Zhuo Liu

Rochester, NY • +1 (585) 410-8951 • zhuo.liu@rochester.edu
Website • GitHub • Google Scholar • LinkedIn

TARGET ROLES Research Intern • Research Engineer • Research Scientist / Applied Scientist

FOCUS **Large Language Models & AI Agents**

Research Interests: I study how large language models and AI agents use external knowledge and feedback to reason and adapt. My goal is to make their reasoning more reliable and efficient.

EDUCATION

University of Rochester Jan 2024 – 2028 (expected)
Ph.D. in Computer Science | Advisor: Hangfeng He

University of Rochester Sep 2022 – Jun 2024
M.S. in Computer Engineering

South China University of Technology Sep 2017 – Jun 2021
B.Eng. in Automation

RESEARCH EXPERIENCE

Research Assistant | He's Lab, University of Rochester Oct 2023 – Present

Compiler-guided adaptive proof search
EMNLP 2026 Findings [Paper]

- Built and released a Python/Lean 4 search system combining generalist and specialist LLMs, compiler-guided repair, pairwise candidate selection, and restarts when refinement stalls.
- Improved average pass rate by **12.8 percentage points** over the two base models' mean pass@32 result across **7 miniCTX-v2 projects**, with **21.9% fewer LLM calls**.

Model priors in inductive reasoning
EMNLP 2025, Oral [Paper]

- Designed controlled evaluations across **5 real-world tasks, 3 LLMs, and 3 reasoning strategies**, varying demonstration availability, label mappings, and presentation formats.
- Found that removing demonstrations largely preserved hypothesis quality in the studied tasks; assessed the finding through downstream accuracy, LLM judging, and human evaluation.

TreeRare: syntax-guided retrieval and reasoning
EMNLP 2025 [Paper]

- Co-developed a framework that decomposes questions with syntax trees, retrieves evidence bottom-up, and aggregates local answers for multi-hop and ambiguous question answering.
- Evaluated across **5 datasets and 3 LLM backbones**; the Llama-3.3-70B variant improved average multi-hop COV-EM by **17.8% relative to the strongest baseline**.

Multimodal hallucination mitigation
NAACL 2025 Findings [Paper]

- Co-developed constraint-aware prompting using bidirectional and transitivity constraints to reduce spatial-relation hallucinations in multimodal models.

Evaluating generalization beyond content shortcuts
arXiv 2026 (preprint) [Paper]

- Designed controlled content-overlap evaluations and representation probes for RoBERTa-large style classifiers; found that higher training overlap improved robustness to changing content cues.

TECHNICAL SKILLS

Programming and systems: Python, Linux, Git, Slurm, Lean 4 / lean-interact.
ML frameworks: PyTorch, Hugging Face Transformers, DeepSpeed, vLLM.
Methods: LLM evaluation, retrieval-augmented generation, inference-time search, supervised fine-tuning, knowledge distillation, multimodal reasoning.

Zhuo Liu zhuo.liu@rochester.edu

SELECTED PUBLICATIONS

Full publication list: Google Scholar

LLM REASONING AND EXTERNAL KNOWLEDGE

On the Role of Model Prior in Real-World Inductive Reasoning.
Zhuo Liu, Ding Yu, Hangfeng He. **EMNLP 2025 | Oral**

Compiler-Guided Adaptive Proof Search with Cross-Model Synergy on Context-Dependent Theorem Proving.
Zhuo Liu, Ding Yu, Hangfeng He. **EMNLP 2026 Findings**

TreeRare: Syntax Tree-Guided Retrieval and Reasoning for Knowledge-Intensive Question Answering.
Boyi Zhang, **Zhuo Liu**, Hangfeng He. **EMNLP 2025**

What Are Step-Level Reward Models Rewarding? Counterintuitive Findings from MCTS-Boosted Mathematical Reasoning.
Yiran Ma, Zui Chen, Tianqiao Liu, Mi Tian, **Zhuo Liu**, Zitao Liu, Weiqi Luo. **AAAI 2025**
Finds that removing natural-language explanations has little effect on the evaluated step-level reward models.

RELIABLE MODELS AND AGENTS

AdvEvo-MARL: Shaping Internalized Safety through Adversarial Co-Evolution in Multi-Agent Reinforcement Learning.
Zhenyu Pan, Yiting Zhang, **Zhuo Liu**, et al. **ICML 2026**
Adversarial co-evolution improves attack resistance in LLM multi-agent systems while preserving task accuracy.

Mitigating Hallucinations in Multimodal Spatial Relations through Constraint-Aware Prompting.
Jiarui Wu, **Zhuo Liu**, Hangfeng He. **NAACL 2025 Findings**

Can CLIP Count Stars? An Empirical Study on Quantity Bias in CLIP.
Zeliang Zhang, **Zhuo Liu**, Mingqian Feng, Chenliang Xu. **EMNLP 2024 Findings**
Identifies quantity bias through controlled evaluations of text, image, and cross-modal representations.

SELECTED PREPRINT

Style or Content? Evaluating Style Classifiers with Controlled Content Overlap.
Zhuo Liu, Haozheng Du, Xiangxiang Xu, Hangfeng He. **arXiv 2026**

RESEARCH SUPPORT AND HONORS

NAIRR Pilot NAIRR250254 2025
Example-Free Test-Time Adaptation from Task Guidelines; **8,000 GPU hours**.

Empire AI Beta 2025
AI-Assisted Formalization of Informal Proofs via Reinforcement Learning; **6,000 GPU hours**.

National Scholarship for Excellent Undergraduates: Top 1% 2020

ACADEMIC SERVICE AND TEACHING

Workshop co-organizer: Trustworthy Foundation Models Workshop, ICCV 2025.

Reviewer: ACL Rolling Review (2024, 2025), NeurIPS (2024, 2025), CVPR (2025).

Teaching Assistant, University of Rochester
CSC 511: Large Language Models Fall 2024, Fall 2025
CSC 247/447: Natural Language Processing Spring 2025