Xuyang Wu

+1-781-541-0920 | elviswu0306@gmail.com | elviswxy.github.io | Xuyang Wu | elviswxy

San Jose, CA - 95133, USA

Research Interests

Deep Learning, Information Retrieval, Search and Ranking, Responsible AI, Large Language Model, Multi-task Learning, Meta Learning, Recommendation System.

Education

- **Santa Clara University** *2019.09 - 2025.07*
  Ph.D, Computer Science Santa Clara, USA
  - Dissertation: Neural Ranking in Sparse Data Environments (Advisor: Prof. Yi Fang)
- **University College London** *2013.09 - 2015.09*
  M.Sc., Web Science and Big Data Analytics London, UK
  - Dissertation: Active Real-Time Bidding for CTR Estimation in Display Advertising (Advisor: Prof. Jun Wang)
- **Coventry University** *2011.09 - 2013.07*
  B.Sc., Computer Science Coventry, UK
  - GPA: First Honer Degree

Experience

- **DOCOMO Innovation, Inc.** *2020.07 - 2025.07*
  Visiting Researcher (Part time) Sunnyvale, USA
  - Developed an end-to-end multi-person head pose estimation system by extending YOLOv5 with a regression branch for pitch/yaw/roll, enabling real-time customer attention analysis for in-store retail advertising.
  - Pioneered HPE-CogVLM, an end-to-end head pose estimation framework using Vision-Language Models with LoRA-based model merging and incremental rehearsal, validating that VLMs can capture physical spatial orientation for robust multi-person detection and pose estimation.
  - Supported Docomo's Responsible AI initiatives by surveying and validating fairness challenges across successive AI paradigms, from ranking models, to vision-language models, to RAG pipelines, and the latest LLM reasoning, delivering actionable insights and mitigation strategies for trustworthy AI deployment. Paper accepted by *EMNLP*, *NAACL* and *COLING*.
- **Walmart Global Tech** *2022.06 - 2022.09*
  Data Scientist (Intern) Sunnyvale, USA
  - Developed a meta-learning based learning-to-rank framework to address sparse supervision challenges in long-tail query settings, enabling knowledge transfer across query-level task distributions and improving model generalization and fast adaptation to new ranking tasks.
  - Outperform LTR methods on sparsely labeled data with different ranking losses. Paper accepted by *ACM TOIS*.
- **Walmart Global Tech** *2021.06 - 2021.09*
  Data Scientist (Intern) Remote
  - Proposed and implemented MLPR, a novel end-to-end multi-task learning framework with domain-specific BERT, multi-expert architecture, and probability transfer, leveraging uncertainty-weighted loss for joint optimization of multiple engagement objectives (click, add-to-cart, purchase) in e-commerce search.
  - AUC of Click was improved by 6.48% over XGBoost on walmart.com Dataset. Paper accepted by *ACM WWW*.
- **Beijing QingLan Information Technology Co., Ltd.** *2016.08 - 2019.08*
  The Technical Director (Full time) Beijing, China
  - Designed and deployed News Recommendation System / Online Advertising System, REST API based service for over 30 media websites/mobile App.
  - Developed and productionized a semantic vector–based recommendation engine for content feeds and article pages, leveraging machine learning and ranking models to improve user engagement and ad targeting effectiveness.
- **Beijing Ruangao Information Technology Co., Ltd.** *2015.11 - 2016.07*
  Senior Algorithm Engineer (Full time) Beijing, China
  - Developed budget-constrained bidding algorithms for real-time bidding (RTB) advertising, combining CTR prediction with pacing strategies and user feedback signals (clicks, conversions) to improve targeting precision, reduce advertiser spend, and ensure effective delivery to high-intent user segments in production.

PUBLICATIONS

C=CONFERENCE, J=JOURNAL, W=WORKSHOP, S=IN SUBMISSION, *=FIRST/CORRESPONDING AUTHOR

[C.1] **Wu, X.***, Nian, J., Wei, T. R., Tao, Z., Wu, H. T., & Fang, Y. (2025). Does Reasoning Introduce Bias? A Study of Social Bias Evaluation and Mitigation in LLM Reasoning. *Findings of the Association for Computational Linguistics: EMNLP 2025*.

[C.2] **Wu, X.***, Wang, Y., Wu, H. T., Tao, Z., & Fang, Y. (2025). Evaluating Fairness in Large Vision-language Models Across Diverse Demographic Attributes and Prompts. *Findings of the Association for Computational Linguistics: EMNLP 2025*.

[C.3] Liu, H., **Wu, X.**, Sun, G., Tao, Z., & Fang, Y. (2024). RaCT: Ranking-aware Chain-of-Thought Optimization for LLMs. *Proceedings of the 2025 Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region*.

[C.4] **Wu, X.***, Li, S., Wu, H. T., Tao, Z., & Fang, Y. (2024). Does RAG Introduce Unfairness in LLMs? Evaluating Fairness in Retrieval-Augmented Generation Systems. In *Proceedings of the 2025 Joint International Conference on Computational Linguistics (COLING 2025)*.

[C.5] Wang, Y., **Wu, X.**, Wu, H. T., Tao, Z., & Fang, Y. (2024). Do Large Language Models Rank Fairly? An Empirical Study on the Fairness of LLMs as Rankers. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers) (NAACL 2024)*, 5712-5724.

[C.6] **Wu, X.***, Magnani, A., Chaidaroon, S., Puthenputhussery, A., Liao, C., & Fang, Y. (2022). A Multi-task Learning Framework for Product Ranking with BERT. *Proceedings of the ACM Web Conference 2022 (WWW 2022)*, 493–501.

[C.7] Chen, H., Fei, X., Wang, S., Lu, X., Jin, G., Li, W., & **Wu, X.** (2014). Energy Consumption Data based Machine Anomaly Detection. *In 2014 Second International Conference on Advanced Cloud and Big Data (CBD 2014)*, 136–142.

[J.1] **Wu, X.***, Puthenputhussery, A., Shang, H., Kang, C., & Fang, Y. (2024). Meta-Learning to Rank for Sparsely Supervised Queries. *ACM Trans. Inf. Syst (TOIS)*. 43, 1, Article 14 (January 2025), 29 pages.

[J.2] Peng, Z., **Wu, X.***, Wang, Q., & Fang, Y. (2024). Soft Prompt Tuning for Augmenting Dense Retrieval with Large Language Models. *Knowledge-Based Systems (KBS)*, 112758.

[J.3] Vincent, S., **Wu, X.**, Huang, M., & Fang, Y. Could Quoting Data Patterns Help in Identifying Journalistic Behavior Online?. In *International Symposium on Online Journalism (ISOJ)*, (p. 33).

[W.1] **Wu, X.***, Peng, Z., Sai, K. S. R., Wu, H. T., & Fang, Y. Passage-specific Prompt Tuning for Passage Reranking in Question Answering with Large Language Models. In *The Second Workshop on Generative Information Retrieval (Gen-IR)*.

[W.2] **Wu, X.***, Gao, X., Zhang, W., Luo, R., & Wang, J. (2019). Learning over Categorical Data Using Counting Features: With an Application on Click-through Rate Estimation. In *Proceedings of the 1st International Workshop on Deep Learning Practice for High-Dimensional Sparse Data (DLP-KDD)*, 1-9.

[S.1] Wei, T. R., Liu, H., **Wu, X.**, & Fang, Y. (2025). A Survey on Feedback-based Multi-step Reasoning for Large Language Models on Mathematics. arXiv preprint arXiv:2502.14333.

[S.2] Peng, Z., **Wu, X.***, Wang, Q., Rajanala, S., & Fang, Y. (2024). Q-PEFT: Query-dependent Parameter Efficient Fine-tuning for Text Reranking with Large Language Models. arXiv preprint arXiv:2404.04522.

[S.3] Tian, Y., Shao, T., Demizu, T., **Wu, X.***, & Wu, H. T. (2024). HPE-CogVLM: New Head Pose Grounding Task Exploration on Vision Language Model. arXiv preprint arXiv:2406.01914.

[S.4] Wei, T. R., Liu, H., Hu, H. C., **Wu, X.**, Fang, Y., & Wu, H. T. (2024). CLERF: Contrastive LEaRning for Full Range Head Pose Estimation. arXiv preprint arXiv:2412.02066.

[S.5] Hu, H. C., **Wu, X.**, Wang, Y., Fang, Y., & Wu, H. T. (2024). Mathematical Foundation and Corrections for Full Range Head Pose Estimation. arXiv preprint arXiv:2403.18104.

RESEARCH PROJECTS

- **Search, Ranking & LLM-based Optimization** *2019.09 - present*
  Tools: PyTorch, LLMs, Python
  - **Multi-Task Product Ranking** (*Published at ACM WWW 2022*): Developed an end-to-end multi-task learning framework combining BERT for product search ranking, significantly improving relevance and click prediction accuracy on a major e-commerce platform.
  - **Meta Learning-to-Rank for Sparse Data** (*Published at ACM TOIS*): Introduced a meta-learning approach to adapt ranking models effectively for sparsely supervised queries, achieving substantial improvements on public benchmarks and industrial datasets with limited user feedback.

- **Soft Prompt Tuning for Dense Retrieval** (*Published in Knowledge-Based Systems*): Proposed a soft prompt tuning technique augmenting dense retrieval with LLM-generated pseudo-queries, significantly boosting retrieval quality in low-data scenarios.
- **Passage-Specific Prompt Tuning for Reranking** (*Published at Gen-IR Workshop*): Designed a passage-specific prompt tuning strategy for question answering reranking, enhancing performance efficiently without extensive retraining.

- **Fairness in Large Language Models** *2023.09 - present*

Tools: PyTorch, LLMs, Python

- **Social Bias in LLM Reasoning** (*Published at EMNLP 2025*): Analyzed bias amplification during multi-step reasoning by large language models, developed lightweight bias-detection techniques, and achieved significant reductions in stereotype-aligned outputs.
- **Fairness in Large Vision-Language Models** (*Published at EMNLP 2025*): Investigated fairness across demographic attributes in vision-language models, revealing biases in zero-shot visual recognition tasks, and proposed a chain-of-thought-based bias mitigation approach.
- **Fairness in Retrieval-Augmented Generation** (*Published at COLING 2025*): Evaluated fairness risks in retrieval-augmented generation systems (RAG), identified bias propagation from retrieved documents into LLM-generated responses, and proposed a comprehensive bias evaluation framework.
- **Bias in LLM-Based Rankers** (*Published at NAACL 2024*): Conducted empirical analyses of demographic biases in LLM-based ranking, introducing novel evaluation methods to measure and mitigate unfairness, informing responsible AI guidelines.

Skills

- **Specialized Area:** Information Retrieval, Search and Ranking, Large Language Models, Responsible AI, Recommendation System.
- **Data Science & Machine Learning:** Meta learning, Multi-task learning, Recommendation algorithms, NLP, Online advertising algorithms.
- **Deep Learning Framework:** PyTorch, TensorFlow.
- **Programming Languages:** Python, Java, Spark, Hadoop, HTML5, Javascript, C/C++, etc.
- **Database Systems:** FAISS, MySQL, MongoDB, PostgreSQL, Redis, ElasticSearch, etc.
- **Other Skills:** Large-scale data analysis, excellent communication and organizational skills, outstanding team work ability.

Professional Activities

- **Journal Reviewer**
  - IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)
  - Neurocomputing
  - IEEE Access
  - Journal of the Frontiers of Computer Science
  - Connection Science
  - Complex & Intelligent Systems (CAIS)
- **PC Member / Reviewer**
  - International Conference on Learning Representations (ICLR)
  - ACM Special Interest Group on Information Retrieval (SIGIR)
  - ACM The Web Conference (TheWebConf)
  - ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)
  - ACL Rolling Review (ARR)
  - International Conference on Web Search and Data Mining (WSDM)
  - Conference on Information and Knowledge Management (CIKM)
  - ACM SIGIR Conference on Information Retrieval in the Asia Pacific (SIGIR-AP)
  - ACM International Conference on the Theory of Information Retrieval (ICTIR)
  - Conference on Language Modeling (COLM)
  - KDD Workshop on Deep Learning Practice for High-Dimensional Sparse Data (DLP-KDD)
  - The Second Workshop on Generative Information Retrieval (Gen-IR)