

# Egocentric 3D Capture and Understanding

## 第一人称视角下的三维捕捉与理解

Fangzhou Hong (洪方舟)

Advisor: Prof. Ziwei Liu (刘子纬)

S-Lab, Nanyang Technological University

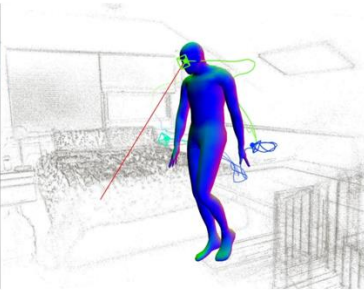


S-LAB  
FOR ADVANCED  
INTELLIGENCE

# Why study egocentric perception?

## Egocentric Perspective Provides:

### 1. Natural Human Experience and Cognition



*Gaze, Attention*

*Navigation,  
Spatial Awareness*

### 2. Better Context for Interaction



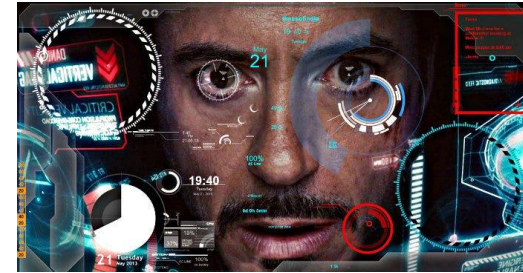
*Object Manipulation  
Hand-Eye Coordination*

*Social Interaction*



## Egocentric Perception Enables:

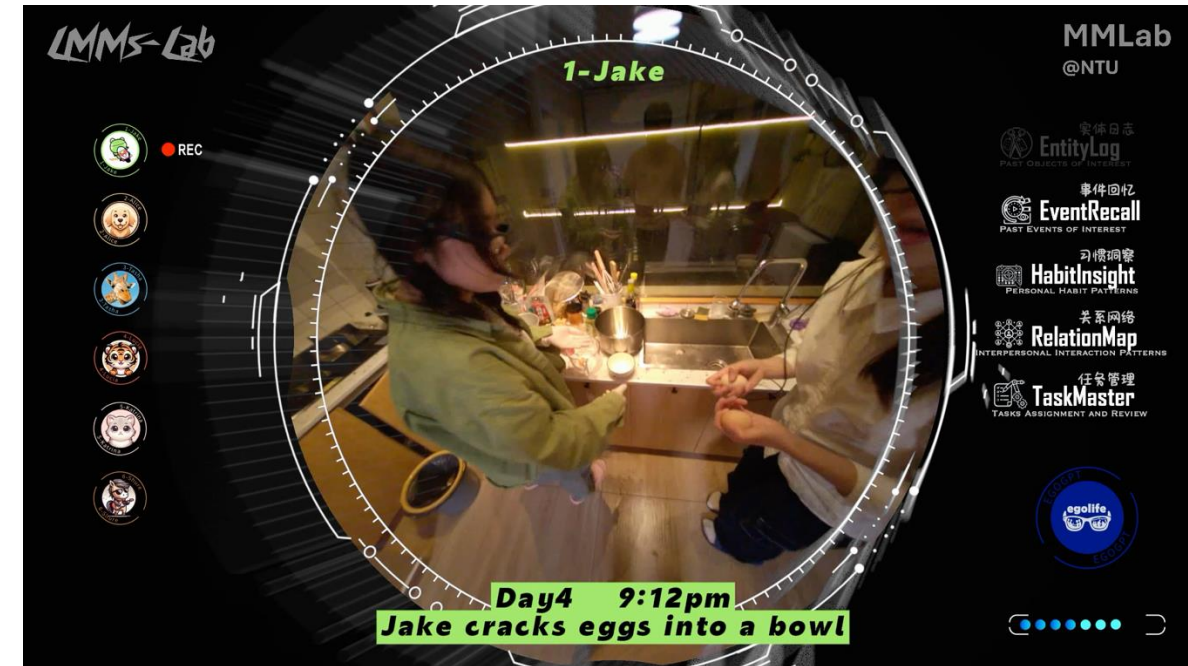
### 1. Personal AI Assistant



*EgoLife, CVPR2025  
EgoLM, CVPR2025*



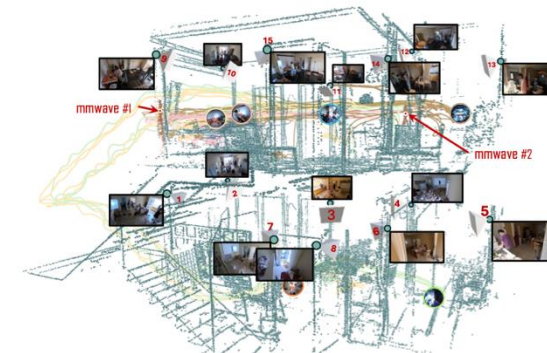
# Egocentric Personal AI Assistant



## EgoLife: Towards Egocentric Life Assistant

Jingkang Yang, Shuai Liu, Hongming Guo, Yuhao Dong, Xiamengwei Zhang, Sicheng Zhang, Pengyun Wang, Zitang Zhou, Binzhu Xie, Ziyue Wang, Bei Ouyang, Zhengyu Lin, Marco Cominelli, Zhongang Cai, Yuanhan Zhang, Peiyuan Zhang, **Fangzhou Hong**, Joerg Widmer, Francesco Gringoli, Lei Yang, Bo Li, Ziwei Liu

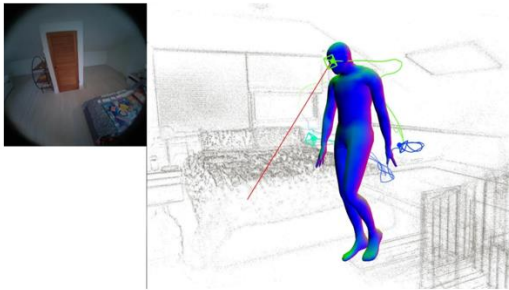
CVPR 2025



# Why study egocentric perception?

## Egocentric Perspective Provides:

### 1. Natural Human Experience and Cognition



*Gaze, Attention*

*Navigation, Spatial Awareness*

### 2. Better Context for Interaction



*Object Manipulation  
Hand-Eye Coordination*

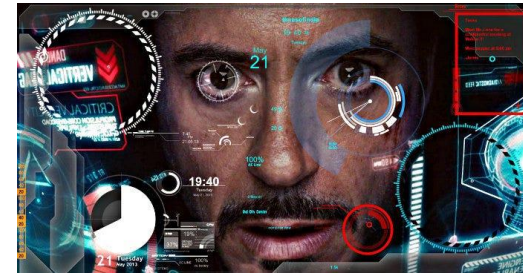


*Social Interaction*



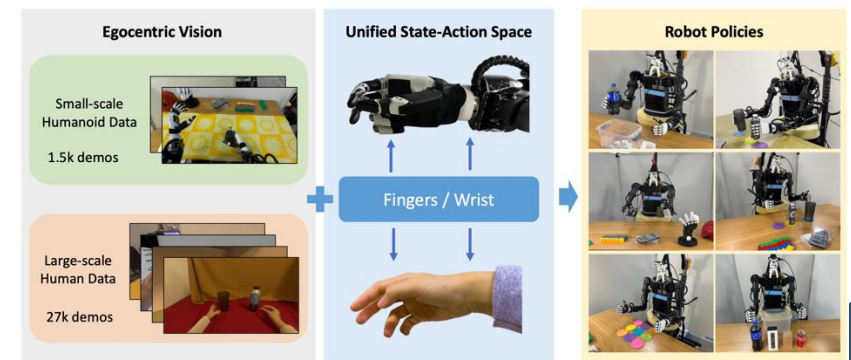
## Egocentric Perception Enables:

### 1. Personal AI Assistant



*EgoLife, CVPR2025  
EgoLM, CVPR2025*

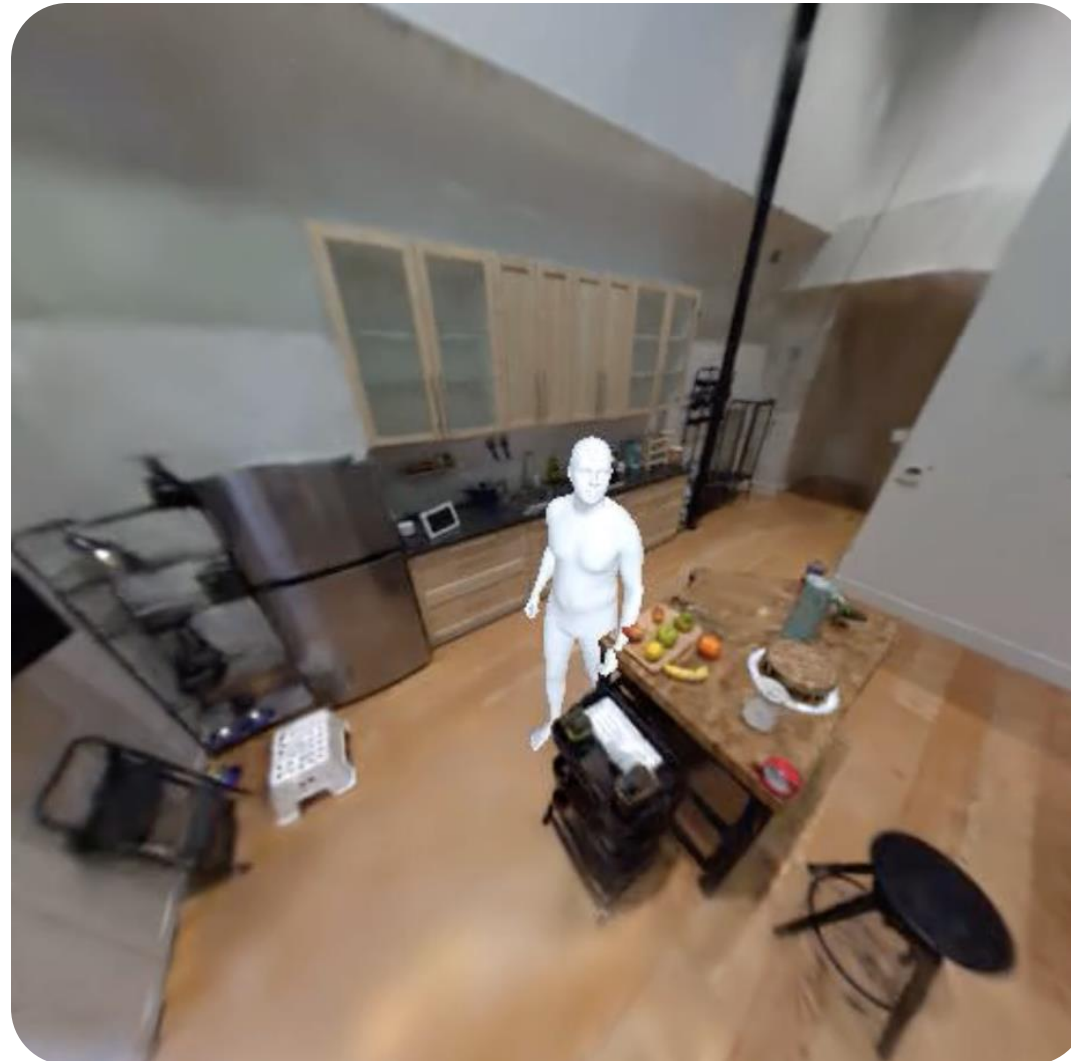
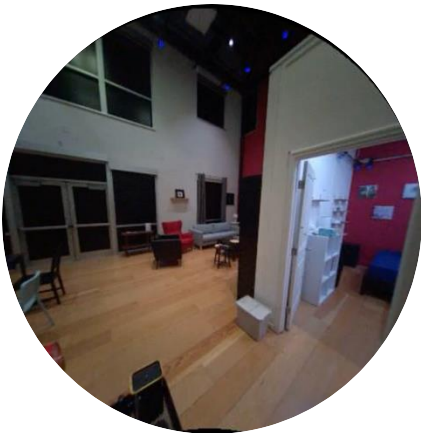
### 2. Embodied AI Learning



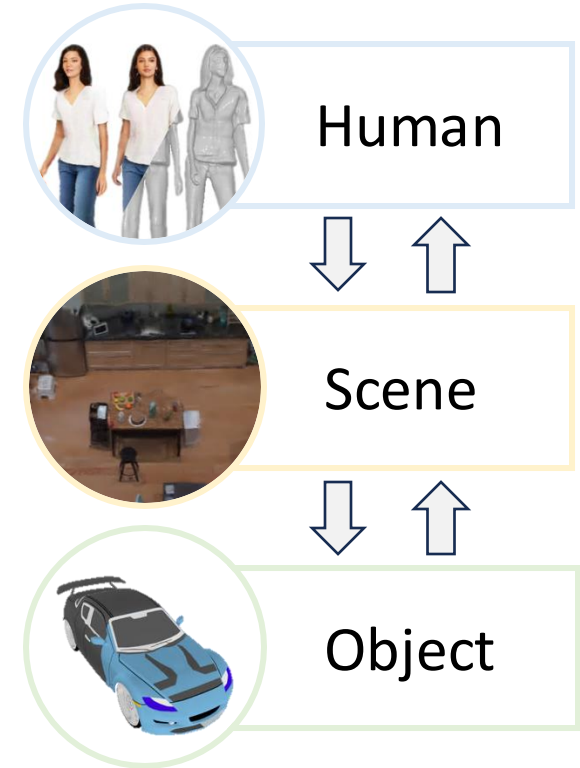
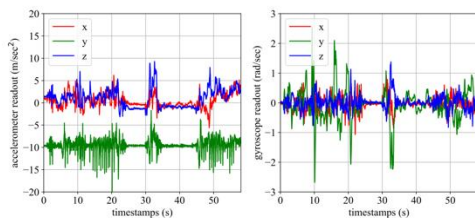
*DexCap, EgoMimic, PH<sup>2</sup>D*

# Holistic Capture from Egocentric Perceptions

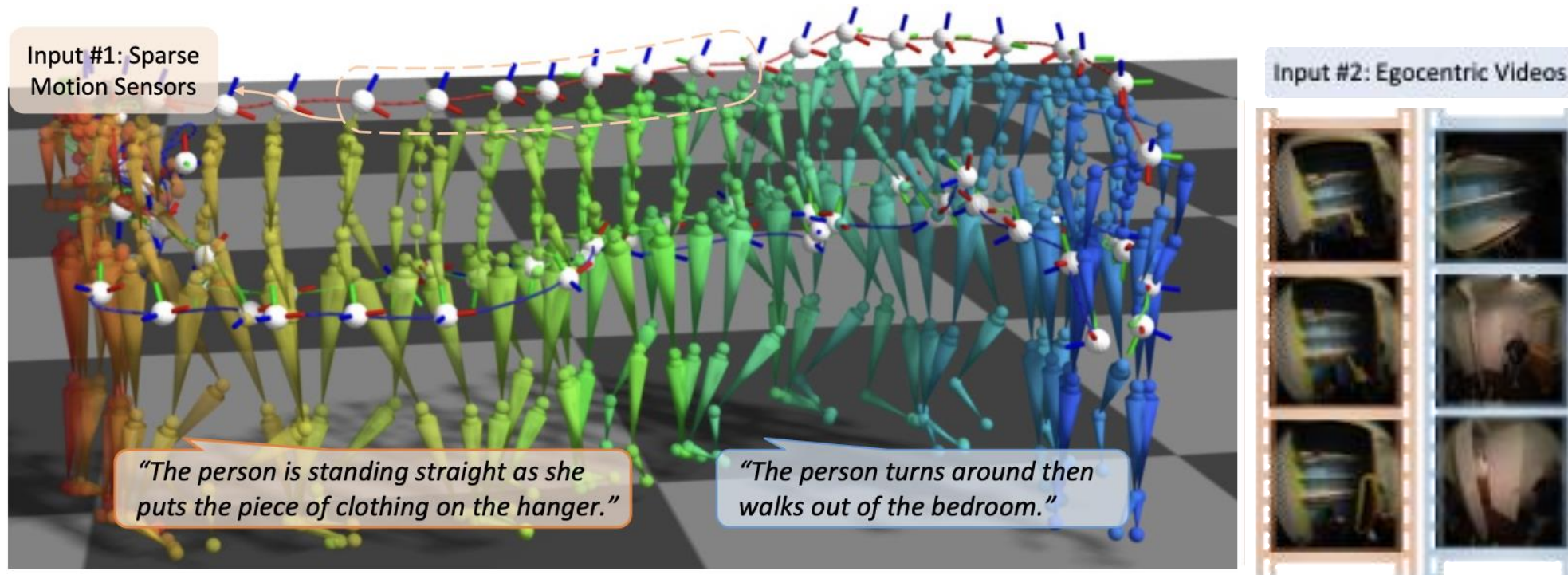
Egocentric Video



Sparse Motion Sensor



# Egocentric Motion Learning



## EgoLM: Multi-Modal Language Model of Egocentric Motions

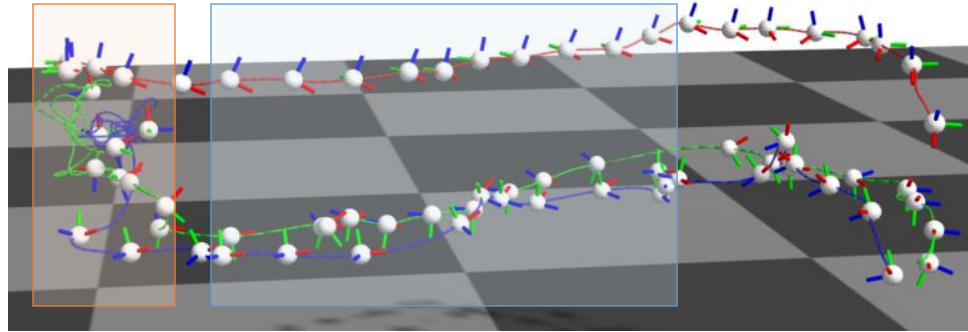
*Fangzhou Hong, Vladimir Guzov, Hyo Jin Kim, Yuting Ye, Richard Newcombe, Ziwei Liu, Lingni Ma*

CVPR 2025 (Oral)



# Egocentric Motion Tracking and Understanding

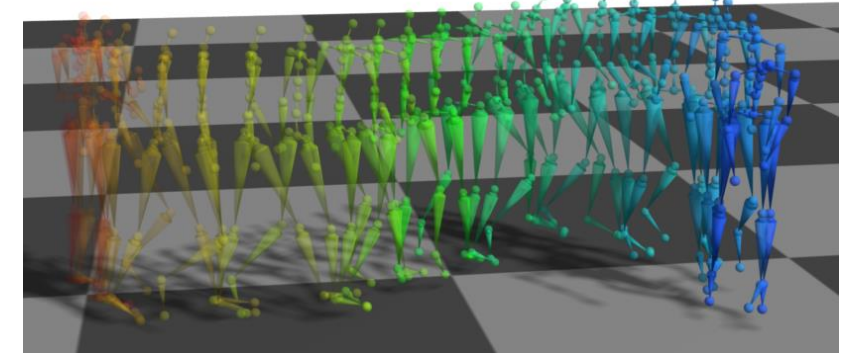
## Sparse Motion Sensors



## Egocentric Videos



## Motion Tracking



## Motion Understanding

*“The person is standing straight as she puts the piece of clothing on the hanger.”*

*“The person turns around then walks out of the bedroom.”*

# Key Challenge 1

## Lack of Direct Observation of the Wearer



Sparse Motion  
Sensors



Lack Lower Body  
Information



Egocentric  
Video



Rarely Visible  
Hands





# Insight 1

## Two Indirect Observations are Complementary



Environment Clues



Body Parts Tracking



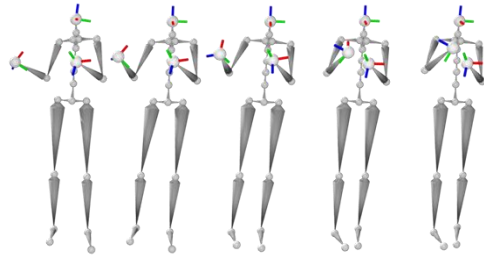
# Key Challenge 2

## Modalities and Tasks with Large Disparities

### Motion Sensors



### Human Motion



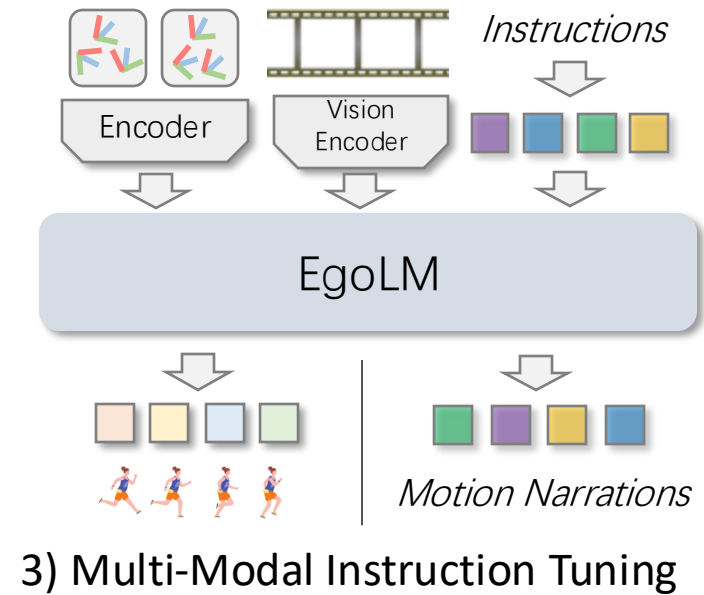
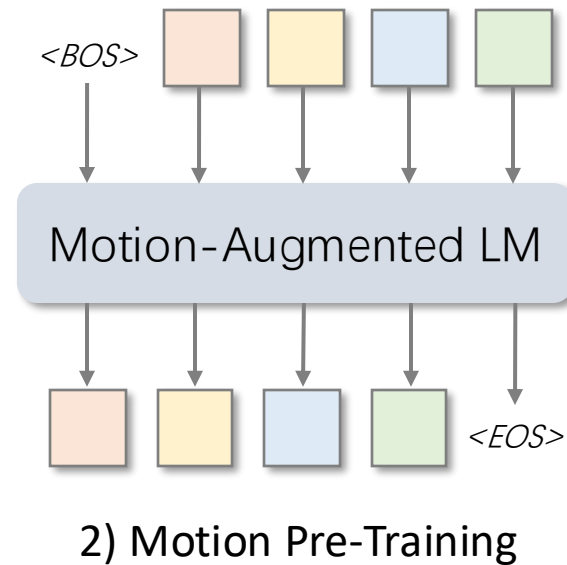
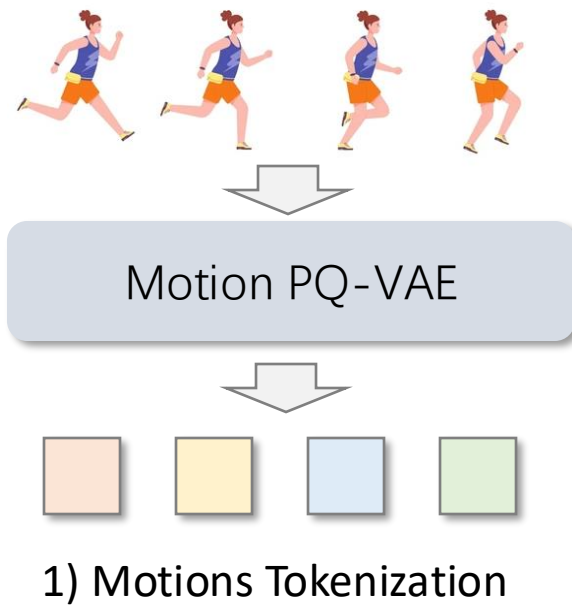
### Egocentric Videos



### Natural Languages

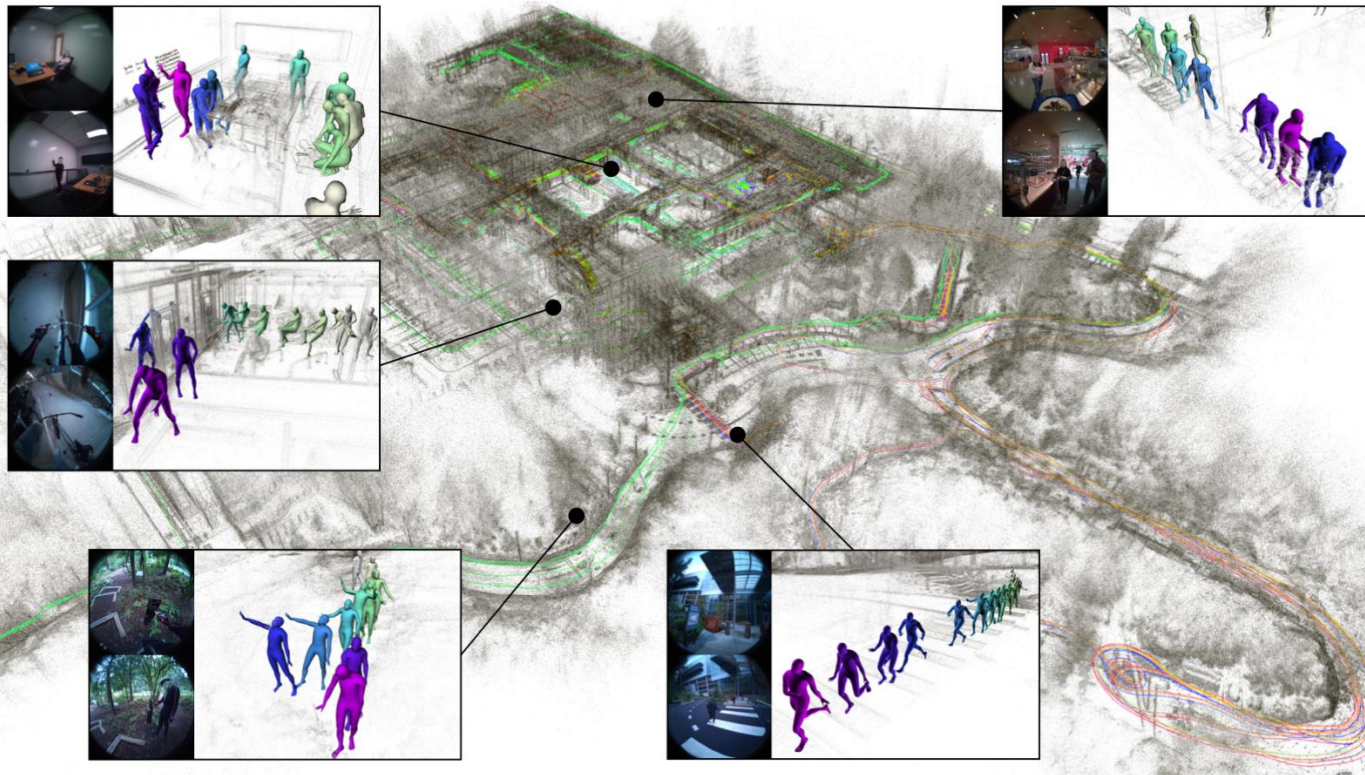
*“The person is standing straight as she puts the piece of clothing on the hanger.”*

# Overview of EgoLM



# Dataset

**Nymeria: A Massive Collection of Multimodal Egocentric Daily Motion in the Wild**, Ma et al., ECCV 2024

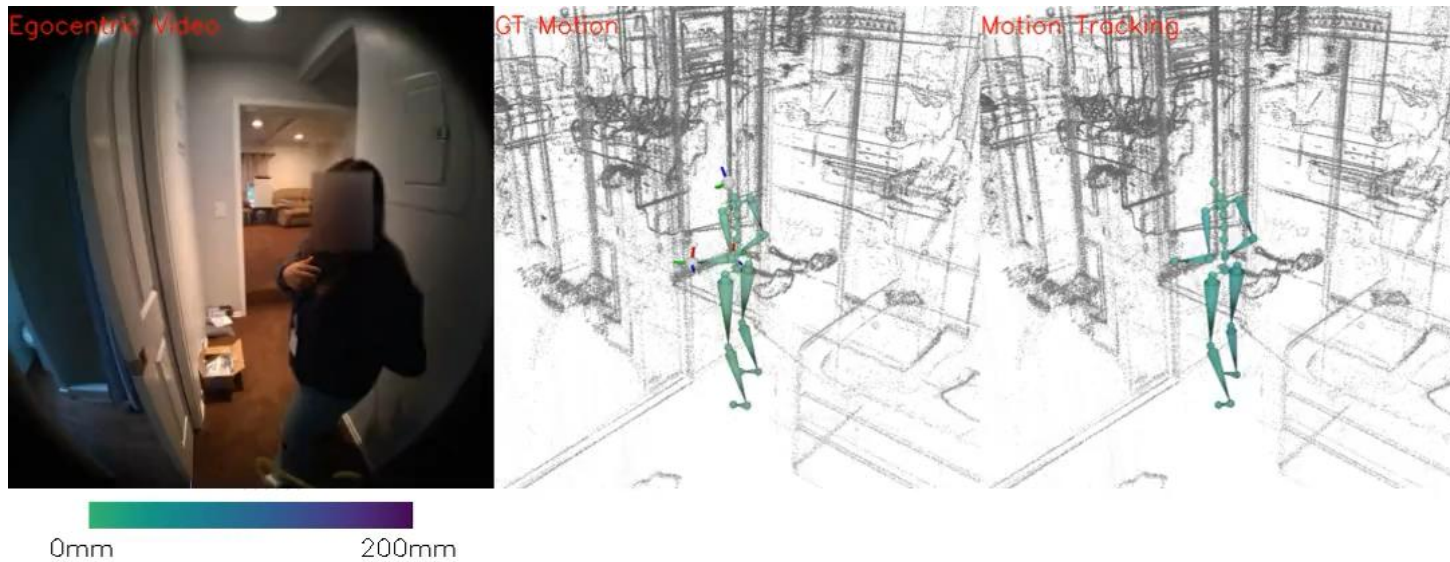


## Dataset Highlights

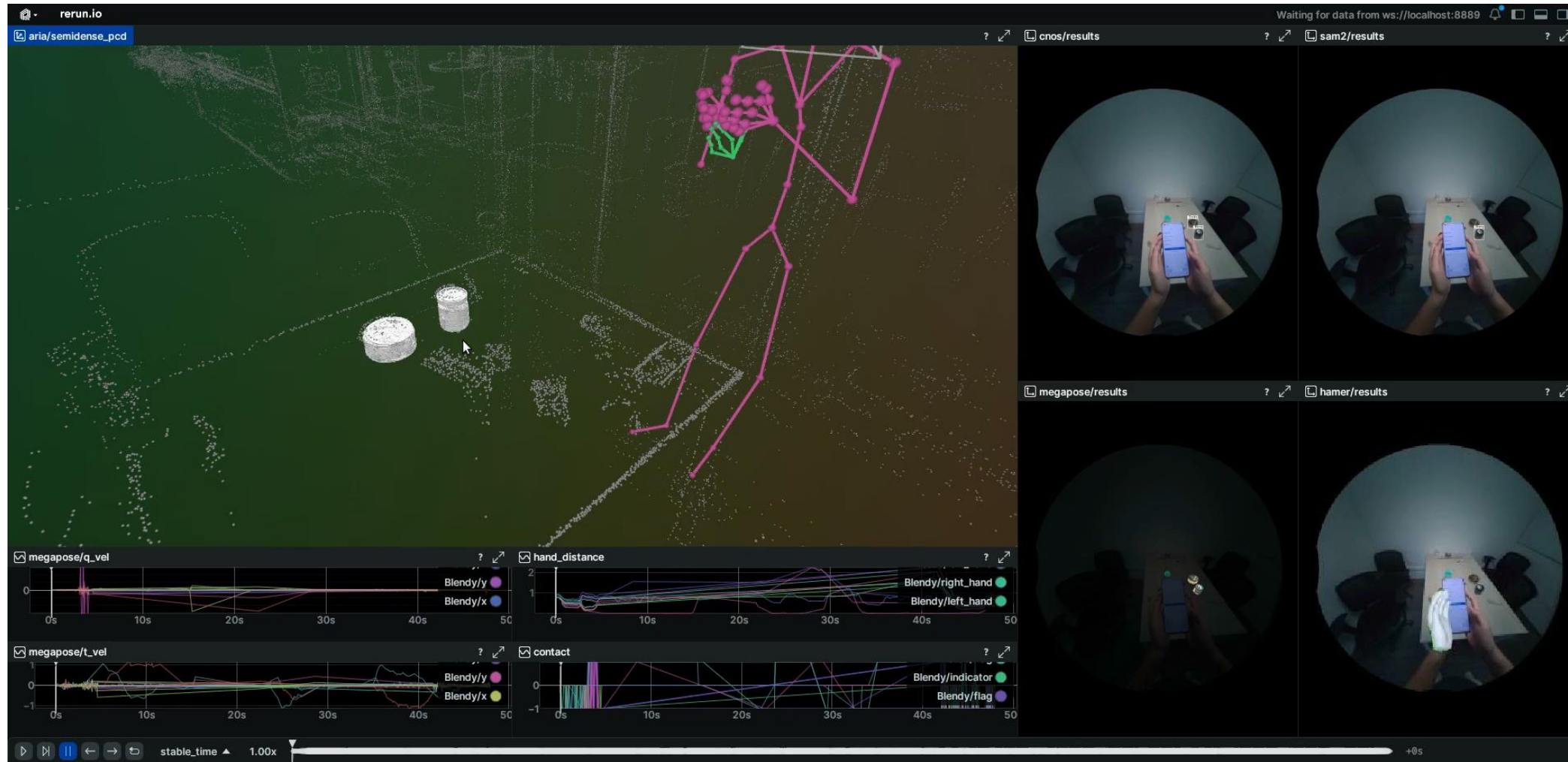
- 300 hours of daily activity
- 3600 hours of video data
- 1200 sequences
- 264 participants
- 50 indoor and outdoor locations
- 20 scenarios
- **230 hours motion with natural language descriptions**
- 310.5 K sentences
- 8.64 M words
- **400 Km traveling trajectory**
- **1053 Km wrist motion**

# Experiments

## Motion Narration from Three-Points and Ego Videos

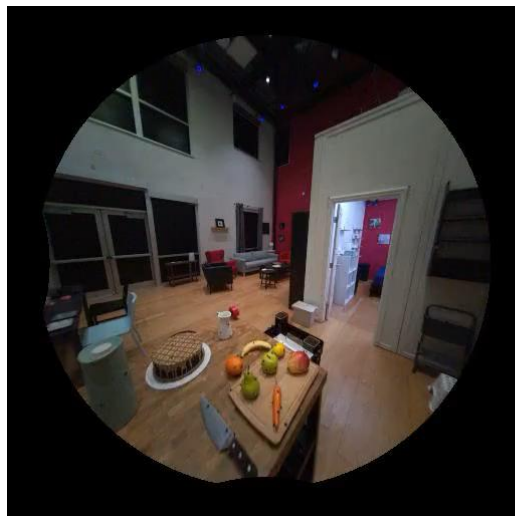
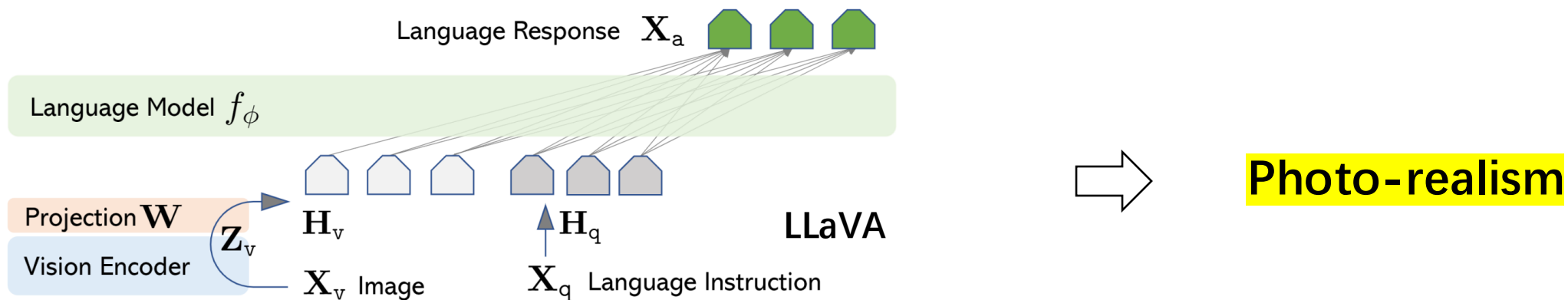


# Extend to More Holistic Capture



# New Challenges

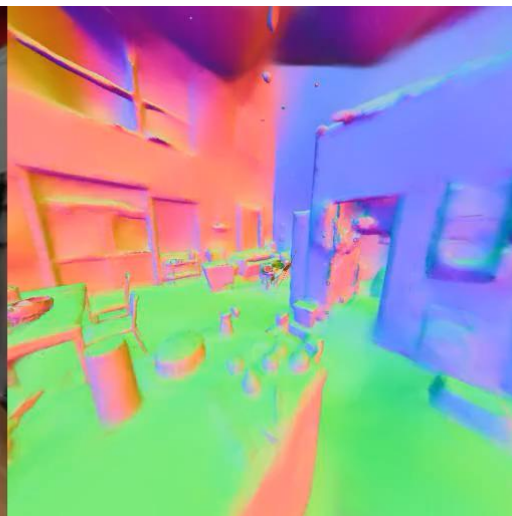
## Egocentric Life Assistant



Input Ego Video



Rendered RGB



Rendered Normal

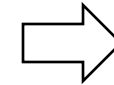


# New Challenges

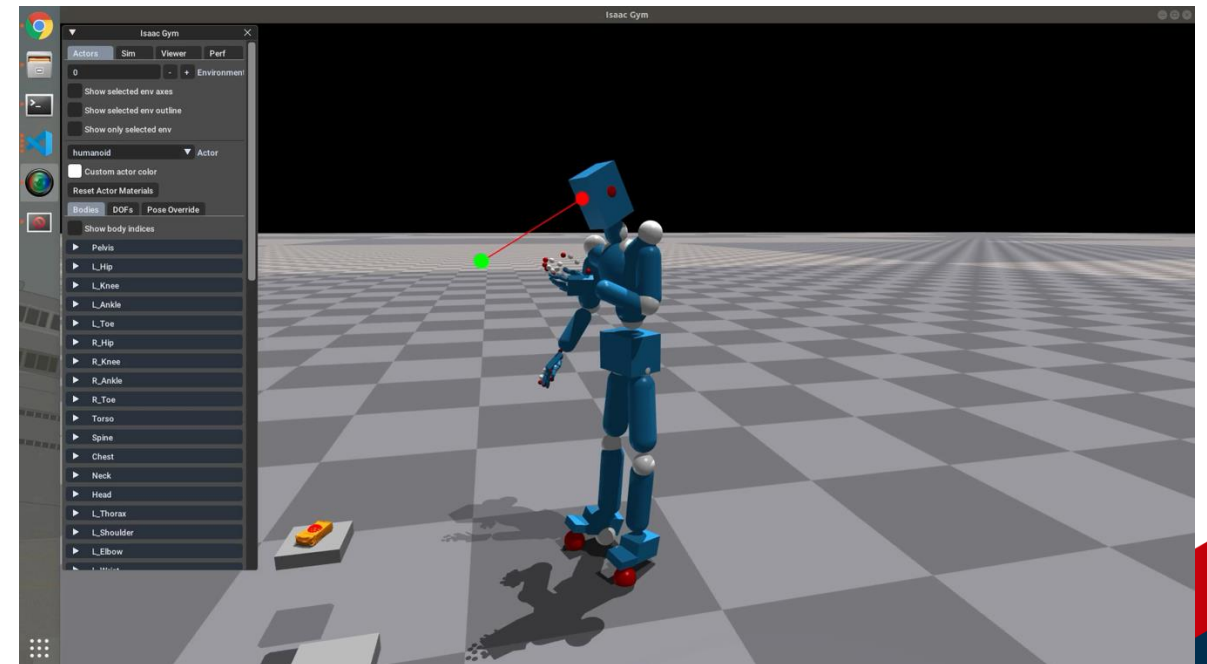
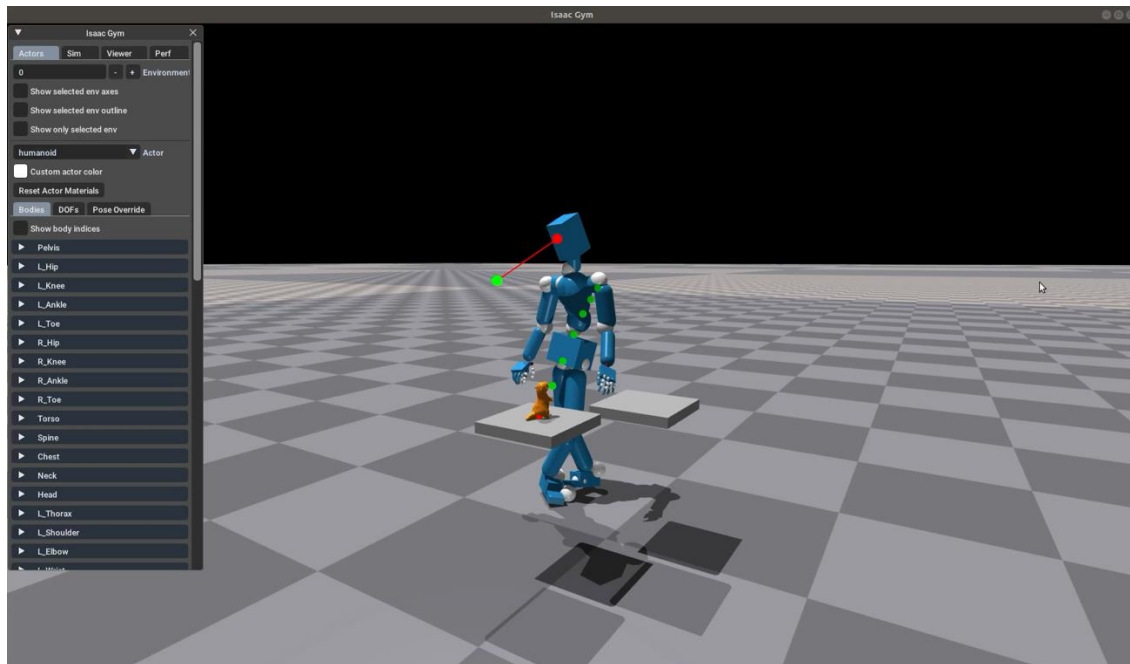
## Embodied AI



DexGraspNet



**Physics Accuracy**





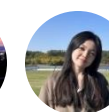


EgoLM  
(CVPR 2025)

EgoLife  
(CVPR 2025)

Nymeria  
(ECCV 2024)

HMD^2  
(3DV 2025)



Thank you!  
Q&A

