, there's a lot of debate. In fact, just in the last multiple weeks, there's a lot of debate, or have we hit the wall with scaling laws? Is it going to continue? I mean, the thing to remember, at the end of the day, these are not physical laws. These are just empirical observations that hold true, just like Moore's Law did for a long period of time. And so, therefore, it's actually good to have some skepticism, some debate, because that I think will motivate, quite frankly, more innovation on whether it's model architectures or whether it's data regimes or even systems architecture. So it's a good thing to have.

In that context, though, if anything, we are seeing the emergence of a new scaling law which test time or inference time compute. In fact, OpenAI's o1 is a good example of it. And features like the Copilot Think Harder is built on o1, is all about using test time to solve even harder problems.

So ultimately, though, all these breakthroughs manifest in three capabilities that are exponentially getting better. The first is this new, universal interface that's multimodal. It supports speech, images, videos, both as input and output. Second, we have these new reasoning and planning capabilities, which essentially we have new neural algebra to help solve complex problems. We can detect patterns involving people, places and things. You can even find relationships between people, places and things using this new algebra.

And third, we now have this capability to support long-term memory, rich contexts, and in fact teach these models to use tools. So if you put all those things together, you can build a very rich agentic world defined by this tapestry of AI agents, which can act on our behalf across our work and life, across teams, business processes as well as organization.

Now, you know I love this stuff. We are going to celebrate a lot of technology over the next multiple days, but it's worth reflecting. In fact, given we are in Chicago, there was a professor of philosophy right here in University of Chicago some 50 years ago, John Haugeland, who said the following: "The trouble with artificial intelligence is that computers don't give a damn, but we do." And that's what really grounds us. Amidst all this rapid change, we remain grounded in our mission to empower every person and every organization on the planet to achieve more using this technology to make a difference for themselves, for their teams and for the world.

It's not about tech for tech's sake, but it's about translating it into real outcomes. And today, I want to focus on AI and this transformational power as it drives growth in business. It improves efficiency; it improves operating leverage. And to do that, we are building out three platforms. Copilot, Copilot Devices, and Copilot and AI Stack. That's it. Those are the three platforms. But before I dive into all of the news across the platforms, I want to talk about that something that we are prioritizing above all else. That is security.

At this conference, you'll hear us talk a lot about our Secure Future Initiative, the progress we are making, the principles of secure by design, secure by default, secure by operations, but most importantly, our commitment to continuous improvement. Because this is not just a destination or stand down, one milestone thing. We'll never be done. In fact, we're only as good as our ability to defend against the next novel attack. And we are seeing real momentum, even with our customers and partners around implementation of zero trust using our tools and practices. In fact, a great example of this is what the U.S. Navy has been able to do. They've been able to respond to all the executive orders and mandates and meet, in fact, their zero-trust goals years ahead of schedule. And it's fantastic to see that.

We continue to invest in security. You can see the updates. We're making tons and tons of announcements that you'll hear about across the security stack. The one thing, though, I want to point out is Purview. It's probably the product for this conference because in the age of AI, data governance takes on an even more critical, central, important role. And in Purview, we're introducing updates to prevent everything, from oversharing risky use of AI such as malicious intent detection, prompt injections, misuse of protected materials. So there's a lot in Purview.

At the end of the day, though, we recognize that when it comes to security, it's fundamentally a team sport. And that's why we want to partner. And we are partnering broadly with the security community. And today, I'm really excited to announce our Zero Day Quest. This is the first big announcement that I'm excited about because it's a new hacking event. There will be $4 million in rewards focused on securing cloud and AI. It's the highest rewards of any public hacking event in the industry, and the quest starts today and will culminate in an in-person hacking event next year. So we're very, very excited about it.

So with that, let's just dive into each of the platforms, and starting with Copilot. Copilot is the UI for AI. It's rapidly becoming an organizing layer for work and how work gets done. Every employee will have a Copilot that knows them, their work, helping them unlock productivity, enhancing creativity and saving time. And Copilot Studio will allow you to create agents that automate business processes. And every IT department will have a control system to manage, secure and measure impact. That's it. Those are the three basic concepts of the Copilot ecosystem.

Over the past year, we've seen incredible momentum of what lean did for manufacturing, AI will do for knowledge work. It's all about increasing value and reducing waste. Just take risk analysis at Bank of Queensland Group. In the past, when an incident occurred, they would have to go through thousands of documents and write reports, and now they use Copilot to be able to synthesize everything that happened and create that first draft. That means analysis that took weeks, it's just taking a day.

Vodafone is another great example. Their legal team used to manually analyze and draft and renegotiate thousands of contracts that govern their massive base station network. And now, they use Copilot to determine which contracts need to be renewed, which need to be scrapped and to keep track of all the expiry dates. Vodafone is also personalizing all of their customer service by leveraging Copilot, as well as Azure AI to manage customer inquiries. Their virtual assistant engages in more than 45 million monthly customer conversations and is reducing the average hold time by more than one minute.

And of course, we're just getting started. We're continuously innovating, shipping hundreds of updates since we made Copilot generally available a year ago. And it starts with actually work on the fundamentals. In fact, you could think about the Copilot responses today. They are more than two times faster than on average, and the response satisfaction is improved nearly threefold. And going forward, we think about Copilot adoption in three fundamental ways. The more employees use Copilot and Copilot agents, the faster they'll be able to realize value. And second aspect is: how do you ground Copilot or extend Copilot in your world, in your ecosystem? So Copilot is a platform you can extend with agents to scale what you can do, right? So if you have one employee, one Copilot, and one Copilot can have thousands, thousands of agents. And finally, it's about measuring ROI.

And let's start with Copilot Pages. Now, Pages is this first artifact of the AI age. And it's pretty magical, right? You see, in this video, we're bringing rich artifacts to Pages. You can add interactive charts, tables, code blocks, math equations, complex diagrams. You can use Copilot directly on the page to iterate on the content, and also control what happens on the page directly from a chat. It's truly this multiplayer canvas that enables you to ideate with AI and collaborate with other people.

In fact, I use Pages all the time. In fact, if I'm prepping for a meeting, I just give it a prompt. The Copilot combs through everything about the customer, let's say, from the web, LinkedIn, all the business applications like our CRM. In fact, all the work artifacts are documents, emails, Teams messages, presentations. And then I can put all of that just as a first draft into my Pages and share it, let's say, with my account team, everyone working with me in my office. It becomes much more of a real-time way for every one of us who's working on that topic to be on the same page. And then it becomes, in fact, the first draft for what comes next, which is my meeting notes, right? Then again, it's real-time shared with everybody involved.

Copilot is also not just in Pages, but it's deeply integrated, right? It's not just about charts and pages, but it's deeply integrated into the entirety of Microsoft 365 system. Let's say it's starting with Teams. Copilot can reason over all the past meetings and chats and transcripts and get you quickly up to speed. With Teams, in fact, screen understanding, which we are announcing today, it's pretty exciting. Copilot can even answer questions about presentations and documents shared during the meeting as well, right? You are doing a presentation. Here is a Copilot that understands exactly what you're presenting and answer questions about that presentation.

In Word, Copilot creates a draft based on other Word documents, PowerPoint, PDFs, emails, meetings. You never have to start from a blank page. In PowerPoint, you can enter a prompt of what you want to do or what your presentation is about. And the narrative builder generates essentially a spec for your presentation, right? An outline with topics, and you can edit it, refine it, its suggestions, and it will create your first complete presentation.

And in Outlook, this feature is something that I've sort of gotten so used to. It's the quickest daily habit I've built, which is prioritize my inbox, is a total game changer because it helps you quickly get to the messages that matter. Analyzing your inbox based on both the content of the mail, but also the role of the sender and the context. And that's, I think, it's just like someone went and labeled every piece of email with a blue arrow and said: why is this important? Right? I mean, think about it.

And I think what we're doing in Excel, though, is perhaps one of my favorite things. Just like what GitHub Copilot did for software developers, Copilot in Excel will do for data analysts. I can start with a very high-level prompt, that in the past, for example, I would have given a data analyst — let's say I'm in a manufacturing plant, and I want to figure out how to improve production rate, right? So I give a high-level prompt and say, hey, come back with an analysis of what can we do here.

And Copilot uses this advanced reasoning capabilities to build out a full plan for strategic analysis. I can easily change that plan, adjust it as I want to, and then it just goes and executes the plan. And throughout I can see it actually doing the work. It first does visualization. It generates heat maps, scatter plots, histograms. It figures out, in fact, the key drivers of production, does a comparison of all those key drivers, figures, and it calculates and then summarizes the insights and the actions in minutes. Like, how amazing is that, right? It takes data analysis and makes it available to everyone who has an Excel spreadsheet.

If you sort of look back, and you say Excel was one of those products that at population scale improved number sense. I think Copilot in Excel with Python will improve analysis sense across the world. So I'm really, really excited about this.

Now, if you have all of this richness of Copilot that you're using across the length and breadth of Microsoft 365, let's talk a little bit about how you can extend Copilot. Today, I'm really excited to announce Copilot Actions. With these Actions, you can use Copilot to reduce the amount of time you spend on repetitive, everyday tasks that you do. In fact, the best way to conceptualize Actions is for those of you who use Outlook rules, this is Outlook rules for the age of AI, and it works across the entire M365 system, not just in Outlook. So it automates everything from asking for, let's say, a status update from your team, compiling weekly reports, to scheduling emails, requesting feedback on a document. Actions are a very simple, but yet powerful way for you to scale what you do, right?

So whatever was the thing that you had to do multi-step, you just create one of these actions, and it just does it for you. You can discover templates for actions which you can reuse in your everyday work. It's just a simple interface, and we're not stopping there. And today, we're introducing new agents you can use within the context of your team. Again, the best way to think about these are as just your teammates. They're scoped to specific roles with very specific permissions. Just like we have permissions and roles. For example, a facilitator agent is someone you can add to your Teams meeting, and the facilitator will help keep the meeting focused, moderate the meeting, chat, as well as the follow-up and action item.

Our project manager agent in Planner will help automate, in fact, all the key steps in a project management workflow. It'll create a new plan from scratch. It'll help oversee what's happening across the project, task assignments, content creation. And next is even self-service agents, right? So for these agents, provide really useful information, answer questions and policies. But not just that. When it comes to HR and IT, these agents will help you complete the task. Think of these as just augmenting your HR and IT departments.

And we're also announcing SharePoint agents. Every SharePoint site will now have a built-in agent. These agents provide instant access to real-time information and insights from your knowledge base in the flow of your work.

We're also giving you the ability to easily create your own agents using Copilot Studio. Sometimes we sort of mystify these agents as things that somehow require a lot of effort to build, but it's really pretty straight forward. In fact, our vision is that it should be as simple as creating a Word doc or a PowerPoint slide or an Excel spreadsheet, that's it. When you say "agent," think creating a doc.

This example is a good one, right? I can create field service sources, in this case, a SharePoint site in my Dynamics CRM. I can easily configure it to meet my specific needs, and you have an agent in seconds that's just now integrated into Copilot.

You can also make these agents autonomous using Copilot Studio, and they can always raise an exception in Copilot for input. Remember, even an autonomous agent from time to time will need attention, and it will need UI, and that UI for interacting with us is Copilot.

Now, just last month, we introduced 10-plus autonomous agents in Dynamics 365, that do everything from optimizing supply chain, to helping customer service teams resolve issues. For example, take a look at a sales qualification agent, autonomously researches all the leads that are there in your system, and flags the best prospects for you, and then grabs a personalized email that you can, of course, edit and send. And we are already seeing customers using capabilities like this.

In fact, McKinsey has built an autonomous agent that reduces client onboarding time by as much as 90%. DAO has built agents to optimize their shipping process or freight shipping process and projecting millions of dollars of savings even in just the first year.

And of course, when we talk about extensibility that includes changing how you interact with your bespoke business applications too, that's fundamental. The idea that I have to go to one business application at a time just goes away in this world of agents, and so we are very excited to share. Many of our partners have built their own agents and connectors in Microsoft 365 Copilot. That includes Adobe, and obviously LinkedIn, SAP, ServiceNow, Workday. Even companies like Cohere, they are building AI-first agents they are also integrating right into Copilot.

And now, to show you all of this in action, I wanted to introduce my colleague, Callie, on stage to take you through the entire Copilot ecosystem. Callie, over to you.

(Applause.)

**CALLIE AUGUST:** Thanks, Satya. Let's imagine for a moment that I work in sales. My day is comprised of a lot of important but time-consuming tasks, which can take away from the time I spend building customer relationships and closing deals. That's where agents come in, created in Copilot Studio. Agents range from simple prompt and response to fully autonomous. They help with everything from prioritizing leads, scheduling meetings, to fulfilling orders. Let me show you.

I have an autonomous agent that monitors and fulfills customer orders on my behalf. It alerts me when there's a cross-sell or up sell opportunity. I can see one of my long time customers place a big order for more products. The agent is able to quickly fulfill this order and send me an alert that this is a great opportunity to offer the customer additional products. In seconds, this deal just got the potential to be a lot more lucrative.

Let’s take the next step and get ready for the pitch meeting. I need to quickly get up to speed on the latest product updates and roadmaps. For that, I go to SharePoint. Over the years, SharePoint has become the most used solution in the enterprise for knowledge and business process. This site is where we keep all of the latest product specs, roadmaps and training materials.

Now, every SharePoint site has an agent, so all this rich information is immediately accessible and useful. I can find just what I need in seconds by simply asking the agent. Now, let’s jump back to this chat. From here, I can interact with the same SharePoint agent by simply @mentioning it. I can ask it to give me an overview of the relevant products, availabilit, and pitch deck.

Because the agent is grounded in this SharePoint site, it pulls from the right source of truth and gives me just what I’m looking for.

Now, it’s been a while since I’ve met with this customer and, of course, I want to put my best foot forward and understand where they’re at. Copilot points me to the sales agent, which surfaces all of the details that I need, like account highlights and opportunities, our most recent sales engagements and interactions with the customer.

I can see that I have a new counterpart on the customer side, and I want to learn more about them. For that, I’m going to turn to LinkedIn. I can simply @mention the LinkedIn Sales Navigator agent to see their background and experience, and who else I might know in their network. It looks like we have several connections in common.

Within seconds, I’m up to speed on my new partner, right in the flow of work. Now, I’m ready for the meeting, and I never even opened my CRM.

I’m going to fast forward to after the meeting. It was super successful, and now I’m going to work up a quote. Let’s step over to SAP’s Copilot Joule. I asked it to reference the latest pricing sheet, and I want it to reflect the appropriate discount given the size of the order. SAP Joule generates a detailed quote I can quickly review and send to the customer.

Now, I’m in great shape to land the business and it’s not just SAP. Many of our largest partners are creating agents to bring their unique knowledge right into the flow of work, from HR to finance to sales.

Take ServiceNow, which covers customer service, human resource management and workplace services. Or, Adobe, which can help your team build world-class marketing, design and visual content.

Copilot will empower every employee, and Copilot Studio Agents will transform every business process. Back to you, Satya.

**SATYA NADELLA:** Thank you so much, Callie, thank you.

(Applause.)

**SATYA NADELLA:** Now, let’s take a look at one more very important consideration, which is measurement. After users start using Copilot and all these agents, one of the fundamental things that all business leaders want to do is to figure out, how do we measure ROI?

And so today, we are very excited to announce Copilot Analytics. Here, if you take, let’s say, a sales territory manager, they can now correlate the specific Copilot usage to a business metric, like their win rate over time. And it’s not just Copilot, it’s Copilot and all the agents that you have built. You can look at their usage and start tuning even the usage to the business KPIs.

Our goal is to show how Copilot usage is ultimately directly translating into business outcomes across sales, marketing, finance and more, and there are many, many more examples. This is fundamentally the process, because it’s really a question of change management. Think of Copilot Analytics as a tool for all of us to change how work, work flow and work artifacts are all getting done. That, I think, is ultimately how we get ROI.

So, that’s a look at Microsoft 365 Copilot, Copilot Studio and Agents and Autonomous Agents, that end-to-end system for AI-driven business transformation.

Now, let’s move to the next platform, which is devices. In the age of AI, even the devices are fundamentally getting transformed, with both AI and cloud. What is happening here is that we are fundamentally taking what’s happening in the cloud with AI to the edge, and think of all of this as one, continuous, distributed computing fabric.

Over the past year, we have introduced an entirely new class of Windows PCs designed to unleash the power of that distributed computing fabric across the cloud and the edge. We call these Copilot+ PCs.

We are working across the entire ecosystem. It’s fantastic. Just like in the cloud, there’s all the silicon innovation. Silicon innovation is back in a big way in the client, whether it’s Qualcomm, AMD, Intel, all building fantastic systems for PCs going forward. And of course, we’re working with all of the OEMs, and now you can see the real manifestation of all of that in these 40-plus FLOPS, all on the client with the NPUs.

Now, when it comes to the fundamentals of these PCs, also, whether it’s battery life or performance, they’re best-in-class. That was the other thing. This is just not adding an NPU, but it’s about making Windows and Windows PCs just fantastic on fundamentals.

And, of course, ultimately, it’s about developers, right? What happened 32 years ago with 3.1 was about applications, and here we are with Copilot+ PCs.

We are back at it with Adobe, and WhatsApp, all who are seeing the capabilities of these new PCs. They’re bringing their best applications, leveraging these NPUs to really deliver breakthrough AI experiences.

We’re also delivering entirely new end points. Three years ago, we introduced the Cloud PC category with Windows 365, which securely streams your personal Windows desktop from the cloud to any device, whether it’s iOS, Android or even a mixed reality, let’s say a headset from like something like Meta Quest. We’ve seen unbelievable momentum and adoption for these Cloud PCs to remote workers, temporary workers, IT developers, all of the frontline scenarios in particular, even around disaster recovery.

In fact, it’s my go-to developer desktop. It gives full access to GitHub, Code Spaces, VS Code Azure SDKs, is all the CLIs set up in one place, and I can access it everywhere. Our Windows App, now, gives us one-click access to all of these Microsoft virtualizations on any device.

Today, we are announcing, in fact, Windows App is coming to Android. We’re excited about that. We are announcing mobile application management. I know this is something that IT has wanted for a long time, and it’s both to iOS and Android. This means any employee can work on Windows 365, even on their personal devices like this iPad, because your corporate apps and files stay secure, all managed on the Cloud PC.

Windows 365, itself, is growing rapidly. It’s grown by triple digits year-over-year, and it’s now used by some of the world’s largest companies, including Wells Fargo, Johnson & Johnson, Siemens.

And today, though, I’m really excited about the next big step here. I’m really thrilled to announce Windows 365 Link. You can see it right here. It’s a simple, secure, purpose-built device for Windows 365. It’s admin-less, password-less, and security configurations are enabled by default and cannot be turned off.

Windows 365 Link expands the PC category, or the Cloud PC category, by connecting you directly to your productivity in the cloud with no data or information left on any device.

Let’s take a look. Let’s play the video.

(Video segment.)

Now, I’m really excited about this device, and I’m also pleased to say that it’s going to be available in April of next year, so really looking forward to it. The form factors on Windows 365 Link give you another choice for Windows endpoints. But look, we fundamentally recognize how mission critical Windows is, and we are committed to both its security and resilience as a first-class priority.

The latest release of Windows 11 has over a dozen new security features, and most importantly, they're turned on by default, including device encryption enabled across all devices. In addition, we are excited to announce this new Windows Resiliency Initiative. Super important. We are doubling down on our commitment to make Windows secure and reliable for customers for all their mission-critical workloads. As part of this work, we are making changes to low-level operating system access. We're introducing new features in partnership with the entire ecosystem, establishing new guidelines for safe deployment practices.

One example of this, which is, I think, something that is really exciting is Windows Hotpatch, which works across your entire Windows estate to apply critical security updates without requiring a restart. We also continue to push the envelope on Windows security and resilience across both the cloud and the client. And thanks to point-in-time restore, customers who use Windows 365 can be up and running in minutes and roll back a cloud PC to its exact earlier state, so it's pretty awesome to see all of this support for all the mission-critical workloads that run on Windows.

That's what we're doing with Copilot devices. And now I want to get to the final third platform, which is Copilot and AI stack. Now, the way we are approaching this is pretty simple. What we're doing is we're taking all of these apps we're building with Copilot and agents in Copilot Studio and exposing every layer of that tech stack so that you can use that to build your own Copilots and agents. That's it. That's as simple as it is. Every app is becoming an AI app, and over the past year we've seen unbelievable momentum in what people have been able to build.

Take a great example like NASA. Data scientists who built the Earth Copilot have all these enormous amounts of geospatial data, contains tremendous insights. Everything from climate and air quality that can be super helpful for urban planning or disaster response, but its scale and complexity are quite difficult to analyze. And so, the Earth Copilot makes it possible for anyone to navigate all this data using just natural language for the first time. For example, you can see how air quality right here in Chicago has changed over the years. That's an analysis you can now just do in natural language.

On the other side of the world, Toyota has built Obeya, meaning big room in Japanese. Their ambition is to create this big room of AI agents that are accessible 24/7 for all their engineers. It's all grounded in their engineering designs, regulatory information, even handwritten docs. You can OCR them, put them into this Copilot. Engineers can ask it anything from how to make a car run faster or something super specific about some emissions output. And so, these are all the examples.

In fact, right after my keynote, you're going to hear from Lance from Blackrock in conversation with Judson on what they're doing with Aladdin and Azure. It's really an exciting work that they're doing, so it'll be great to listen to Lance. These are some examples on how the stack is being used today, and we're just getting started. Adding, in fact, capabilities across every layer of the tech stack here, starting right at the infrastructure layer. We continue to build out Azure as the world's computer. Over the past year, we have added and made so many data centers in 15 countries, data center investments in 15 countries across six continents. We now have 60+ data center regions, more than any other provider.

We're innovating to build these data centers sustainably. I'm really proud of this. In fact, we just announced two data centers in Northern Virginia built completely with low-carbon, cross-laminated timber to reduce embodied carbon footprint. This new construction model will really reduce the carbon footprint of our data centers by 35% compared to any conventional steel construction, so it's exciting to see this. — (Cheers, applause)— And when people say data center is the computer, increasingly we really think of the system starting right from the construction all the way.

At the network level, we're delivering innovation in hollow-core fiber. This technology delivers absolute breakthroughs, whether it's speed or bandwidth or power efficiency, in fact, compared to traditional fiber. Obviously, photons travel faster in air compared to glass, and earlier this year, we demonstrated fiber loss at the lowest level ever achieved in optical fiber. This low fiber loss is absolutely critical for data center to data center connectivity, and we now have production routes of hollow fiber running. In fact, we're going to add 15,000 additional kilometers planned over the next 24 months.

We're not stopping there. We're extending our cloud to the edge, and today we're going further and announcing Azure Local. — (Cheers, applause)— This is, again, something that many of you have asked us to do, which is bring Azure Arc all the way to all of the edge with Azure Local extends Azure services across hybrid, multi-cloud and edge locations with one central control plane. It brings Azure services to customers' distributed locations, whether they're in retail, hospitality or manufacturing, so that they can run their mission-critical workloads, some of these new AI workloads across cloud and edge.

A great example of this is how Armada has taken Azure Local and in fact, with Starlink connectivity, is helping Marriott with full resilience even in the midst of real extreme weather events. Now, when it comes to silicon, our Cobalt 100 VMs became generally available last month, and customers and partners, whether it's Databricks or Siemens, Snowflake, are seeing up to 50% improvements in price performance with these virtual machines. In fact, our own media processing capabilities in Teams are now all 100% on Cobalt 100 as well, and we're not stopping there with silicon innovation.

Our approach to silicon includes deep commitment to security, and I'm excited to announce our first in-house security chip, Azure integrated HSM. This is it. — (Cheers, applause) — This is a dedicated hardware security module that hardens key management, managing encryption and key signing, that can remain within the bounds of the device without compromising performance or security. And starting next year, it'll be part of every new server deployed on Azure, enhancing security for both confidential computing as well as general-purpose virtual machines and containers. We are very, very excited about this new silicon innovation.

Beyond hardening security, ultimately, when we think about systems innovation at the silicon layer, it is about removing bottlenecks that stand in the way, whether it's performance, latency or resource constraints. That's the opportunity we have as a hyperscaler. When you look at the workload, you look at the system, you say, what can we offload to silicon just to kind of get rid of some of these constraints. And so, that's why today we're very excited about expanding Azure boost with our first in-house DPU. DPUs are processors specifically architected to accelerate data-centric workloads, absorbing multiple components of a traditional server into a single piece of silicon.

It runs, in fact, cloud storage workloads at three times less power and four times the performance. I mean, what this will do for storage is what SmartNIC did for hosts in the network, these are going to do for storage. And of course, when it comes to AI, we are continuing to build out these new data center intelligence factories. We're extending all of Azure as the world's computer to basically be these intelligence factories. Tokens for Watt plus dollar is the best way to think about the new currency of performance. It's all about maximizing that value and doing it in the most efficient way.

The pace of innovation across the industry is simply phenomenal, and we are working with our partners to bring you more choice across cost and performance. That includes our deep partnership with Nvidia, which of course spans bringing their own workloads, whether it's Omniverse or DGX cloud onto Azure, but of course, working with them at the core system level on AI infrastructure.

In fact, last month we brought new clusters with H200 that became available. We're very excited about it. And our systems stack optimization, we have done between H100 to H200 continues to push on the total performance that we can deliver to anyone doing inference or training. And today we are announcing the preview of Nvidia Blackwell AI infrastructure on Azure.

(Applause)

Now, Blackwell is pretty amazing. It's got this 72 GPUs on a single NV Link domain, and then you combine it with InfiniBand on the back end. These racks are optimized for the most cutting-edge training workloads and inference workloads, so we are very excited about having Blackwell. And we're also continuing to work very closely with AMD. We're the first cloud that offered VMs powered by AMD's MI300X GPU, and we're using that infrastructure to power Azure OpenAI, so that's in our fleet today. And today we're introducing, in fact, Azure HB v5, which we co-engineered with AMD. It's up to eight times faster than any other cloud virtual machine, setting a new standard for high-performance computing, and it will be generally available next year.

(Applause)

And now, of course, when it comes to our own AI accelerator, MAIA 100, I'm really excited to say that it's right now live in the US East region supporting Azure OpenAI inferencing one. In fact, one of the most impactful applications we built is our customer service today, and it's all running today on MAIA 100 during all of the customer support workloads. It's exciting to see MAIA make it into the fleet, and we'll continue to improve it and scale it. We are contributing, in fact, our own systems innovation to the industry.

Last year at Ignite, we showed our first-generation MAIA RAG, which was that liquid cooling sidekick. These can support cooling of GPUs and AI accelerators, not just our own MAIA systems, but we're going to bring that innovation to Nvidia GB200. And you'll see this in the show floor because this is the type of systems innovation to get that performance. Ultimately, this is all about really being able to bring compute storage network edge silicon to deliver that TCO performance for all of your workloads.

Now, with that infrastructure layer, let's move to the next one, which is data. There's no AI without data. In order for you to build your AI applications, you need to be able to rendezvous your data with your AI compute effectively, and we are building out this full data estate to just do that. At the core of our platform is Microsoft Fabric, which we introduced last year at Ignite. It brings together all of your data, as well as all of your analytical workloads, into one unified experience with OneLake.

In fact, in Fabric, you can easily unify your data no matter where it lives in Azure or whether it's on premise, whether it's on Amazon or GCP, you can create this data layer for AI workloads in one place. The momentum actually has been incredible. We have over 16,000 Fabric customers, including 70% of the Fortune 500. Today, we're taking the next step with it. Until now, if you think about a typical data architecture, you required separate services for your operational stores and your analytical stores, and a lot of data had to be shuffled between these two.

We are very excited to announce that we are bringing our flagship operational database, SQL Server, natively to Fabric with Microsoft Fabric Databases. Just like Fabric simplified every aspect of an organization's analytical needs, we want to do the same for operational databases. Now, with Microsoft Fabric, customers have an enterprise data platform that serves all of their use cases. Whether it's batch data, real time or even massive transactional performance, all in one unified product. And all of that data is in open-source formats in Fabric's OneLake. This new database experience enables you to autonomously provision a database that's secured by default in seconds for OLTP applications, while simultaneously creating that connection to analytical workloads right there inside of Fabric. You have both your operational and analytic workloads, essentially like a SaaS service.

It integrates with developer tools like VS Code and GitHub, so you can now utilize your unified data estate and build applications. This creates a unified data platform with the ability to apply AI across all of your data, operational and analytical data. And for those that need additional customizations and control, our Azure databases offer just that and are optimized for AI.

In fact, one of the things that we're very, very focused on is vector search at scale because that's one of the most important operations when it comes to AI and data. The DiskANN was developed by Microsoft Research to power that low-latency, high-scale, cost effective vector search. We ourselves have been using it for a while now for our 400+ billion vector indexes in Bing, and with 10,000 or so real time updates and query latencies of less than even ten milliseconds.

And today at Ignite, we are bringing this very powerful DiskANN technology to our Azure databases, including both PostgreSQL as well as Cosmos DB. Excited about this really cutting-edge stuff coming to databases.

Now, let's talk about so you have your infrastructure, you have your data, and now it's time to build some apps. When it comes to applications, every application is an AI application. And every new generation of apps has brought a changing set of needs. Right? When it is the web or mobile or cloud you need to build a new app platform. The same thing is happening with AI. AI is transforming how we design and customize and manage apps today, and that's why we are building out a first-class app server for the AI age, announcing Azure AI Foundry.

With Foundry, we are unifying all of our models, tooling, safety and monitoring solutions into a single experience integrated with the most popular developer tools available as a standalone SDK and a portal. It all starts, in fact, with our approach to models. We know models and model choice obviously sit at the core of every use case. You want to be able to optimize for COGs, latency and performance, and we want to help you choose the right model for the right job. In fact, we now have 1,800 models in our catalog.

OpenAI continues to make unbelievable innovation. They're setting the pace around model innovation in everything they're doing right. Even the latest open models are all available on Azure. In the last six months, Azure OpenAI consumption has more than doubled. But we also support all these other models, like open-source models from Meta and Mistral, as well as providers like Cohere, so that you all can choose the right model.

In fact, no application uses one model. If you look at even Microsoft apps, they're all thousands of models that are being optimized, fine-tuned, distilled, that all come together. And so new categories of models are also emerging. In fact, we've added over 20 industry models designed for very specialized use cases from partners like Bayer, Page, Rockwell, Siemens, Site Machines and others. With all of this model diversity, it's never been more important to have the right tools, though, to choose the models for the job at hand.

With Foundry, we are adding model experimentation capabilities. Meaning for the first time, you'll be able to experiment with several of these models, compare the outcomes, and choose the best model that works for you. In addition to our own offerings, we're also announcing new collaboration to help developers accelerate this model customization. From data prep and generation to training eval experimentation with fine-tuned models, these are all the considerations of an app server.

Our work with Gretel Labs and Scale AI helps developers remove data bottlenecks, make data AI ready for training. We are working with Statsig to help customers configure run these fast A/B tests using different models. New integrations with weights and biases brings a comprehensive suite of tools for tracking, evaluating, optimizing models using Azure OpenAI service.

This is all available on Foundry. We are excited also to introduce a new service to help you simplify the creation of all of these AI-powered agents. Our new agent service helps developers build, deploy and scale AI apps that automate business processes. I showed you before how you can use Copilot Studio to build agents just with a few clicks, but as developers you want a code first approach to building an agent, and that's what the agent service really enables. You can build agents that are grounded in data wherever it lives. Public data from the web, enterprise data in Microsoft 365, SharePoint. Or you can leverage, even Fabric's OneLake to unify your data across all of your clouds.

These agents then can take action, right? So, you want to be able to give action space to these agents. And you can take those 1,200-plus connectors we have in Logic Apps that we have been using in our app services, and you can connect it to the agent runtime. We also know that the multi-agent stuff is becoming pretty exciting. It's evolving very, very quickly. And we are making sure that even at this very early stage, our agent service supports all the multi-agent frameworks, right? Which is Magentic-One, Autogen, Semantic, Kernel. You can use all of these within Foundry and agent services to build out your applications.

AI apps have specific operational considerations as well. Right? Because when you build an application at your organizational level, you want to be able to manage AI costs, performance, safety and security. And that's why we are very excited to share that we are bringing really new management capabilities to Foundry. For example, we are introducing AI reports that will help developers document and share their applications, use cases, and most importantly, eval results. Because I think one of the ways we're going to be able to think about, and talk about, and reason about what your application does, is through the eval of the models on which it's built.

Safety is the most important feature of any AI app, and we'll continue to ensure we have the best tools to build these secure AI apps, including things like prompt, shield, and so that you can detect and block manipulation of outputs for your business content. And today also we are announcing risk and safety evaluations for image content. That's become a very important consideration right in Foundry. Now to show all of this AI Foundry and how you want to build applications, I wanted to invite up on stage Seth, to show you all of this in action. Seth, take it away.

**SETH JUAREZ:** Thank you so much, Satya. Let me show you how Azure AI Foundry offers everything developers, AI engineers, and AI and IT pros need to build and manage transformative AI apps. AI Foundry makes all of this easier with three groundbreaking capabilities. Our model experimentation tooling, our new agent service, and observability features like evaluations that help manage these apps even after they're in production. And by the way, everything I'm going to show you is actually live. OK.

In this demo I will use GPT-4o, our new real time audio model, GPT-4o audio, the real time API, and image processing models all seamlessly integrated into the experience I'm going to show you in a little bit. Foundry offers over 1800 models for you to choose from. It's pretty amazing. Once I've chosen my model, I can use Azure AI agent service, call APIs when needed, and perform actions with function calls. In this example, I have a main customer facing agent that is then delegating tasks to other agents to understand what the customer wants, what their purchase history is and to recommend products based on multimodal conversation with the customer.

These agents can run micro-intelligent tasks to provide a cohesive, personalized experience. These agents can also be grounded to a variety of data sources, including Fabric; it's pretty awesome. Fabric's OneLake makes grounding your agents really easy because it brings together your data across not only Azure, but also data in AWS, GCP, or from SaaS platforms such as SAP and many others.

And for CIOs, IT pros, business decision makers, we know compliance, measurement and observability are paramount. With Foundry, you get evaluations with your own data. All evaluation results are securely stored, ensuring you have the data you need for audits, compliance checks and continuous improvement. You also get experimentation and AI tracing that captures precise information, allowing you to correct issues proactively. Let me show you how all of these capabilities I just showed can create a hyper personalized customer service experience that solves real world problems in ways that we couldn't even have imagined a year ago.

OK, so I'm going camping, and my friend sent me this picture, and I want to make sure I have the right gear for a winter camping adventure, so I started a chat. Now notice this is something that's now routine, AI interacting with me with chat and images. The chat is great and we all love chatting, but what if we can break out of the actual chat box and do a more personalized experience? Because it feels like AI today it's all just like inside of your chat box. It feels like there's a lot of Copilot and stuff, but what if we could do a little bit more and maybe make it a little more personal, hopefully with a phone call? Hello? Is somebody there?

**COPILOT AGENT:** Hi, Seth. Yes, I'm here. I see you're getting ready for a winter camping trip. Can I send the concierge page to your browser to provide more details?

**SETH JUAREZ:** You certainly can. Send me to the concierge page, please.

**COPILOT AGENT:** Sending you to the concierge page now. You'll find more detailed guidance there for your winter camping trip.

**SETH JUAREZ:** This is awesome. What do you recommend for this camping trip?

**COPILOT AGENT:** Hey, Seth: Super exciting that you're gearing up for a camping trip. Since you've already got the Skyview two-person tent, let's build on that. Here are a few recommendations. First, Mountain Dream sleeping bag.

**SETH JUAREZ:** Is there a way that you can write it up for me so that I can see it, maybe with pictures?

**COPILOT AGENT:** Absolutely.

**SETH JUAREZ:** Right now, what it's doing it's going to take a couple of seconds to write an entire article for me, just for me. And you're probably wondering, well, hey, it's taking a couple of seconds. When's the last time you wrote a personalized article for someone that calls you in 10 seconds?

**COPILOT AGENT:** I'll put together a personalized recommendation article with all these products and their details, complete with pictures. Give me a moment and it should pop up on your screen shortly.

**SETH JUAREZ:** You already did it. This is amazing. I love everything here.

**COPILOT AGENT:** Fantastic to hear, Seth. I'm glad you love the recommendations. Enjoy your camping trip! If you need any more help, feel free to reach out. Have an awesome adventure out there!

**SETH JUAREZ:** How about them apples? Remember, like I'm serious, this was running right here because I wanted to show you something real. This isn't a peek into the future. It's happening right now with Azure AI Foundry. Back to you, Satya.

**SATYA NADELLA:** Thank you so much, Seth. Yeah, that's pretty awesome. You know, taking one of those multimodal models and then the app server Foundry and to be able to build these sophisticated applications. That's the app platform today.

Now, let's go to dev tools. We have the best developer tools for this AI era. GitHub Copilot is by far the most widely adopted, used AI developer tool, and we are focused on making it even more of a game changer.

Until recently, if you think about my own usage of Copilot, you can use the editor with completion, and you use chat no more. Now, with Copilot Edits, you get both of them to come together. We are bringing chat and inline editing together, so that you can easily make inline changes across a set of files. You can have a working set of files and then just use natural language to be able to change across all files. Copilot has gone full multifile.

And with Workspaces, Copilot is the most advanced in the first agentic AI-native IDE. Copilot Workspaces leverages agents from start to finish. You can go from sort of basically, an issue to a spec to plan to code, all in natural language. And just last month at GitHub Universe, we added agents on both ends of the workflow, one for ideation and one for automatically building and repairing code.

And we're not just stopping with just writing code. We are building software agents across the entire lifecycle, from testing to deployment. And we can tackle these complex code maintenance tasks, like upgrading an app framework. Right here, you have this example of a Java framework where you're applying updates, iterating until, in fact, that all of your code builds and all the tests pass. And so, that's just an agent that's doing that for you.

Agents can also improve, in fact, performance. This is pretty cool, which is performance engineering agent effectively, by creating performance benchmarks, running them, iterating on the code, until, in fact, you find a solution, and you pass the performance evals.

And finally, agents can also help you go seamlessly from idea to implementation all the way to production by creating all the resources on Azure and deploying it. Think of all of the DevOps functions and having an agent for it.

And all of this innovation is what we're working on. It will just ship in the months to come. We're very excited about it.

Now so far, we've talked a lot about how AI can drive productivity by understanding, fundamentally the language of business, but AI can also drive fundamental business transformation by understanding the language of nature and science.

Science itself is becoming computed science, and that's one of the reasons why we are focusing on delivering the systems and AI innovation to power breakthroughs in material science, chemistry, physics and more. And our new frontier in AI-powered science is moving from static prediction to dynamic prediction, meaning not just predicting the shape of the molecule, but understanding the dynamics their motion and how they interact, which is a critical step in developing new materials and new medicine.

Earlier this month, in fact, Microsoft Research published in *Nature* this AI-driven simulation system that can accurately model protein behavior down to individual atom orders magnitude faster than ever before. This is, I think, a real breakthrough that will then help biomedical research and really, advances in areas such as drug discovery and protein design and enzyme engineering, all of these functions, because you need those dynamic systems to drive them.

And this is not theoretical. We're bringing together these advanced AI models and agents to help scientists reason over and orchestrate across the entire scientific method effectively. And we are already delivering on this vision of a platform for scientists with customers around the world.

Novartis, for example, is using generative AI to design hundreds of new molecules for drug discovery projects, helping accelerate the process. Nissan has partnered with us to create a model to predict EV battery performance over time, improving it by something like 80%. Unilever is running lots of simulations to really accelerate their R&D, using AI. In fact, the Institute for Protein Design at the University of Washington, scientists are using our cloud to engineer new proteins from scratch that promise to be absolute game-changers in medicine and sustainability and other fields.

Let's just take a quick look.

(Video segment.) (Applause.)

**SATYA NADELLA:** It's amazing to see it. Congratulations again to David, who was just awarded the Nobel Prize in Chemistry for his work. And it's great to see RoseTTAFold is open source, and it's available, in fact, right in that Azure AI Foundry catalog.

(Applause.)

Yeah. It's the most accessible tool for scientists in the field. And so, it's great to see all of the progress.

Talking about AI for science, though, this new era of discovery can go into hyperdrive with quantum computing. To simulate the world, you have to fundamentally break free from the limitations of the von Neumann architecture and classical computing. And for that to happen, we need reliable qubits, not just noisy qubits that are currently available. And our Azure Quantum provides that unique virtualization technology that can extend any type of qubit, detect errors in physical qubit, correct them and reliably help compute all over them.

In fact, earlier this year, our partner, Quantinuum, and us, we achieved a first-time ever record of reliable logical qubits; a huge, huge milestone for the industry. And then just a couple of months ago, we went further. We announced 12 reliable logical qubits. That was the record. And today, I'm really thrilled to announce yet another milestone, this time with Atom Computing. We just doubled the previous record, creating a machine of 24 logical qubits.

(Applause.)

I mean, to just kind of put this in perspective, these logical qubits are all entangled, making it the foundation for the world's most powerful quantum machines. And to give you a sense for why this matters, if you add 100 of these reliable qubits, you will have scientific quantum advantage, could be achieved. And so, and that will unlock, as you can imagine, the computing power to go solve some of the most pressing challenges we have.

Discovering these solutions, or discovering these types of workloads, or running these type of workloads on classical computers is obviously impossible. And so, therefore, we are excited about these next generation of quantum computers that go beyond today's noisy qubits.

We are building, in fact, to that end, a first-of-a-kind, commercial offering with our partner, Atom Computing, which will combine our science solution in Azure and our quantum computing platform, that virtualization layer I talked about, with Atom's quantum hardware into a full discovery suite for companies, as well as labs. This is going to be available to really accelerate scientific discovery.

I want to close out by talking about sort of our core mission. This was a preview of what you'll see throughout this week, across all of the three platforms, but at the same time, our mission is to empower every person and every organization on the planet to achieve more, one community and one country at a time.

As we enter this middle innings of AI, it's up to us to empower human achievement. That's why, over the past year, we have helped train over 23 million people in AI and digital skills. And we are very committed to working to help millions more learn how to use AI, because learning these skills will change people's lives. And in fact, it already has. Let's roll the video.

Thank you all very, very much. And have a fantastic Ignite.

(Applause.)

END