

Microsoft CEO Tour FY26
AI Tour Keynote
Satya Nadella, Chairman & CEO, Microsoft
London, United Kingdom
Tuesday, February 24, 2026

(Applause.)

SATYA NADELLA: Good morning. It's wonderful to be back in the UK and in London.

In times like this, where there are platform shifts happening, it's always wonderful to be able to have a chance to ground yourself in both the innovation and the impact the innovation is having broadly.

When I think about Microsoft and our core identity, it always starts with our mission, our mission to empower every person and every organization on the planet to achieve more. There's a lot of talk about what AI can do and what it means. To us, it fundamentally is a technology that needs to deliver on that mission of empowering us and the institutions and organizations that we build to be able to achieve more. It's not anything more than that for us. In fact, grounding ourselves in ensuring that AI does translate into real world impact is perhaps the most important work that we collectively all have to do.

When talking about that impact, we have this framework around what does it mean to be a frontier firm, and the best way to describe that is to really talk about the outcomes that any firm cares about.

When you think about any technology that's being applied, you want to make sure that your ability to deliver the product or the service is getting better, your experience for your own employees is now much more superior than it was before, your internal operations are getting more efficient. These real world outcomes is what we really want to move the frontier on.

When AI comes inside of an organization, can we measurably look at what we were doing before, and then look essentially at what in AI you talk about as evaluations, can I, in fact, have a metric that says, here is how my customer service got better, or here is how my employee experience got better, or any particular supply chain operation that I may have had got better. Really having that ability to move the frontier is what really the name of the game is.

To do that, I do believe that with this particular technology, you want to approach it differently, and it starts with us thinking about the mindset with which we think about this technology. This is not about just deploying another new digital system that just

looks like the previous generation of digital systems. These are end-to-end agentic systems that will require us to think about the workflow as first class, not just any given function or any given role or any given business process. In fact, the interconnectedness of the business process inside the company becomes very critical.

That's where the mind-shift shifts, especially functionally when marketing and sales and engineering are coming together, then to observe even how work and work artifacts and workflow happens between these functions, and then to be able to think about AI augmenting it is very, very critical.

Obviously that means we have to really ensure that the skills of our entire employee base are being really re-looked at in order to be able to use these new tools. The tool set is something that you want to deploy and diffuse broadly. Everyone should have it. It's not about a few people inside the organization who have the tools. In some sense, the first thing you want to do is to diffuse the tools, and then that way, the in-job training and picking up the skills becomes very, very critical.

The other piece is data. In other words, it's context, because when you think about AI, it's about AI that really is becoming sharper in your context. That means the data that you have, the data asset we have always talked about, needs to be structured in a way that it truly then allows you to really achieve that change in frontier for any outcome.

Now, to us, to realize these outcomes, we are building out three parts to an AI stack, and I was going to use the rest of my time to just really describe each of these layers and go into a little bit of detail on the experience layer, the platform layer to build agents, and then, of course, the token factory.

But perhaps before I dig into each of these layers of the stack, it's probably important for us to talk a little bit about two things that perhaps matter the most, which is intelligence and trust.

Now, when we say intelligence, it's not the intelligence of any AI model or any AI system. It's the intelligence of the corporation, the organization, which is the combination of the digital systems and the human capital and how does it compound inside of your institution, inside of your organization.

That is really at the end of the day, the goal. The goal is not to celebrate some AI model on the outside. It's to really make sure concretely you're able to point to how your organization has been able to build more intelligence inside the enterprise. In some sense, the enterprise value creation has to be related to your own ability to increase what you describe as the intelligence layer inside, which again, is a combination of the human capital that gets augmented because of AI, not somehow replaced by AI. That, to us, is number one.

Then in order to ensure that that happens, you need to have trust in these AI stacks and AI platforms that come inside of the enterprise, because one of the key things you want to ensure is that there's no leakage of that enterprise value.

You can't have what is unique about your firm, what is that tacit knowledge that you have as a firm to produce the goods and services that you produce somehow to leak out. You want to retain that as the core IP, obviously with the people you have and their know-how and their knowledge, but also as embeddings in some weights and of a model that you have.

So to us, ensuring that there is absolute trust that ultimately, when you deploy AI, you are building that next layer of intelligence and trust is what is the most important outcome, perhaps, that we want to drive towards.

Now, if you start with the experience layer, one of the things that you want to ensure is you're bringing AI to where people are naturally working today, because the easiest way for anyone to be able to use any new platform, any new tool, is to have it integrate into existing workflows and then thereby change, in fact, the workflow, the work artifact. That's what we're doing with all of the things we are building, whether it's on the consumer side, whether it's for information work or for science, for security, all of these high value domains.

But if you take information work or knowledge work, which is at the core of what Microsoft has always done with things like Office, we have really built this Copilot layer across the entire ecosystem of applications you use every day, and also built new applications like Copilot itself.

If you think of this as a graph, now this graph gets connected, not just by you moving from app to app, but by this intelligence layer being persistent, stateful, that is there with you across the entire system.

Of course, chat was the first thing that we all got used to over the last couple of years. What we have done is we have built applications. We use chat to be able to query and get back responses and information, and then you move applications. Whether I'm in Teams, whether I'm in Outlook, whether I'm in Excel, I have my chat history and it roams with me.

But the thing that in the last, let's say year and a half or so, we've also gotten very used to is invoking agents. It's no longer just chat, but you're also able to hand off tasks to agents asynchronously, and then have the agents work on our behalf and come back with results that then we can continue with the work in the workflow.

A good example of this is what we do with Researcher. Researcher is pretty interesting, right? It not only does the deep reasoning over a bunch of your data, but it also uses tools. For example, it can use the Computer Use Tool to go do its research. That's one of the

more classic agentic tools that now has become pretty much every-day in terms of how you go about your work.

Analyst is another agent that now is like invoking 24/7 a data science expert who can take any spreadsheets that you have or any data you have and then analyze them on the fly. It's like having a data scientist with you all the time.

Now, of course, we even have agent mode inside of Excel or PowerPoint and Word. In particular, Excel is a great example. What happened with software development is now happening even with people who are in Excel, because I may generate in chat a model of sophisticated model signal prompt. It'll come back with a pretty sophisticated spreadsheet. But having a zero shot spreadsheet is interesting, but it's your ability to use the spreadsheet which has all sorts of interesting positional context.

When you're looking at a spreadsheet, the reason why spreadsheets have stayed with us all these 40-plus-years is because they communicate a lot. As a human UI, to make sense of numbers, it's a very powerful tool. You can see the formulas. You can see by the position of rows and columns what it may mean.

Having a spreadsheet that you can reason over, and then you can iterate, the agent mode is, I can continue to refine it, iterate it. And if it is a sophisticated model, I can make sense of it. That's, I think, what the agent mode really allows you to do. It's about creation of artifacts and augmentation of these artifacts.

In fact, we are seeing, because of the AI proliferation, the number of artifacts, whether it's spreadsheets, whether it's documents, whether it's PowerPoint slides, the volume of these artifacts has increased, and thereby what we are seeing also is the need for precision augmentation of these AI-generated artifacts and the ability to understand them.

Nobody wants anything that is sloppy in terms of AI creation. They want to be able to make sure that what is being created is something that the next person who receives it, understands it, can see it, that you have done the work to make it much more explainable. That's the type of work flow.

Now, to show you all of this in action. I wanted to invite up my colleague Anne to give you a flavor for how this agentic experience and the overall Copilot experience comes to life.

Anne, take it away.

(Applause.)

ANNE KRUPKE (Sr. Product Marketing Manager, Microsoft): Thank you, Satya. I am so excited to show you how the entire Copilot stack is changing the way people work.

Let's imagine I'm an accounts receivable manager, and my job is to secure our inbound cash flow, or, in plain speak, to make sure customers pay their bills on time. I start my day in Copilot, and the first thing I'm going to do is look at this model selection that's available to me. I'm going to choose GPT-5.2 quick response today, because I'm doing a live demo and I've got a lot to show you.

Then I'm going to ask Copilot to prep me for my day. Copilot uses Work IQ to look at my meetings, emails and chats, and it knows the context of our previous interaction to build a tailored response that's really going to effectively help me prep. Let's take a look at that answer, quick response, indeed.

First up, it looks like I have a meeting the AI tour London demo. Really glad I'm in the right place with you all today. And then at the top of my inbox, I have an ongoing contract conversation going with a customer, and it's my job to incentivize this customer to pay their bills even more than on time, early. And we do that by offering a cash discount.

Instead of having to come up with policy ideas or offer options on my own, I can use Copilot again to help me out. Let's ask, what are my policy compliant options to offer this customer a cash discount for early payments on this contract. This is where the power of Work IQ really shines. Copilot is able to pull from our standard policy documentation. It understands my role as an accounts receivable manager and the context and the history of this customer and this particular contract.

And we can already see here how that's come through in the response. It's called out my role as an accounts receivable manager. It's sourced our payment terms and cash discount policy documentation, which I can choose to verify and make sure it's coming directly from our finance SharePoint. And it's summarizing my options. It recommends, based on the customer context and the contract size that we offer, our standard 2% discount for payment within 10 days of invoice, or a 2/10 cash discount.

But before I go offer this to the customer, I want to make sure the profitability of the contract isn't too much of a trade off, and where better to do that analysis than Excel? Instead of doing this all out manually myself, instead I'm going to turn to Copilot and turn on Agent Mode and put it to work doing the math for me.

I'll ask the agent to model that impact of a 2/10 cash discount on contract profitability. And because the agent is grounded in the same context that I am, I can give it access to that policy document and have it pull all those details to create a rich analysis. And then we'll kick it off.

Agent Mode is going to work in this spreadsheet, just like a coworker would or I would, and just like us, but a lot better, it works fast, but not that fast. We're going to use this time we have while it works to work on some other things.

Now, all those insights that the business gets, whether it's coming from Copilot, Excel Outlook, are based on our underlying data estate. And when it comes to finance, for our business insights to be valuable, our data needs to be accurate and up-to-date. Dynamics 365 is our platform for intelligent business applications that enables us to do just that.

Today, when I log into Dynamics, I see I have a notification from the account reconciliation agent, an out-of-the-box agent in Dynamics, that notifies me there's a couple discrepancies in our books that need my attention. For each of these discrepancies, I get a natural language summary that tells me exactly what's going on.

In this case, it looks like somebody posted a transaction directly to our ledger. I can see the posting details here and nothing in our subledger. And if you're not a finance person, that's really bad. We need them in both places so our books balance.

And it's not just this insight that the agent is giving me. It's also recommending an action. It knows our best practices, finance standard best practices, and it's able to make a recommendation that we reverse the transaction and repost through the subledger so it's in both places.

From there, it's just a button, click and a confirmation, and the agent goes to work fixing it. And what's so critical about this is we're fixing this directly in our financials, so that our source of truth can remain just that.

And there we go. We've got a confirmation. The reversal has been completed, and our book accuracy has been improved. And for my audit folks out there, everything I do, everything the agent does, gets logged in our audit trail, so we have that accuracy and accountability that is so critical in finance processes.

Okay, we took care of a discrepancy. Let's go back in and check on our coworker agent in Excel. It looks like we're still working, but we can already see what the agent's done, is created a brand new sheet in the workbook, and we just saw it add some tables, starting to model out the impact of that discount.

What you're seeing is how Agent Mode works live, right in front of you, and it's able to source information from that policy document, from the prompt I provided, and the context that was already in the workbook to figure out we wanted to offer a 2% discount for payments within 10 days of invoice. It pulled an estimated uptake of 40% from that standard policy document, and now we just saw it create a table that's modeling out how that changes our revenue and profitability in a month-over-month basis.

Again, live demo, crunched for time, I want to jump ahead to a prompt that I ran earlier, same exact prompt, same exact core data, and show you the final output.

Here, we can see the agent has developed that full model, and it's not just static text and cells. These are dynamic formulas. As I continue to iterate with the agent, everything

updates with me. And it's not just these tables, it's also charts that really visualize the impact of offering the discount. I can review and make sure this is a good deal.

And again, from an audit ability and visibility perspective, I have a total summary of everything, assumptions, data sources, calculations that the agent did, and key insights that offer me some support to make this decision about the tradeoff between collections risk and profitability. With that, I can keep working with the agent and go back and make a great deal with my customer.

And to recap what we just saw, we saw how Copilot and Work IQ are providing me not just efficient but effective responses to help me do my job, how the Agent Mode in Excel is helping me make great decisions that impact business profitability, and how Dynamics 365 is helping us create an accurate, up-to-date financial data state that empowers the business to make great decisions.

And I want to emphasize all of this is 100% available to you today. And even if you're not a finance person, there's something out there for you. Go home and check it out in Copilot, Excel and Dynamics 365.

Back to you, Satya.

(Applause.)

SATYA NADELLA: Thank you so much, Anne. And when you see that demo, you see a new metaphor of how we work with applications in computers. In fact, if you think back, we've had a couple of very powerful ways we have thought about what computing means to us. There was the Jobs one, which was, there are "bicycles for the mind." There was Bill, I think, in the mid-90s, in Comdex, he came up with this "information at your fingertips."

This one, you can think of as when you work with agents, as us working with an infinite set of minds. It's a good way to even think about how you organize your work. And then the next piece is you're doing this macro delegation, but micro-steering. Literally, when you saw Anne demo, you see a new way the work and the workflow happens where AI does the assist between the steps of the workflow, but you are in the loop as a human, managing many agents as you really streamline the operation.

These are the new ways we'll do information work. And you see that evolution start to emerge.

I'm seeing fantastic – this is not just theoretical talk. I had a chance to meet with many customers right here in the UK, using all this to all these tools at scale. In fact, Lloyds Bank, which is 250 years old, has one of the largest deployments of Copilot. It is great to see how the diffusion of that technology and what they are able to get out of it in terms of time save, how they're then being able to allocate their attention on things that matter. In

fact, their use of even GitHub was fantastic to see, and how that's changing their ability to modernize their code bases.

I had the chance to meet with the Councils from Wales. It's great to see when technology is also coming into public sector, because that means they're able to then improve all of the services we care about. They were explaining in childcare, that it would take them four days to create a report about a child who is receiving services, which now has been reduced to a day. I mean, think about what they're able to do. That means they're able to allocate their time on the core service that they're providing, and it's just fantastic to see that type of real-world impact.

But one of the things I had a chance was to see the NHS in deployment in Manchester University, and it's pretty inspiring to listen to the emergency physician talk about how this tech, in this case, Dragon Copilot has been transformative. Let's roll the video.

(NHS & Dragon Copilot video segment.)

SATYA NADELLA: Yeah, it's really wonderful to listen to their experience with this deployment, because one of the things that the doctor was explaining to me was when in an emergency room, when you're going from bed to bed, one of the things that is a massive cognitive burden is to remember everything, the last time you were at the bed, you learned about the patient. And to be able to count on this ambient intelligence to be there as the rails that you can then query is such a cognitive offload that is super helpful.

It's great to see, again, how these technologies, are really getting embedded in the real world, and then fundamentally helping, in this case, the physicians spend more time with their patients, which is, after all, the core purpose.

Now, the next section of the stack that I want to talk about is the agent platform. Now, we talked about some of these Copilot systems we have built that are getting deployed, that are getting modified inside of all of your enterprises, but the other thing that's exciting to see is all the agent and agentic systems you're building. And the first fundamental bootstrap for any agentic system, as I said earlier, is data.

And so, what we have done is brought this, what we describe as the IQ layer, where you are able to harness the intelligence that already exists enterprise wide, starting with perhaps the most important database in many instances that you have, which is the data underneath your Microsoft 365.

Think about it, the data underneath Microsoft 365 represents your people, their relationships to other people inside the organization, their work artifacts, their projects, their files, their calendars, their communications data. This is massive information, and to be able to then have essentially a stateful system with AI reasoning that then translates all of that into what we describe as Work IQ can be a bootstrap for any agent you build. Going forward, you would literally go in and say, can I have a skill in my agent to look up Work IQ or an MCP call to Work IQ? That's, I think, the way to think about it.

The second piece is all the data in Fabric. When I look at all those Excel spreadsheets or Power BI dashboards, they all have underneath them a semantic model that, over the years, have been created. Whenever you have a dashboard, someone went in and really created a semantic model to make sense of complex data that you're tracking. And to be able to take all of that rich data and then expose it, again, to your agent system is what Work IQ does.

And then Foundry IQ can then take any other enterprise data and essentially think of it as a planner. You can use Foundry to plan your ability to reason over, in fact, Work IQ and Fabric IQ, and any other new system, which has data. To me, that data layer and thinking of it, not just as building an index on data or doing search on data, but to be able to think of your own data layer as a stateful, agentic layer with full reasoning capabilities that can then be skills for additional agents you build is what we're trying to really achieve in that space.

Now, once you have data, the next thing is a classic middle-tier application platform. Every time you had a new platform, whether it was the Web or .NET, or what have you, you built an app platform. And similarly, you now have an application platform with Foundry, which comes with everything that you need in order to build agents, comes with frameworks for you to be able to build agents or multiagent systems. It comes with your ability to do evals. It comes with services for you to have guardrails for AI safety. All of the runtime services that are required for you to build full production ready AI systems is what Foundry does.

Now, one of the core things, of course, that Foundry also delivers, is models. We love to talk about the core foundational models, whether it's the OpenAI 5.2 or 5.3, the Anthropic models, Opus 4-6, but the reality is we have 11,000+ models in Foundry.

And you may ask, why do we need – is it all about one model? And the real world is about many, many, many models. And that, I think, is what is being proven out every day as you deploy these systems. I had a chance, even this morning, to see the deployment of models that are by domain, by industry, by role, that are all needed in order to build a more sophisticated agentic systems.

One example was recently, we released a model called GigaTIME. It's a model that fundamentally uses the ability of Gen AI to simulate, to take a simple pathology slide and translate it into what used to be a sophisticated test that was done previously, which cost a lot of money, took a lot of time for anybody who was going through immunotherapy for cancer treatment, and to be able to then make this just literally a simulation that could bring the test, the pricing down all the way, such that every secondary hospital in every town can have it. Let's roll the video to give you a flavor for it.

(GigaTIME video segment.)

SATYA NADELLA: Yeah, I mean talking about using Gen AI as a technology, and as being able to generate essentially, what is a simulation of a medical test and reducing time and cost, I think, is a fantastic use case. And it just speaks to, I think, the power of what this generation of technology can achieve in the real world.

Now, let's say the runtime is there. All the models are there. You've got your data structure. The next thing is you start building these agents. And even to build agents, in fact, the place where we want to start is us in our own everyday use of information or knowledge work tools to be able to build agents. Agent Builder, built into Copilot, is the place where we are seeing tremendous activity.

In fact, some of the examples I described earlier, of people deploying Copilot are not just talking about using chat in Copilot. They're talking about building agents, using agents, and a lot of it is done, just like if you can build an Excel spreadsheet, you can build an agent. And that's what Agent Builder does. In this case, this is a simple app that they were able to build.

Now, we also have, say, IT has always used things like Power Apps. We have Copilot Studio to be able to then, for IT organizations, to be able to build applications, which are grounded in knowledge, which then are using these models we talk about to be able to build applications. In this case, this is a Copilot Studio app that was built, which is to onboard an HR use case.

Now, for professional developers, what we have is GitHub in Visual Studio. And it's great to see GitHub in the United Kingdom really continue to grow in momentum. We have 5.2 million developers with 18% year over year growth. The community continues to grow here.

And in GitHub, our goal is to not only have the repo, but have all the agents, and have the agent organizing layer close to your repo. In fact, Agent HQ is the concept that we have that you can now go in. In fact, if you go to GitHub, you'll be able to see all of the agents, whether it's Claude, whether it's Codex, whether it is DeepSeek, what have you. You can see them all and use them in the context of your repo. And multiple of them, you can, in fact, have them. And so, obviously, with GitHub Copilot, with slash models, you have all of the models.

And so, what we're doing is building out into GitHub Copilot, all the form factors, whether it's CLI, whether it is IDE or fully autonomous, where you're literally delegating to the agent. All the three form factors matter. In fact, one of the debates is it one form factor to rule them all. The reality is you want, in fact, all the three form factors to compose. I may start in a CLI, or I may start with an autonomous agent, then drop back into an interactive pair programming session in CLI, and then I want to go in, open a file in an IDE. And you want to be able to support all of these together.

And then, as I said, the ability to get to all the models, the most powerful thing is to be able to now, you're able to set up your GitHub repos such that you can even switch

model families. We now are even seeing the emergence of where is Codex great at, where is Opus good at, how do I switch between the two? In fact, how do I set up my harness in such a way that my skills all transfer over between the two, or even doing slash models and then having multiple models reason simultaneously across your code base and come back with their suggestions on what you should do?

We are building all of this out into this one Copilot system that is embedded across all these form factors with all these models. To show you all of this, let me invite up on stage my colleague, Marlene.

Marlene, take it away.

(Applause.)

MARLENE MHANGAMI (Sr. Cloud Advocate, Microsoft): Thanks, Satya, and hi, everyone. Let's say I'm a developer that works at a toy company called Tailspin Toys. A few days ago, I got an email from the Search Product Management team asking me to add some new search and filter features to the site. They've given me a short checklist of what to build, including a search bar with text search for simple searches and Azure AI search for more complex ones. They've also asked me to add in a sidebar so customers can filter by both category and price.

I'd like for Copilot to help me with this task. And so, to do that, I'm going to navigate to my terminal. I'm going to run the command to get GitHub Copilot started.

Oh, sorry. (Laughter.) I'm going to run the command to get GitHub Copilot started. With Copilot CLI, I can now kick off any agentic task directly in my terminal.

For us to be able to get Copilot to help me with my task. I need to be able to have Copilot find the information from the PM team and pull it into my terminal. Earlier, Satya talked about Work IQ, which allows us to connect the M365 suite to our CLI so Copilot can be able to use the data.

In my prompt, I've asked Copilot here to be able to use the Work IQ MCP server, which we're already connected to, and to find the email from the PM team and to list out the things we need to build.

Copilot has connected to the Work IQ skill, which, under the hood, uses Graph API to triage through my emails. And if the demo gods are on my side this morning, it's going to find the email, and then we'll list out exactly what we need to be able to build this morning.

Sometimes it takes a bit of time with Work IQ, but hopefully it will find the email. There we go. Thank goodness, it found it. (Laughter.) And we can see those are requirements that we need to be able to build.

Now that we have the requirements for the search and filter features, we are almost ready to get started. But I want to point out that right now, I'm connected to Claude Opus 4.6, but I'm not sure if this is the right model for the job. Let's see what options we have available.

With a single GitHub Copilot subscription, I have access to all the latest models, including like I can see here, GPT-5.3 Codex. I'm not sure how many of you are chronically on social media like I am, but if you are, you probably have already seen this model come out. It only came out just a few days ago, and it's already available for us to use in GitHub Copilot today. This is really exciting.

I'm going to choose this model, and then I'm going to run a new prompt. In a new prompt, I'm going to tell Copilot to go ahead and start to build out that search bar and the filter that the PM team requested.

You'll see that the model begins by thinking. It's going to analyze both our prompt and my code base to understand what it needs to add to be able to build out these features. This process is going to take a few minutes, so while we wait, let's go ahead and switch over to a new tab.

Earlier today, I asked Copilot to build the same search and filter features and to add the code that it made to a new PR. You'll notice that Copilot is already connected to the playwright MCP server. I'd like for Copilot to use this server to test out the search bar and the filters, just like a human user would.

Playwright is an open source testing framework built by Microsoft that allows users to be able to validate that their features work. To test a Web app, Playwright begins by going ahead and opening up a browser, and it will programmatically control that browser to start to test out the different features. It's going to run a script, clicking buttons, typing in inputs.

And you'll notice that my hands are not on the keyboard. This is all Copilot and Playwright testing the features. We've just seen it running through, manually, typing in, searching for the different toys and also filtering for different categories and ranges.

Perfect. Our tests pass out. We can see that our search bar and our filter options are working as expected.

Now, I have one more thing to show you, and that's documentation. We now have the PR ready to go with our new features, but for developers, documentation has always been important. And I would say in the age of AI, where we're generating code faster than ever, it's probably more important now than ever before.

To be able to help me automate this process, I've created a PR visualizer app, which I've built with a Copilot SDK, which allows me to be able to generate architecture diagrams based off of a PR. You'll see here that I've passed in a PR data tool that's going to take in

details from my PR, and it's going to pass that through to the Excalidraw MCP server, which will generate an image for us. Let's see this in action by passing through the PR that Copilot made in the terminal just a few seconds ago.

We'll see that here in the SDK, it's going to start by pulling in the data from our PR and getting those details. And then Copilot is going to analyze the Excalidraw format, and thankfully, it's been able to generate a diagram for us. On the screen, we can see exactly which code has been added to our code base and which parts of our code base have been modified. I can even go ahead and change, move the demo image around and so on.

This is great. In the past few minutes, we've seen Copilot with Work IQ, pull in a spec and build out the features based off of that spec. We've also seen Copilot use Playwright to functionally test the features of our Web search. And finally, we just saw Copilot SDK generate an architecture diagram to help us visualize the changes in a PR.

Whether you're using the terminal or editing code in the code editor or adding AI to your applications, give GitHub Copilot a try today.

Thanks, everyone. Over to you, Satya.

(Applause.)

SATYA NADELLA: Thank you, Marlene.

That demo gives you a great flavor for one of the things I was mentioning, which is take Work IQ as the stateful agent into itself, and then to be able to use it in another agentic system like Copilot.

One of the things I was doing was there was a design meeting, and I literally said, "Go to Work IQ, pick up the transcript, create a plan based on the conversation we had, to change my repo."

I mean, think about that, right, which is to be able to have a reference to a design meeting in Teams, and then to be able to have that change your code base. That's the type of interplay that you can have between your data, your ability to then parlay that into your own agents that then work in concert with other agents. I think that that shows the power of these systems.

The other fun thing is, who would have thought, right, here we are in 2026 and the coolest thing we have to show you is a CLI, right? It's back to the terminal. But it speaks to, I think, the power of some of these UIs, too. Even in the age of AI, the idea that we will want that chat interface, we will want, ultimately, even a sophisticated numerical thing to show up as an Excel spreadsheet because I know how to reason about it, or even like one of the things is, all of us now we're coding, we've got, what 16 CLIs open. The next thing is, of course, an inbox, which is an agent inbox. I think some of the UI

elements are also becoming clear in terms of how you really are going to work with agents, this idea of having those infinite minds.

That was the agent platform, and the last layer of the stack that I want to talk about is the token factory. This, to us, is about building out Azure throughout the globe, and starting right here in the UK, and making sure that we're really ensuring that the token factory delivers on the core formula, which is tokens per pound per watt, and really optimizes for this performance.

Ultimately, it's about being able to have an efficient token factory, and in order to do that, we have to build out these data centers. We are building out Azure as the world's computer. In the UK, we have two regions. The last time I was here, we talked about how we are expanding that with significant investment right here in the United Kingdom, and we are continuing to grow that out by also making sure that we have renewable energy powering all of this data center footprint.

For us, we have 100% renewable power today that is really powering all of Azure, and we are very proud to be essentially creating the stimulant for renewable energy around the world, even right here in the UK.

The other element of this is, in order to have the best efficiency is to take advantage of the new systems. Because in fact, there's an unbelievable renaissance happening with systems. These workloads now are at scale. Whether they're training workloads or inference workloads, they're unlike anything we've seen in the past. They're data parallel, synchronous workloads.

As you look at them, you now can characterize them, build new systems that support them. They have a different memory hierarchy. You can now think about, like even since December to now, we've all fallen in love with this low latency tier of AI inference. You even saw when Marlene showed Slash models, there were a couple of models which were low latency.

The point is that we now need to have a fleet that truly brings together the greatest innovation from Nvidia or AMD or anyone else, and then into that, we are very thrilled to be making progress with Maia. Maia 200 today delivers 35% or so more superior performance per TCO, when it comes to any leading AI accelerator.

Our goal is to have all of the innovation at the system's level available in the token factory. That way then you can use software and the power of software to be able to actually have the ability to use all of that capability to both train models and deliver models for inference.

Now, the next commitment we have is all around sovereignty, because as you build out the data center, the requirement that now is absolutely front and center for everybody is how to ensure that in every region and every country you have the sovereign controls.

The best way to think about sovereignty is think of it as a portfolio. There is public cloud with sovereign controls. That means you have all of the key management, data encryption, data residency requirements, right in the public cloud itself.

In many, many workloads, that will be a fantastic solution for you to be able to have full visibility. In fact, take an AI workload. You can have local AI data processing. You can use confidential computing, even on GPUs, and you can have all the key management with you. That basically gives you all of what is required in many of the regulations for data residency, local processing of AI, and confidentiality.

Now, the next option, of course, is to have a private cloud, a sovereign private cloud where we have both Azure Local and all of Microsoft 365 Local running on Azure local. That's another option for you to be able to use.

In fact, we are very thrilled today to announce that we have Foundry Local coming to Azure Local as well. That means all those models I talked about, you can now run them in a local data center, in a private data center. We're thrilled about that new capability, and that extends that portfolio of sovereignty.

Not only that, we're also thrilled to announce today that you now have the ability to completely have disconnected operations of your local data centers. Azure Local with disconnected operations, or hybrid operations, gives you even the next level of sovereignty controls that give you the optionality in that portfolio approach to managing workloads.

We also have several local partners when it comes to sovereign cloud, who then have the options of Azure Local, M365 Local, Foundry Local, and this is something that we will continue to build on. This is what will help us ensure that we are able to build out this entire code base.

The next thing I want to talk about, though, is a consideration that sometimes gets missed in sovereignty discussions, which is, as you think about sovereignty, you also want to make sure that you have a first-class consideration around cyber-resilience. Cyber-resilience is a signals game. Cyber-resilience, fundamentally, is about having global intelligence to help you with your cyber-resilience.

This is something that you want to think about, which is, whenever you have a sovereign operation, how do you ensure that the sovereign operation does not lead to some cyber exposure. Again, thinking about all the options as a portfolio, and then really having cyber-resilience, which comes through because of global intelligence, is going to be a very, very important consideration as you think about the placement of your workloads across all these options.

I want to close out where I started. Ultimately, for us, it's about being able to translate this next generation platform, this powerful new technology, ultimately to have a real world impact right here, one community at a time in the United Kingdom.

To us, to be able to make a small business more productive, to make the public sector more efficient, the startups in the ecosystem more globally competitive, the large companies to be more globally competitive, these outcomes, these transformational stories that we talked about is what it's all about.

We're very, very thrilled to have a chance to have you all as partners on this journey of taking this technology and creating more technology on top to bring innovation to the real world.

Thank you all very, very much. Have a great rest of the conference.

(Applause.)

END