# XIANFENG TANG

Homepage ◇ Email ◇ LinkedIn ◇ Google Scholar

## INTRODUCTION

Xianfeng Tang is an experienced researcher with five years of hands-on work in large foundation models and their real-world applications, including Amazon Rufus and Amazon Ads foundation models. His research spans a broad range of topics such as supervised fine-tuning, reinforcement learning, reasoning, agentic capabilities, multi-modality, knowledge editing and unlearning, trust and safety, retrieval augmentation, and e-commerce applications.

## RECENT EMPLOYMENT

| | |
|---|---|
| **Senior Applied Scientist**, founding member of the Amazon ADs foundation model team | 03/2025 - Present |
| **Senior Applied Scientist**, built Rufus, Amazon's Shopping AI | 02/2023 - 02/2025 |
| **Applied Scientist**, improved Amazon search experiences | 01/2021 - 01/2023 |

## EDUCATION

| | |
|---|---|
| **Ph.D. in Computer Science**, Penn State University | 08/2016 - 12/2020 |
| Improve the robustness, fairness, and explainability of graph neural network models. | |
| **B.S in Computer Science**, University of Science and Technology of China | 08/2012 - 06/2016 |
| Excellent Graduate Student (honored class) | |

## PUBLICATIONS

See the complete publication list on Google Scholar.

**Post-training**

- *Mitigating Heterogeneous Token Overfitting in LLM Knowledge Editing*, PDF
- *SFT or RL? An Early Investigation into Training R1-Like Reasoning Large Vision-Language Models*, PDF
- *Keeping an Eye on LLM Unlearning: The Hidden Risk and Remedy*, PDF
- *A General Framework to Enhance Fine-Tuning-Based LLM Unlearning*, PDF
- *Does your LLM truly unlearn? An embarrassingly simple approach to recover unlearned knowledge*, PDF
- *Divide-Verify-Refine: Aligning LLM Responses with Complex Instructions*, PDF

**Reasoning**

- *Efficient Long CoT Reasoning in Small Language Models*, PDF
- *Stepwise Perplexity-Guided Refinement for Efficient Chain-of-Thought Reasoning in Large Language Models*, PDF
- *A Theoretical Understanding of Chain-of-Thought: Coherent Reasoning and Error-Aware Demonstration*, PDF
- *Graph Chain-of-Thought: Augmenting Large Language Models by Reasoning on Graphs*, PDF
- *m1: Unleash the Potential of Test-Time Scaling for Medical Reasoning with Large Language Models*, PDF
- *Reasoning with Graphs: Structuring Implicit Knowledge to Enhance LLMs Reasoning*, PDF

**Agentic Capability**

- *Attention Knows Whom to Trust: Attention-Based Trust Management for LLM Multi-Agent Systems*, PDF
- *Comprehensive Vulnerability Analysis is Necessary for Trustworthy LLM-MAS*, PDF
- *RRO: LLM Agent Optimization Through Rising Reward Trajectories*, PDF

- *Cite Before You Speak: Enhancing Context-Response Grounding in E-Commerce Conversational LLM-Agents*, [PDF]

**Efficiency**

- *Examples as the Prompt: A Scalable Approach for Efficient LLM Adaptation in E-Commerce*, [PDF]
- *Resmoe: Space-efficient compression of mixture of experts LLMs via residual restoration*, [PDF]
- *Unlocking Efficient, Scalable, and Continual Knowledge Editing with Basis-Level Representation Fine-Tuning*, [PDF]
- *Blendfilter: Advancing retrieval-augmented large language models via query generation blending and knowledge filtering*, [PDF]
- *Learning with Less: Knowledge Distillation from Large Language Models via Unlabeled Data*, [PDF]
- *Simrag: Self-improving retrieval-augmented generation for adapting large language models to specialized domains*, [PDF]
- *Harnessing the Unseen: The Hidden Influence of Intrinsic Knowledge in Long-Context Language Models*, [PDF]

**Applications**

- *Exploring Query Understanding for Amazon Product Search*, [PDF]
- *Towards Unified Multi-modal Personalization: Large Vision-Language Models for Generative Recommendation and Beyond*, [PDF]
- *Sequential LLM Framework for Fashion Recommendation*, [PDF]
- *IHEval: Evaluating Language Models on Following the Instruction Hierarchy*, [PDF]
- *Shopping MMLU: A Massive Multi-task Online Shopping Benchmark for Large Language Models*, [PDF]
- *Large Language Models in the Clinic: A Comprehensive Benchmark*, [PDF]
- *Amazon-M2: A Multilingual Multi-locale Shopping Session Dataset for Recommendation and Text Generation*, NeurIPS 2023 [PDF]

## HONORS AND AWARDS

| | |
|---|---|
| - Travel Awards from AAAI, KDD, CIKM, etc. | 2019 - 2020 |
| - Rednor Graduate Fellowship, Penn State | 2016 |
| - 3rd Prize of Student Cluster Competition, SC14 | 2014 |
| - First Prize, Olympiad in Mathematics, Anhui Province, China | 2011 |