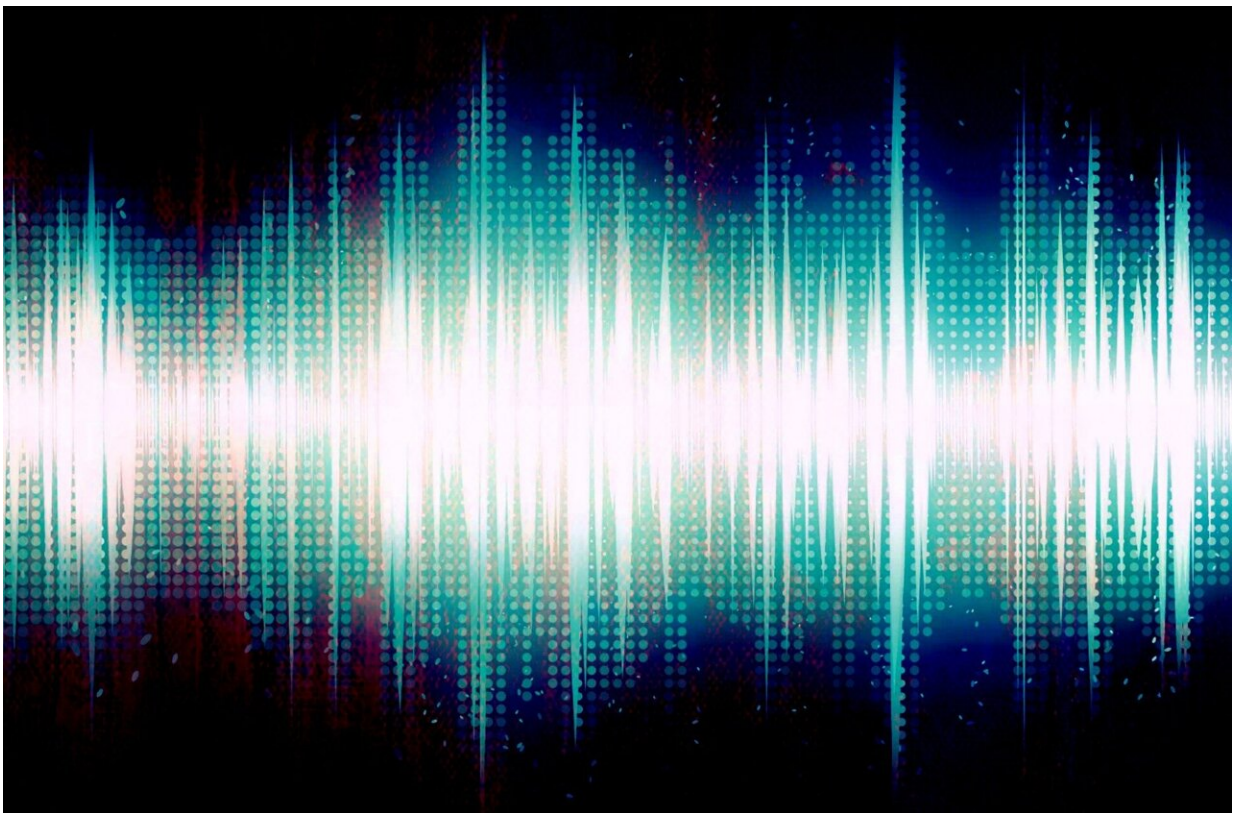# Scarlett Johansson's complaint to OpenAI is a new benchmark in the development of machine intelligence

May 23 2024, by David Reid

Credit: Pixabay/CC0 Public Domain

Over 2,000 years ago, the ancient Greek philosopher Aristotle came up with a way to build arguments. [He called this "rhetoric"](#) and described how logic in the text of an argument or speech, the needs and understanding of the audience, and the authority of the speaker could be used as strategies to persuade others.

Rather than just relying on logic in the argument or trust in the speaker, politicians and actors have long recognized that there is nothing as effective as using emotion to win the hearts and, consequently, minds of an audience.

With the launch of [GTP-4o](#) last week, we may have just seen a machine ideally suited to this task. While most see this as a fantastic breakthrough, having the potential to [benefit very many people](#), some view it with more caution.

Despite having previously declined OpenAI's request to sample her voice, actress Scarlett Johansson said [she was "shocked" and "angered"](#) when she heard the new GTP-4o speak.

One of the five voices used by GTP-4o, called Sky, sounded uncannily like the actress in her role as the [AI Samantha in the 2013 film Her](#) —about a man who falls in love with a virtual assistant. Adding to the discussion, OpenAI founder and CEO Sam Altman appeared to play up the comparison between Sky and Samantha/Johansson, [tweeting "her"](#) on the launch day of GPT-4o.

OpenAI later posted on X that it was ["working on pausing the use of Sky"](#) and created a web page on May 19, explaining that a different actress had been used. The company also expanded on [how the voices were chosen](#).

The fact that the film Her was almost immediately referenced when GPT-4o was launched has helped raise awareness of the technology among the general public and, perhaps, made its capabilities seem less scary.

This is fortunate because rumors about partnering with Apple have ignited privacy fears, with [iOS18](#) coming out next month. Similarly, OpenAI has partnered with Microsoft with its new generation of AI powered Windows system called Copilot + PC.

Unlike other [large language models](#) (LLMs), GTP-4o (or omni) has been built from the ground up to understand not only text but also vision and sound in a unified way. This is true multi-modality going far beyond the capabilities of "traditional" LLMs.

It can recognize nuances in speech such as emotion, breathing, [ambient noise](#), birdsong and it can integrate this with what it sees.

It's a unified multi-modal model (meaning it can handle photos and text), is quick—responding at the same speed as normal human speech (at an average of 320 milliseconds)—and can be interrupted. The result is unnervingly natural, altering tone and emotional intensity appropriately. It can even sing. Some have even complained about how "flirty" GTP-4o is. No wonder some actors are worried.

It genuinely is a new way to interact with AI. It represents a subtle shift in our relationship with technology, providing a fundamentally new type of "natural" interface sometimes referred to as EAI, or empathetic AI.

The speed of this advance has unnerved many government organizations and police forces. It's still unclear how best to deal with this technology if it is weaponized by rogue states or criminals. With audio deepfakes on the rise, it is becoming increasingly difficult to detect what is, and is not,

real. Even friends of Johansson thought it was her.

In a year when elections are due to be held involving more than 4 billion potential voters, and when fraud based around targeted deepfake audio is on the rise, the dangers of weaponised AI should not be underestimated.

As Aristotle discovered, persuasive capability often isn't about what you say, but in the way you say it. We all suffer from [unconscious bias](), an interesting [report from the UK about accent bias]() highlights this. Some accents are more believable, authoritative, or even trustworthy than others. For this precise reason, people working in call centers are now using AI to "westernize" their voices. In GTP-4o's case how it says things may be just as important as what it says.

If the AI understands the audience's needs and is capable of logical reasoning, then perhaps the final piece that's needed is the manner in which the message is delivered—as Aristotle identified 2,000 years ago. Perhaps then we will have created an AI that has the potential to become a superhuman master of rhetoric and with persuasive powers beyond the ability of audiences to resist.

This article is republished from [The Conversation]() under a Creative Commons license. Read the [original article]().

Provided by The Conversation

https://techxplore.com/news/2024-05-scarlett-johansson-complaint-openai-benchmark.html