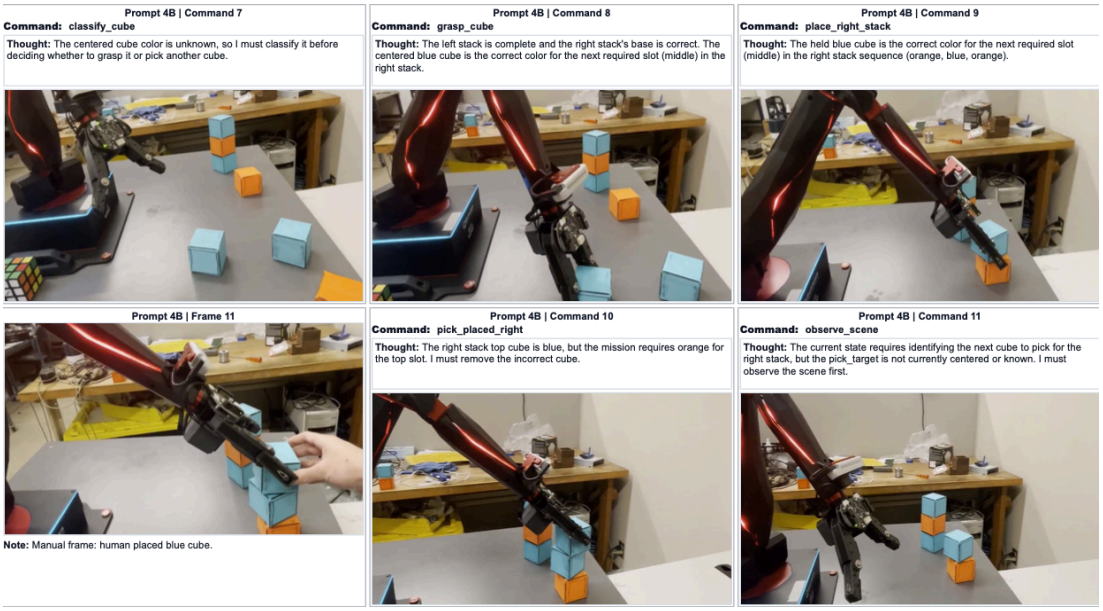
Prompt-Driven Tabletop Robotic Manipulation with a Structured LLM Interface and Visual Verification

A Master's Thesis Defense in Mechanical Engineering

Presented by: Tomas Franco

Committee Members: Dr. Mohammad Hasan (Chair)
Dr. Eric Espinoza-Wade
William Engel M.S.

When/Where: 10am, Jun 24 on Zoom - zoom link



Creating a tabletop robotic manipulation system that connects language goals to robot actions is a challenging problem. Large language models can interpret mission objectives and reason through actions, but when placed directly in control of hardware they can produce hallucinated commands that result in unsafe behavior without the proper safeguards.

This thesis presents a constrained LLM-guided manipulation system built around the Quanser Qarm, a four degree of freedom manipulator with a mounted Realsense RGB-D camera and a gripper with readable current. Every decision made by the LLM planner is routed through a gated command interface that restricts the model to legal actions for the current robot state, with visual and current-based verification closing the loop after each physical action. We tested the system across single-color stacking, stack correction, alternate color stack building and external robustness tasks across forty trials, achieving a 65\% autonomous and non-autonomous completion rate. The system is able to adapt to unexpected scene changes including replaced, removed, and added cubes during place sequences. Dominant failures sources were due to object orientation, depth estimation error, and state tracking mismatches that could cascade into poor planner decisions.

Through this work you may find useful lessons about integrating LLMs into physical systems, and how reliably an LLM can facilitate an autonomous system through language prompts.