Xiangru Jian

David R. Cheriton School of Computer Science
University of Waterloo
Waterloo, Canada

E-mail: xiangru.jian@uwaterloo.ca
Website: Xiangru Jian
Google Scholar: Xiangru Jian

Education

- **University of Waterloo, Ph.D. in Computer Science.** Waterloo, Canada
  Supervised by Prof.Tamer Özsu. Focus on VLM/LLM agents and multimodal data management. *Sept. 2022 – Present*
- **City University of Hong Kong, M.Sc in Data Science.** Hong Kong SAR, China
  Supervised by Prof. Yu Yang. Focus on data mining and graph learning. *Aug. 2020 – Sept. 2021*
- **Tongji University, B.Eng.** Shanghai, China
  Major in Civil Engineering and minor in Mathematics. *Sept. 2014 – July. 2019*

Work Experience

- **ServiceNow Research, Visiting Researcher (Internship)** Montreal, Canada
  Focus on multimodal learning and GUI understanding. *April. 2024 - May. 2025*
- **Hong Kong Institute of Data Science at CityU, Research Assistant** Hong Kong SAR, China
  Focus on order prediction for industrial partners, such as Alibaba and HKTV Mall. *July. 2021 - June. 2022*
- **City University of Hong Kong, Research Assistant** Hong Kong SAR, China
  Focus on the landscape of DNNs. *June. 2021 - Aug. 2021*

Research areas

- **LLM and Multimodality**: Multimodal Reasoning; GUI Agents; Document Understanding; Natural Language Interface for Databases; Multi-agent System; Cross-modal Retrieval.
- **Machine Learning**: Graph Learning; Optimization; Set Learning and Generation.

Selected Publications and Preprints (*Co-first authors, †Corresponding author)

- Graph **Xiangru Jian***, Xinjian Zhao*, Wei Pang*, Chaolong Ying, Yimu Wang, Yaoyao Xu, Tianshu Yu. Rethinking Spectral Augmentation for Contrast-based Graph Self-Supervised Learning. **TMLR**, 2025. (Paper Link)
- Agent Multimodal Shravan Nayak*†, **Xiangru Jian***† et al. UI-Vision: A Desktop-centric GUI Benchmark for Visual Perception and Interaction. **ICML 2025**. (Paper Link, Project Site)
- Agent Multimodal Wei Pang*, Kevin Qinghong Lin*, **Xiangru Jian***, Xi He, Philip Torr. Paper2Poster: Towards Multimodal Poster Automation from Scientific Papers. Accepted by **NeurIPS** 2025. **Oral** at MAS Workshop at **ICML** 2025. (Paper Link, Project Site, 2.4k stars on GitHub)
- VLM/LLM Multimodal Juan Rodriguez*, **Xiangru Jian*** et al. BigDocs: An Open Dataset for Training Multimodal Models on Document and Code Tasks. **ICLR**, 2025. (Paper Link, Project Site)
- Retrieval Multimodal Graph **Xiangru Jian** and Yimu Wang. InvGC: Robust Cross-Modal Retrieval by Inverse Graph Convolution. Findings of **EMNLP** 2023. (Paper Link)
- Retrieval Multimodal Yimu Wang, **Xiangru Jian** and Xue Bo. Balance Act: Mitigating Hubness in Cross-Modal Retrieval with Query and Gallery Banks. **EMNLP** 2023. (**Oral**, Paper Link)
- Optimization Keke Wu, **Xiangru Jian**, Rui Du, Jingrun Chen and Xiang Zhou. Understanding Loss Landscapes of Neural Network Models in Solving Partial Differential Equations. **IEEE BigData** 2023. (**Oral**, Paper Link)

- Graph Xinjian Zhao, Wei Pang, Zhongkai Xue, **Xiangru Jian**, et al. The Underappreciated Power of Vision Models for Graph Structural Understanding. Accepted by **NeurIPS** 2025.
- Optimization Qixin Zhang, Zengde Deng, **Xiangru Jian**, Zaiyi Chen, Haoyuan Hu and Yu Yang. Communication-Efficient Decentralized Online Continuous DR-Submodular Maximization. **CIKM** 2023. (Paper Link)
- Graph Lyuyi Zhu, Qixin Zhang, **Xiangru Jian**, Yu Yang. Graph convolutional network for traffic incidents duration classification. Engineering Applications of Artificial Intelligence, Volume 151. (Paper Link)
- Data Management Agent **Xiangru Jian**, Zhengyuan Dong, M. Tamer Özsu. InteracSPARQL: an Interactive System for SPARQL Query Refinement Using Natural Language Explanations. (Preprint, Submitted to VLDB 2026)
- Data Management Multimodal **Xiangru Jian**, Wei Pang, Zhengyuan Dong, Chao Zhang, M. Tamer Özsu. LazyVLM: Neuro-Symbolic Approach to Video Analytics. (Preprint, (Paper Link)
- VLM/LLM Graph Hao Xu*, **Xiangru Jian***†, Xinjian Zhao*, Wei Pang*, Chao Zhang, Suyuchen Wang, Qixin Zhang, Joao Monteiro, Qiuzhuang Sun, Tianshu Yu†. GraphOmni: A Comprehensive and Extendable Benchmark Framework for Large Language Models on Graph-theoretic Tasks. (Preprint, Paper Link, Project Site)
- Data Management Optimization Yu Yin*, **Xiangru Jian***, Haoxiang Liu, Wei Gong, Yu Yang. A Pattern-based Subset Choice Model. (Preprint, Under review at ESWA)
- Graph Lyuyi Zhu, Qixin Zhang, **Xiangru Jian**, Yu Yang, Lishuai Li. Spatio-temporal Traffic Accidents Detection via Graph-based Generative Adversarial Network. (Preprint, Under review at EAAI)

Selected Research Projects

- **Research Intern at ServiceNow Research** — Montreal, Canada
  Multimodal Reasoning and GUI Agents — *April 2024 – Present*
  - **UI-Vision: A Desktop-centric GUI Benchmark for Visual Perception and Interaction**
    Developed an open-source benchmark for evaluating autonomous GUI agents in desktop environments. Designed and curated a richly annotated dataset spanning 83 software applications to assess visual perception, layout grounding, and action prediction. (ICML 2025, Paper Link, Project Site)
  - **BigDocs: An Open Dataset for Training Multi-Modal Models on Document and Code Tasks**
    Co-led the effort of BigDocs—a 7.5-million sample, license-permissive dataset for multimodal document understanding. The dataset supports tasks such as document parsing and long-format code generation (e.g., Screenshot2HTML and Table2LaTeX), advancing research in structured output and document reasoning. (ICLR 2025, Paper Link, Project Site)
- **Graduate Research Assistant at University of Waterloo** — Waterloo, Canada
  Multimodal Data Management, Cross-Modal Retrieval and Graph Learning — *Sept 2022 – Present*
  - **Paper2Poster: Towards Multimodal Poster Automation from Scientific Papers**
    Proposed and developed Paper2Poster, the first benchmark and metric suite for academic poster generation, and PosterAgent, a visual-in-the-loop multi-agent pipeline. The system transforms research papers into editable posters, outperforms GPT-4o, and was accepted for Oral at the MAS Workshop ICML 2025. (Paper Link, Project Site).
  - **GraphOmni: Benchmarking LLMs on Graph-Theoretic Reasoning**
    Developed GraphOmni, a comprehensive and extensible evaluation framework that systematically benchmarks large language models on graph reasoning tasks by varying graph types, serialization formats, and prompt schemes. Introduced an RL-based selector to dynamically choose optimal serialization–prompt combinations, yielding significant accuracy gains. (Preprint, Paper Link, Project Site)

- **Robust Cross-Modal Retrieval**
  1. Formulated the representation degeneration problem in cross-modal retrieval and proposed InvGC, the first post-processing solution. (EMNLP 2023, Paper Link)
  2. Developed DBNORM, a unified framework mitigating hubness by reducing similarity between hubs and queries. (EMNLP 2023, Paper Link)
  3. Further proposed DREAM, a relevance-based augmentation framework using language and visual generative models, significantly improving video-text retrieval. (NAACL 2025, Paper Link)
- **Graph Self-Supervised Learning**
  Conducted empirical studies revealing that simple edge perturbations (edge dropping and adding) offer comparable or superior performance to complex spectral augmentations in graph self-supervised learning, significantly reducing computational costs. Provided theoretical analysis supporting these findings and challenging conventional assumptions in spectral augmentation methods. (TMLR 2025, Paper Link)
- **Application of Foundation Models on Data Management** (Supervised by Prof. Tamer Özsu)
  Proposed a datalog-like logic form that unifies various graph query languages (e.g., SPARQL) as an intermediate language for natural language interfaces to graph databases. Designed a pipeline for open-vocabulary video query across multi-frame systems leveraging scene graph generation and a streaming graph OLAP architecture. (In submission to VLDB 2026)

- **Graduate Research Assistant at City University of Hong Kong** Hong Kong SAR, China
  Data Mining and Optimization in E-commerce and Deep Learning *Sept 2020 – Sept 2022*
  - **Submodular Maximization** (Supervised by Prof. Yu Yang)
    Developed one-shot decentralized algorithms (Mono-DMFW and DOBGA) that reduce per-round communications and gradient evaluations from $T^{3/2}$ to 1, achieving $(1-1/e)$-approximation with $O(T^{4/5})$ and $O(\sqrt{T})$-regret bounds, respectively. (CIKM 2023, Paper Link)
  - **Landscape of DNN** (Supervised by Prof. Xiang Zhou)
    Introduced the *Roughness Index* to quantitatively analyze the loss landscape of deep neural networks, revealing novel insights into optimization dynamics. (IEEE BigData 2023, Paper Link)

Professional Services

- Reviewer for: EMNLP 2023; ICLR 2025, 2026; ACL ARR; TKDD; TBD; ACI Material Science Journal.
- External Reviewer for: VLDB 2023, 2024; SIGMOD 2024; CIKM 2022, 2023; SIGIR 2022; KDD 2022.

Selected Honors and Awards

- Graduated with distinction, ranking 1st in the master's program.. City University of Hong Kong.
- Outstanding Graduate of Tongji University (Highest Honor, Top 5%), 2019.
- Second Prize of China Undergraduate Mathematical Contest in Modeling (Top 5%), 2016.
- National Scholarship (Top 2%), 2015.

Skills

- Programming Languages: Python, MATLAB, R, C#, SQL, Java, C++/C.
- Development and Deployment: Git, GitHub, Torch, Vim, Emacs, Linux/Unix, IntelliJ IDEA, Valgrind.
- Mathematics and Statistics: ODE and PDE, Dynamic Programming, Convex Optimization, Tensor Analysis, Numerical Analysis, Monte Carlo