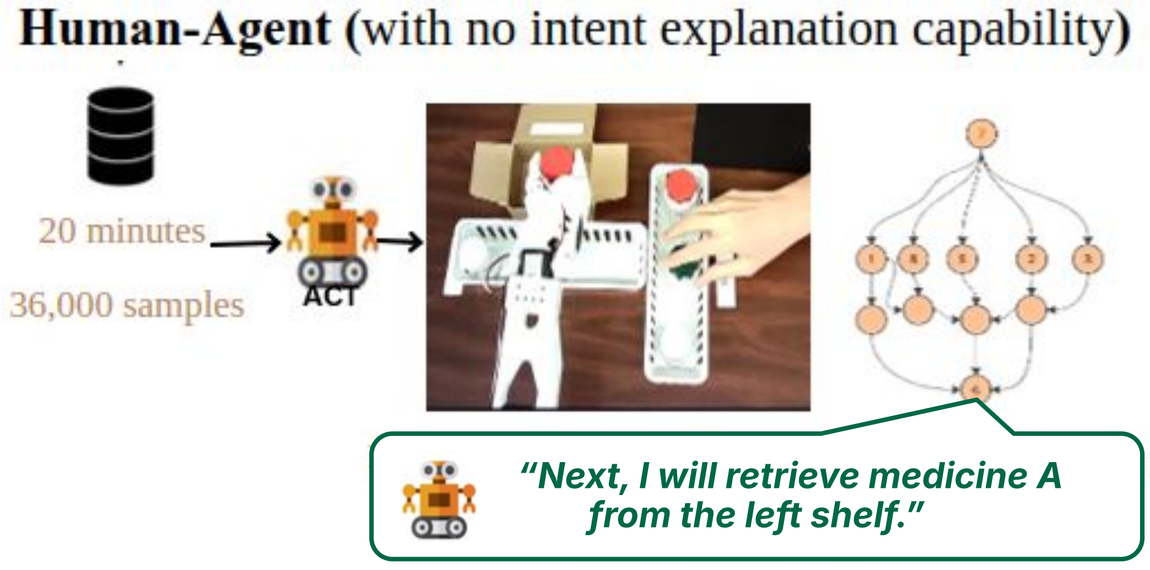
Towards Embodied Agent Intent Explanation in Human-Robot Collaboration: ACT Error Analysis and Solution Conceptualization

RARE LAB

UNIVERSITY of SOUTH FLORIDA

Amanuel Ergogo Zhao Han

RARE Lab, Bellini College of Artificial Intelligence, Cybersecurity and Computing, University of South Florida, USA



- **Learned policies are opaque**
- **Humans may struggle to predict robot actions**

Would a high-performing robot fail at teamwork?

We show that even high-performing robot policies (e.g., ACT) struggle to coordinate with human teammates as their the policies are opaque.

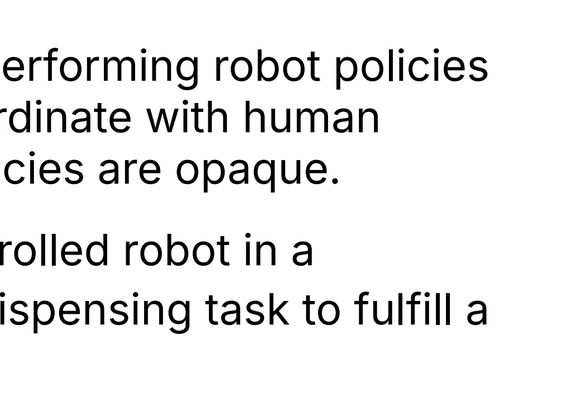
We evaluated an ACT-controlled robot in a collaborative medication-dispensing task to fulfill a 2-item prescription.

Specifically, we designed the task with 2 conditions:

1. Human-Human (baseline)
2. Human-Agent (ACT, no explanation)

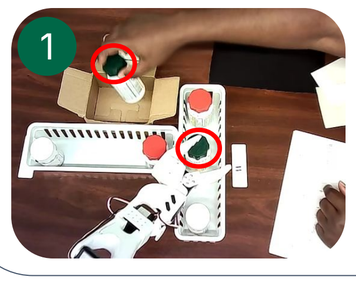
We measured task success rate, number of safety incidents, and completion time.

Teamwork Performance (15 matched trials)



Common Failures

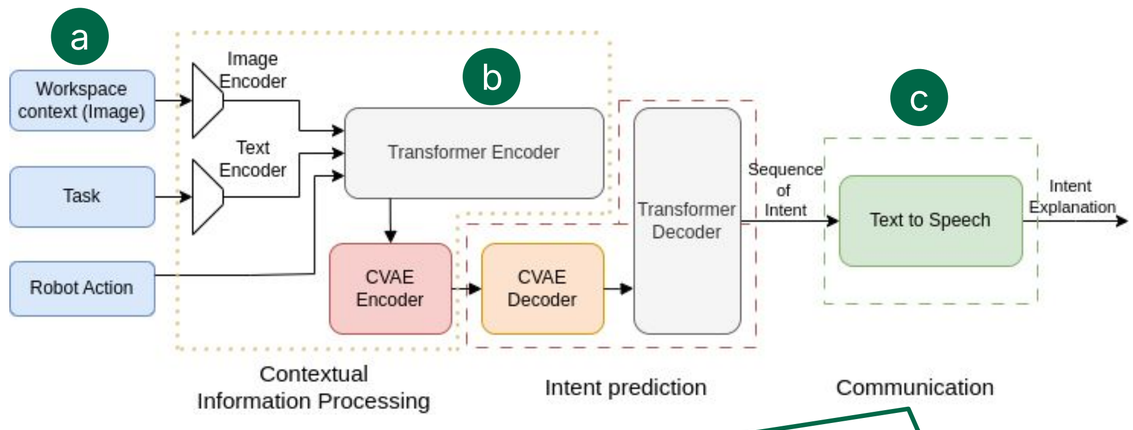
1. Redundant retrievals
2. Safety conflicts
3. Delays & hesitation



What if robots can explain their next move—on the fly?

We conceptualize a model-agnostic **Contextual Robot Intent Explanation (CRIE)** system. It

a. Encodes actions, context, goals & progression
b. Uses Transformer and CVAE to process contextual inputs into a latent representation of subtask intent and decode it into a symbolic subtask labels
c. Outputs natural language explanations via speech from the predicted subtask label



"Next, I will deliver the labeled order to the packing area."

Takeaways

- State-of-art robot policies limit coordination and safety during collaboration
- Transparent robot intent is essential for teamwork
- We hope CRIE will enable policy-agnostic intent explanations for fluent collaborations

RARE LAB Visit therarelab.com *or*

Amanuel Ergogo

Read the paper: