An embodied conversational agent that merges large language models and domain-specific assistance

September 13 2023, by Ingrid Fadelli

The FurChat system. Credit: The National Robotarium, where Oliver Lemon is academic co-lead for Heriot-Watt University.

Large language models (LLMs) are advanced deep learning techniques

that can interact with humans in real-time and respond to prompts about a wide range of topics. These models have gained much popularity after the release of ChatGPT, a model created by OpenAI that surprised many users for its ability to generate human-like answers to their questions.

While LLMs are becoming increasingly widespread, most of them are generic, rather than fine-tuned to provide answers about specific topics. Chatbots and robots introduced in some airports, malls and public spaces, on the other hand, are often based on other types of natural language processing (NLP) models.

Researchers at Heriot-Watt University and Alana AI recently created FurChat, a new embodied conversational agent based on LLMs designed to offer information in specific settings. This agent, introduced in a paper pre-published on *arXiv*, can have engaging spoken conversations with users via the Furhat robot, a humanoid robotic bust.

"We wanted to investigate several aspects of embodied AI for natural interaction with humans," Oliver Lemon, one of the researchers who carried out the study, told Tech Xplore. "In particular, we were interested in combining the sort of general 'open domain' conversation that you can have with LLMs like ChatGPT with more useful and specific information sources, in this case, for example, information about a building and organization (i.e., the UK National Robotarium). We have also built a similar system for information about a hospital (the Broca hospital in Paris for the SPRING project), using an ARI robot and in French."

The key objective of the team's recent work was to apply LLMs context-specific conversations, In addition, Lemon and his colleagues hoped to test the ability of these models to generate appropriate facial expressions aligned with what a robot or avatar is communicating or responding to at a given time.

"FurChat combines a large language model (LLM) such as ChatGPT or one of the many open-source alternatives (e.g., LLAMA) with an animated speech-enabled robot," Lemon said. "It is the first system that we know of which combines LLMs for both general conversation and specific information sources (e.g., documents about an organization) with automatic expressive robot animations."

The responses given by the team's embodied conversational agent and its facial expressions are generated by the GPT 3.5 model. These are then conveyed in spoken terms and physically by the Furhat robot.

The FurChat system. Credit: The National Robotarium

User interacting with the FurChat system. Credit: Cherakara et al.

To evaluate FurChat's performance, the researchers carried out a test with human users, asking them to share their feedback after they had interacted with the agent. They specifically installed the robot at the UK National Robotarium in Scotland, where it interacted with visitors and offered them information about the facility, its research endeavors, upcoming events, and more.

"We are exploring how to use and further develop the recent AI advances in LLMs to create more useful, useable, and compelling systems for collaboration between humans, robots, and AI systems in general," Lemon explained. "Such systems need to be factually accurate, for example, explaining how the information they present is sourced in specific documents or images.

"We are working on these features to ensure more trustworthy and explainable AI and robot systems. At the same time, we are working on systems which combine vision and language for embodied agents which can work together with humans. This will have increasing importance in the coming years as more systems for human-AI collaboration are developed."
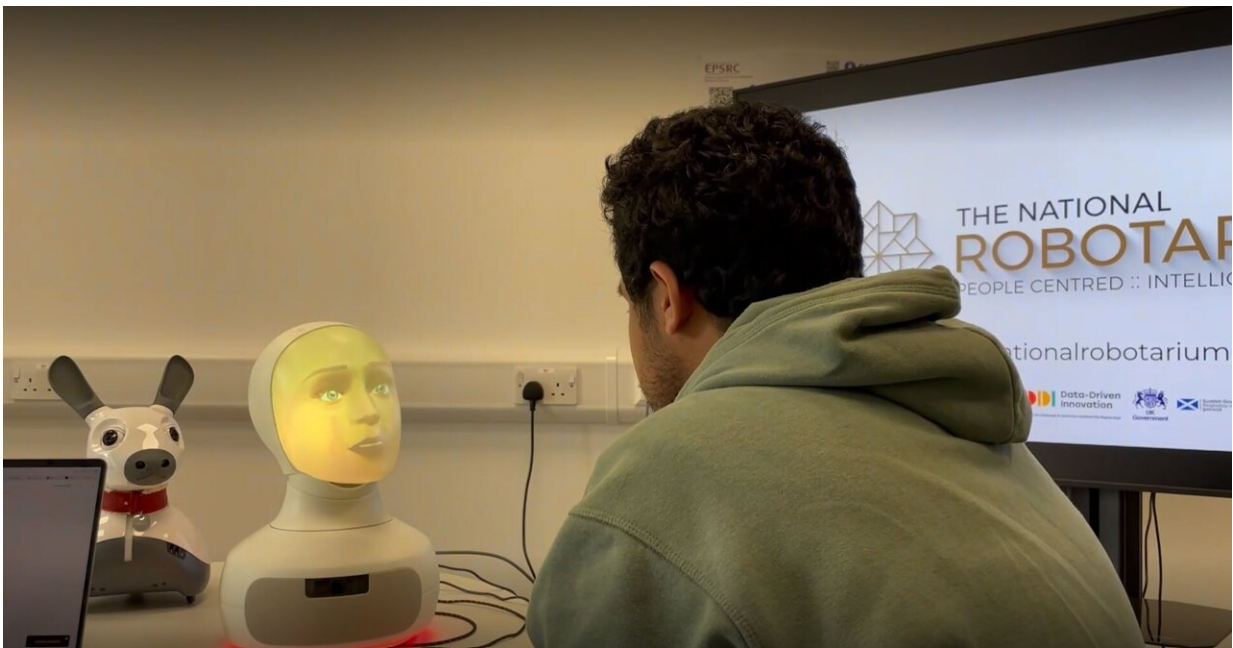
User interacting with the FurChat system. Credit: Cherakara et al.

In the team's initial real-world experiment, the FurChat system appeared to be effective in communicating with users both smoothly and informatively. In the future, this study could encourage the introduction of similar LLM-based embodied AI agents in public spaces or at museums, festivals and other venues.

"We are now working on extending embodied conversational agents to

so-called 'multi-party' conversations, where the interaction involves several humans, for example when visiting a hospital with a relative," Lemon added. "Then we plan to extend their use to scenarios where teams of robots and humans collaborate to tackle real-world problems."

More information: Neeraj Cherakara et al, FurChat: An Embodied Conversational Agent using LLMs, Combining Open and Closed-Domain Dialogue with Facial Expressions, *arXiv* (2023). DOI: 10.48550/arxiv.2308.15214