Anqi Li

andyhandsom6.github.io | lianqi0417@stu.pku.edu.cn

Education and Research Experience

Peking University Sep 2023 - present

B.S. in Intelligent Science and Technology & B.B.A in Management
Advisor: Prof. He Wang; GPA: 3.70/4.00

Galbot Feb 2025 - present

Research Intern, Department of Large Models
Working on multimodal foundation models for embodied navigation and manipulation.

Publications and Preprints

UrbanVLA: A Vision-Language-Action Model for Urban Micromobility *(preprint)*
Anqi Li*, Zhiyong Wang*, Jiazhao Zhang*, Minghan Li, Yunpeng Qi, Zhibo Chen, Zhizheng Zhang†, He Wang†.

- Focusing on **urban micromobility**, UrbanVLA aligns noisy navigation-tool routes with visual observations to enable **scalable, long-horizon** urban navigation.
- Through a **two-stage training pipeline** including supervised fine-tuning and reinforced fine-tuning, UrbanVLA demonstrates the ability to operate **safely** and in compliance with **social norms**.

Embodied Navigation Foundation Model *(preprint)*
Jiazhao Zhang*, **Anqi Li***, Yunpeng Qi*, Minghan Li*, Jiahang Liu, Shaoan Wang, Haoran Liu, Gengze Zhou, Yuze Wu, Xingxing Li, Yuxin Fan, Wenjun Li, Zhibo Chen, Fei Gao, Qi Wu, Zhizheng Zhang†, He Wang†.

- A **cross-embodiment and cross-task** navigation foundation model encompassing quadrupeds, drones, wheeled robots, and vehicles, spanning tasks including VLN, ObjectNav, tracking, and self-driving.

TrackVLA: Embodied Visual Tracking in the Wild *(CoRL 2025)*
Shaoan Wang*, Jiazhao Zhang*, Minghan Li, Jiahang Liu, **Anqi Li**, Kui Wu, Fangwei Zhong, Junzhi Yu, Zhizheng Zhang†, He Wang†.

- Simultaneously conducting **object recognition** and **visual tracking**, TrackVLA demonstrates robust tracking, long-horizon tracking and cross-domain generalization.

Skills and Interests

Technical Skills:

Programming: Python, PyTorch, C/C++, Cuda, LaTeX
Languages: English (TOEFL: 111), Chinese (Native)

Research Interests:

My long-term research goal is to build intelligent robotic systems that operate reliably in complex, human-centered environments. Currently, I focus on **embodied navigation** with **vision-language-action (VLA)** models, exploring how general intelligence enables robust and adaptive navigation behaviors. In the future, I hope to extend these ideas toward a more unified framework of human-centered robotic intelligence that seamlessly integrates perception, reasoning, and action.

Last updated: November 30, 2025