# Tanqiu Jiang

tanjiang@cs.stonybrook.edu | Cell: (610)-674-8718 | tanqiujiang.github.io

## Education

**Stony Brook University** — Stony Brook, NY
*Ph.D. in Computer Science* — *Aug 2023 – Present*

**Pennsylvania State University** — State College, PA
*Ph.D. in Informatics (transferred to Stony Brook)* — *Aug 2022 – May 2023*

**University of Rochester** — Rochester, NY
*Master of Science in Data Science* — *Aug 2020 – Dec 2021*

**Lehigh University** — Bethlehem, PA
*Bachelor of Science in Computer Engineering (Minor: Data Science)* — *Aug 2016 – May 2020*

## Publications

### Under Review

[1] **Jiang, T.**, Bai, M., Pappas, N., Qi, Y., & Swamy, S. Cross-Modal Content Optimization for Steering Web Agent Preferences. (**Submitted to ICLR 2026**).[pdf]

[2] **Jiang, T.**, Liang, J., Zhu, R., Zhou, J., Ma, F., & Wang, T. Robustifying Vision-Language Models via Dynamic Token Reweighting. (**Submitted to ICLR 2026**). [pdf]

### Conference Papers

[3] **Jiang, T.**, Wang, Z., Liang, J., Li, C., Wang, Y., & Wang, T. (2025). RobustKV: Defending Large Language Models against Jailbreak Attacks via KV Eviction. In *International Conference on Learning Representations* (**ICLR 2025**). [pdf]

[4] **Jiang, T.**, Li, C., Ma, F., & Wang, T. (2025). RAPID: Retrieval Augmented Training of Differentially Private Diffusion Models. In *International Conference on Learning Representations* (**ICLR 2025**). [pdf]

[5] **Jiang, T.**, Li, Y., Lin, H., Ruan, Y., & Woodruff, D. P. (2020). Learning-Augmented Data Stream Algorithms. In *8th International Conference on Learning Representations* (**ICLR 2020**), Addis Ababa, Ethiopia. [pdf]

[6] **Jiang, T.**, Bendre, S. K., Lyu, H., & Luo, J. (2021). From Static to Dynamic Prediction: Wildfire Risk Assessment Based on Multiple Environmental Factors. In *2021 IEEE International Conference on Big Data* (**IEEE Big Data**), [pdf]

[7] **Jiang, T.** & Xiong, Z. (2021). Rule-Based Approach to the Automatic Detection of Individual Tree Crowns in RGB Satellite Images. In *2021 IEEE International Conference on Computer Science, Artificial Intelligence and Electronic Engineering (IEEE-CSAIEE)*, pp. 132-135. [pdf]

[8] Wang, X., **Jiang, T.**, Cai, H. II. (2021). Human epithelial-2 cell image classification using deep unsupervised learning and gradient boosting trees. In *Proc. SPIE 11601, Medical Imaging 2021*. [pdf]

## Experience

**Amazon** — Seattle, WA
*Applied Scientist Intern, Agentic AI Lab* — *May 2025 – Sep 2025*

- Developed cross-modal adversarial attack framework to identify vulnerabilities in VLM-based web agents, achieving up to 71% preference manipulation rate across GPT-4.1, Qwen-2.5VL, and Pixtral-Large models.
- Implemented PGD-based visual perturbation attacks exploiting CLIP encoder vulnerabilities, achieving >50% black-box transfer success while maintaining imperceptibility ($\epsilon$=8/255).
- Designed iterative textual refinement algorithms leveraging RLHF-induced biases to steer agent preferences through semantically plausible content modifications.
- Conducted comprehensive evaluation on ControlNet-100k and VisualWebArena benchmarks, demonstrating 3.5$\times$ improvement over baseline methods while evading detection.

**Stony Brook University** Stony Brook, NY
*Research Assistant* *Aug 2023 – Present*

- Conduct advanced research in computer science under the supervision of Dr. Ting Wang, focusing on enhancing the robustness and privacy of machine learning models.
- Lead and manage two ongoing research projects:
  * **RobustKV: Defending Large Language Models against Jailbreak Attacks via KV Eviction**
  * **RAPID: Retrieval Augmented Training of Differentially Private Diffusion Models**
- Coordinate with a multidisciplinary team to design experiments, analyze data, and draft research manuscripts for publication.
- Present research findings at internal seminars and collaborate with external researchers to refine methodologies and approaches.

**Pennsylvania State University** State College, PA
*Research Assistant* *Aug 2022 – May 2023*

- Led a research project developing "Unlearnable Examples" aimed at minimizing the performance of contrastive learning models by introducing adversarial noise into training images.
- Attended and presented in weekly individual and group meetings with Dr. Ting Wang, Dr. Dongwon Lee, and lab mates.

**University of Rochester, Goergen Institute for Data Science** Rochester, NY
*Teaching Assistant* *Jan 2021 – Dec 2021*

- Held office hours averaging over 2 hours weekly to assist students with Python, Linux, SQL, PySpark, and R.
- Collaborated with Professors Lloyd Palum and Brendan Mort to test and refine course materials on Databricks.
- Conducted review sessions to reinforce course material and prepare students for assessments.
- Graded assignments and projects, providing constructive feedback to students.

**Vista Lab, University of Rochester** Rochester, NY
*Student Researcher* *Dec 2020 – Dec 2021*

- Led a research project on analyzing and predicting California Wildfire Incidents under the mentorship of Prof. Jiebo Luo and Hanjia Lyu.
- Integrated multiple environmental factors to assess wildfire risk using Logistic Regression, SVM, Neural Networks, etc.
- Utilized LSTM/RNN for time-series analysis to perform dynamic risk predictions.
- Co-authored a research paper published in the IEEE International Conference on Big Data.

**IEEE Big Data 2021 Conference** Remote
*Student Volunteer / Cohost* *Dec 2021*

- Cohosted the special session "S29: Contrastive Learning" alongside the session chair.
- Ensured speakers' presentations adhered to the schedule and were displayed correctly.

## Competitions

**Kaggle: Google Landmark Retrieval 2020**, Awarded Silver Medal (Top 3%, 16th/541).

## Grants and Scholarships

**Graham Endowed Fellowship** ($4,000), Pennsylvania State University, Aug 2022

**NSF Student Travel Grant** ($500), IEEE-Big Data Conference, Dec 2021