# Human Collaboration Dataset for Collaborative Al Assistants in the Real World

parsing

**ETH** zürich and Geometry Lab

Dingxi Zhang<sup>1</sup>, Peiyu Liu<sup>1</sup>, Jingyuan Li<sup>1</sup>, Zhao Huang<sup>1</sup>, Alexey Gavryushin<sup>1\*</sup>, Xi Wang<sup>1\*</sup>

### 1 Introduction

Purpose: Create a large-scale dataset to capture daily collaborative interactions.

**Device:** Aria Glasses

## **Key Features:**

- Simultaneous dual-viewpoint
- Multimodal data
- Fine-grained collaboration

## **Current Dataset Sample Statistics**

Task types: 6

Task scenarios: 3



## Time length: ~110 min **Total Participants: 5**

### Hand-off Detection 1. Build Lego and Coordination Watch and Help bricks **VISOR** 2. Assemble HaMeR monitor arm Assembly Tasks Gesture and Communication Holo Assist Hand mesh & mask Communication Person B Recognition Person A 555 Dish making Partner 1. Ingredients Aria MPS preparation Collaboration service 2. Prepare Point cloud Dynamics Analysis CoELA tableware Cooking Tasks 3. Cooking EPIC-KITCHENS Progress Tracking Use rice cooker to cook rice and Goal Alignment Position tracking Open Ego-Exo4D could you please help Whisper Cut vegetable **Error Prevention** and Correction Packing Tasks IMU Audio

## 2 Background

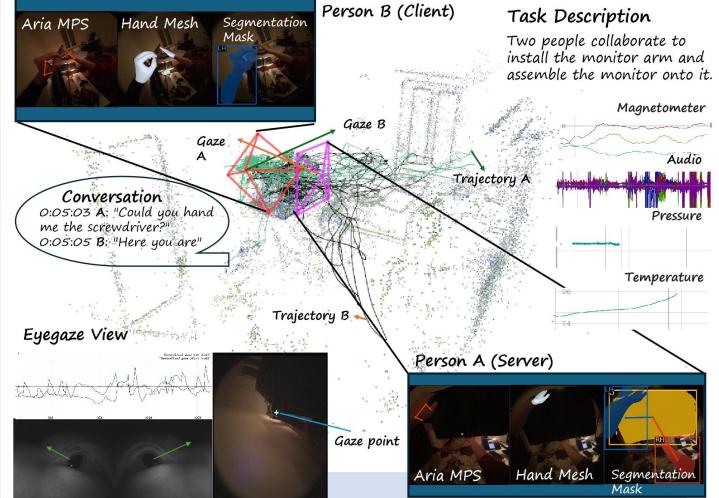
Current datasets like Epic-Kitchen-100, Assembly101, and Ego4D have advanced our understanding of task-oriented behaviors. These include cooking, toy assembly, and daily-life activities. However, they have limitations:

- No multi-agent collaboration contexts
- No real-time verbal communication
- Limited task diversity under shared visual feedback Collaborative datasets like PARTNR and CoELA address some gaps but lack realistic scenarios.

Dataset	Settings	Collaboration	Verbal Interaction	Physical Interaction	Realistic
Epic-Kitchen-100	Cooking	×	×	×	$\checkmark$
Assembly101	Toy assembly	X	$\checkmark$	X	$\checkmark$
Ego4D	Daily-life task	+	+	X	$\checkmark$
Ego-Exo4D	Daily-life task	+	+	X	$\checkmark$
HoloAssist	Assistive task	+	$\checkmark$	X	$\checkmark$
VirtualHome	Household task	×	<b>√</b>	Х	X
ALFRED	Daily-life task	X	$\checkmark$	×	X
WAH	Cooperative task	$\checkmark$	X	X	X
PARTNR	Cooperative task	✓	$\checkmark$	X	X
CoELA	Cooperative task	$\checkmark$	$\checkmark$	X	X
Ours	Cooperative task	✓	✓	✓	✓

Table 1: Comparison of datasets for collaborative and instructional tasks. Note: '+' indicates dataset partially support these attributes.





## 5 Discussion

Our dataset bridges gaps in collaboration studies, offering multimodal recordings for diverse human-human interaction patterns, such as:

- Object handover
- Coordinated tool usage
- Error correction and recovery
- Real-time feedback exchanges
- Role-switching during collaboration
- Verbal negotiation

## References

[1] Wang, Xin, et al. "Holoassist: an egocentric human interaction dataset for interactive ai assistants in the real world." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

[2] Chang, Matthew, et al. "PARTNR: A Benchmark for Planning and Reasoning in Embodied Multi-agent Tasks." arXiv preprint arXiv:2411.00081 (2024).

[3] Pavlakos, Georgios, et al. "Reconstructing hands in 3d with transformers." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern

[4] Zhang, Hongxin, et al. "Building cooperative embodied agents modularly with large language models." arXiv preprint arXiv:2307.02485 (2023).