

(no title)

57 MIN · YOUTUBE · [HTTPS://WWW.YOUTUBE.COM/WATCH?V=D0PFu6B7gIM](https://www.youtube.com/watch?v=D0PFu6B7gIM)

<https://www.youtube.com/watch?v=d0PFu6B7gIM>

SUMMARY

Satya Nadella discusses Microsoft's strategic investments in AI, particularly the \$1 billion bet on OpenAI in 2019, and the company's ongoing commitment to developing frontier AI technologies. He emphasizes the importance of creating an ecosystem where companies can leverage AI while retaining their intellectual property and value. Nadella also highlights the need for a growth mindset within organizations to adapt to technological advancements.

- Microsoft's investment in OpenAI has significantly influenced the AI landscape.*
- The company aims to create a frontier intelligence ecosystem that allows businesses to build and retain their own IP.*
- Nadella emphasizes the importance of clean data and ethical practices in AI model development.*
- New tools and models announced at the Build conference aim to empower companies to utilize AI effectively.*
- The integration of AI into consumer products is a focus, with innovations in hardware and software to support this.*
- Nadella advocates for a growth mindset within organizations to navigate technological changes.*
- Microsoft is exploring quantum computing as a complementary technology to classical computing.*
- The company is committed to open licensing for AI models to ensure accessibility and innovation across industries.*

Welcome. Uh, today we have Sachin Nadella. Thank you so much for coming. >> Absolutely. It's my pleasure. >> And, you know, realizing it's the finals week, so it's a different type different time of the day, whatnot, and but you know, I appreciate the students that y'all made it here. Um, you know, I was thinking this morning how it's it's kind of fitting that you're the last person uh that we're having to this class. And I say that because if I think back like your bet of putting a billion dollars into OpenAI in 2019 feels like that really set the stage for like this Cambridge explosion, if you will, around AI. So I'm curious to just kick it off like how like what was the thought process to take that to make that bet?

>> Yeah. I mean I think it is fascinating to look back what now six seven eight years in some sense. Um I think the the thing that I I feel at least got me convinced that this was the right thing to go try at that time because um is quite frankly what I describe as a prepared mind right was Microsoft's obsession um has always been in natural language um and um and of course we were mostly focused on uh trying to get to natural language with you know some machine learning some NLP uh but fundamentally if you had asked us even in 2017 2018 it'll be some combination of some symbolic logic plus um uh machine learning, right? So, we were perhaps at that stage not the truest believers that deep learning can even get you NL uh NLP breakthroughs. Uh but that was something we wanted to have happen. So, I would take shots. In fact, most people talk about just the open AI bet, but you know, we bought bunch of companies. I've invested in a whole lot uh of others uh because the fundamental thing we were conditioned to do was anyone who had an ambitious angle uh on natural language

irrespective of what sort of lineage they came from we would always take it whether it was in organically inside the company or outside um and that is when when I you know when Sam and in fact you know we were one happy family at that time right so Dario was there [laughter] >> that's I'm saying that it really did kick off so many So you know everyone was in the same place and so to some degree um the the scaling laws paper came out and you know their their ambition on pushing the transformer with more compute um and data uh was an appealing thing for us to take a shot at. And of course what's happened uh it's pretty stunning that the fact that the capability graph sort of has just stayed at that um scaling law is just pretty amazing. Mhm. It's I mean as someone who worked at Microsoft I guess 20 years ago um you've changed the culture like so much and I think it's it's striking to me that when you took that bet was there an uprising within Microsoft saying hey like we can do this ourselves like yeah I mean I think you see Microsoft over the years I I always say that at the end of the day the core bet has to be the organic bet in what one does inside and then there are partnerships uh there's M&A Okay. Um and I think any company like in some sense you when you grow up at Microsoft you learn that you can create a lot of enterprise value uh by building by partnering right if I look back even you know if you take what you know the PC revolution wouldn't have been possible but for what people describe as the Gates Grove model right which is Intel Microsoft coming together um to create essentially what was a PC ecosystem. So I you know or you know I worked on SQL server and so what we did with SAP to build our database business and for them to build the ERP application right so we are conditioned quite frankly for these type of ecosystem partnerships um as well as organically built and so I I would say there was not an uprising in some sense you know there would have always been like hey you know whenever you allocate your scarce resource or whether it's capital or more in this case it's more than capital the biggest decision was about compute concentration on a particular effort right I mean that was more uh the big uh bet and that's why um you know we made it because this was the group that wanted to sort of go drive it and you know we benefited obviously immensely from it >> and now um one of the reasons why he's in town is there was a big developer conference called build and yesterday you announced this frontier intelligence ecosystem which kind of right in line what you're saying and you had a bunch of pretty huge announcements. Um, and I'd love to talk about some of those.

So, um, you launched seven new models and I thought what was really interesting about at least Mustafa's kind of description of how you did those models is that, you know, all the data is very clean. There was like a lot of focus on um, we say not breaching any copyright things. I'm just kind of interested to hear from you why seven. What was kind of the thought behind that? Obviously, you want your own models which makes tons of sense. Yeah, I think it's um uh if you step back, in fact, I think you call this class the frontiers uh class. And so I think that one of the the challenges of this uh um conceptualizing how does anyone any individual entrepreneur a developer company participate at the frontier, right? There are frontier models, but how does one have real agency um to add value, derive value and protect value, right? Because that's the question, right? If you have a model that basically learns from data, uh what's the future of the firm even, right? Which is the firm today is about tacet knowledge inside the company that comes about because of its operations and human capital. And uh in a world where there is going to be tokens uh and humans collaborating together uh what's the future of the firm right? So there are some substantially big questions and so what our vision for this is simple which is we believe a frontier ecosystem is one where every company can actually operate at the frontier with their own IP compounding over time. not just the human capital but even their token capital. So that is the motivation we have right just like um so for example when the models we built u I'll come back to the lineage there's a nice technical report that I would encourage uh folks to go read uh because I think it's probably one of the most

transparent good detailed document on the entire pipeline uh that's been written lately uh from a a model of this size and I think you'll learn a lot from it but the purpose of both let's take our thinking in coding models was to be able to do this in such a way that we can license it uh along with the weights in and and really allow every company to build their own hill climbing machine.

Right. So we ourselves and I you know climbed our hill using as you said very clean lineage of data making sure we were not you know adding a bunch of synth data in in the mix. Uh so all that was very much true so that we could truly have a model uh that where reasoning emerged. Um and uh and so we now have a fantastic good efficient model but inside of a hill climbing machine that any company sets up it can go learn using the traces uh of that company and those tasks. Right? So our goal is every company starts thinking strategically uh about what's the role environment that they set up what is the private evals that they have how do they then welcome any model uh into that gym so to speak uh and then allow them to retain the IP uh and not leak value. Uh so to me that's kind of what I think every company will need to start doing because if you're just a consumer of a foundation model um then I'm not sure how you can retain uh enterprise value uh let alone create right so so the only way I see this ecosystem quite frankly being non zero sum or positive sum where lots of participants can all uh be at the frontier is they're able to take frontier models, take open weight models, take a model like ours which is a licensed IP uh and then hill climb on their own environment uh and then build out their own IP. So that's the core premise and in and we unpacked that in a lot of detail and all the tooling around it how for example one of the coolest things is if you're a Microsoft 365 customer we can bootstrap even right because after all what is Microsoft 365 today you use it uh to run your business right people communicate with other people related to a business process so you can imagine we can bootstrap the role uh in fact we can even generate the eval uh for let's say an HR onboarding process based on the observation of what you're doing that's unique to the company and first of all it's their data it's just that this think of we built a multi-tenant SAS application we now can turn that into a multi-tenant hill climbing service where the data and the environment and the models and the traces and the outcomes are owned by the company.

>> Do you think that most companies though have the right talent to be able to build those hill climbing machines? Yeah, it's a great one. So that's why I think this is the easy button on it, right? So we are now not saying you need to build so you have the hill climbing machine uh that has been instantiated for you. All you need is a bit of strategic discipline in making sure that these models, the harness, the context, the eval are all artifacts and constructs that you understand and you manage them as assets.

uh just like how you have historically done where you cared about privacy, you cared about confidentiality, you cared about security. I think in a world where AI comes into your company, these things will become as important architectural and strategic considerations. And one of the other products you announced was Scout uh was like around enterprise clause. I'm kind of interested to hear the vision thinking behind that. >> One of the things that uh we're very excited about is like when I look at co-pilot and its evolution, right? It started at uh at with chat um and chat became very powerful especially with reasoning models because you could not um not only you know get uh just essentially use it more like search but you could now really do it as you can use it as a thinking assistant essentially and so that became powerful.

Then co-work was the next form factor and co-work is pretty neat as a way to delegate tasks, right? It's a multi-step uh reasoning tool calling agentic loop. Uh and so therefore you're able to do uh longrunning or a short task assignment. It's very much like what we were doing uh with GitHub copilot let's say even two years ago when the agent loop started coming right. So we now doing it for knowledge work but now with scout you essentially have the third form factor which is autopilot. So now you have the long running agent uh where um it continuously is operating uh it's monitoring it's got a heartbeat it's got you know it's dreaming all the things that you expect from a claw um you can now have uh and you can create it you can have one that with your identity right so I can essentially if I have an entra ID I can give scout my entra ID as a delegated ID and it's sort of essentially my digital uh twin that's uh working on my behalf continuously. Uh but not that just that but we can also allow you to mint more uh autopilots uh and those things can have their own identity and their own sandboxes and so so it's a pretty complete system. So I think of it as an OP enterprise open claw um and a UI that fits in nicely with the rest of the copilot system >> and it makes sense because you have those identities you can really you can address the security question or I mean obviously I don't know how many of you have you know set up openclaw but like I kind of struggled to yolo and and give my credentials cuz I'm like I I don't trust it.

>> Yeah. I mean I think that the most uh yeah we even announced in fact Peter was on stage with us at uh build as well because one of the other things is we were even working with the open claw foundation to make sure uh that it can be run securely. We will have in fact on Windows an out of the box uh experience uh where you can install open claw and have it secured or contained in what is this new um essentially a container called MXC uh which is essentially a way to sandbox the environment right and uh so I think containment is key right because after all you now have these longrunning agents that are able to generate code and execute code um and so therefore or it'll become very important to govern the execution. And so we have a container uh that then you can set policy and isolation boundaries, right?

It can be process level isolation, session level isolation. You can even have a WM boundary if you wanted. Uh for me, for example, I run if anything I wanted to ever run even uh I'll just run it on Windows 365 which is my cloud instance, right? So a complete cloud instance uh that's fully isolated for longrunning agents. So I think we are all going to learn how to work with many agents. Uh and we are also going to learn how to isolate the environments for these agents. Just like how we you know back in the day we thought about processes we're going to think about the process boundaries, session boundaries uh and container boundaries for agents.

>> Yeah. One of the other things that you also announced was around bringing uh we'll say a AI to consumers and um and I'm kind of curious like what does that mean? I mean there's a lot of big announcements on with Nvidia. Yeah. And then there couple of things on that right one is um we're very excited about this concept of unmetered intelligence. uh so you know if you think about it right every PC uh has you know historically the install base had a lot of GPUs if you count the number of PCs with GPUs it's pretty substantial uh you know the dGPU install base so one of the things that we're trying to make sure is that in in a world where um you know these models are there there is applications that are being written the tokens are in short supply we want to tap into essentially the edge compute uh silicon Um and uh so and in that context obviously Nvidia announced a new SOC which we are very excited about their RTX. So we have a Surface laptop which is going to come out um in the fall which is built on it. In fact all our OEMs will have fantastic u you know designs for it. Uh we also

announced a dev box. I mean think about it. It's going to have a pedaflop of AI compute.

It's going to have 20 CPU cores, 128 GB of memory, unified memory for both the CPU and the uh the AI compute. Um and uh uh and it's going to run something like a a trillion parameter model locally, right? I mean, think about uh and by the way, we and and Jensen even and we also worked with Jensen to get uh Windows working on a GB300. So we even have a DGX workstation. So I think of it as a a data center desktop, right? I mean um and so we're I think that there's going to be real demand for all this right because people will want especially when you install something like scout or claw or what have you and I want it to just keep working um you know 24x7 uh and I don't want to get bu build for it the best way to do that is to run it on your laptop or on your desktop uh so we are very excited about just even the rebirth of the existing PC form factor with this new unbelievable functionality brought uh because brought forth because of both the silicon innovation and the model capabilities that now we can have locally. So that was sort of a lot of what we talked about. But the other thing uh that we also said is uh just as there's new functionality in the old form factors uh I think there's a real opportunity to create new form factors for the agent era. Uh so that's where project solara comes in and uh what our goal there is to say you know we showed two uh reference designs one is a badge uh and the other one was a desk u companion if you will um but the the badge is pretty interesting right so you can imagine an agent that has a fingerprintter read fingerprint reader and uh or badge that has a fingerprint reader as well as a camera uh and has enough onboard compute it's a MediaTek uh processor um to be able to wake up something like copilot uh and I can literally get notified. I can you know in fact give it like I can even give it say a coding task or whatever I can dictate to it. it will take the input uh and then go execute it uh in the cloud notify me back uh you can imagine in healthcare a nurse if I was a nurse I was moving uh station to station I could use that to badge in the data uh right versus the phone like right now we're conditioned either we're entering in the PC or we are you know using the phone and in an agent era where you really have ambient intelligence and ubiquitous computing uh you can imagine these form factors now uh that are just endpoints for longunning agents uh that wake up notify um and help you get both output input u that's right there in the real world and so we're very excited about sort of bringing even a platform for it right so we will build some but the goal here is also to have even by the way new platform rules right so Windows has always been it's kind of you know fascinating that we are the only open platform out there right there is no you you can go through our app store or hot you can install uh anything on Windows right it's always had that ethos of being uh not something that only micros you need you don't need to call Microsoft to build applications for Windows right how about that so that's the uh the openness we want even in this new agent platform so that we don't have the carryover of these platform rules that were written um uh for the previous era I'm going to switch gears a little bit so here we're at Stanford University um probably you know the center of the world in terms of AI peel pill and um when you get outside of the you know the Bay Area Seattle um you know people are looking at AI and saying like what what's good for me I think there was a prior speaker that used a metaphor which I found quite powerful which was as electricity came about we didn't sell electricity we sold light >> and what do you think is that equivalent for AI because right now there's not a lot of good messaging around AI of how it's going to benefit people. Yeah, I think that I think that's right in the sense that we have um perhaps gone too into the uh you know the bubble that I guess we all live in um is more about hyping the tech and the tech progress for its sake. Um, and it, you know, we live in it and it's great to be impressed by it and push the frontiers of it and what have you. But at the end of the day, the world will evaluate us in what was the value we created for the world one community at a time. Uh, I mean that's kind of how it's it should always be the case. And so unless I can see the true benefits uh

of this technology be broadspread uh right health we talked about healthcare uh when we suddenly start seeing AI in healthcare change the cost equation the care uh one can get uh not in an abstract you know sense but when it happens to someone in our community in our family uh when even take economic opportunity right talking about this as something where um it takes away jobs.

Uh it's it's clear that any technology that's disruptive will have you know real displacement. Uh but at the same time there's going to be new economic activity where humans will have agency which will have wages which in fact if you think about it right if what is current intelligence gets commoditized humans are the one species that are most adaptive in the sense of creating new value on top of what's the new commodity. Uh and so the way I look and and and that has to not be abstract but it has to be real. Um and it will happen uh but until that transition happens to your point as we go from electricity to light and the light is not seen only by the AGI pill people in the zip code uh but it's seen by the world uh as something that they can thrive in and even my point about that frontier ecosystem right that when every company is not sitting there thinking that oh my god I'm just going to be you know if I let any one of these frontier models into my organization it's just going to run over my, you know, all the IP I've created. That's not a why would they welcome that they, you know, by definition they should not. And so, uh, I think that's why as entrepreneurs, as students, uh, I think we have to shape and and as incumbents, uh, we have to shape this to an ecosystem which is positive sun by definition. If you're not and it's about like a a few firms that have all the returns and everybody else it's all you know um uh in bad shape that's just you you'll by def you will absolutely lose social permission or we will lose social permission. Yeah.

>> Okay. I'm going to switch over to questions. So I'm generally curious about their custom silicon program and I'll use them for their hyperscalers like UML and um Amazon who have kind of made some progress there. It seems like at the hardware side right uh you know their chips that they offer are actually pre- bifurcated like they have like training chips and chips versus Nvidia and AMD have kind of kept a unified chip. On the networking side, we had on Invaded class talk about like the optical men's uh system that they built.

That was really interesting. And at the software side, it seems like they've built their own kind of versions of CUDA with Euron and XLA respectively. Um whereas you guys are kind of building on something uh based off of Trident for mining. So I'm curious like given these sort of different design decisions that these guys have made what have you learned and where are you taking uh your own customers? Yeah, I think I think the the key thing is to sort of recognize um what are the new workloads, right?

Whenever you think about any new um uh uh system, you want to be motivated by what's the new software or what's the new workload. And the good the good news here is that there are these three dominant new workloads, right? there's the training workload, there's the inference workload and now we can sort of say there is the longunning agent uh that uses inference um uh and regular compute. So if you sort of said that's kind of what uh you have then you can start from a first principles uh like looking at hey and these are interesting type of workloads right they're not like the previous scale out workloads these are synchronous data parallel workloads um where you need to to your point to a means I guess point which is you got to even think about the scale up part uh some of the tricks that worked for us for scale out in the past won't work so therefore you now

need to even innovate on the scale up and the scale out to really keep things coherent um and uh you know uh and the MFUs on a training run are maximized and so on. So therefore the way we come out of this and say okay uh even just last yesterday we announced there's a there's Maya 200 that's essentially being co-designed with our own uh uh models plus the open AI uh models because we have that IP and right now in fact my 200 is running uh GPD55 uh you know in multiple data centers powering copilot right so that and giving us total TCO advantage so that's a great way to roundtrip uh for an inference workload code. What's the advantage of that is? Uh we not only did that but we also for um uh uh uh built cobalt uh which is our ARM uh uh processor uh for compute and we're benchmarking it to improve both for latency performance uh when it comes to you know for example the agentic loop right because the place where you need great cores uh are for these agent loops uh so we're using all the GitHub copilot traces to optimize our ARM processor even and bringing all this together with even the networking stack. Um, and so our approach would be to not and at the same time we love to have the GPUs because they're general purpose to your point, right? Which is in fact we're using the GPUs. In fact, we're using the old GPUs in our fleet to accelerate um our data warehouse. Uh so fabric uh is seeing 7x plus performance gains uh because of GPU acceleration. So we think of our fleet as a heterogeneous fleet where we will use software to get the maximum benefit out of it and do smart workload placement. Um at the same time optimize for the high volume workloads like inference and training and agent loop with our own groundup system. And there's lots of design points, right?

Most people get fixated on the AI accelerator, but the AI accelerator is one, the CPU is one, the network, you know, accelerator, the storage accelerator, the AI WAN is another one. Uh, right, you want to be able to sort of really do multi- data center uh hops even. So, lots of stuff. Uh, it's a great time, by the way, to be in computer architecture. Um, you know, I think, you know, when I started in, you know, in the industries when the Patterson book first came out, and that was the risk versus CISK debate. I feel like we're back at uh sort of a time like this where uh you can really rethink uh from the physical design of a data center uh to uh to by the way the electrons. I mean one of the places where I'm very excited about is the the efficiency with which we can bring the electrons all the way uh to the CPU uh so that the tokens are uh that much more efficient right without all the losses in between. So I think there's just a tremendous design space >> and I know another announcement you had yesterday was around quantum yeah >> which is kind of adjacent to our time.

So I'm kind of curious what was the announcement and what was the the kind of recent advances there. >> Yeah. So I so look this quantum you know uh we have been on at this for now the last 20 plus years and um uh it's sort of really exciting to see the progress. I'll just say one of the things is u even independent of the quantum program um even with what we were able to achieve in the last couple of years uh with even the natural atom based uh quantum computers with our stack we've worked with partners on it we're able to generate now these very good traces um and which those traces like for basically if you think about what's the purpose of a quantum computer uh a quantum computer can simulate nature right I mean it's since the the the nature is quantum and so therefore if it's a so instead of you know relying on DFTs or what have you can now have a lot better fidelity um of say chemistry or um or molecular dynamics or what have you and those traces then can be taken back and you can train a model. In fact, we uh are doing that with our material science models uh where you can take the traces from a even whatever an early stage quantum computer uh to improve the data on which you train a model uh for something like uh material science or chemistry. Um now our quantum program itself is as I said there's a software side to it which we will put on ion trap machines which we're putting with partners. We're

putting it on a photonixbased machine. Uh we're also putting it on natural atoms. We have a partnership with uh you know in Denmark called Q North where we believe to have uh a quantum computer powered by atom computing with our stack uh within the year and so on. So that's sort of one side of it. The second side is ultimately in order to build a quantum computer at scale at utility scale um you need fall tolerance. Um our bet on that has been that uh we h there was a a theoretical physicist um uh who theorized essentially um a a state of matter uh called myana uh in the 1930s. Um and so one of the things that we felt that that was the state of matter that we needed to make in essentially fabricate and make real. So we launched our first QPU which was Myurana 1 a year ago which essentially proved out the fundamental physics breakthrough uh that you can actually have this and then um instantiate it. And now we've got Myana 2 which allows this uh to you know be built at industrial scale. Uh and so there's a lot of detail in terms of you know how long these cubits can be stable for you know uh and by the way one of the other things is we perfected the digital control of this quantum computer because that's going to be super important. Um so overall we feel that the quantum program at Microsoft's progressing on two dimensions. one is in the near term uh with even what are the quantum computers that I think are most easy to fabricate and build today with these things like natural atoms uh and then in the long run we want to build out uh what we think is uh you know the what is needed in order for true uh quantum computers to act uh like uh utility scale computers >> on on that ladder if you had to guess a timeline >> you know you know I'm the third CEO at Microsoft uh to keep going on the quantum journey. Um, I would say that the what I'm now a lot more bullish is it may not it's kind of like the AI it's kind of like the previous discussion. I think of quantum as the new accelerator.

Uh, and and remember by the way quantum is not going to replace classical, right? Quantum is not going to be great at storage and memory and so on. It's going to be great at computation. And so you kind of have to marry classical plus quantum in order to do things. And so therefore I think of this as uh maybe a lot more staged even. So if you have a 100 logical cubits with good error correction, we can start using it to generate synth data for uh science models like that'll be a pretty important milestone that may be even more achievable in the short run. Um so we'll see. But I'm like you know we we have you know we we I think made the claim even yesterday that by the end of the decade uh we believe we will be able to build a quantum computer that starts solving some real uh challenges.

>> Real problems. >> Yeah. Amazing. Next question. >> Thanks so much for coming. Um I spent 20 years at Microsoft. Um I joined the year after you joined as after you and became a CEO. um as a market. So I went through cloud transformation, went through pandemic, went through AI transformation and I think um like M for was such a such an amazing experience for me and it shaped the person um how do you think it contributed um and is it still contributing to the culture and uh and success of M? Yeah, I mean first of all we're very very thrilled about obviously students coming in and joining and having essentially the Mac program.

There are a couple of programs like that at Microsoft we created where people can even rotate through uh various functions. Um you know at the end of the day any company for it to be at the frontier so to speak has to be able to get people coming with fresh ideas fresh energy and sh reshaping. I always say to anyone joining Microsoft, you of course you want to come in and learn about how Microsoft works, but we also want the Microsoft to learn from you. Um, and more importantly for you to have the agency to reshape uh what is Microsoft's culture. It's not a static thing. Um, it's a it's an organic thing that gets shaped by the behaviors, the

decisions uh of people at the company. And so we always would uh welcome students coming in uh building their career uh at Microsoft. uh people you know one of the things as a 50-year-old company I mean Mike's an alum at Microsoft and still engaged with us uh we have people who come had a tour of duty gone out come back and so at this point um I think you know I think what the way to think about it is uh the uniqueness of the Microsoft is our core DNA has remained uh right we are a developer tools platform form knowledge worker tools company. That's kind of what we've done for 50 years. But the interesting thing about is us is that we have been able to reinterpret that with every new platform, right? In fact, I joined the company back in the '90s when my existential competition was Noel. Uh and now, you know, it's some foundation lab, right? I had not even heard of 5 years ago. uh but that is the thing that I think keeps us vibrant which is our existential challenge or what we need to compete with uh is new and fresh versus it's the same old and I think that that's sort of an attractive part right when and you come to Microsoft um you will be able to sort of go at that mission of being able to empower people and organizations all over the planet which means a lot um uh to us uh but to do So, uh, you know, recognizing that we as a company can bring a lot to that mission.

>> I have a follow-up question. So, one of the many attributes I admired about you is that you have a growth mindset and you really look at like your leadership team and drive it. How have you instilled that across the company because you clearly have like you just pointed that out that you've been able to deal with these platform transformations? Yeah, I mean at some level the I think it's not uh something you instill per se, Mike. It's sort of you invoke what is innate in all of us.

I think I think the I mean and by you have to do it more out of practice, right? It's it's not like uh you know mostly what I have to exhibit more than anything else is my um ability to confront my fixed mindset, right? Because at the end of the day all of us it's easy to talk about growth mindset but it's very difficult to exercise it individually right has somebody said to me which I have always liked as a sort of a nice quip is everybody likes um uh to change everybody likes change except they want the other person to change not themselves right um and that I think is the challenge of growth mindset so it's not about growth talking about growth mindset it's about having the courage to confront one's own fixed mindset so it can't become corporate dogma to your point right which is so one of the keys the reason why it's worked at Microsoft is we never made it like oh you know some mandate it started with you >> yeah it is not and and also it's like it does it's not like trademarked Microsoft right I mean if you exercise growth mindset or you confront a fixed mindset you'll be a a better human being first you'll be a better colleague a better friend a better neighbor a better parent a better student everything So you're not even doing this for Microsoft's sake. You're doing this for yourself.

And I think giving that oxygen, leaving that at that as opposed to some new corporate thing, uh has been very very helpful. So I'm an advocate of it uh not just at Microsoft anywhere. uh and more importantly it's sort of that practice of uh there are two things that I feel as uh that were pretty influential for me which I learned through sort of my wife's uh you know readings quite frankly one was this thing uh called non-violent communications which is also another form uh of un having developing a sense of empathy understanding where the other person is coming from not having your amygdala always triggered and what have you um so that's sort of one which I think is uh it's a great read if you've not read it. Uh and then of course Carol Dwek's work um on growth mindset. These are two things that you know are I think relevant for children and students and child psychology. uh but I think they apply to corporate uh cultures uh because I think one of the

fascinating things is what Herbert Simon described as the bounded rationality right um I think humans are great uh but we have this unfortunate um um uh you know you you know we don't see what's in our interest all the time uh we get hijacked uh often uh without sort of being able to do uh the the simple calculus of what's the to what does it mean to be at the frontier of our own behavior. Um and I think that these are nice practices that gets us and pushes us. So it's think of it as your training run that you need.

>> That's great guidance. >> So uh thank you so much for the answer and uh I have my piece presentation. I started 10 10 10 minutes ago. I really want to meet you. Thank you so much. >> All right. Good luck. We realize it's finals. Realize it's finals week. Um, next question. >> Yeah. Thanks so much for coming. I was wondering like how do you become such a good public speaker and what do you [laughter] >> I don't know man. I mean I uh [laughter] I'm glad [clears throat] you think uh I'm a good public speaker. Let's leave it at that. [laughter] you know, I I mean, look, I think like anything else, um uh it's not that I I I think about public speaking as sort of a key thing that I'm trying to develop or what have you. Uh but uh uh uh the the the lucky part that I find myself is it um in particular even I think one of the things is when I became CEO um you kind of had to talk about things perhaps that you didn't get the opportunity to talk about previously. Maybe that's a better way to characterize it. But the good news is it's not as if I was not thinking about those things previously.

Uh and that you know I reflected on it is that why is it that I was thinking about those things previously. Uh and I think that comes out of just natural interest. Uh right for example uh thinking about technology but its impact. Uh what does it mean? uh you know I have like many pet sort of passions like what does any technological progress uh like uh AI mean to the global south uh what does it mean to even uh what has been a dream I mean I grew up as a son of a development economist so he instilled in me that hey this convergence growth is going to happen and it's going to be great and so on and so I'm like obsessed about it uh and so as long as I think you have these passions uh that allow you to think broadly. Um, and then for you to be able to talk freely about it, write is the other one. Um, and so, uh, I'm not particularly an expert on public speaking. Uh, but I think the more any of us, uh, can have broad interests that we can articulate and in today's day and age, the media allows us, um, to be able to have our own, uh, outlets. And so I think that this is a great time to both build that interest and then to be able to have different medium whether it's speaking whether it's writing whether it's podcast what have you uh there are variety of ways I think we can express reach debate uh which I think is fantastic >> thank you for coming I was wondering is your undergraduate cel sitting in the audience right now what advice would you give him and like knowing what you know now I guess and what would you tell him like to put his energy into and to like many avoid.

>> Yeah, I mean it's a great question. I mean it's such a you know such a privilege in some sense. I wish I could um right uh because everything is in front of you. You're risk 100% of the time. Um um maybe that's what it is right which is um the let me just say two interesting things. last yeah last yesterday on hacker news I came across I forget one of the CS classes here it had the guidelines on how to use coding agents um which I thought was well done right they had the dos and don'ts and um and u and so the the fascinating thing I find right now is the ability to learn uh new things has become so much easier right uh because you have this very accessible, personalized tutor that's deep uh that you can go uh and work with. And so I would say more so than any assignment anxiety or I don't know grade anxiety or what have you uh you can have one of the terms one of my colleagues uses um real cognitive coverage right uh like test coverage u you can now have cognitive coverage

that really follow you through your curiosity right if I were a back as an undergrad, I would be trying to like it's kind of like what I do with GitHub sessions today, right? GitHub app, which is what are all the coding agents and what are all they doing, right? So that's what I would do. I would be sitting with what are the 10 hundred agents and me at Stanford learning, but I need to have cognitive coverage. It's not I've offloaded to the hundred, right? the the key is what am I instructing them and then when they get something done can I understand what they did uh in order to have learned right it's kind of like a like it's 100 classes uh it's a fascinating and that I think is what I think will happen one of these days somebody's going to break a new uh pedagogy that goes with um like the the tools for example are evolving right like think about what happened in developer tools We went from saying hey we have like 100 CLI to now we need a thing that to manage our CLI CLI complexity which is the new ADE which is kind of like for example the GitHub app is fantastic in that context because it's like the new inbox for managing my sessions right what's the moral equivalent of that that allows a student to navigate through their learning experience um and be max curious uh but really getting deeper faster uh on things that uh you're trying to cover. Uh and I think that that's what I would sort of do and not have anxiety, right? Because you know, you can always push a button to get an assignment done, right? So that's no longer the case that the grains may or may not matter, right? Um and so therefore there's a lot of relitigation on the things we valued.

That's a really good answer. Next question. Hey, thank you so much for coming. So, like whenever we're interacting with a computer, pretty much we always interact with a fooy interface, but uh agents, they they just so happen to be good at coding, not so happen to be not that good at interacting with guey interfaces. So for example, if I want to design a poster and I want a to do it, it is easier for me to get it to generate like HTML CSS to render than to use like a GMI design product. In that case like what do you think uh like implication of like an abilities for like uh buoy interfaces versus a seal? Yeah, I mean I think you're bringing up a couple of different things. one is uh it is essentially codegen is powerful and for HTML and um web UI uh as an artifact creation process uh I think is going to really you know proliferate right so basically we've gone from you know we in fact always bill used to have this thing where what's the difference between you know building an app writing a document or creating um a website you know uh at this point there's none. You can just sort of basically do all three by using code. Uh so that's sort of one side of it. Um uh but I think that you know the direct manipulation is the challenge, right? But but I think the in in the intermediate time frame what's going to happen is you're going to have an intermediate format, right? So you're going to do the HTML and then you can convert into Excel, PowerPoint, PowerPoint into intermediate format and then have agents. So I think that that's what you see in Copilot and elsewhere um when you think about artifact creation.

But the ultimate thing is can you truly uh teach uh even the agent on the model uh the canvas and the direct manipulation of the canvas uh which has to be done fundamentally by teaching it um the semantics of that canvas um and so it has to be exposed whether it's through APIs whether it is through um uh a protocol or what have you and so therefore I think you'll see innovation like that but it is true that direct manipul By the way, you you talked about one of the other things that struck me is one of the nice little features we added to GitHub yesterday was a thing called canvas. Uh and the reason was not because um we wanted to uh the agents need UI but we need UI right because it's now become too dense to just keep tracking my CLI session or the chat session uh because it's you know it's kind of first of all it's linear and uh it's painful like you're trying to sort of scroll through it and so one of the things that we uh said is now we can for example I can have a conbon board as a visualization uh which both the agent and I are working on Uh right. So I think that this idea of generated UI uh

becoming the new way for human agent interaction uh might be one of the coolest things that will happen across all product lines.

>> Next question. >> Uh hi S. Thank you so much for coming. Um if you were at our age like college freshman and you have like the world at your fingertips, what's a problem that you would um encourage us to attack? Oh man, I wish I wish I was that, but um um I don't know. I mean, I it's sort of always interesting to look back and say what would I pick, right? Um I don't know. No, I mean I you know in an in an interesting world like ours right now uh they I think you have to go and say what is the thing that you have inherent interests in uh and the world will value right I think always whenever people are making choices I think they have two things they're trying to intersect right they're trying to intersect something that they believe uh they have a real passion for but they're also doing quite frankly the calculus on what does the world value? Uh right, they always have some destination in mind. I want that career. I want that job. I want to start that company or what have you. And so I would sort of focus on that like answering those two questions uh is what I think will lead and in my case I would probably go get back and look at that right you know it may may be even outside quite fr like right you know one of the things I if I go back to the computer industry I was an electrical engineer and sort of I then drifted into software uh but if I went back now I may go back into hardware um you know just because uh there's just such an unbelievable time um in uh like there are a lot of things that you know I would love to go deep on understanding what sort of uh the optical side of networking would look like uh some of the system so I think that's kind of how I think things will pan out people will pick uh the thing that they're good at then they see the trajectory of what they're good at and say wow I'm going to bet on myself uh to get good at this and start something uh and um and oh or policy right uh um and you know people talk about safety engineering but I kind of was thinking wow there's so many aspects of um what does it mean to have safety around AI uh that require people to think through deeply um and so anyway so there are lots of choices out there >> along those lines have you guys hired like philosophers to help you with the AI kind of guidance >> Mustafa is a philosopher dropout or something. Okay. Yeah. So he's at least [laughter] so we have a quasi quantity philosopher.

Um um but you know but he thinks very you know clearly uh uh from that st I mean he's been obviously since being a founder of deep mind to now he's always thought about it um you know u and we've always had folks in MSR who brought a real deep multidisciplinary approach to it whether it's the economists the moral philosophers um the sociologists um uh and I think I think we'll will always have that.

>> Mhm. >> Next question. >> Hi S. I was curious what do you think about space data centers because I've had a lot of startup uh you know CEOs that come here and talk and you know try to convince me that we should be looking at that and the research I'm doing is showing that like Elon's probably the only one that's going to be able to you know do it profitably just from the launch cost you per kilo. I was just curious what you hear. I'm I'm I'm I'm not an expert sort of on um any of that in both the supply side of it and the economic side of it. Um but it's possible at least what I've read and what I've talked to who are people who are experts in it seems like um you know it makes sense. Um the question really is you now need to not only solve both how do you get there uh but how do you build the stack uh that operates there and then solve all the practical issues of RMA and others uh right because therefore I think there's a whole supply chain uh right when you think about the data center most people you know it's a complex project uh right we built on the shoulders of unbel unbelievable engineering depth, right? Starting from civil engineering uh on and to electrical engineering to mechanical

engineering to then ultimately have it meet um uh the needs of computing. Um and so that level of sophistication for this new payload in space has to get built and it could get built. Um and as far as I'm concerned from a Microsoft standpoint, I would love to. I mean uh uh I think we have a few right instances where we have already had some you know Azure SKS we had a program where we even um put Azure SKS in space and what have you but they were more like edge um and so to the degree to which you know um if somebody says to me that there's a uh there's you know gigawatts available or even megawatts available I'm happy to plug myself in.

>> Thanks. Yeah. So, um it seems that Meta is now pulling back from building French share open models and I understand that Google or Kal are doing some work with open models like flash and the five class of stuff but they tend to be like pretty small like you know for for like running local. Um given how you've personally embraced open source the biggest example being winners. I'm curious, do you see Microsoft MAI open AAI building frontier open models like at the count you know?

>> Yeah, I think the thing that uh we're focused on we definitely will always have open weight models and to your point they will be more for what we will ship in fact we launched two um even yesterday both um an instruct model and a plan model for local agent loop and what have you. So they are derivatives of what we have done with fire silica before uh called ion instruct and ion plan and they'll run on windows and they'll be open way. Um the uh thing that we have focused on with the MAI lineage of models is again think of them as licensed but we are going to license them pretty broadly right so for example you can go to base 10 you can go to fireworks and you can then fine-tune even using their inference stack um and so on. So we are but the reason why we're doing that is because we want quite frankly every company whether it's a SAS company an AI native or an enterprise company to have their own uh model uh that they can then uh postrain that they can RL uh and what have you right so therefore that'll be our goal um is to build an ecosystem around uh the MI lineage of models um um and so and the reason why we want to make sure there's still lic just because at this point you know there's going to be real need for inspections safety uh you know so there's the accounts at which uh we are and even if you look at the Chinese models they're also quickly becoming close source and so on so I think there I think there will be and you know I know Jensen's working on uh some openweight models so and we're definitely supporters of it but uh we want to make sure uh that we are all leading the ultimate goal here is to have everyone have real agency in being able to take some model and to be able to then add to it uh and then protect it from having it sort of go back.

>> So like >> it won't be open. It'll be licensed. >> What do >> So we license the weights? >> Okay. Okay. And and then how does if someone's using it on fireworks or together or something like that, how does that help y'all at at Microsoft? you know, it will be licensed and so therefore we will have an economic model uh in all of us. >> Well, I think we're out of time, but thank you so much for coming. Thank you so much. Uh fantastic.

And