

Zhaoyang Li

6083205509

zhl165@ucsd.edu

LinkedIn

Github

Homepage

Education

MS University of California San Diego

- Electrical and Computer Engineering (Intelligent Systems, Robotics & Control)

Sep 2023 - Jun 2025

BS University of Wisconsin-Madison,

- Double Major in Computer Science & Mathematics

Jan 2021 - May 2023

Research Interests

My research focuses on developing embodied agents that perceive, reason, and act in the physical world. I work at the intersection of **multimodal learning**, **robot learning**, **VLA integration**, and **preference alignment**, with a focus on **(1)**. improving the robustness of vision–language models and their applications, **(2)**. developing efficient control and planning policies, **(3)**. integrating vision–language–action systems into unified embodied agents, and **(4)**. aligning agent behavior with human intentions through feedback and preferences. Ultimately, I aim to build reliable, interpretable agents that generalize to real-world environments.

Publications & Preprints

* indicates equal contributions.

- 1. **Zhaoyang Li***, Zhan Ling*, Yuchen Zhou, Litian Gong, Erdem Biyik, Hao Su. “ORIC: Benchmarking Object Recognition under Contextual Incongruity in Large Vision-Language Models.” ICCV 2025 Workshop on Multi-Modal Reasoning for Agent Intelligence (MMRAgI); extended version in submission. [\[Paper\]](#) ↗ [\[Code\]](#) ↗
- 2. **Zhaoyang Li**, Sushaanth Srinivasan, Ninad Ekbote, Pengtao Xie. “A Multi-modal Large Language Model for Predicting Mechanisms of Drug Interactions.” Under review, 2025.
- 3. Tongzhou Mu*, **Zhaoyang Li***, Stanisław Wiktor Strzelecki*, Xiu Yuan, Yunchao Yao, Litian Liang, Aditya Gulati, Hao Su. “When Should We Prefer State-to-Visual DAgger Over Visual Reinforcement Learning?” AAAI 2025. [\[Paper\]](#) ↗ [\[Code\]](#) ↗
- 4. Yifei Zhang, Yusen Jiao, Jiayi Chen, **Zhaoyang Li**, Huaxiu Yao, Jieyu Zhang, Frederic Sala. “Just Select Twice: Leveraging Low-Quality Data to Improve Data Selection.” ATTRIB Workshop at NeurIPS 2024; extended version in submission. [\[Paper\]](#) ↗

Research Experience & Project

Contextual Override in VLA Agents: Analysis and Mitigation (Project Lead)

Advisor: Prof. Erdem Biyik, USC

Sep 2025 – present
Los Angeles, CA

- Studied default-action bias where scene affordances override language under noisy and irrelevant commands, and built a conflicting-affordance benchmark to measure instruction adherence vs bias in VLA agents.
- Validated findings in simulation and on real robots, with fail-safe action checks and multi-sensor synchronization.
- Project ongoing; targeting a submission to the International Conference on Intelligent Robots and Systems (IROS) 2026.

Context-Incongruity Robustness for Large Vision Language Models (Project Lead)

Advisor: Prof. Hao Su, UC-San Diego

Aug 2024 - Sep 2025
San Diego, CA

- Developed ORIC to generate context-incongruous object-context data for training and evaluation, and built ORIC-Bench via dual LLM and CLIP sampling.
- Evaluated 18 LVLMs and 2 open-vocabulary detectors; improved Qwen3-VL-8B-Instruct with Visual-RFT on ORIC-style data for stronger human-aligned predictions.
- Submitted the extended ORIC study to CVPR 2026.

Multi-modal Language Model for Drug Interaction Prediction (Project Lead) Advisor: Prof. Pengtao Xie; UC-San Diego	Dec 2024 - May 2025 San Diego, CA
<ul style="list-style-type: none"> Fine-tuned a multi-modal LLM with SMILES inputs to predict drug interaction status, degree, and mechanisms, integrating chemical informatics and NLP. Achieved strong performance: METEOR 0.42, BLEU-1 0.25, semantic similarity 0.57; outperforming GPT-4o (METEOR 0.16, BLEU-1 0.11, semantic similarity 0.30). 	
Empirical Analysis of State-to-Visual (S2V) Imitation vs. Visual RL (Project Co-Lead) Advisor: Prof. Hao Su, UC-San Diego	Feb 2024 - Sep 2024 San Diego, CA
<ul style="list-style-type: none"> Benchmarked State-to-Visual DAgger vs. visual RL across 16 tasks from ManiSkill, DMControl, and Adroit. Analyzed performance trade-offs, efficiency, and computational costs. Built a standardized S2V pipeline and derived practical recommendations. 	
Modality Transfer for PET and MRI Images Advisor: Prof. Vikas Singh; UW-Madison Medical Science Center, Computer Vision Group	May 2022 - Jun 2023 Madison, WI
<ul style="list-style-type: none"> Enhanced image translation with self-attention, MobileNetV2, and total variance loss in the pix2pix framework. Proposed U-TransGan model achieving PSNR 32, 0.98 correlation, and 0.92 SSIM. 	
Simulation of the Connected and Automated Driving Systems Advisor: Prof. Bin Ran; The Connected Automated Vehicle Highway System Group	Sep 2021 - May 2022 Madison, WI
<ul style="list-style-type: none"> Simulated and optimized traffic systems in CARLA, enhancing traffic management models for improved efficiency. Refined object detection algorithms, including YOLO and Faster R-CNN, to improve vehicle detection and traffic control systems. 	

Industry Experiences

Computer Vision Algorithm Engineer Mech-Mind Robotics Technologies Ltd.	Jun 2023 - Sep 2023 Beijing, China
<ul style="list-style-type: none"> Developed algorithms for structured light 3D cameras, improving image accuracy and optimizing point cloud reconstruction for laser systems. Led the refinement of internal camera distortion models, significantly enhancing imaging fidelity and calibration precision. 	
Backend Engineer Quanzhou YouGouZan Network Technology Co., Ltd.	Jun 2020 - Aug 2020 Quanzhou, China
<ul style="list-style-type: none"> Developed an online shopping mall on WeChat using SQL and Java, enabling functionalities like product search, browsing, recommendations, ordering, and payment. 	

Teaching

Teaching Assistant at UW-Madison	Spring 2023
<ul style="list-style-type: none"> CS540: Introduction to Artificial Intelligence 	
Peer Mentor at UW-Madison	Fall 2022
<ul style="list-style-type: none"> CS537: Introduction to Operating System 	

Professional Services

Reviewer	
<ul style="list-style-type: none"> AAAI Conference on Artificial Intelligence 	
<ul style="list-style-type: none"> AAAI Workshop: Large Language Models and Generative AI for Health 	

Technical Skills

<ul style="list-style-type: none"> Languages: English (Proficient; TOEFL 108, GRE 325 + 3.5), Chinese (Native) Programming: Python, C++, C, Java, Matlab, R, LaTeX, SQL ML and Vision Frameworks: PyTorch, TensorFlow, OpenCV, scikit-learn, SimpleITK, SPM12, TorchIO Simulation and Robotics: CARLA, MuJoCo, ManiSkill, DMControl, RoboMimic, WidowX, Adroit. Developer Tools: VS Code, Vim, IntelliJ IDEA, Visual Studio, Git, Docker, Kubernetes
--