

Jiakang YUAN

Fudan University | (+86)-199-2126-1301 | jkyuan112@gmail.com
[Homepage](#) | [GitHub](#) | [Google Scholar](#)

ABOUT ME

I am a Ph.D. candidate in Electronic Engineering at Fudan University, expected to graduate in June 2027. My research specializes in agentic foundation models, multi-agent systems, long-horizon agents, and multimodal reasoning. I have authored 14 first- or co-first-author research papers and contributed to 4 technical reports, with work accepted at ACL, ICML, ICCV, NeurIPS, CVPR, ECCV, and ICLR.

Research Interests: Agentic Foundation Models, Multi-Agent Systems, Long-Horizon Agents, and Multimodal Reasoning.

EDUCATION

Fudan University , Ph.D. Candidate in Electronic Engineering GPA: 3.68/4.00; Rank: 2/30	Sep. 2022 - Jun. 2027 Shanghai, China
Fudan University , B.Eng. in Electronic Engineering GPA: 3.77/4.00; Rank: 2/77	Sep. 2018 - Jun. 2022 Shanghai, China

RESEARCH EXPERIENCE

Tongyi Lab, Qwen Post-Training Group. Research Intern. Developing general-purpose agentic models for long-horizon tasks, mainly focusing on PaperBench and PostTrainBench as contributions to the Qwen 3.8 series, along with other post-training efforts.	Apr. 2026 - Present
Hunyuan Multimodal Foundation Group, Tencent TEG. Research Intern. Worked on agentic captioning for multimodal foundation models, exploring how structured visual descriptions can improve generation quality and agent workflows.	Mar. 2026 - Apr. 2026
Shanghai Artificial Intelligence Laboratory. Research Intern. Developed multi-agent systems and agentic foundation models for automated research, deep research, survey generation, and ML engineering, including Agents-A1 , InternAgent , Dolphin , SurveyForge , MCP-R1, AutoMLGen , and FlowSearch . Also contributed to multimodal reasoning and captioning projects, including MME-Reasoning , Omni-Captioner , and Chimera .	Oct. 2024 - Oct. 2025
Shanghai Artificial Intelligence Laboratory. Research Intern. Worked on 3D perception, autonomous driving, and test-time adaptation, contributing to Bi3D , Uni3D , AD-PT , Reg-TTA3D , and ReSimAD [ADLab].	Aug. 2022 - Feb. 2024

SELECTED PUBLICATIONS AND PRE-PRINTS * EQUAL CONTRIBUTION

Agentic Foundation Models and Multi-agent Systems

- **Scaling the Horizon, Not the Parameters: Reaching Trillion-Parameter Performance with a 35B Agent.**
- InternAgent Team, Shanghai AI Lab [[Technical Report](#) | [Paper](#) | [HF](#)]
- [Summary]: Open-sourced a 35B agentic model for long-horizon tasks.
- **From Static Context to Calibrated Interactive RL: Mitigating Distribution Shift in Multi-turn Dialogue with Aligned Simulator.**
- X. Wang*, [J. Yuan*](#), Z. Huang, M. Tian, C. Lv, K. Song, C. Tao, X. Zheng [[Under Review](#) | [Paper](#)]
- [Summary]: A unified framework that couples interactive RL with simulator alignment.
- **InternS1-Pro: Scientific Multimodal Foundation Model at Trillion Scale.**
- InternS1 Team, Shanghai AI Lab [[Technical Report](#) | [Paper](#) | [GitHub](#) | [HF](#)]
- [Summary]: A scientific multimodal foundation model at trillion scale.
- **InternAgent-1.5: A Unified Agentic Framework for Long-Horizon Autonomous Scientific Discovery.**
- InternAgent Team, Shanghai AI Lab [[Technical Report](#) | [Project](#) | [Paper](#) | [GitHub](#) | [HF](#)]
- [Summary]: A frontier multi-agent system for automated research.
- **MCP-R1: Generalized Real-World Task Agent Mastering Dozens of Tools.**

- T. Peng*, J. Yuan*, Z. Zhong, Y. Jiang, B. Wang, P. Ye, T. Chen, L. Bai, et al.
- [Summary]: An MCP-based agentic model for real-world multi-tool tasks.
- **InternAgent: When Agent Becomes the Scientist – Building Closed-Loop System from Hypothesis to Verification.**
- B. Zhang*, S. Feng*, X. Yan*, J. Yuan*, et al. [Technical Report | Project | Paper | GitHub | HF]
- [Summary]: A closed-loop multi-agent framework for scientific discovery.
- **DOLPHIN: Moving Towards Closed-loop Auto-research through Thinking, Practice, and Feedback.**
- J. Yuan, X. Yan, B. Shi, T. Chen, W. Ouyang, et al. [ACL 2025 | Paper | Project]
- [Summary]: First closed-loop open-ended auto-research framework.
- **SurveyForge: On the Outline Heuristics, Memory-Driven Generation, and Multi-dimensional Evaluation for Automated Survey Writing.**
- X. Yan*, S. Feng*, J. Yuan, R. Xia, B. Wang, L. Bai, B. Zhang [ACL 2025 | Paper | GitHub | HF]
- [Summary]: Comprehensively improves the quality of AI-generated survey papers.
- **AutoMLGen: Navigating Fine-Grained Optimization for Coding Agents.**
- S. Du, X. Yan, D. Jiang, J. Yuan, Y. Hu, X. Li, L. He, B. Zhang, L. Bai [Paper]
- [Summary]: Coding agent workflow achieves 1st on MLE-Bench.
- **FlowSearch: Advancing deep research with dynamic structured knowledge flow.**
- Y. Hu, R. Ma, Y. Fan, J. Shi, Z. Cao, Y. Zhou, J. Yuan, et al. [ACL 2026 | Paper]
- [Summary]: A high-performance open-source workflow for deep research.

Multimodal Foundation Models and Reasoning

- **MME-Reasoning: A Comprehensive Benchmark for Logical Reasoning in MLLMs.**
- J. Yuan*, T. Peng*, Y. Jiang, Y. Lu, R. Zhang, K. Feng, et al. [ICML 2026 | Paper | GitHub]
- [Summary]: A benchmark for logical reasoning in multimodal LLMs.
- **Segment-Aligned Policy Optimization for Multi-Modal Reasoning.**
- L. Gao, Z. Li, M. Jia, J. Yuan, H. Sun, H. Sun, X. Li [ICML 2026 | Paper | GitHub]
- [Summary]: Optimizes multimodal reasoning at the segment level.
- **OMNICAPTIONER: One Captioner to Rule Them All.**
- Y. Lu*, J. Yuan*, Z. Li, S. Zhao, Q. Qin, X. Li, L. Zhuo, L. Wen, et al. [Paper | GitHub | HF]
- [Summary]: Captions diverse visual domains, including natural, structured, and visual-text images and videos.
- **HL3DLLM: Human-like 3D Understanding and Reasoning with Retrieved Images.**
- J. Yuan, M. Li, L. Zhang, B. Zhang, T. Chen
- [Summary]: Retrieves images to improve indoor 3D understanding with LLMs.
- **Chimera: Improving Generalist Model with Domain-Specific Experts.**
- T. Peng*, M. Li*, J. Yuan, H. Zhou, R. Xia, R. Zhang, L. Bai, et al. [ICCV 2025 | Paper | GitHub]
- [Summary]: Improves visual reasoning by combining knowledge from multiple visual experts.

Other Direction

- **Trust Your Instincts: Confidence-Driven Test-Time RL for Vision-Language-Action Models.**
- S. Chen, J. Yuan, J. Wang, T. Chen [ECCV 2026]
- **All-in-One: Transferring Vision Foundation Models into Stereo Matching.**
- J. Zhou*, H. Zhang*, J. Yuan*, P. Ye, T. Chen, H. Jiang, M. Chen, Y. Zhang [AAAI 2025 | Paper]
- **Causal Sequence Modelling for End-to-End 3D Object Detection.**
- M. Li*, J. Yuan*, S. Chen, L. Zhang, A. Zhu, X. Chen, T. Chen [NeurIPS 2024 | Paper]
- **Training-Free Adaptive Diffusion with Bounded Difference Approximation Strategy.**
- H. Ye*, J. Yuan*, R. Xia, X. Yan, T. Chen, J. Yan, B. Shi, B. Zhang [NeurIPS 2024 | Paper | Project | GitHub]
- **Better Regression makes better test-time adaptive 3D Object Detection.**
- J. Yuan, B. Zhang, K. Gong, X. Yue, B. Shi, Y. Qiao, T. Chen [ECCV 2024 | Paper]
- **AD-PT: Autonomous Driving Pre-Training with Large-scale Point Cloud Dataset.**
- J. Yuan, B. Zhang, X. Yan, T. Chen, B. Shi, Y. Li, Y. Qiao [NeurIPS 2023 | Project | Paper | GitHub]
- **Bi3D: Bi-domain Active Learning for Cross-domain 3D Object Detection.**
- J. Yuan, B. Zhang, X. Yan, T. Chen, B. Shi, Y. Li, Y. Qiao [CVPR 2023 | Paper | GitHub]
- **Bi3D++: Hybrid Bi-domain Active Learning for Cross-domain 3D Object Detection.**
- J. Yuan, X. Yan, B. Zhang, B. Shi, Y. Li, Y. Qiao, T. Chen [T-PAMI 2026]

PROJECTS

- **Agentic Foundation Models and Multi-agent Systems** Sep. 2024 - Present
 - Developed long-horizon agent workflows with planning, reflection, tool use, and multi-agent collaboration.
 - Built agent systems for automated research and survey generation, including **InternAgent**, **InternAgent-1.5**, **Dolphin** ([ACL 2025](#)), and **SurveyForge** ([ACL 2025](#)).
 - Extended agent workflows to deep research and ML engineering, including **FlowSearch** and **AutoMLGen**.
 - Worked on agentic model post-training, tool use, and interactive RL, including **Agents-A1**, **MCP-R1**, and Paper-Bench/PostTrainBench for the Qwen 3.8 series.
- **Multimodal Foundation Models and Reasoning** Mar. 2024 - Present
 - Worked on multimodal foundation models, reasoning benchmarks, captioning, and expert-composition methods.
 - Representative works include **InternS1-Pro**, **MME-Reasoning** ([ICML 2026](#)), **SAPO** ([ICML 2026](#)), **OmniCaptioner**, **HL3DLLM**, and **Chimera** ([ICCV 2025](#)).
 - Explored structured visual representations, captioning, and policy optimization for reasoning and generation.
- **Other Direction** Aug. 2022 - Mar. 2025
 - Conducted research on diffusion acceleration, 3D perception, autonomous driving, and test-time adaptation.
 - Representative works include **T²VLA** ([ECCV 2026](#)), **AdaptiveDiffusion** ([NeurIPS 2024](#)), **3DET-Mamba** ([NeurIPS 2024](#)), **AD-PT** ([NeurIPS 2023](#)), **Bi3D** ([CVPR 2023](#)), **Bi3D++** ([T-PAMI 2026](#)), **Reg-TTA3D** ([ECCV 2024](#)), and **AIO-Stereo** ([AAAI 2025](#)).

SCHOLARSHIPS AND AWARDS

BYD Scholarship of Fudan University.	Sep. 2025
National Scholarship (Rank 1st among all Ph.D. candidates in EE.).	Sep. 2024
Huatai Scholarship of Fudan University.	Sep. 2023
Second prize of the Scholarship for Outstanding Students at Fudan University.	Sep. 2021 / 2020 / 2019

ACADEMIC SERVICE

Reviewer of Conferences: CVPR, ICCV, ECCV, NeurIPS, ICML, ICLR, AAI, ACM MM
Reviewer of Journals: T-PAMI, T-IP, T-CSVT, T-MM, Pattern Recognition, Neurocomputing.