Microsoft Ignite
Satya Nadella – CEO, Microsoft
Tuesday, November 17, 2026

(Video segment.)

VO: We are in a moment unlike any other. It is bigger than any one product, company, or technology. It's about ensuring the benefits of AI reach us all. That's why we're building a frontier ecosystem, one where you own your intelligence, continuously improve the outcomes that matter and accelerate your advantage in a world that never stops moving. This is how we create technology that amplifies human ambition, and turn intelligence grounded in trust into real world impact. More people building. More people innovating. More people shaping the future together. Welcome to Microsoft Ignite.

(End Video segment.)

SATYA NADELLA: All right. Good morning. Good morning. It's fantastic to be back here at Ignite and welcome. We have a fantastic show for you lined up for the rest of the week, and I want to kick it off by highlighting what we are going to unpack throughout the week. This week is about two things. First, how do we capitalize on all this new capability that's growing exponentially? And second, how do we build a capability inside of our organizations to deal with the consequences of all this capability that's growing exponentially?

So last week, in some sense, was really a reminder of how critical digital infrastructure is as the dial tone for the world. We know the trust you place in our services and take that responsibility extremely seriously. And in that context, I know we have work to do to provide that resilience that you can count on, the world can count on, and we are committed to ensuring that we learn and we continuously improve.

And I also want to recognize – yeah.

(Cheers, applause.)

But more than that, I want to recognize the hard work that each of you do inside of your organizations to ensure that continuity, that resilience. And so, really, a big, big thank you from the bottom of my heart for all that you do for your organizations and the entire ecosystem.

(Applause.)

I always find it grounding helpful to always start with our mission, which is to empower every person and every organization on the planet to achieve more. And in today's world, that means one singular thing, which is how do we collectively build out that frontier ecosystem? We need to be able to parlay all of these technological advances into real outcomes, new points of

leverage for people, the organizations that we build, as well as the society at large. That's the goal.

And for us, that comes down to three things. One, keeping the human at the center of it all, ensuring that you have ownership over your organization's intelligence and learning, and trust. AI in the flow of human ambition is perhaps the biggest design challenge we have in front of us. How do you do that? How do you make sure that individual roles, organizations, systems, and products are all being designed to amplify human intent and ambition and agency so that it can flourish going forward.

And we are seeing this change in some sense unfold when you look at, say, a software engineering team today. As an example, developers are not just writing code or doing PRs, but they are building these systems around code that give them new leverage. You can think of software engineering scaling from building software artifacts to building software factories that build those artifacts. The important thing here, though, is to remember still that human intent, that judgment, that experience, skills, expertise all very much matter.

And it's not just about the point – so it's mostly about the change in the point of leverage. What you did before changes, but your expertise matters, your skill matters, your craft matters. And this is not just limited to software teams. Every function will go through a similar change across their work, their workflow, and their work artifacts.

The second pillar is intelligence. In the age of AI, the key question is how do individual firms continually build and own their learning and tacit knowledge. For an economy to remain an economy where many firms exist and thrive, intelligence, by definition, needs to be broadly distributed versus just collapsed into a couple of models.

And the third pillar is trust. I mean, it's an interesting time. Scaling laws continue to work, you have all these emerging capabilities from each scale up, and yet we find ourselves in a place where alignment is elusive, mechanistic understanding of these models is elusive. And so the question is, how can you even trust all this powerful technology? And the reality is, by definition, you can't. And in other words, it becomes an architectural imperative to separate the supply of intelligence from the authority over it.

And so, our goal is to build a stack that addresses these three architectural imperatives, so let's start with the first one: Copilot. Copilot is the new operating system for work. It brings together all of the form factors of human computer interaction, whether it's chat or Cowork or Code or Autopilot all into one experience. It also brings together the disparate categories that we have all referenced like knowledge work, coding, business process, applications.

Apps are just plugins to Copilot. That's why we in fact have brought all of Office, all of its richness built into Copilot, and underlying this is a model system that matches the right intelligence to the right context and the right form factor to accomplish any task that you may have. Home gives you one place to start, whether you need a quick answer or want to really complete a complex, multi-step task. And Code brings you an infinite SaaS factory to every

individual inside the organization. You can build your applications and any dashboard and much more, and all of this can be hosted, managed securely inside of your tenant.

And Autopilot is an autonomous agent that keeps moving your work 24/7. It continues to work in all the places where you and your teammates collaborate today. You essentially have an additional teammate. Underlying all of this is this very powerful database, Work IQ. You can think of this as perhaps the most important data asset you have. It has a deep understanding of the people, its relationships to other people, your emails, your meetings, chats, projects that make up all of the fabric of work.

And today, we are thrilled to announce that all of this that I just talked about is broadly available to you to use inside of your organizations.

(Applause.)

And so we are going further. We're extending Copilot to specific roles. That means if you're in sales, legal, finance, service, we are organizing the skills, the connectors, the plugins to all of the business applications so that you can access them from the Copilot surface areas. This includes, in fact, third-party systems too. In fact, we're announcing today business plugins for Dynamics 365. It's just not Dynamics. We're excited to partner with Salesforce to bring Salesforce to Copilot.

(Applause.)

And you'll hear a lot more from Judson shortly. And of course, we also recognize the need to continue to have the right form factors, the ergonomics for heads-down work, whether you are an analyst in Excel or legal professional in Word or a designer in PowerPoint, a scientist using Discovery, a doctor or nurse using Dragon or software developer in GitHub.

Now, speaking of GitHub. At Universe, a couple of weeks ago, we announced two new primitives. Projects lets you define a goal, connect one or more repos, and have Copilot coordinate task sessions and recurring work all driving towards that goal. It's pretty magical. And second, we're also creating a new composite model family in GitHub Copilot Fusion. It orchestrates multiple models to deliver frontier-level results at a significantly lower cost.

In fact, just today, Copilot Fusion was ranked as the number one coding agent on Terminal-Bench, scoring five points over Claude Code at less than half the cost. It's exciting to see that.

(Cheers, applause.)

And of course, we're going beyond that as well. Up until now, when you think about Copilot, it's been about the single-player world. Of people and their coding loops or agent loops and swarms. But time has come for a true multiplayer form factor where people, their agents, and other people are driving towards that common goal with shared context. This is Copilot Workspaces. Let's take a look.

(Video segment.)

(Cheers, applause.)

Yeah. I'm really, really excited about it. Even in the few weeks that I've been using it, it's really changed my working habits. Just like how when you first had Chat show up and it became the place where you started all work; I think Workspaces and Today are going to be very similar.

What you just saw though, was just what points to essentially organizational intelligence. All that work we've done building out that database and Work IQ lights up. The cloud enabled us to move client software to have this very rich representation of organizational data enabling deep collaboration. So the cloud was about moving from just client software to essentially this cloud data gravity that allowed that collaboration.

And now AI is allowing us to take that data that you have in the cloud and connect it to intelligence, and that's what we're doing with Microsoft IQ. The idea behind IQ is to help every organization build, retain, and continuously improve your own intelligence across the organization. It brings together your action space, that means all of your data estate, all of your application estate, the context that means all of the IQ that you have created over the years, and the additional intelligence that you have with a multi-model system.

Essentially, we are taking the system we built underneath Copilot, and we are externalizing it as a first-class platform that embodies your organizational intelligence as an asset that you can continually own, continually build on. And we're bringing together all of the data you have. Whether it's the data underneath Microsoft 365 that's Work IQ, Fabric IQ, Foundry IQ, the data inside the – across the web with Web IQ.

And of course, third-party context too. In fact, Microsoft IQ comes with hundreds of new connectors we're announcing today so that agents that you build can act with full enterprise context where businesses get done. And all this manifests as simply as business plugins that can be shared and inherited across the entire organization.

And finally, there is a multi-model system. Now, for us, it's about making sure that we are able to map intelligence to a given task with the best economics. That's the optimization problem we are trying to solve. This is something we can uniquely do in the context of the real-world task by tuning the product, the harness, the tools, and the model system all as one end-to-end optimization. Now, speaking of models, we're very excited about a new class of models called Decision Models.

These models allow you to embed intelligence into software as a core primitive. They're quickly becoming key to how software is built. In fact, just like how databases back in the day or relational databases back in the day changed software engineering, I think of decision engines similarly.

And today, we are thrilled to announce Microsoft Decision 1.1, which is now generally available and ready for enterprise deployment.

(Applause.)

We have added multi-model support to it. We have improved performance, making decisions more accurate and the responses faster. Already, over 4,000 customers have used it. In fact, Epic is using it to make decisions across their patient experience, clinical care, as well as their back office apps. Copilot system itself uses it to route users to that right mode, model, the effort level with millisecond precision. And our Xbox team is using it to sort tens of thousands of open-ended pieces of feedback into what people are saying about all the different game launches.

And here's the fun thing: It's more than 10x faster and 200 times less expensive than using an LLM.

(Applause.)

And Foundry, of course, has all of the popular closed and open weight models. In fact, today, we are very excited to bring GPT 6.2 to Foundry. It's a great new model. It sets a new bar for AI professional work. On a benchmark for complex professional work, GPT 6.2 Astra outperformed its predecessor by more than half the estimated cost. And we also brought Claude Fable 5.2, Mythos 5.2 last week. And today, we are also announcing Meta's latest models are all coming to Foundry as well, so a great choice of models for you.

(Applause.)

And of course, we are also innovating at the frontier with our own model family, the MAI model family, that's both compute efficient when it comes to pre-training and token efficient when it comes to deployment. Our goal is to put additional points on the frontier curve, really informed by these real-world RLEs around task performance that matters to you inside of all the products, and we have great momentum.

MAI Transcribe 2 requests are up 12x since it launched at OpenRouter. MAI Code Flash on GitHub Copilot last month had three million users. MAI Cyber now powers over 40% of the scans across Perception code scanning, and we are going further.

We are announcing MAI licensed weights, meaning trusted partners will now be able to get access to Microsoft AI model weights through our Frontier Co. And we are committed to doing what it takes to ensure that everyone can own their own intelligence and thereby, drive the broad benefits of this platform shift.

And today, we're also introducing lots of new models. And Mustafa will come up and talk a lot more about this in more detail, but I wanted to highlight two in particular.

First, Code. We're introducing a new flagship family of agentic coding models, Code 2.

(Applause.)

Code 2 is at the frontier of cost performance for everyday enterprise coding tasks. It excels in taking code to production, multi-file implementations, repo-wide reasoning, and UI at a fraction of the cost. And Code 2 Flash is designed for fast, high-volume work like code explanations, targeted edits, rapid iterations, and it really brings mere flagship quality with much lower cost.

And we are taking the same approach with cyber as well. Cyber 2 Flash is our most capable security model yet. It's flash size model that delivers performance ahead of frontier models at a fraction of the cost per task. And just as we have done with coding, we are co-designing the model, the tools, and the systems all as one system.

And we are integrating Fiber 2 with Perception, and once you have the models and your context in Foundry, our managed agent harness is available to you to be able to deploy and build these long-running agents.

The same harness we use to power Copilot and GitHub Copilot is now available to you, and you can connect it to a continuous improvement loop that learns from experience. You get to define your own private benchmarks and evals, and then Foundry measures your agents against them and improves the agent or the model underneath it based on what you see in production. That means every execution is an opportunity to learn, improve, and deliver a better outcome.

We're creating ultimately, that always improving, always self-healing system. That's how your organizational intelligence becomes an asset and more valuable over time, and becomes your IQ.

Now, let's move to the third pillar, which is all about trust. It starts by making sure that you have observability at every layer of the stack. That's where Agent 365 comes in. It's the control plane that extends your existing governance, identity, security, and management systems to the agents across your enterprise. We already have 50 million agents in Agent 365, and it includes FinOps, too.

And as AI scales, every organization needs a clear line between what it spends and the value AI creates inside the organization. And that's why today, we are very excited about announcing Agent 365 Business Analytics.

(Applause.)

You can simply think of this as an ERP system of AI usage. You can see exactly what's driving the spend. You can connect the spend to the outcomes that drive business value. You can optimize the ROI across different scenarios.

And the next thing I want to talk about is how do you take an environment you've set up from Copilot to IQ to all the agents registered in Agent 365 and secure them. That's where Perception, our security stack, comes in.

We are building this end-to-end trust platform. We are helping customers both secure their AI deployments and use AI to strengthen their security posture, from Agent 365 and Entra giving every agent a unique durable identity, to MXC and cloud containers providing strong tool and

network boundaries, to Purview governing all of the data and information flows, and Defender disrupting and containing threats in real time.

In fact, we saw a fantastic example of this recently. An independent security researcher used ARTEX, that's in the news everywhere nowadays, an autonomous AI attack framework to test an environment protected by Defender. The agent kept trying to move laterally across the network, but Defender automated attack disruptions really kept containing those compromised accounts and isolating devices even as the agent continued its attack.

Attacks like this one are unfortunately, increasingly agentic and autonomous, and that's why I think it's so important that we have that machine speed in our own agentic response. That's why we built Perception as this complete agentic security system.

And the best way to think about Perception is that you have these Red team agents that simulate attacks, Blue team agents that detect attacks, Green team agents that fix and remediate. A good example of a Blue team agent would be what you had in Defender that protected against the ARTEX attack.

And another example would be a pen test Red team agent. It takes, let's say, an entire surface area of something like a Web app, the auth, the APIs, injection attacks, business logic, and creating all of the evidence you need to confirm that you're not exploitable. In fact, it's not just any individual agent; it's the playbooks you can compose teams of agents to tackle the complex jobs that trigger and take action autonomously.

Perception also helps essentially connect the signals, the sensors, the actuators across all of the assets. You have identity, your client, your cloud, your app, of course, agents now, and really secure your end-to-end environment. And that's why I am very excited to share that Perception is now generally available.

(Applause.)

At the end of the day, we believe there should never be a choice between innovation and safety. That means, in addition to the work we've been doing across our products, we're also working across the ecosystem to ensure that everyone can innovate with speed and scale safely.

Today, we can see almost everything that models touch, but you cannot see what happens inside a model. And really, what's worse is that we know that what a model tells us about itself is not even independent evidence, and therefore cannot be trusted. And one model monitoring another model has the same issue. You essentially have this unacceptable nested set of black boxes, which we now need to make transparent and verifiable. That's the technical challenge.

And so, therefore, we need standards that allow us to examine what AI sees and does, verify its actions independently and control what it's allowed to do. And that's why today, we are also announcing the Microsoft AI Transparency Initiative.

(Applause.)

There are three parts to this very important initiative. First is we will expand our collaboration in NIST's Agent Standards Initiative, drawing on its research to help the development of open industry-wide standards for AI transparency, security, and governance.

Second, we are taking an ecosystem approach. That means together, we will enable AI systems that are accountable, independently verifiable, and safe by design, regardless of which model it is or who the provider is. Already, 40 organizations are working with us on defining and applying these standards.

And third, we are going to incorporate the same standards across all our products. That means our MAI models, Foundry, and Copilot.

Our ambition through all of this work, the research work, the product work, our ecosystem work, is to help establish an open industry standard that makes AI safer, more transparent, and more accessible.

Now, underneath every one of these experiences that I've just talked about is infrastructure to deliver them at global scale. And when it comes to infrastructure, it's always shaped by the primary workloads of any new platform era. That means these are being shaped; today's infrastructure is being shaped by large-scale training, and of course, now agentic inference workloads at scale.

A great example of this is what we are seeing with GitHub. GitHub is seeing exponential increases, the pushes, the commits, the actions, the PRs, all being driven by agents. And we're rebuilding; that means all of GitHub's core, starting with the Git architecture, simplifying the system primitives. In fact, decoupling storage and compute, minimizing the coordination by shrinking that critical path of a push to the smallest step of integration.

After doing all of that and testing, in testing we pointed, I think, 5,000 agents to one repo, and we're able to sustain 7,200 writes per second. That's 180x the current throughput. That's the improvement and scale we need for an agentic workload.

And of course, what worked for GitHub now needs to be pervasive throughout the Azure platform, and that's what we are doing in Azure for Agent primitives. We now have more than 500+ data centers, 80 Azure regions, and we are continually working to make them more efficient, more sustainable, and partner deeply with the communities in which we operate.

We absolutely believe that technology should strengthen and not strain any community and the systems on which they depend on, and that belief underpins our Community-First AI Infrastructure commitments. And we're doing all of that sustainably, from matching 100% of our annual global electricity consumption with renewable energy, or reusing or recycling 92% of all the servers and components, and of course, all of the cooling designs that conserve water.

And customers are also asking for more control over where and how they run their most important workloads, and that's where sovereign cloud options provide a continuum of security, control, resilience capabilities without compromising on innovation.

And of course, we are also excited about Windows extending all of that infrastructure to the edge. We now have very capable models that can run locally, and we really see the opportunity with Windows to deliver unmetered intelligence. We are working across the hardware ecosystem to bring those capabilities to a new generation of PCs. With new devices, you can run, in some sense, last year's frontier models on your laptop and last quarter's models on your desk.

And we are also rearchitecting, of course, the core compute, storage, and networking primitives in Azure to execute, connect, share context at that agent scale I talked about when it comes to GitHub.

On top of all of this, you have these micro-VMs now that can come up in less than a second with full sandbox isolation guarantees on top. And all of this can run inside of confidential computing enclaves. You can then put an agent grid on front of it that brings identity and policy-aware networking controls, where each agent, what you can control what each agent can access, where it can connect, in fact, which other agents can it communicate with. And then this context cache will even accelerate and model loading.

Of course, you can now then take all this and wrap it into AKS or AKS Flex, or even hosted agents inside of Foundry, giving you all of the agentic primitives for you to be able to build your next generation of applications.

We're also reimagining the data layer. Agents need databases that can keep up with their speed and scale while giving them the context they need for all of the reasoning and actions. And that's why we are building Horizon DB, our new distributed –

(Cheers.)

Yeah, it's exciting. (Laughter.) It's a new distributed, managed PostgreSQL database with ultra-low latency, scalable reads, built-in search, and AI capabilities. I mean, just the performance is mind-boggling. It's 8x faster performance than self-managed Postgres.

And today, I'm excited to share that Horizon DB is generally available in Azure.

(Applause.)

And this next part is what I've been working on for, I think, 25 years of my life, which is we are bringing all of Horizon DB into Fabric. This represents the complete unification of transactional and analytical data for agentic era.

Think about what this means for a second. In the past, you had your transactional databases running your business, and then a separate set of analytical systems to understand what was

happening. But now, imagine an e-commerce business processing millions of transactions, every order, every purchase, every inventory change, all happening obviously, in real time.

And then with Horizon DB and Fabric coming together, agents now can reason over those live transactions along with petabytes of historical records in real time. And now you can just simply go to Copilot and just tell it, tell me what's happening in my business right now, and Copilot can create a live dashboard of your real-time enterprise.

This is really, really exciting that this vision that we have had for 20+ years is now finally coming together, and for you to be able to build the next generation of systems.

(Applause.)

Underneath all of this is also unbelievable systems and silicon innovation. We are building a heterogeneous fleet, a platform where we are able to tap into the best innovation coming across the ecosystem. And we're also advancing our own hardware systems, co-designing the data center, the servers, the silicon, all mapping to our workloads.

We just announced Maia 300, which will deliver over 40% more tokens per watt compared with leading GPU systems. And Azure Boost and Azure Integrated HSM take that critical infrastructure and security functions and put them into purpose-built silicon.

And today, we're very excited to announce the general availability of our next generation of Cobalt 200 VMs, which are our highest performance, general purpose VMs with leading price performance features. Thank you so much for that.

(Applause.)

That's just a very quick walk across all of the layers of what we are building. And when you look at all of what is coming, what is clear is we are entering an era of ubiquitous intelligence, but the key imperatives for us are to ensure that human ambition remains central, affirms control of its own intelligence remains central, and trust in technology that gets broadly diffused is maintained.

This is the promise of the frontier ecosystem, and that's what we are focused on building together with you.

And now, let me hand it over to Judson to take all of this and show you how we're bringing it to life. Thank you all very much, and have a fantastic Ignite.

(Applause.)

END