Workshop on Embodied Intelligence with Large Language Models In Open City Environment

1 Workshop Summary

This workshop is motivated by a fact: human beings have strong embodied intelligence in an open environment, but it is still challenging for large language models and LLM agents. Despite some progress on embodied AI in static and indoor environments, the LLM agents still struggle with tasks in large-scale outdoor environments, such as navigation, search, spatial reasoning, task planning, etc. Therefore, we propose this workshop to discuss the recent advances in the related research area and look forward to future development. Specifically, it delves into topics of outdoor embodied intelligence, such as spatial intelligence and embodied perception, reasoning and planning, decision-making and action, multi-agent and human-agent collaboration, and the development of simulators, testbeds, datasets, and benchmarks. This comprehensive exploration of embodied LLM agents in an open city environment holds the potential to advance the field of artificial intelligence and open up new applications in various domains. We also have a special poster/short paper session for those solutions that perform best in the Open Urban Environment Embodied Intelligence Competition.

We would like to discuss the following topics in this workshop:

(1) Spatial intelligence and embodied perception with LLM agents in open city environment:

a) How LLM agents can develop a sense of space and time in open city environments.

b) The role of embodied perception in enhancing the performance of LLM agents in outdoor environment.

c) Techniques for integrating spatial intelligence and embodied perception for LLM agents in outdoor environment.

d) Other related topics.

(2) Reasoning and planning with LLM agents in open city environment:

a) How LLM agents can use reasoning to make decisions in open city environment.

b) Strategies for planning actions and sequences of tasks for LLM agents in city environment.

c) Analysis on the bias and limitations of reasoning and planning of LLM in city environment.

d) Other related topics.

(3) Decision-making and Action with LLM agents in open city environment:

a) How LLM agents can make decisions based on outdoor context and goals.

b) Combination of large language models and small machine learning models for decision-making in outdoor environment.

c) Techniques for evaluating and improving the decision-making and action capabilities of LLM agents in outdoor environment.

d) Other related topics.

(4) Multi-agent and human-agent collaboration in open city environment:

a) How multiple LLM agents can collaborate to achieve common goals in outdoor environment.

b) The challenges and opportunities of human-agent collaboration in open city environment.

c) Strategies for designing effective multi-agent systems in open city environment

d) Perspectives on human-AI system for outdoor applications.

e) Other related topics.

(5) Simulator, testbeds, datasets, benchmark for embodied LLM agent in city environment:

a) The development and use of simulators and testbeds for evaluating embodied LLM agents in outdoor environment.

b) The creation and curation of datasets for training and testing embodied LLM agents in outdoor environment.

c) The establishment of benchmarks and evaluation metrics for embodied LLM agents in outdoor environment.

d) Other related topics.

(6) Short paper / poster presentation for winners of Open Urban Environment Embodied Intelligence Competition: The organizer team have recently launched the competition with the benchmark platform named EmbodiedCity (https://embodied-city.fiblab.net/), a realistic, dynamic, and open platform for embodied intelligence, aimed at evaluating and researching its applications in complex urban scenarios. The benchmark consists of two components: a simulator and a dataset. The simulator is based on a large business district in Beijing, one of the largest cities in China, with 3D models of buildings, streets, greenery, statues, traffic flow, and other elements built using Unreal Engine 5. On the dataset side, we provide data and code for tasks related to multimodal scene understanding, question answering, dialogue, navigation, and task planning. The platform supports both online access and offline operation.

The competition on the EmbodiedCity platform focused on "Intelligent Urban Navigation and Interaction with LLM Agents". Participants will be challenged to develop and optimize embodied agents for efficient navigation and multimodal interaction in complex urban environments. Tasks include, but are not limited to:

a) Urban navigation by aerial/ground agents,

b) Multi-agent natural language dialogue,

c) Completing designated tasks in diverse urban scenarios.

Outstanding participants will be featured in a special poster session showcasing the most advanced urban embodied agent designs and corresponding insights. The are also encouraged to submit the solution via a short paper (will be peer-reviewed before acceptance). The platform does not needs huge compuation power for partipants, which accords to one of original purposes: to encourage researchers who lack the computing power to publish long papers.

2 Tentative schedule

(1) Six Invited talk
(2) Four oral presentations for accepted long paper
(3) One poster session for accepted short paper

(4) One poster session for those solutions that perform best in the Open Urban Environment Embodied Intelligence Competition
(5) One panel session (invited speaker and additionally invited panelists)

- 9:00 AM to 9:10 AM: Opening remarks
- 9:10 AM to 9:30 AM: Invited talk 1 (20 min)
- 9:40 AM to 10:00 AM: Invited talk 2 (20 min)
- 10:10 AM to 10:25 AM: Oral presentation 1 (15 min)
- 10:25 AM to 10:40 AM: Oral presentation 2 (15 min)
- 10:45 AM to 11:45 AM: Poster session for best solutions & coffee socials (1 h)
- 11:45 AM to 1:00 PM: Lunch break
- 1:00 PM to 1:20 PM: Invited talk 3 (20 min)
- 1:30 PM to 1:50 PM: Invited talk 4 (20 min)
- 2:00 PM to 2:15 PM: Oral presentation 3 (15 min)
- 2:15 PM to 2:30 PM: Oral presentation 4 (15 min)
- 2:35 PM to 3:35 PM: Poster session for short paper & coffee socials 2 (1h)
- 3:40 PM to 4:00 PM: Invited talk 5 (20 min)
- 4:10 PM to 4:30 PM: Invited talk 6 (20 min)
- 4:40 PM to 5:10 PM: Panel discussion (30 min)
- 5:10 PM to 5:25 PM: Awards and conclusive remarks

3 Organizers and Biographies

(1) Chen Gao is now an Assistant Professor (research-track) of Tsinghua University. His research primarily focuses on large language model, embodied agent, etc., with over 60 papers in top-tier venues, including ICLR, NeurIPS, KDD, Nature Communications, etc., attracting over 3,800 citations. His work on large language model agents won the ACL 2024 Outstanding Paper Award. He is at the list of Elsevier/Stanford Top 2% global scientists 2024.
(2) Yitao Liang is Assistant Professor of Peking University. His research interests span knowledge reasoning, machine learning and AI agents. His work has received recognition from top AI conferences; for example, the best-paper honorable mention from AAMAS 2016, the best paper from RL for Real Life workshop in ICML 2019, a best paper runner-up from the LLD workshop in NeurIPS 2017, a best paper from the TEACH workshop in ICML2023. He regularly serves as area chairs and senior area chairs in top venues e.g., NeurIPS, ICML.
(3) Xin Wang is currently an Associate Professor at the Department of Computer Science and Technology, Tsinghua University, China. His research interests include machine intelligence, media big data analysis, machine learning and its applications.
(4) Yu Zheng is am currently a postdoctoral fellow in Massachusetts Institute of Technology, USA. His research interests lie in the interdisciplinary area of artificial intelligence and urban science, and its applications to real-world complex systems. His work is selected as cover article of Nature Computational Science.
(5) Tong Xia is a postdoctoral research associate at the University of Cambridge. Her research interests include machine learning and trustworthy AI. She has published in prestigious venues such as NPJ Digital Medicine, IEEE JBHI, NeurIPS, KDD and AAAI. She has been nominated as a

Rising Star Women in Engineering at Asian Deans' Forum 2024. She co-organized the FairComp Workshop at UbiComp 2023, and she has been co-organizing the Mobile and Wearable Health Seminar Series at the University of Cambridge since 2023.

(6) Fengli Xu is an Assistant Professor at the Department of Electronic Engineering at Tsinghua University. His research explores the interdisciplinary realms of Artificial Intelligence, Data Science, Social Computing and Urban Science, aiming to develop scientific methods and algorithmic tools to address the long-standing puzzles in the complex networks arising from human dynamics and social interactions. Recently, he is particularly interested in the emergence of role-playing and commonsense reasoning capabilities in Large Language Models (LLMs).

(7) Yong Li is a Tenured Full Professor at the Department of Electronic Engineering at Tsinghua University. His work is published in top venues including NeurIPS, ICLR, KDD, Nature Machine Intelligence, Nature Computational Science, Nature Human Behavior, Nature Communications, etc., with over 25,000 citations.

4 Invited speakers

(1) Tat-Seng Chua (National University of Singapore)

Dr Chua is the KITHCT Chair Professor at the School of Computing, National University of Singapore. He is also the distinguish Visiting Professor of Tsinghua. Dr Chua was the Founding Dean of the School from 1998-2000. His main interests are in multimedia information retrieval and social media analytics. In particular, his research focuses on the extraction, retrieval and question-answering of text, video and live media arising from the Web and social networks. He is the Director of a joint research Center between NUS and Tsinghua (NExT) to research into big unstructured multi-source multimodal data analytics.

(2) Guy Van den Broeck (UCLA)

He is a Professor of Computer Science and Samueli Fellow at UCLA, where I direct the Statistical and Relational Artificial Intelligence (StarAI) lab. His research interests are in Machine Learning (Tractable Deep Generative Models, Statistical Relational Learning, Probabilistic Programming), Knowledge Representation and Reasoning (Probabilistic Inference, Probabilistic Databases), and Artificial Intelligence in general.

(3) Xing Xie (Microsoft)

Dr. Xing Xie is a partner research manager at Microsoft Research Asia. He received his B.S. and Ph.D. in Computer Science from the University of Science and Technology of China in 1996 and 2001, respectively. Since joining Microsoft Research Asia in July 2001, Dr. Xie has focused on data mining, social computing, and responsible AI. His work has been recognized with several prestigious awards, including the IEEE MDM 2023 Test-of-Time Award, ACM SIGKDD 2022 Test-of-Time Award, ACM SIGKDD China 2021 Test-of-Time Award, ACM SIGSPATIAL 2020 10-Year Impact Award Honorable Mention, and ACM SIGSPATIAL 2019 10-Year Impact Award. He has delivered keynote speeches at notable conferences such as MDM 2019, ASONAM 2017, and W2GIS 2011. He served as program co-chair of ACM Ubicomp 2011, PCC 2012, UIC 2015, SMP 2017, ACM SIGSPATIAL 2021, ACM SIGSPATIAL 2022, IEEE MDM 2022, PAKDD 2024, and IEEE BigData 2025. Dr. Xie is a Fellow of the ACM and IEEE.

(4) Qianru Sun (Singapore Management University)

She is an Associate Professor of Computer Science in Singapore Management University. Her recent research is mainly about multi-modal LLMs and their applications. She was a research fellow at

National University of Singapore and MPI for Informatics, focusing on interesting machine learning problems such as few-shot learning, meta-learning and continual learning and their implementations on computer vision tasks.

(5) Stella Christie (Tsinghua University)

She is a tenured full professor for the department of psychology at Tsinghua University, a research chair of Tsinghua laboratory of brain and intelligence, and the director of the child cognition center of Tsinghua laboratory of brain and intelligence. The goal of her research is to chart this fundamental aspect of cognition—the relational mind—in three lines of investigation: (1) The initial state: what are the basic relations that humans and/or other animals possess? (2) The learning tools: how do we learn to become relational thinkers? (3) An application—the social relational mind. How do we acquire and process complex social relations?

(6) Yaoyao Liu (UIUC)

He is an Assistant Professor in the School of Information Sciences at the University of Illinois Urbana-Champaign. His research lies at the intersection of computer vision and machine learning – with a special focus on building intelligent visual systems that are continual and data-efficient. His research interests include continual learning, few-shot learning, semi-supervised learning, generative models, 3D geometry models, and medical imaging.

5 Anticipated audience size

Based on our experience in hosting workshops, our attendance experience in past ICLR conferences, and the attractiveness of the speakers, we expect the workshop to host 150-200 people in the room at all times and cover 300-400 audiences in total throughout the event.

6 Plan to get an audience for a workshop (advertising, reaching out, etc.)

(1) We are going to establish the official social media account (X, Facebook, etc.), along with the official website, to help reach out the audience.
(2) We will continuously advertise the workshop in related venues including NeurIPS 2024, etc.
(3) The call-for-paper message will be broadly sent via emails.

7 Diversity commitment

We strive to achieve balance and parity across multiple levels–from seniority levels, field/community, position/location, to personal backgrounds for invited speakers, organizing team, covered topics, and targeted audiences.

(1) Seniority level

The organization team has spanned the full range of seniority from full professor to postdoctoral researcher. The invited speakers/panelists span the full range of seniority from the assistant professor to the full professor level.

(2) Field/community

The organization team and the invited speaker/panelists cover the various community from computer science, artificial intelligence to psychology and behavioral science (Stella). As for computer science and artificial intelligence, the research area is also very diverse, covering from the

machine learning theory, computer vision, large language model, data mining, machine learning engineering, etc.

(3) Position/location and personal backgrounds for invited speaker

We have organizers and invited speakers from different continents, speaking different languages. Our female organizers and invited speakers covers from different seniority level (postdoc, assistant professor, and professor), which we believe well encourage different background female partipants to join our workshop. We have also speaker from industry (Microsoft).

(4) Covered topics and targeted audiences

The invited talks are from diverse speakers, and it naturally attract diverse audiences. In addition, in our call-for-paper, we setup multiple topics related to the embodied intelligence in open city environment.

8 Virtual access to workshop materials and outcome

We build one official website: embodied-city-workshop.fib-lab.net, which help easily access all related materials.

9 Previous related workshops

There is one related workshop: Workshop on Open-World Agents (OWA-2024) in NeurIPS 2024, of which we have one overlapped organizer: Litao Liang. But, however, our proposed workshop is largely differnet from multiple aspects. First, the proposed workshop focues on the open city environment while "open" in OWA-2024 only refers to open-ended tasks. Second, the proposed workshop pay attention to embodied intelligence and large language model while OWA-2024 is different.