# Tao Jin

*Email:* taoj@outlook.com *Links:* Homepage, LinkedIn, Google Scholar, GitHub

## Education

**University of Virginia**  Feb. 2018 - May 2025
*Ph.D. in Computer Science*  Charlottesville, Virginia

**Zhejiang University**  Sept. 2012 - June 2017
*B.S. in Computer Science*  Hangzhou, Zhejiang, China
Excellent Student Awards (top 10%)

## Publications (∗ indicates equal contribution)

- **Tao Jin**∗, Yue Wu∗, Quanquan Gu, Farzad Farnoud: *Ranking with Multiple Oracles* [In submission]
- Yue Wu, **Tao Jin**∗, Qiwei Di∗, Farzad Farnoud, Quanquan Gu: *Borda regret minimization for generalized linear dueling bandits* [ICML'24]
- Qiwei Di∗, **Tao Jin**∗, Yue Wu, Farzad Farnoud, Quanquan Gu: *Variance-Aware Regret Bounds for Stochastic Contextual Dueling Bandits* [ICLR'24]
- Hao Lou, **Tao Jin**, Yue Wu, Pan Xu, Farzad Farnoud, Quanquan Gu: *Active Ranking without Strong Stochastic Transitivity* [NeurIPS'22]
- Yue Wu∗, **Tao Jin**∗, Hao Lou, Pan Xu, Farzad Farnoud, Quanquan Gu: *Adaptive Sampling for Heterogeneous Rank Aggregation from Noisy Pairwise Comparisons* [AISTATS'22]
- **Tao Jin**∗, Pan Xu∗, Quanquan Gu, Farzad Farnoud: *Rank Aggregation via Heterogeneous Thurstone Preference Models* [AAAI'20, Oral]
- Chenghao Liu, **Tao Jin**, Steven Hoi, Jianling Sun, Peilin Zhao: *Collaborative Topic Regression for Online Recommender Systems: An Online and Bayesian Approach* [ACML'16, Machine Learning Journal]

## Experience

**Applied Scientist** *Amazon*  July 2024 -
- Research and application for efficient training and data selection for fraud detection use cases.
- Integrate few shot learning with large language models for fraud detection.

**Graduate Research Assistant** *University of Virginia*  Feb. 2018 - June 2024
- Published 5 peer reviewed conference papers. 1 additional paper under review.
- Main research focus is on efficient estimation of ranking or utilities given noisy preferential data from multiple information sources.
- Developed an adaptive algorithm within this model assumption for active ranking use case.
- Developed adaptive learning strategies for noisy ranking problems under multi-armed bandit and dueling bandit settings that has 40% increase in benchmark compared to baseline methods.

**Applied Scientist Intern** *Trinity AI*  Jan. 2024 - May 2024
- Increased recall rate by 30% when using open large language models (LLMs) for protecting sensitive information by prompt engineering and fine-tuning using a small dataset with efficient training strategies.

**Applied Scientist Intern** *Amazon*  June 2023 - Aug. 2023
- Trained and deployed sequential recommendation transformer models that increase the recall rate by 8% in offline benchmark, increasing the click-through-rate by 5% in A/B testing.
- Investigated the model capabilities with respect to the model size (10M - 100M) and training data size (5M - 200M). Designed deployment size strategies given the computational budget and latency requirements.
- Profiled and optimized the training data feed system to reduce the data preparation time to 10%.

**Computer Vision Scientist** *Momenta Technology Ltd.*  Feb. 2017 - Feb. 2018
- *Lead a team of 4 engineers and 8 data operation interns.*
- Developed traffic light and traffic sign detection and tracking algorithm for autonomous driving with iterative data collection and model training cycle.

    Coordinated with data operation team, enacted labeling rules and optimized operation strategy and management tactics for the collection of million-scale data. Delivered models trained this data after quality control and cleaning.

    Deployed traffic light and traffic sign detection and tracking algorithm to the road test car and assisted decision module to achieve 100% correctness in closed road test area.