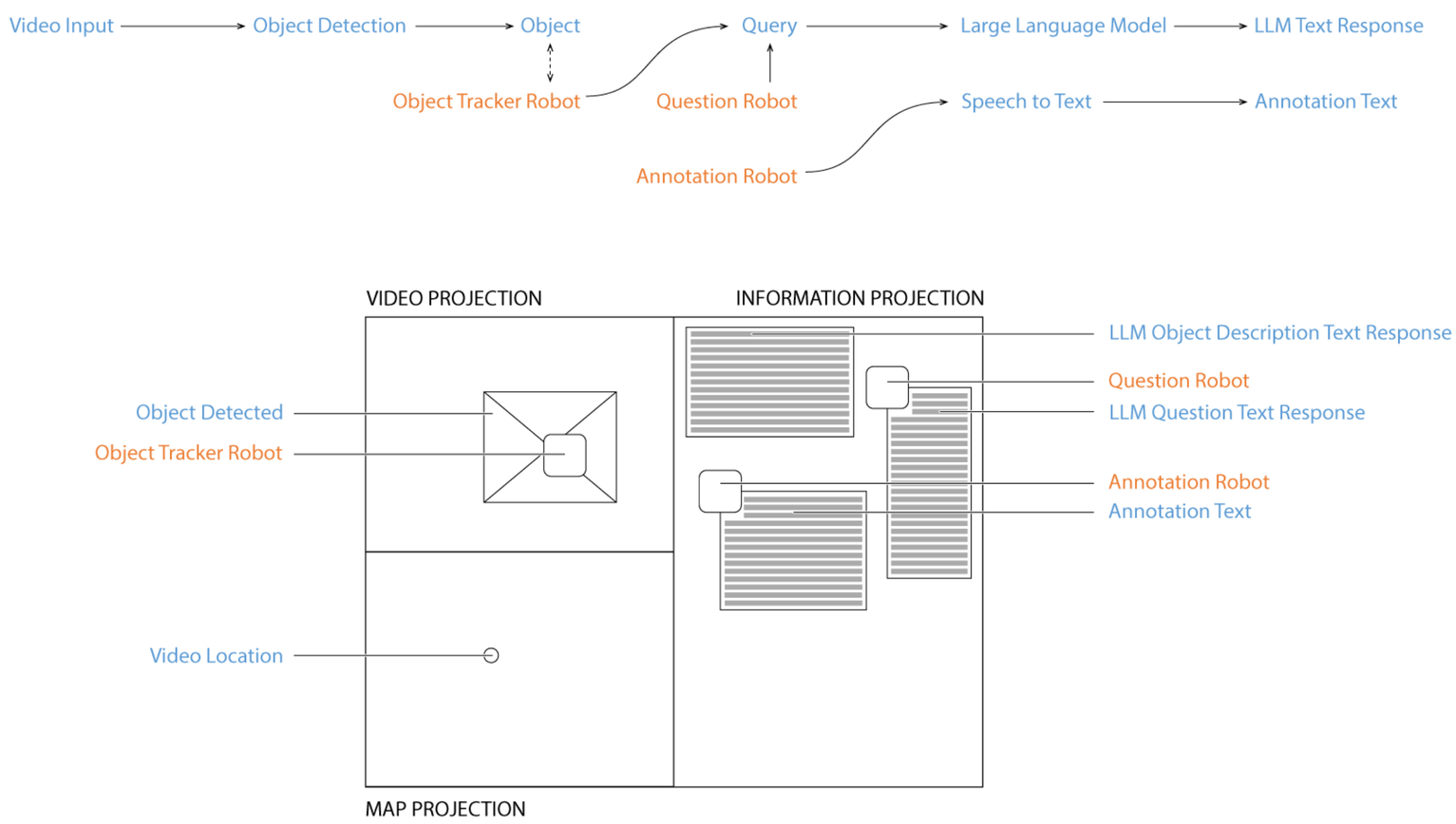
CENTER ON Pervasive Personalized Intelligence

Together We Go Farther

Tangible Interaction with Object Detection and Large Language Models

David Hunter, Professor Ellen Do
ATLAS, University of Colorado Boulder



Problem and Industrial Relevance

The use of generative AI such as Large Language Models (LLMs) have exploded in recent years, as well as Machine Learning across a range of tasks.
However, LLMs typically use text-based input, which does not work for or stimulate all users in all contexts. Text input keyboards don't lend themselves to in-situ collaboration.

Therefore… can tangible interaction be used to bridge interaction between data input and LLM output and create a novel method for information discovery?

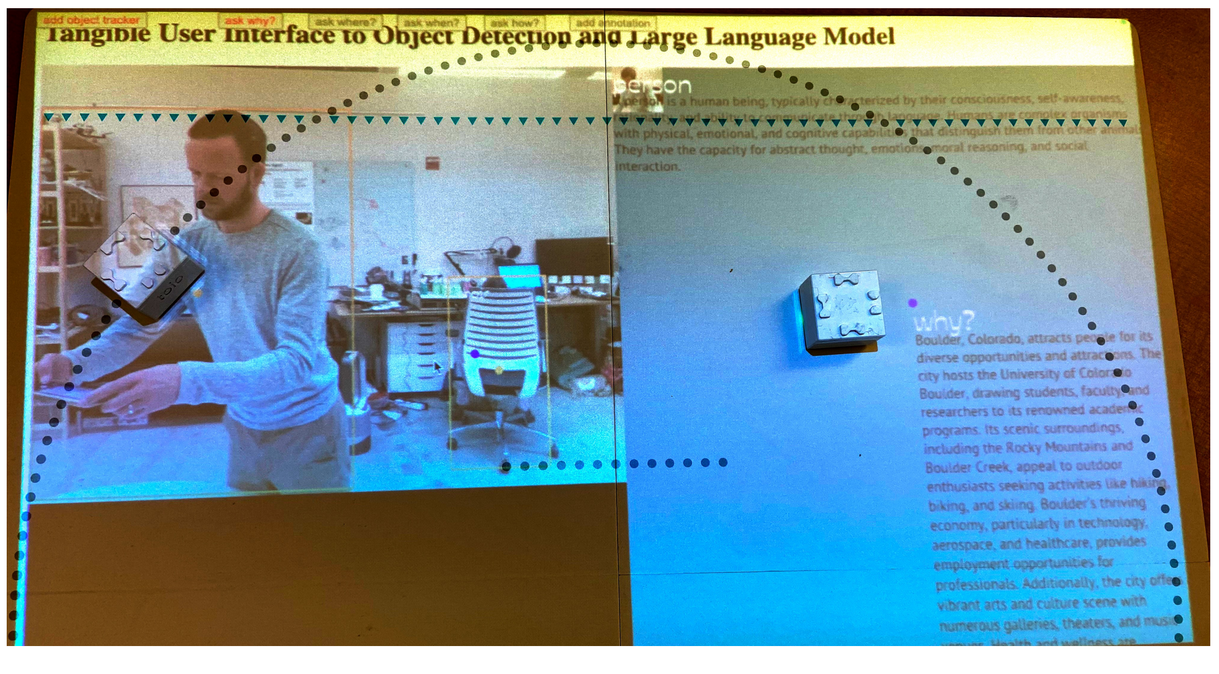
Answering this question:

- Opens ML and AI to other audiences or new contexts

- By combining different types of AI/ML models and interaction modalities we may synthesize new ways for users to discover and engage with information

Contributions and Applications

- **Advancing research on tangible interaction with AI**
- **Explore how AI and humans can collaborate**
- **System with range of applications:**
  - Public outreach for wildlife centers
  - Citizen Participation
  - Land management
  - First responder scenarios like wildfires and disaster relief



What Follows / Evaluation Plan

- Implement Object Permanence for reliable tracking
- Connect input data to LLM
- Add annotation function

Deploy prototype with small group of people and document their usage, cataloguing:

- what data they uncovered
- what LLM returned
- what users annotated
- when they used the system
- what positions and movements the interface makes

Executive Summary

1. A system that empowers users to tangibly interact with real time video, extract information using Object Detection and generate new information and query that information using a Large Language Model.

2. Robots enable both human input and system response, facilitating a tangible collaboration between human and AI and dynamic layout and adjustment of information.

3. Uses the Journalists 5 W's + 1 H: Who, What, Where, When, Why, How to interrogate Object Detection results.