

耿欣宇

邮箱: xgengad@connect.ust.hk | 手机: +86-13325300719 | 谷歌学术: [谷歌学术主页](#)

研究方向

多模态大语言模型 (MLLM)、网页智能体 (Web Agent)、多模态推理 (Multimodal Reasoning)、深度研究 (Deep Research)

教育经历

香港科技大学, 中国香港
计算机科学与工程博士
2025年9月 | 至今

哈尔滨工业大学, 中国深圳
电子信息硕士
2022年9月 | 2025年6月
累计绩点: 3.365/4.0 | 排名: 12/82

电子科技大学, 中国成都
自动化学士
2018年9月 | 2022年6月
累计绩点: 3.61/4.0

实习经历

阿里巴巴, 通义实验室
研究型算法实习生
中国杭州
2024年7月 | 2025年8月

导师: 王新宇; 团队负责人: 蒋勇、黄非、周靖人

- 参与构建 Multimodal Deep Research Agent (Web Agent + RAG) 系统, 负责 tool-use 轨迹生成与策略优化
- 设计并实现 agent tool-use pipeline (search / browse / reasoning), 提升多步信息检索与决策能力
- 构建大规模 SFT 训练数据 (数十万级轨迹), 用于多轮推理与工具调用能力对齐
- 研究成果发表了多篇论文 ([C1], [C6], [C7]), 并在技术报告 ([R1]) 中做出重要贡献

腾讯, AI Lab / 混元, Frontier 组

人才计划实习生 (青云实习)
中国深圳
2025年11月 | 至今

导师: 梁振文、方天庆; 团队负责人: 米海涛、俞涛、姚顺雨

- 参与研发 Deep Research Agent (多模态 + Web), 聚焦复杂推理与工具调用能力
- 设计 agent trajectory + reflection 机制, 优化规划与执行策略
- 探索 on-policy self-distillation / RLVR 在 agent learning 中的应用
- 研究成果发表论文 ([C2], [C5])

技术报告

[R1] TDR Team, Tongyi DeepResearch Technical Report[R]. 2025

论文

[C1] **X Geng**^{*}, P Xia^{*}, Z Zhang^{*}, et al. WebWatcher: Breaking New Frontier of Vision-Language Deep Research Agent[C]. ICLR 2026. (Accept, Poster)

[C2] **X Geng**, T Fang, Z Liang, et al. Search, Distill, Evolve: Self-Distillation of Web Agents in Verifiable Virtual-Tool Environments[C]. NeurIPS 2026. (Submitted)

[C3] **X Geng**, J Wang, J Gong, et al. OrthCaps: An Orthogonal CapsNet with Sparse Attention Routing and Pruning[C]. CVPR 2024. (Accept, Poster)

[C4] **X Geng**, J Wang, J Xu, et al. ParseCaps: An Interpretable Parsing Capsule Network for Medical Image Classification[C]. AAAI 2025. (Accept, Poster)

[C5] **X Geng**^{*}, Y Xiao^{*}, et al. GeoBrowse: A Geolocation Benchmark for Agentic Tool Use with Expert-Annotated Reasoning Traces[C]. ECCV 2026. (Under review)

[C6] Z Chen^{*}, **X Geng**^{*}, X Wang, et al. Efficient Multimodal Planning Agent for Visual Question Answering[C]. ICML 2026. (Under review)

[C7] Z Chen, X Wang, Y Jiang, Z Zhang, **X Geng**, et al. Detecting Knowledge Boundary of Vision Large Language Models by Sampling-Based Inference[C]. EMNLP 2025. (Accept, Main)

[C8] Z Su, J Gao, H Guo, Z Liu, L Zhang, **X Geng**, et al. Agentvista: Evaluating multimodal agents in ultra-challenging realistic visual scenarios[C]. ICML 2026. (Under review)

[C9] F Zhang, G Zhang, S Qian, H Li, **X Geng**, et al. Struct-Searcher: Agentic Structural Thinking Advances Multimodal Deep Information Seeking[C]. ICML 2026. (Under review)

学术服务

审稿人: ICLR, ICML, ECCV, AAAI, CVPR

受邀报告:

- MLNLP 学术报告
2025年9月
- 智源论坛
2025年9月
- AAAI Imageomics 研讨会
2025年2月

荣誉奖励

- 国家奖学金
2024
- 哈尔滨工业大学特等学业奖学金
2024
- 哈尔滨工业大学三等入学奖学金
2022
- 电子科技大学校级奖学金
2018-2021

技能

- 编程: Python (PyTorch, TensorFlow, SFT, GRPO), MATLAB, LaTeX
- 语言: 中文 (母语)、英语 (CET-4: 615, CET-6: 563, IELTS: 7.0)