Towards Physically-grounded Human-AI Collaboration

Xuhui Kang, Yen-Ling Kuo
University of Virginia
{xuhui, ylkuo}@virginia.edu

Abstract

Creating physically-grounded human-AI collaboration remains challenging because of continuous state-action spaces, constrained physical transitions, and diverse human behaviors. Successful collaboration in physical environments requires an agent to generalize their learned policies across three key collaboration modes: *coordination*, where agents must coordinate subtasks or movements and avoid collision; *awareness*, where agents need to recognize when another agent needs help and offer assistance; and *action consistency*, where agents must align their actions toward the same goals when engaging in joint actions. We designed *Moving Out*, a physical human-AI collaboration environment to illustrate these challenges and collaboration modes. We observe that existing AI agents often fail to assist appropriately, align actions, or generalize to unseen physical settings. Our findings suggest future research directions in physical reasoning, behavior adaptation, and reliable and scalable evaluation of human-AI collaboration.

Challenges in Physically-grounded Human-AI Collaboration

Humans can quickly adapt their actions to the physical attributes (e.g., sizes, shapes, weights, etc.) and constraints (e.g., force application, narrow paths, etc.) in an environment when collaborating with others. This ability is crucial for embodied tasks such as assembly and household chores, where human-AI collaboration must account for the increased complexity of the continuous state-action space and dynamics caused by physical constraints. However, existing works (Carroll et al. 2019; Papoudakis et al. 2021; Ng, Liu, and Kennedy 2023; Bard et al. 2020) only focus on human-AI collaboration in discrete or task-level settings, which simplifies interaction dynamics and overlooks the complexities of real-world cooperation. For example, while physical constraints, e.g., narrow passages, restrict movement and require precise coordination, there are still a large number of rotations and ways to hold objects that can lead to successful collaborations. Minor variations in human actions, e.g., lifting angles or applying force, can significantly affect collaboration outcomes.

Figure 1: The Moving Out Environment.

To study how these physical challenges can affect human-AI collaboration, we designed *Moving Out*, a task environment inspired by the game Moving Out (SMG Studio 2020). It is built on top of a 2D physics engine with dynamic physical interactions and diverse collaboration scenarios. As in Fig. 1, it requires two agents to collaboratively move all objects to the goal region (i.e., blue squares) while navigating around obstacles (i.e., walls). The movable objects have different shapes and sizes. A single agent can move small items quickly, but as the object size increases, the agent needs to work together to move an object effectively. Successful collaboration in this kind of physical environment requires an agent to generalize their learned policies across physical properties and constraints.

Potential Physical Collaboration Modes

The continuous space and physical constraints introduce new ways of collaboration. We illustrate the potential modes in Fig. 2 using Moving Out examples. For instance, when organizing a target area, objects must be placed based on their physical properties like size and shape, requiring agents to reason about spatial arrangements. To capture these collaboration modes, we categorize them into three key aspects: *coordination*, where agents must synchronize actions to achieve shared goals; *awareness*, which involves understanding the physical environment and partners' intentions; and *action consistency*, ensuring that movements and forces applied are aligned for smooth cooperation.

Coordination means that agents work together to complete tasks more efficiently, even if one agent could do it alone. For example, in narrow spaces, passing objects may be the only viable option to avoid congestion and en-

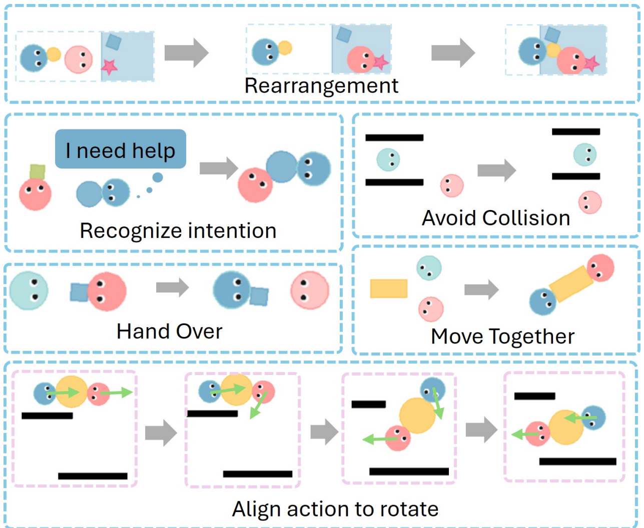

Figure 2: Illustration of various physical collaboration modes using Moving Out, including organizing objects in a limited space (e.g., goal region), recognizing when help is needed, avoiding collisions, passing objects, moving items together, and aligning actions.

sure smooth movement. However, when space is sufficient, agents have more flexibility. They can choose to pass objects or transport items themselves. Both can be viable strategies. AI must learn to decide when to coordinate based on human actions.

Awareness focuses on whether and how an agent recognizes the need to assist its partner. Sometimes, there's no clear best strategy. Agents must decide whether to move small items alone or go to help with the other agent for larger ones. It requires dynamic adjustment of strategies rather than following a fixed action sequence. In human-AI collaboration, this is even more important because humans may have diverse strategies, e.g., starting with small items and then waiting for help with heavy ones, or vice versa. An AI agent with strong awareness can observe behavior cues and adapt its actions to align with human strategies.

Action Consistency checks whether agents align their actions when working toward the same goal. For example, two agents must continuously align their forces to move and rotate large objects through narrow openings or around obstacles. Minor deviations in force or direction can cause items to get stuck or result in wasted effort. By maintaining consistent actions , agents demonstrate a better understanding of physical teamwork. Failure to understand physics can lead to mistakes that are costly to recover, e.g., dropping a rectangular object around a narrow opening of the walls.

Observations from Human-Human vs. Human-AI Rollouts

We tested Moving Out in human-human and human-AI collaboration to show the improvements needed in AI agents to engage in physically-grounded collaborations.

Adaptive Human-Human Strategies

While humans do not work together smoothly at the beginning, they adapt quickly. They can learn to adapt to the new physical configurations created by their partner based on their partner's past behavior. For example, if one player places an item in the middle of the path, the other may recognize this as a request to pass the object. Humans also learn through try-and-error when dealing with physical constraints. The first time they attempt to pass through a narrow space, they may get stuck due to improper angles or positioning against the walls. They quickly adjust their position or angle until they succeed. In future attempts, they remember the correct positioning and apply it smoothly. AI will similarly need to adapt to new physical configurations made by partners and physical constraints.

Failure Cases in Human-AI Collaboration

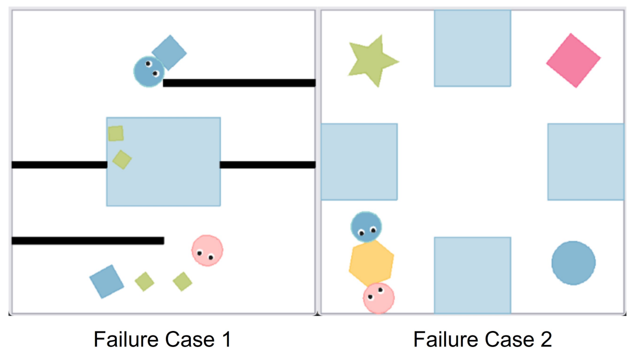

Figure 3: Failure case study: 1) Not responding when assistance is needed, and 2) Inability to grasp large items upon approach.

Existing AI agents such as state-of-the-art behavior cloning agents (e.g., diffusion policy (Chi et al. 2023)) show significant difficulty in collaborating in a physically-grounded environment. Fig. 3 presents two common failure cases. In Case 1, the red agent (AI) fails to recognize that the blue agent (Human) needs help to move the medium square and instead chooses to move a smaller item first. This failure is a result of increased behavior diversity in continuous state-action space. A small variation in human action during the test time can create trajectories that are unseen by AI in training, making it hard for the AI to recognize when the human needs assistance. In Case 2, the AI fails to grasp the object. While the AI has learned to grasp squares and circles in training, it does not know the right positions and angles to pick up a hexagon. This shows that the AI does not generalize the understanding of physical attributes (e.g., similar shapes) from one to another.

Conclusion and Future Directions

In this paper, we discuss several challenges for AI agents to successfully collaborate with humans in physically grounded environments. We show that in a simulated physical environment, Moving Out, current embodied AI agents frequently struggle to offer assistance in time, adapt to unseen situations, and effectively generalize learned skills to diverse and dynamic physical constraints. Addressing these issues is critical for achieving reliable human-AI collaboration in real-world physical tasks.

Providing AI models with the capability to better reason about physical constraints and possible outcomes of collaborative actions is the first step to improving their effectiveness in physically-grounded collaborative tasks. However, repeated and reliable evaluation with real humans is challenging given the variable human behavior and the high costs associated with repeated human experiments. Future directions also include developing automated methods can efficiently and consistently evaluate physically grounded human-AI interactions. Additionally, methods to augment behavior data, utilize agents' prior experiences to better anticipate and model human behaviors, or reason about physical properties and dynamics of objects will also be important to improve AI models' adaptability to diverse human behaviors in physically grounded settings.

References

Bard, N.; Foerster, J. N.; Chandar, S.; Burch, N.; Lanctot, M.; Song, H. F.; Parisotto, E.; Dumoulin, V.; Moitra, S.; Hughes, E.; et al. 2020. The hanabi challenge: A new frontier for ai research. *Artificial Intelligence*, 280: 103216.

Carroll, M.; Shah, R.; Ho, M. K.; Griffiths, T.; Seshia, S.; Abbeel, P.; and Dragan, A. 2019. On the utility of learning about humans for human-ai coordination. *Advances in neural information processing systems*.

Chi, C.; Xu, Z.; Feng, S.; Cousineau, E.; Du, Y.; Burchfiel, B.; Tedrake, R.; and Song, S. 2023. Diffusion policy: Visuomotor policy learning via action diffusion. *The International Journal of Robotics Research*, 02783649241273668.

Ng, E.; Liu, Z.; and Kennedy, M. 2023. It takes two: Learning to plan for human-robot cooperative carrying. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 7526–7532. IEEE.

Papoudakis, G.; Christianos, F.; Schäfer, L.; and Albrecht, S. V. 2021. Benchmarking Multi-Agent Deep Reinforcement Learning Algorithms in Cooperative Tasks. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 1)*.

SMG Studio, D. G. 2020. https://store.steampowered.com/app/996770/Moving_Out/.