

Weiyuan Chen

Tel: (+1) 5402393145 | Email: cweiyuan9@vt.edu

Education

Virginia Tech

Computer Science and Applications Ph.D.

- GPA: 4.0 /4.0

Blacksburg, U.S.
August. 2025 - Present

Zhejiang University

Bachelor's Degree in Computer Science

- GPA: 3.88/4.0
- GRE: 330/340
- TOEFL: 116/120

Hangzhou, China
Sept. 2021 - June. 2025

Skills: Deep Learning (Stanford CS224n), Natural Language Processing (CMU ANLP2024), LLM Agents (Berkeley RDI Agentic AI MOOC 2025 Fall), AI and Machine Learning (VT Frontier AI Systems), Human-computer Interaction (VT Advanced Topics in HCI)

Research Experience

LLMs Lab at Virginia Tech

Advisor: Prof. Tu Vu

Research Assistant

May. 2025 - Present

PRISM: Pushing the Frontier of Deep Think via Process Reward Model-Guided Inference

- **Co-lead research on advancing LLM inference-time reasoning**, conducting a comprehensive bottleneck analysis of current deep think frameworks and population enhancement methods.
- **Developed PRISM**, a novel framework using PRMs to guide score-based resampling and stochastic solution refinement, achieving consistent net-directional correction during the reasoning process.
- **Achieved SOTA performance on AIME (90%) and GPQA-Diamond (71.4%)**, enabling a 20B model to outperform a 120B zero-shot baseline, and maintaining a position on the compute-accuracy Pareto-frontier.

SealQA: Raising the Bar for Reasoning in Search-augmented Large Language Models

- **Contributed to a high-quality Question-answering dataset with noisy search results** by curating data including questions and annotated answers from tables from websites such as wikipedia.
- **Contributed to benchmarking and error analysis**, validating the effectiveness of the dataset in testing Frontier Large Language Models.

Yale NLP Lab

Advisors: Prof. Arman Cohan, Prof. Chen Zhao at NYU and NYU Shanghai

Undergraduate Research Intern

Mar. 2024 - Sept. 2024

AbGen: Evaluating Large Language Models in Ablation Study Design and Evaluation for Scientific Research

- **Co-lead the curation of a large-scale benchmarking dataset** by extracting 1500+ structured ablation studies from 800+ peer-reviewed AI papers.
 - **Benchmarked state-of-the-art open and closed-sourced LLMs** on their ability to autonomously generate scientific ablation studies, identifying critical gaps in the capabilities of AI research assistants.
 - **Built a LLM-as-a-Judge evaluation framework**, implementing a templated Likert-scale scoring system to automate the assessment of generated studies based on **soundness, importance and faithfulness**.
-

Notable Projects

AgentX AgentBeats Competition | Project Lead

Oct. 2025 - Jan. 2026

- **Challenge:** Developed "LLM-Core-Wars," a novel multi-agent benchmark inspired by the classic "Core Wars" game to evaluate how LLM agents compete for a shared, limited memory context.
- **Action:** Engineered a system using the *a2a* Python library featuring "Purple Agents" (competitors) and a

"Green Agent" evaluator powered by Gemini-3; implemented a structured asynchronous polling mechanism to manage memory requests and attribution.

- **Result:** Successfully simulated and analyzed complex behaviors in a multi-agent framework; established a winning condition based on 80% context occupancy, providing a measurable framework for evaluating agentic safety and resource contention.

ModelScope Open Source Challenge | Project Lead

Sept. 2023 - Oct. 2023

- **Challenge:** Directed the expansion of a Multimodal GenAI open-source repository and augmented an AI-portrait generation pipeline to support diverse stylistic outputs.
- **Action:** Led a technical team of 3 through the full development lifecycle; engineered custom prompt templates and stylistic logic to enhance model versatility.
- **Result:** Orchestrated workload distribution and milestone tracking, resulting in the high-quality, on-time delivery of 14 unique AI portrait styles and earning a National Third Prize.

Honors and Awards

- | | |
|---|-----------|
| • Mastery Tier — UC Berkeley Center for Responsible Decentralized Intelligence Agentic AI MOOC (25 Fall) | Feb. 2026 |
| • National Third Prize --- ModelScope AIGC OpenSource Challenge | Oct. 2023 |
| • First Prize (55th rank) --- National Collegiate Mathematics Competition | Jan. 2023 |
| • National Scholarship for Chinese College Students | Oct. 2022 |
| • First Prize (28th rank) --- China Mathematical Olympiad (Guangdong Divisional Round) | Oct. 2019 |
| • First Prize (21st rank) --- 18th China Girl's Mathematical Olympiads (CGMO) | Aug. 2019 |
| • Bronze medal --- the 34th China Mathematical Olympiad (CMO) | Nov. 2018 |
| • First Prize (18th rank) --- China Mathematical Olympiad (Guangdong Divisional Round) | Oct. 2018 |
| • First Prize (18th rank) --- 17th China Girl's Mathematical Olympiads (CGMO) | Aug. 2018 |

Publications

(1) Pham, T.; Nguyen, N.; Zunjare, P.; **Chen, W.**; Tseng, Y.-M.; Vu, T. *SealQA: Raising the Bar for Reasoning in Search-Augmented Language Models*. ICLR 2026 (poster).

(2) Zhao, Y.; **Chen, W.**; Xu, Z.; Patwardhan, M.; Wang, C.; Liu, Y.; Vig, L.; Cohan, A. *AbGen: Evaluating Large Language Models in Ablation Study Design and Evaluation for Scientific Research*. ACL 2025 (poster).

(3) Zhao, Y.; Long, Y.; Jiang, Y.; Wang, C.; **Chen, W.**; et al. *FINDVER: Explainable Claim Verification over Long and Hybrid-Content Financial Documents*. EMNLP 2024 (poster).

Services

Virginia Tech Teaching Assistant Served as a Graduate Teaching Assistant (GTA) for Data Structure and Algorithms class during Spring 2026 semester, host office hours to respond to inquiries and grade assignments.

Zhejiang University Peer Mentoring Star Served as a mentor in the Computer Science Department Peer Education Mentor Program for multiple semesters, achieving the title of 'Peer Mentoring Star'.