Channel Islands Currents
2024 - 2025, Vol. 1, No. 1

CI CURRENTS
UNDERGRADUATE
RESEARCH JOURNAL

Robotic Control for Object Manipulation Using Artificial Intelligence

Andrew Ge,[1] Sheldon Peters, [1] & Alejandro Antonio [1]

Summer Undergraduate Research Fellow (SURF) Project, 2024

[1]*California State University Channel Islands, Camarillo, California*

Instructor's Introduction:
This Summer Undergraduate Research Fellowship (SURF) research project combined the reasoning power of Large Language Models (LLMs) and robotics to understand the capabilities of ChatGPT-4o in multimodal learning. The student researchers combined the Sawyer robotic arm, ChatGPT-4o, and Google Speech-To-Text services on a ROS Melodic system to develop a closed-loop system. A human may teach the robot a task and test the robot's understanding of that task. As the research continues, the students hope to conduct human studies to qualify the ChatGPT-4o's reasoning abilities in this context and assess its usability and adaptability. The research was started in Summer 2024 and is currently continuing through Fall 2024.

- Bahareh Abbasi, Assistant Professor in Mechatronics Engineering, October 2024.

The integration of Large Language Models (LLMs) with robotics offers new possibilities for developing autonomous interactive systems for manufacturing environments and has the potential to impact and transform how robots and humans collaborate on factory floors. This SURF project aims to develop a mechanism that enhances the process of teaching a robot to perform object manipulation tasks through interaction with a human expert. By leveraging the capabilities of LLMs to comprehend human conversations, we strive to ensure seamless and effective human-robot interaction. We have integrated the state-of-the-art LLM, ChatGPT-4o, and Google Speech-To-Text services with the Sawyer robotic arm running on ROS Melodic. We developed a closed-loop demonstration where the robot learns a task through human multimodal input, combining both language and visual cues. The LLMs acts as the robots' action planner and, according to the perceived multimodal human actions, decides how to execute the demonstrated task. In the future, the team plans to conduct human studies to evaluate the performance of the proposed system in real human-robot scenarios and assess its adaptability.

CI CURRENTS
UNDERGRADUATE
RESEARCH JOURNAL

Robotic Control for Object Manipulation Using Artificial Intelligence

Andrew Ge | Sheldon Peters | Alejandro Antonio • Dr. Abbasi | Dr. Isaacs

Introduction

Despite significant advancements in robotics, achieving full autonomy for robots in various domains, such as manufacturing, remains a significant challenge. This SURF project aims to develop a mechanism that enhances the process of teaching a robot to perform object manipulation tasks through interaction with a human expert. By leveraging the capabilities of Large Language Models (LLMs) to comprehend human conversations, we strive to ensure seamless and effective human-robot interaction.

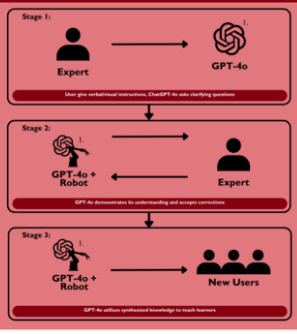
While most of the previous work focuses on using LLMs for task planning, here we take a new approach to use it for understanding of human expert's demonstration using multimodal (vision and speech) data and convert it to an actionable program.

The applications of this work can be utilized in various fields such as healthcare, emergency response, manufacturing, hospitality, and more.

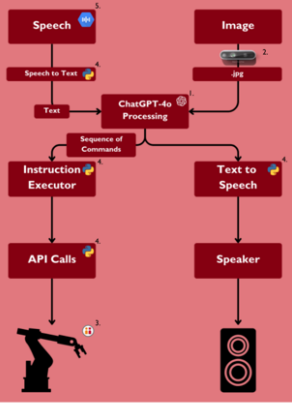
Proposed System

In this project, we utilize Sawyer, a robotic arm equipped with a state-of-the-art vision language model, ChatGPT-4o as its brain and planner. This LLM is capable of processing and understanding language and vision, while operating within Robot Operating System (ROS). With this architecture, our team hopes to train the LLM not only to perform a task, but also to train new users through a combination of expert instructions and LLM processing.

Human-Robot Interaction



System Components



Results

Here is a summary of the achievements of our team up to this point:

- Integrated ChatGPT-4o and Google Speech-To-Text services with the Sawyer robotic arm running on ROS Melodic
- Created a closed loop demonstration where a robot is taught a task using multimodal input in order to receive corrections and attempt to show understanding

SURF 2024 Github/Demo

https://tinyurl.com/summersurf2024

Future Objectives

In the future, the team hopes to continue our research to accomplish the following goals:

- Experiment to find optimal prompts for LLM understanding and comprehension
- Use more complex shapes/objects, (ex: foods, branded objects, faces, etc)
- Teach increasingly complex tasks such as using tools
- Optimize human user experience
- Recruit human participants for additional testing
- Use videos as part of multimodal input

Acknowledgements and Citations

SURF Staff | Maximilian Seligman
Cobot Team Representatives | Google Speech API
Rethink Robotics | OpenAI | Intel | ROS | Ubuntu

1: https://openai.com/, 2: https://www.intelrealsense.com/depth-camera-d435/,
3: https://www.rethinkrobotics.com/, 4. https://www.python.org/community/logos/
5. https://www.capterra.com/p/254094/Google-Cloud-Speech-to-Text/