

C.H.A.N.D.R.A Project

Phan Anh Duong and Cat Luong

College of Engineering and Applied Science, University of Cincinnati

Objectives

In movies and video games, interactions with robots are intuitive and easy. CHANDRA brings that vision closer to reality by letting you talk to robots like you'd talk to a friend. Instead of typing complex commands or using defined syntaxes, you can simply say something in natural language, like "Hey, walk ahead for a little bit" and the robot will understand.

Solution

Imagine talking to a robot like you'd talk to a helpful assistant. CHANDRA enables seamless communication between humans and robots by translating spoken language into precise robotic commands. When a user gives a command, the system first converts the spoken words into text through advanced speech recognition, ensuring accuracy even in noisy environments. The AI-driven model then analyzes the text, understanding the user's intent and breaking it down into specific instructions, such as the direction, speed, and purpose of movement. For example, a command like "Move forward and check what's around the corner" is translated into a series of robotic actions, directing the robot to move and use its sensors to gather information about the surroundings.

By functioning as a bridge between human language and robotic actions, CHANDRA simplifies the complexity of robotic control, enabling users to interact with robots without the need for technical knowledge. This makes robots more accessible and intuitive, allowing them to be used by a broader range of people for various tasks, from everyday household assistance to more complex applications in industries like healthcare and logistics.



Figure 1: Brutus, GhostRobotic's V60 Robot

Technology

- ① **Frontend:** A web-based interface that supports both voice and text input, built using *Streamlit*.
- ② **Backend:** A *FastAPI* server to receive data. Chosen due to universality and flexibility.
- ③ **AI Components:** Uses OpenAI's *Whisper* for advanced speech recognition. Uses compact language models and *smolagents* to elicit agentic AI command selection.
- ④ **Hardware:** CUDA-enabled *Nvidia Jetson Orin Nano* for accelerated model inference via GPU.
- ⑤ **Robot:** *Turtlebot3 Burger & GhostRobotic's V60*, running on *ROS2 Humble*. We simulate the robots on *Gazebo* and *Bullet Sim*.

Architecture

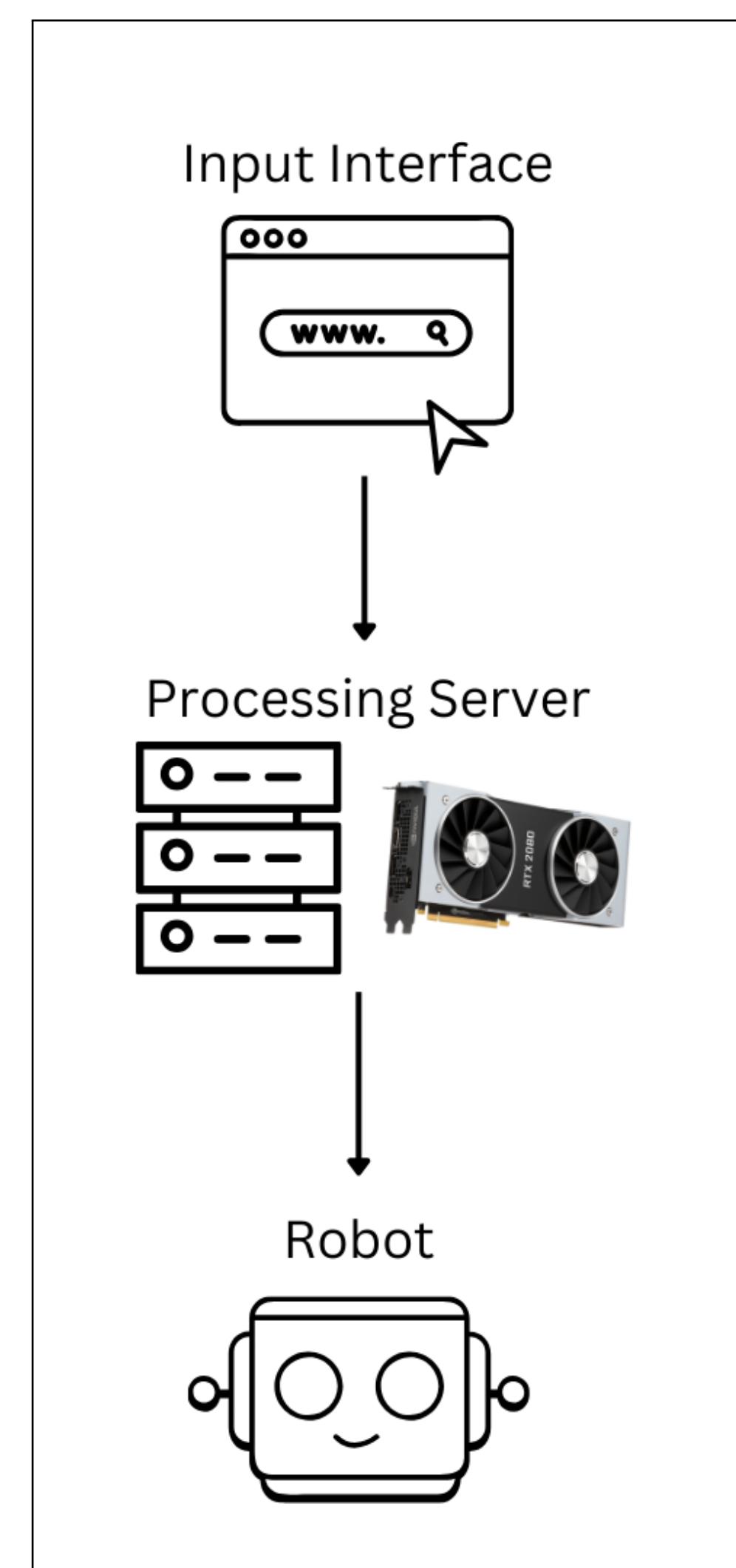


Figure 2: Architecture Diagram

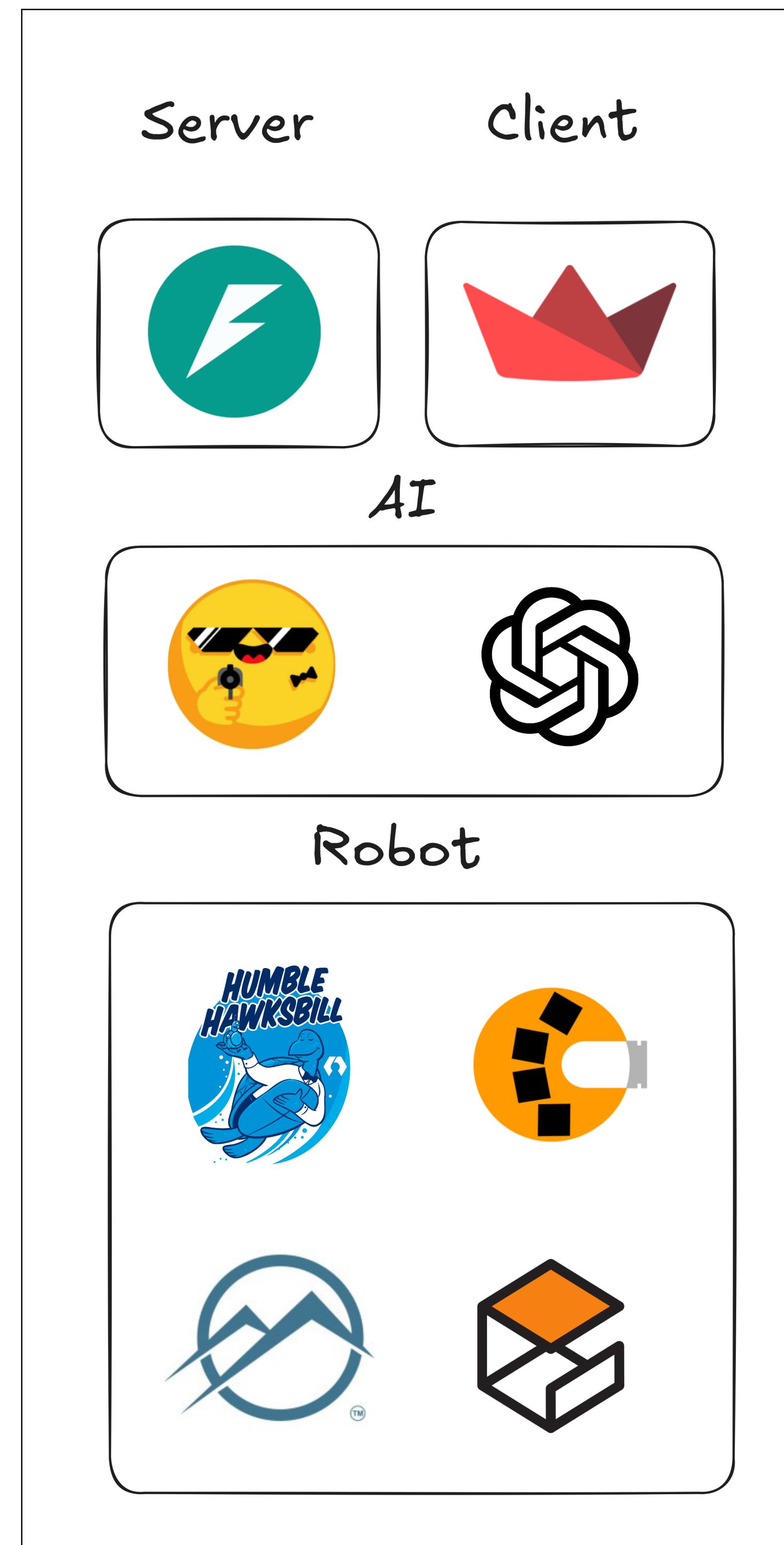


Figure 3: Tech Stack

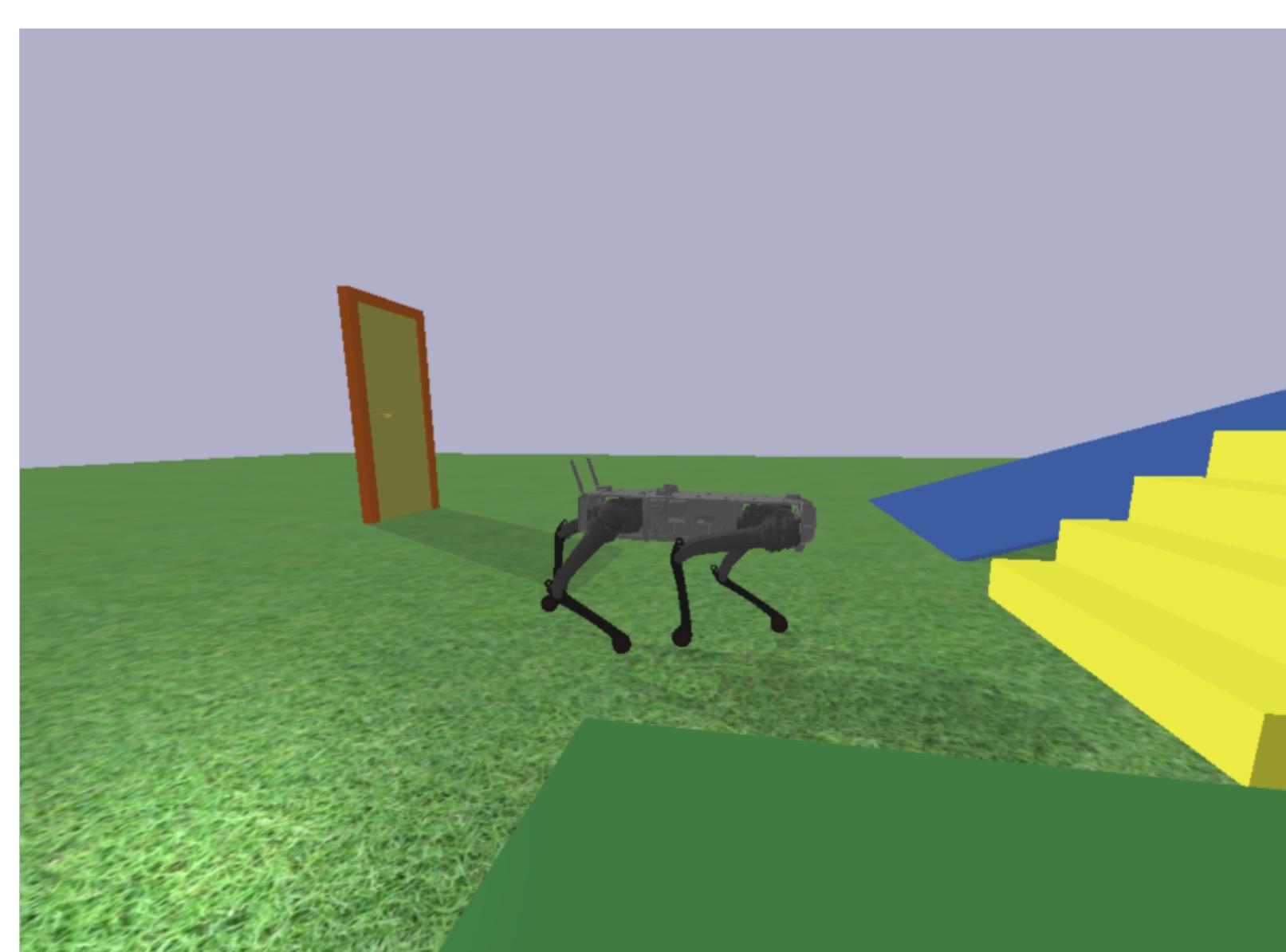


Figure 4: Brutus In Bullet Simulator

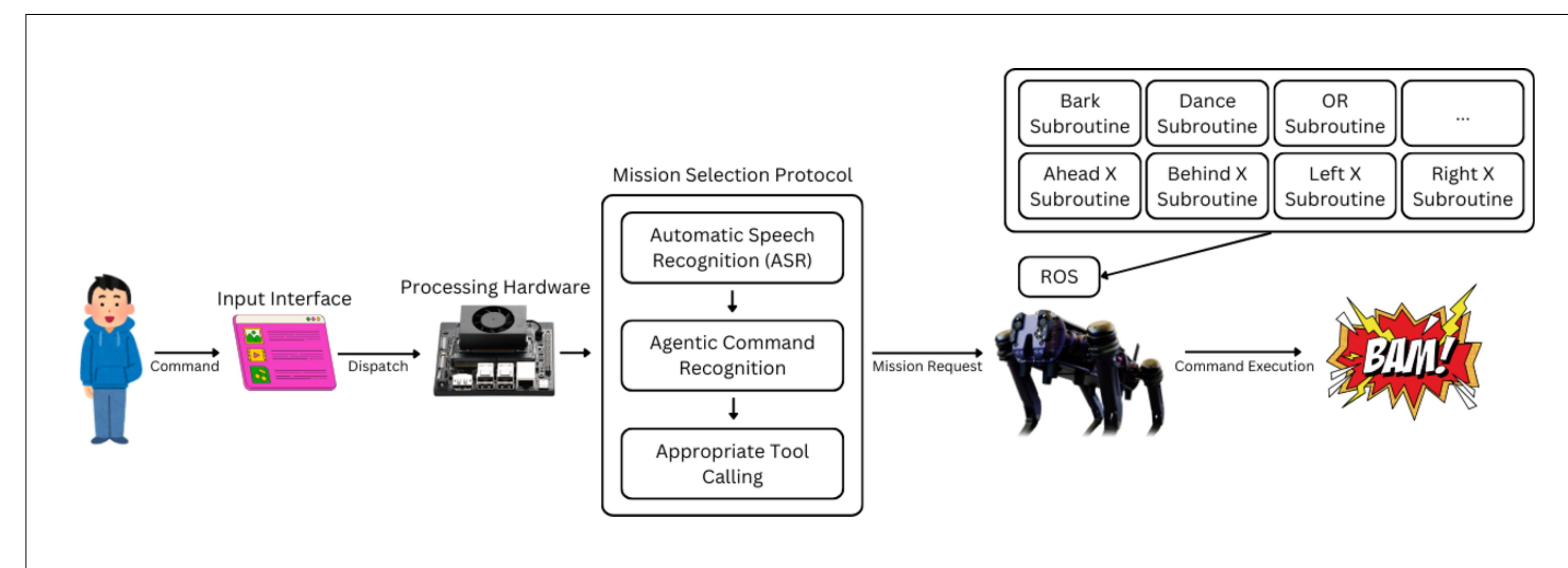


Figure 5: Execution Flow

Impacts

- **Accessibility:** Accessible to non-technical users, especially individuals with disabilities.
- **Automation:** Interpret spoken commands and translate them into robotic actions driving automation in diverse environments.
- **Scalability:** Adaptable across industries like manufacturing, agriculture, and healthcare.
- **Companionship:** Provide emotional support and companionship to the elderly.

Future Work

- **Contextual Memory:** Enabling the retention of past interactions, tasks, and environmental changes to enhance decision-making.
- **Multimodal Instruction:** Integrating visual and spoken inputs for real-time decision-making (e.g., "Move forward, avoid obstacles").
- **Personalization:** The robot will be trained to specialize in specific tasks, using advanced AI to adapt and make tailored decisions.
- **Human-Robot Collaboration:** Supporting collaborative tasks in shared spaces, such as assisting with physical tasks or organizing items.

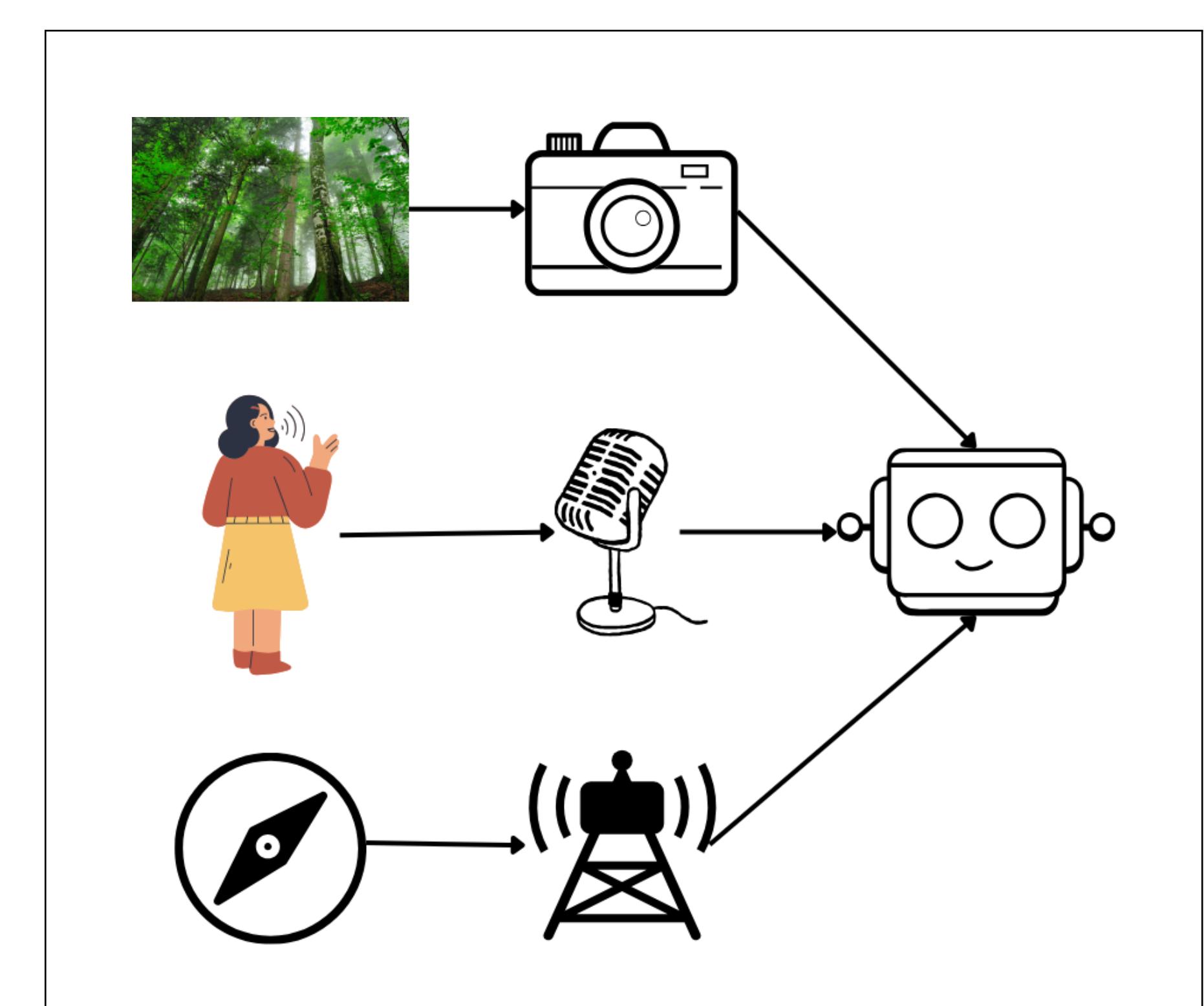


Figure 6: Ingesting Multimodal Data

Advisors

- Dr. Justin Zhan, CS Department Head
- Bryan Kowalczyk, AA Lab Director
- Arnav Komaragiri, Solutions Architect