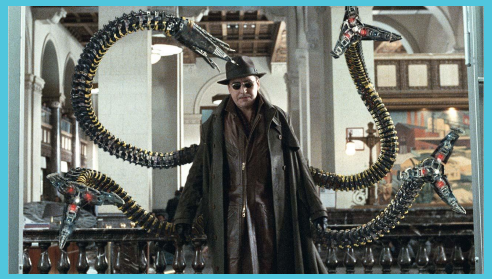
Harvard John A. Paulson
School of Engineering
and Applied Sciences

VLA-Controlled Supernumerary Arm

Alex and Charlie

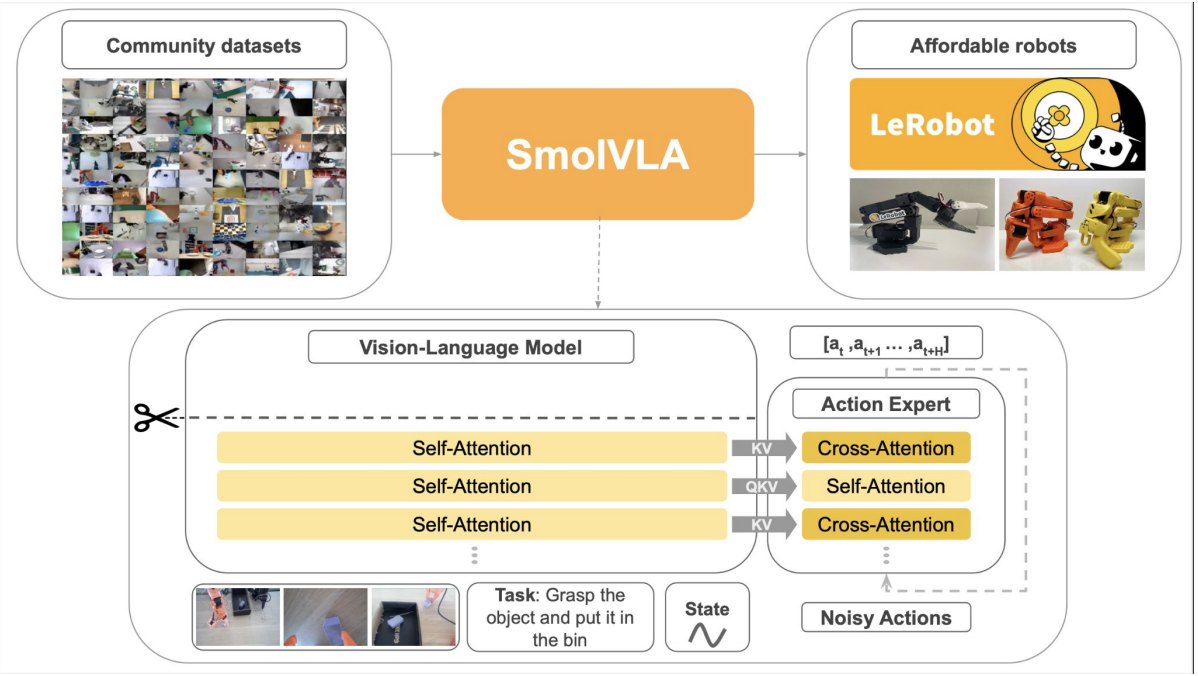


DISCLAIMER: Not an accurate representation. But one day…

This but a third one for Professor Slade

Vision-Language-Action Models (VLAs)

End-to-end **observation** → **inference** → **action** loop to perform desired tasks

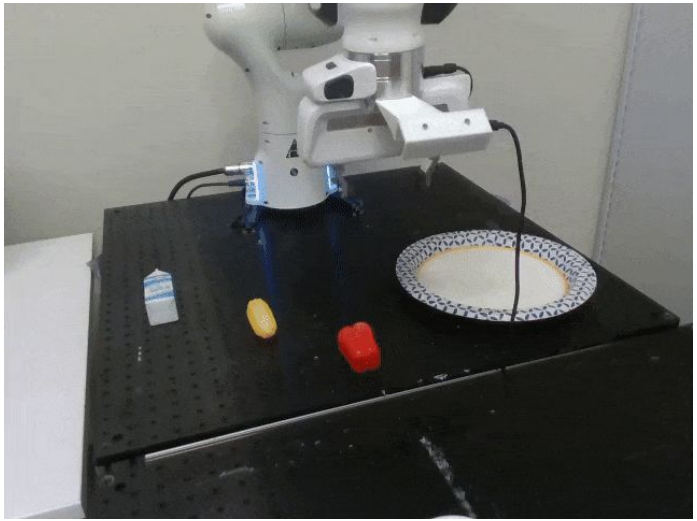


https://arxiv.org/abs/2506.01844v1

Limitations

1) Fixed-based robots and cameras
2) Controlled environments with minimal human-robot interaction

https://www.physicalintelligence.company/blog/pi0



https://openvla.github.io/

Motivation

Evaluate VLA performance on:

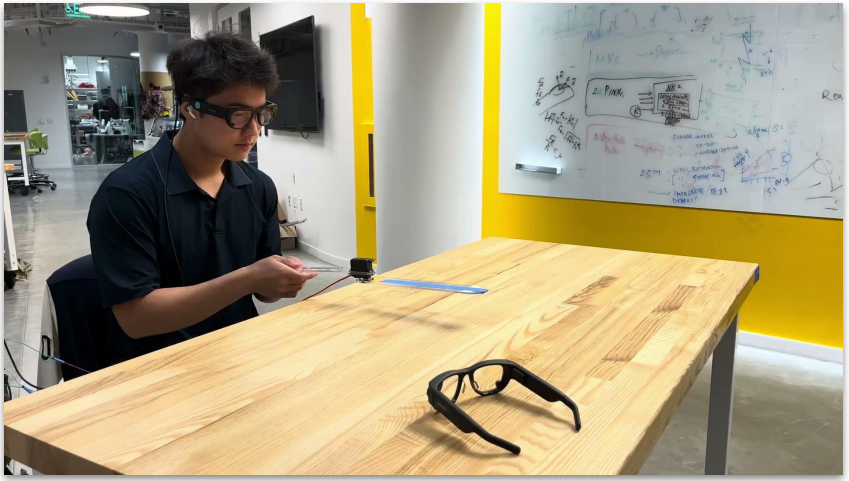
- Human-mounted supernumerary robotic limbs
- Tasks that require human-robot interaction (HRI)

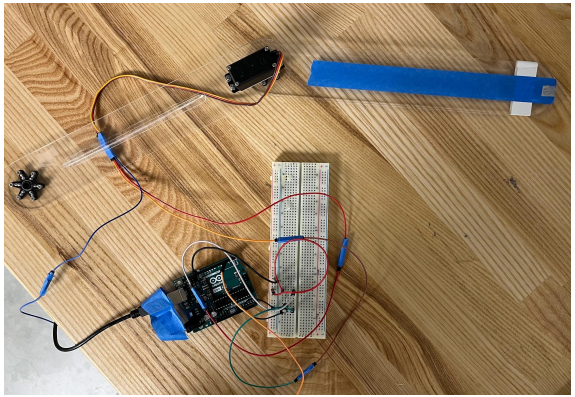
Tasks should assist with activities of daily living. Ex:

> “Pick up the glasses and hand them to me”

Custom 1-DOF Arm

- Proof-of-concept build
- End-to-end teleoperation recording and autonomous inference action loops
- Tabled due to library compatibility and feasibility issues :(



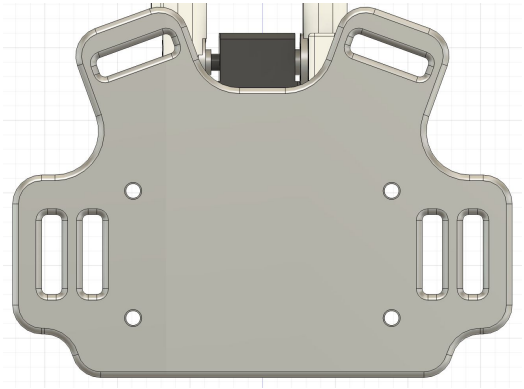


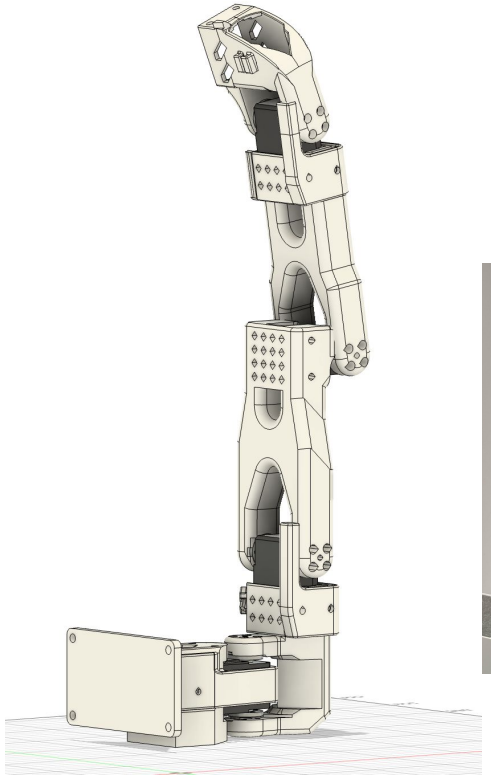
Wiring + Messy Circuitry

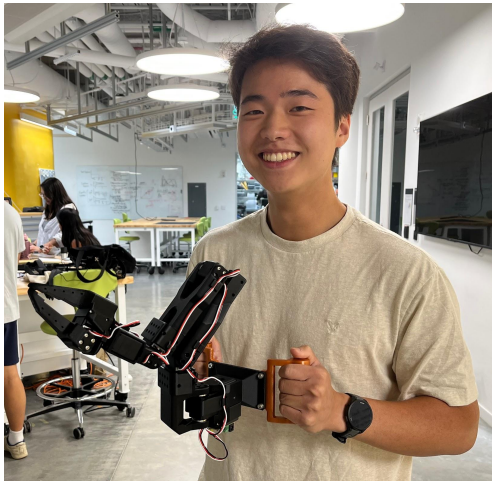
Human-mounted SO-101 Arm

Physical changes:

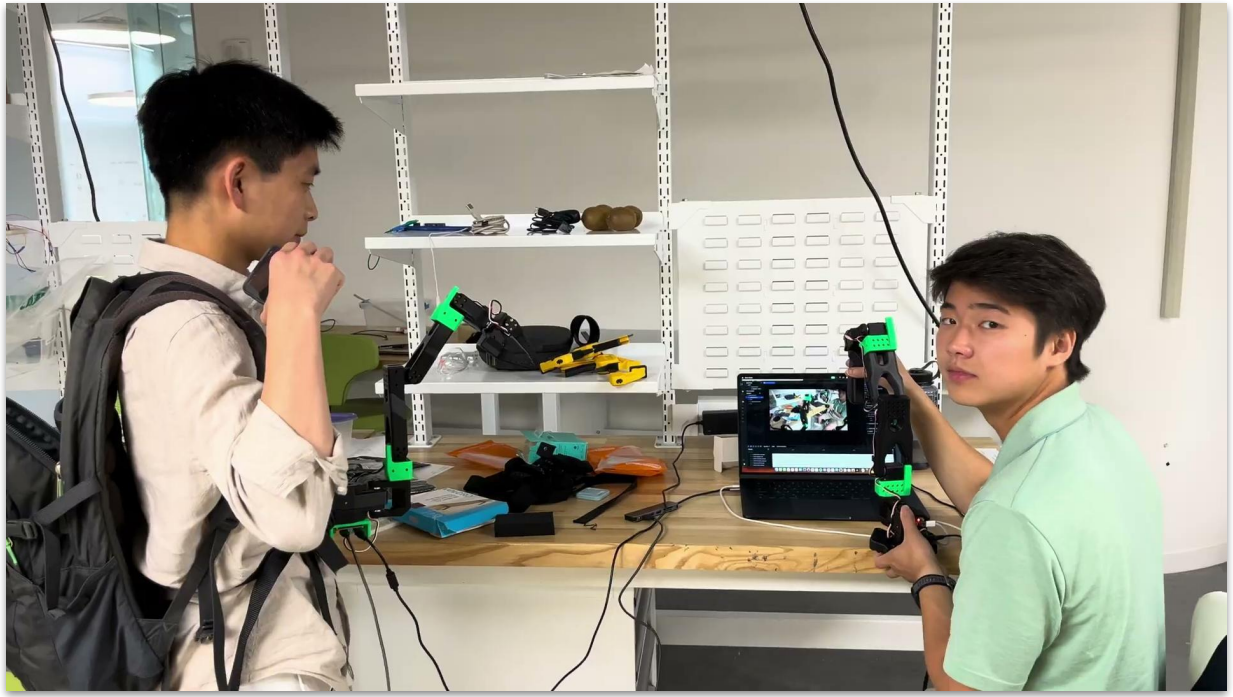
1) Handles
2) Mounting board + straps
3) Physical ROM stops





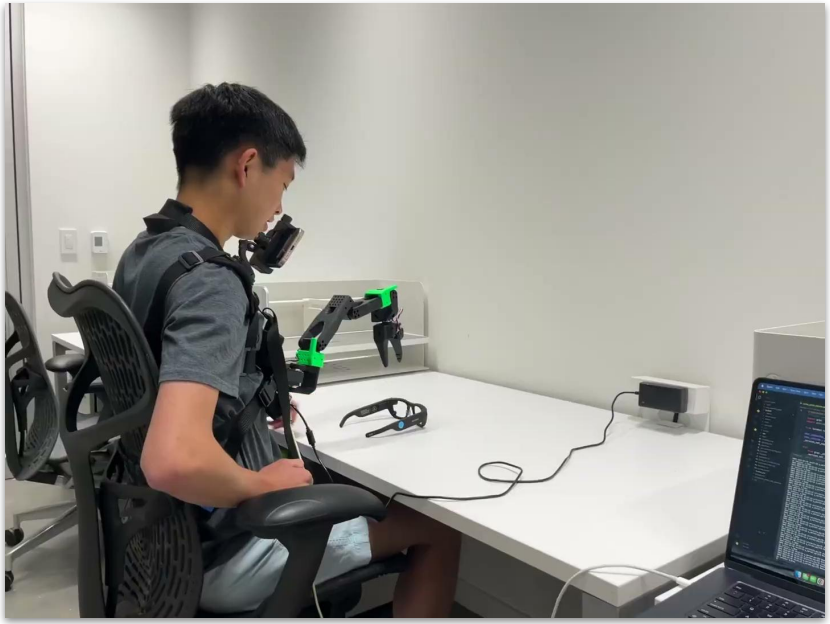


Teleoperation with Physical Stops



Single Shoulder Camera Setup

Poor quality + limited training data → very overfit initial model



- ➔ Only learned the movement pattern
- ➔ Lack of object recognition
- ➔ Failed if glasses moved at all
- ➔ Limited view of end-effector

Dual Camera Setup

Added wrist camera to include view of glasses and hand (and collected better training data)

- ➔ Consistently moved too far right
- ➔ Lack of depth perception
- ➔ Failed to pick up glasses

Triple Camera Setup

Added a third environment camera for depth perception

Next Steps

- Learn how to improve models performance
- Collect more data (on more tasks!)
- Tune hyperparameters more

Future Experimental Trials

Task categories:

- Basic non-HRI (partial credit via checkpoints)
- One-time interaction (partial credit via checkpoints)
- Continued interaction (time-based)

Variation dimensions:

- Human movement (measured with OpenCap)
- Workspace clutter
- Object orientation & identity
- Initial arm position

What we Learned!

Charlie:

- How to read research papers!
- Very simple CAD/3D printing
- Engineering prototyping process
- How to think about work through a research/experiment design context

Alex:

- Research project design and conception process
- Basic coding (and how to leverage LLMs for support)
- Working independently and managing a project

Thank you!

Week by Week Layout:

1. Read papers, created project proposal
2. First meeting with slade, investigated Project Aria code base
3. Built 1 DOF servo system
4. Debug pyvirtualcam and other issues
5. Wrote custom scripts for data collection, collected data, first fine tuning of smolvla
6. Get FASRC to work, retrained smolvla with new data, met with patrick and decided to pivot to SO-101
7. Adapted SO-101 to mobile form, researched ADLs, and thought about basic study params
8. Finalize safety measures, try out different camera setups, make QoL changes with arm, collect initial data
9. Test arm for first time, collect better/more data, retrain
10. Final week clean things up with the project, prepare for final presentation, do small amounts of testing