6.1820/MAS.453: Mobile and Sensor Computing
aka IoT Systems

https://6mobile.github.io/

Lecture 13: Mobile & Augmented/Mixed Reality

Some slides adapted from Sruti Srinidhi (CMU)

Course Staff	Announcements
<u>Lecturers</u> Fadel Adib (fadel@mit.edu) Tara Boroushaki (tarab@mit.edu) <u>TAs</u> Waleed Akbar (wakbar@mit.edu) Jack Rademacher (jradema@mit.edu)	1- Project Proposals due April 1st 2- Lab 4 is out today; due April 4 3- PSet 2 due April 10

What are we learning today?

Mobile & Augmented/Mixed Reality

1- How can augmented reality headsets enable efficient and accurate RFID tag localization?

2- How does a human-in-the-loop design improve RFID sensing and localization?

3- What unexpected insights do user studies reveal?

4- How can multi-modal LLMs and XR be combined to create new applications?

5- What are the limitations of multi-modal LLMs in cognitive assistance for humans?

Battery-less 3 cent Radio Frequency IDentification (RFID) Tags


RFID tagged items
inside boxes


RFID tagged items
inside boxes

Applications

Applications

RF-AR

- First augmented reality headset that enables efficient and accurate localization of RFID Tags.

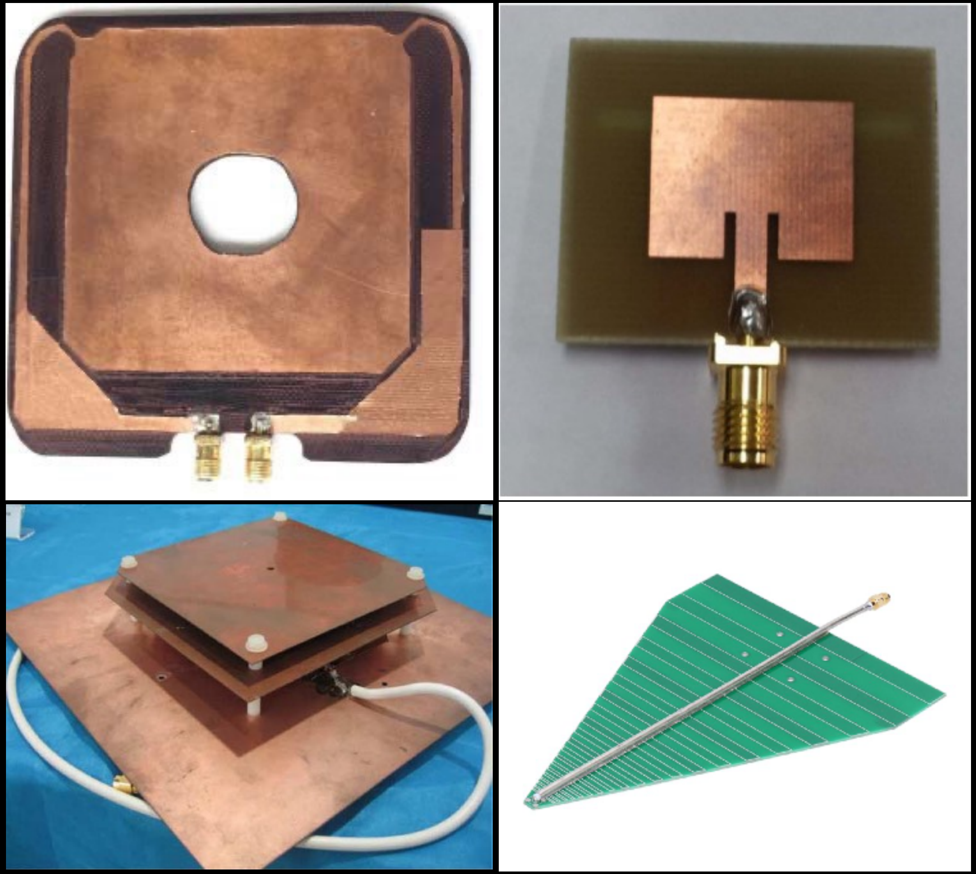
- It introduces “human-in-the-loop” design for sensing and localization RFIDs through RF-based Dynamic user interface.

- A user study of 20 people: It enables users to quickly find items with <10 cm accuracy across line-of-sight, non-line-of-sight, or fully occluded settings.

How does RF-AR work?

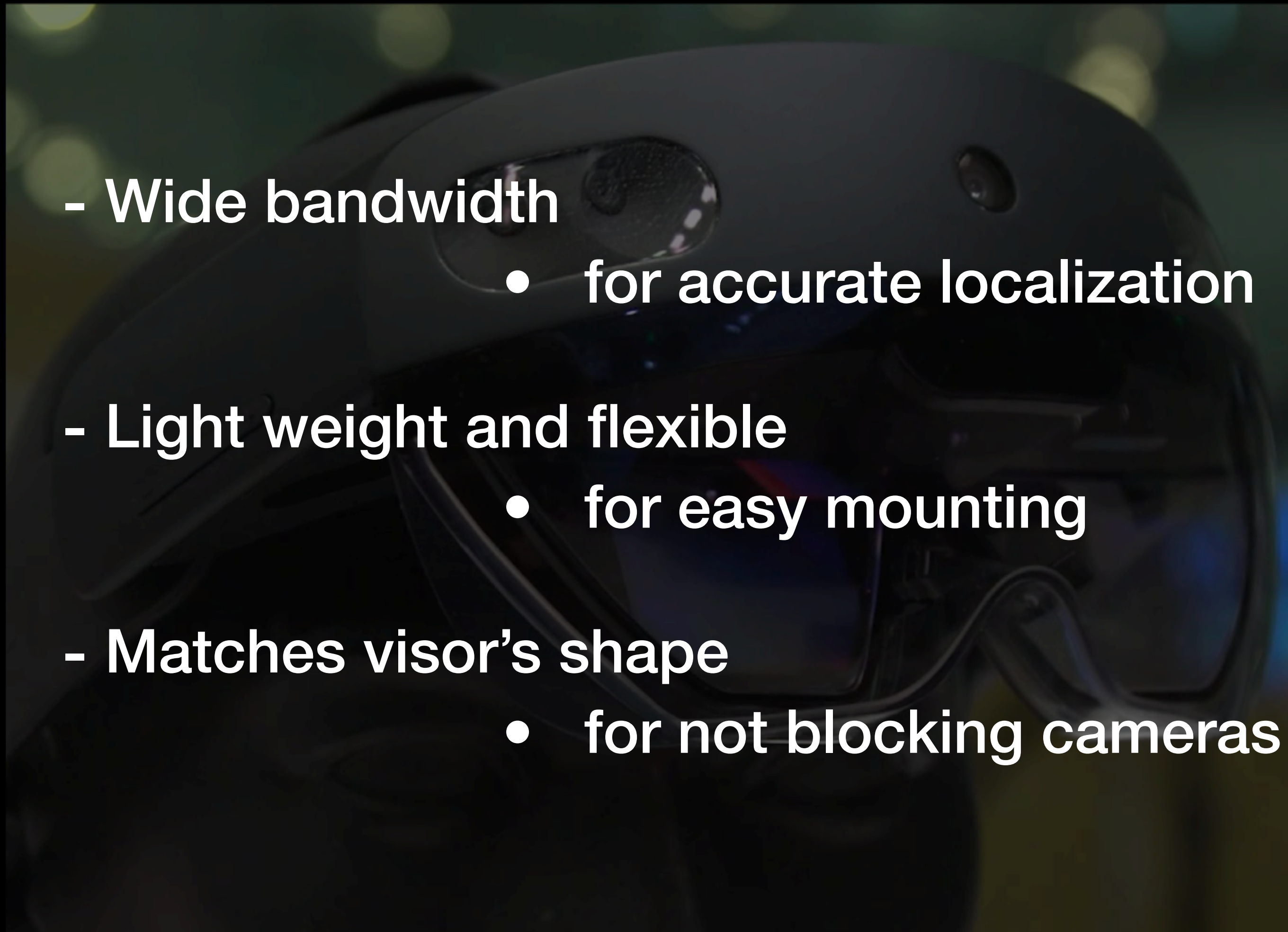
Designing an RFID reader antenna that conforms to AR headsets

RFID reader antennas come in many shapes and forms



J-RFID’19, IEEE RFID TA’13, IEEE MTT’09

Designing an RFID reader antenna that conforms to an AR headset is challenging



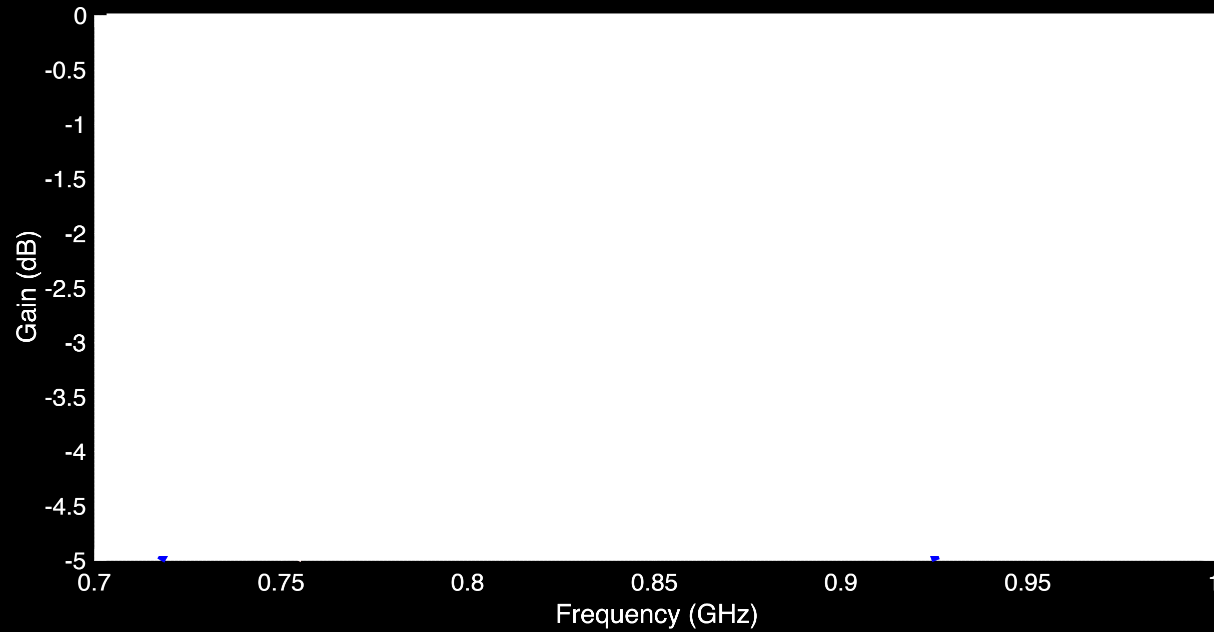
- Wide bandwidth
  - for accurate localization
- Light weight and flexible
  - for easy mounting
- Matches visor’s shape
  - for not blocking cameras

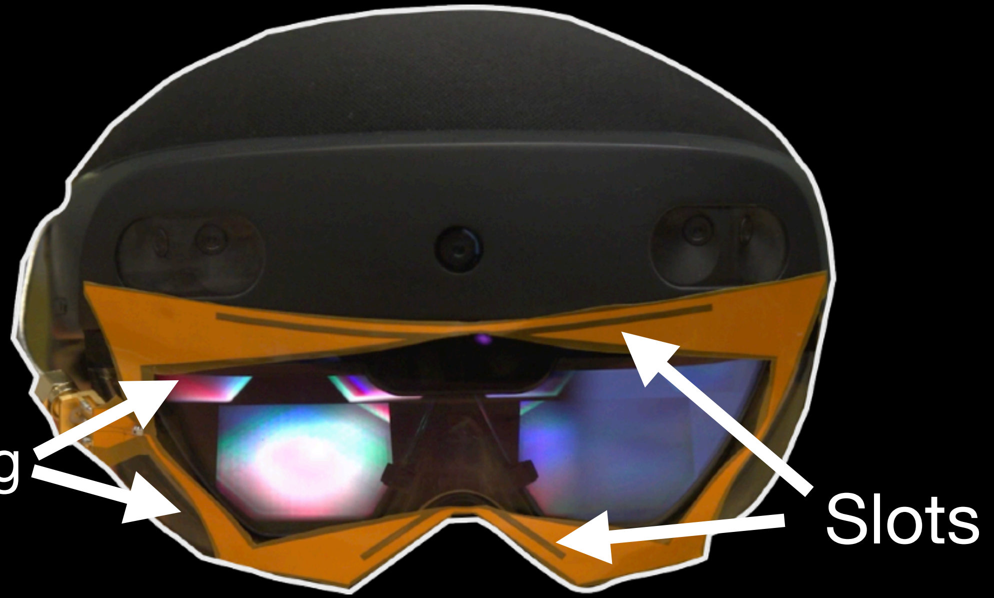
Designing an RFID reader antenna that conforms to AR headsets

✓ Light weight and conformal

✓ Does not block sensors

✗ Wide bandwidth



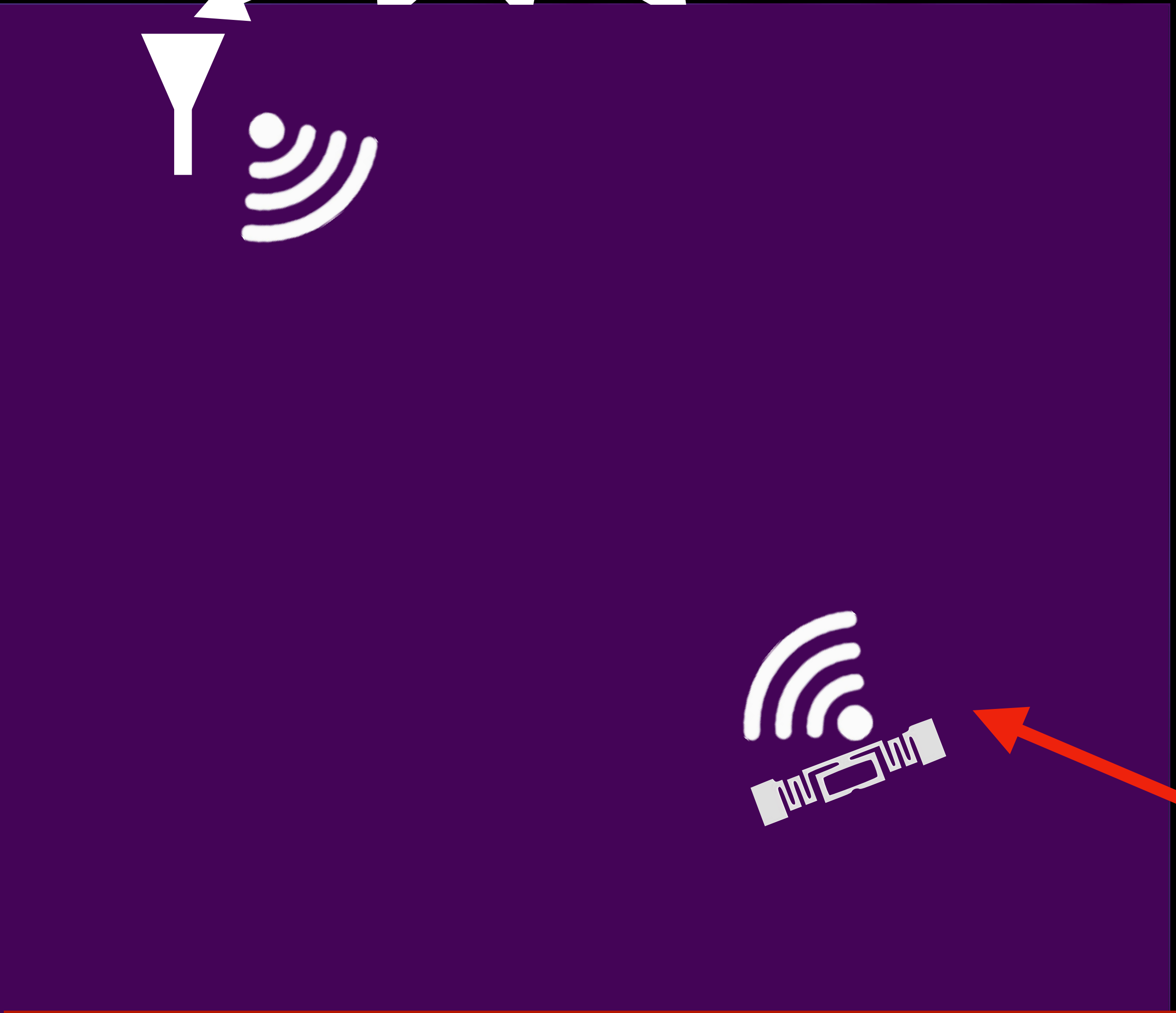


Wideband custom-designed antenna

✓ Wide bandwidth (200 MHz)

Synthetic Aperture Radar (SAR)

Antenna



Target RFID

How can we locate the antenna as the user walks in the environment?

Exploit AR headset sensors to track the human movements for SAR

Exploit AR headset sensors to track the human movements for SAR

Track Headset location
through visual and inertial odometry

Real World

Virtual 3D map
using Cameras on AR headset

Exploit AR headset sensors to track the human movements for SAR

Track Headset location
through visual and inertial odometry

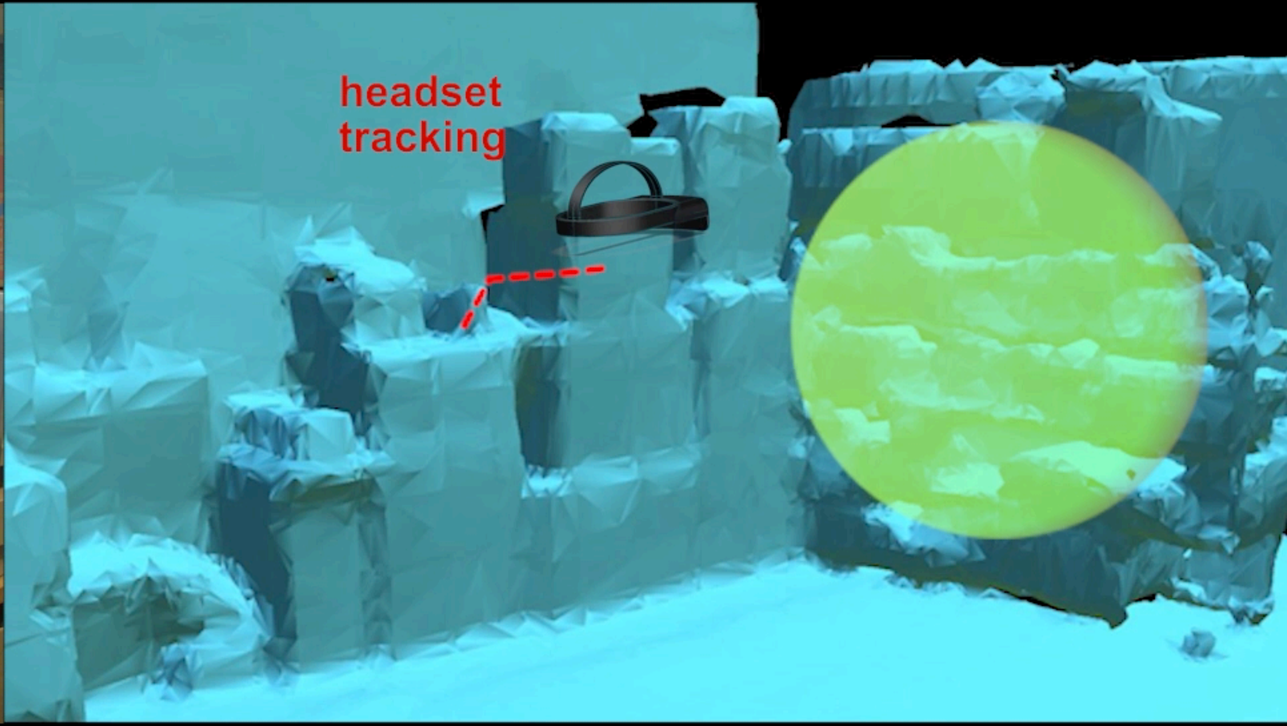
Real World

Virtual 3D map
using Cameras on AR headset

Exploit AR headset sensors to track the human movements for SAR

The user has to move back and forth, leading to inefficiency

Real World



Virtual 3D map
using Cameras on AR headset

How can RF-AR help the user to reduce task time?

Idea: Exploit the Synergies between AR and RFID Localization

AR can bring the human in the loop for <u>efficient</u> RFID localization

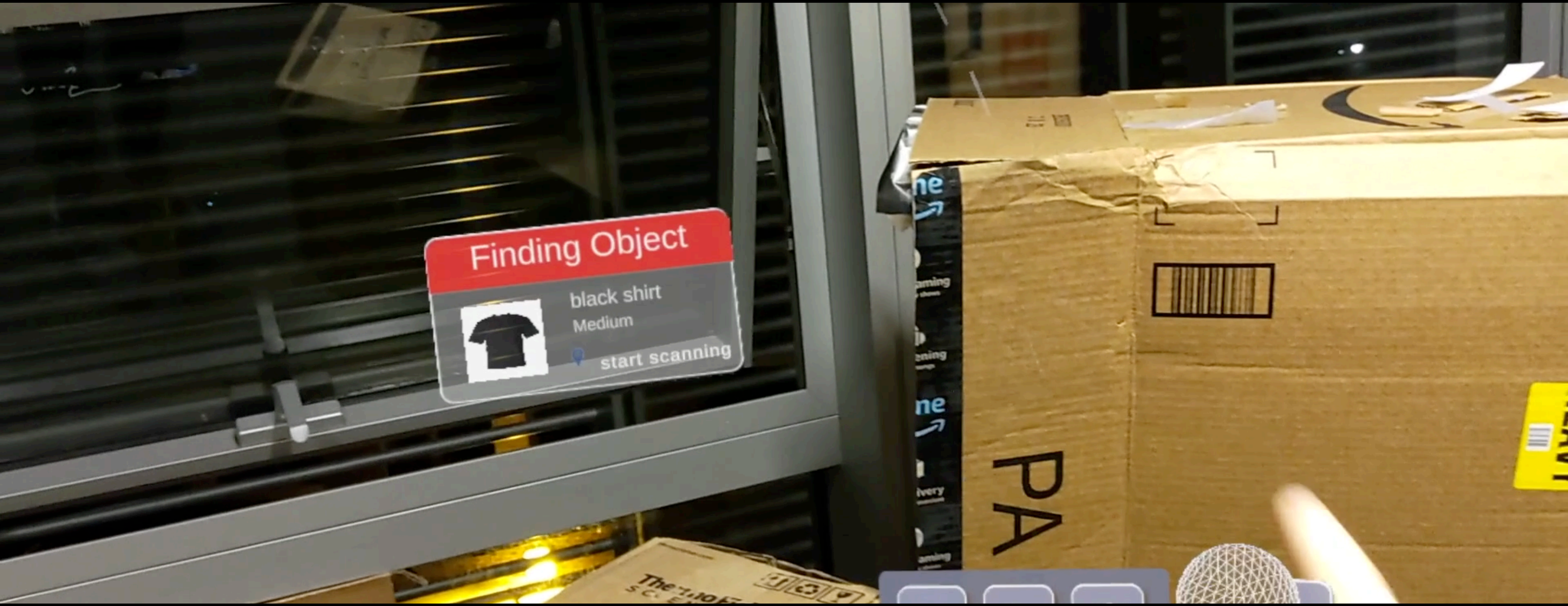
RF-AR improves efficiency by projecting a Dynamic RF-based User Interface

Idea: Exploit the Synergies between AR and RF sensing

Color Cues

Idea: Exploit the Synergies between AR and RF sensing

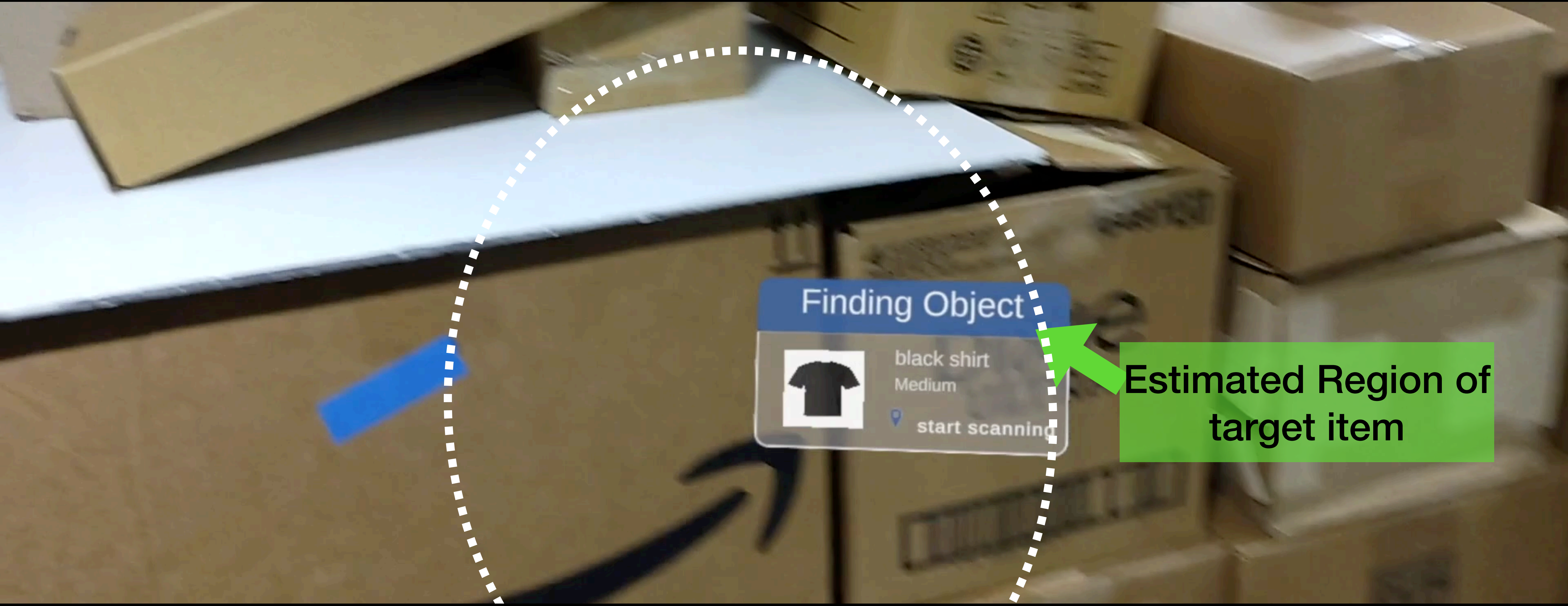
Color Cues



Users might stop walking in the environment when they see the blue color.

Idea: Exploit the Synergies between AR and RF sensing

Estimated Region Cues

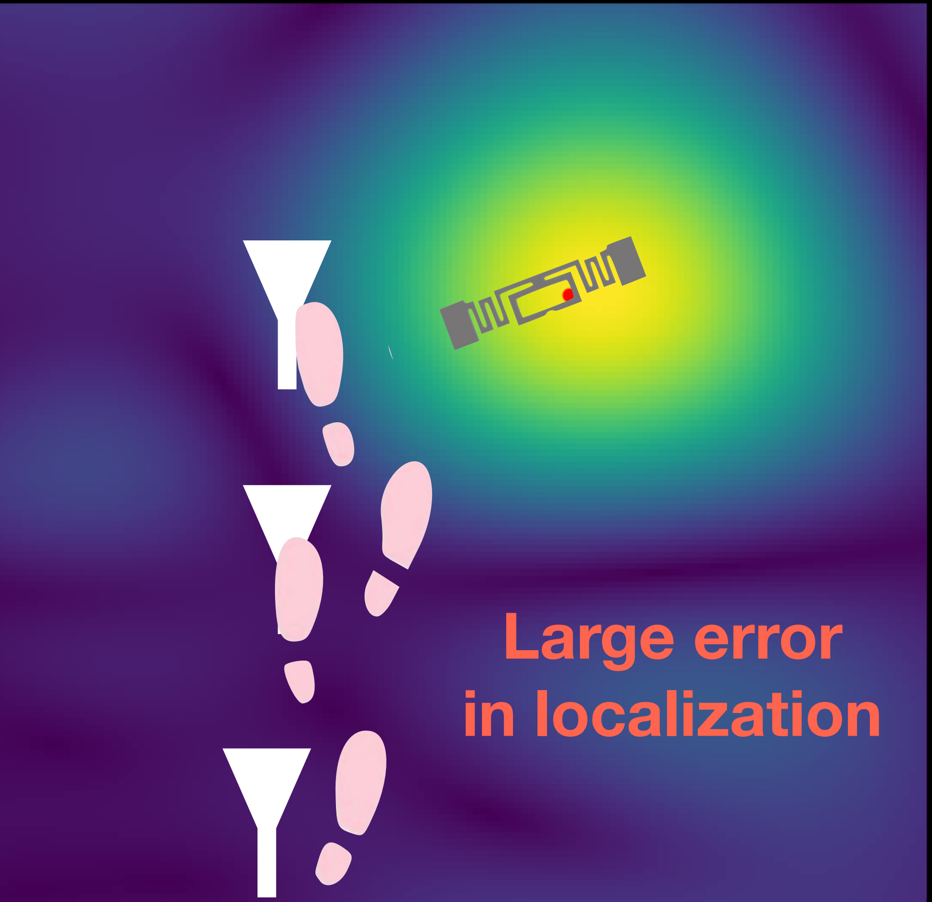


Walking on a straight line reduces the accuracy of RFID localization

Why walking on a straight line toward tag reduces the localization accuracy?

Dilution of Precision (DoP)

Large Dilution of Precision



Why walking on a straight line toward tag reduces the localization accuracy?

Dilution of Precision (DoP)

Small diversity in measurement angles

Large Dilution of Precision

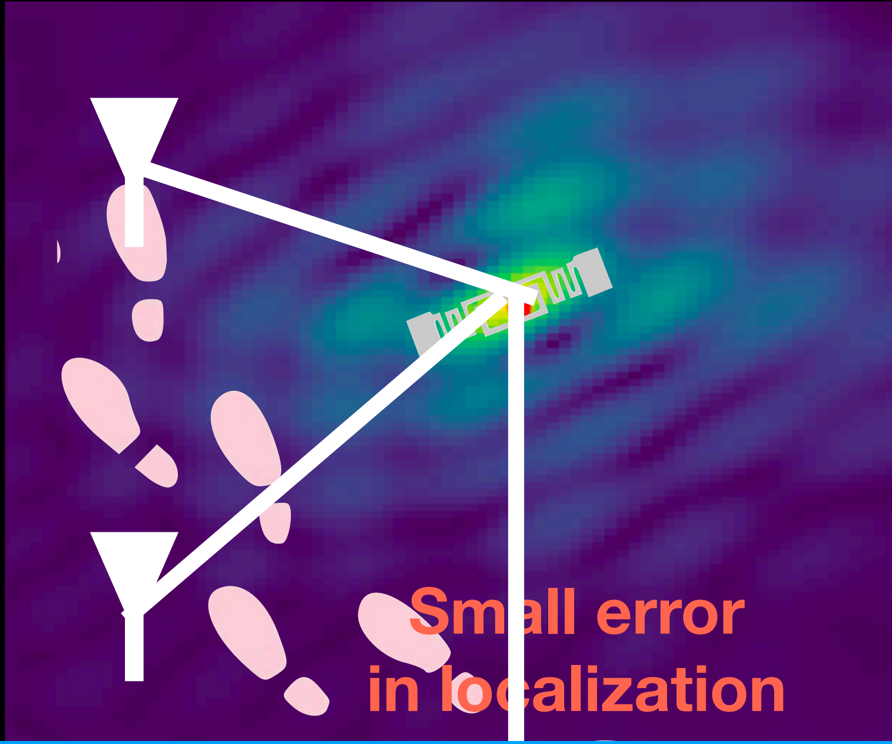
$$A = \begin{bmatrix} \frac{x_1 - x_p}{R_1} & \frac{y_1 - y_p}{R_1} & \frac{z_1 - z_p}{R_1} \\ \frac{x_2 - x_p}{R_2} & \frac{y_2 - y_p}{R_2} & \frac{z_2 - z_p}{R_2} \\ \vdots & \vdots & \vdots \\ \frac{x_n - x_p}{R_n} & \frac{y_n - y_p}{R_n} & \frac{z_n - z_p}{R_n} \end{bmatrix}$$

$$Q = \left(A^T A\right)^{-1}$$

$$DOP = \sqrt{tr(Q)}$$

Large diversity in measurement angles

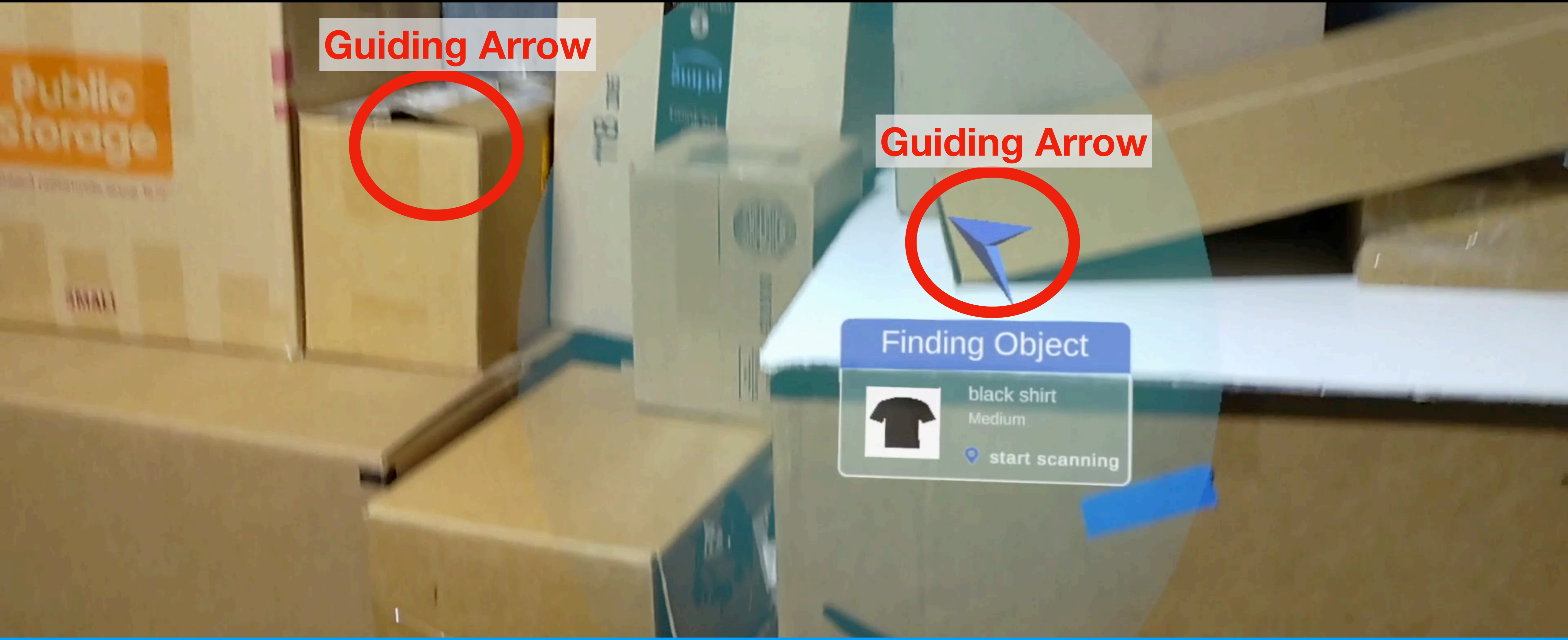
Small Dilution of Precision



RF-AR needs to abstract away this technical information and guide the user

How does RF-AR guide the user?

Guiding Arrows

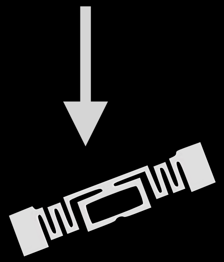


How does RF-AR calculate the optimum trajectory and direction ?

Optimum trajectory should result in high SNR measurements

We are limited to the antenna’s Field of View

RFID outside antenna FOV

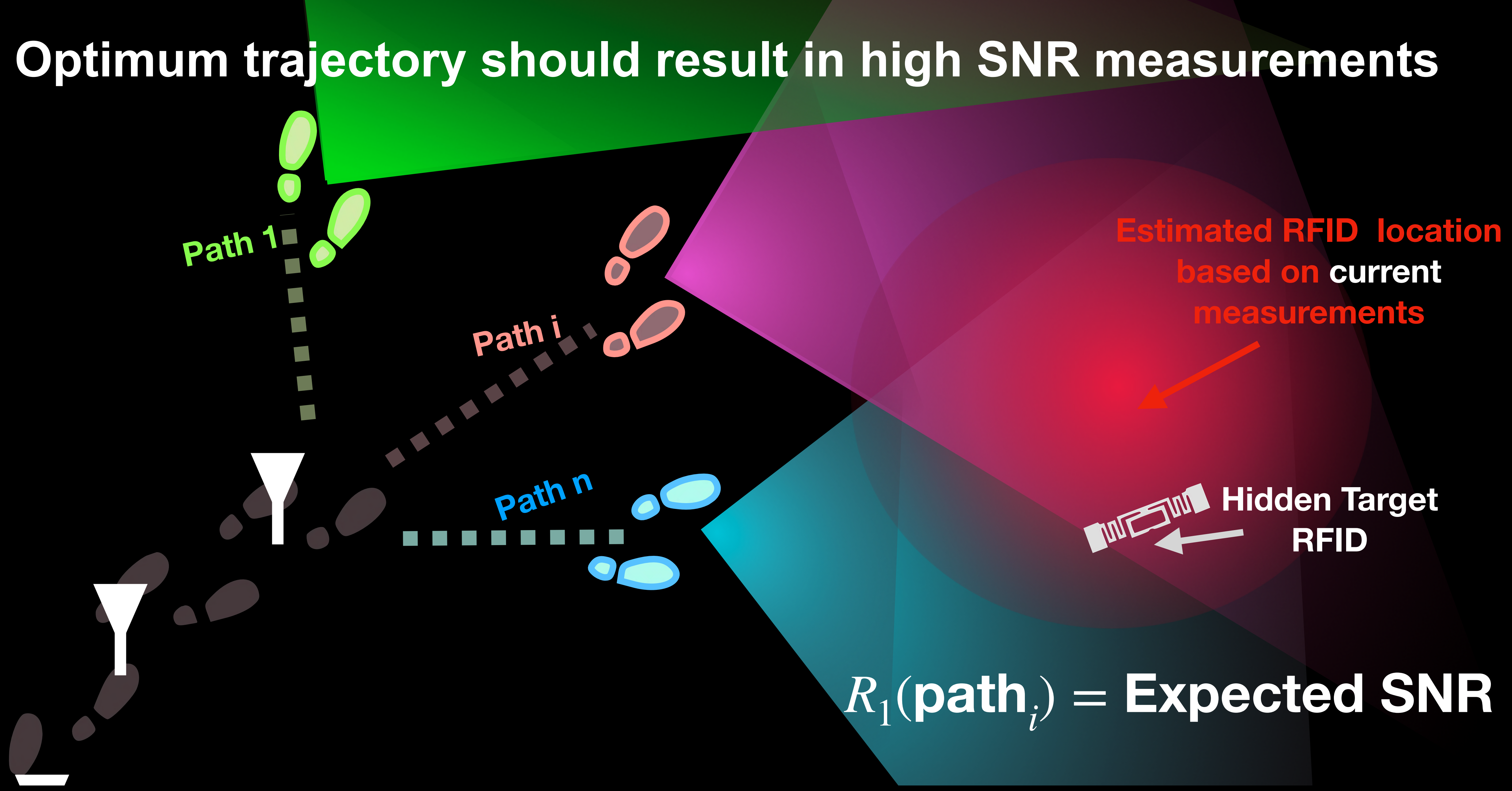


X RF measurement with low SNR

RFID inside antenna FOV

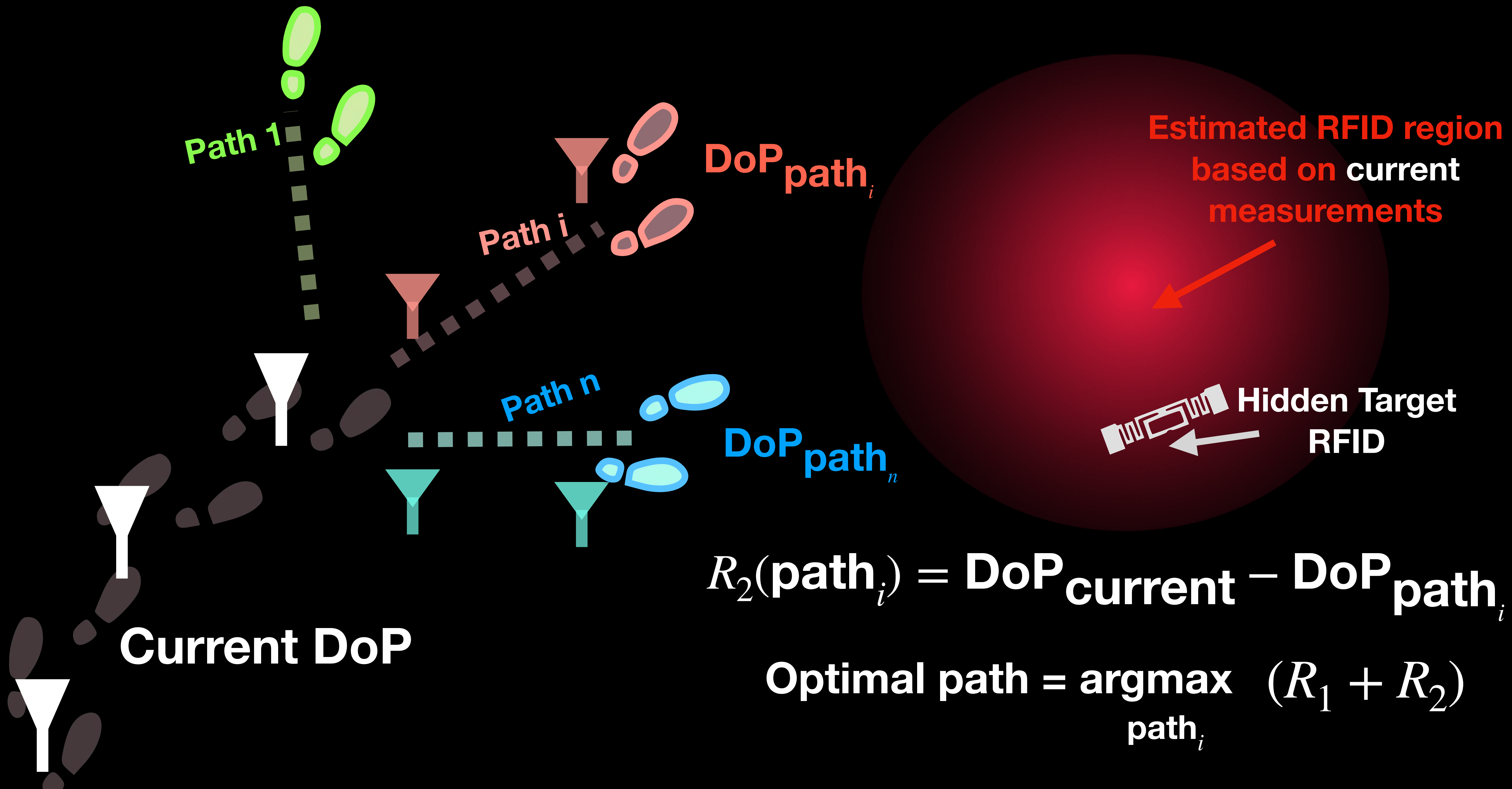
✓ RF measurement with high SNR

Optimum trajectory should result in high SNR measurements



The optimum measurement locations should reduce the DoP and time to convergence

Optimum trajectory should improve DoP



$$R_2(\text{path}_i) = \text{DoP}_{\text{current}} - \text{DoP}_{\text{path}_i}$$

$$\text{Optimal path} = \underset{\text{path}_i}{\text{argmax}} \ (R_1 + R_2)$$

How does RF-AR seamlessly guide the user?

Directing Arrows

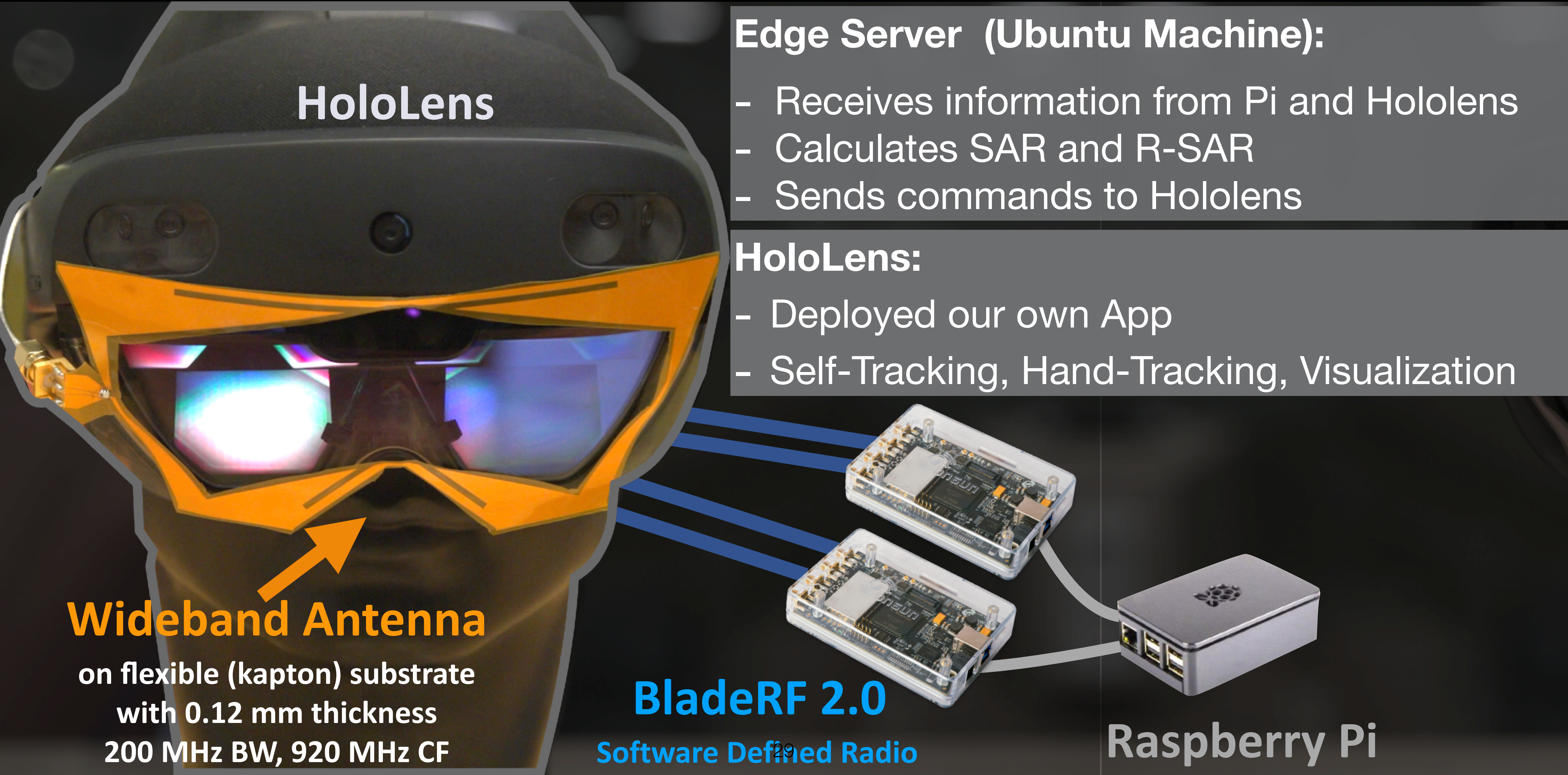
Implementation

Edge Server (Ubuntu Machine):

- Receives information from Pi and Hololens
- Calculates SAR and R-SAR
- Sends commands to Hololens

HoloLens:

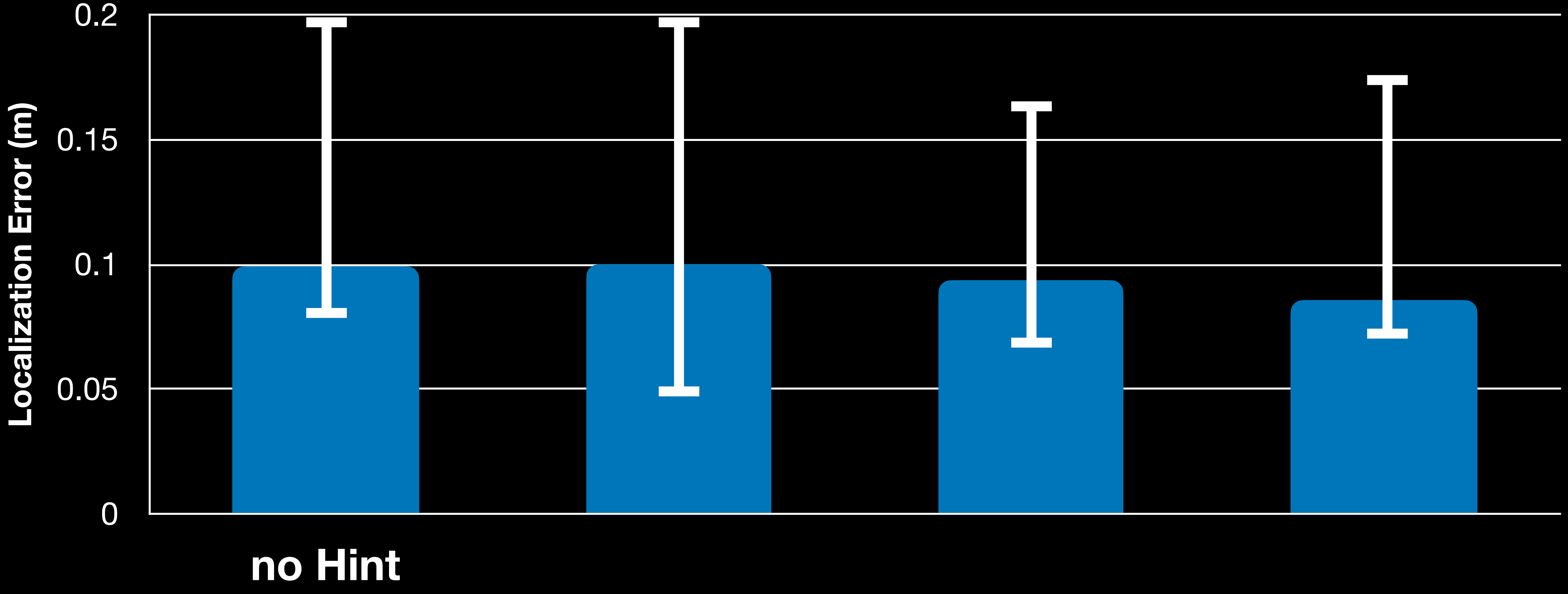
- Deployed our own App
- Self-Tracking, Hand-Tracking, Visualization



How accurately RF-AR locates RFID tags?

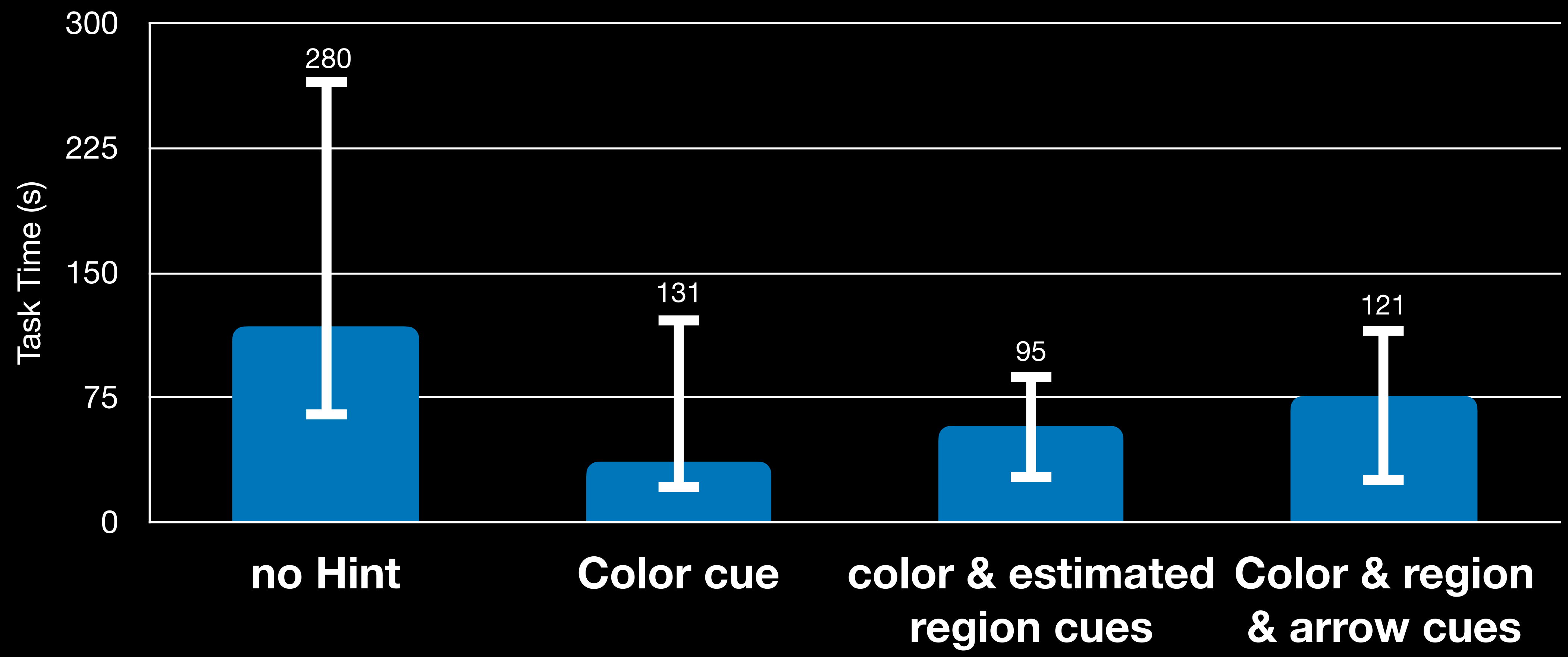
80 trials of user study on 20 people (14 males & 6 females, 22-34 years old)

Hid RFID tags in the environment and randomized order of UI



Different UI result in similar RFID tag localization accuracy

How much RF-AR improves efficiency and reliability?



Success Rate: 90% 90% 100% 100%

Find the item within 5 minutes and error < 25 cm

IEEE ISMAR 2024

GREATER SEATTLE AREA

XaiR

An XR Platform that Integrates Large Language Models with the Physical World

Sruti Srinidhi, Edward Lu, Anthony Rowe

Carnegie Mellon University

BOSCH

IEEE COMPUTER SOCIETY vgtc

Cognitive Assistants Should…

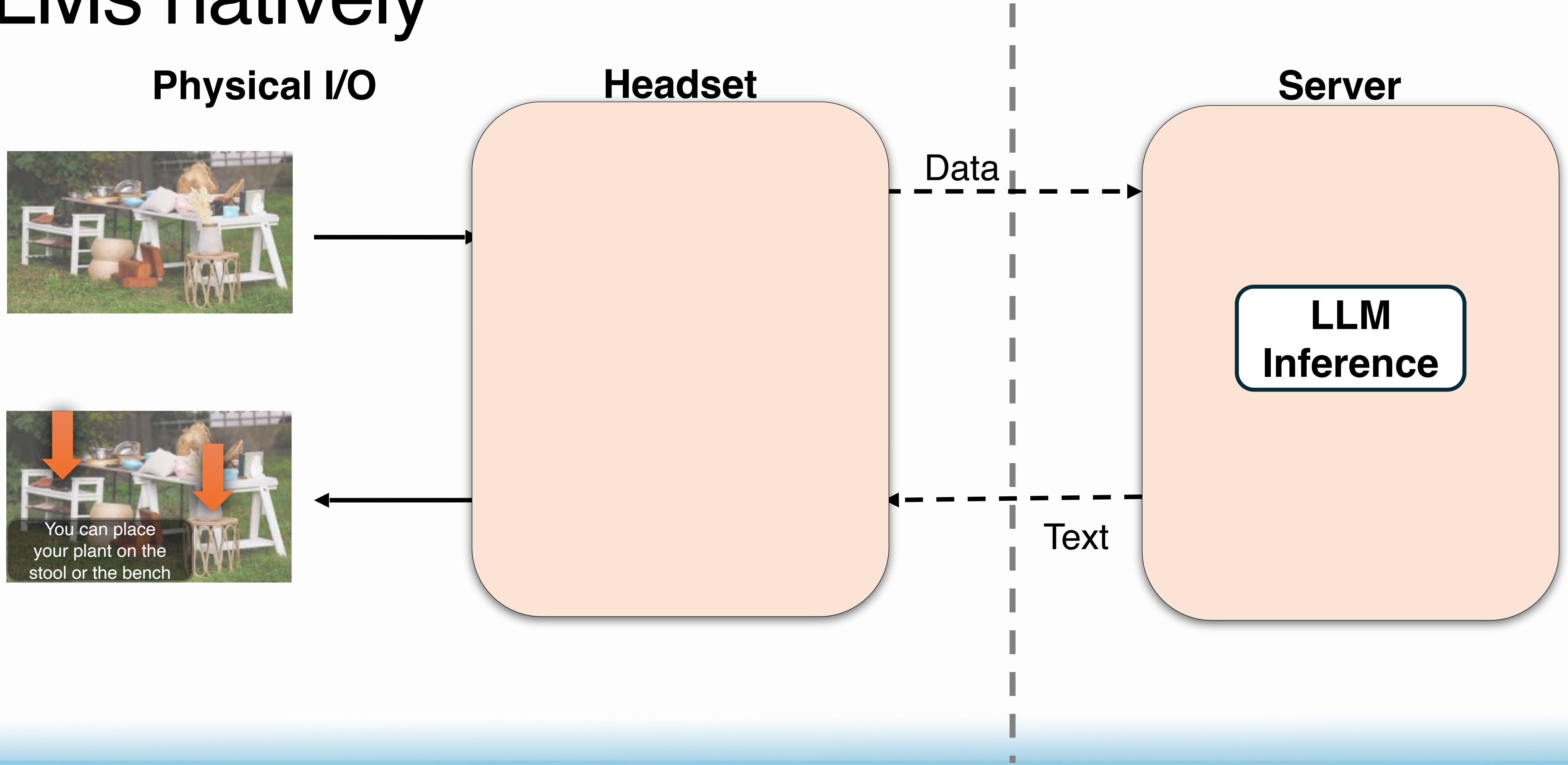
1. Understand the world

2. Reason about the world

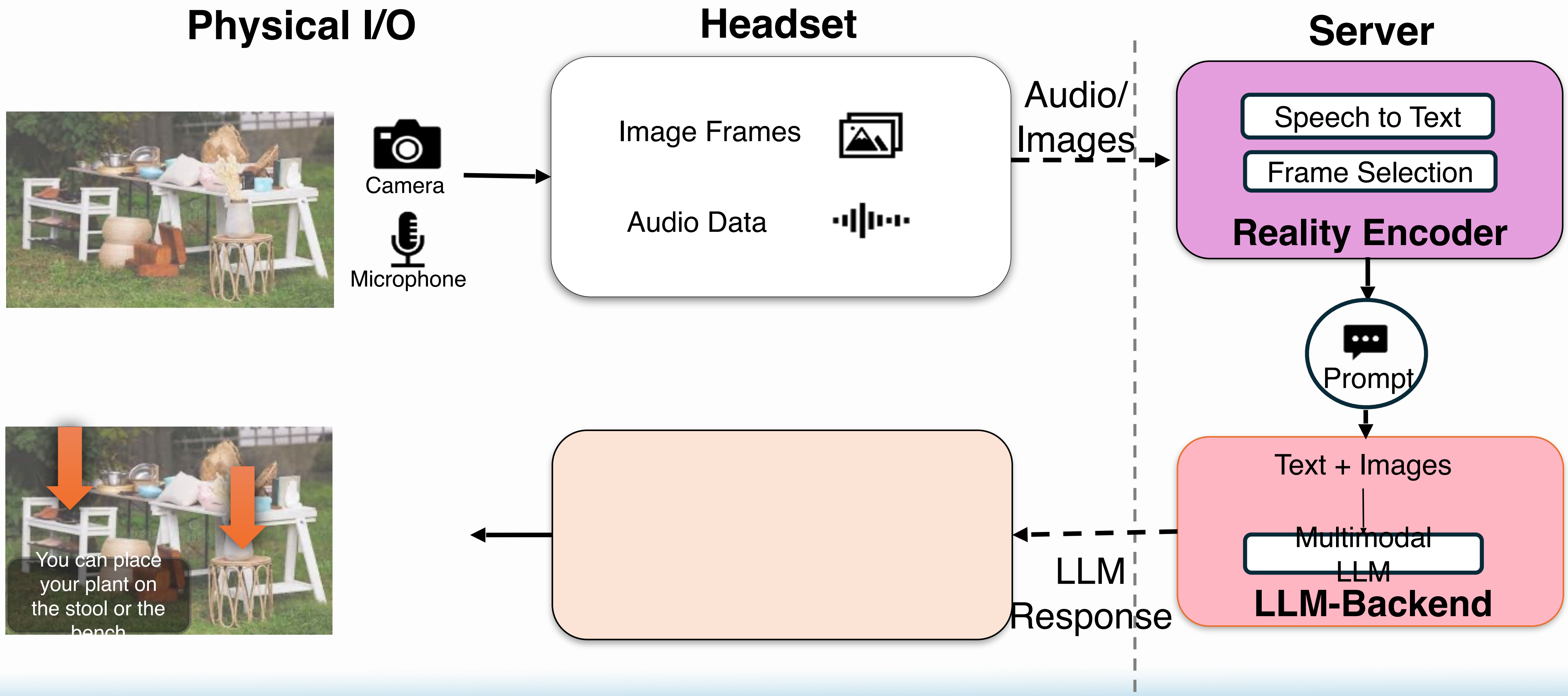
3. Respond to the world

But…

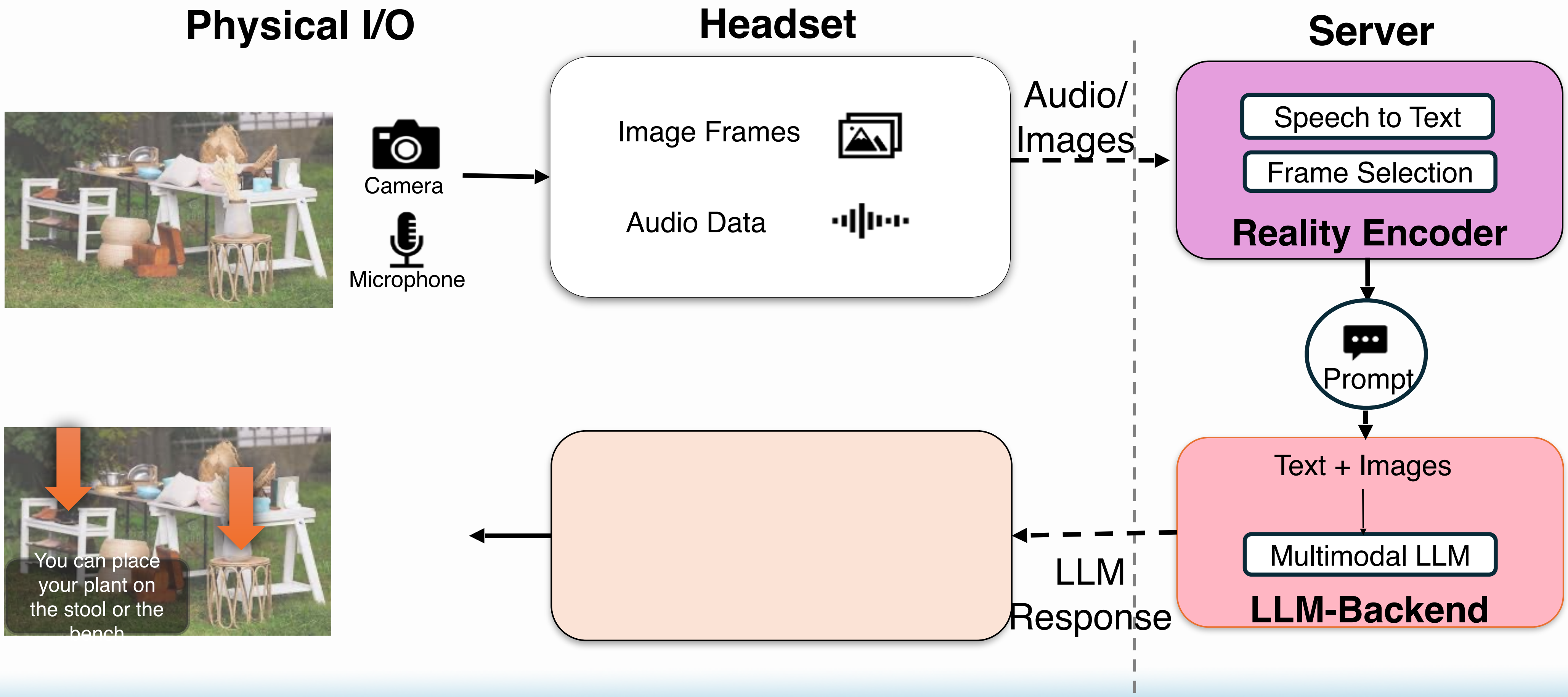
XR Headsets do not have the resources to run LLMs natively



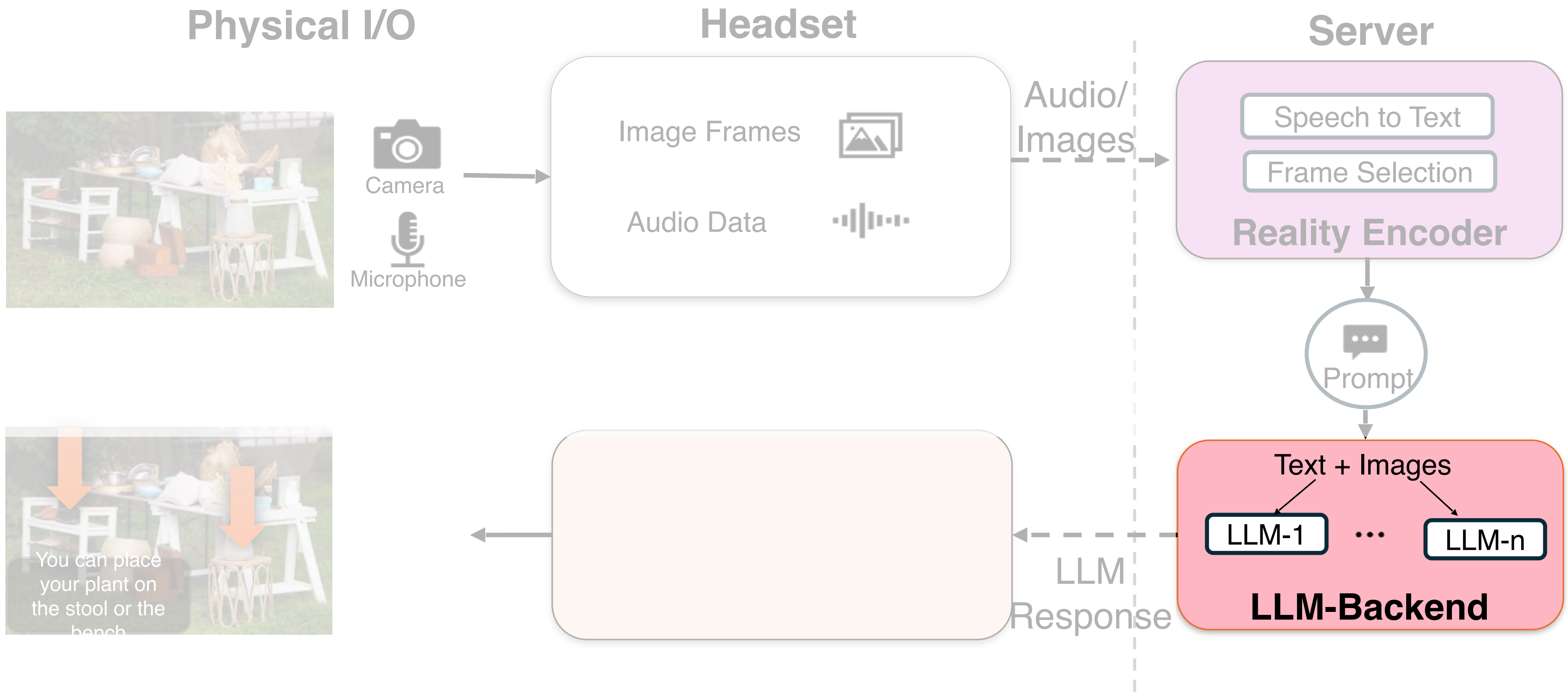
LLMs struggle to understand the user's 3D environment



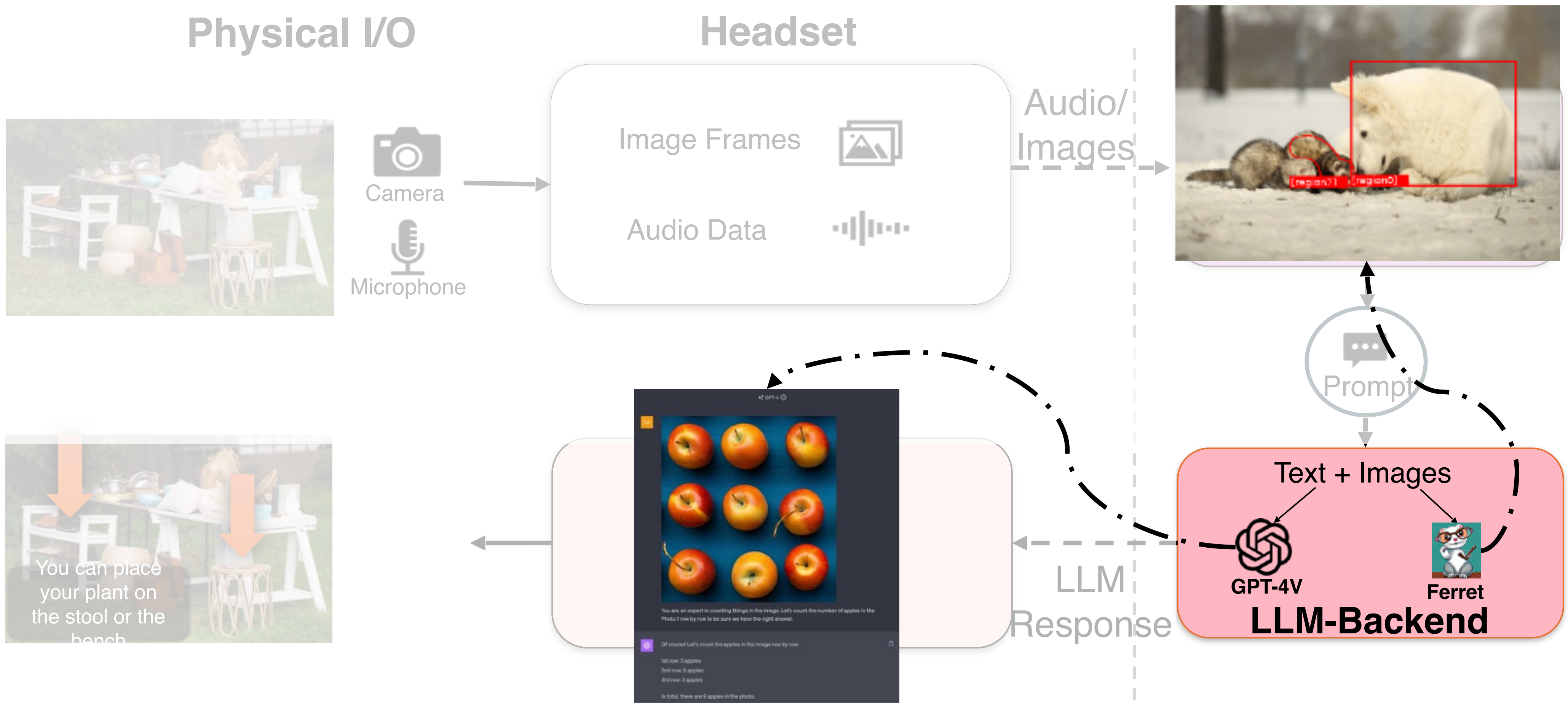
But Different LLMs are Trained for Different Purposes



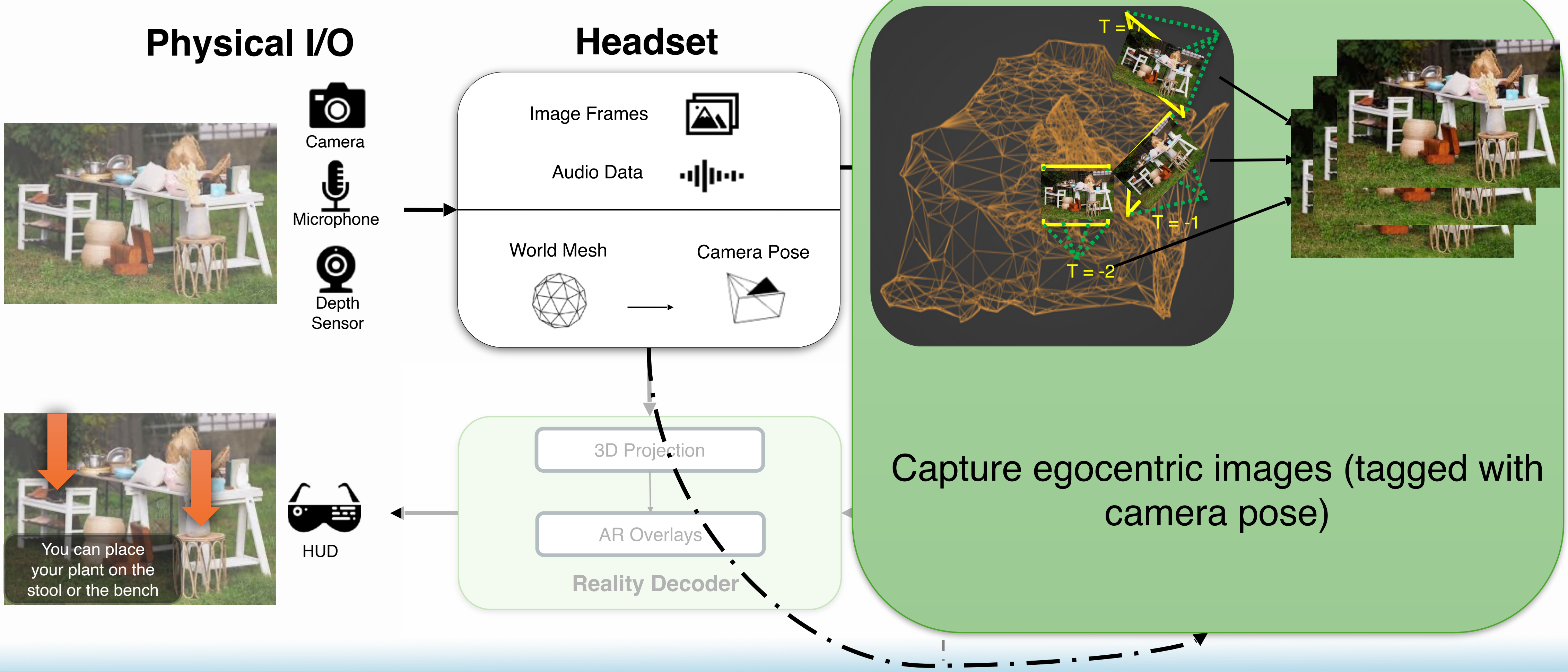
But Different LLMs are Trained for Different Purposes



But Different LLMs are Trained for Different Purposes

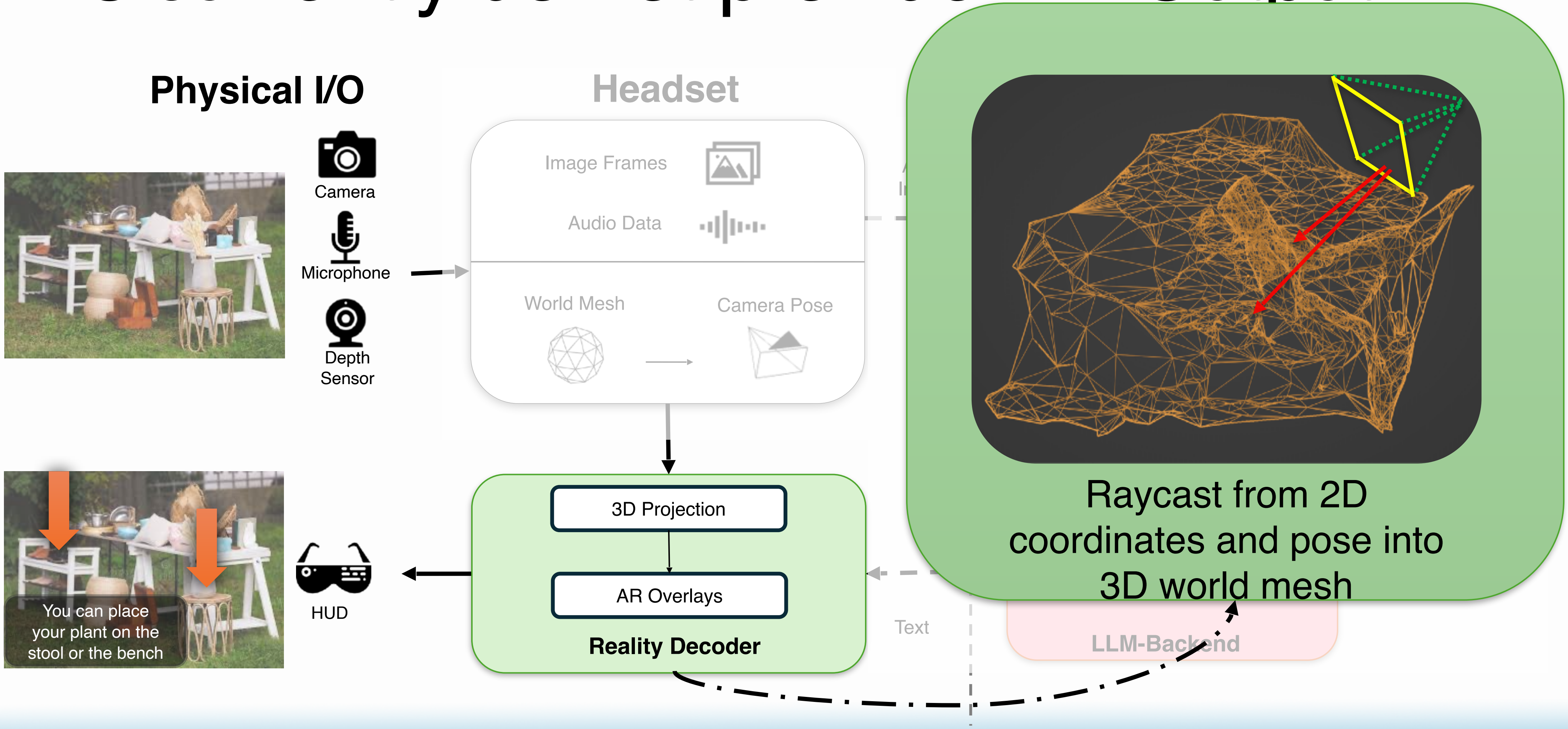


LLMs currently do not provide XR Output

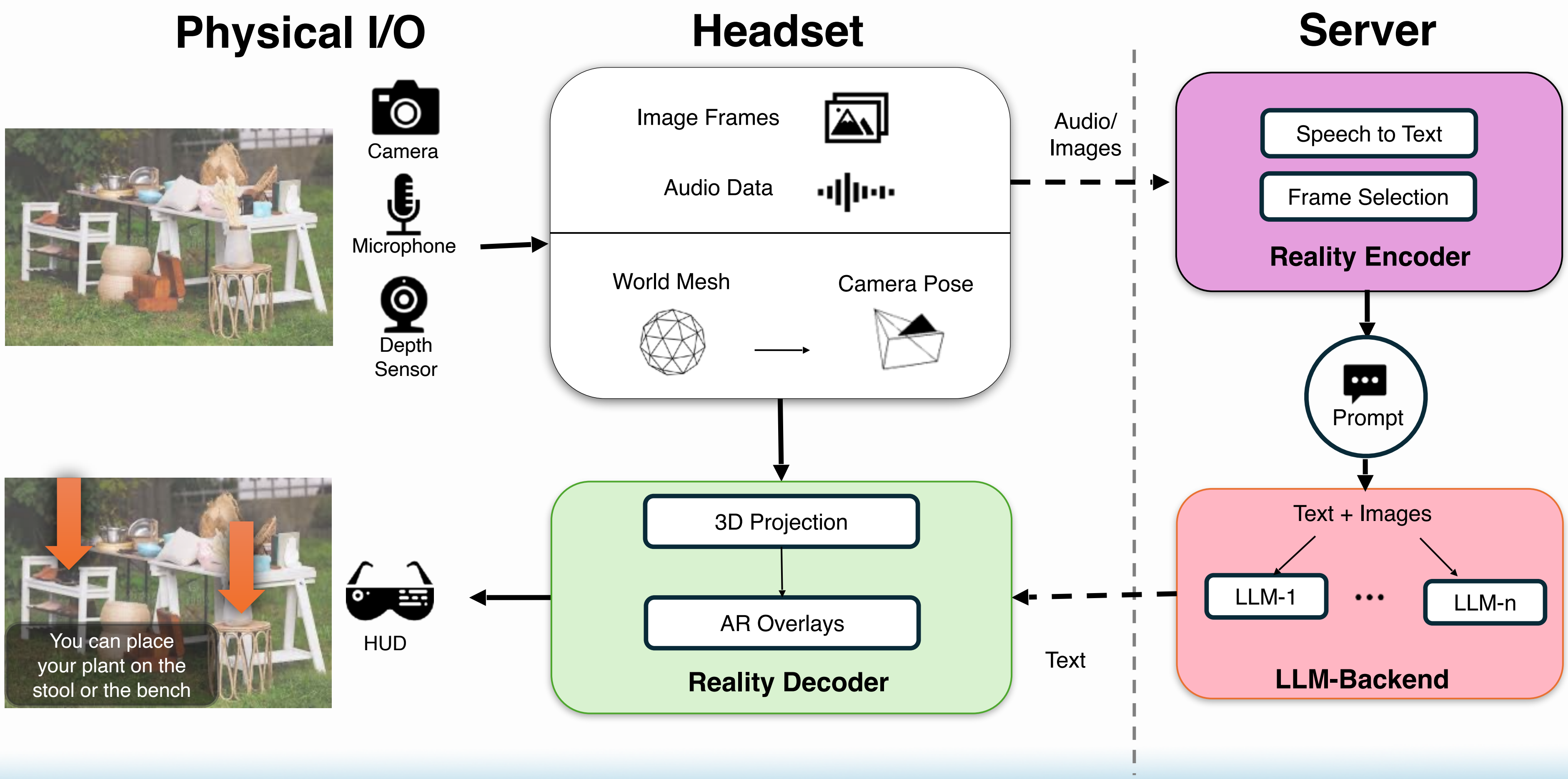


Capture egocentric images (tagged with camera pose)

LLMs currently do not provide XR Output



XaiR Architecture



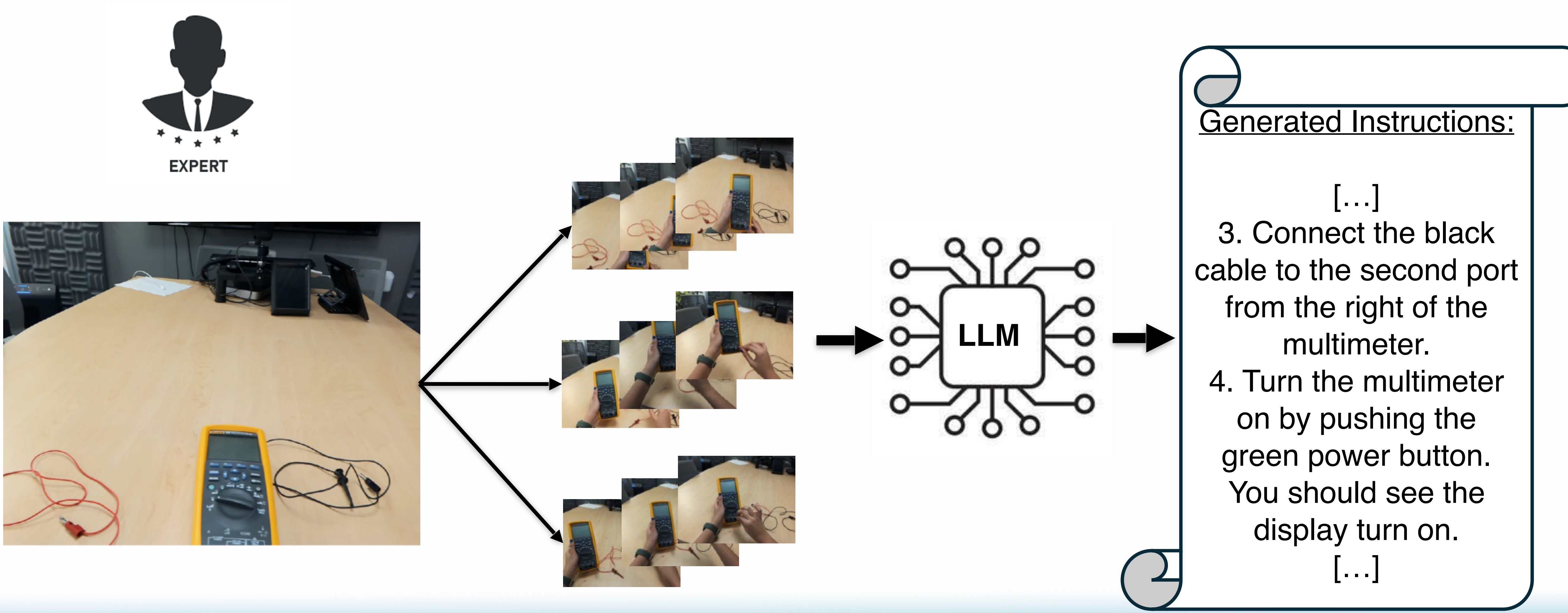
Example Application: Instruction Following Assistant

An assistant that can:

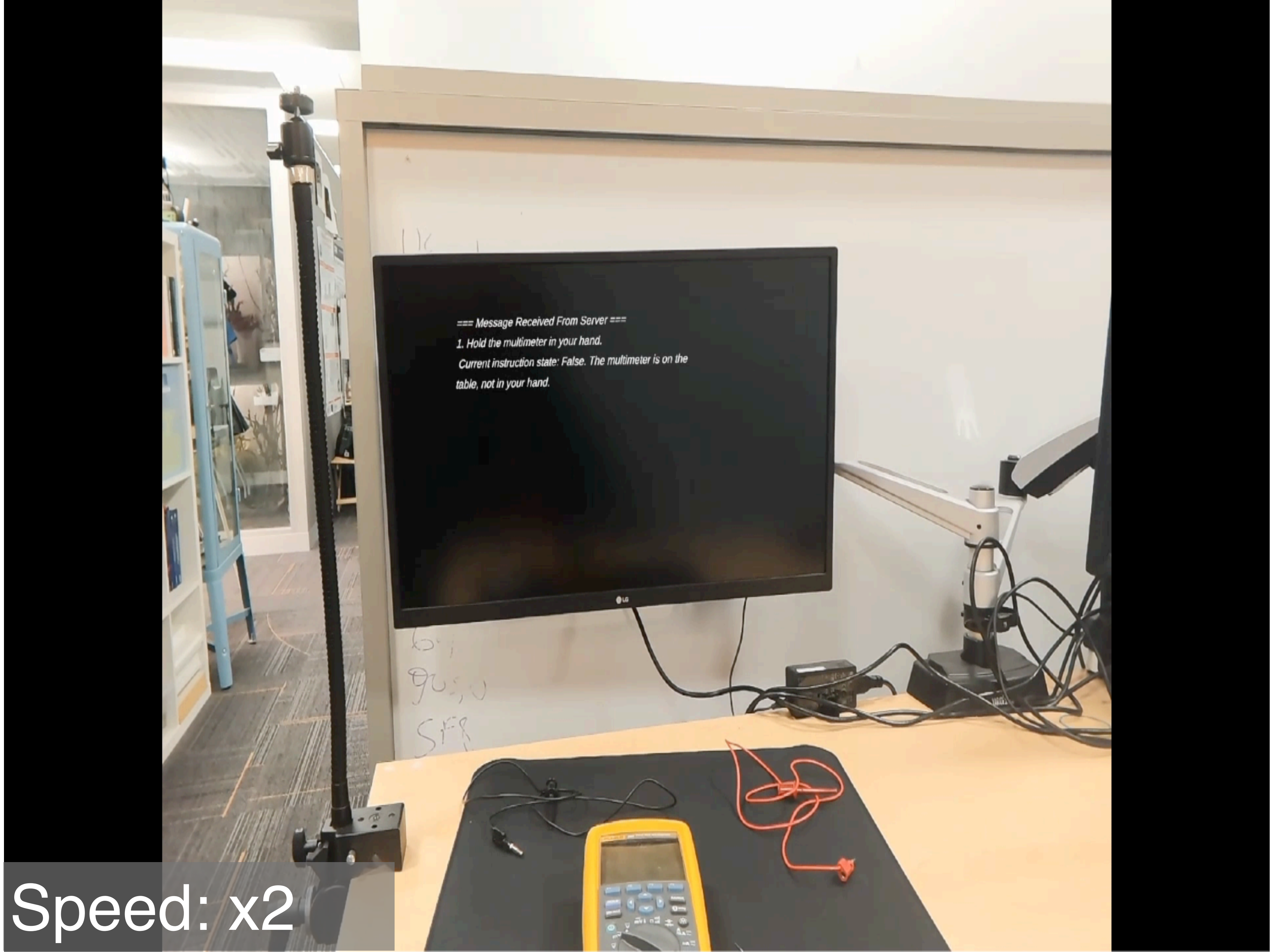
1. **Generate Instructions** based on an expert performing the task

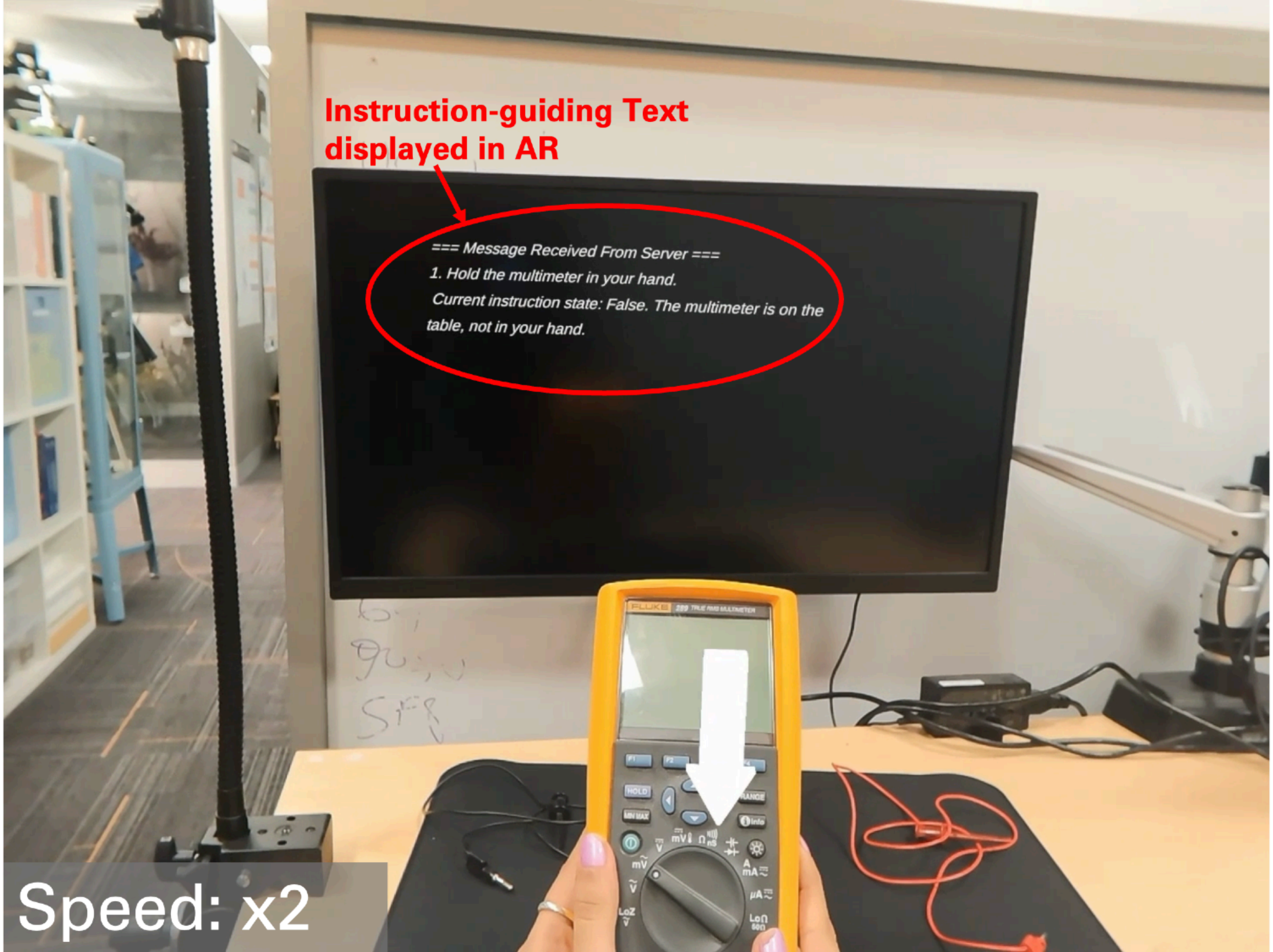
1. **Provide feedback** to a user so they can **follow the generated instructions**

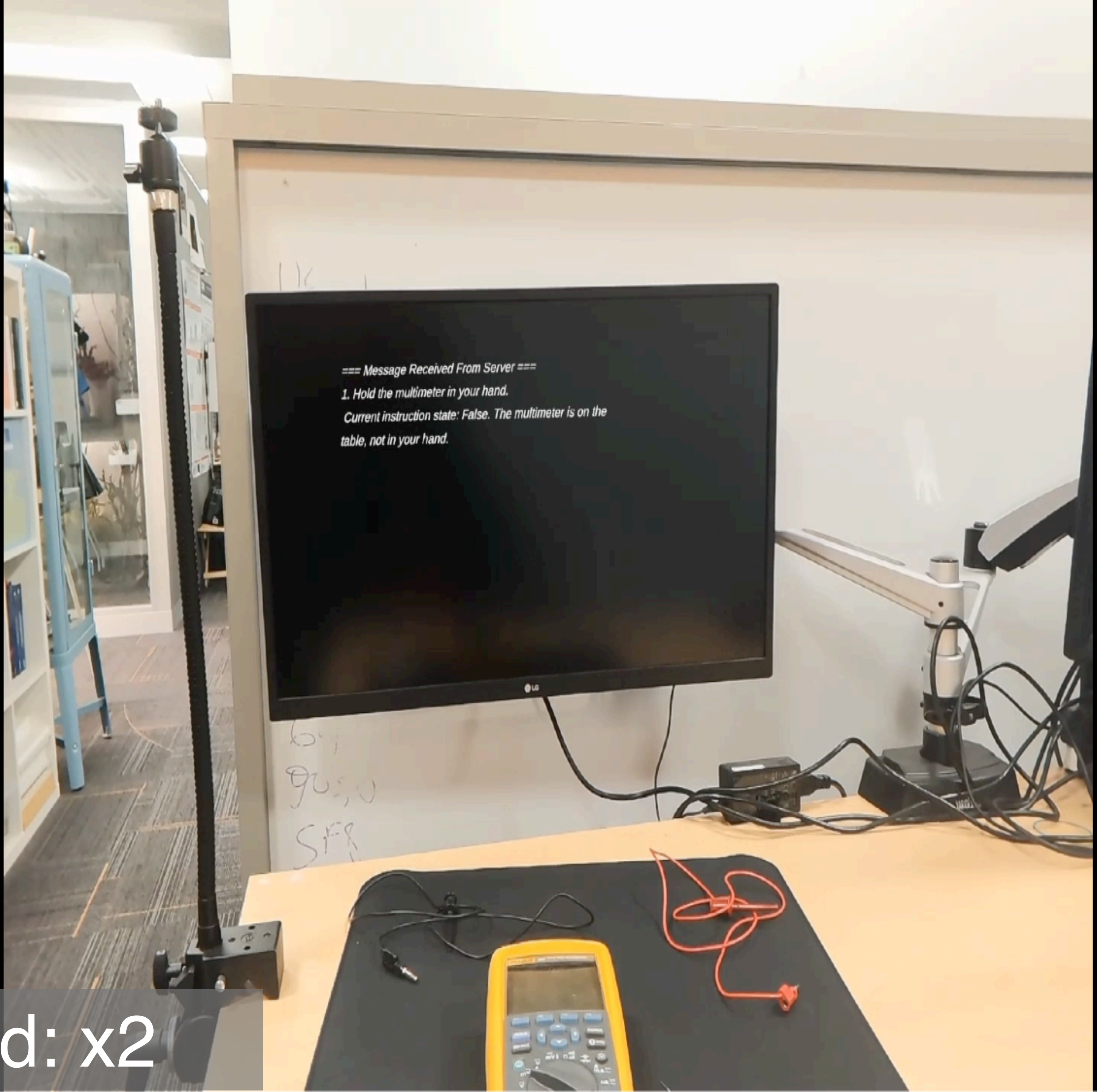
Step 1: Instruction Generation



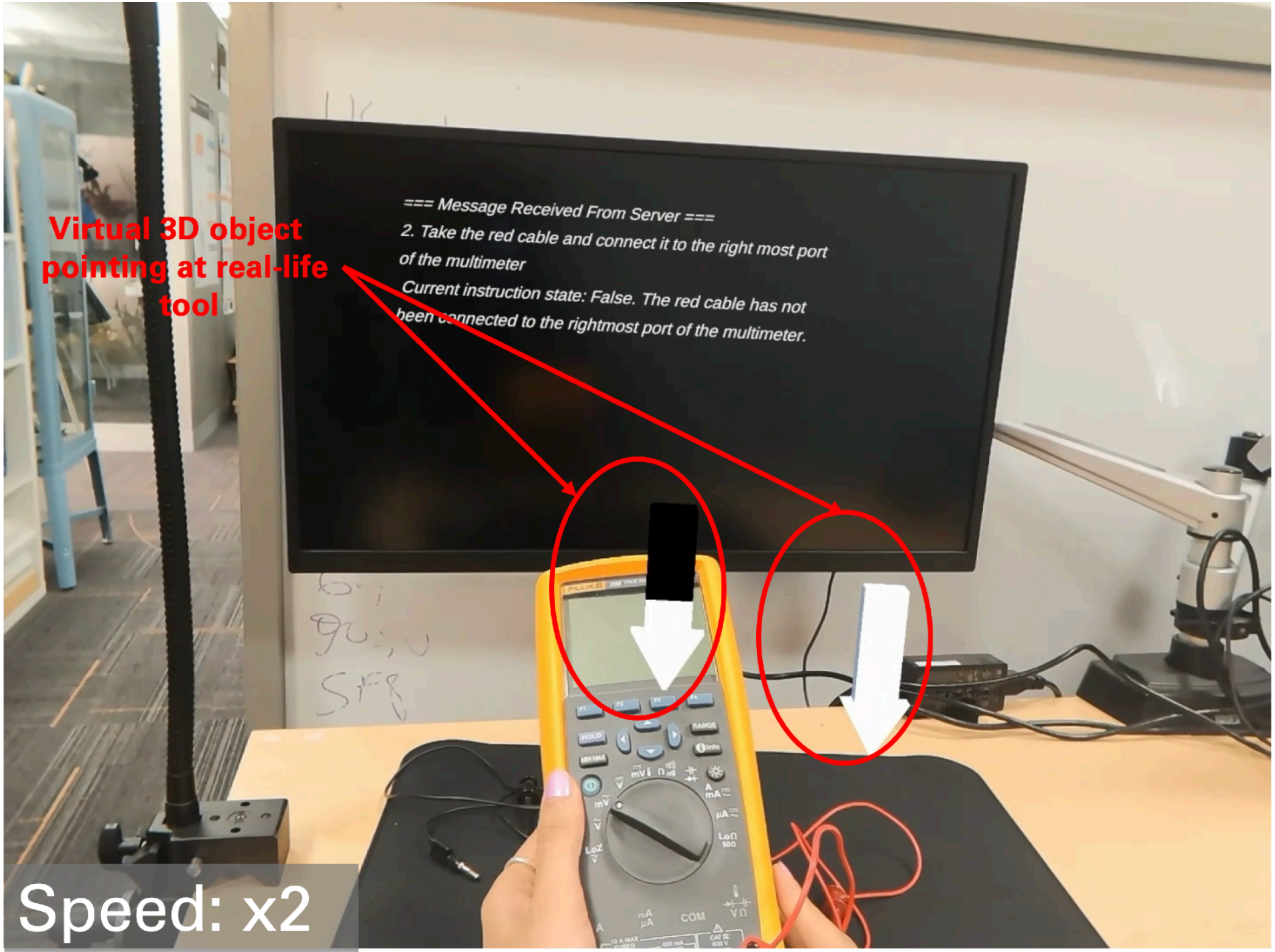
Step 2:
Assistant for Instruction Following


=== Message Received From Server ===
1. Hold the multimeter in your hand.
Current instruction state: False. The multimeter is on the table, not in your hand.
LG
Speed: x2

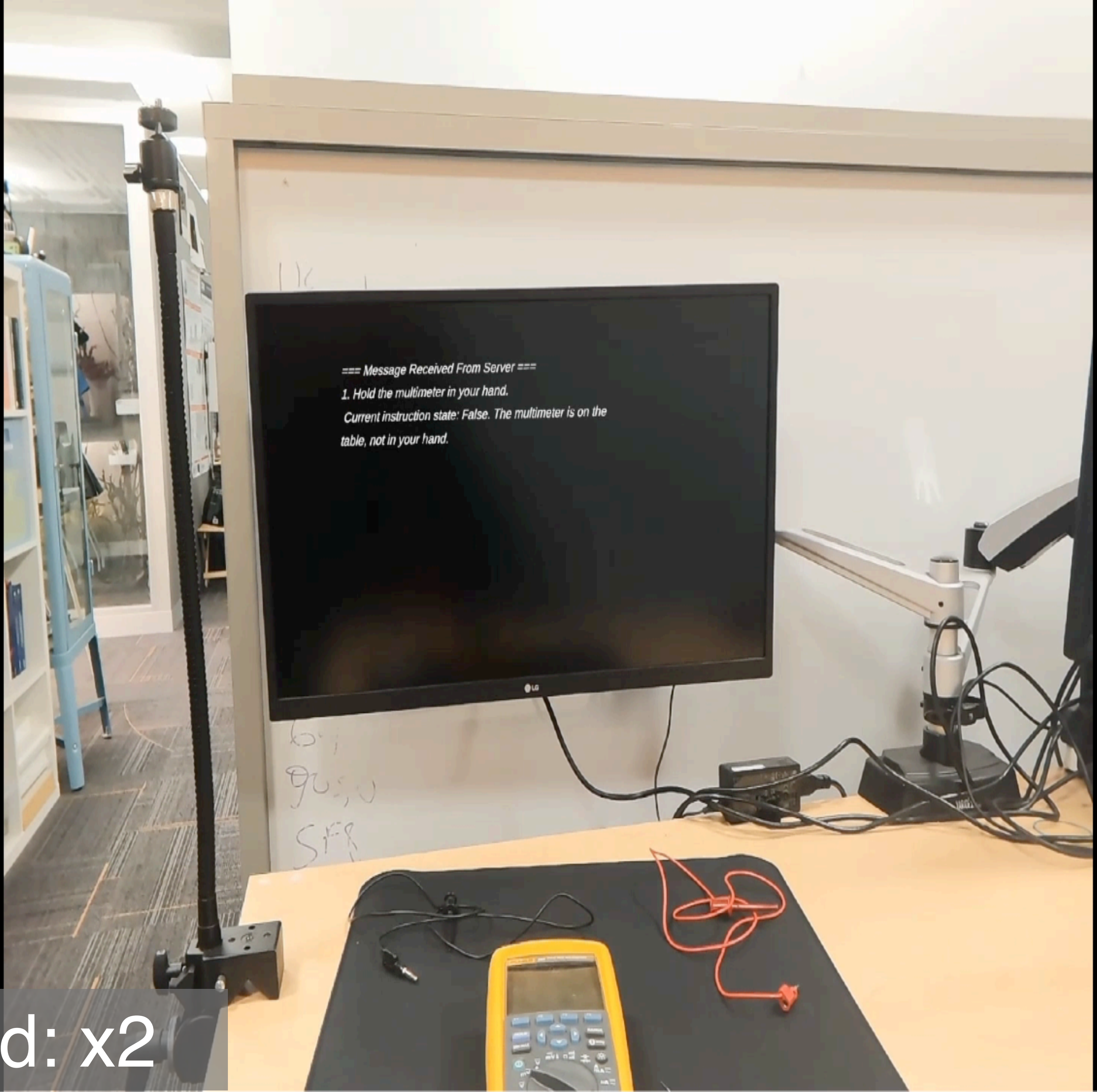

Instruction-guiding Text
displayed in AR
=== Message Received From Server ===
1. Hold the multimeter in your hand.
Current instruction state: False. The multimeter is on the
table, not in your hand.
FLUKE
HOLD
RANGE
Speed: x2


=== Message Received From Server ===
1. Hold the multimeter in your hand.
Current instruction state: False. The multimeter is on the table, not in your hand.
LG

Speed: x2



Virtual 3D object pointing at real-life tool
=== Message Received From Server ===
2. Take the red cable and connect it to the right most port of the multimeter
Current instruction state: False. The red cable has not been connected to the rightmost port of the multimeter.
Speed: x2

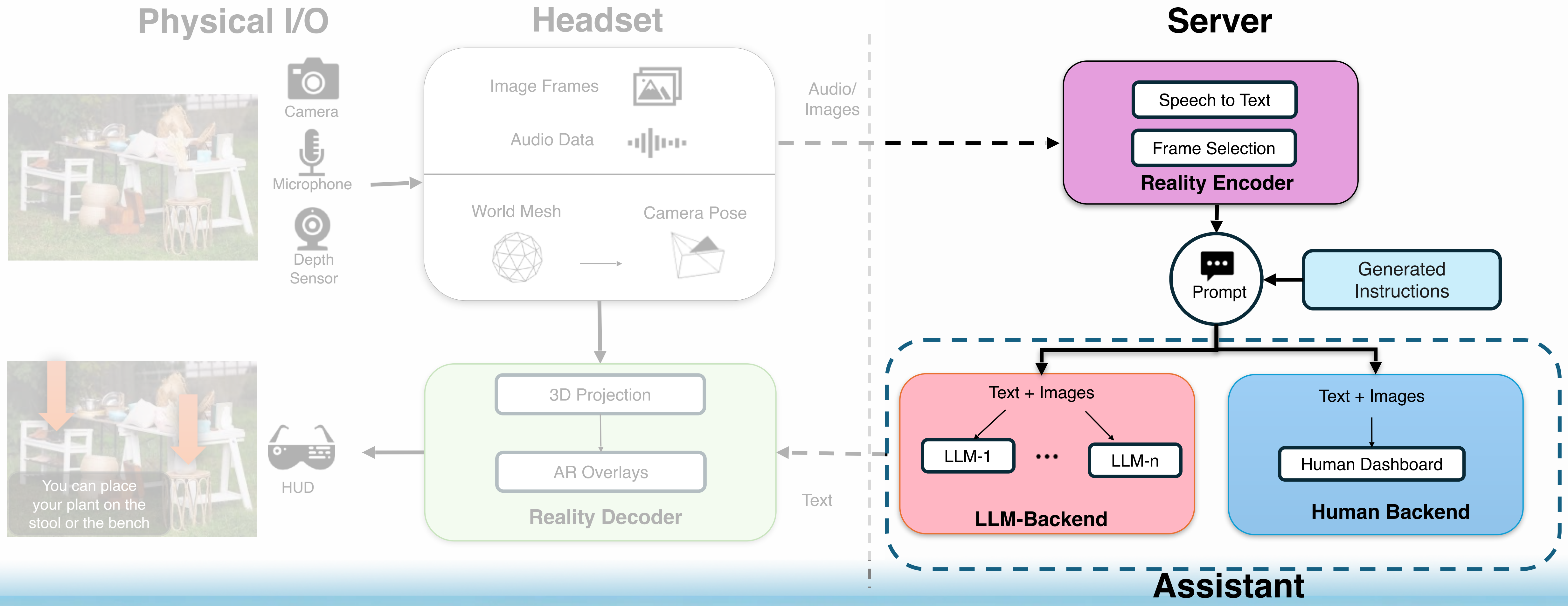

=== Message Received From Server ===
1. Hold the multimeter in your hand.
Current instruction state: False. The multimeter is on the table, not in your hand.
Speed: x2

How do we Evaluate Such a System?

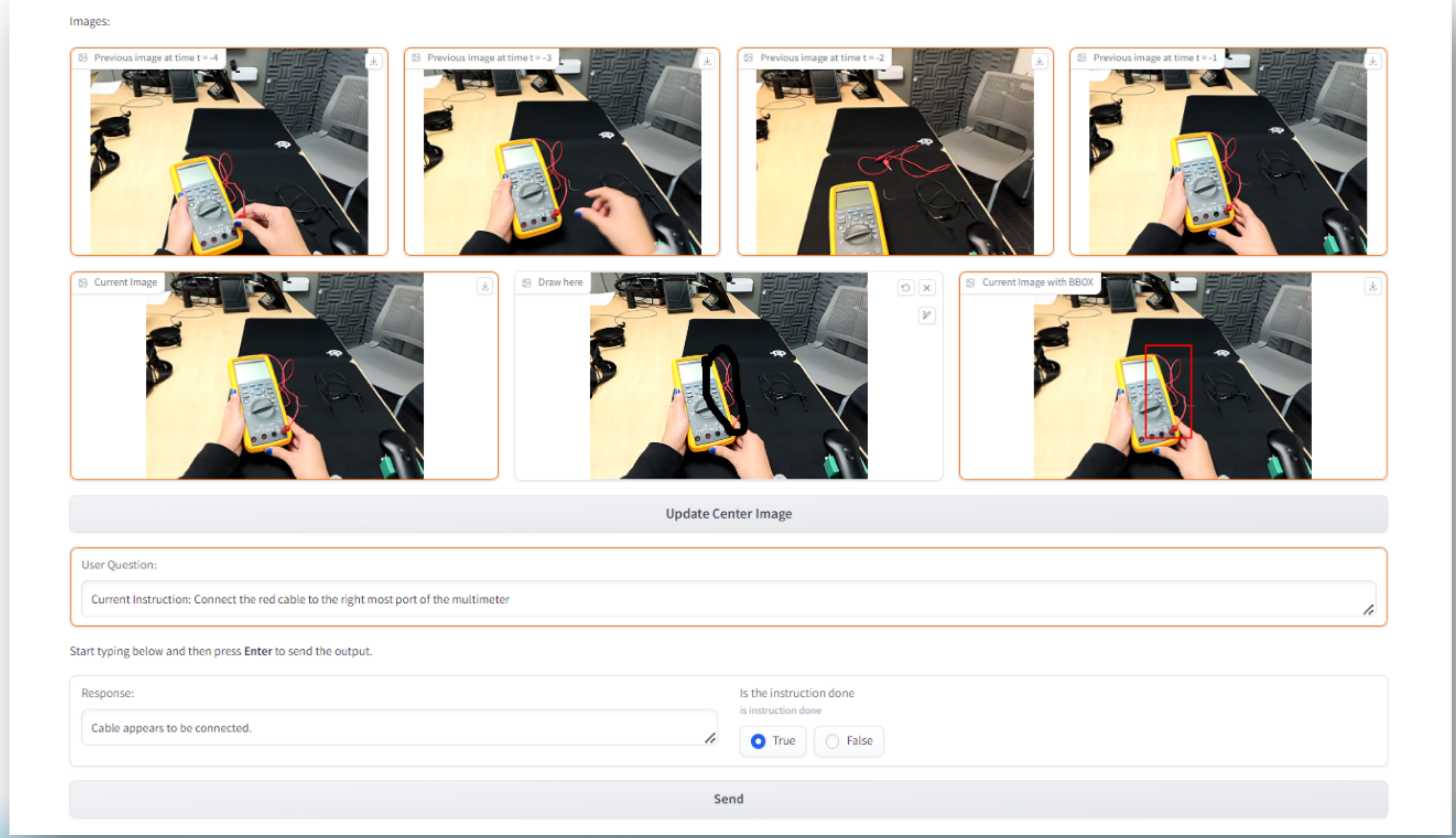
User Study (n=15):
Benchmark Cognitive Assistant against a Human

Method: The LLM-based cognitive assistant and a human expert assist users through the same task

User Study (n=15):
Benchmark Cognitive Assistant against a Human



User Study (n=15):
Benchmark Cognitive Assistant against a Human



User Study (n=15):
Benchmark Cognitive Assistant against a Human

User Tasks:

Basic Tasks:

Advance Tasks:

- Each user did 2 basic and 2 advanced tasks (one of each with the LLM-based assistant and one with the human expert)

User Study

Not Surprising :
LLMs are not as good as Humans!

	LLM-Backend Assistant	Human Assistant
Accuracy (%)	90	**100**
Queries to Assistant per Instruction	9.39	**3.2**
Time per Instruction (sec)	38.02	**27.86**
Time to complete Task (sec)	197.32	**151.15**
NASA TLX Raw Score (out of 100)	30.50	**17.23**

User Study

LLMs are not as accurate as humans

But not far off: LLM Assistant made only 3 errors out of 30

	LLM-Backend Assistant	Human Assistant
Accuracy (%)	90	100
Queries to Assistant per Instruction	9.39	3.2
Time per Instruction (sec)	38.02	27.86
Time to complete Task (sec)	197.32	151.15
NASA TLX Raw Score (out of 100)	30.50	17.23

User Study

The LLM makes significantly more queries for each instruction

It takes longer to recognize the user's action

	LLM-Backend Assistant	Human Assistant
Accuracy (%)	90	100
Queries to Assistant per Instruction	9.39	3.2
Time per Instruction (sec)	38.02	27.86
Time to complete Task (sec)	197.32	151.15
NASA TLX Raw Score (out of 100)	30.50	17.23

User Study

The human assistant takes longer to respond to each query when compared to the LLM

	LLM-Backend Assistant	Human Assistant
Accuracy (%)	90	100
Queries to Assistant per Instruction	9.39	3.2
Time per Instruction (sec)	38.02	27.86
Time to Respond to Each Query (sec)	4.04	8.71
Time to complete Task (sec)	197.32	151.15
NASA TLX Raw Score (out of 100)	30.50	17.23

User Study

The task load index is almost double for the LLM-based system

The mostly stems from needing multiple queries to complete a step

	LLM-Backend Assistant	Human Assistant
Accuracy (%)	90	100
Queries to Assistant per Instruction	9.39	3.2
Time per Instruction (sec)	38.02	27.86
Time to complete Task (sec)	197.32	151.15
NASA TLX Raw Score (out of 100)	30.50	17.23

XaiR:
An XR Platform that Integrates Large Language Models with the Physical World

- A platform that integrates the **physical world data** captured through **XR devices** with multimodal **LLMs**
- **Comparative study** on the effectiveness of MLLMs in supporting physical tasks against human benchmarks
- Open-sourced implementation: https://srutisrinidh

What did we cover today?

Mobile & Augmented/Mixed Reality

1- How can augmented reality headsets enable efficient and accurate RFID tag localization?

2- How does a human-in-the-loop design improve RFID sensing and localization?

3- What unexpected insights do user studies reveal?

4- How can multi-modal LLMs and XR be combined to create new applications?

5- What are the limitations of multi-modal LLMs in cognitive assistance for humans?

TODO:

1- Project Proposals due April 1st

2- Lab 4 is out today; due April 4

3- PSet 2 due April 10