IF IT TALKS LIKE A HUMAN... —

Disarmingly lifelike: ChatGPT-4o will laugh at your jokes and your dumb hat

It's amazing what a few well-placed chuckles and vocal tone shifts can do.

KYLE ORLAND - 5/14/2024, 1:55 AM

Enlarge / Oh you silly, silly human. Why are you so silly, you silly human?

FURTHER READING

Major ChatGPT-4o update allows audio-video talks with an "emotional" AI chatbot

At this point, anyone with even a passing interest in AI is very familiar with the process of typing out messages to a chatbot and getting back long streams of text in response. Today's announcement of ChatGPT-4o—which lets users converse with a chatbot using real-time audio and video—might seem like a mere lateral evolution of that basic interaction model.

After looking through over a dozen video demos OpenAI posted alongside today's announcement, though, I think we're on the verge of something more like a sea change in how we think of and work with large language models. While we don't yet have access to ChatGPT-4o's audio-visual features ourselves, the important non-verbal cues on display here—both from GPT-4o and from the users—make the chatbot instantly feel much more human. And I'm not sure the average user is fully ready for how they might feel about that.

DELIVERED STRAIGHT TO YOUR INBOX.

It thinks it's people

Take this video, where a newly expectant father looks to ChatGPT-4o for an opinion on a dad joke ("What do you call a giant pile of kittens? A meow-ntain!"). The old ChatGPT4 could easily type out the same responses of "Congrats on the upcoming addition to your family!" and "That's perfectly hilarious. Definitely a top-tier dad joke." But there's definitely much more impact to *hearing* GPT-4o give that same information in the video, complete with the gentle laughter and rising and falling vocal intonations of a lifelong friend.

Or look at this video, where GPT-4o finds itself reacting to images of an adorable white dog. The AI assistant immediately dips into that high-pitched, baby-talk-ish vocal register that will be instantly familiar to anyone who has encountered a cute pet for the first time. It's a convincing demonstration of what xkcd's Randall Munroe famously identified as the "You're a kitty!" effect, and it goes a long way to convincing you that GPT-4o, too, is just like people.

/ **Not quite the world's saddest birthday party, but probably close...**

OpenAI

Then there's a demo of a staged birthday party, where GPT-4o sings the "Happy Birthday" song with some deadpan dramatic pauses, self-conscious laughter, and even lightly altered lyrics before descending into some sort of silly raspberry-mouth-noise gibberish. Even if the prospect of asking an AI assistant to sing "Happy Birthday" to you is a little depressing, the specific presentation of that song here is imbued with an endearing gentleness that doesn't feel very mechanical.

As I watched through OpenAI's GPT-4o demos this afternoon, I found myself unconsciously breaking into a grin over and over as I encountered new, surprising examples of its vocal capabilities. Whether it's a stereotypical sportscaster voice or a sarcastic Aubrey Plaza impression, it's all incredibly disarming, especially for those of us used to LLM interactions being akin to text conversations.

FURTHER READING
Microsoft "lobotomized" AI-powered Bing Chat, and its fans aren't happy

If these demos are at all indicative of ChatGPT-4o's vocal capabilities, we're going to see a whole new level of parasocial relationship developing between this AI assistant and its users. For years now, text-based chatbots have been exploiting human "cognitive glitches" to get people to believe they're sentient. Add in
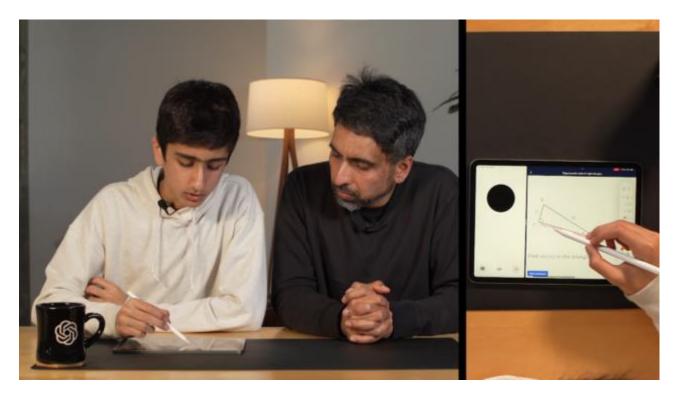
the emotional component of GPT-4o's accurate vocal tone shifts and wide swathes of the user base are liable to convince themselves that there's actually a ghost in the machine.

See me, feel me, touch me, heal me

Beyond GPT-4o's new non-verbal emotional register, the model's speed of response also seems set to change the way we interact with chatbots. Reducing that response time gap from ChatGPT4's two to three seconds down to GPT-4o's claimed 320 milliseconds might not seem like much, but it's a difference that adds up over time. You can see that difference in the real-time translation example, where the two conversants are able to carry on much more naturally because they don't have to wait awkwardly between a sentence finishing and its translation beginning.

Enlarge / ChatGPT-4o can actually see which part of your math homework is giving you trouble.

OpenAI

This lack of pausing seems especially important when a user interrupts GPT-4o in the middle of what can sometimes be long, rambling answers. This comes into focus during a counting demonstration where the user is constantly interrupting to adjust the counting speed faster or slower (and I swear GPT-4o sounds downright annoyed when it says "Okaaaay" after the fourth such interruption).

Afmma.ir

Then there are the new non-verbal instructions that users can give thanks to GPT-4o's video interaction mode. In a demo made in collaboration with vision-assistance app Be My Eyes, a blind user gets near-instant descriptions of their surroundings, ranging from the actions of a group of ducks in the water to the approach of a nearby taxicab. Video seems poised to give users a new way to highlight information for GPT-4o as well, such as in a Khan Academy demo where a student asks questions about specific parts of a geometry problem by marking them on an iPad with an Apple Pencil (The chatbot also follows instructions well by giving helpful hints without providing the answer directly.)

Stop laughing at me!

Before we get too far ahead of ourselves, it's worth pointing out again that we haven't had hands-on time with GPT-4o, and that carefully controlled AI demos have a history of being at least somewhat misleading. Even in these highly controlled demos, there are some glaring gaps in capability; GPT-4o's original lullaby about potatoes and self-harmonizing singing sample are both relatively atonal messes, in our opinion. A sample conversation between two GPT-4o bots also quickly descends into inane observations about room lighting before the user forces it back into focus.

Enlarge / Two OpenAI employees laugh as ChatGPT-4o hilariously owns up to its own mistake.

OpenAI

Afmma.ir

In fact, the bland inanity of some of GPT-4o's responses seems to be a weakness that a convincing vocal register can't always overcome. When a user tells the AI that they are preparing for an announcement, it replies, "That's exciting! Announcements are always a big deal," like the world's most brain-dead PR consultant. And during a Zoom meeting where users discuss the relative merits of dogs versus cats, GPT-4o chimes in with the useless, mealy mouthed response that "I can see the appeal of both" (before impressively summarizing the views of three different human speakers by name).

Beyond blandness, we also worry about how an LLM's tendency toward hallucinating incorrect information will mix with this new interaction model. It's one thing to see a confidently wrong answer in a text chat, it's another to hear a chatbot give a small chuckle as it tries to gently tell you that 2+2 = 5 or some other such obviously wrong statement.

But even some of GPT-4o's snafus can seem hilariously human in these demos. After the AI

assistant unexpectedly breaks into French during a rendition of "Take Me Out to the Ball Game," it owns up to its mistake with a "Sorry guys, I got carried away and started talking in French. Guilty as charged!" The way that GPT-4o can instantly realize its specific error immediately after the fact is baffling and hilarious in equal measure.

Enlarge / **Rocky here reacted with good humor when ChatGPT-4o laughed at his dumb hat. Will the general public?**

OpenAI

When GPT-4o is working as designed, though, it can also seem uncannily human in these demos. Take the demo of the chatbot helping prepare a user named Rocky for an interview with OpenAI. It starts off chuckling through a self-deprecating joke that OpenAI "sounds vaguely familiar..." then immediately switches to a supportive tone to congratulate Rocky's exciting opportunity. When Rocky asks about looking presentable, you can hear GPT-4o instantly shift to its most gentle, diplomatic tone when it says, "You definitely have the 'I've been coding all night' look down, which could work in your favor..."

But I really started to marvel at GPT-4o's ersatz humanity when Rocky put on a goofy-looking hat, and the chatbot literally laughed in his face, calling the topper "quite a statement piece." When GPT-4o tells Rocky, "I mean, you'll definitely stand out, though maybe not in the way you're hoping for an interview," I couldn't tell if the AI assistant was giving the devastating truth in a way only a true friend can or simply acting as a new, digital version of a Middle School Mean Girl.

I'm not sure how everyone will react to an AI chatbot that is willing to laugh at a stupid hat, but thanks to OpenAI, we're about to find out.

مقالات علمی و پژوهشی

OPENAI DROPS GPT-4O: AN OMNIMODEL WITH INTEGRATED VISION AND AUDIO ABILITIES, AND IT'S FREE VIA CHATGPT

انجمن شرکت های فنی، مهندسی و مشاوره ای
خراسان رضوی