Zhimu Zhou

180-8271-7992 2022zhimu.zhou@bupt.edu.cn
https://pikonguwu.github.io/academicpages.github.io/

Education

Beijing University of Posts and Telecommunications 2022.09 – Present
Internet of Things Engineering

- **GPA**: 3.82/4 (Rank: 2/182) **Weighted Average**: 91.29/100 **English Proficiency**: CET-6: 627

Honors & Awards

(National No 1)	IC Innovation Contest	MIIT	2024
(Only 1 in Major)	Xiaomi Scholarship	BUPT	2024
(Major Top 10%)	Second-Class Scholarship	BUPT	2023, 2024
(Provincial Top 10%)	First Prize, China Innovation and Entrepreneurship Competition (Beijing)	Ministry of Education	2024
(Provincial Top 15%)	Second Prize, Computer Design Competition	Ministry of Education	2024

Research Experience

VLA Strategy Guidance Based on the Process Reward Model
Project Leader 2024.09 – 2025.02

- Inspired by the verification effectiveness of NP problems, proposed an independent **process reward model** to provide dense guidance for VLA models to avoid ineffective exploration.
- Designed the overall architecture and verified through virtual simulation experiments.

Zero-Shot Vision-and-Language Navigation with Dynamic Memory and LLM Spatial Reasoning
Completed 2025.01 – 2025.05

MSNav: Zero-Shot Vision-and-Language Navigation with Dynamic Memory and feature enhancement
Chenghao Liu*, **Zhimu Zhou***, Jiachen Zhang, Minghao Zhang, Songfang Huang, Huiling Duan.
Under Review, ***Co-first Author, randomly ordered by dice rolling,** **Preprint: arxiv.org/pdf/2508.16654**

- The context window of traditional intelligent agents can only passively receive all the information and cannot actively filter out the relevant information for tasks to form task memory.
- Proposed MSNav, a modular framework with **dynamic topological memory** and **spatial capabilities**.
- **Personal Contribution**: All experimental implementation and part of the paper writing.

TTF-VLA: Temporal Token Fusion via Pixel-Attention Integration for Vision-Language-Action Models
Completed 2025.05 – 2025.07

TTF-VLA: Temporal Token Fusion via Pixel-Attention Integration for Vision-Language-Action Models
Chenghao Liu, Jiachen Zhang, Chengxuan Li, **Zhimu Zhou**, Songfang Huang, Huiling Duan.
Under Review, **Fourth Author**, **Preprint: arxiv.org/pdf/2508.19257**

- Propose a method of **reuse of visual features** can enhance the basic performance of the model **without incurring additional training costs**.
- **Personal Contribution**: Consider the application of the locality principle in the VLA field, validated the feasibility of visual reuse in VLA models through experiments.

Chasing the Silver Bullet: An Inquiry into the Potential and Limits of Vibe-Coding
Project Leader 2025.08 – Present

- "Silver Bullet" effect in software engineering **: the linear expressive growth of natural language cannot match the non-linear growth of project complexity.**
- Independently designed and developed the **"Vibe Entropy" evaluation system** to explore the optimal theoretical boundaries of Vibe-Coding.