Poster: Towards Federated Embodied AI with FEAI

Leming Shen, Yuanqing Zheng
The Hong Kong Polytechnic University, Hong Kong SAR, China
leming.shen@connect.polyu.hk,csyqzheng@comp.polyu.edu.hk

ABSTRACT

Embodied AI (EAI) transforms our daily lives by bridging intelligent agents with various sensors and actuators. Large Language Models (LLMs) further enhance EAI agents in environment comprehension, task decomposition, and action execution for robotic manipulation. However, developing a general EAI agent capable of adapting to and continuously learning from diverse operating environments is extremely challenging: 1) Robots with mobility capture environments from multiple perspectives, leading to heterogeneous semantic interpretations, particularly in large or open settings. 2) Heterogeneous environments further exacerbate the variability of decomposed tasks and corresponding actions required for robotic manipulation. To address these challenges, we propose *FEAI*, a novel paradigm to enhance the adaptability and self-learning capabilities of EAI agents in heterogeneous environments via federated embodied learning. Specifically, *FEAI* shares and constructively aggregates environment semantic maps, decomposed task templates, and action-reward rules from federated EAI agents. The aggregated information can further enhance EAI agents' local models through continuous tuning or dynamically updated knowledge databases. We believe that *FEAI* has significant potential to integrate more advanced technologies, further advancing performance and innovation in the field of EAI.

CCS CONCEPTS

• Human-centered computing → Ubiquitous and mobile computing; **• Computer systems organization → Embedded and cyber-physical systems**.

KEYWORDS

Large Language Model, Embodied AI, Federated Learning

ACM Reference Format:
Leming Shen, Yuanqing Zheng. 2025. Poster: Towards Federated Embodied AI with FEAI. In *The 23rd ACM International Conference on Mobile Systems, Applications, and Services (MobiSys '25), June 23–27, 2025, Anaheim, CA, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3711875.3734561

1 INTRODUCTION

Embodied AI (EAI) empowers intelligent agents to interact with the physical world by integrating perception, reasoning, and action [1]. By grounding decision-making in sensory inputs and real-world feedback, EAI systems can perform complex tasks such as robotic manipulation and autonomous navigation, positioning EAI as a cornerstone for context-aware applications in smart homes and automated industries.

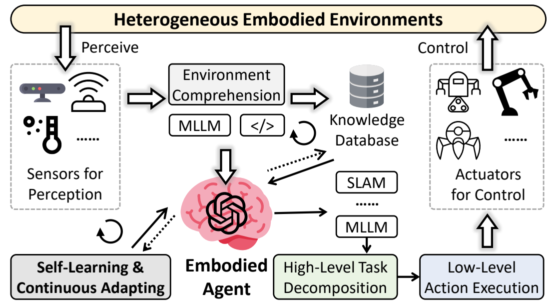

Figure 1: The workflow of an LLM-powered EAI agent.

Large Language Models (LLMs) and multimodal LLMs (MLLMs), serving as the core of modern EAI agents, exhibit exceptional intelligence in environmental understanding via multimodal sensor data [2], as well as in robotic task decomposition and manipulation. As shown in Fig. 1, an LLM-powered EAI agent typically involves three stages: 1) *Environment Comprehension*: Equipped with various IoT sensors, the agent utilizes MLLMs or programs synthesized by LLMs to process multimodal sensor data [3, 4], thereby understanding 3D relations within a given scene; 2) *High-Level Task Decomposition*: Based on the user instruction and environment information – typically represented in textual or embedding forms – the agent utilizes LLMs to decompose the task into multiple manageable subtasks; 3) *Low-Level Action Execution*: For each subtask, the agent invokes various actuators (*e.g.*, robot arms) to perform different operations (*e.g.*, grasping and placing objects). As such, EAI systems gain enhanced capabilities to understand complex instructions and reason over multimodal inputs by combining LLM-driven reasoning with sensory-grounded interaction.

Though armed with zero-shot generalizable LLMs, existing EAI systems struggle to adapt to diverse environments due to two main limitations: 1) *Environmental Heterogeneity*: Mobile robots often operate in expansive environments, observing their surroundings from diverse angles. Leveraging pre-trained Simultaneous Localization and Mapping (SLAM) models or MLLMs for environmental comprehension may be inadequate to adapt to heterogeneous environments; 2) *Task & Action Heterogeneity*: Environmental heterogeneity further amplifies the variability of the decision-making process during task decomposition and action execution for robotic manipulation. For example, to grasp an apple for the user, the decomposed tasks can be "move to the table → find the apple → grab the apple" in one environment, whereas in another, they may be "move to the fridge → open the fridge → find the apple → grab the apple." Since LLMs and the physical world are loosely bridged through intermediate language descriptions or embeddings, relying solely on the general reasoning capabilities of LLMs for embodied decision-making and action execution may be insufficient to address such cascaded heterogeneity.

MobiSys '25, June 23–27, 2025, Anaheim, CA, USA

ACM ISBN 979-8-4007-1453-5/25/06.
https://doi.org/10.1145/3711875.3734561

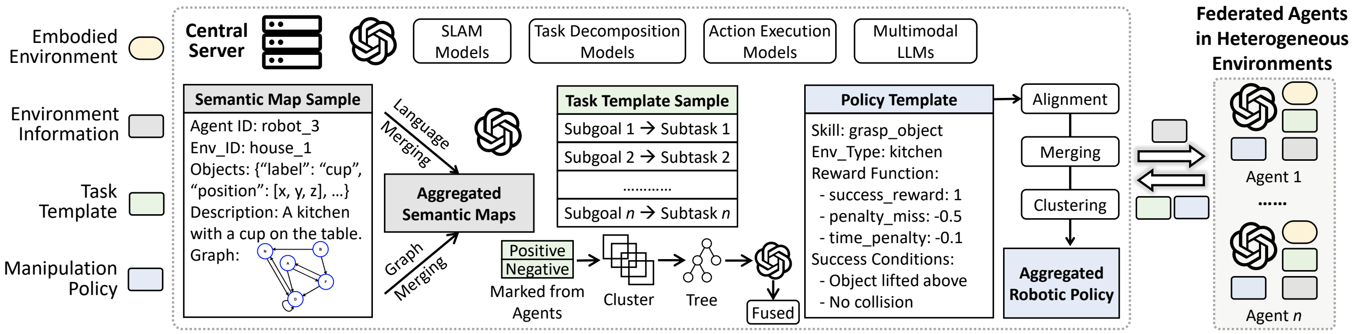

Figure 2: The overall workflow of *FEAI* with a central server and multiple federated EAI agents.

These challenges motivate us to ask: ***Can multiple EAI agents collaboratively enhance each other's adaptability to heterogeneous embodied environments?*** To solve this research question, we propose *FEAI*, a new paradigm that enables federated EAI agents to collaboratively share environmental knowledge and task-planning experiences. By absorbing the aggregated information, *FEAI* facilitates EAI agents to continuously learn and adapt to heterogeneous embodied environments.

2 SYSTEM OVERVIEW – BLUEPRINT

Fig. 2 illustrates the system overview of *FEAI*. Specifically, different from traditional federated learning that trains a global model by sharing its parameters [5], *FEAI* shares and constructively aggregates environment information, task templates, and robot manipulation policies among federated EAI agents.

Environment Sharing. Federated EAI agents can collaboratively enhance their perception and understanding of the physical world by sharing environmental information such as semantic maps and spatial relationships. Instead of transmitting raw sensor data, each agent constructs a structured semantic map – including semantic relations among objects – which is then abstracted and shared with a central server. The server performs graph-based and language-level fusion to align overlapping environments and generate aggregated representations, allowing agents to benefit from each other's experiences with enhanced collective situational awareness.

Task Template Sharing. To support generalizable task planning, federated EAI agents can share abstract representations of how complex instructions are broken down into manageable steps. By exchanging these task templates, agents contribute their experience in translating high-level goals into actionable plans across different environments. The server leverages techniques like clustering and tree structuring to organize and refine the shared templates into generalized patterns. This collaborative sharing allows agents to benefit from diverse real-world experiences, improving their ability to understand and plan for new tasks in unfamiliar settings.

Policy Sharing. Federated EAI agents can also improve their capability to interact with the physical world by sharing manipulation policies, which capture how specific actions are performed and evaluated in different contexts. Each agent contributes high-level descriptions of successful behaviors – such as how to grasp an object or navigate a cluttered space – along with the conditions that define success or failure. These policies are then organized and generalized to create shared guidelines that reflect best practices across environments. By learning from each other's experiences, agents develop more robust and adaptive manipulation strategies that transfer across heterogeneous real-world scenarios.

Local Knowledge Transfer. Federated EAI agents further transfer knowledge from the received aggregated information to their local systems to enhance embodied performance. Two directions can be considered: 1) *Dynamic Knowledge Database*: Agents may store this knowledge in dynamically updated local databases for real-time reference. 2) *Fine-Tuning*: Agents can augment the aggregated information and fine-tune their local models for enhancement. As such, federated EAI agents can achieve ongoing adaptation to diverse environments via database updates or model refinement.

3 CONCLUSION & FUTURE WORK

FEAI is an innovative framework to enhance federated EAI agents' adaptability and self-learning capabilities in heterogeneous environments. By sharing aggregated environment information and action policies, *FEAI* enables EAI agents to learn from diverse real-world contexts collaboratively.

Looking ahead, four directions can expand the vision of *FEAI*. First, advancing *aggregation strategies* is critical. We should explore more fine-grained, context-aware methods for merging environment maps, task templates, and action policies, enabling agents to reconcile diverse knowledge with minimal conflict. Second, improving *scalability* is essential to support large networks of heterogeneous agents operating under varying resource constraints. Third, ensuring *privacy protection* in knowledge sharing is vital for real-world deployment, including detecting and mitigating noisy, malicious, or biased contributions. Lastly, developing *standardized benchmarks and simulation platforms* for federated embodied learning helps catalyze progress, allowing researchers to evaluate in controlled yet realistic settings. These directions point toward a future where EAI agents evolve together – continually learning, adapting, and collaborating at scale.

ACKNOWLEDGMENTS

We sincerely thank our anonymous reviewers for their constructive comments and invaluable suggestions that helped improve this paper. This work is supported by Hong Kong GRF Grant No. 15211924 and 15206123. Yuanqing Zheng is the corresponding author.

REFERENCES

[1] J. Duan, S. Yu, H. L. Tan, H. Zhu, and C. Tan, "A survey of embodied ai: From simulators to research tasks," *IEEE TETCI*, 2022.
[2] H. Xu, L. Han, Q. Yang, M. Li, and M. Srivastava, "Penetrative ai: Making llms comprehend the physical world," in *ACM HOTMOBILE*, 2024.
[3] L. Shen, Q. Yang, Y. Zheng, and M. Li, "Autoiot: Llm-driven automated natural language programming for aiot applications," in *ACM MobiCom*, 2025.
[4] L. Shen, Q. Yang, X. Huang, Z. Ma, and Y. Zheng, "Gpiot: Tailoring small language models for iot program synthesis and development," in *ACM SenSys*, 2025.
[5] L. Shen, Q. Yang, K. Cui, Y. Zheng, X.-Y. Wei, J. Liu, and J. Han, "Fedconv: A learning-on-model paradigm for heterogeneous federated clients," in *ACM MobiSys*, 2024.