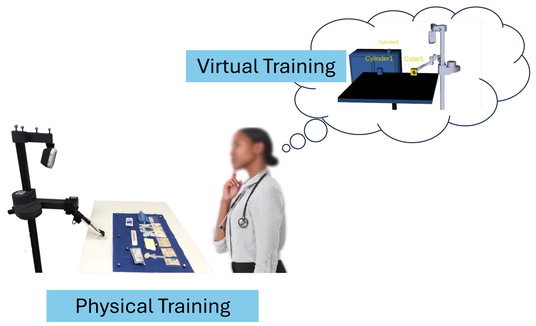
Interactively Explaining Robot Policies to Humans in Integrated Virtual and Physical Training Environments

RICE UNIVERSITY

Peizhu “Pam” Qian and Vaibhav Unhelkar. {pqian, unhelkar}@rice.edu
peizhuqian.com

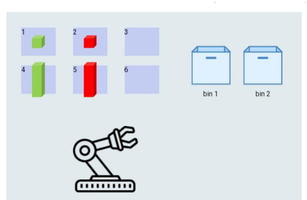
TL; DR: Virtual + Physical + Interactive Training = Better XAI for Real-World HRI

Motivating Example



A nurse wants to deploy a robot to gather medical supplies. During interaction with the physically embodied robot, she notices that the robot’s sensor sometimes makes mistakes. However, this was not reflected in the virtual training she has received.

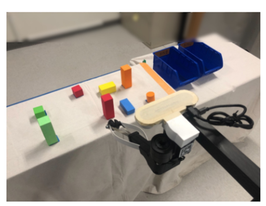
Limitations of Prior Work



Virtual Training:

✅ Provides explanations quickly

❌ Limited by fidelity of simulation environment

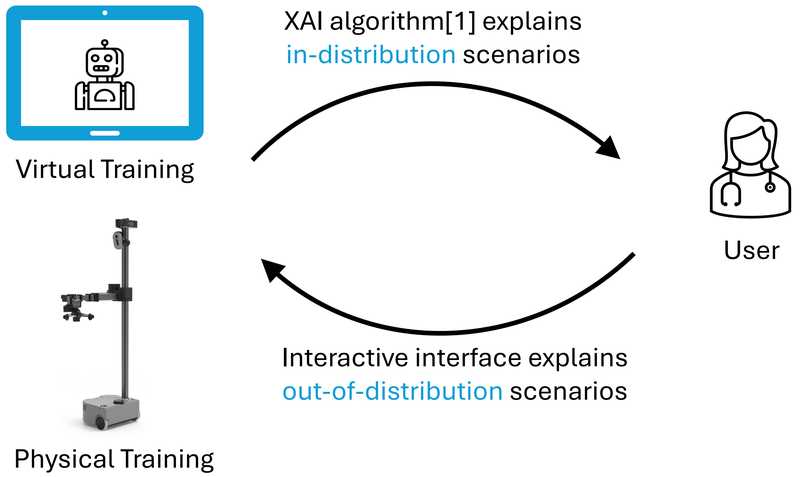


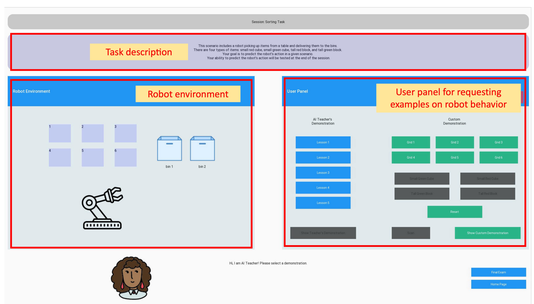
Physical Training:

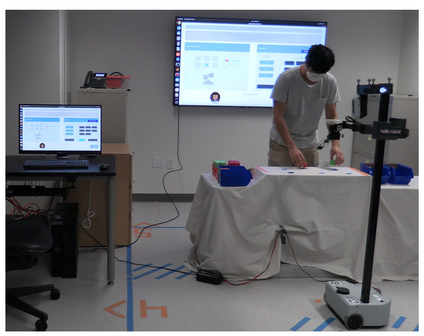
✅ Demonstrates limitations of physical system

❌ Resource- and time-intensive

Our Solution: Interactive XAI Using Combined Training Modality

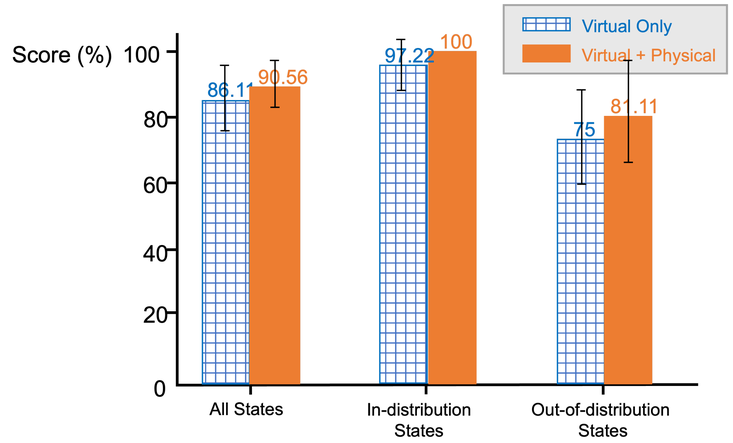


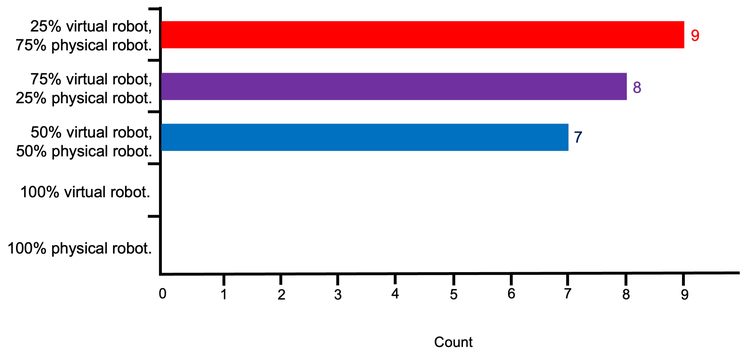




Our integrated interface allows users to switch between virtual and physical training in one place, interactively learning the robot's behavior.

Preliminary Results (N=24)





Participants prefer an integrated system and perform marginally better in predicting the robot’s out-of-distribution behavior when using virtual + physical training.

References

[1] Peizhu Qian and Vaibhav Unhelkar. Evaluating the Role of Interactivity on Improving Transparency in Autonomous Agents. (AAMAS’22).