Relevance Statement

We would like to state that our work on the Embodiment-Aware Heterogeneous Multi-Robot Operating System (EMOS) aligns closely with the scope of ICLR, particularly in areas like robotics, embodied AI, and benchmarking.

According to statistics about **ICLR accepted papers from 2022 to 2024** (data source is from: https://papercopilot.com/statistics/iclr-statistics/), there has been a growing focus on these topics, which can referred to the figure 1. This figure shows the filters relevant past ICLR papers and visualizes trends in topics such as **Robotics** and **Benchmarks**, verifying the alignment of our work with the increasing focus of ICLR community.

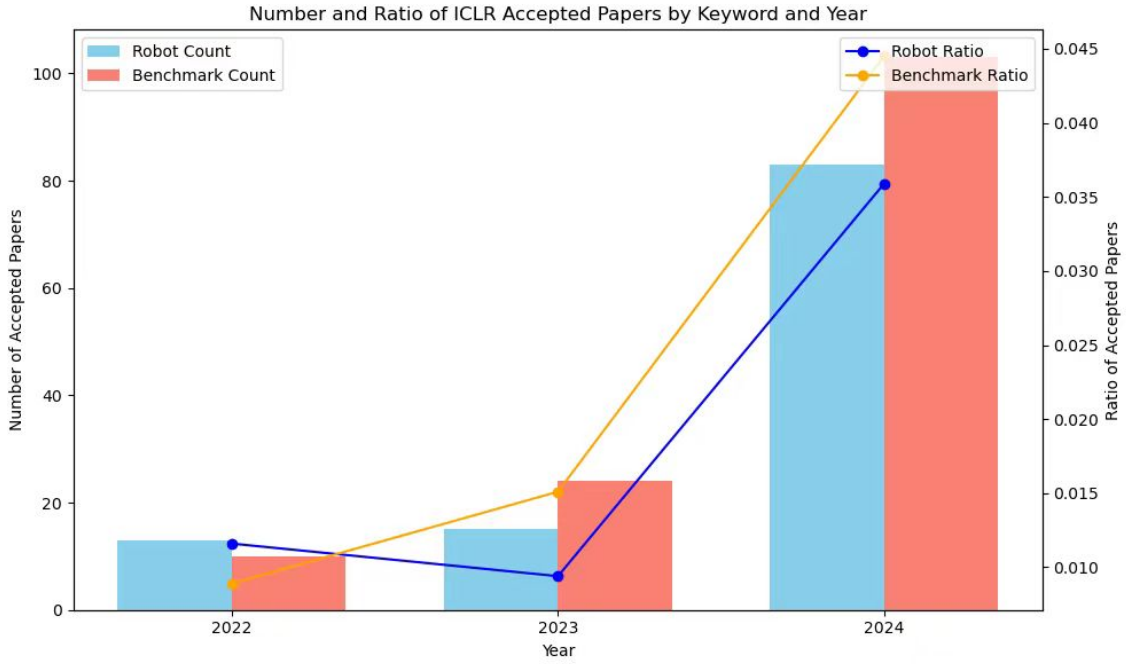

Figure 1. Number and Ratio of ICLR Accepted Papers by Keyword and Year

Then, we will specifically list and analyze some of the works in the past few years at ICLR that are similar to the topic or contributions of our article.Large-scale pretrained models, especially LLMs, trained on internet-scale datasets, bring strong priors and reasoning capabilities to everyday tasks. Many interesting ICLR papers have explored LLM applications for robotic manipulation and other embodied agent tasks. Notable among ICLR 2024:

- **"Programmatically Grounded, Compositionally Generalizable Robotic Manipulation"**, which leverages pre-trained model modularity to aid robotic manipulation.
- **"Building Cooperative Embodied Agents Modularly with Large Language Models"**, which uses LLMs as different modules within multi-agent systems.

- **“Steve-Eye: Equipping LLM-based Embodied Agents with Visual Perception in Open Worlds”**, an end-to-end LLM-based model enabling embodied agents to perceive their environment.
- **“Habitat 3.0: A Co-Habitat for Humans, Avatars, and Robots”** presents a simulation platform for human-robot interaction tasks.

Additionally,

- “**GenSim: Generating Robotic Simulation Tasks via Large Language Models**” applies LLMs to robotic task generation and policy learning.
- “**Vision-Language Foundation Models as Effective Robot Imitators**” fine-tunes large models for robot gripper control, both highly judged as spotlight papers by the ACs.

In the area of agent benchmarking and multi-agent systems, exciting works include:

- **"LoTa-Bench: Benchmarking Language-oriented Task Planners for Embodied Agents"**
- and **"SmartPlay: A Benchmark for LLMs as Intelligent Agents**", all explore various aspects of LLM evaluation.
- Of particular note, the oral paper "**MetaGPT: Meta Programming for A Multi-Agent Collaborative Framework**” introduces an innovative role-play approach for multi-agent collaboration.

Our work aims to identify the problem of *embodiment-aware task planning* that is still missing in the Embodeid-AI community, providing a platform to study this problem and provide an initial method to try to address these challenges in heterogeneous multi-robot systems. While it’s true that a large proportion of our efforts fall in simulation engineering for the benchmark, one of the main motivations is to provide a platform to share with the Embodied AI community to work on this direction of further automation of complex multi-robot systems with a more general problem setting. In the methodology part, by leveraging the priors that LMs hold about robot capabilities parsing from URDF and their decision-making capabilities in complex tasks, we address a practical issue of embodiment-aware understanding innovatively. By building a hierarchical multi-agent system that leverages the synchronized communication and distributed execution, we provide a practical approach that is tailored for real-time multi-robot system settings, and this could serve as a starting point for more comprehensive system development.

In conclusion, our work on EMOS aligns with ICLR's growing focus on robotics and embodied AI, addressing the crucial gap in embodiment-aware task planning for heterogeneous multi-robot systems, while providing a benchmark platform and innovative methodology leveraging large language models.