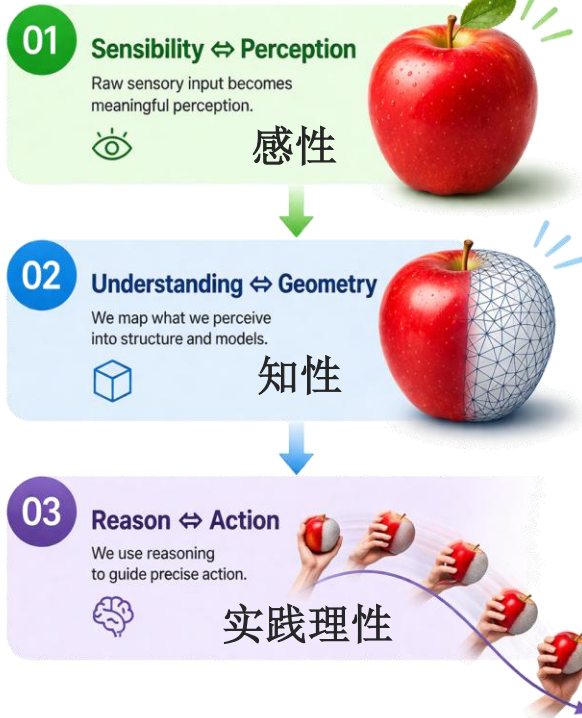


# From Egocentric Experience to Interactive Intelligence: Building the World in First Person

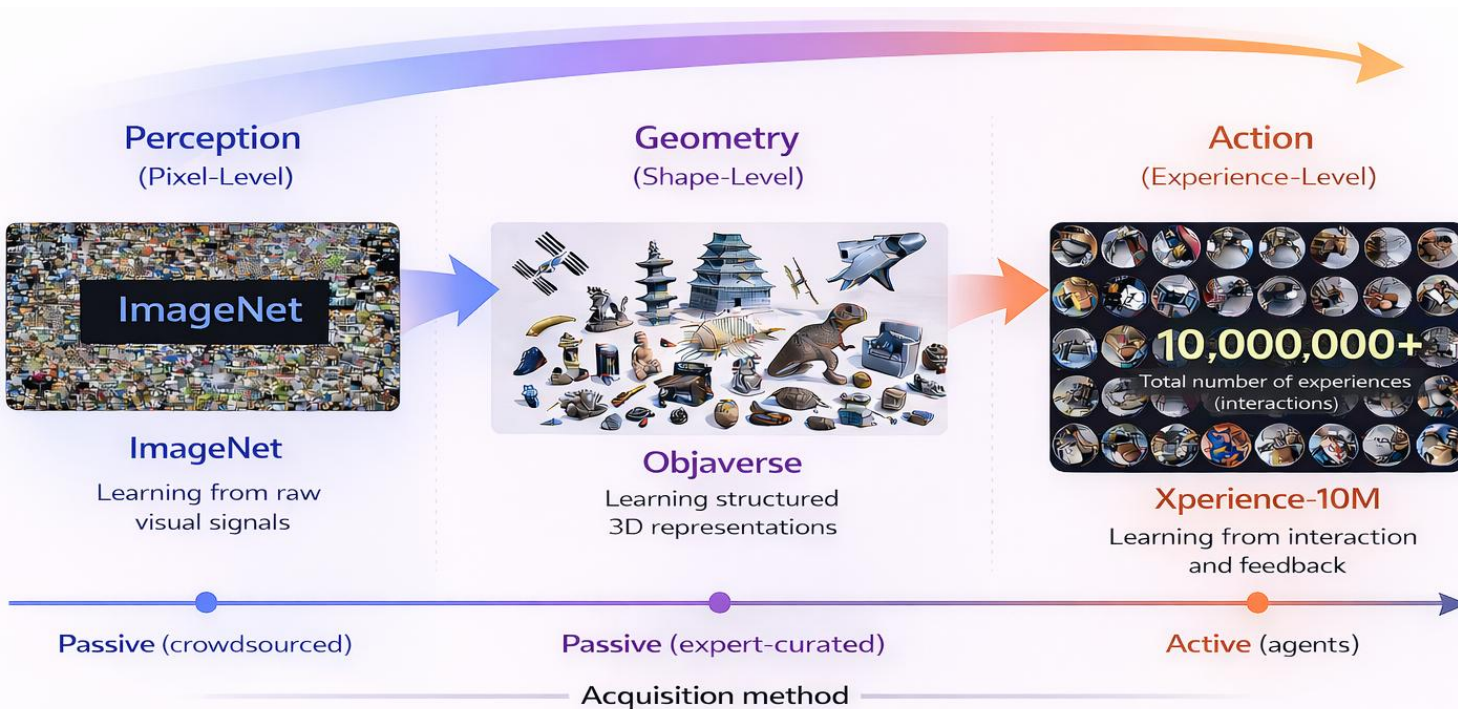
Ziwei Liu

Nanyang Technological University

# Three Dimensions of Kant's Laws of Nature



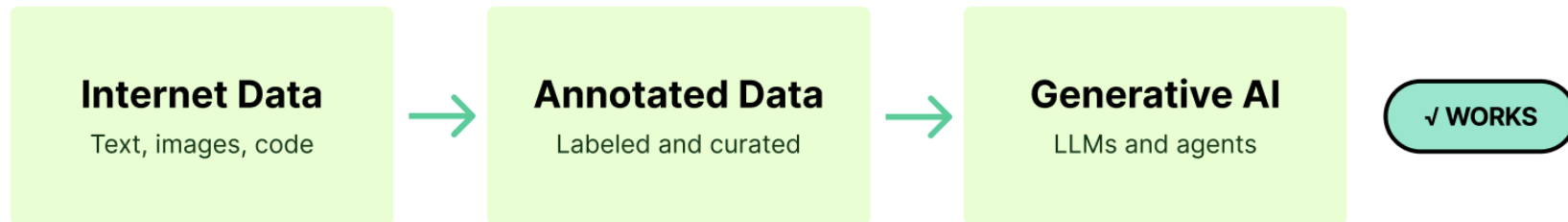
# From Perception to Action: The Evolution of Visual Intelligence



# The Real-World Data Bottleneck

The same scaling law that powered LLMs is now entering the physical world.

## GENERATIVE AI PIPELINE

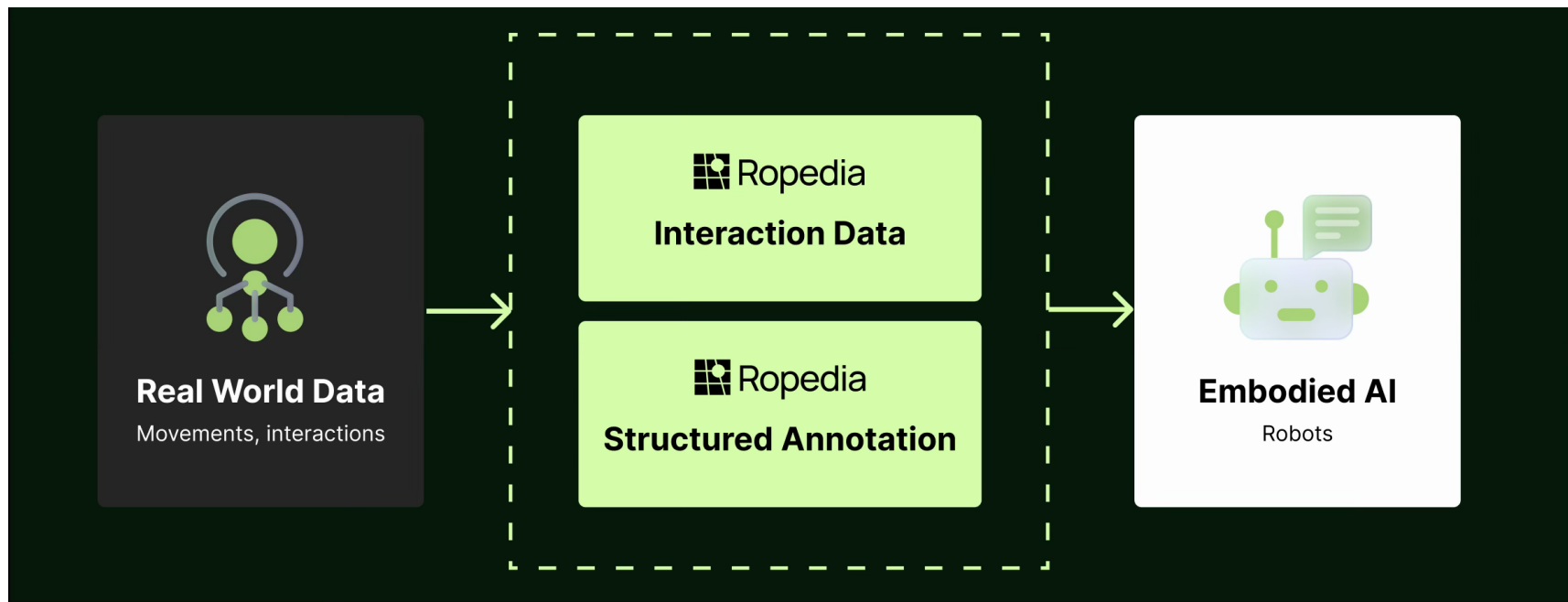


## EMBODIED AI PIPELINE



# Our Solution

We capture structured real-world interaction at **scale** and **low cost**.



# Full-stack Low-cost Portable Data Capture

## HOMIE

320g

Super lightweight

360°

Field of View

10x

Faster Deployment

50 $\mu$ s

Synchronization

13h

Storage Duration

12.5x

Cost Reduction



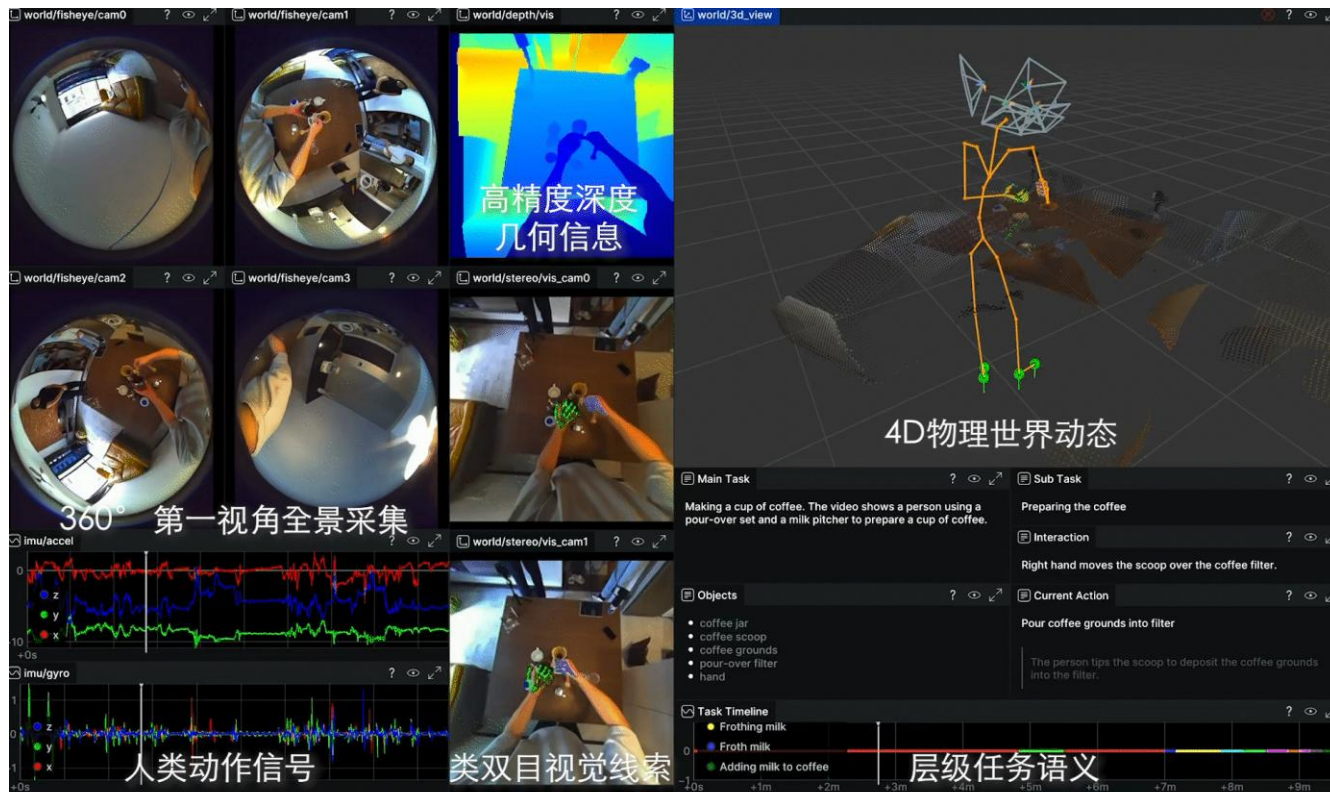
# Our Dataset: In-the-Wild Human Xperience-10M

2.1M Total Downloads in 1 Week

Top 3 Over ALL Datasets on Hugging Face

The image shows a screenshot of the Hugging Face datasets interface. On the left, the navigation sidebar includes sections for 'Main', 'Tasks', 'Libraries', 'Languages', and 'Licenses'. Under 'Modalities', 'Video' is selected. The 'Format' section shows 'parquet' as the chosen format. The 'Evaluation' section has a 'Benchmark' button. The main content area displays a list of datasets, with 'ropedia-ai/xperience-10m' highlighted in blue. To the right of the dataset list is a large grid of 360-degree camera views showing various indoor scenes. Overlaid on this grid is the text '10,000,000+' in large green font, with 'Total number of experience (interaction)' written below it. The Ropedia logo is visible in the bottom right corner of the grid.

# High-Quality Data from Anywhere by Anybody



INTERACTION EPISODES

10M

VIDEO W/O AUDIO

10,000h

RGB FRAME NUMBER

2.88B

MOCAP FRAME

576M

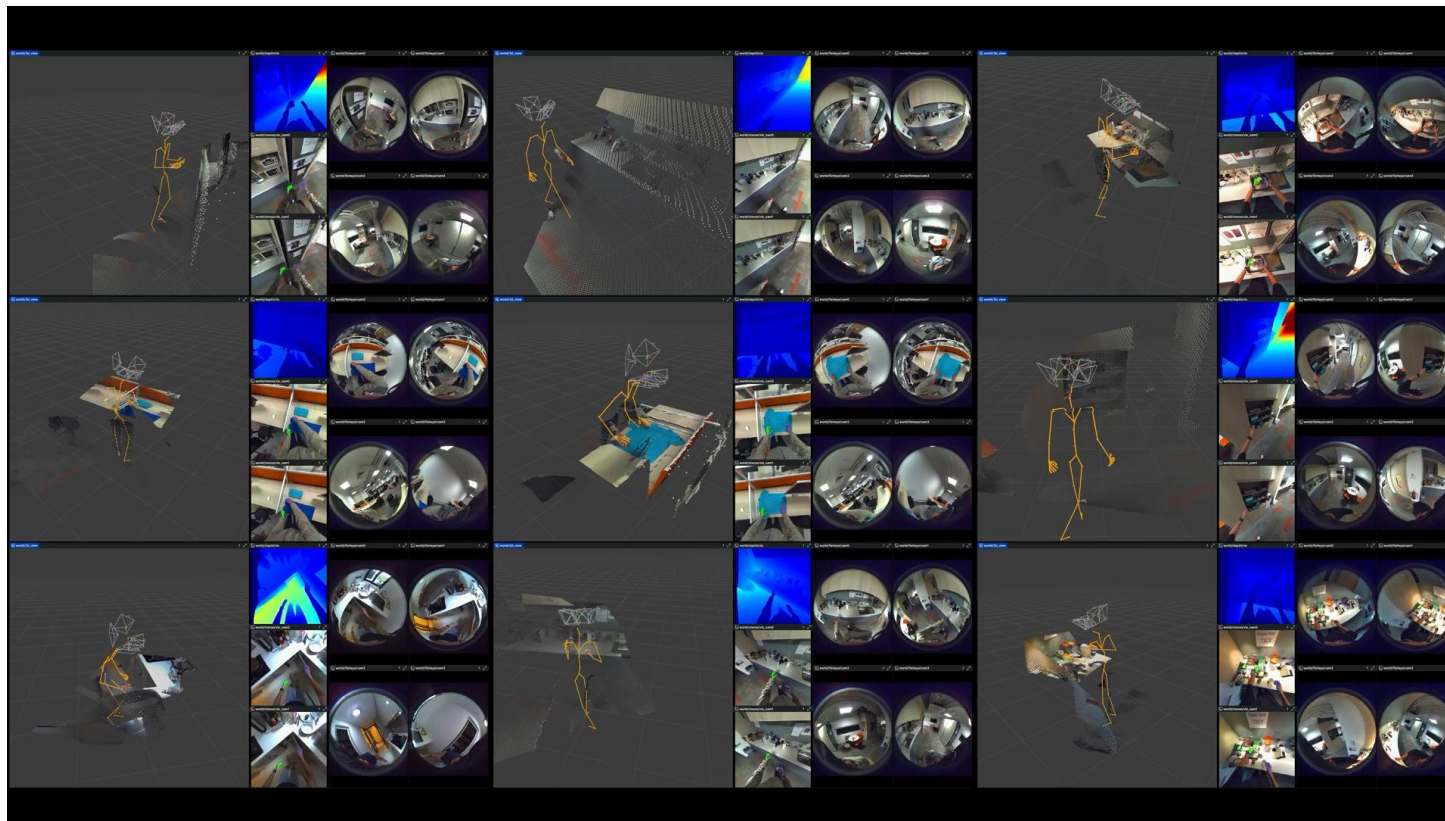
NO. OF OBJECTS

350K

TOTAL STORAGE

~1PB

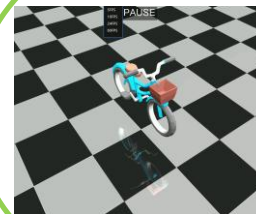
# Demo: In-the-wild Data Capture



# From **Interactive Experience** to **Simulation-Ready** Modeling



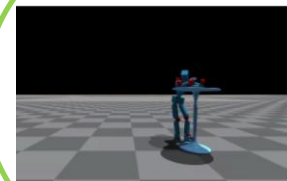
Xperience-10M



Object



Human-Object  
Interaction

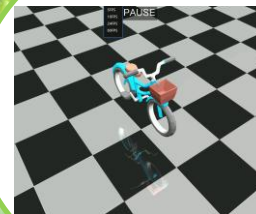


Human-Scene  
Interaction

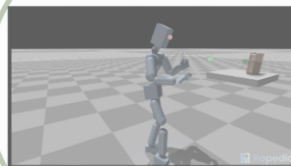
# From **Interactive Experience** to **Simulation-Ready Modeling**



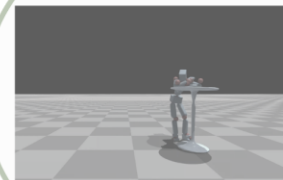
Xperience-10M



Object



Human-Object  
Interaction



Human-Scene  
Interaction

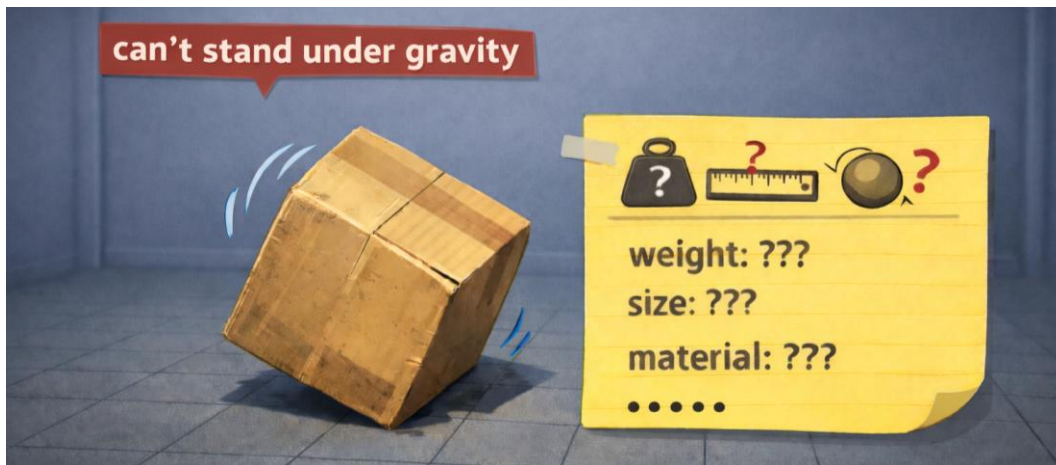
# PhysX-Anything: **Simulation-Ready** Physical 3D Assets from Single Image

Ziang Cao, Fangzhou Hong, Zhaoxi Chen, Liang Pan, Ziwei Liu

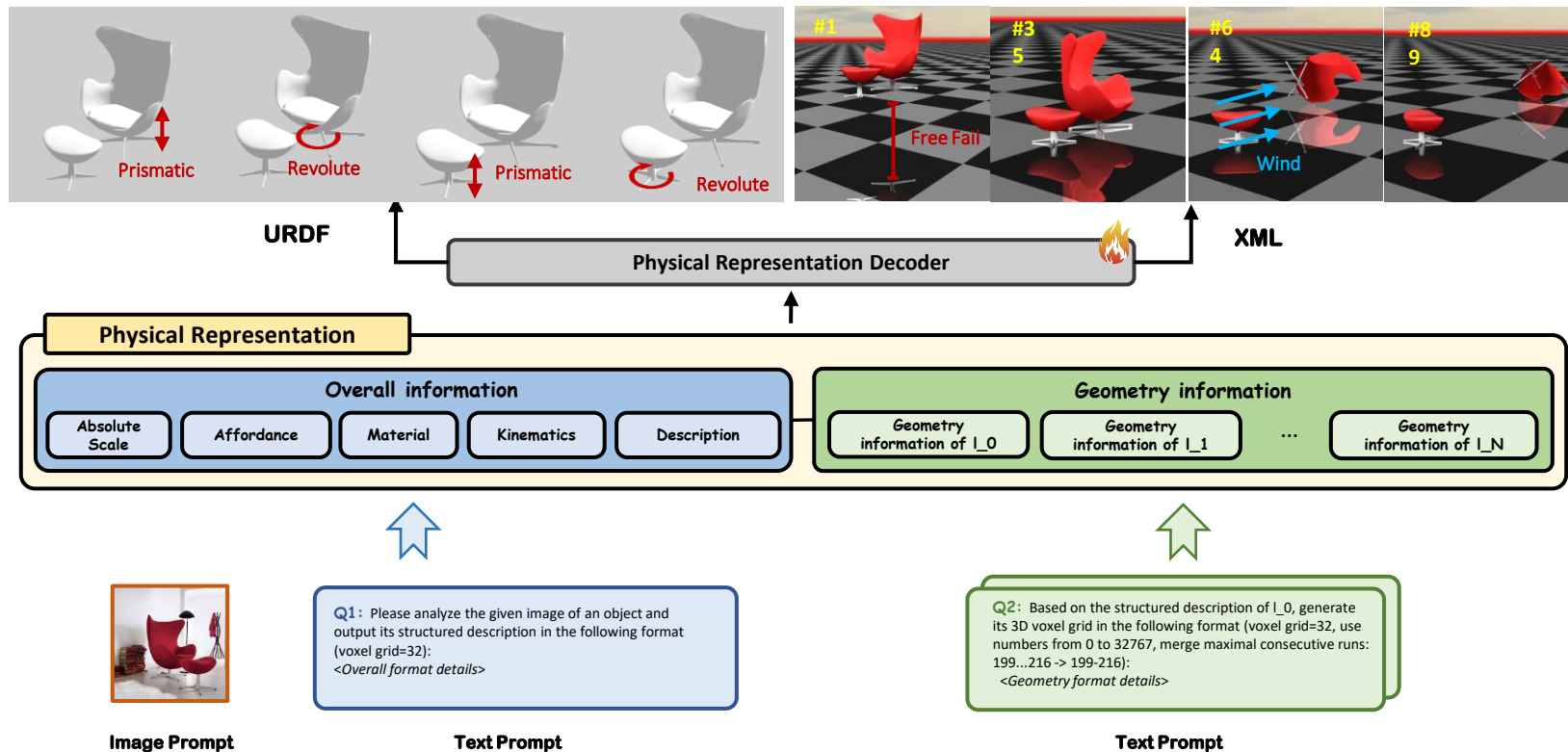


# Challenge

- ▶ Existing 3D generation primarily emphasizes geometries and textures while **neglecting physical-grounded modeling**, hampering their real-world application in physical domains like simulation and embodied AI.



# Key Idea: Interactive 3D Object as **Physical Code**

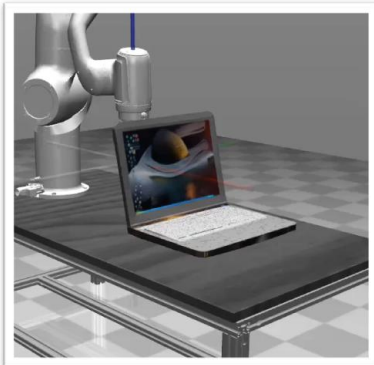


## Results: Simulated Objects

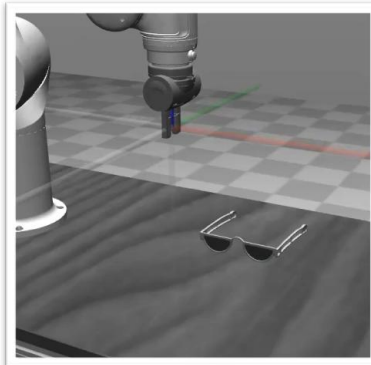


# Results: Simulated Interactions

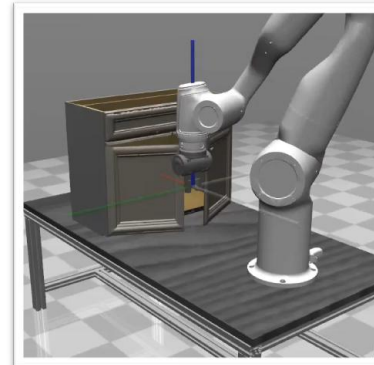
Laptop Closing



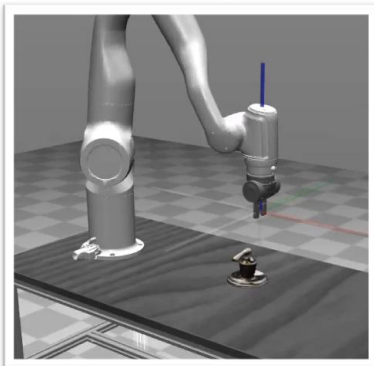
Eyeglass Temple Folding



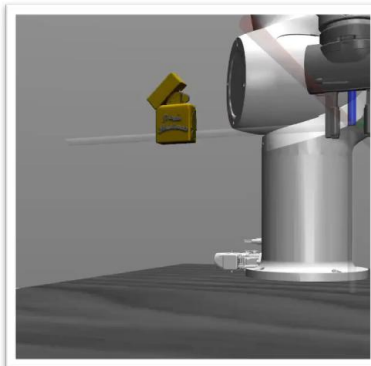
Door Opening and Closing



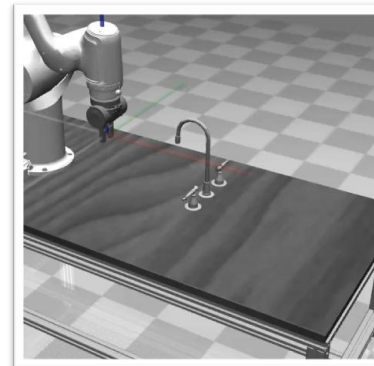
Handle Manipulation



Lighter Snapping Open



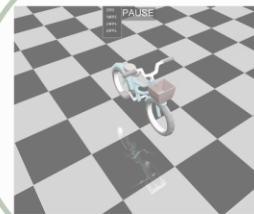
Faucet Switch Manipulation



# From **Interactive Experience** to **Simulation-Ready** Modeling



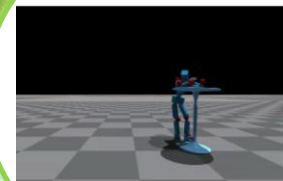
Xperience-10M



Object



Human-Object  
Interaction



Human-Scene  
Interaction

# HSImul3R: Physics-in-the-Loop Reconstruction of **Simulation-Ready** Human-Scene Interactions

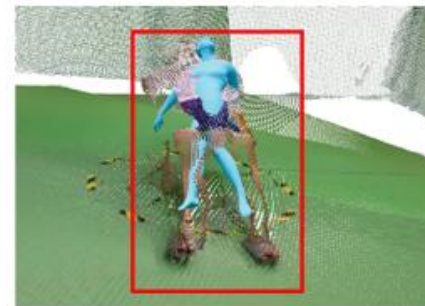
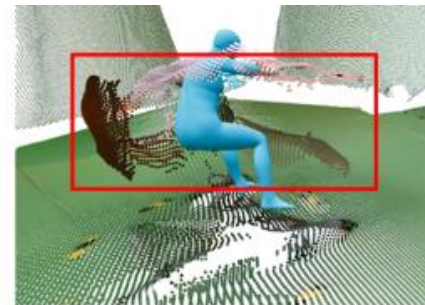
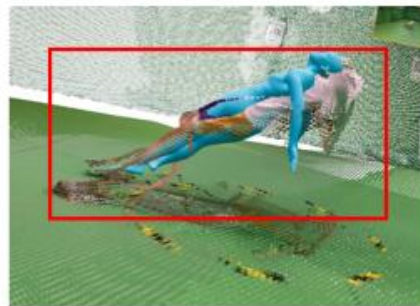
Yukang Cao, Haozhe Xie, Fangzhou Hong, Long Zhuo, Zhaoxi Chen, Liang Pan, Ziwei Liu



# Challenge

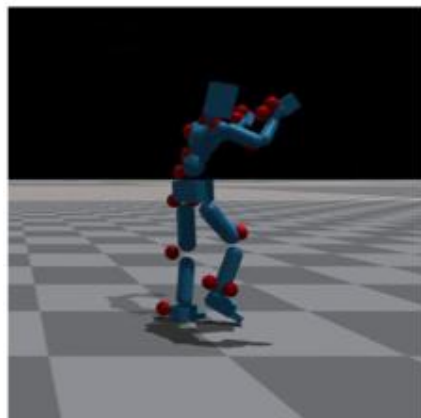
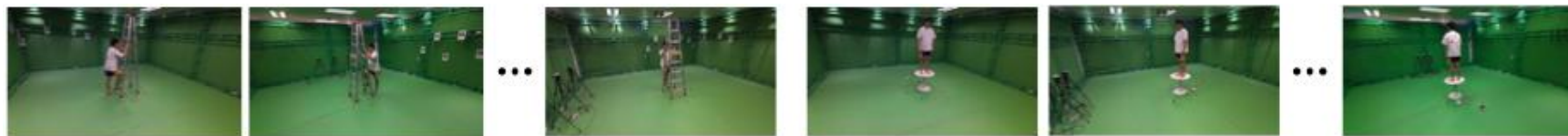


Inputs

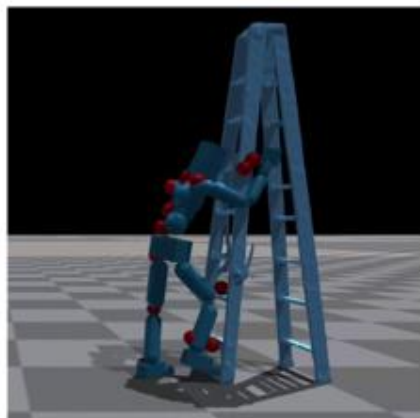


HSfM

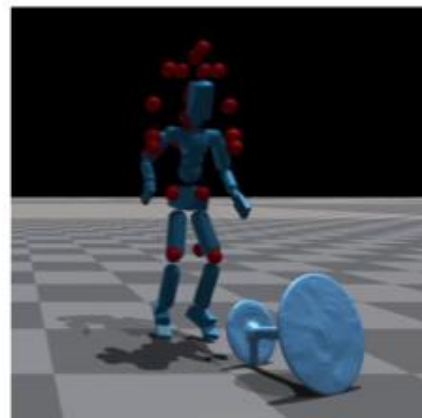
# Challenge



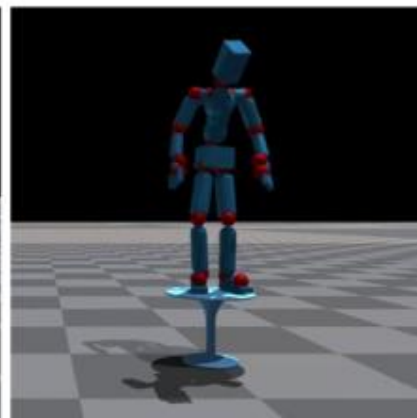
HSfM



Ours



HSfM



Ours

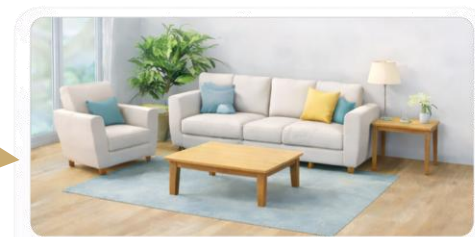
# Key Idea: Leverage **Simulator Interaction Feedback** to Optimize Human Motion and Scene Geometry



Casual Capture

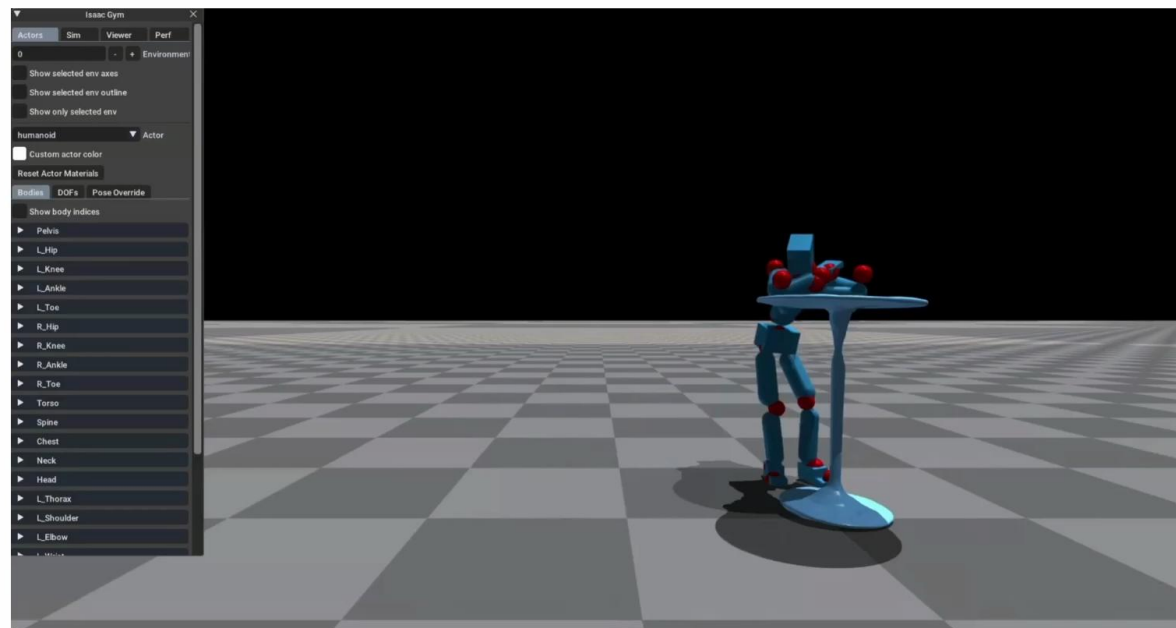
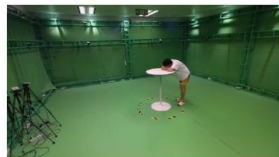


Optimize Human Motion



Optimize Scene Geometry

# Results: Real-to-Sim



# Results: Sim-to-Real



We are moving from a world that is seen,  
to a world that is **experienced**

