# FurChat: An Embodied Conversational Agent using LLMs, Combining Open and Closed-Domain Dialogue with Facial Expressions

**Neeraj Cherakara, Finny Varghese, Sheena Shabana, Nivan Nelson**
**Abhiram Karukayil, Rohith Kulothungan, Mohammed Afil Farhan,**
**Birthe Nesset, Meriam Moujahid, Tanvi Dinkar, Verena Rieser, Oliver Lemon†**
Interaction Lab, Heriot-Watt University ; †Alana AI
{nc2025, fv2002, ss2022, nn2023, ak2120, rk2065, mf2034, bn25
mm470, t.dinkar, v.t.rieser, o.lemon}@hw.ac.uk

## Abstract

We demonstrate an embodied conversational agent that can function as a receptionist and generate a mixture of open and closed-domain dialogue along with facial expressions, by using a large language model (LLM) to develop an engaging conversation. We deployed the system onto a Furhat robot, which is highly expressive and capable of using both verbal and nonverbal cues during interaction. The system was designed specifically for the National Robotarium to interact with visitors through natural conversations, providing them with information about the facilities, research, news, upcoming events, etc. The system utilises the state-of-the-art GPT-3.5 model to generate such information along with domain-general conversations and facial expressions based on prompt engineering.

## 1 Introduction

The progress in robotics and artificial intelligence in recent decades has led to the emergence of robots being utilized beyond their conventional industrial applications. Robot receptionists are designed to interact with and assist visitors in various places like offices, hotels, etc. by providing information about the location, services, and facilities. The appropriate use of verbal and non-verbal cues is very important for the robot's interaction with humans (Mavridis, 2015). Most research in the field has been mainly focused on developing domain-specific conversation systems, with little exploration into open-domain dialogue for social robots.

Conventional agents are often rule-based, which means they rely on pre-written commands and keywords that are pre-programmed. This limits the
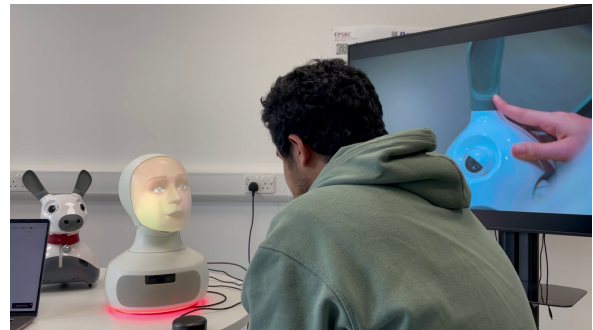


Figure 1: A user interacting with the FurChat System.

interaction with humans to little or no freedom of choice in answers (Tudor Car et al., 2020). The advancement of large language models (LLMs) in the past year has brought an exciting revolution in the field of natural language processing. With the development of models like GPT-3.5[1], we have seen unprecedented progress in tasks such as question-answering and text summarization (Brown et al., 2020). However, a question remains about how to successfully leverage the capabilities of LLMs to create systems that can go from closed domain to open, while also considering the embodiment of the system.

In this work, we present FurChat[2], an embodied conversational agent that utilises the latest advances in LLMs to create a more natural conversational experience. The system seamlessly combines open and closed-domain dialogues with emotive facial expressions, resulting in an engaging and personalised interaction for users. The system was initially designed and developed to serve as a recep-

---

[1] https://platform.openai.com/docs/models/gpt-3-5

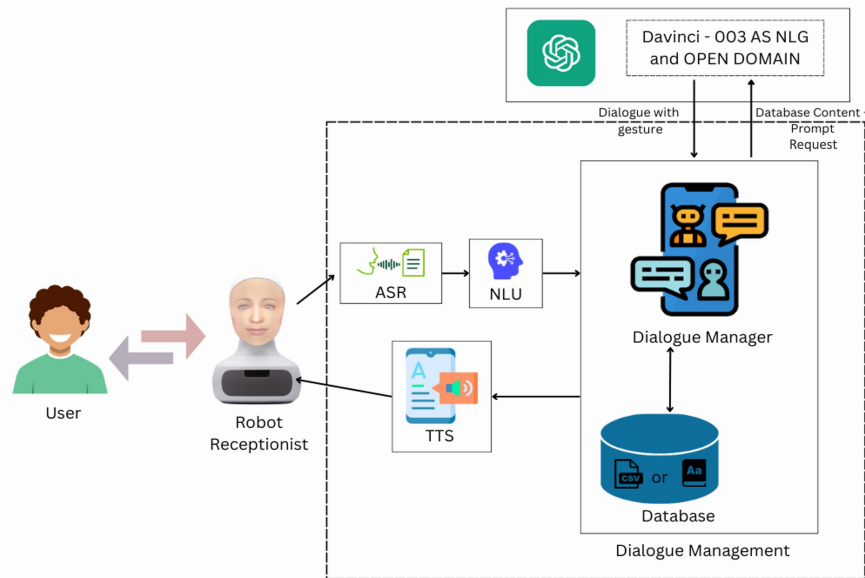[2] A demonstration video of the system is available here.

Figure 2: System Architecture of the current FurChat system.

tionist for the National Robotarium, in continuation of the multi-party interactive model developed by Moujahid et al. (2022b), and its deployment shows promise in other areas due to the LLMs versatile capabilities. As a result, the system is not limited to the designated receptionist role, but can also engage in open-domain conversations, thereby enhancing its potential as a multifunctional conversational agent. We demonstrate the proposed conversational system on a Furhat robot (Al Moubayed et al., 2013) which is developed by the Swedish firm Furhat Robotics[3]. With FurChat, we demonstrate the possibility of LLMs for creating a more natural and intuitive conversation with robots.

## 2  Furhat Robot

Furhat is a social robot created by Furhat Robotics. To interact with humans naturally and intuitively, the robot employs advanced conversational AI and expressive facial expressions. A three-dimensional mask that mimics a human face is projected with an animated face using a microprojector (Al Moubayed et al., 2013). A motorised platform supports the robot's neck and head, allowing the platform's head to spin and nod. To identify and react to human speech, it has a microphone array and speakers. Due to the human-like appearance of Furhat, it is prone to the uncanny valley effect (Ågren and Silvervarg, 2022).

## 3  System Architecture

As shown in Figure 2, the system architecture represents a conversational system that enables users to interact with a robot through spoken language. The system involves multiple components, including automatic speech recognition (ASR) for converting user speech to text, natural language understanding (NLU) for processing and interpreting the text, a dialogue manager (DM) for managing the interaction flow, and natural language generation (NLG) powered by GPT-3.5 for generating natural sounding responses(Ross et al., 2023). The generated text is then converted back to speech using text-to-speech (TTS) technology and played through the robot's speaker to complete the interaction loop. The system relies on a database to retrieve relevant data based on the user's intent.

### 3.1  Speech Recognition

The current system uses the Google Cloud Speech-to-Text[4] module for ASR. This module, which transcribes spoken words into text using machine learning algorithms, is integrated into the system by default through the Furhat SDK.

### 3.2  Dialogue Management

Dialogue Management consists of three sub-modules: NLU, DM and a database storage. The NLU component analyses the incoming text from

---

[3]https://furhatrobotics.com/

[4]https://cloud.google.com/speech-to-text

the ASR module and, through machine learning techniques, breaks it down into a structured set of definitions (Otter et al., 2021). The FurhatOS provides an NLU model to classify the text into intents based on a confidence score. We provide multiple custom intents for identifying closed-domain intents using Furhat's NLU capabilites.

The in-built dialogue manager in the Furhat SDK is responsible for maintaining the flow of conversation and managing the dialogue state based on the intents identified by the NLU component. This module is responsible for sending the appropriate prompt to the LLM, receiving a candidate response from the model, and subsequent processing of the response to add in desired facial gestures (see §3.4).

An open challenge faced by present-day LLMs is the *hallucination of nonfactual content*, which potentially undermines user trust and raises concerns of safety. While we cannot fully mitigate hallucinated content in the generated responses, in order to tone-down this effect, we create a custom database following suggestions from Kumar (2023). We do so by manually web-scraping the website of the National Robotarium[5]. The database consists of a dictionary of items with the intents as keys and scraped data as values. When an appropriate intent is triggered, the dialogue manager accesses the database to retrieve the scraped data, which is then sent with the prompt (further details in §3.3)) to elicit a response from the LLM.

## 3.3 Prompt engineering for NLG

The NLG module is responsible for generating a response based on the request from the dialogue manager. Prompt engineering is done to elicit an appropriate sounding response from the LLM, which generates natural dialogue that results in engaging conversations with humans. The current system uses `text-davinci-003`, which is one of the most powerful models in the GPT-3.5 series and it is priced at $0.0200 per 1000 tokens.

Producing relevant responses was achieved using the combined technique of few-shot learning and prompt engineering, which enabled us to try different variations in techniques and produce a variety of output by the LLM.

During prompt engineering, the personality of the robot and the context of the application are described, along with the past few dialogue histories and scraped data from the database in a particular response format. Moreover, the prompt engineering methodology involves using the LLM to generate an appropriate emoticon based on the conversation. In the context of emotional expression during an interaction, selecting an appropriate emoticon depends on understanding the underlying emotions being conveyed by the visitors and adhering to the display rules of the specific social situation. If the dialogue reflects joy or humor, a happy facial gesture might be fitting. On the other hand, if the conversation conveys empathy or sadness, a sad face could be more suitable. These emoticons are then integrated with the robot's facial gestures to generate facial expressions (see §3.4), thereby enabling a text-based LLM to integrate in the embodied Furhat robot. The explicit specification of the personality and context in the prompt aids in creating a natural conversation between the robot and the human that is coherent and relevant to the topic. The sample format of the prompt used is as follows:

*This is a conversation with a robot receptionist, <Robot Personality>, <Data from the Database>, <Dialogue history>, <Response Format along with sample emoticons>.*

## 3.4 Gesture Parsing

The Furhat SDK offers a range of built-in facial gestures that can be enhanced by custom facial gestures that meet specific needs. The latest GPT models have the ability to recognise emotions and sentiments from text, which is used in the system (Leung et al., 2023). Rather than simply recognising sentiments in the text, the model is tasked with generating appropriate emotions for the conversation from the text After receiving the response from the model, the matched conditional clause in the dialogue manager will trigger an expression from the pre-developed set of gestures, which will be triggered along with the generated speech.

## 3.5 Text-To-Speech Generation

For converting the text to speech, the Amazon Polly[6] service is used. This service is available within FurhatOS by default.

## 4 Example Dialogue

An example dialogue between a human and the robot is illustrated in Table 1.

---

*At the outset, the robot remains in an idle state. Once the user enters the vicinity of the robot, the conversation begins.*

R: [robot] Hello, I am the Receptionist here at the National Robotarium. Would you like to know about this facility? <Robot smiles>

U: [user] Yes, tell me about this facility.

R: [robot] Sure. <Robot smiles>

R: [robot] The National Robotarium is a state-of-the art research facility located on the Heriot-Watt campus in Edinburgh. . . .

U: [user] That's great. Is there any ongoing healthcare researches here

R: [robot] Yes. <Robot smiles> The National Robotarium is part of a multimillion-pound collaborative project called SPRING and is focused on developing socially assistive robots for healthcare settings. SPRING will focus on supporting elderly patients and improving psychological well-being and the relationship between patients and hospital professionals. <Robot nod>

U: [user] That's nice. Can you tell a joke?

R: [robot] <Robot smiles> Sure, Why did the robot go to the bar? <Robot winks> To get a screwdriver!

U: [user] Can you suggest me a movie about robots?

R: [robot] Sure thing! <Robot smiles> How about "I, Robot"? It's a classic that deals with the exploration of robotics and AI. <Robot nod>

Table 1: Sample Conversation between the user and the robot. For a full system description, please refer to §3.

## 5 Conclusions and Future Work

We demonstrate FurChat, an embodied conversational agent with open and closed domain dialogue generation and facial expressions generated through LLMs, on a social robot in a receptionist environment. The system is developed by integrating the state-of-the-art GPT-3.5 model on top of the Furhat SDK. The proposed system uses a one-to-one interaction method of communication with the visitors. We plan on extending the system to handle multi-party interaction (Moujahid et al., 2022a; Addlesee et al., 2023; Lemon, 2022; Gunson et al., 2022), which is an active research topic in developing receptionist robots. It is also crucial to address the issue of hallucination from the large language model and this problem can be mitigated by fine-tuning the language model and directly generating conversations from it without relying on any NLU components which we plan to implement in the future.

We plan to showcase the system on the Furhat robot during the SIGDIAL conference to all the attendees and show them the capabilities of using LLMs for dialogue and facial expression generation as described in this paper.

## References

Angus Addlesee, Weronika Sieinska, Nancie Gunson, Daniel Hernandez Garcia, Christian Dondrup, and Oliver Lemon. 2023. Data collection for multi-party task-based dialogue in social robotics. In *IWSDS 2023: International Workshop on Spoken Dialogue Systems Technology*.

Isabella Ågren and Annika Silvervarg. 2022. Exploring humanlikeness and the uncanny valley with furhat. In *Proceedings of the 22nd ACM International Conference on Intelligent Virtual Agents*, pages 1–3.

Samer Al Moubayed, Jonas Beskow, and Gabriel Skantze. 2013. The furhat social companion talking head. In *INTERSPEECH*, pages 747–749.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

Nancie Gunson, Daniel Hernández García, Weronika Sieińska, Christian Dondrup, and Oliver Lemon. 2022. Developing a social conversational robot for the hospital waiting room. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1352–1357. IEEE.

Krishna Kumar. 2023. Geotechnical Parrot Tales (GPT): Overcoming GPT hallucinations with prompt engineering for geotechnical applications. *arXiv preprint arXiv:2304.02138*.

Oliver Lemon. 2022. Conversational AI for multi-agent communication in Natural Language. *AI Communications*, 35(4):295–308.

John Kalung Leung, Igor Griva, William G Kennedy, Jason M Kinser, Sohyun Park, and Seo Young Lee. 2023. The application of affective measures in text-based emotion aware recommender systems. *arXiv preprint arXiv:2305.04796*.

Nikolaos Mavridis. 2015. A review of verbal and non-verbal human–robot interactive communication. *Robotics and Autonomous Systems*, 63:22–35.

Meriam Moujahid, Helen Hastie, and Oliver Lemon. 2022a. Multi-party interaction with a robot receptionist. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '22, page 927–931. IEEE Press.

Meriam Moujahid, Bruce Wilson, Helen Hastie, and Oliver Lemon. 2022b. Demonstration of a robot receptionist with multi-party situated interaction. In *Proceedings of the 2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 1202–1203. IEEE.

Daniel W. Otter, Julian R. Medina, and Jugal K. Kalita. 2021. A Survey of the Usages of Deep Learning for Natural Language Processing. *IEEE Transactions on Neural Networks and Learning Systems*, 32(2):604–624.

Steven I. Ross, Fernando Martinez, Stephanie Houde, Michael Muller, and Justin D. Weisz. 2023. The programmer's assistant: Conversational interaction with a large language model for software development. IUI '23, page 491–514, New York, NY, USA. Association for Computing Machinery.

Lorainne Tudor Car, Dhakshenya Ardhithy Dhinagaran, Bhone Myint Kyaw, Tobias Kowatsch, Shafiq Joty, Yin-Leng Theng, and Rifat Atun. 2020. Conversational agents in health care: Scoping review and conceptual analysis. *J Med Internet Res*, 22(8):e17158.