

# Humanoid Loco-Manipulation from Human Interaction

Sirui Xu

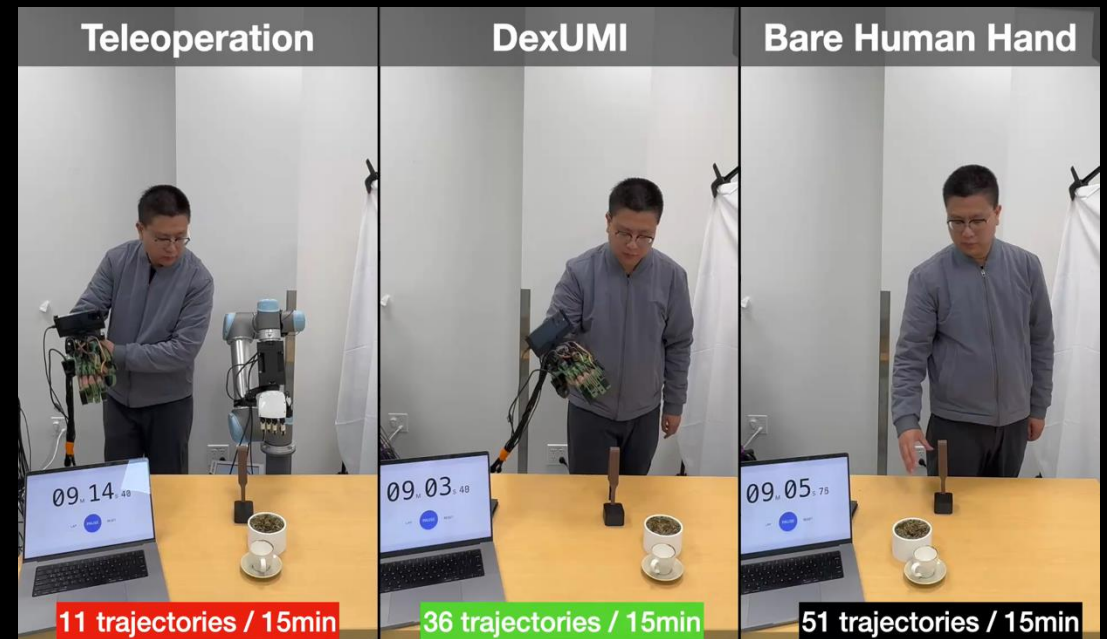


# Real bottleneck

- Robot learning remains data-hungry
  - Needs diverse interactions across objects, scenes, and tasks
  - Real-world robot collection is expensive, slow, and embodiment-specific



Teleoperation



DexUMI

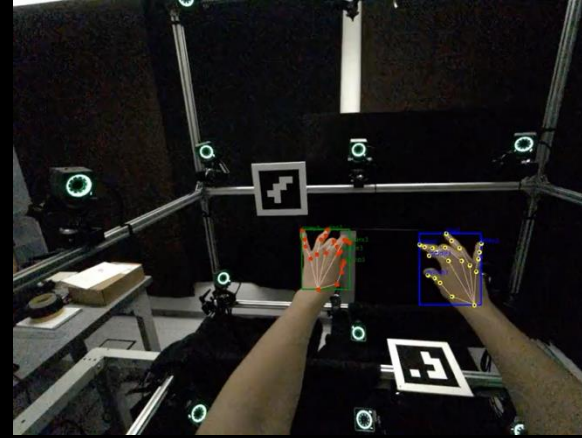
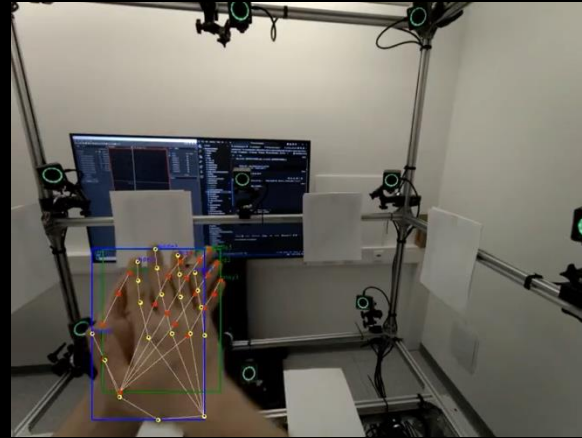
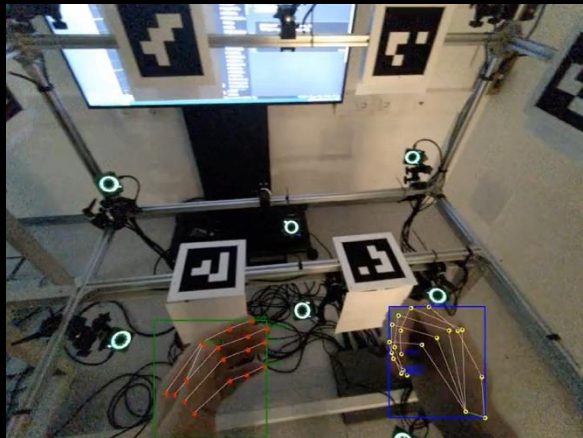
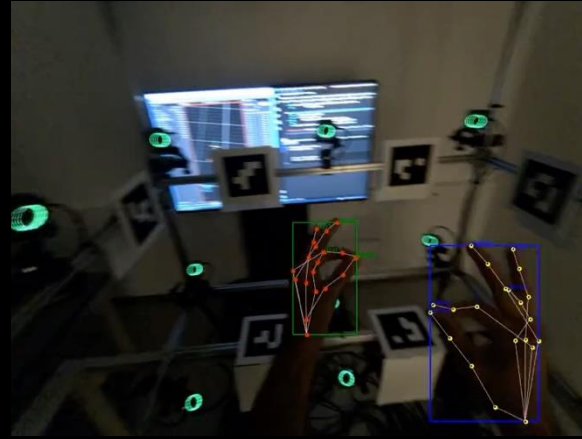
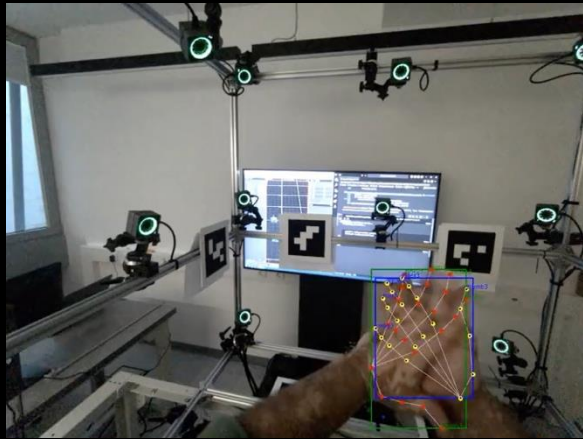
Yanjie Ze, et al. Twist2: Scalable, Portable, and Holistic Humanoid Data Collection System. ICRA 2026.

Mengda Xu, et al. DexUMI: Using Human Hand as the Universal Manipulation Interface for Dexterous Manipulation. CoRL 2025.

# Human interaction data from MoCap



June 5, 10:45  
AM – 12:45 PM  
Poster #210

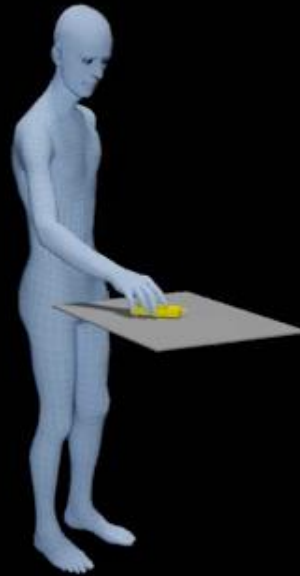


Zimu Zhang\*, Yucheng Zhang\*, Xiyan Xu, Ziyin Wang, **Sirui Xu**, Kai Zhou, Bing Zhou, Chuan Guo, Jian Wang, Yu-Xiong Wang\*, Liang-Yan Gui\*. **HandX**: Scaling Bimanual Motion and Interaction Generation. **CVPR 2026**

# MoCap data is fragmented and limited in scale



Full-body Interaction



Dexterous Manipulation

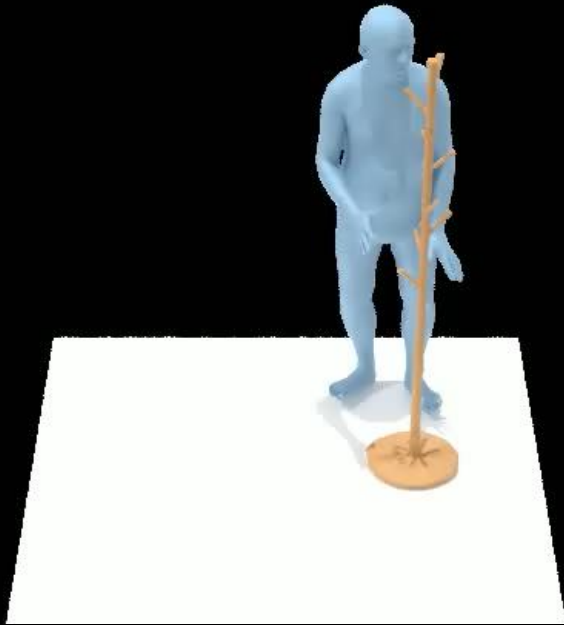


Scene Interaction

Omid Taheri, et al. GRAB: A Dataset of Whole-Body Human Grasping of Objects. ECCV 2020.  
Jiaman Li, et al. Object Motion Guided Human Motion Synthesis. SIGGRAPH Asia 2023.  
Jiaxin Lu, et al. HUMOTO: A 4D Dataset of Mocap Human Object Interactions. ICCV 2025.

# Raw interaction data is not automatically physical

- Mocap noise create penetration, jitter, and floating contact



Imprecise hand poses



Floating & Penetration



Jittering

Bharat Lal Bhatnagar, et al. Behave: Dataset and method for tracking human object interactions. CVPR 2022.

Jiaman Li, et al. Object Motion Guided Human Motion Synthesis. SIGGRAPH Asia 2023.

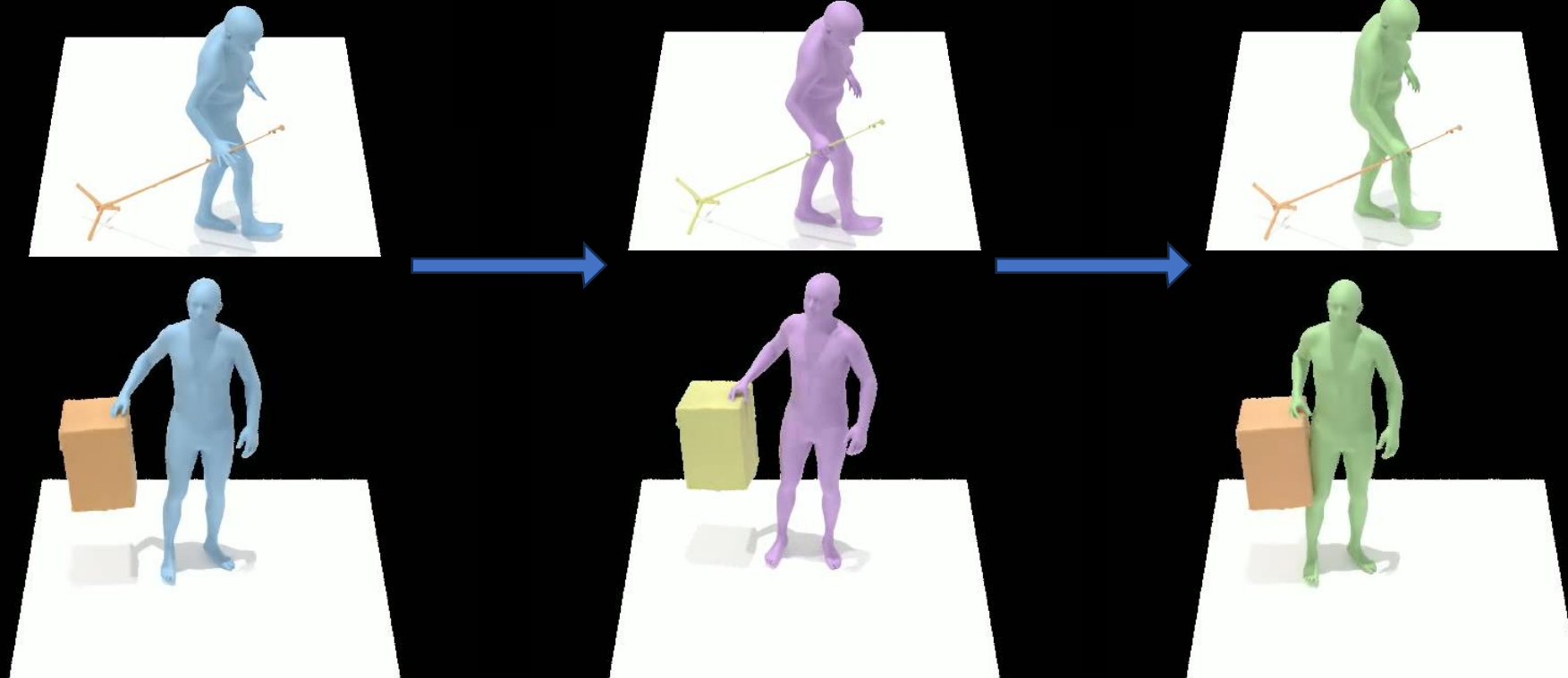
# A scalable data engine



Raw Data

Correction

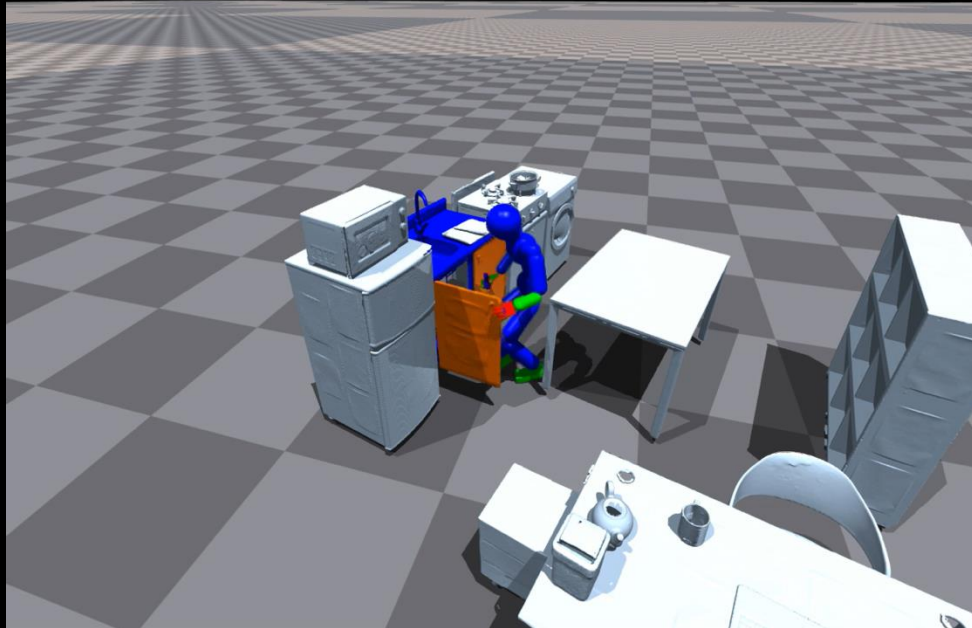
Augmentation



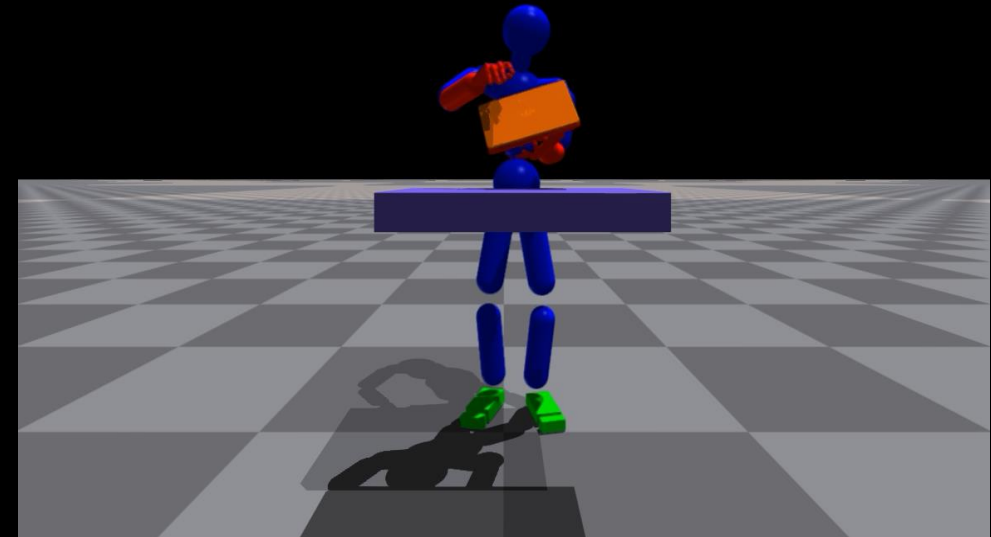
**Sirui Xu\***, Dongting Li\*, Yucheng Zhang\*, Xiyan Xu\*, Qi Long\*, Ziyin Wang\*, et al. InterAct: Advancing Large-Scale Versatile 3D Human-Object Interaction Generation. CVPR 2025

# Go beyond single rigid objects

Interaction with scene



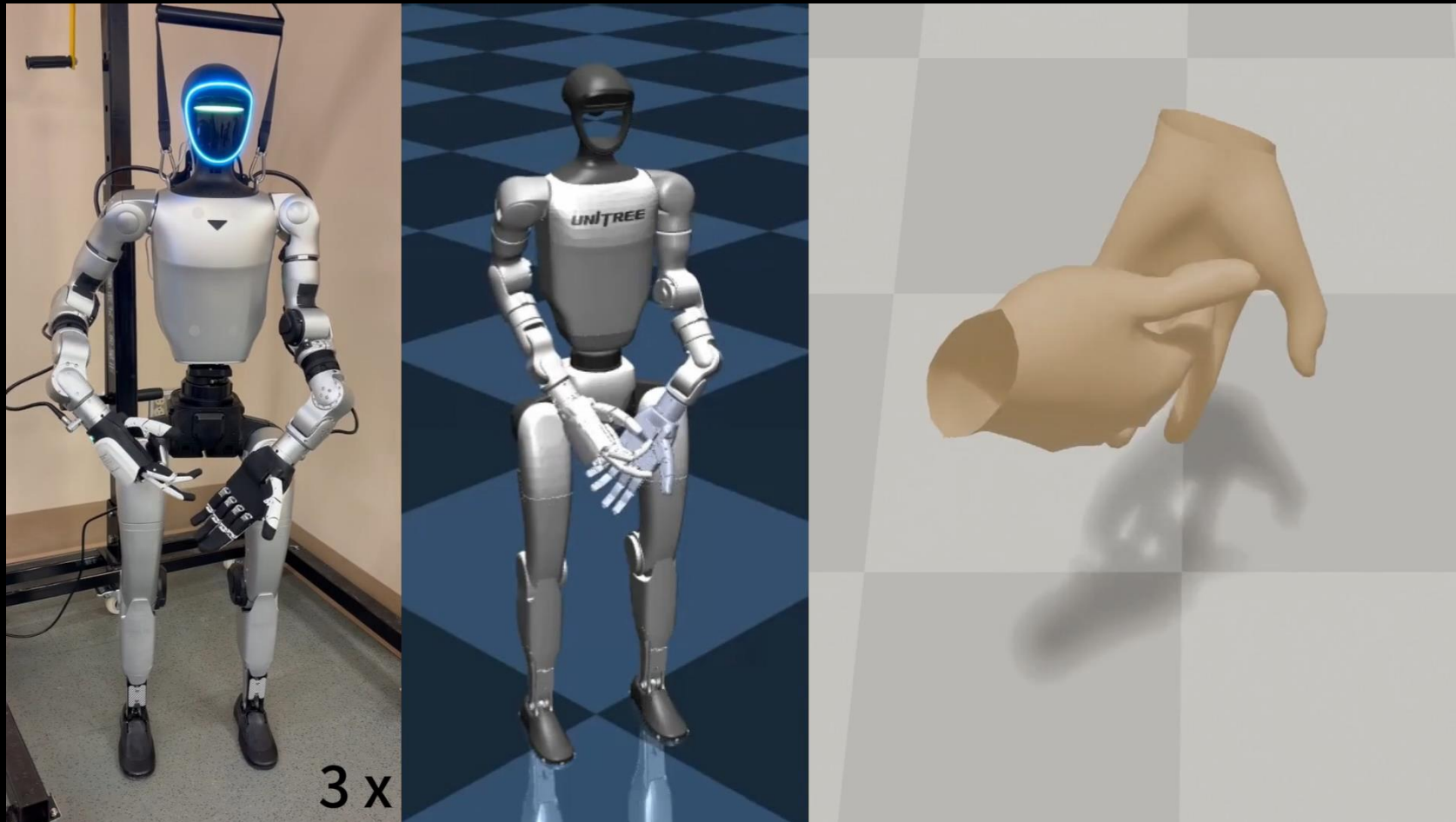
Interaction with articulated object



Jeonghwan Kim, et al. Parahome: Parameterizing Everyday Home Activities Towards 3D Generative Modeling of Human-Object Interactions. CVPR 2025.

Zicong Fan, et al. ARCTIC: A Dataset for Dexterous Bimanual Hand-Object Manipulation. CVPR 2023.

# HandX scales data with dexterity



June 5, 10:45  
AM – 12:45 PM  
Poster #210

Zimu Zhang\*, Yucheng Zhang\*, Xiyan Xu, Ziyin Wang, **Sirui Xu**, Kai Zhou, Bing Zhou, Chuan Guo, Jian Wang, Yu-Xiong Wang\*, Liang-Yan Gui\*. **HandX: Scaling Bimanual Motion and Interaction Generation. CVPR 2026**

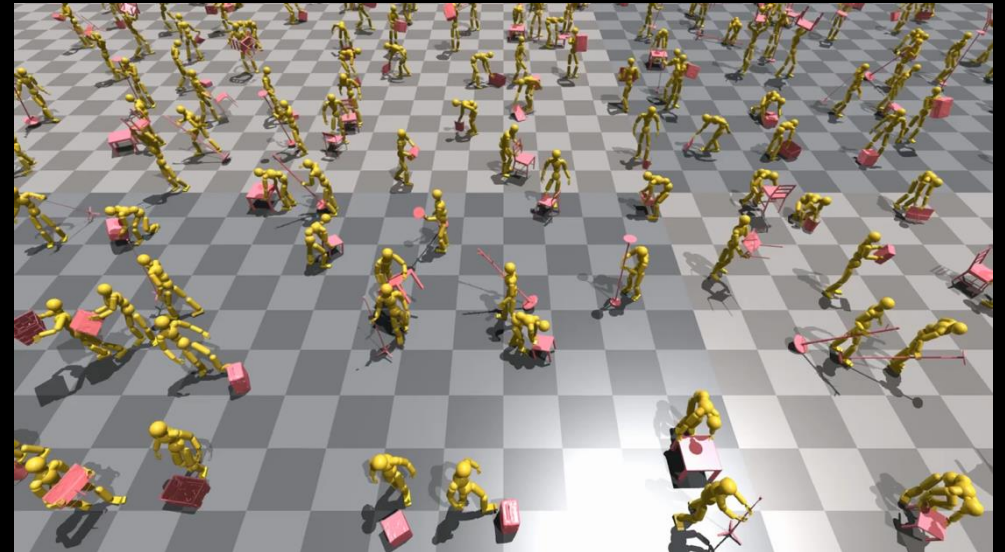
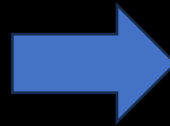
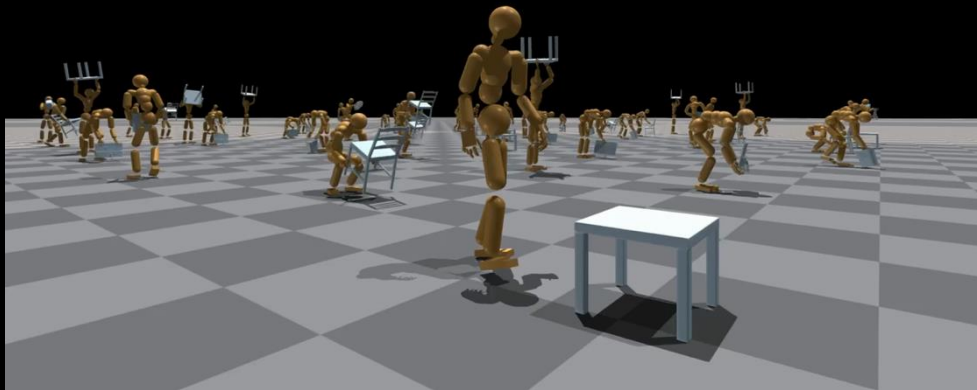
# MoCap is a bridge, but not robot data

- No action signal
- Not embodiment-aware

# Translate MoCap into actions

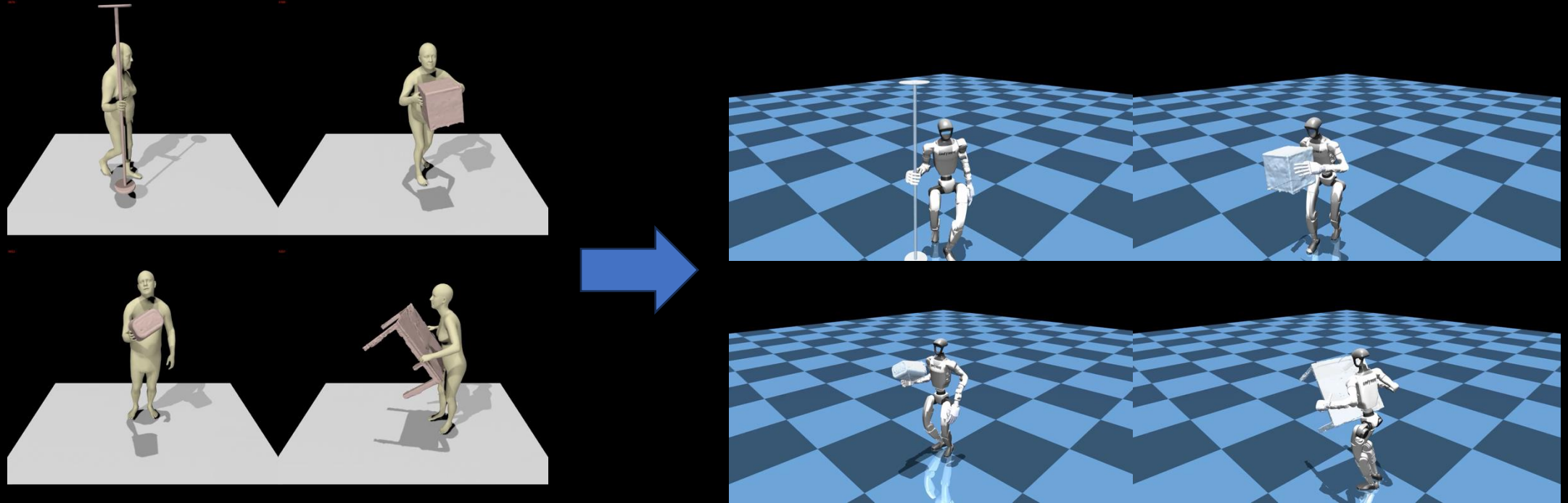


- Large-scale co-tracking of human and object dynamics in simulation



**Sirui Xu**, Hung Yu Ling, Yu-Xiong Wang\*, Liang-Yan Gui\*. InterMimic: Towards Universal Whole-Body Control for Physics-Based Human-Object Interactions. **CVPR 2025 Highlight**

# Translate MoCap across embodiments

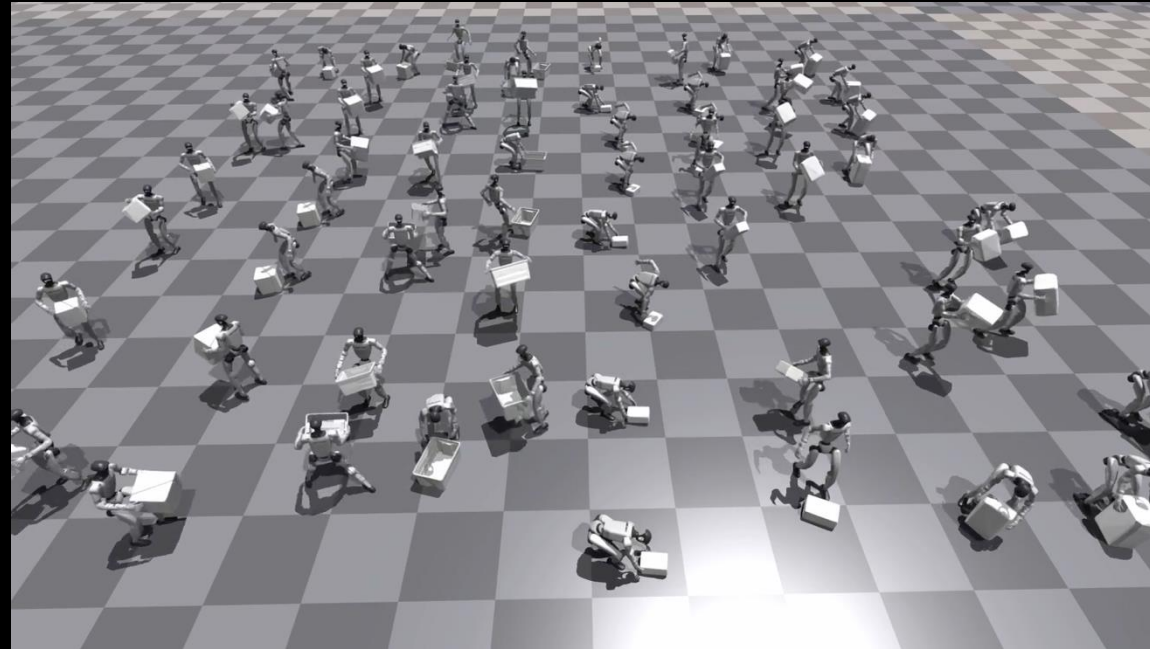


Sirui Xu\*, Dongting Li\*, Yucheng Zhang\*, Xiyan Xu\*, Qi Long\*, Ziyin Wang\*, et al. InterAct: Advancing Large-Scale Versatile 3D Human-Object Interaction Generation. CVPR 2025

# Formulate retargeting as motion tracking



- Retargeting solves a constrained optimization problem
  - Inverse kinematics
  - Motion tracking under forward dynamics (physics simulation)

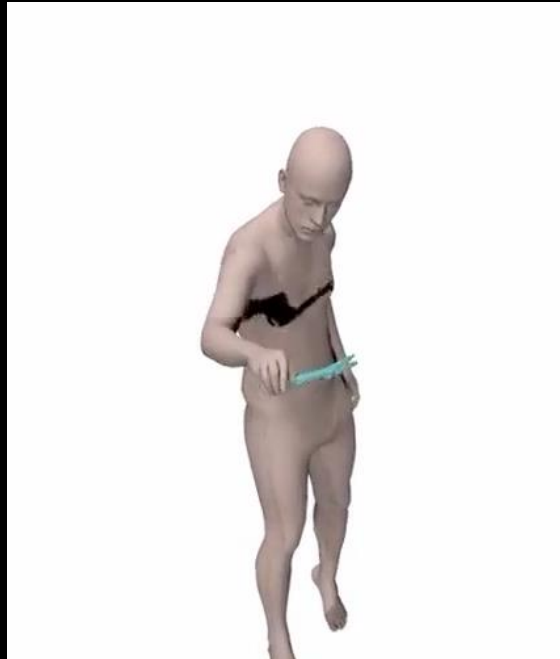


**Sirui Xu**, Hung Yu Ling, Yu-Xiong Wang\*, Liang-Yan Gui\*. InterMimic: Towards Universal Whole-Body Control for Physics-Based Human-Object Interactions. **CVPR 2025 Highlight**

# Neural retargeting is task- and physically aware



Raw MoCap



Kinematics-based retargeting

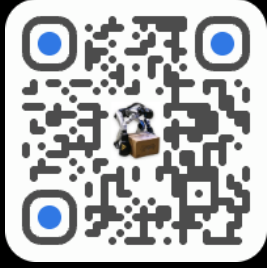


Physics-based retargeting  
(Ours)

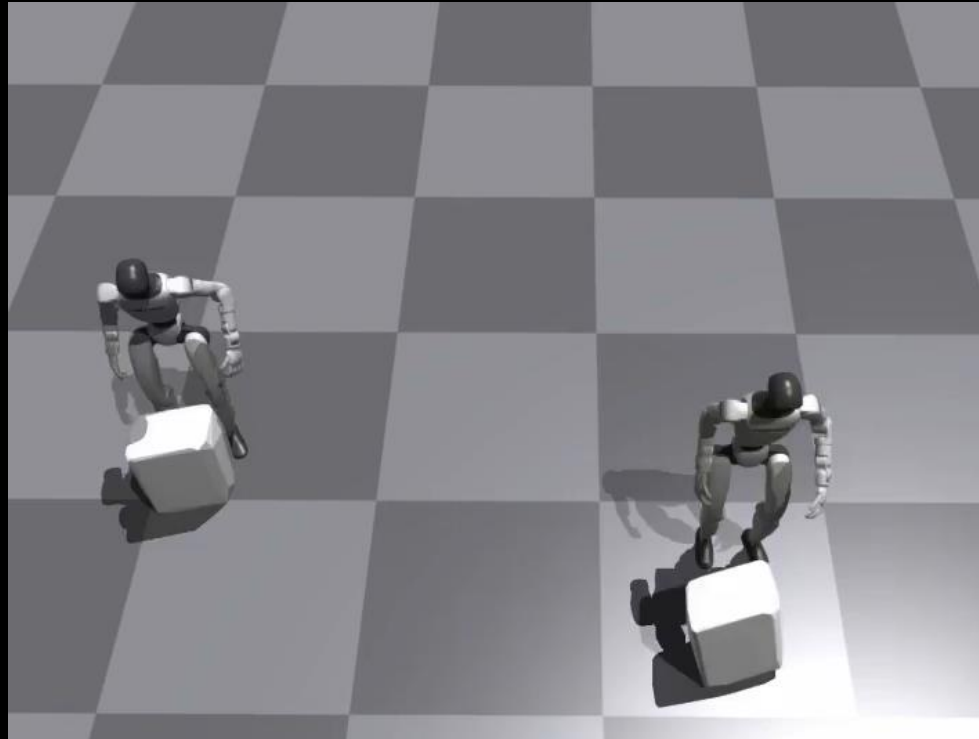


**Sirui Xu**, Yu-Wei Chao, Liuyu Bian, Arsalan Mousavian, Yu-Xiong Wang\*, Liang-Yan Gui\*, Wei Yang\*. **Dexplore: Scalable Neural Control for Dexterous Manipulation from Reference-Scoped Exploration. CoRL 2025**

# Neural retargeting scales to unlimited robot data



Same human reference; Different rollout trajectories

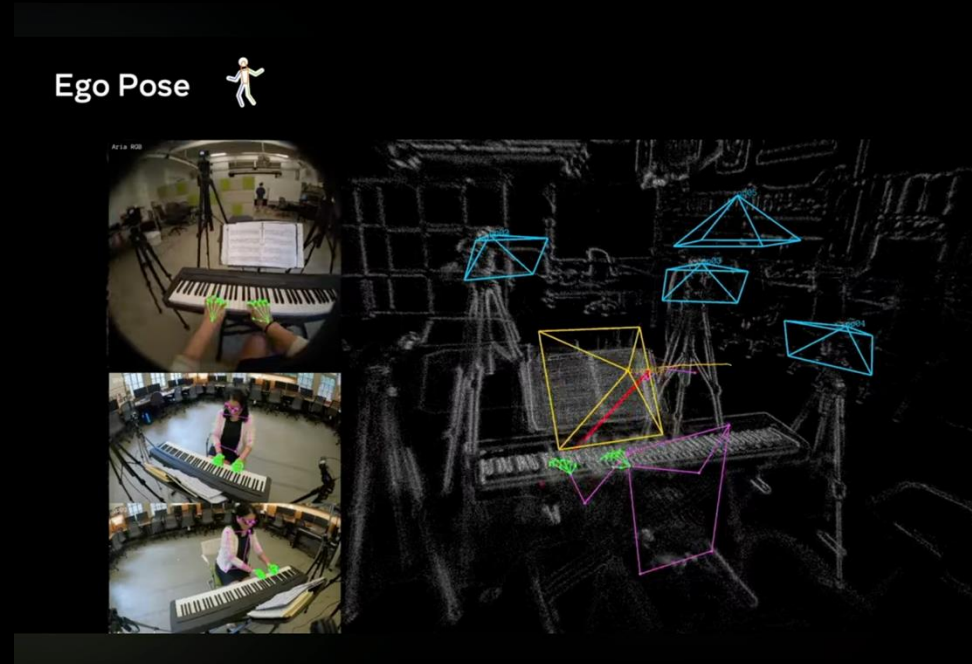


Augmentation  
from InterAct



Xialin He\*, **Sirui Xu\***, Xinyao Li, Runpei Dong, Liuyu Bian, Yu-Xiong Wang\*, Liang-Yan Gui\*. ULTRA: Unified Multimodal Control for Autonomous Humanoid Whole-Body Loco-Manipulation. ArXiv 2026

# Humans plan sparsely; robots should too

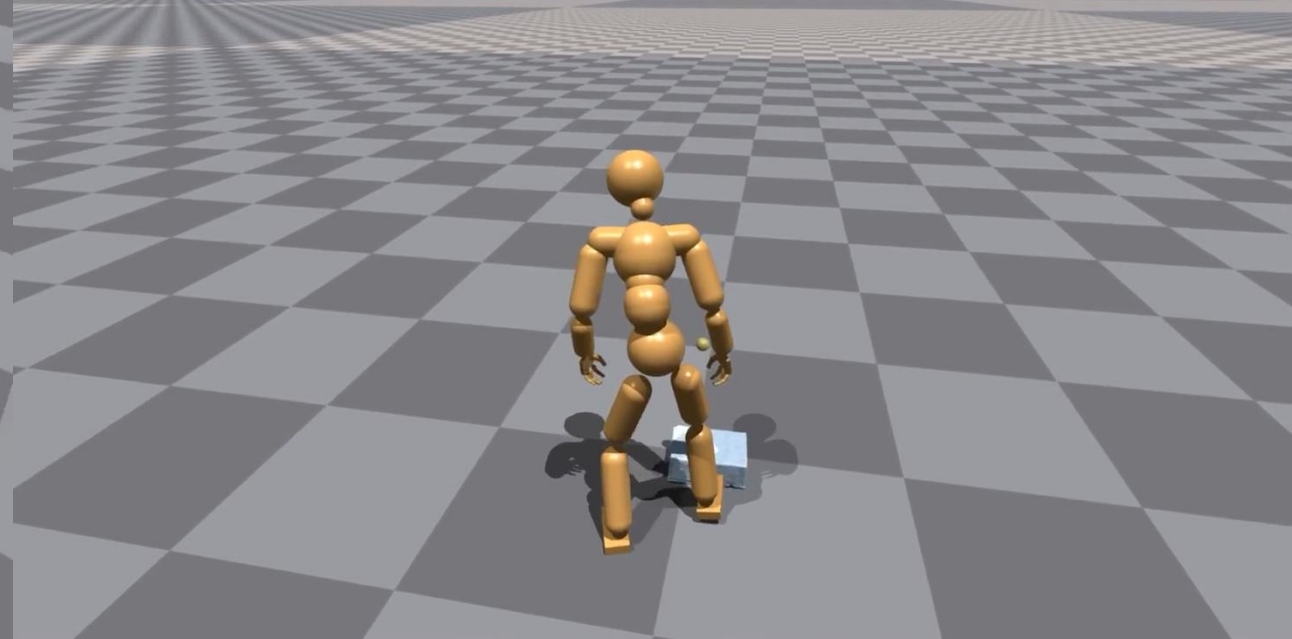
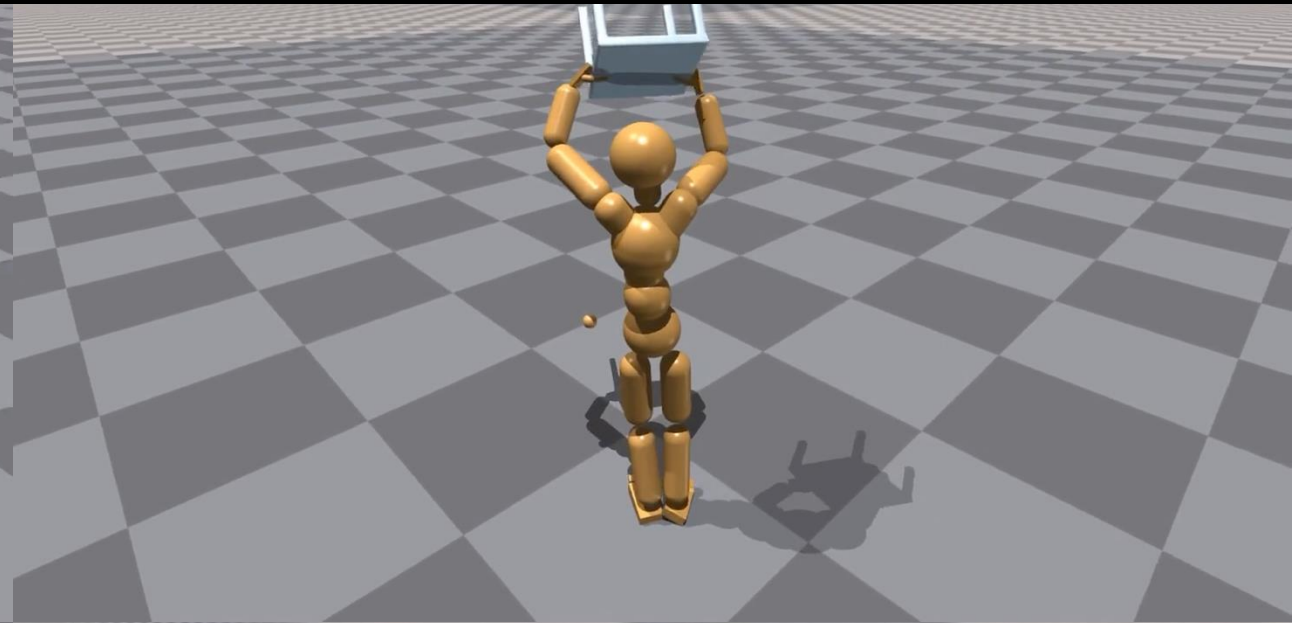
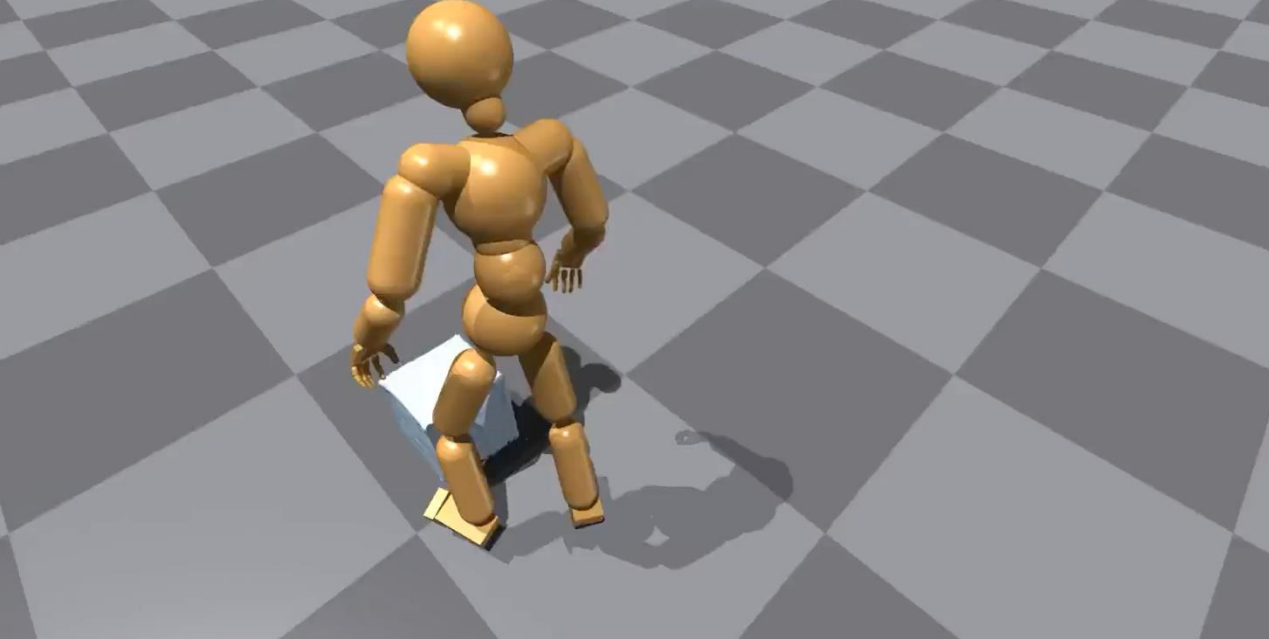
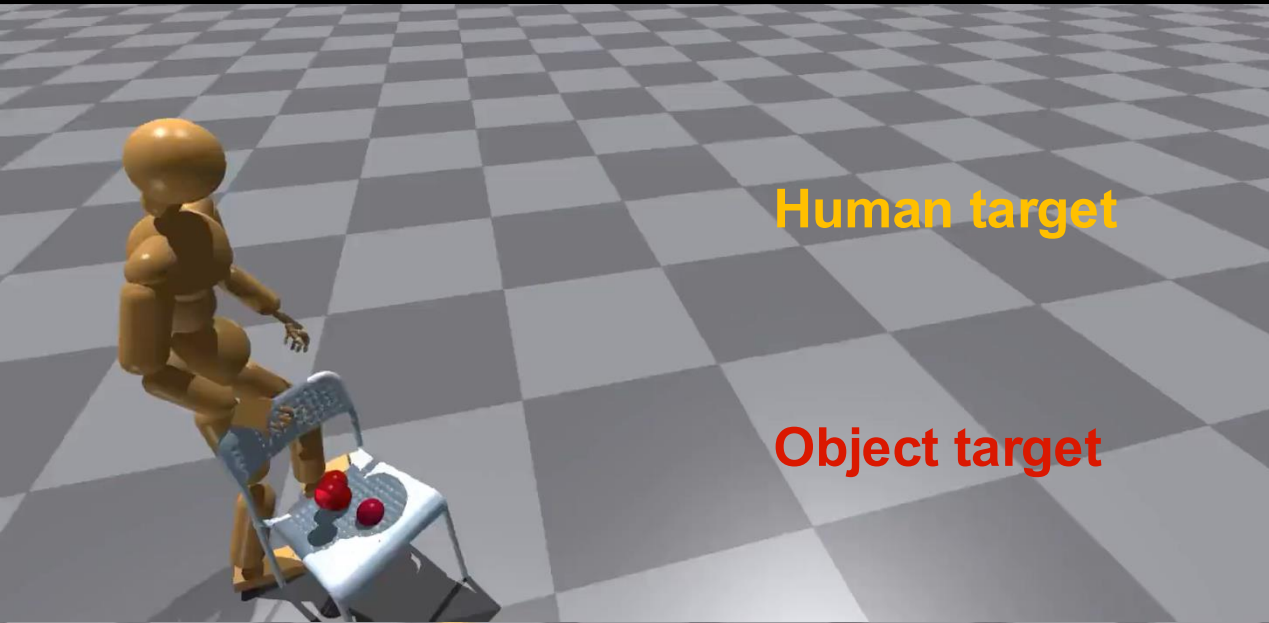


Plan from **sparse intentions**; coordination and contact emerge through **fast, intuitive motor response**

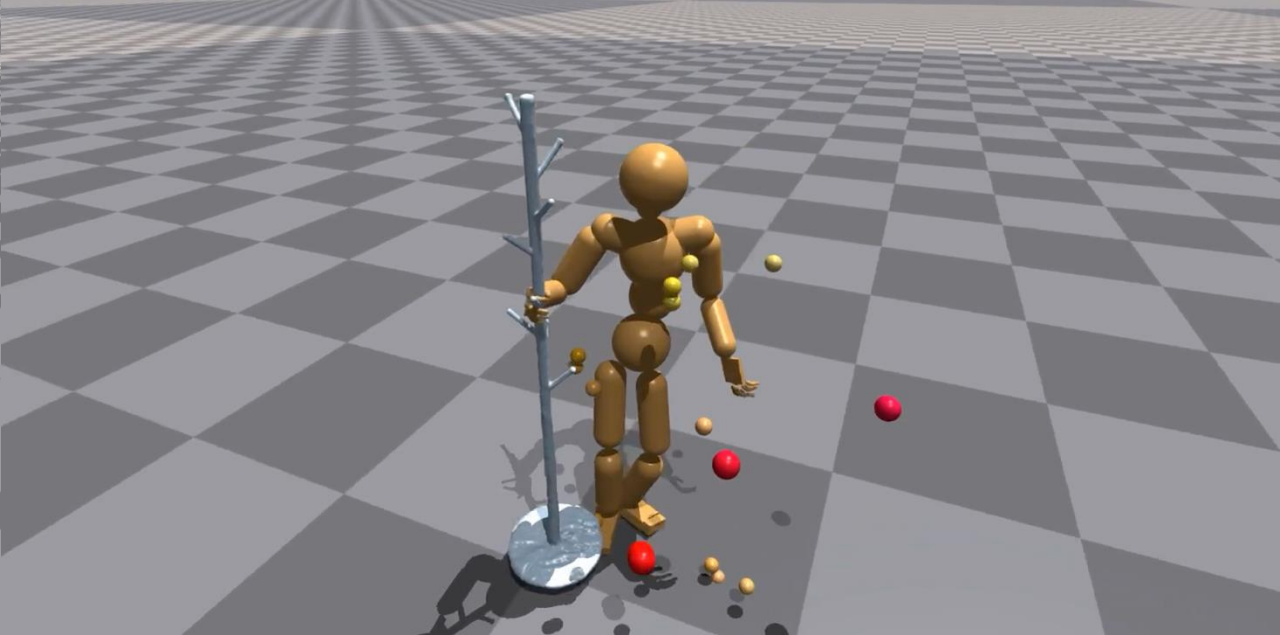
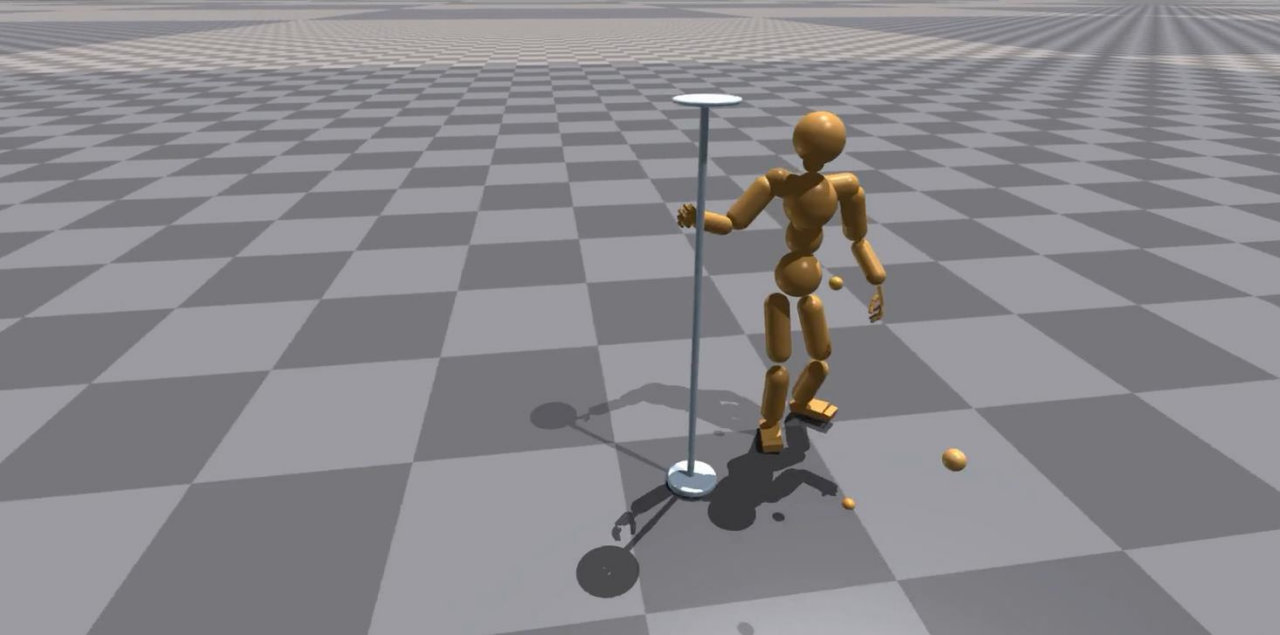
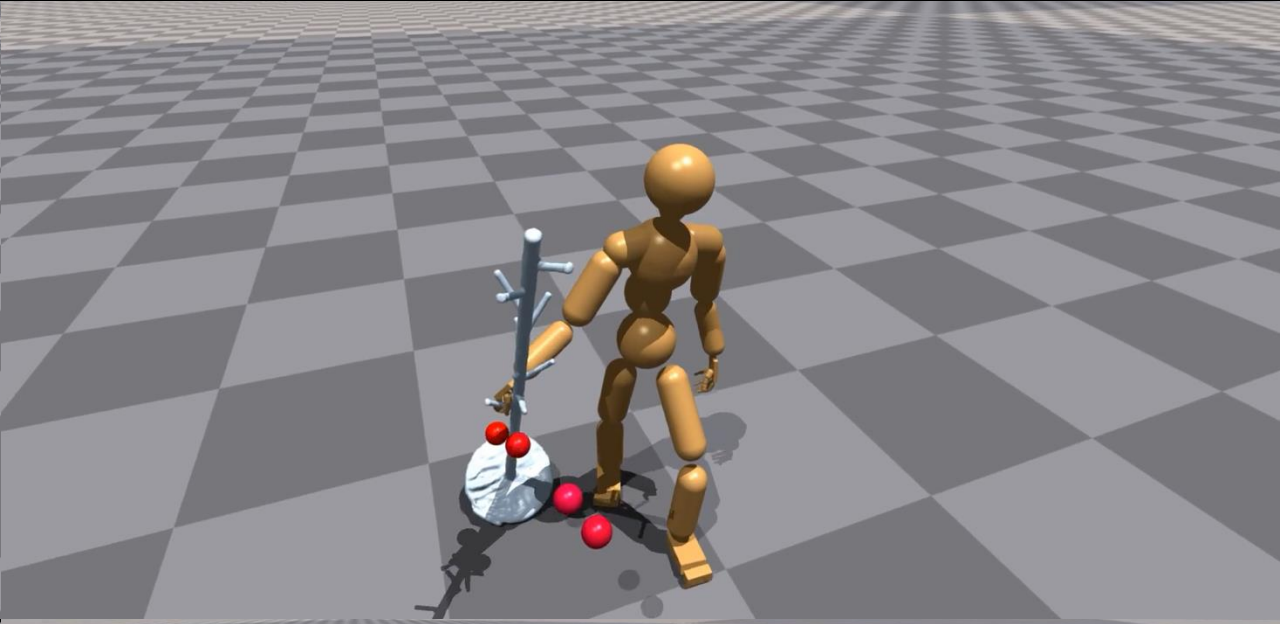
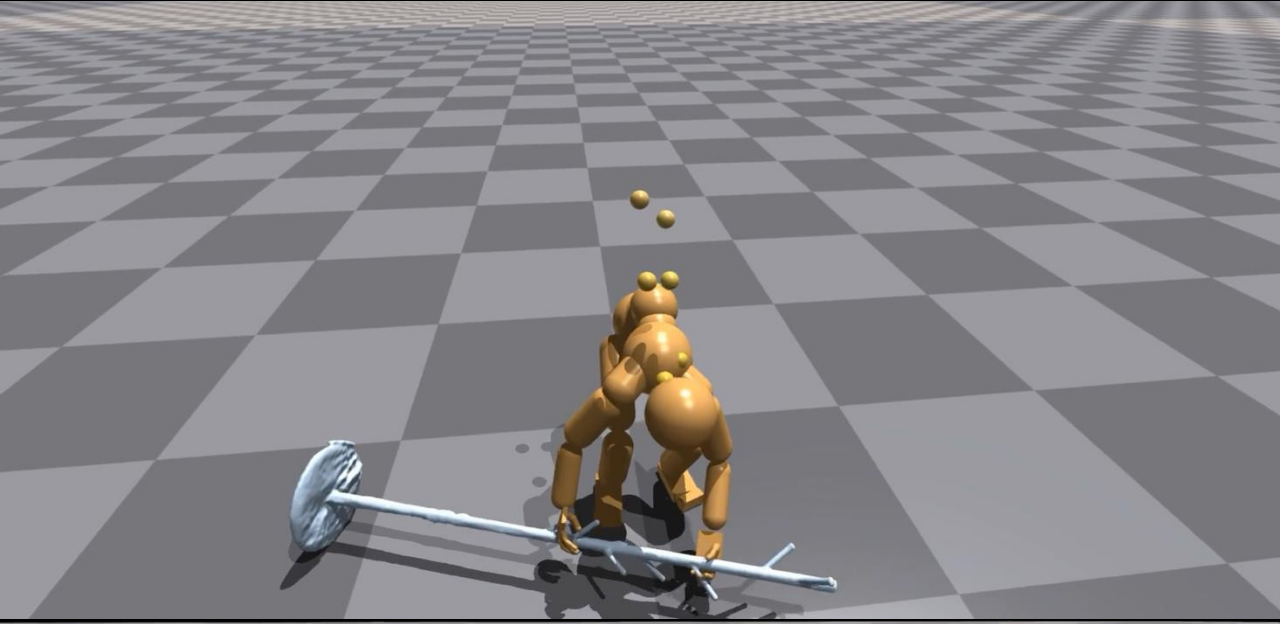
Grauman et al. Ego-exo4d: Understanding skilled human activity from first- and third-person perspectives. CVPR 2024

Humanoids need **muscle memory** that compress demos into reusable skill manifold

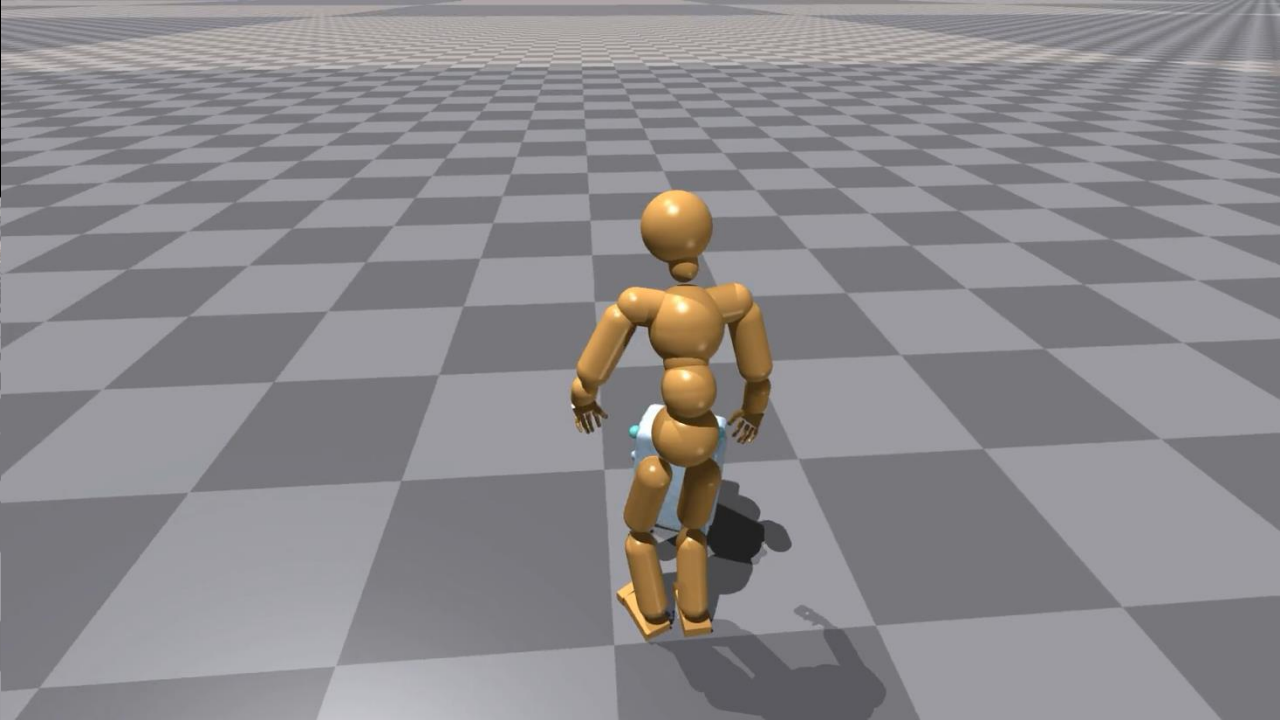
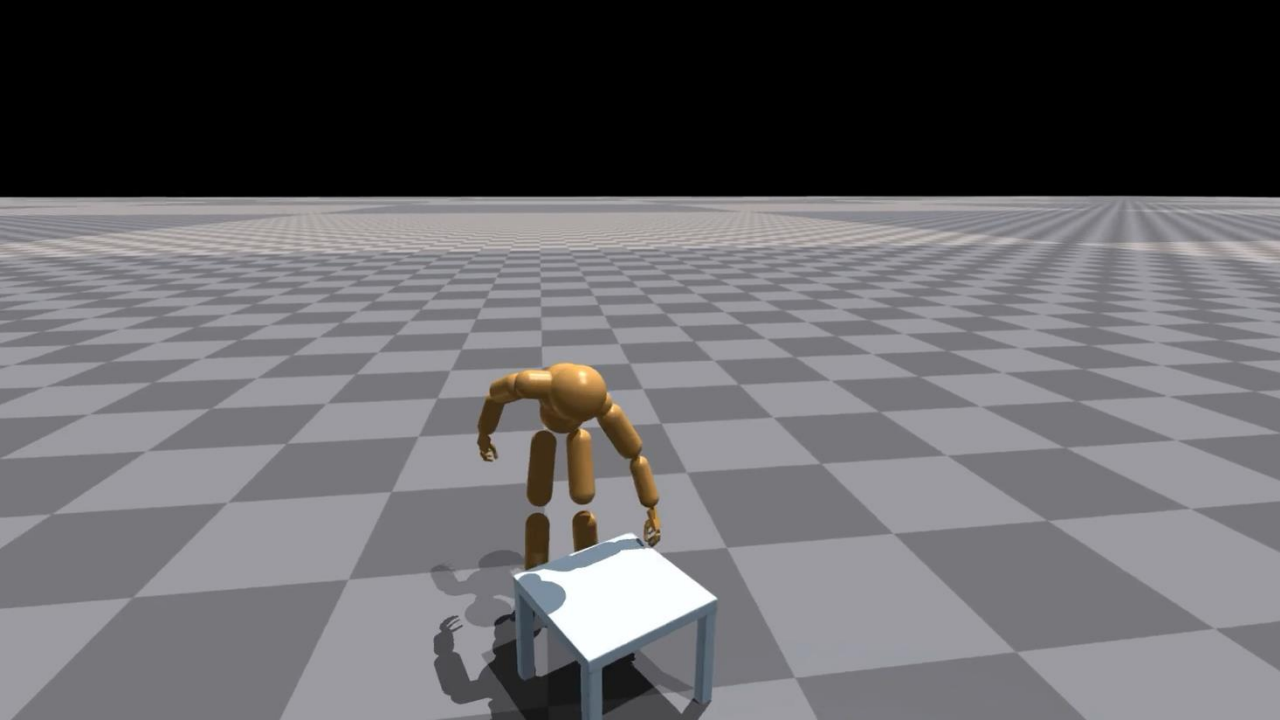
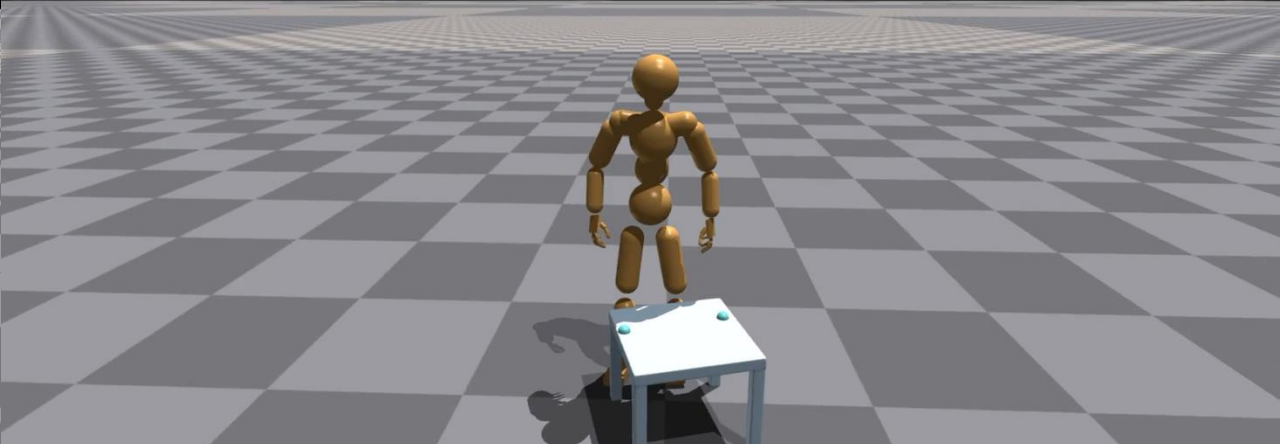
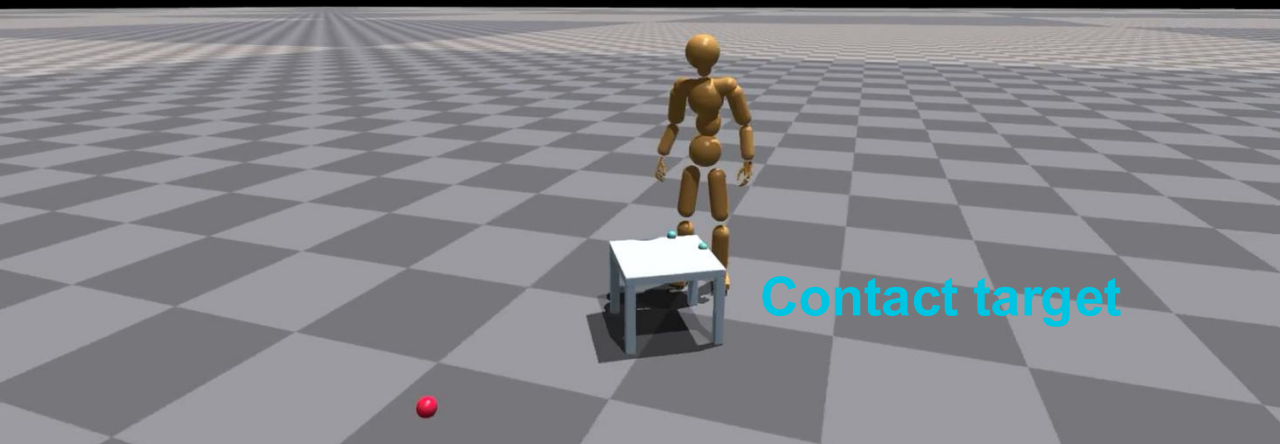
# From a future snapshot to whole-body skill



# From a target trajectory to whole-body skill



# From a contact target to whole-body skill



# How we build the muscle memory

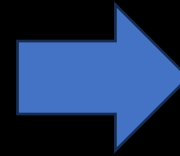


Three stages turn expert demonstrations into a reusable motor prior



**Sirui Xu**, Samuel Schulter, Morteza Ziyadi, Xialin He, Xiaohan Fei, Yu-Xiong Wang\*, Liang-Yan Gui\*. InterPrior: Scaling Generative Control for Physics-Based Human-Object Interactions. **CVPR 2026 Highlight**

# Stage I: Imitation of expert skills



## Human-Object Interaction Reference

<https://afterjourney00.github.io/IM-HOI.github.io/>

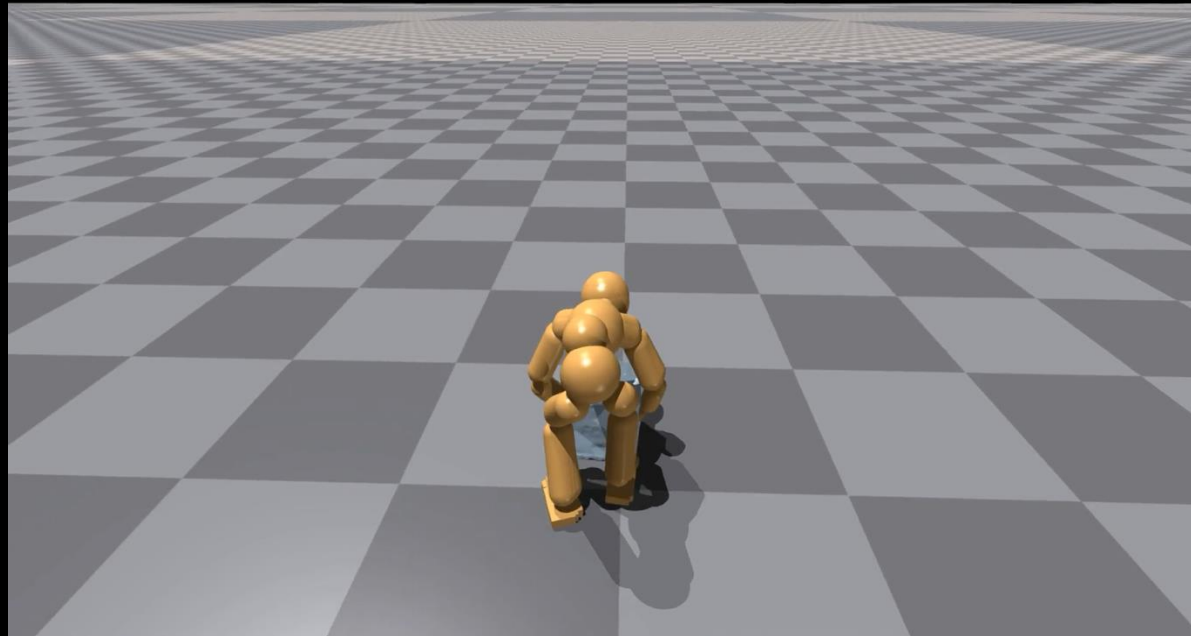
## Physical Motor Skills

Sirui Xu, Hung Yu Ling, Yu-Xiong Wang\*, Liang-Yan Gui\*. InterMimic: Towards Universal Whole-Body Control for Physics-Based Human-Object Interactions. **CVPR 2025 Highlight**

# Stage II: Distillation into one policy



- Distill the expert into a **masked variational** policy that maps **any** sparse goal-level inputs to a distribution

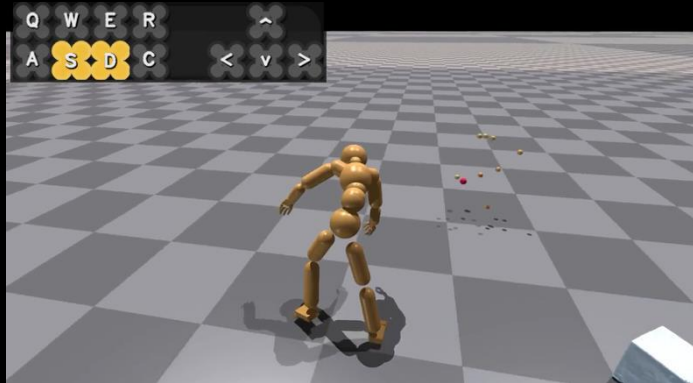


Sirui Xu, Samuel Schulter, Morteza Ziyadi, Xialin He, Xiaohan Fei, Yu-Xiong Wang\*, Liang-Yan Gui\*. InterPrior: Scaling Generative Control for Physics-Based Human-Object Interactions. **CVPR 2026 Highlight**

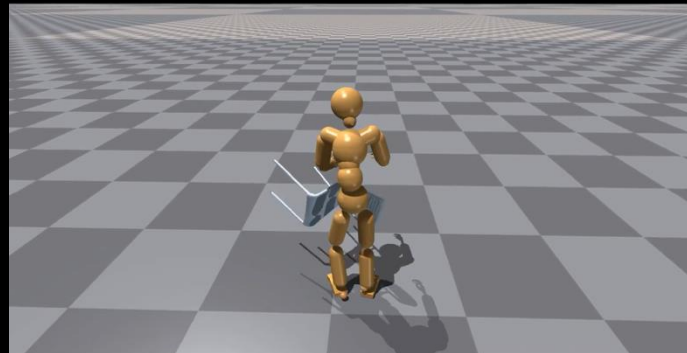
# Stage III: Post-training with RL



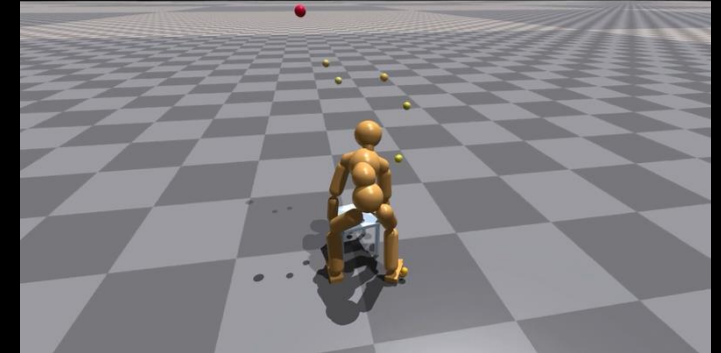
- **Fine-tune** the goal-conditioned policy under unseen configurations



**Getup**



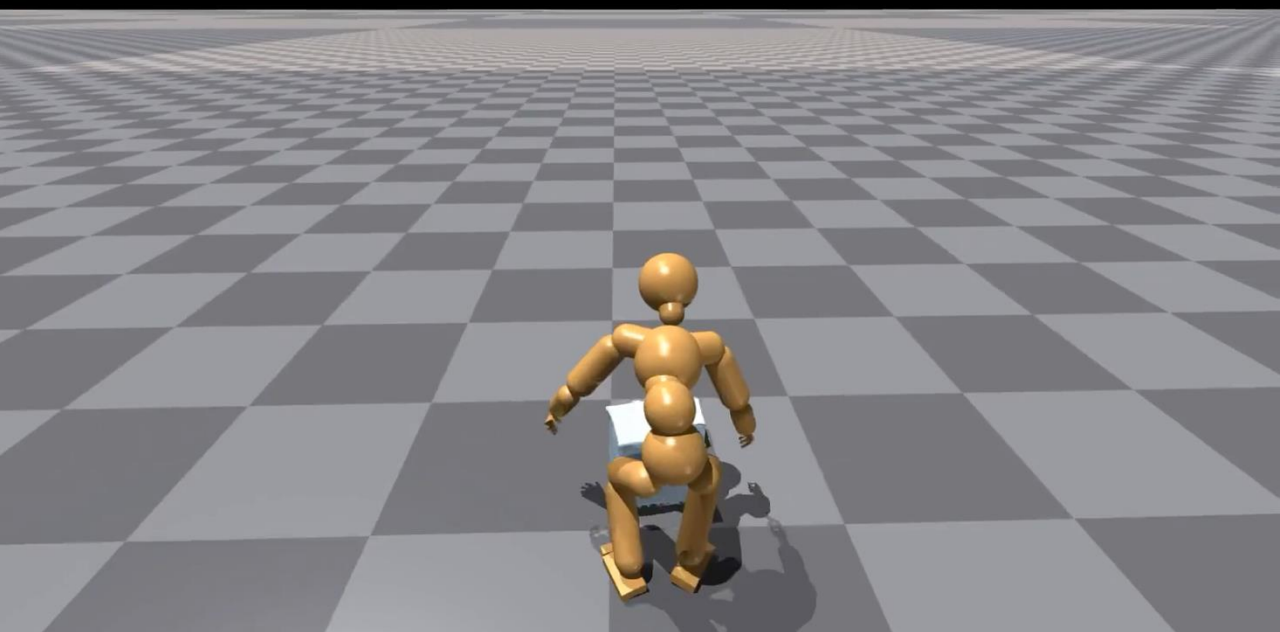
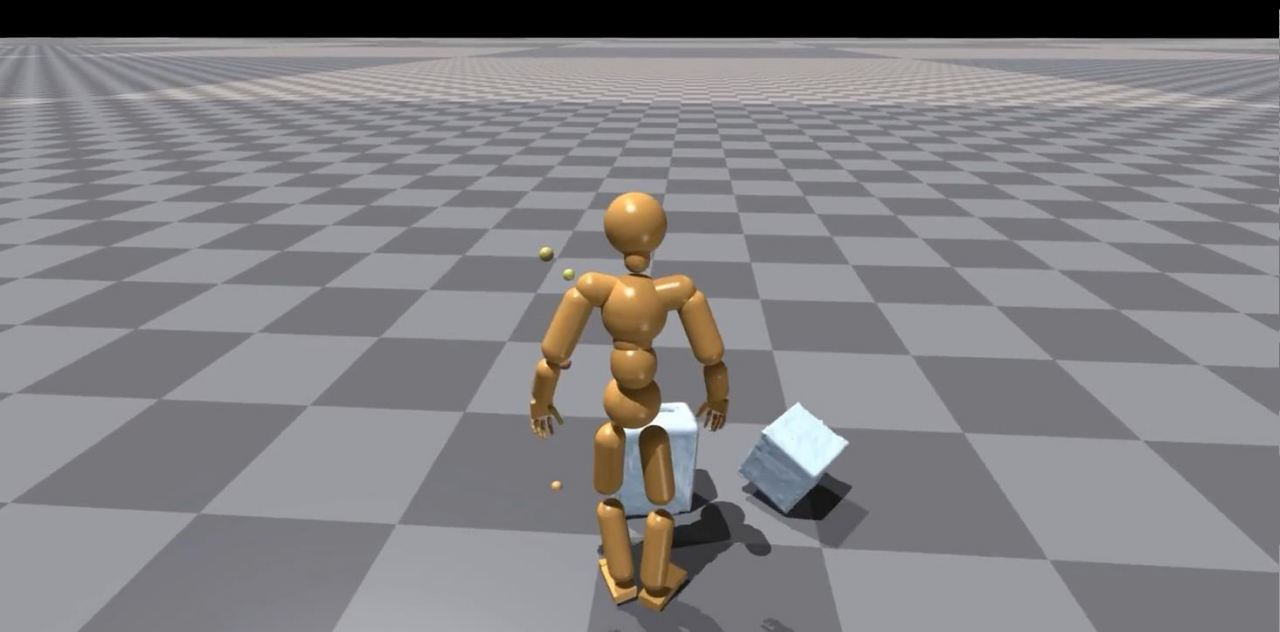
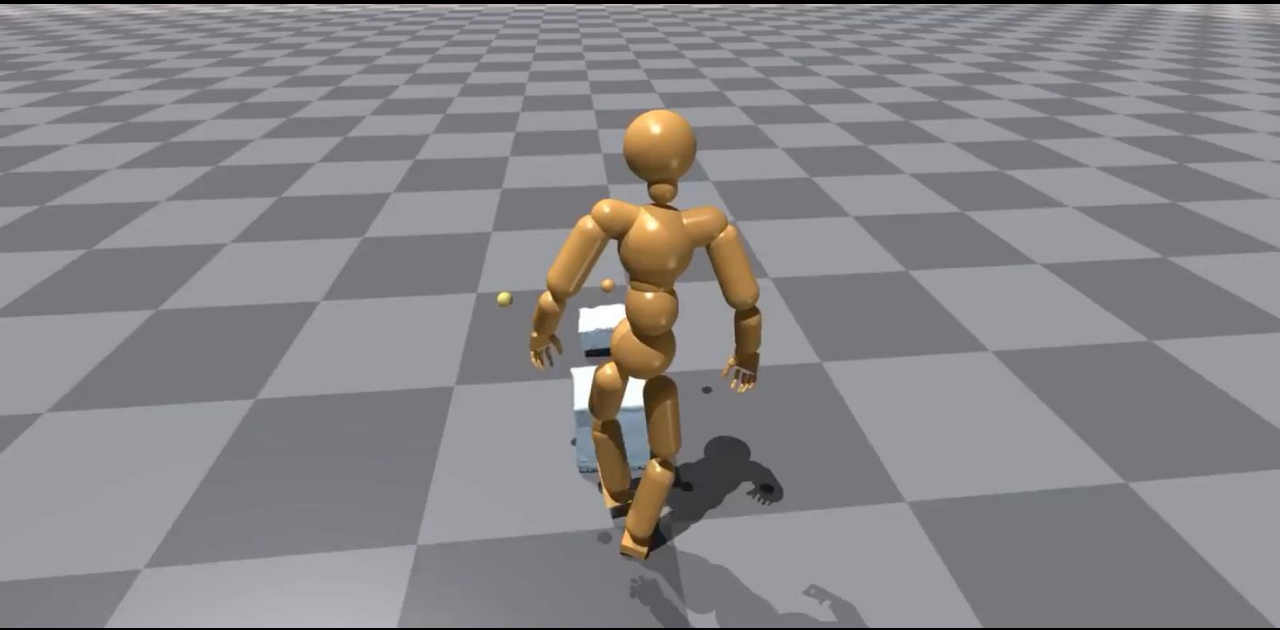
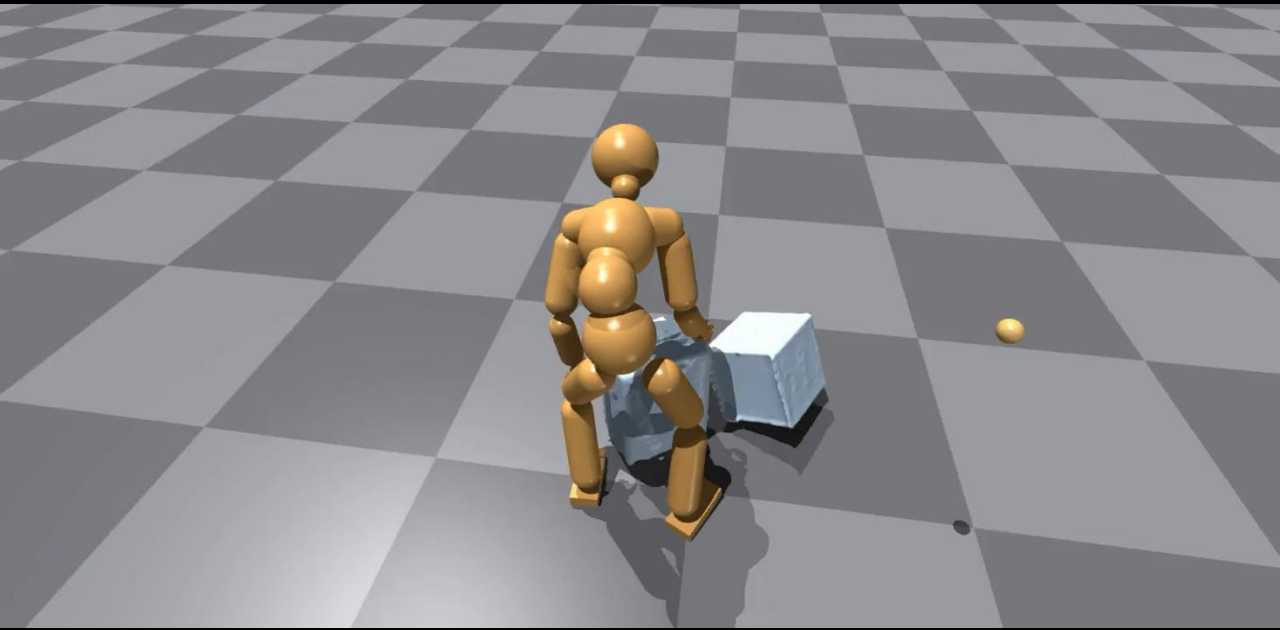
**Regrasp**



**Long-Horizon  
Composition**

Sirui Xu, Samuel Schuller, Morteza Ziyadi, Xialin He, Xiaohan Fei, Yu-Xiong Wang\*, Liang-Yan Gui\*. InterPrior: Scaling Generative Control for Physics-Based Human-Object Interactions. **CVPR 2026 Highlight**

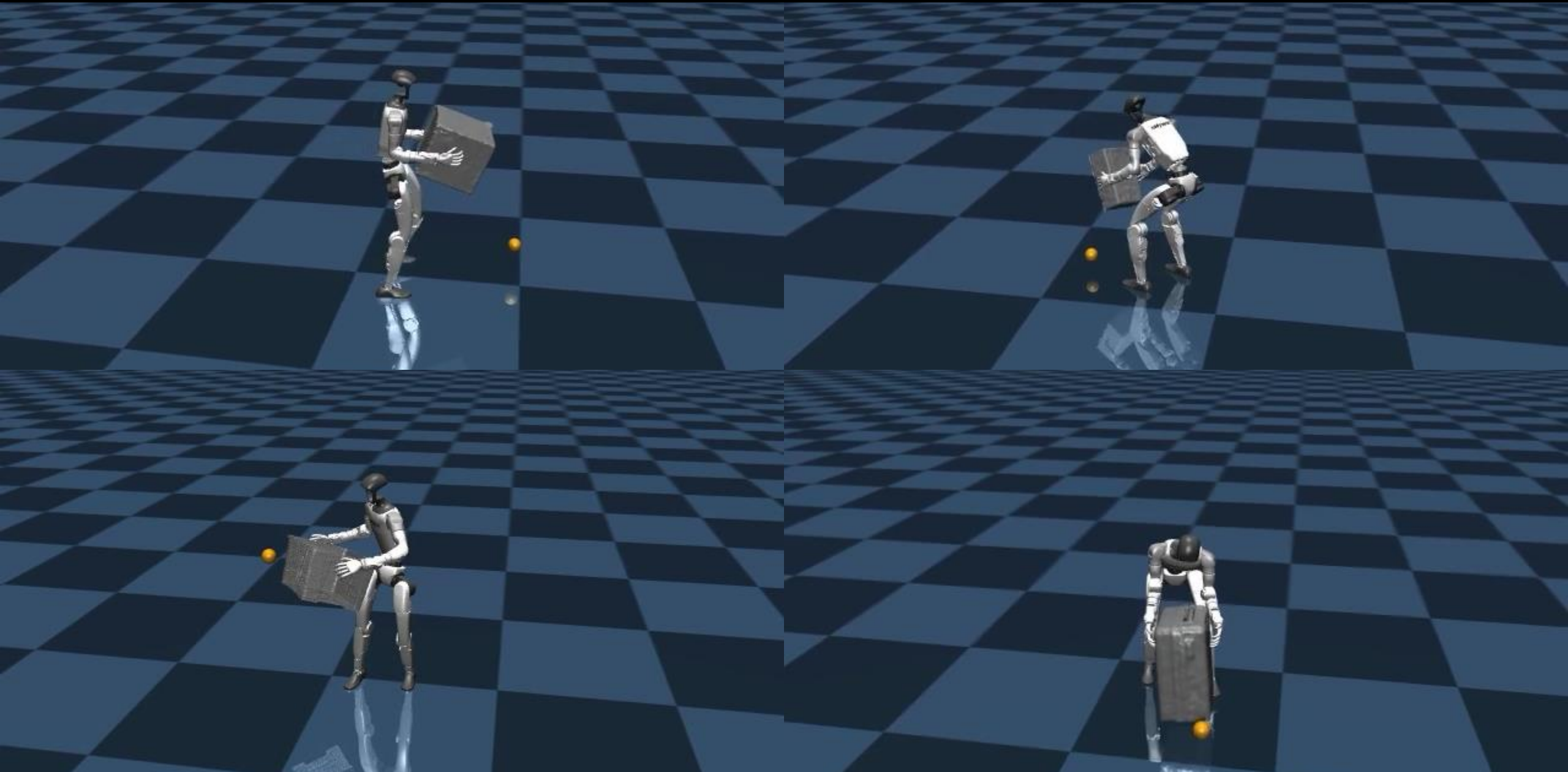
# Interacting with multiple objects



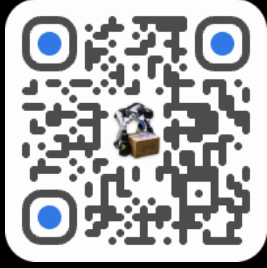
# Controllable with user's input



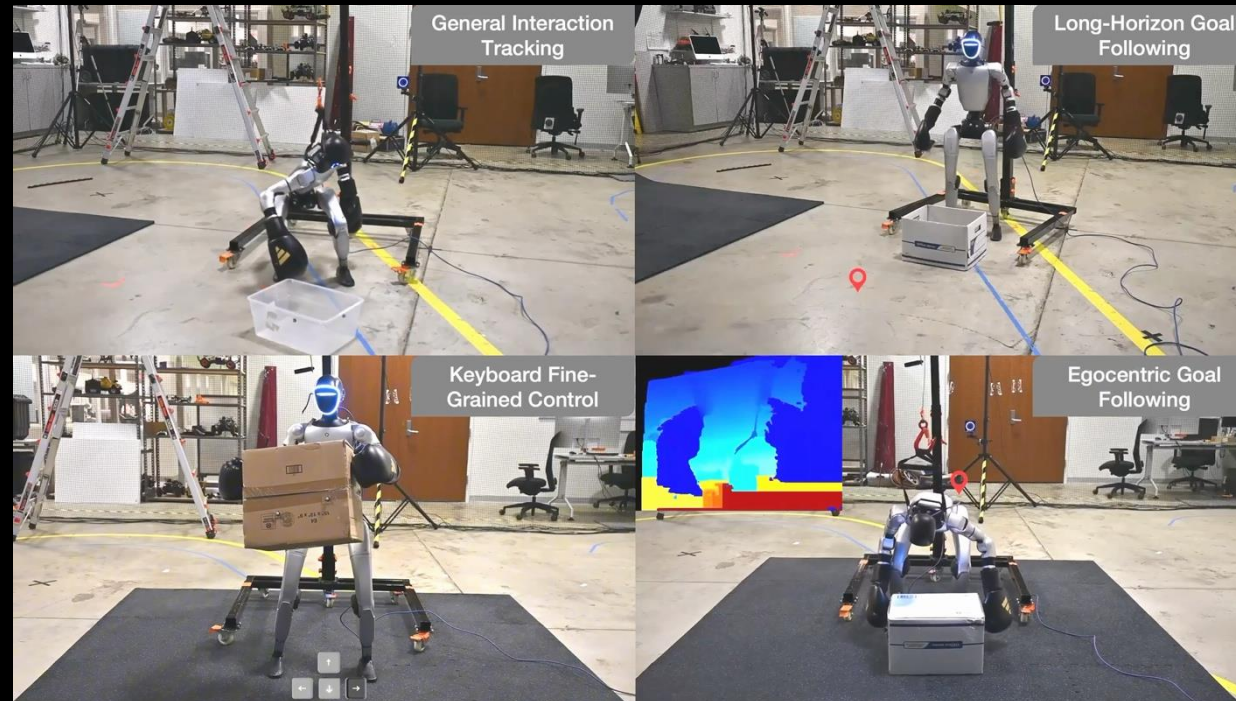
# Sim-to-sim transfer



# All-in-one in deployment: multi-task and multi-modal



- NO privileged object information in real world
- Integrate perception into the loop



Xialin He\*, Sirui Xu\*, Xinyao Li, Runpei Dong, Liuyu Bian, Yu-Xiong Wang\*, Liang-Yan Gui\*. ULTRA: Unified Multimodal Control for Autonomous Humanoid Whole-Body Loco-Manipulation. ArXiv 2026

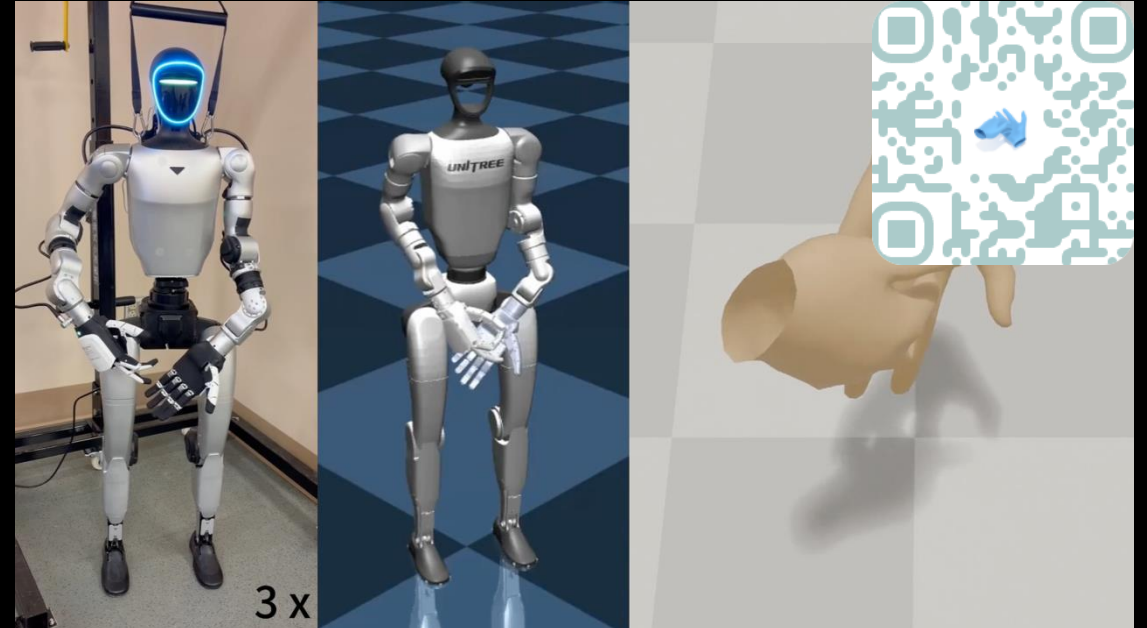
# Deploying with dexterity at scale

**Dexplore:** generative control for dexterous robot hand



**Sirui Xu**, Yu-Wei Chao, Liuyu Bian, Arsalan Mousavian, Yu-Xiong Wang\*, Liang-Yan Gui\*, Wei Yang\*. **Dexplore:** Scalable Neural Control for Dexterous Manipulation from Reference-Scoped Exploration. **CoRL 2025**

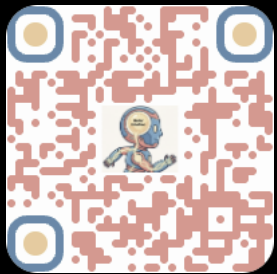
**HandX:** scaling bimanual interaction for a universal model



Zimu Zhang\*, Yucheng Zhang\*, Xiyan Xu, Ziyin Wang, **Sirui Xu**, Kai Zhou, Bing Zhou, Chuan Guo, Jian Wang, Yu-Xiong Wang\*, Liang-Yan Gui\*. **HandX:** Scaling Bimanual Motion and Interaction Generation. **CVPR 2026**

# Scale at every stage, but how you scale matters

- Turn finite captures into an unlimited resource: **scaling** dynamics, behavior, and dexterity, not just volume
- From human data to robot data: imitate, but also retarget, correct, and augment human data at **scale**
- **Scale** motor prior, “*Intelligence is useful compression in service of action*”



June 6, 4:45 –  
6:45 PM  
Poster #194

InterPrior

June 5, 10:45  
AM – 12:45 PM  
Poster #210



HandX



Thank you!