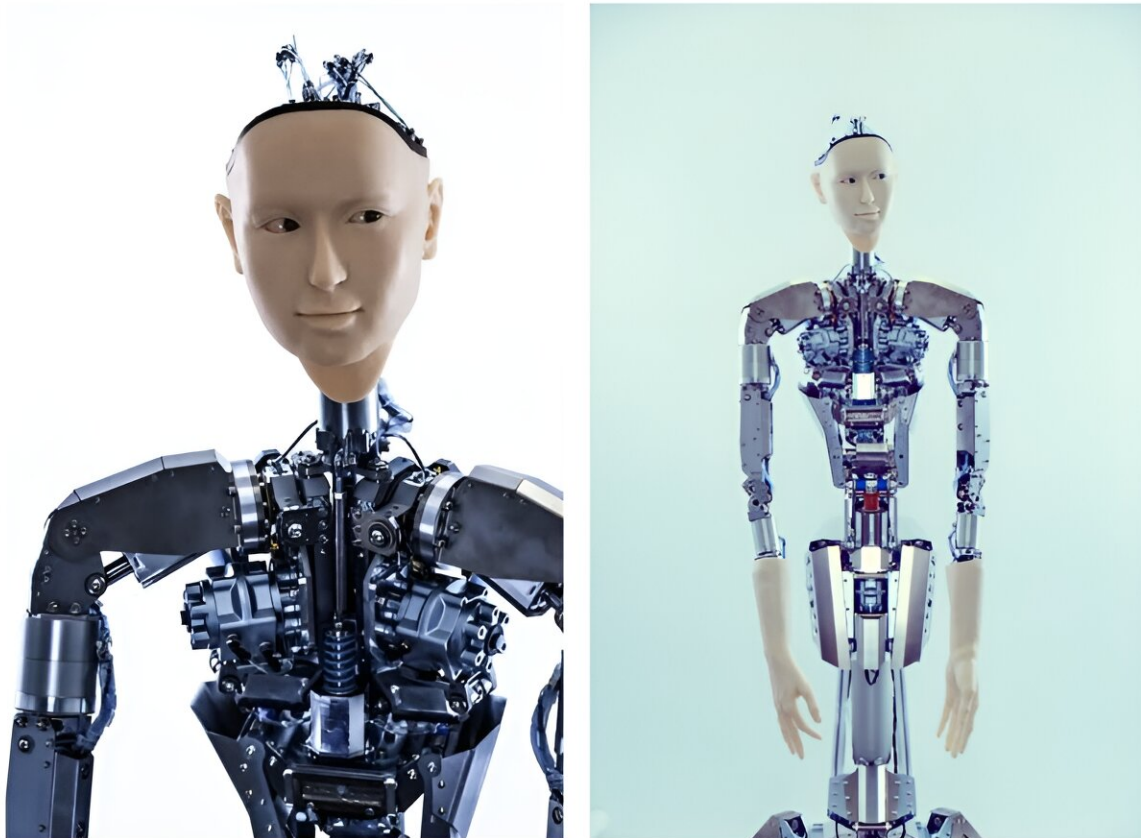# GPT-4 driven robot takes selfies, 'eats' popcorn

December 19 2023, by Peter Grad

Body of Alter3. The body has 43 axes that are controlled by air actuators. It is equipped with a camera inside each eye. The control system sends commands via a serial port to control the body. The refresh rate is 100–150 ms. Credit: *arXiv* (2023). DOI: 10.48550/arxiv.2312.06571

A team of researchers at the University of Tokyo has built a bridge between large language models and robots that promises more humanlike gestures while dispensing with traditional hardware-dependent controls.

Alter3 is the latest version of a humanoid robot first deployed in 2016. Researchers are now using GPT-4 to guide the robot through various simulations, such as taking a selfie, tossing a ball, eating popcorn, and playing air guitar.

Previously, such actions would have required specific coding for each activity, but incorporating GPT-4 introduces broad new capabilities to robots that learn from natural language instruction.

Robots powered by AI "have been primarily focused on facilitating basic communication between life and robots within a computer, utilizing LLMs to interpret and pretend life-like responses," the researchers said in a recent study.

"Direct control is [now] feasible by mapping the linguistic expressions of human actions onto the robot's body through program code," they said. They called the advance "a paradigm shift."

Alter3, which is capable of intricate upper body movement, including detailed facial expressions, has 43 axes simulating human musculoskeletal movement. It rests on a base but cannot walk (although it can mimic walking).

The task of coding the coordination of so many joints was a massive task involving highly repetitive motions.

"Thanks to LLM, we are now free from the iterative labor," the authors said.

Now, they can simply provide verbal instructions describing the desired movements and deliver a prompt instructing the LLM to create Python code that runs the Android engine.

Alter3 retains activities in memory, and researchers can refine and adjust its actions, leading to faster, smoother, and more accurate movements over time.

The authors provide an example of the natural language instructions given to Alter3 for taking a selfie:

Create a big, joyful smile and widen your eyes to show excitement.

Swiftly turn the upper body slightly to the left, adopting a dynamic posture.

Raise the [right hand](right%20hand) high, simulating a phone.

Flex the right elbow, bringing the phone closer to the face.

Tilt the head slightly to the right, giving a playful vibe.

Utilizing LLMs in robotics research "redefines the boundaries of human-robot collaboration, paving the way for more intelligent, adaptable, and personable robotic entities," the researchers said.

They injected a little humor into Alter3's activities. In one scenario, the robot pretends to consume a bag of popcorn only to learn it belongs to the person sitting next to it. Exaggerated facial expressions and arm gestures convey surprise and embarrassment.

The camera-equipped Alter3 can "see" humans. Researchers found that Alter3 can refine its behavior by observing human responses. They

compared such learning to neonatal imitation, which child behaviorists observe in newborns.

The "zero-shot" learning capacity of GPT-4 connected robots "holds the potential to redefine the boundaries of human-robot collaboration, paving the way for more intelligent, adaptable, and personable robotic entities," the researchers said.

The paper, "From Text to Motion: Grounding GPT-4 in a Humanoid Robot 'Alter3'," written by Takahide Yoshida, Atsushi Masumori and Takashi Ikegami, is available to the preprint server *arXiv*.

More information: Takahide Yoshida et al, From Text to Motion: Grounding GPT-4 in a Humanoid Robot "Alter3", *arXiv* (2023). DOI: 10.48550/arxiv.2312.06571

Project page: tnoinkwms.github.io/ALTER-LLM/