Commercialization Leader - Azure OpenAI Service (Artificial Intelligence) at Microsoft

Expert Details
Compliance
Consultant
2 Other Interviews on Tegus

Commercialization Leader - Azure OpenAI Service (Artificial Intelligence) at Microsoft. The expert works with OpenAI Models, and can speak to LMM's.

Commercialization Leader - Azure OpenAI Service (Artificial Intelligence) at Microsoft. The expert is responsible for leading Microsoft's commercialization of the Azure OpenAI Service. The expert bootstrapped customer and partner adoption and is now programmatically driving customer and partner wins, activating Microsoft's direct and channel sales channels, and driving engineering feedback back to the development teams. OpenAI is one of Microsoft's top strategic investments. Accordingly, this work has Microsoft's senior leadership visibility. All activities are performed in close collaboration with the OpenAI leadership, engineering, and sales organizations.

Prior to Microsoft, the expert was the Former VP/Global Managing Director, of Cloud Strategic Alliance Sales at Equinix (EQX), leaving in January 2021. The expert was responsible for managing a team of engineers, solution architects, business development, sales, and marketing professionals that managed Equinix's 360-degree business relationship with the top global hyper-scale Cloud Service Providers. Expert's 360-degree engagement framework included global buildout of hyper-scale and edge data centers, API level integration of Equinix's interconnection and edge services platform into public and hybrid cloud services, and leveraging the platform integration work to drive joint sales and marketing between Equinix sellers and sales teams within the hyper-scale CSPs.

Expert has over 20 years of experience in the Data Centers/Telecommunications Industry. Their current role is managing a new sales organization that brings together critical players across Equinix into a focused Alliance sales organization that serves as one-stop resource for global hyper-scale cloud partners. The expert oversees +$100MMs in P&L and their focus is on emerging technologies: 5G/edge ecosystem (MNO, cloud, MSP, ISV, edge infrastructure, edge data center), autonomous driving/ADAS, hybrid multi-cloud, management of complex and distributed environments across on-premises, edge and multi-cloud (Arc | Anthos), hyper-scale cloud data centers.

The expert can speak in-depth to the competitive environment in the data center industry with the biggest providers Google, Amazon, Microsoft, and Azure. The expert can also speak to tier 2 companies like Apple and Salesforce and smaller companies like Vantage, Extent, Vertical Bridge, and Databank. The expert can speak very knowledgeable about trends in the 5G ecosystem, technological innovation in the data center space, and the competitive landscape (market share, market outlook, etc.).

See Less

Screening Questions (1)
Q.

Can you please rank the following AI vendors on a scale of 1-5?
Microsoft Azure OpenAI, Google AI team, Google Brain team, OpenAI,
Cohere

A.
OpenAI/Huggingface/AI21. (5/4/1) I am the product manager for Azure
OpenAI Service. I run global product planning, customer and partner
engagement bringing OpenAI Models to enterprise customers, SIs, and
ISVs through Azure. My organization is running projects with 100s of
customers and partners, which provides me a unique view of the market,
use cases, competitive environment, customer deployment trends, etc.

Interview
Call on January 25, 2023
TEGUS CLIENT
Thank you for taking the time to speak with me. So we're spending a lot
of time in and around AI, specifically on the foundational model, and
so wanted to have this conversation with you to go one layer deep and
understand some kind of foundational questions. So let's start off with
your background and then let's go from there, if that works for you.
COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL
INTELLIGENCE) AT MICROSOFT
Sure. Great. So I am at Microsoft, running product planning and go-to-
market commercialization of Azure OpenAI Service. So been running that
for about 1.5 years now. Before that, I was at Equinix. I was there for
about four years as VP and running all their business with the
hyperscale's. Before that, I was at Microsoft, 16 years, a whole slew
of different roles, mainly in sales and in the engineering or doing
commercialization of new products. That's really what I am. I'm an
incubation, commercialization person. And right now, I'm 100% focused.
I've got a whole pretty big team focused on OpenAI, getting Azure
OpenAI Service to market and adopted by customers and ISVs and SIs. So
that's what I do.
TEGUS CLIENT
Perfect. And do you have a perspective on the larger platforms of
Microsoft, Google, Amazon as you fast forward into the future, whether
or not the foundational models are going to be primarily owned by
potentially those three big platforms? Will it be owned by one
platform? Or do you think there'll be a lot of different kind of stand-
alone foundational model or foundational model companies that emerge as
you kind of look into a crystal ball for the future?
COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL
INTELLIGENCE) AT MICROSOFT
Yes, absolutely. So if you take a look at the world, it's divided up
right now. You still have some companies like Amazon that really
believe in smaller models. And when I say that it's really not models
as a platform. But these are all very task-specific, highly trained
small models that do one thing particularly well, like speech-to-text
translation, facial recognition, things of that nature. So the world
has shifted tectonically since OpenAI came out in the market. I mean
you have some smaller players like Stable Diffusion that has got nice
valuations even though they're sort of in that category I talked about.
Even though it's cool, still a small task-specific model. But the world
has tectonically shifted even so in the past couple of months with
ChatGPT. And I mean that's, the days have been very long for me. It has
been a wild month.
And the tectonic shift has happened because of the success, but you
still sort of have this thing between, and you have this in every

company. You have this in Google. You have this in Microsoft. Not so much in Amazon. But it's the sort of like, what do we do? Are we going to build relatively low COGS, meaning low-cost, smaller GPU footprint, task-specific models? Are we going to really invest in large language models? And really, the way that the market is evolving, it's like, yes, we're going to do both. And let me kind of explain that to you because like the best way to take a smaller task-specific model and make it really good is to train a large language model to do one specific thing, well, carve the model down in terms of GPU footprint. Highly tune it on to do one thing, and you get all the benefits of the large language model and a smaller footprint with a much lower cost than the approach that we're seeing today with OpenAI, which is a general-purpose, large language model, foundational model that does like everything really well. So you're still going to see this sort of yin and yang, this push and pull between customers that have task-specific needs and those that want to have a model that does 80% of everything extremely well, which is generally what you get with OpenAI. So who's going to win. I really think the main thing with Google is they're just scared of their own shadow right now. They're really concerned about responsible AI concerns and upsetting their search franchise. So they're kind of paralyzed right now.

They have great large language models, but yet they haven't commercialized them. They haven't built it into Google Cloud, and Google Cloud's in vapor lock right now just because of revenue. I think they're going to have a relatively tough time and reach a point of desperation before they really start to monetize some of their crown jewels that they have with their LLMs. So it going to be a small company thing or a large company thing? I really think that the best way to kill a really good innovation train with large language models is for a large cloud to buy it and consume it into the machine because OpenAI is not a very big company.

I mean they've probably grown. But when they built GPT-three, there were less than 100 people. They're substantially larger, but they're not 1,000 people. But still, they're very well-funded. They have a very unique culture. And with a company like that, you just don't want to mess it up by acquiring because then everyone's got handcuffs for two or three years and then all your best developers go onto the next start-up. And that's what would happen. I think the model of either acquiring a company, keeping it as a separate entity or what's going on in my former life in the data center business where you're going to have a lot of private equity, very well-funded start-ups that Microsoft is doing with OpenAI, just really funding the company very well. I think that is the best model because it keeps the creativity train going.

And then it just results, as you've seen with Microsoft and OpenAI publicly announced just this past Monday, that there is an exclusive relationship. The large language model of ISV they really can't be anything other than exclusive because like who in their right mind, if you're Google and Amazon, is going to tie up with OpenAI given that everything is already built on Azure and all you're doing is carving up the market with a single provider? It will be the addressable market, the TAM isn't going to go up. It's just you're going to take the pie and cut it into two or three. I can't talk from a proprietary standpoint, but once a company becomes ingrained in your cloud infrastructure, it's very difficult to decouple it.

Just the way that the architecture is built and the way that the models are scaled out, it's just difficult and very expensive to stand up another supercomputer in another cloud. I think the bottom line is that

Google has got a, got to get a little more aggressive and risk-averse if they're going to really compete and win in the space and embrace the fact that they can enhance their search dominance by implementing a really good AI strategy. I can't tell you how many people, like all I say is ask ChatGPT a question about current events, and it will fail out on you. We could talk about that, but that's my general view.

TEGUS CLIENT

Let me ask a follow-up question. So you mentioned that there are smaller task-specific or use case-specific models that are much cheaper. Are they also better?

COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL INTELLIGENCE) AT MICROSOFT

The answer is no. They're not better. Like if you take a look at.

TEGUS CLIENT

Let me make one other point there because another point you made was if you're looking to do something task-specific, you can use smaller models, and that's fine. But some customers want to do many things very well. And ChatGPT is great. You can get 80% of the way there with just using ChatGPT. And if you're using a single use case or a specific task on using a smaller model, does that 80% go to 95%? Does that 80% go to 85%? Does it go to 99%? That's the question. And then I'm going to have follow-up. I don't want to lead.

COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL INTELLIGENCE) AT MICROSOFT

And it really depends on the industry vertical you are in and how domain-specific your information is. Because a task-specific model will do something very well. It will give you positive, neutral, negative sentiment. It will be entity extraction. But it's not going to do any generative things very well. But if you're in classic NLP type stuff, classification sentiment, embeddings like doing search embeddings and machine learning, what else, yes, I mean conference call summarization, I mean you could develop some models that really shine in that. Like with Microsoft did with text analytics for health, you can create a domain-specific model to do language text analytics like really well for medical.

However, what you run into is that companies that are using these types of smaller models in like conversational AI, they're finding that this doesn't produce human-like responses in the way that they used to, or the sentiment isn't nuanced enough, or the classifications are not as clear-cut as it, they are more complex than it seems. So what I find is that particularly companies that are using spaCy models, BERT models, things of that nature, get much better results out of the box with OpenAI that's just been, not even fine-tuned just with few-shot learning, for instance. And in a head-to-head comparison, I see OpenAI doing a better job than a lot of the language models that are out there, certainly a lot of the open-source stuff that's out there.

Then you get into the whole cost performance thing because the cost of running a large language model, I can't tell you how much more expensive it is, but it's pretty significantly more expensive in terms of the price you're going to charge for the model. You could take a look at pricing that's published today and look at OpenAI pricing versus some of the published pricing for Amazon, Google and Microsoft model. It's pretty significantly higher.

However, people are really willing to pay that price premium for text-davinci or ChatGPT when that gets out just because of the improvements that they get in results quality even for specific tasks. But people are not just using it for sentiment. Like for example, if you're in a contact center or a call center, you're going to use a speech to text

to get the call transcript into text format. And then you're going to summarize it. You're going to classify it. You're going to look at the customer sentiment. You're going to use it to write a response back. There's so many different things you can do that if you try to do it with multiple models, each is a different API call. It requires significant coding and integration. And simplifying your application pipelines just using a large language model makes a hell of a lot of sense. I'm seeing are just loving it, and more demand than I've ever seen in my long and storied career. I still think there's going to be a place for smaller models. However, I think that the use of the large language model is to make these smaller models even more effective than they are today, so they can produce human-like responses for like a Q&A bot, for instance, because right now, they don't work very well. Or you augment the search capabilities and then use this task-specific model for conversational AI.

That's what I think certainly the case of what's going on with Microsoft if you read the blogs that are out there. I think Google needs to do that, too. Amazon, I was kind of waiting for the shoe to drop there. I think they're a little concerned about Alexa right now.

TEGUS CLIENT

And as you think about the future and if we go one step further there and think about completely vertically integrated models and use cases what's your viewpoint on those sort of companies and their ability to build something that is, stands the test of time?

COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL INTELLIGENCE) AT MICROSOFT

When you say vertically integrated, meaning like an industry vertical type model or.

TEGUS CLIENT

Yes, they'll have a specific use case. So the example would be they would have their own kind of foundational model. They wouldn't open it up to third parties and let third parties build on top of it. They would build their own applications on top of it or application on top of it and then feed that specific data, that specific proprietary data back into their dataset.

COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL INTELLIGENCE) AT MICROSOFT

Yes, building a model, incorporating into an application, selling the app. Yes, I get what you're saying. Increasingly, you're seeing those companies move to large language models. I see hundreds and even thousands of ISVs wanting to use OpenAI because it allows them to not focus on the AI component of it, but yet focus on the other aspects of the application. So I do see still a lot of ISVs out there that aren't a sales culture. They're a coding or engineering culture. I see those companies continuing to do what you described and build their own models, very task specific. But the potential for disruption is huge because a company that has 1/10 the burn rate because they're using OpenAI as opposed to hiring 30 developers is going to stand the test of time.

So I see ISVs that are particularly entrepreneur, like a sales-oriented culture, not as much of an engineering culture, really gravitating pretty significantly towards large language models and OpenAI right now. So I can't quite call them dinosaurs, but I get very concerned with companies if they think they're differentiating based on their AI model. I'd rather see them focus on customer acquisition and building that application layer that consumes the model. That's what I'm seeing right now.

TEGUS CLIENT

Can I ask one related question to this? And so just kind of want to understand that relationship between sort of maybe OpenAI customer and OpenAI as it relates to this. So let's say X company builds a product on top of GPT-three that allows them to create, I don't know, sales and marketing copy, for instance. When that company X, goes out and sells to their end customers, gets a lot of product feedback, get a lot of iteration. They're seeing a lot of usage and getting a lot of feedback. Does that feedback go back to OpenAI and sort of enrich GPT-three in the next one? Or is that sort of product feedback encapsulated within company X?

COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL INTELLIGENCE) AT MICROSOFT

That is a fantastic question. I just got briefing, just got done briefing 1,200 Microsoft salespeople on this very topic. So, it's something there's definitely a difference because any customer can work with OpenAI or Azure OpenAI. It's basically the same models. You have text-davinci-003. Both companies will be offering ChatGPT.

The pricing for inference is about the same. So what's truly the difference? With OpenAI, you're going to get if you work with them, you're going to get the people that built the models. It's like playing the piano with the piano instructor as opposed to the students. You're going to the primary source.

But you're also going to get an architecture that isn't as scalable, resilient, doesn't have a service-level agreement. If you tried to get on ChatGPT yesterday, it was down. It was out of capacity. And you're not going to get the same capacity planning as you do with the Microsoft. Now we're very proud of OpenAI. They sit on Azure infrastructure, but the point is it's not enterprise-grade. It's not enterprise ready. Azure is or Google will be or Amazon would be. So any time you're working with a Tier one cloud, the way that they approach uptime, service architecture is going to be the same as they do for any of other cloud services.

The second thing to answer your question directly on data and data governance and data security and IP, that's another pretty significant difference because with OpenAI, yes your data can be, go back and used in model train. So Jasper, big OpenAI customer, their data can go back and train the next version of ChatGPT, which some people perceive is competing with Jasper. So there's a lot of complexity there.

With Tier one clouds with Google, Amazon, Microsoft, the sort of the data security policy is that your data is your data. If you're an enterprise customer or an ISV that's putting your data into the cloud to train, to fine-tune a large language model, that data is never accessible by the cloud. Amazon, Google or Microsoft will never access your data unless the models are being used to generate porn or hate speech or something like that. That's a different story. And there's content moderation services built into large language models. But when you, the big difference is you work with a Tier one cloud. You don't have to have any concerns about your data being used for model training. It's just on a policy to do so.

TEGUS CLIENT

So you used the Jasper example. So Jasper is leveraging kind of OpenAI model. They're also generating their own kind of those models are based on the data that kind of OpenAI has. Jasper has also generated its own data through its application. Is that data then going back into the data set? I guess the answer to that is no. The question is, does the next iteration of ChatGPT, will start, will it incorporate the data that is being produced by the applications that sit on top of it? Or will it always have its own data set and not use the data, the

proprietary data that is being generated from the applications built on top of it?

COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL INTELLIGENCE) AT MICROSOFT

But what I can say is that most enterprise customers read OpenAI's terms and conditions and raise that as a concern that they have, that their data will be used to train the next model. And it just completely violates that risk and violates their corporate data governance policies. I don't really know, and I don't think anyone with Microsoft, really knows how OpenAI uses customer data. But the terms and conditions that their customers leave them to make the decision that they can't use OpenAI in production just because they can't use it on their own data.

They can't go in and upload their data to do fine-tuning. It violates their corporate governance policies. So that would lead me to believe that, yes. The risk is either there or the risk is heavily perceived and deemed too risky by their legal departments to enter into an agreement.

TEGUS CLIENT

We'll just use a fabricated company. We'll call it like J Company for instance. You're saying J Company, it's, the data that J Company produces based on the contract and the terms and services is not being fed back into ChatGPT or the model to fine-tune the model. Or are you saying it is?

COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL INTELLIGENCE) AT MICROSOFT

Yes. I mean I don't know if it is or it isn't. I got to be frank with you. can tell you is that enterprise customers that review OpenAI's contract basically say that they have concerns that their data will be used to train the next version of the model. I don't, never sat down and nor has anyone at Microsoft taken part of the training of the models, but a couple of hundred customers have told us, we want to use you, Azure, because you have your Azure data security policy and so forth. It's different than OpenAI. So yes, I'm kind of walking a fine line here.

TEGUS CLIENT

I think the statement that you said was that OpenAI does not receive product feedback, and I thought early in the conversation, it was stated that it is fed back to OpenAI. So I couldn't figure out what's the.

COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL INTELLIGENCE) AT MICROSOFT

So the reality is that enterprise customers will not use OpenAI if their company data, call center logs, financial reports, et cetera, are used to train models or are uploaded in terms of doing few-shot learning or something like that because when they read OpenAI's terms and conditions, they have concerns that OpenAI is going to be using their information to train their models. That's the situation.

TEGUS CLIENT

In your perspective, or your point that you make is if you're working with Azure, that will not happen?

COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL INTELLIGENCE) AT MICROSOFT

Well, it's any Tier one cloud I mean GDPR, for instance, in Europe, that data can be disclosed and observed or monitored or logging and so forth. It's just PPI in there, it's like you're violating, you're causing your customer violate GDPR. So yes, any Tier one cloud. It's not just Azure but Amazon, Google, Microsoft, IBM, DeepMind. I mean all

of them have the same type of data governance policies, data security policies that your data is your data. It's not used. The cloud doesn't have access to it, and it is not used to train models.

TEGUS CLIENT

Perfect. And as you think about some of the companies that are building on top, so they're building on, let's just use the fictional company. Let's call it J Company again. And so are they then fine-tuning and creating their own model to continue to improve their product and using their own proprietary data to enhance the product?
Meaning if another company, let's say, K Company wanted to come in, they would have an inferior product because J Company had a head start and has all this proprietary data and has been fine-tuning its own proprietary models on top of ChatGPT and other OpenAI models?

COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL INTELLIGENCE) AT MICROSOFT

So like the way that the foundational, large language models work is that the company that produces them, OpenAI, they take public data sets. They actually license data like image libraries for DALL·E, et cetera. And they pretrain the model. So if there's 100 ISVs, 100 start-ups that have access to OpenAI, they all start with the level playing field. They start with exactly the same model as everyone else has.
What companies can do is they can take the models without changing the models themselves. They're not changing the model weights, but they can fine-tune the way that the model will operate on their data by uploading and by fine-tuning the model based on their data.
It's not changing the underlying model. You're just fine-tuning the model based on your particular content. So like J Company, if they really have talented developers and are very skilled at prompt engineering and fine-tuning and they have really good data scientists, they will make better use of the model than Company K will if they have an inferior data science team because like the Company J will be able to finesse the models and get them to do things really super well.
Like Jasper is a perfect example of something like that. They just are like rocket scientists in terms of using it for content generation. Another company to take a look at, Zammo. I mean they're a great company. There's conversational AI companies that have a very surgical approach for how they're going to use large language models. Keep their conversational AI bot, digital assistant, whatever you want to call it, but then augment specific functions of that by bringing in a self-fine-tuned version of the large language model.

TEGUS CLIENT

And what does fine-tuning mean?

COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL INTELLIGENCE) AT MICROSOFT

Yes. Fine-tuning means that what you're doing is you're bringing up, so when you use a large language model, if you get on the OpenAI playground environment, what you're doing is you're just inferencing against a model that's multi-tenant, that's shared across like hundreds of thousands of customers. So there's a pool of GPUs. You got the model up. And you're inferencing against it and someone else is. What you do with the fine-tuning is you take sort of an instance of that model that sits on a defined set of GPUs. So it's like, it's almost like setting up a server for yourself. Remember the days of web hosting. You have shared web hosting, $19.95 a month, and there's like 10,000 customers on a server. And then you had dedicated hosting, where you have your own server. It's very similar with that.
But, and that enables you to take a data set you have a medical terminology data set. You can just basically upload that data into the

model and train the model to do language understanding, to do NLP, NLU, NLG. That's highly adept at doing it on your domain-specific data set, your company terminology, your company acronyms, medical terminology. Like you can train it on specific languages. Like the degree to which a large language model will support a language really well depends on the degree to which that model is on the Internet. But if your language doesn't do very well with OpenAI, you just train it on a bunch of large data files based in that language.

So yes, that's really what fine-tuning is. It's getting the model to do what you really want it to do on your domain-specific topic by uploading a representative data set up into the model, occupying that dedicated GPU cluster and then coming out with an instance of the model that's very tuned to what you wanted to do. That's kind of what Jasper's done.

TEGUS CLIENT

And then you're saying in that scenario, that fine-tuning stays with Jasper and does not go back into OpenAI, and OpenAI will not be able to use that fine-tuning data for their next version?

COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL INTELLIGENCE) AT MICROSOFT

You see, that's the gray area that's exactly where you take a large bank or a medical health care company. That's when they get really concerned because when they read the contract, they get the impression that OpenAI can take that data that I spoke about and use it to fine-tune the GPT-4. I'm just like making stuff up here. But they don't have that same concern with a Tier one cloud because of the precedent has been set years ago that the Tier one cloud doesn't have access to that data, will not use that to train models. It's contractually committed to in the cloud.

TEGUS CLIENT

So for instance, can someone purchase through Microsoft, fine-tune their model, and that data never goes to OpenAI?

COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL INTELLIGENCE) AT MICROSOFT

That is what happens in 100% of the cases.

TEGUS CLIENT

So if we fast forward and let's assume the data privacy piece becomes increasingly important and no enterprise wants to share their data back to OpenAI just a hypothetical scenario, no one ever wants to share any data back, how does that model eventually train over time then?

COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL INTELLIGENCE) AT MICROSOFT

Yes. I mean periodically OpenAI will be GPT-two, GPT-three, the GPT-3.5 model series, which are the text-davinci-002 and 003. And then next on the train is GPT-4. I mean they will go through basically a development exercise and pull in like a larger corpus of information or more public up-to-date information.

And that's what happened with ChatGPT. And you can ask the questions, and it will literally give you the data. I can't remember what the cutoff data, which you can't comment on things because that was the date at which the data source stopped the end of it. So it's just subsequent releases of the model that you should bring in more training data and train the data up to a certain point in time.

TEGUS CLIENT

Got it. So the way that the model gets trained is through kind of these public-facing releases of GPT. They're not. In that scenario, they're not getting trained from any of the enterprise use cases kind of built on top of OpenAI?

COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL INTELLIGENCE) AT MICROSOFT

Well, in, yes. I mean it's a self-fulfilling prophecy that customers, enterprise customers that have data governance policies do not put their data on OpenAI. Like you just don't see it. I mean banks, insurance, regulated industries. I mean anybody who has SEC reporting requirements, they just don't work on OpenAI with their data. They just use synthetic data, dummy data, public domain data. And they hit a brick wall because OpenAI is going to use our, potentially use our data in model training. That violates our corporate data governance and security policies.

I mean is it's a self-fulfilling prophecy. They just don't do it. But you take the thousands of start-ups that are using OpenAI in production as there are thousands, hundreds, whatever the number is. They don't have that same level of concern just because they're a start-up. They're not as concerned about that type of stuff, I guess. Even companies like Jasper. But then again, I can't comment on the individual contractual relationships between a Jasper and an OpenAI. I frankly don't know what their arrangement is. They may have an arrangement where no data is to be used. I hear and see is like companies are concerned about that, enterprise companies. So they choose not to use OpenAI for anything other than ideation and experimentation.

TEGUS CLIENT

I guess one just related to this topic question is when you actually look at what customers are trying to build on top of whether it's OpenAI directly or OpenAI plus Azure, what are you seeing? And the question is around like is it just a company that's using one model like GPT-three to do the whole service? Or are they interweaving two, three, four, five different models to actually create the service that they want to create?

COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL INTELLIGENCE) AT MICROSOFT

That's, yes, companies will start off with a single model and a single task. Like they'll say, I'm using a BERT model for entity extraction. I'm going to compare the results I get with OpenAI with what I get from my BERT model that I'm using. And they'll say, wow, OpenAI is a lot better. And they start off just by comparing what they have today to what can be produced by the large language model. Then as they like, and there's a few use cases that are very common report summarization, customer feedback sentiment, call center transcript summarization, writing of e-mail, direct marketing e-mail tag lines. These are very task-specific things that summarization of customer service commentary out of Google or Yelp.

And then they start to say I could daisy-chain instead of using six different models, my contact center. I'm using OpenAI to summarize. Why don't I also use it to detect sentiment of the customers and classification of the incoming requests by classifying them by product, route the angry customers to the A team in my call center? Oh, and then I can use it to summarize the call, and then I can use it to write an e-mail response back. So suddenly, they're making six API calls to the model to do six different things, all on text-davinci-003 fine-tuned with their call, with 100,000 of their call center records. And that's what people are doing and increasingly. They're starting off small, and then they're daisy-chaining things together.

We also see companies, ISVs, usually start-ups, starting to do, to group together GPT-three with DALL·E for direct, for marketing campaign personalization. I want to do this marketing campaign in 18 different

countries, but I have a stock image like a Getty Images of like a palm tree on the beach. Well, you can take DALL·E, and you can modify that image to cover different cultures for instance. You could give me a picture of this widget in South Africa, give me a picture of this widget in France. And actually personalize or even hyper-personalize. Like let's say high net worth individuals. You can create a marketing candidate specific for a billionaire.

So that's going to combine GPT-three with DALL·E for instance. And all these models are going to be multimodal anyway. But the trend is companies start small. And then they prove themselves internally that these models are useful. And then they expand across multiple use cases. So it's pretty much a very well-established pattern that we see.

TEGUS CLIENT

So they expand against across mostly use cases, but do they expand across multiple models as well?

COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL INTELLIGENCE) AT MICROSOFT

That's a really good question. And I don't see that all that much. I see, and it's really because it's different teams that are going to be using DALL·E for text image. It's just going to be the creative ideation people.

Like there's a public case study about Mattel using it to, you remember, Hot Wheels cars, those little die-cast cars. There's a great case study out there that my team did on designing Hot Wheels cars using DALL·E. There, you're going to have the creative ideation team for like natural language to code. There's another case study with a company called Trovant And they're basically using the Codex model to develop, to generate what are called docstrings, which are like in-code explanations of what that piece of code does.

They're developers. They're going to use Codex. And then you get all the language people. And it's a skill. It's an art. It's a specialty. The people that are doing NLP, NLU are not going to be the people that are going to do Codex and coding. Just different organizations within the company. And you really don't see it all that often. But what you do see is like one large multinational having there's some systems integrators where there's 20 different people working on the service that have no idea what each other are doing. And you see that in large multinationals, different divisions that are siloed or using it for completely different purposes, unrelated and unknown to each other.

But what you don't see is one development team going and doing work on Codex and then doing work on DALL·E and then GPT-3. What everyone's interested in is ChatGPT. That seems to, like I really think that world hunger will get solved because of ChatGPT. I'm being facetious.

TEGUS CLIENT

It was kind of the extension of my colleague's question, which is the difference between working directly with OpenAI and working with OpenAI and Azure. And what I mean by that is it seems like with OpenAI, there's some language in the contract, and there's something happening functionally where there may be some visibility into what inferences that they're sending out to their customers and what feedback they're getting from the customers.

Like there's sort of something architecturally or technically that allows them to have visibility into that because that's the thing that's up for question is like can you use that or not? So the question I have, just from your vantage point is that true? Do they, do you have like visibility into that? And then the second question is like with Azure, how is it different? If you build on Azure, my sense is like you

have visibility as well. You just do something different from a governance perspective. But yes, that's sort of the question.

COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL INTELLIGENCE) AT MICROSOFT

Yes. So it all comes down to contractual commitment. The Azure terms and conditions, what a customer signs there clearly state that Microsoft does not have access to your data. Your data is your data. And it's a contractual commitment. And the systems are architected, our people just don't have access to the data. There's no way for people to gain access to the data. It's like architecturally built in, the safeguards are built in.

The only sort of edge case that, I think I mentioned this at the early part, was there content filtering, content moderation that is a service that runs within the Azure OpenAI Service that detects hate speech and so forth.

I mean if someone is using it to develop political commentary that's going to incite violence and attack the Capitol in the next election, that data that is identified by the content moderation like Google, Microsoft, Amazon, we're going to want to see that. But companies that have purely internally focused use cases, they're going to say, we're not going to generate this type of speech because it's our employees, and we know what they're doing. Well, then we'll relax that policy.

It's not the same with OpenAI. I mean people have the perception that their data will be used to fine-tune the next instance of the models.

TEGUS CLIENT

And then there's the underlying implication that they think there is visibility into that. Like it's a switch they can turn on if they wanted to. So.

COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL INTELLIGENCE) AT MICROSOFT

Yes, it's a perception, yes.

TEGUS CLIENT

Got it. I just wanted to do one more clarifying point on this because from OpenAI's perspective, there's visibility as I think about it in two ways. One is just the raw data that's being generated for Jasper, for instance, the raw data that's being generated from the application, and I think this is, the raw data that's being generated from the application. Does OpenAI have visibility or access to that data? And then the second is when Jasper uses its own proprietary data to fine-tune its models, does Jasper have access to that data? So I bifurcated between those two points.

COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL INTELLIGENCE) AT MICROSOFT

I even make a third leg is like the user interaction and what happens there.

TEGUS CLIENT

Yes, I think that was my first bucket I meant. Just the raw data that's generated from the application. Do they have access to that? And then separately, it's in and around the fine-tuning. Do they have access to the fine-tuning data? But I'm getting the sense that we have a suspicion. We don't know for sure.

COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL INTELLIGENCE) AT MICROSOFT

Let me jump in. I've had many customers tell me that they read OpenAI's terms, and they're concerned that the prompts and completions that are generated by the enterprise are going to go into fine-tuning the OpenAI models. So there's a lot of angst and concern on that just from like the way that the terms and conditions are written. So I think it's

pretty clear that there's like at least a strong perception that there's an issue. Every Wall Street bank is telling me this.
TEGUS CLIENT
And you're saying if you're with Azure, that concern is completely removed?
COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL INTELLIGENCE) AT MICROSOFT
That is correct, unless your people are using it to generate hate speech, porn, that type of thing.
TEGUS CLIENT
But it's still not being said, even in that scenario, it's not being fed back to OpenAI?
COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL INTELLIGENCE) AT MICROSOFT
No, never, ever. There's no data exchange.
TEGUS CLIENT
Can you still fine-tune the models if, without feeding the data back to them?
COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL INTELLIGENCE) AT MICROSOFT
Oh, absolute, yes. Because when you're doing model fine-tuning, you're not changing the underlying model. Like you're not changing GPT-three, text-davinci-003. What you're doing is you're enhancing the capabilities of that prebuilt model to do something you really needed to do based on you.
TEGUS CLIENT
And you can do that in a silo with, you can do that in a silo without sending them the data?
COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL INTELLIGENCE) AT MICROSOFT
Yes. You are not modifying the underlying model. You're just loading the data in the memory. It's being used to tune the instance of the model you've spun up. So what you do is you will create a resource. And that instance of your model is up and running on the GPU infrastructure. And then you can tune that model, bring it down and store it and then bring it back up and inference against it.
So these are sort of the femoral resources that are created by you. It's the same thing is creating a virtual machine. You spin your virtual machine up. You spin your model instance up. It's not like when you tear down that VM, you have to recreate it. It's there. It's just not being used. And so that's the way it is. You're not changing the underlying fabric of the VM platform. It's just your VM is customized, yes.
TEGUS CLIENT
Thank you. Perfect. You said there's a ton of excitement here. We're going to solve world hunger. Where is the density? Where is all the excitement? What are the use cases that you're seeing?
COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL INTELLIGENCE) AT MICROSOFT
Yes. So if you had asked me a few weeks ago the top verticals are going to be financial services, ISVs, telecommunications. Those are really the top three that we've seen. But since the general availability announcement, all industry verticals are extremely excited. Where I don't see excitement is within systems integrators. Like I don't see a lot of progress. I don't think SIs have figured out how they're going to make money because you're not going to have a $500 million engagement fine-tuning the model. But the biggest excitement is on ISVs and start-ups. That is number one. Financial services is huge. Health

care is quite big. Retail is emerging. There's some really interesting stuff with retail.

Now on the use cases that we're seeing the highest volumes, really what you're seeing is contact centers, very, very big. Large subject matter expert document summarization is really big, especially in financial services. And you've got like 60,000 analyst reports. Just like using a large language model to summarize that takes your analysts and puts them in a position of being able to spend more time doing deep analysis as opposed to poring through a document. So that's a huge one. The other is natural language to code. With Codex, we see a lot of companies, particularly gaming companies, doing that, building a plug into a game development platform that allows citizen developers to develop games.

That's a huge one. Let's see. Search using OpenAI to basically enhance the search capabilities. Another one is conversational AI. That one is huge. Chatbots, avatars and virtual assistance. That's probably the biggest thing that we see right now, especially with ChatGPT. It seems that every CEO in corporate America has asked their CTO when they can get ChatGPT. And I'm not kidding you.

TEGUS CLIENT

As a chatbot and like an avatar?

COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL INTELLIGENCE) AT MICROSOFT

Yes. I've got some companies that said we want to take and have every employee in the bank have it on their desktop. Like a few hundred thousand people, yes, yes. It's become, like people think of it as the new Google Search I can train this thing with all my information and current events, and then like it will replace Google. It's pretty interesting.

TEGUS CLIENT

Was that a related comment to the avatar piece, the CEO and the avatar piece? Or no?

COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL INTELLIGENCE) AT MICROSOFT

No. Avatars are a real tough one. I'll be frank with you. Like taking technology and having it interact as a human, I know we don't have time, but most of the stuff that we're seeing right now is straight conversational AI bots like a chatbot. Either one that's built from the ground up, so the three types are, if it's built from the ground up or if they want to like rip and replace one of their existing language things or what we call domain failover.

Like a travel site. Someone's asking you what's the weather like in Paris in January. Now that bot will like book tickets, but it won't give you information. We call that domain failover. So it's basically augmenting the bot to be more robust in terms of what it can do. But that's a huge one.

TEGUS CLIENT

And when you're, I'm just going to rapid fire right now with kind of yes, no. So when you would say ISVs, are these like public software companies, not these companies specific but companies like or similar to Salesforce, ServiceNow, Workday, big public SaaS companies, things like that? Is that.

COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL INTELLIGENCE) AT MICROSOFT

No. These are friends and family through seed, through Series D, all the way up to small, like mid-cap ISVs. The big whales have their own AI stacks. And like they almost look at it as competitive in some ways, but yes.

TEGUS CLIENT
So when you're out in the ecosystem talking about this, you're talking about you selling ChatGPT or you're selling OpenAI, something within OpenAI, what are the other names that pop up? Like whom else has a lot of notoriety right now or heat in and around them?
COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL INTELLIGENCE) AT MICROSOFT
So if you take a look at the Microsoft CEO's presentation last April, there are quite a few customers that were mentioned. And I'm going to read off a bunch of names to you because it's publicly available. BMW, Shell, PwC, Cognizant, Accenture.
TEGUS CLIENT
Just in the ecosystem when people are, not customer. When people are talking to you and they're thinking about using X or thinking about using Y or what do you think about Z, what are the other names in and around AI that have a lot of heat right now?
COMMERCIALIZATION LEADER - AZURE OPENAI SERVICE (ARTIFICIAL INTELLIGENCE) AT MICROSOFT
Yes. I mean Stable Diffusion is sort of come and gone. I've been doing this type of thing for years, and you're always compared to your competitor. I don't get questions on how it compares with like Google, for instance. So it comes from somebody that doesn't know what they're talking about. Like when I was doing Azure Stack, it was all about AWS Outposts. When I was doing IoT, it was about like Google and Amazon. People don't ask me about the competitors on this thing because the competitors don't have anything. It's a really strange world we're in right now. The OpenAI has so much mind share right now. I've never seen something like this before.
TEGUS CLIENT
Yes, it's wild. It was really great connecting. Appreciate it. Good luck with everything.