

Wenhao Li

📞 (+86) 15194386577

@ weli0055@uni.sydney.edu.au

🏛️ The University of Sydney

📅 2025.03–2027.09 (Expected)

🎓 Computer Science and Technology · Ph.D. Candidate

4 first-author papers accepted by CCF A / CORE A* conferences, including 2 oral presentations. Research interests include VLA / embodied intelligence (dynamic manipulation, error monitoring, dynamic reasoning, and test-time scaling), hallucination mitigation for large vision-language models, efficient diffusion models, and agents. The open-source project Sibly has received 240+ stars.

🎓 Education

- ▶ **The University of Sydney** Computer Science and Technology Ph.D. Candidate 2025.03–2027.12 (Expected)
Advisor: Chang Xu
- ▶ **Beihang University** Software Engineering B.Eng. GPA: 89/100 2019.09–2023.06

💼 Internships and Work Experience

- ▶ **Agibot, Embodied Research Center** Research Intern, Embodied Large-Model Algorithms 2025.04–2025.10
 - Conducted research on state monitoring, error handling, and test-time scaling for vision-language-action (VLA) models.
- ▶ **SenseTime** Full-Time Algorithm Researcher 2023.02–2025.02
 - Worked on hallucination mitigation for large vision-language models and image generation algorithms based on lightweight diffusion models.

⚙️ Research Experience

- ▶ **Vision-Language-Action Models for Dynamic Manipulation Scenarios such as Conveyor Belts** 2025.11–2026.05
 - **Background:** Existing VLA models lack dynamic signals in visual inputs, show limited robustness to action obsolescence caused by inference latency, and cannot respond to dynamic environments in real time. They also lack explicit modeling of spatio-temporal affordances for dynamic manipulation.
 - **Research:** Introduced optical flow in a polar-map form to provide VLA models with motion perception, delay-aware training to improve robustness, and an RL-trained residual correction network for real-time high-frequency action correction in dynamic environments. Also explicitly modeled spatio-temporal affordances in dynamic manipulation tasks, enabling automatic selection of optimal execution time windows and spatial coordinates so that VLA models act in the least difficult spatio-temporal windows.
 - **Outcomes:** First-author paper **DSDyn-VLA: A Dual-Stream Dynamic Manipulation Framework with Motion Perception, Future Awareness, and Realtime Correction** submitted to TPAMI; first-author paper **Kairos-VLA: Spatio-Temporal Affordance-Guided Execution Scheduling for Dynamic Manipulation** submitted to NeurIPS 2026.
- ▶ **Sibly: An Automated Research System with Active Exploration and Experience Accumulation** 2026.03–2026.04
 - **Research:** Built an agent system based on Claude Code that can automatically propose ideas, conduct exploratory experiments, accumulate experience to guide later exploration, design solutions, implement code and experiments, and generate full papers end to end.
 - **Outcome:** Open-source repository with 240+ stars.
- ▶ **State Monitoring, Error Recovery, and Test-Time Compute Scaling for Vision-Language-Action Models** 2025.04–2025.10
 - **Background:** VLA models lack execution-state monitoring and error-handling capabilities, and have limited deliberation ability in difficult scenarios.
 - **Research:** Studied active state monitoring, dynamic reasoning, error recovery, and adaptive test-time deliberation for VLA models to improve stability and generalization in embodied tasks.
 - **Outcomes:** First-author paper **Sentinel-VLA: A Metacognitive VLA Model with Active Status Monitoring for Dynamic Reasoning and Error Recovery** accepted by ICML 2026; first-author paper **VLA-ATTC: Adaptive Test-Time Compute for VLA Models with Relative Action Critic Model** accepted by ICML 2026.
- ▶ **Hallucination Mitigation for Large Vision-Language Models** 2024.12–2025.04

- **Background:** Large vision-language models are prone to hallucinations, which affect model trustworthiness and practical deployment stability.
 - **Research:** Analyzed and isolated hallucination components in the internal knowledge distribution of LVLMs, and proposed a self-evolving distillation strategy to purify models.
 - **Outcome:** First-author paper **Identify, Isolate, and Purge: Mitigating Hallucinations in LVLMs via Self-Evolving Distillation** accepted as an **Oral** paper by ACM MM 2025, CCF A / CORE A*.
- **Architecture Search and Efficient Training for Diffusion Models** 2023.02–2024.02
- **Background:** Diffusion models incur high training and inference costs, requiring more efficient architecture designs and training strategies.
 - **Research:** Used large language models as agents and leveraged their reasoning capabilities to iteratively predict better diffusion-model architectures. Further analyzed task characteristics and difficulty patterns across diffusion timesteps, and assigned customized denoising models to improve training and inference efficiency.
 - **Outcomes:** First-author paper **DiffNAS: Bootstrapping Diffusion Models by Prompting for Better Architectures** accepted as an **Oral** paper by ICDM 2023, CORE A*; first-author paper **Adaptive Training Meets Progressive Scaling: Elevating Efficiency in Diffusion Models** accepted by ICME 2025.

Publications

- | | |
|--|--|
| ► VLA-ATTC: Adaptive Test-Time Compute for VLA Models with Relative Action Critic Model | ICML 2026, CCF A, First Author |
| ► Sentinel-VLA: A Metacognitive VLA Model with Active Status Monitoring for Dynamic Reasoning and Error Recovery | ICML 2026, CCF A, First Author |
| ► Identify, Isolate, and Purge: Mitigating Hallucinations in LVLMs via Self-Evolving Distillation | ACM MM 2025, (Oral), CCF A, First Author |
| ► DiffNAS: Bootstrapping Diffusion Models by Prompting for Better Architectures | ICDM 2023, (Oral), CORE A*, First Author |
| ► Adaptive Training Meets Progressive Scaling: Elevating Efficiency in Diffusion Models | ICME 2025, First Author |
| ► DSDyn-VLA: A Dual-Stream Dynamic Manipulation Framework with Motion Perception, Future Awareness, and Realtime Correction | Submitted to TPAMI, First Author |
| ► Kairos-VLA: Spatio-Temporal Affordance-Guided Execution Scheduling for Dynamic Manipulation | Submitted to NeurIPS 2026, First Author |
| ► StateTrace: An Object-Centric Framework for Hidden-State Spatiotemporal Reasoning in Long Videos | Submitted to ACM MM 2026, Co-First Author (Second-Listed) and Corresponding Author |
| ► PyrMem-VLA: Plug-and-Play Pyramid Memory for Vision-Language-Action Models | Submitted to ACM MM 2026, Co-First Author (Second-Listed) and Corresponding Author |
| ► Anomaly Detection and Generation with Diffusion Models: A Survey | Submitted to TPAMI |

Honors and Awards

- Excellent Completion for 2 National Undergraduate Innovation and Entrepreneurship Training Program projects, as project lead.
- First Prize, Beihang University Science and Technology Innovation Scholarship, awarded 3 times.

Technical Skills

- Proficient in Python, C, PyTorch, JAX, DeepSpeed, Hugging Face Transformers, Diffusers, and related development frameworks and tools.
- Familiar with research on VLA models, embodied intelligence, agents, large vision-language models, and diffusion models.