

## **Transforming Large Language Models into Interactive Agents in the 3D Visual World**

**Liwei WANG**

**Department of Computer Science and Engineering, The Chinese University of Hong Kong, China**

**Email: [lwwang@cse.cuhk.edu.hk](mailto:lwwang@cse.cuhk.edu.hk)**

The rapid progression of Large Language Models (LLMs) has led to significant advancements in numerous language-related tasks. However, human intelligence extends beyond mere language understanding and expression, and is not solely derived from static textual resources. Indeed, humans acquire knowledge through engaging with their surrounding environment, particularly by visually perceiving the three-dimensional world. To address this challenge, it is essential to explore ways in which large language models can broaden their competencies. In this presentation, the speaker will discuss three key components that his research team is investigating to transform large language models into more powerful interactive agents: 1) Enhancing the training of large AI models by integrating visual and linguistic modalities; 2) Devising strategies to augment the reasoning capacities of large models, and extending such proficiencies to interactions within the 3D visual world; 3) Establishing a more comprehensive framework for understanding and evaluating the capabilities of large models.