Boston University
Electrical & Computer Engineering
EC463 Capstone Senior Design Project

Test Report

Visually Impaired AI Wearable

By

Team #32
Mimir

Team Members

Louis Jimenez-Hernandez louisjh@bu.edu
Heather Li hli9753@bu.edu
Dylan Ramdhan dylram01@bu.edu
Houjie Xiong xhj@bu.ed

Required Materials

Hardware:

- Raspberry Pi 5
- Raspberry Pi Camera
- Raspberry Pi Fan
- Bluetooth Connected Speaker
- 3D Printed Chassis
- LiDAR camera

Software:

- Python
- Server

Set Up

The equipment and setup are divided into 2 parts: the device with the Raspberry Pi with the camera and the laptop.

Raspberry Pi:

- Download the Raspberry Pi Imager tool from the official website: Raspberry Pi OS.
- Use the Imager to flash the Raspberry Pi OS (choose the desktop version if you need a GUI) onto a microSD card.
- Insert the microSD card into the Raspberry Pi and power it up.
- Connect to BU (Unencrypted) WiFi
- Attach the camera to the Raspberry Pi's camera port.
- Set up Bluetooth with Bluetooth-pairing device (external speakers)
- Download required libraries

Laptop:

- Install Docker and Pull Ollama Container
- Run Ollama Container
- Download the Llama3.2-Vision Models

Pre-testing Setup Procedure:

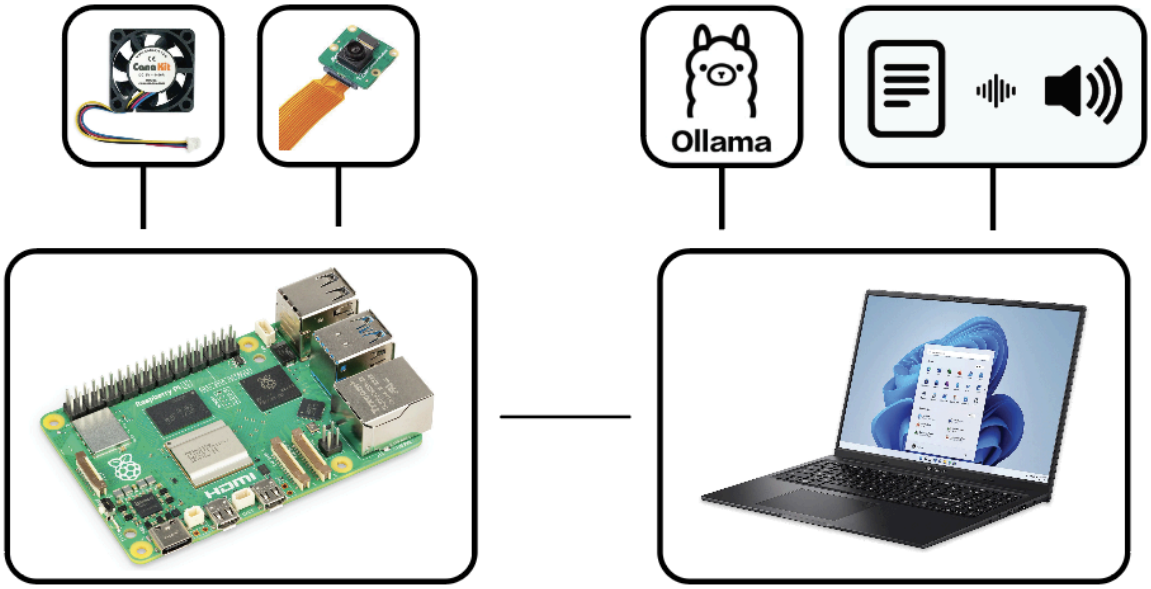
Raspberry Pi Side:

Set up device to be ready to take a photo
Create script to take the photo

Laptop Side:

Run Ollama llava

IMAGE:



Testing Procedure:

1. Run the Llama3.2-Vision Models
2. Set up device onto table
3. Hold item in front of camera and run the script to take an image
4. Upload it to Google Drive
5. Download image locally from Drive
6. Run Llama locally.
7. Use the pre-set prompt “Describe the the item, no need to be accurate in xxxxxxxxx” (where xxxxxxx is the path of the image file)
8. Get the result of the model

Measurable Criteria:

The criteria for successful running and output is as follows:

1. The Raspberry Pi should be able to capture an image, have it processed by OCR, and have the output converted to a wav file

2. The Pi should be able to capture an image and have it processed by LlaMa and have the output converted to a wav file
3. The LiDAR camera should accurately track the movement and position of one's hands.
4. The script will process the nearby words to the indWex fingertip and read it back to the user.

Score Sheet:

OCR

Object	**Image-OCR Time Output (min:second)**
Gluten-Free Pasta	0:55
Test Sheet	0:46
FujiFilm Manual	1:00
FujiFilm Box	0:59

LlaMA

Object	**LlaMa Output Time (min:second)**
Gluten-Free Pasta	1:31
Test Sheet	1:22
FujiFilm Manual	1:45
FujiFilm Box	1:39

LiDAR

Task	**Successful? (y/n)**
Hand Tracking	y
Word Detection	y
Text-to-Speech	y

Discussion:

For our OCR testing, we found that on average the time it takes to get the OCR processed is only about a minute. This is the fastest time which makes sense since we were able to just extract out the text and only have to analyze that which is considerably faster than doing an entire image.

For LlaMa, we had a huge improvement over our initial prototype testing. Originally, we took an average of about 4:30 just to do the LlaMa processing alone. Now the entire process from taking the picture to getting the output is about 1:30 which is a pretty big improvement. We anticipate seeing even more once we move to a cloud server as for now we are running it on a local desktop.

For LiDAR, the program is very accurate at being able to identify the word we are pointing at and outputting that within seconds. Based on this, we want to be able to extrapolate that to be able to do more complex functions like being able to identify the object that we have our finger pointed at.