Ziyang Tang

Email: ztang@utexas.edu LinkedIn: ziyang-tang-858535160

SUMMARY

Applied Scientist with 5+ years of industry experience (Microsoft, Amazon) training LLM-based agents and recommendation systems with RL. Currently working on RL post-training for code generation models at Microsoft AI. Previously led end-to-end agent model training (SFT + RL) achieving frontier-model-level performance on web shopping tasks, and built reward and evaluation infrastructure for recommendation RL. PhD in CS (UT Austin) with 10+ publications at top-tier conferences such as NeurIPS, ICML, and ICLR in off-policy RL, uncertainty estimation, and approximate inference.

EXPERIENCE

Microsoft AI, Seattle, WA
Member of Technical Staff 2026.05 – Present

RL post-training for MAI-Code-1.x-flash

- Built synthetic coding data pipeline for RL post-training of MAI-Code-1.x-flash.
- Created and optimized behavior rubrics used in RL training to shape model coding behavior.

Amazon, Seattle, WA
Applied Scientist 2024.01 – 2026.05

Amazon "Buy For Me": Shopping agent model training (2025.06 – 2026.05)

- Led end-to-end design of agent model training program: led the design document, aligned science roadmap with leadership. Trained SFT reasoning model achieving task completion rate comparable to frontier LLMs on web shopping tasks, over 10× improvement from the open-source base model, targeting significant inference cost reduction.
- Built scalable model inference and evaluation framework: standardized data formats, unified LLM interfaces, dockerized the stack, enabling collection of 20k+ training trajectories and systematic evaluation of 10+ backend models across 10k+ third-party e-commerce sites.
- Designed step-wise reasoning benchmark with LLM-as-Judge evaluation; trained model achieved substantial accuracy improvement over baseline. Evaluation framework informs reward signal design for ongoing RL training.

Amazon Rufus: Recommendation RL (2024.10 – 2025.06)

- Developed multi-dimensional reward models and LLM-as-Judge evaluation with rubric-based critics (relevance, hallucination, diversity) for ASIN recommendation quality. Contributed to reward model training deployed in production PPO pipeline.
- Built live traffic data collection and evaluation pipeline for Rufus recommendation, enabling offline policy evaluation from production signals.

Amazon Ads: Bidding Optimization (2024.01 – 2024.10)

- Designed RL-based dynamic bidding optimization and multi-agent coordination framework for Sponsored Products advertising across 20+ bidding agents.

Research Scientist 2021.09 – 2023.12
Amazon Ads: Long-term Treatment Effect Estimation

- Built production ML pipeline for estimating long-term treatment effects in ad recommendation systems, deployed to two internal product teams; validated using 27 hold-out experiments.
- Developed novel nonstationary algorithm for long-term causal impact prediction; published at RecSys 2022 and presented at KDD 2023 workshop (50+ audience).

Bytedance, Seattle, WA
Research Intern 2020.06 – 2020.08

- Explored Product-of-Experts model design for large-scale recommendation retrieval (Douyin Ads), improving model capacity over standard mixture-of-experts approaches.

University of Texas at Austin, Austin, TX
Graduate Research Assistant 2017.09 – 2021.08

- PhD research on off-policy RL, uncertainty estimation, and approximate inference. Published 10+ papers at top-tier conferences such as NeurIPS, ICML, and ICLR.

TECHNICAL SKILLS

- **Languages**: Python, C/C++
- **ML/DL Frameworks**: PyTorch, TensorFlow
- **LLM & Agent Frameworks**: vLLM, BrowserGym, Verl, SkyRL
- **Tools**: AWS, Docker, Git, Claude Code, LaTeX

EDUCATION

University of Texas at Austin 2016.09 – 2023.05
Ph.D., Computer Science

Cornell University 2015.09 – 2016.05
M.Eng., Computer Science

Peking University 2011.09 – 2015.07
B.S., School of EECS

SELECTED PUBLICATIONS

Google Scholar link for full list.

- **Ziyang Tang**, Yiheng Duan, Steven Zhu, Stephanie Zhang, Lihong Li. Estimating Long-term Effects from Experimental Data. *RecSys 2022*.
- Yihao Feng*, **Ziyang Tang***, Na Zhang, Qiang Liu. Non-asymptotic Confidence Intervals of Off-policy Evaluation: Primal and Dual Bounds. *ICLR 2021*.
- **Ziyang Tang**, Yihao Feng, Na Zhang, Jian Peng, Qiang Liu. Off-Policy Interval Estimation with Lipschitz Value Iteration. *NeurIPS 2020*.
- **Ziyang Tang***, Yihao Feng*, Lihong Li, Dengyong Zhou, Qiang Liu. Doubly Robust Bias Reduction in Infinite Horizon Off-Policy Estimation. *ICLR 2020*.
- Dilin Wang*, **Ziyang Tang***, Chandrajit Bajaj, Qiang Liu. Stein Variational Gradient Descent with Matrix-Valued Kernels. *NeurIPS 2019*.
- Qiang Liu, Lihong Li, **Ziyang Tang**, Dengyong Zhou. Breaking the Curse of Horizon: Infinite-Horizon Off-Policy Estimation. *NeurIPS 2018*.

SERVICE

Reviewer: NeurIPS, ICML, ICLR, AAAI, AMLC, JMLR, TMLR, TPAMI.