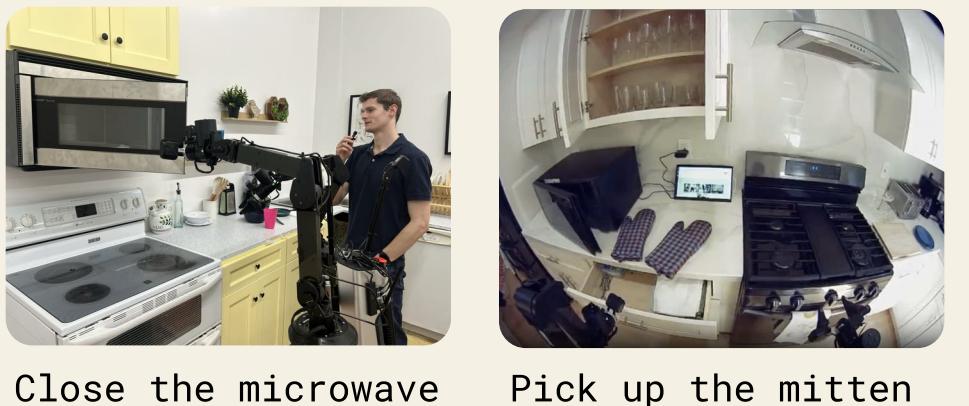


Multimodal Data

Subtask Commands



Object Detection



Multimodal Web Data



Q: How many desks are in the image?
A: 12

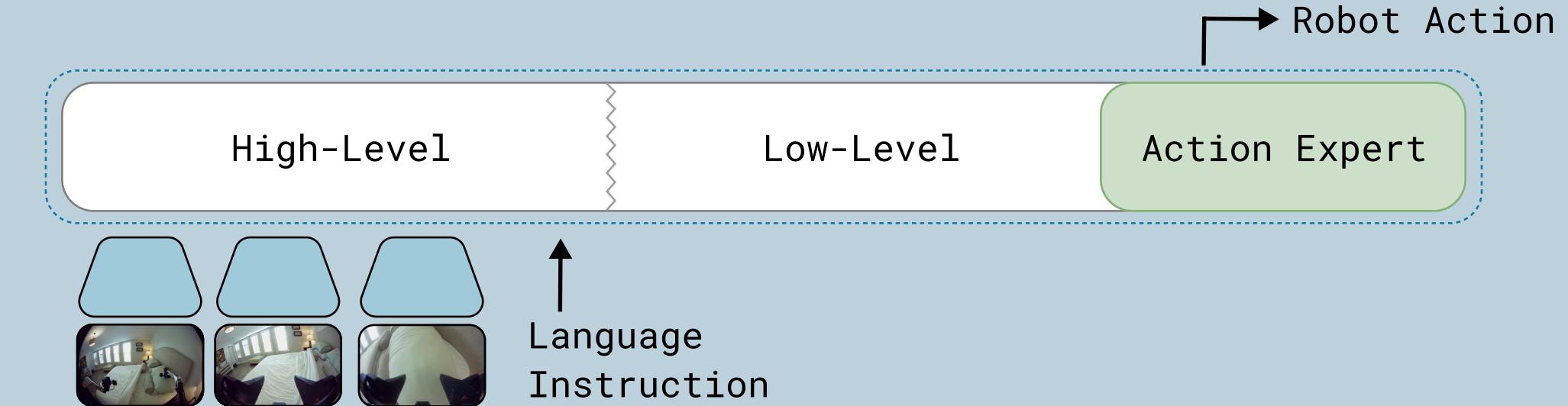


Q: Detect and label all objects in the scene.
A: <loc0112> <loc0234>...

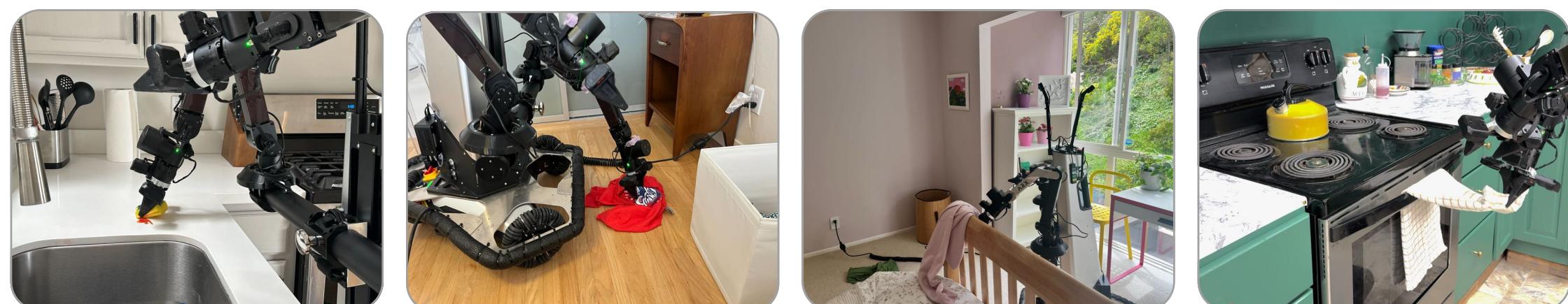
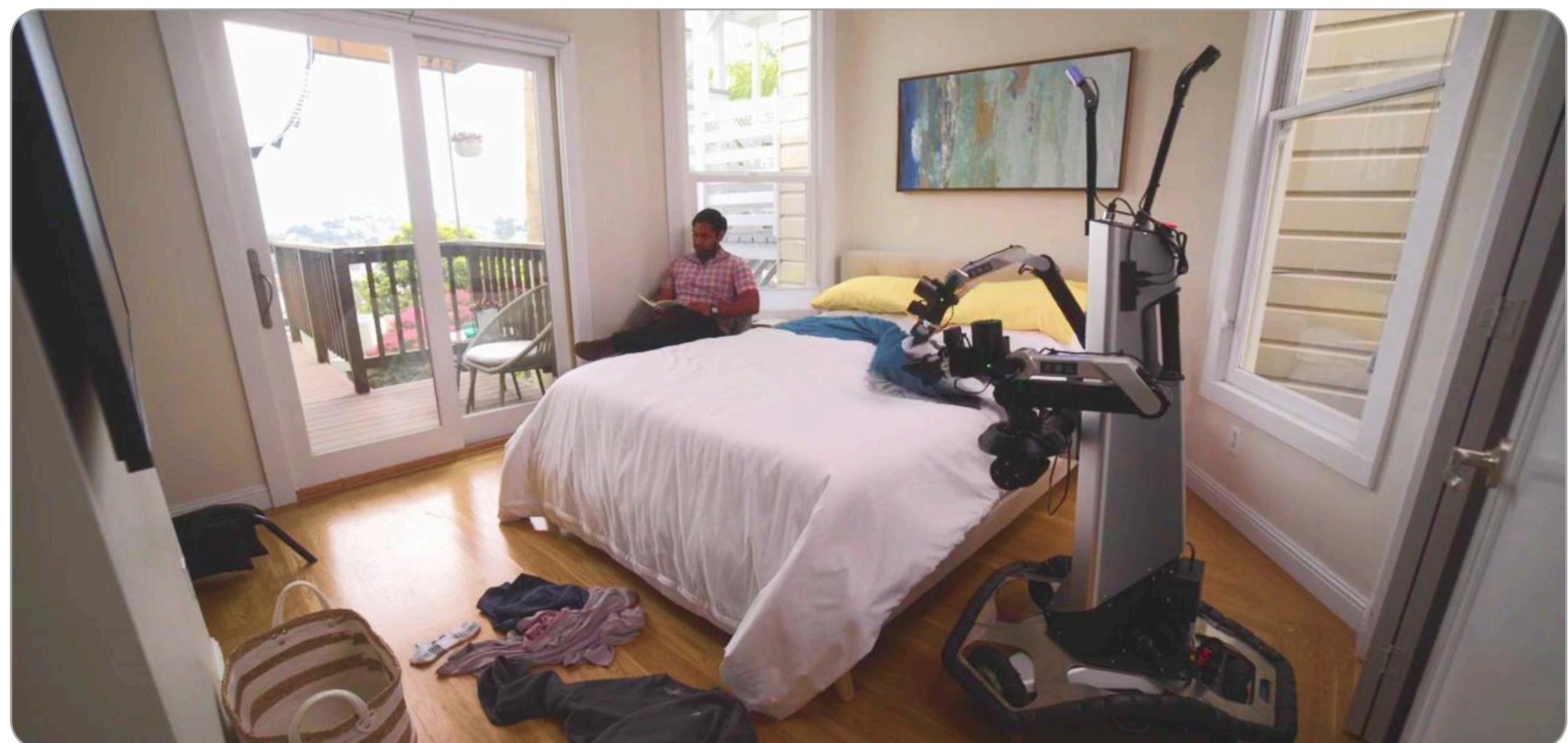


Q: What kind of pie is on this plate?
A: Chocolate

$\pi_{0.5}$ Vision-Language-Action Policy



Deploy out-of-the-box in new homes



Robot Action Data

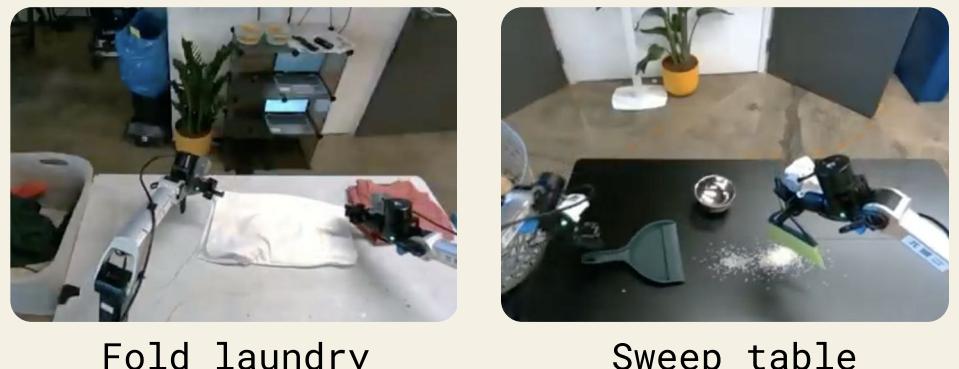
In-the-wild Mobile Robot



In-the-wild Static Robot



In-Lab Static Robot



General Robot Data

