Tianruo Rose Xu

Mobile: +1 607 262 1199 Email: tx88@cornell.edu

Education

Cornell University, College of Engineering Aug 2024 – May 2027 (expected)
B.S. in Computer Science Minor: Mathematics
Relevant coursework: data structures, algorithms, operating systems, computer systems, machine learning, natural language processing, quantum computing, probability and statistics, multivariable calculus, linear algebra, differential equations, logic, cognitive science, linguistics, language and thought.

Achievements and Awards

- Hunter R. Rawlings III Cornell Presidential Research Scholar Award
- Clare Booth Luce (CBL) Undergraduate Research Award
- Received a $7,000 stipend through the Bowers Undergraduate Research Experience (BURE) scholarship program and continued to be supported by BURE academic year expansion program
- Platinum division, Achieved Full Score in USA Computing Olympiad (USACO) 2023 January Gold Contest
- USA Mathematical Olympiad Qualifier (USAMO)

Publication

ACCEPTED. **Tianruo Rose Xu**, Anirudh Atmakuru, Priyan Pattnayak, Tanya Goyal. *Can You Trust What I Think? Analyzing and Improving Verbalized Uncertainty and Factuality in Reasoning-Based Large Language Models.* AAAI-26 Undergraduate Consortium, Singapore, January 2026.

PUBLISHED. **Tianruo Rose Xu**, Vedant Gaur, Liu Leqi, Tanya Goyal. *The Progress Illusion: Revisiting Meta-Evaluation Standards of LLM Evaluators*. Findings of the Association for Computational Linguistics: EMNLP 2025, Suzhou, China, November 2025. [Paper] [Poster].

PUBLISHED. Yingke Ding, Jiankai Tang, Wanying Mo, **Tianruo Rose Xu**, Yuanchun Shi, Yuntao Wang. *NFCMTL: Auto NailFold Capillaroscopy through a Multi-Task Learning Model.* MICCAI 2025 Workshop MSB EMERGE, July 21, 2025. [Paper]
Code: GitHub Repository
Award: Runner-up Best Paper, MICCAI 2025 Workshop MSB EMERGE.

PUBLISHED. Tianruo Xu. *Elder Depression Detection by Multimodal Means.* BCP Education & Psychology, November 7, 2022. [Paper]

Research Experience

Undergraduate Researcher @ Cornell University: Computer Science Department, Natural Language Processing Group

Analyzing Verbalized Uncertainty and Factuality in Reasoning-based LLMs: 2025 – present
Supervised by Professor Tanya Goyal (Cornell University); collaborating with Anirudh Atmakuru.
Serve as project lead: design a multi-step pipeline to analyze verbalized uncertainty in LLM thinking traces, implement data processing and FactScore-based factuality evaluation, run large-scale experiments across multiple reasoning models, and train/analyze lightweight classifiers for uncertainty and factuality prediction. Coordinate research direction, experiment design, and writing of research updates and poster materials.
Accepted for presentation at the *AAAI-26 Undergraduate Consortium* (Singapore, 2026).

Revisiting Meta-Evaluation Standards of LLM Evaluators: 2025
Supervised by Professor Tanya Goyal (Cornell University), in collaboration with Professor Liu Leqi (UT Austin) and Vedant Gaur (University of Pennsylvania).
Co-led the project developing fine-grained meta-evaluation methods to audit automatic benchmarks for large language models (LLMs): designed the evaluation framework, implemented data collection and scoring pipelines for multiple LLM-as-judge benchmarks, and ran large-scale experiments comparing evaluator behavior across models and settings. Analyzed evaluator reliability in distinguishing similar-capability models, studied the impact of reference model choice, and conducted correlation and sensitivity analyses against human judgments. Contributed substantially to paper writing, figure design, and result interpretation.

Content and Style Evaluation in NLG Systems: 2024
Supervised by Professor Tanya Goyal and Professor Claire Cardie.
Served as a core student researcher designing and running experiments to evaluate content and style separately in NLG systems: developed the experimental framework, implemented data processing and evaluation pipelines, and ran model-based on system outputs. Conducted error analysis and ablation studies to refine task definitions and metrics, and contributed to writing and visualizing results.

Undergraduate Research Assistant @ Tsinghua University: Pervasive Human-Computer Interaction Lab

Intelligent AI Tool Benchmarking and Orchestration Platform: 2025 – present
Supervised by Primary Mentor Yingtian Shi (Georgia Tech) and Professor Chun Yu (Tsinghua University).
Designing and implementing the back-end tool matching and execution module for an intelligent AI tool platform that standardizes tool capability descriptions, matches tools to user intents, and orchestrates multi-tool workflows. Build APIs and infrastructure for safe, containerized tool execution and monitoring, and collaborate with the front-end and evaluation teams to support personalized AI tool selection and benchmarking.

NFCMTL: Auto NailFold Capillaroscopy through a Multi-Task Learning Model: 2025

Supervised by Primary Mentor Yingke Ding, Associate Researcher Yuntao Wang, and Professor Yuanchun Shi.
Proposed the first multi-task learning framework for nailfold capillaroscopy, integrating segmentation, classification, and keypoint detection into a unified model. Leveraging a Multiscale Vision Transformer (MViT) backbone with uncertainty-weighted task heads, NFCMTL achieves sub-pixel accuracy in capillary parameter estimation and outperforms prior single-task baselines. Demonstrated strong performance on a reorganized public dataset, providing a more accurate and efficient solution for automated, non-invasive microvascular diagnostics.

NFCMTL: Auto NailFold Capillaroscopy through a Multi-Task Learning Model: 2025
Supervised by Primary Mentor Yingke Ding, Associate Researcher Yuntao Wang, and Professor Yuanchun Shi.
Proposed the first multi-task learning framework for nailfold capillaroscopy, integrating segmentation, classification, and keypoint detection into a unified model. Leveraging a Multiscale Vision Transformer (MViT) backbone with uncertainty-weighted task heads, NFCMTL achieves sub-pixel accuracy in capillary parameter estimation and outperforms prior single-task baselines. Demonstrated strong performance on a reorganized public dataset, providing a more accurate and efficient solution for automated, non-invasive microvascular diagnostics.

Undergraduate Research Assistant @ Cornell University: Information Science Department 2024
Supervised by Primary Mentor Keigo Kusumegi and Professor Yian Yin.
Explored the impact of diverse collaboration on Wikipedia content. Processed data and used K-means clustering to enhance article quality.

Multimodal Transformer Elder Depression Detection: 2022 - 2023
Used NLP, audio, and video data to detect depression in the elderly.

The Kinematics, Physical Conditions, and Chemical Abundances of Ionized Gas in the Andromeda Galaxy and a Comparison to the Triangulum Galaxy: 2023
Supervised by Primary Mentor Aparajito Bhattacharya and Faculty Advisor Professor Raja GuhaThakurta at UCSC.
Analyzed FITS images and spectrograph data from the Keck II telescope to study the ionized gas in galaxies.

CAPTCHA - Taxonomy, Challenges, Attacks, and Potential Counter Defense: 2022
Supervised by Professor Suleyman Uludag at the University of Michigan.
Investigated cybersecurity and machine learning for digital security.

A Manipulator for Material Classification Based on Deep Learning: 2021 - 2022
Led a team to develop a robotic system for planetary exploration and material identification using deep learning for real-time detection.

STEM GIRLS Program, SAMSUNG & China Women's Foundation: 2020 - 2022
Developed environmental solutions, including a composting device and waste interceptor. Featured in People's Daily for innovation.

Teaching Experience @ Cornell University

Teaching Assistant: CS 4820 Introduction to Analysis of Algorithms 2025 Fall
Duties: Participate in weekly grading sessions and staff meetings, hold office hours, answer student questions on Ed Discussion, grade homework problems, assist with exam proctoring, and contribute to review sessions and exam proofreading.

Teaching Assistant: CS 2800 Mathematical Foundations of Computing (Discrete Mathematics) 2025 Spring
Duties: Lead discussion sections, grade homework and exams, hold office hours, and assist with course-related questions.