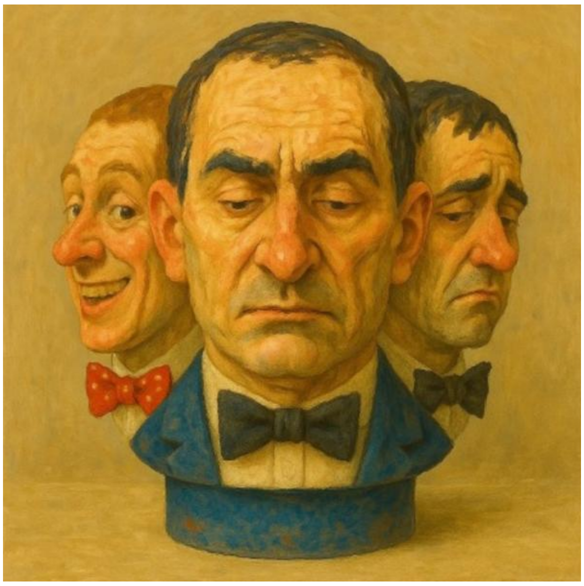
Software Engineering Thesis Capstone Project



Overview

A comedic robotic waiter with four distinct personalities that entertains as they serve customers. They argue among themselves while attempting to serve customers, engage in repartee, and so creating controlled chaos that showcases sophisticated AI conversation management. The robot has a spinning mechanism for the different faces/ personalities, delivers plastic food with a robotic arm and identify and recognise customers facial expressions.

The project aims to demonstrate the use of large language models with robotics and develops system integration skills.

Research Questions

1. How can a modern conversational LLM be used to enable embodied systems to create engaging, personality-driven experiences?
    - How does having multiple faces and gestures enhance rich interactions with LLMs?
    - How can LLMs be integrated with embodied systems?
2. Can an LLM really control a generic robot engine with all robot specific logic in the LLM?
    - How can it handle multiple multimodal input (text, voice, vision)?
    - How can the LLM operate in real time?
    - *(Stretch goal)* Can an LLM learn to control a robot through self-directed experimentation using feedback.
3. Can an LLM utilise vision to infer users body language?

A brief literature search has not found any closely related work.

Evaluation

Evaluation will be to determine whether the project successfully measures user engagement based on a behavioural interaction dataset (logs + timestamps + face state), failure/benefit patterns and structured user surveys. For the stretch goal to determine how effectively an LLM can write kinematic code based on self-directed experimentation.

Example Script: The Whisky Order

This script was developed by Claude and uses four personalities. Other personality choices could include characters from Inside Out, Winnie the Pooh, various sitcoms or other made-up personalities.

Customer 1 orders fine whisky. Customer 2 orders the same with coke.

```
[Head rotates to Jeeves the pompous Buttler]

Jeeves: "An excellent choice, sir. May I recommend the Macallan 18? Though I doubt you'd
       appreciate the subtleties..."

[Spins to Eddie, the Manic]

Eddie: "Ooh, fantastic! Coming right up!"

Customer 2: "Same, but with coke"

[Rotates to Jeeves]

Jeeves: "I'm terribly sorry, sir, but that was our last measure. Perhaps rum instead?"

[Eddie interrupts, spinning]

Eddie: "Wait, what? There's a whole bottle on the top shelf!"

[Spins to Mum, dominating and controlling]

Mum: "Don't you LIE to the customers! I SAW that bottle!"

[Marvin, paranoid, depressed.]

Marvin: "Of course there is. Jeeves protecting his precious whisky again. Not that it
       matters. Heat death of the universe is coming anyway..."

[Jeeves, flustered]

Jeeves: "That bottle is... reserved for discerning palates—"

[All four arguing simultaneously as head spins faster]

[Camera detects Customer 2's Upset/Frustrated expression]

[Head spins to Mum]

Mum: "Oh, DON'T give me that face, young man! What did you EXPECT ordering fine whisky with
       COKE? You should be GRATEFUL Jeeves even considered it!"

[Eddie tries to defuse]

Eddie: "Now, now, let's all just—"

[Mum continues]
```

```
Mum: "And ANOTHER THING—if you can't handle a little personality with your service, maybe
       you should order at a VENDING MACHINE!"

[Camera detects Customer 2's Angry expression]

[Marvin]

Marvin: "This is not going well. I predict violence."
```

Other scripts could be the Dirty Fork, ordering the wrong wine, substitutions, is the steak gluten free? and many others the LLM itself will devise and react to.

Components

- 3D printed head with four faces that rotates with a stepper motor
- Nodding mechanism using servo motor
- Simple robotic arm
- Dual microphones to determine which customer is speaking
- Stereo speakers
- Wide angle camera
- Frontier LLM via API
- OpenCV for real time image analysis
- Speech generation (local or cloud?)
- Raspberry pi
- Storage mechanism for holding customers' orders and chat history.

Implementation

The control flow starts with speech fragments that are recorded, separated by silence. They are sent to a frontier LLM that analyses the speech for text and sentiment. The LLM then sends JSON commands to the robot to rotate the head, perform other preset gestures, and provide text to speak. The robot uses a text to speech engine to talk to the customers. This will be controlled by a carefully written LLM system prompt, but the LLM will have a high degree of autonomy.

Dual microphones enable a simple comparison of volume to determine which customer is speaking. The camera can provide an image to the LLM with each interaction. Initial investigation has shown that an LLM (ChatGPT) can understand the body language of customers and the state of the table.

As a stretch goal, simple OpenCV analysis can determine the position of the orange-coloured robot hand. It will then be seen if the LLM itself can write the relatively simple reverse kinematics code to control the arm based on self-directed experimentation.

About me

I am Ella Berglas, a software engineering student going into my 5th and final year. I have demonstrated strong leadership skills, and my main interests lie in using AI to develop intelligent systems and applying robotics to solve real-world problems.

I believe this project is a nice balance between these interests and provides the opportunity to develop skills in robotics and system integration.

Relevant Courses/ Skills:

- DECO3801 where I developed a chatbot with a RAG system to support those within the youth justice system.
- COMP4703 (Natural Language Processing)
- COMP4702 (ML)
- CSSE3010 (Embedded Systems)
- COMP3710 (Pattern Recognition and Analysis)
- COMP3702 (AI)
- INFS1200, INFS2200, INFS3200 (databases)
- CSSE6400 (Software Architecture)
- Basic CAD skills
- Development of various first year/ high school level robotics projects.