Tianyu (Tiya) Cao

(412)-918-0926 | tianyuca@cs.cmu.edu | Home: tiyacao.com | Linkedin: caotianyu | Github: Springcty

EDUCATION

Carnegie Mellon University Pittsburgh, PA
M.S. in Intelligent Information Systems (Natural Language Processing), School of Computer Science Dec 2025
- Selected Coursework: Advanced NLP, Multimodal LLM, LLMs Methods and Application, ML with Large Datasets, Intro to ML

Zhejiang University Hangzhou, China
B.Eng. in Computer Science and Technology, Chu Kochen Honors College Jun 2024
- Selected Coursework: Data Structure & Algorithm Analysis, OOP, Linear Algebra, Intro to AI, Operating Systems

PUBLICATIONS

- **Tianyu Cao***, Xin Dong*, Hongwei Wang, Zhou Qin, Bingfeng Deng, Xinran Ye, Hongyu Xiong. **Multimodal Post-Training with Sequential GRPO and Consolidation SFT for Unoriginal Classification**. Preprint, 2025
- **Tianyu Cao*,** Neel Bhandari*, Akhila Yerukola, Akari Asai, Maarten Sap. **Out of Style: RAG's Fragility to Linguistic Variation**. Preprint, 2025
- Yixiao Zeng, **Tianyu Cao**, Danqing Wang, Xinran Zhao, Zimeng Qiu, Morteza Ziyadi, Tongshuang Wu, Lei Li. **RARE: Retrieval-Aware Robustness Evaluation for Retrieval-Augmented Generation Systems**. Preprint, 2025
- **Tianyu Cao**, Natraj Raman, Danial Devoric, Chenhao Tan. **Characterizing Multimodal Long-form Summarization: A Case Study on Financial Reports**. In the 1st *Conference on Language Modeling (COLM)*, 2024.
- Junru Chen, **Tianyu Cao**, et al. **Con4m: Context-aware Consistency Learning Framework for Segmented Time Series Classification**. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

SKILLS

Programming: Python, Java, C++, SQL, NumPy, Pandas
ML & AI: PyTorch, Transformers, Huggingface, NLP (RAG, LLMs, MLLMs), Reinforcement Learning (GRPO, DPO, PPO)
Tools & Platforms: Linux, Spark, GCC, AWS, Git, Jupyter

WORK EXPERIENCE

Multimodal Video Understanding Post-Training via GRPO for Feed Quality | TikTok Inc. San Jose, CA
Machine Learning Scientist Intern, TikTok-Data-Responsible Recommendation System-Feed Quality May 2025 – Aug 2025
- Designed video-understanding post-training tasks across three modalities (frames, audio/OCR, ASR) and built structured CoT data, transforming complex video understanding tasks into a combination of visual understanding and reasoning tasks
- Developed a curriculum-style training pipeline with mixed-task cold-start SFT followed by sequential GRPO training
- Specialized the final MLLM through SFT on a large-scale unoriginal dataset, improving OOD performance on business metrics
- Applied Cold-Start + RL pipeline to sludge content detection task, boosting OOD F1 score from 44.17% to 47.16%

RESEARCH EXPERIENCE

Retrieval-Aware Robustness Evaluation (RARE) for RAG Systems | Carnegie Mellon University Pittsburgh, PA
Research Assistant, Language Technologies Institute (LTI), Advisor: Lei Li Feb 2025 – Current
- Introduced RARE, a framework for systematically evaluating RAG robustness under realistic perturbations
- Developed RARE-Met, a diagnostic metric that evaluates robustness against query, document, and simulated real-world retrieval
- Built RARE-Get, a dynamic synthesis pipeline that generates time-sensitive multi-hop queries by knowledge graph extraction
- Released RARE-Set, a domain-specific benchmark with 400+ documents and 48k+ queries in finance, economics, and policy

Robustness of RAG Systems on Linguistic Variations | Carnegie Mellon University Pittsburgh, PA
Research Assistant, Language Technologies Institute (LTI), Advisor: Maarten Sap Sept 2024 – Mar 2025
- Conducted the first systematic analysis of RAG's robustness on linguistic variations across four linguistic dimensions
- Uncovered 15.34% retrieval score degradation, with grammatical modifications most severely impacting recall
- Uncovered 19.60% (F1 score) generation performance drops, while LLM scaling doesn't always help mitigate these gaps
- Identified cascading errors between retrieval and generation components, revealing that RAG systems are more vulnerable to linguistic variations than LLM-only approaches

Characterizing Multimodal Long-form Summarization | University of Chicago Chicago, IL (Hybrid)
Research Assistant, Chicago Human+AI Lab (CHAI), Advisor: Chenhao Tan Jul 2023 – Apr 2024
- Developed an evaluation framework for LLM-generated multimodal long-form financial report summaries
- Compared Claude 2.0/2.1 and GPT-4, identifying position bias and Claude's possible ability to capture key information
- Revealed Claude's superior numeric usage, with 8.37% of summary numbers from report tables vs. 4.98% for GPT-4
- Pioneered the taxonomy of numeric hallucinations, identifying a ~5% ratio and exploring prompt engineering solutions

Segmented Time Series Classification (TSC) | Zhejiang University Hangzhou, China
Research Assistant, AINet Lab, Advisor: Yang Yang Sept 2022 – Sept 2023
- Pioneered *Con4m*, a consistent learning framework for Multi-class with Varying Durations (MVD) segmented TSC
- Designed a progressive harmonization approach, improving model robustness against inconsistent training labels
- Validated *Con4m*'s superior performance through experiments on two public and one private MVD datasets