Overview

On October 9, 2024, the Azure AI Platform team held an AMA on the Microsoft Tech Community. This live hour of Q&A provided members with the opportunity to ask questions and provide feedback to the product team. We hope you join us live next time!

Resources

- GPT-4o-realtime API with Audio announcement blog
- Copilot Voice product information
- RAG + Voice demo utilizing the GPT-4o Realtime API

Introduction

Welcome to the GPT-4o-realtime API with Audio AMA! View the list of introductions in this thread.

General Discussion

Q: I followed the links in the email to install the Android app. It does not match your claims (refuses to state what AI model version it's using but does say that it's NOT GPT-4). The app is also not really ready for release IMHO: it's truncating the end of spoken sentences (missing the last word) and locks up if you go from spoken to text and then back to spoken. The login process has a weird loop that's super-confusing, and voice-input has been disabled for login for no sensible reason.

As a heavy ChatGPT user - the AI seems basically useless in comparison. It refuses to answer basic questions with excuses like "I can't help you with business advice", and does not understand how to follow instructions at all (e.g. tell it to stop asking you questions at the end of every response).

Your email says, "be part of the transformation!" - so I asked it how I can make my value-added services available to users ... and it basically says I cannot.

What exactly does "be part of the transformation" mean? I don't want to add CoPilot to my service - I want my Service to be available to CoPilot users. Is that going to be possible? (link)

A: Allan and I are part of the AI Platform team working on the Azure OpenAI Service capabilities (Copilot is a same-company internal customer of the same capabilities now available to everyone) and thus we're not the best people to comment on Copilot app specifics. I will beg a bit of patience with any rough edges around anything related to the technology, though -- we just simultaneously released this gpt-4o-realtime-preview feature set (the beta /realtime API endpoint) with OpenAI last week and I can vouch for things continuing to change *very* quickly; I'm still just astounded that so many cool experiences were made possible so quickly with underpinnings changing so rapidly! As far as a "transformation" goes: flashy wording aside (hey, it got some attention), there really *is* some amazing potential that this kind of voice-in, voice-out interaction paradigm opens up. When voice assistants first became popularized, many people were understandably disappointed with how "on rails" and ultimately limited some of the capabilities necessarily ended up being, given the constraints in the technology: handling truly natural speech (including interruptions, so-called disfluencies like "ums" and "ahs", speaker variations, etc.) was hard, interactions still felt very "walkie-talkie-like" in how transactional and turn-based things were, you still felt like you were choosing from a short menu of things the assistant was good at, and so on. This new/realtime capability set built around gpt-4o-realtime-audio breaks through a lot of those barriers--I've had several people I've demoed to remark that they didn't believe it wasn't actually a pre-recorded person answering the questions and/or that it wasn't a person replying to them, even when they were trying it live, given how natural the experience felt. Aside from white-lie flattery, nobody ever *really* said that about voice assistants before. Now, that isn't to say that everything's absolutely perfect yet -- this is a beta/preview feature area, after all! -- but even trying it out on the playground or demo apps (or seeing it in action inside of Copilot, OpenAI's Advanced Voice Mode, etc.) really gives a sense for it isn't an unreasonable exaggeration to call this all "transformative."

Q: *Security - How safe is the LLM that sits behind all this awesome technology. For example, if I have a proprietary calculation to calculate something for a business and I want to use it in a query with Chat GPT or Copilot. Will that mean that this will now be available to others who might need the same calculation? Then, will Microsoft be looking at a "reduced" cost version of Copilot that would work only in Teams or only in Power Point? Finally, has there been any improvements on Copilot in Excel to interrogate basic standard data without tables?* (link)

A: Microsoft put security above all else. Security comes first when designing any product or service, including LLM. Security protections are enabled and enforces by default, require no extra effort and are not optional. When you use Azure OpenAI Service, your prompts (inputs) and completions (outputs), your embeddings, and your training data, are all your own data, meaning your data are NOT available to others for any uses. As for pricing, please get the latest from your sales representatives or from our pricing page. As for your last question, we will relay it to our CoPilot team to follow up or stay tuned with the Copilot product updates.

Q: *This looks amazing! Waiting to try it.* (link)

A: The model is available today in AI Studio, whether you want to try it in the Playground or integrate it with the available API. Check out the link below and let us know about your experience!

GPT-4o-Realtime-Preview

Q: *Hi there! I would like to respectfully inquire about Microsoft's plans for supporting the training community in keeping pace with the rapid updates, especially for Copilot. As a trainer, I find it challenging to stay current with existing tools before new ones are released, making it even harder to document changes and effectively transfer knowledge within our industry. Are there any plans to establish industry-specific or license-specific communities focused on Copilot, where we can share experiences, use cases, and learn together? Could Microsoft please share their strategies for transitioning from traditional training methods to fostering continuous learning and effective knowledge transfer, ensuring full adoption of these innovative tools?*

A: We understand the challenges of staying updated with AI technology. To help you stay updated, Microsoft offers a variety of programs and tooling. **MS Learn** is the go-to place for you to get interactive training and learning for a wide range of Microsoft technologies including Azure OpenAI and Copilot.

You can also get training and certifications from **Microsoft Certified programs** including Azure Fundamentals and Power Platform Fundamentals. The **Tech Community** is a good place for you to connect with experts and stay updated for the latest product announcements. As for specific topic trainings, stay tuned with our community event calendars.

Q: *Will this be hard for people to adopt to this new and exciting change? What adoption advice would you give to Organizational Change Managers who come up with the strategy to drive adoption?* (link)

A: Thanks for the question. I've found Jared Spataro's "AI at Work" videos to be useful inspiration for the organizational change guidance you're seeking.

1. employees struggle with the confidence to believe they can successfully master this new AI technology,
2. a lack of knowledge around which tasks are best suited to AI, and
3. the skills needed to maximize AI's potential.

My takeaways from the video were that leadership encouraging for their organization to try AI and in-person trainings were most impactful. What do you think?

Here was the latest video I'm referring to: https://www.linkedin.com/posts/jaredspa_ai-activity-7249443174991998976-dYGM?utm_source=share&utm_medium=member_desktop

A1: Great question! I worked on voice assistants all the way back to the early Windows Phone days (*before* Cortana!) and it's surprising how much is "old" at the same time it's "new" -- if you're familiar with voice assistant paradigms and ever struggled with the technology just "not being quite there yet" for a lot of useful scenarios, then much of this will feel very familiar. Building a rich, voice-in, voice-out product experience can be very complex to get *perfect,* but can also be surprisingly quick to get to a min-viable-product, "already delivers quite a bit of value" state. I'd highly recommend just playing around with either the playground or demo applications to get a feel for what kinds of things it makes possible; everyone I've shown off even the interactive console demos too seems to ask it to do different things, and everyone walks away with a different idea of what they'd like to do with it. But everyone's walking away wanting to make something!

Q: *Is there an SDK or set of APIs to have a GPT-4o audio agent join a Teams call? Could multiple agents join a Teams call? e.g.: a Project manager agent, a QA agent, and a few human developers all in the same Teams call? Is this possible today? If so, which SDKs/APIs would enable this use case?*

A: We don't yet have higher-level abstractions for Teams specifically, but between OpenAI and Microsoft we've started some client library coverage to expose the new capabilities of the gpt-4o-realtime-preview model and the /realtime API:

- The OpenAI .NET SDK (https://github.com/openai/openai-dotnet) (as well as the AOAI companion library, Azure.AI.OpenAI) already has early support for a client integrated into 2.1.0-beta.1
- Python has an early standalone client we're iterating on at aoai-realtime-audio-sdk/python at main · Azure-Samples/aoai-realtime-audio-sdk (github.com)
- JavaScript has an early standalone library: openai/openai-realtime-api-beta: Node.js + JavaScript reference client for the Realtime API (beta) (github.com) and we also one at aoai-realtime-audio-sdk/javascript at main · Azure-Samples/aoai-realtime-audio-sdk (github.com)

We've already seen developers prototype applications with multiple agents talking to people (and each other!) using the /realtime capabilities and the results are very cool. Very possible with the tools we have today!

Q: *Thank you for scheduling this session. We have been experimenting with some of the sample code provided by Microsoft, and it appears to be functioning well. However, we have observed instances where the model generates music-like sounds, although it is not actual music but has a tune. Additionally, there are occasions when the model changes its voice. Could you provide guidance on how we should approach grounding the outputs?* (link)

A: Oh, I know exactly what you mean; the model can get pretty "creative" sometimes. It was even more entertaining a few weeks ago; one of its favorite pastimes was to start -- no joke -- giggling in the middle of a response. Much of this is getting rapidly improved within the model itself and is driven by continual new deployments. From a consumption perspective, you can use system messages ("instructions" inside of "session.update" with the /realtime API) and few-shot examples (conversation items with example input/output) to help prime the model for better output, just like you would with e.g. chat completions. This applies to even mundane things like retaining the same tone or voice -- responses should (and will) do a better job of not "getting distracted" all on their own, but gentle reminders surprisingly do assist, too.

Q: Wow, a voice enabled AI Event using text-based chatting? From Copilot: Why did the text message go to school? To improve its grammar and become a better text! (link)

A: For what it's worth, I promise that my own typos will be 100% human-produced here!

Q: Are there any Visual Studio examples or tutorials that engage with these audio real time chat API's that we can use to get started with on our own projects? (link)

A: For some basic "getting started" resources, check out https://github.com/azure-samples/aoai-realtime-audio-sdk -- this has an interactive localhost web demo using a standalone TypeScript SDK library, an interactive console demo with tools using the official .NET SDK's latest beta, and some non-interactive, file-based demonstrations using a standalone Python library.

A1: We have prepared SDKs and samples to help builders get up and running as quickly as possible -> https://github.com/azure-samples/aoai-realtime-audio-sdk. This repository is actively monitored by our team, and we welcome any suggestions and contributions per these guidelines -> https://github.com/Azure-Samples/aoai-realtime-audio-sdk/blob/main/CONTRIBUTING.md. Our goal is to continuously improve the experience and make getting started with any new model as easy as possible!

Q: *How can no-code developers best utilize the Copilot Voice product and GPT-4 in their applications? Are there specific integrations with no-code platforms (like Make, Bubble, etc.) that you recommend?* (link)

A: I'm sure the Copilot and Copilot Studio teams are cooking up something for no-code and low-code developers to take advantage of this new modality.

Q: *For a non-programmer, what is the setup process like to get started with GPT-4 in Azure? Are there any beginner-friendly resources or templates available?* (link)

A: For a non-developer, we have some helpful documentation to get started: https://aka.ms/oai/docs. Without code you can create a resource by using the Azure Portal and launching the OpenAI Studio once a resource has been created. For now, we'd recommend creating the resource in either East US 2 or Sweden Central Azure regions. Check out this quickstart guide for how to interact with GPT-4o using the Realtime API within the Studio. For more code-first development, you can check out these pre-made samples: https://github.com/azure-samples/aoai-realtime-audio-sdk

A1: Whether you plan to write code or not, the process to get started with the new gpt-4o-realtime-preview model is fairly straightforward: (0) if you don't have one yet, create an Azure account via Azure Portal; (1) create an Azure OpenAI Service resource in one of the two preview regions (eastus2 or swedencentral); (2) using Azure AI Studio, create a gpt-4o-realtime-preview model deployment in your eastus2 or swedencentral; (3) use the "Real-time audio" playground (left navigation bar) to check out the new model with a live, browser-based voice-in/voice-out experience. From there, there are code samples -- including ones that just require setting environment variables and running -- at https://github.com/azure-samples/aoai-realtime-audio-sdk . We don't have much in the way of "build a new experience with no code whatsoever" yet given how new this all is, but we're continually looking for ways to make it easier to integrate this new /realtime feature set and other Azure OpenAI capabilities.

Q: Could you explain how the new multilingual features work in the real-time API? What steps are involved for a no-code developer to incorporate multiple languages seamlessly? (link)

A: Much like the text-based capabilities of LLMs, these models can natively interpret many languages other than English. As always, it is best to test GPT-4o's audio capabilities to ensure the model has the fidelity you need to meet your business requirements.

Q: What are some specific examples of how no-code developers can use GPT-4's real-time API for improving customer interaction experiences, such as through chatbots or voice assistants? (link)

A: This voice capability unlocks a new modality in interacting with applications where AI becomes the universal interface. I've found inspiration in the scenarios our customers we gave early access like Bosch and Lyrebird Health as well as the examples OpenAI demonstrated in their spring update.

Q: *How does Microsoft ensure data security and privacy when using GPT-4 in business applications, especially for industries like healthcare or finance? Are there compliance certifications built into the Azure deployment?* (link)

A: Yes! Azure OpenAI Service is an enterprise-grade platform hosting the latest models from OpenAI. Check out our data, privacy, and security documentation for more information.

Q: For someone just starting with this technology, what are the cost implications of using the GPT-4 API on Azure? Are there recommendations for managing costs while scaling up usage? (link)

A: Cost implications are use case dependent. Items that are likely to impact your costs include: Input/output audio ratio, average length of audio session, number of concurrent connections, deployment type, and throughput requirements. We recommend that developers start with small scale tests and development and evaluate costs before scaling to production.

Q: *Are there any upcoming features for GPT-4 on Azure that could further benefit no-code developers? What should we be looking out for over the next few months?* (link)

A: The "On your data" feature of Azure OpenAI is always a great way to get started quickly, then you can create a web app or a chatbot to share with your team or organization. Check it out!

A1: For getting started with no code, Azure AI Studio (https://ai.azure.com/) includes the new /realtime capabilities in its Playground experience (much like OpenAI's) and allows you to interact with the new gpt-4o-realtime-preview model in the browser without needing to write any lines of code -- it's also the best way to create and manage the model deployments on your Azure OpenAI Service resource. Low- and no-code development of /realtime-based experiences is an area we're actively looking into. The new API endpoint and its WebSocket-based capabilities are considerably more complex than the prior REST-based operations like /chat/completions, and making it as approachable and easy to integrate as possible (including with no code at all, where it makes sense) is a major priority for us when it comes to Developer Experience.

Q: *What are some common mistakes non-programmers might make when setting up or using GPT-4 with Azure? Any tips on how to avoid them?* (link)

A: Common mistakes I've witnessed firsthand typically include:

1. not adding enough of your default quota to your deployment
2. not spending enough time on refining the system message to specify instructions
3. when building retrieval augmented generation (RAG) applications, folks don't spend enough time on their chunking strategies and investigating relevancy of the documents in the vector store.

I'm sure there are plenty of others you'll uncover as you begin your journey. Best of luck!

Q: *Will there be an OotB hardware option similar to Amazon Echo Dot or Google Home voice assistant nodes? (this is our chance to have a second shot at having an awesome Cortana implementation). Or if not, will there be enough API and persistence to implement something like that?* (link)

A: As the Azure AI Platform team, it is our responsibility to ensure that state-of-the-art technology is available for any developer to integrate into their exciting products and applications. With the improvements the GPT-4o-realtime API provides in speech and audio capabilities, there are endless opportunities to integrate speech capabilities into any product.....whether old or new.

Q1: *Even text messages! Any chance this will be integrated into Dynamix or PABX solutions or are we just talking about the API today?*

A1: Today we are focused on the API since it is brand new. Teams all around the world (including Microsoft) are integrating the API into their applications, and we're looking forward to seeing more announcements about their applications in the future!

Q: *Will this new feature be included in PowerApps and Power Automate, or will this only be available in Azure OpenAI as an API for coding? Following on from this, is there an API syntax or* *reference site we can reference for some help on how to interface with the API? Are there any specific code or framework requirements to allow for this to be a part of the project we are working on? Finally what sort of costs are being planned for this (will it be subscription or per voice line or volume of data)? Okay my bad I just saw the SDK discussion earlier. Any chance we can get a view of some of those examples mentioned?* (link)

A: .NET sample is here

JavaScript sample is here

Python sample is here

As for pricing, we will be sharing more details very soon! Similar to other Azure OpenAI models, we will start with Pay-as-you-go pricing and introduce others like provisioned, batch, etc. over time. If there are other pricing models that would better help you scale your applications, we always welcome feedback!

That’s a wrap!

Thank you for joining this fun event! We hope you'll continue to ask questions and share your feedback.

See you next time!