Bowen Jin

Room 1117, Siebel Center for Computer Science, 201 N. Goodwin Ave, Urbana, IL 61801
Email: bowenj4@illinois.edu | Homepage: https://peterjin.me | Phone: 217-819-1796

RESEARCH INTERESTS

Large Language Models • Information Retrieval • Multimodality • Graph Models • Recommender System

My main research lies at the intersection of large generative models (e.g., large language models and diffusion models), multimodal data, and information networks. In particular, I focus on how large models can integrate text, network, and multimodal data for solving real-world problems including information retrieval and knowledge discovery.

- Large Generative Models + IR: How large models can benefit from or help retrieval?
- Large Generative Models + Graphs: How to conduct structured multimodal learning?
- Large Generative Models + Science: How large models could influence scientific discovery?

EDUCATION

University of Illinois at Urbana-Champaign *2021.08 - Present*
Ph.D. Student in Computer Science. Advisor: Prof. Jiawei Han

University of Illinois at Urbana-Champaign *2021.08 - 2023.12*
M.S. in Computer Science. Advisor: Prof. Jiawei Han
Thesis: Language Models as Semantic Indexers.

Tsinghua University *2017.09 - 2021.07*
B.S. in Electrical Engineering & Statistics. Advisor: Prof. Yong Li
Outstanding Graduates (Top 1%).

RESEARCH EXPERIENCE

University of Illinois at Urbana-Champaign *2021.09 - Present*
- Research Assistant. Data Mining Group. Advisor: Prof. Jiawei Han
- Project: Multimodal Representation Learning with Large Language Models (ICLR'23, ACL'23, KDD'23, etc)

Google *2024.05 - Present*
- Student Researcher. Google Cloud Research Group. Mentor: Dr. Jinsung Yoon and Sercan O. Arik.
- Project: Retrieval-augmented generation and Long-context LLMs.

Amazon.com Inc *2023.05 - 2023.12*
- Research Intern. Query Understanding Group. Mentor: Dr. Xianfeng Tang.
- Project: Semantic ID and Generative retrieval (ICML'24).

Microsoft *2022.05 - 2022.08*
- Research Intern. Information and Data Science Group. Mentors: Dr. Chenyan Xiong and Alec Berntson.
- Project: Dense Retrieval for Heterogeneous Data.

Microsoft *2020.09 - 2021.03*
- Research Intern. Social Computing Group. Mentors: Dr. Zheng Liu and Dr. Xing Xie.
- Project: Knowledge-empowered News Recommendation.

Tsinghua University *2018.09 - 2020.07*
- Research Assistant. Future Intelligence Lab. Advisor: Prof. Yong Li.
- Project: Recommendation with Graph Neural Networks (SIGIR'20)

SELECTED PUBLICATIONS

First author papers:

Bowen Jin, Jinsung Yoon, Jiawei Han and Sercan O. Arik.
Long-Context LLMs Meet RAG: Overcoming Challenges for Long Inputs in RAG.
(submitted to ICLR 2025).

Bowen Jin, Ziqi Pang, Bingjun Guo, Yu-Xiong Wang, Jiaxuan You, Jiawei Han.
InstructG2I: Synthesizing Images from Multimodal Attributed Graphs.
2024 Conference on Neural Information Processing Systems (NeurIPs 2024).

Bowen Jin, Gang Liu, Chi Han, Meng Jiang, Heng Ji, Jiawei Han.
Large Language Models on Graphs: A Comprehensive Survey.
IEEE Transactions on Knowledge and Data Engineering (IEEE TKDE 2024).

Bowen Jin, Chulin Xie, Jiawei Zhang, Kashob Kumar Roy, Yu Zhang, Zheng Li, Ruirui Li, Xianfeng Tang, Suhang Wang, Yu Meng, Jiawei Han.
Graph Chain-of-Thought: Augmenting Large Language Models by Reasoning on Graphs.
The 62nd Annual Meeting of the Association for Computational Linguistics (ACL 2024 Findings).

Bowen Jin, Hansi Zeng, Guoyin Wang, Xiusi Chen, Tianxin Wei, Ruirui Li, Zhengyang Wang, Zheng Li, Yang Li, Hanqing Lu, Suhang Wang, Jiawei Han, and Xianfeng Tang.
Language Models as Semantic Indexers.
The 41st International Conference on Machine Learning (ICML 2024).

Yu Zhang*, Xiusi Chen*, **Bowen Jin***, Sheng Wang, Shuiwang Ji, Wei Wang, Jiawei Han.
A Comprehensive Survey of Scientific Large Language Models and Their Applications in Scientific Discovery.
The 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024).

Kerui Zhu*, Bo-Wei Huang*, **Bowen Jin***, Yizhu Jiao, Ming Zhong, Kevin Chang, Shou-De Lin, Jiawei Han.
Investigating Instruction Tuning Large Language Models on Graphs.
The 1st conference on language modeling (COLM 2024).

Bowen Jin, Yu Zhang, Sha Li, and Jiawei Han.
Bridging Text Data and Graph Data: Towards Semantics and Structure-aware Knowledge Discovery.
The 17th ACM International Conference on Web Search & Data Mining (WSDM 2024). **Tutorial.**

Bowen Jin, Yu Zhang, Qi Zhu, and Jiawei Han.
Heterformer: Transformer-based Deep Node Representation Learning on Heterogeneous Text-Rich Networks.
The 29th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD 2023).

Bowen Jin, Wentao Zhang, Yu Zhang, Yu Meng, Xinyang Zhang, Qi Zhu, and Jiawei Han.
Patton: Language Model Pretraining on Text-rich Networks.
The 61st Annual Meeting of the Association for Computational Linguistics (ACL 2023).

Bowen Jin, Yu Zhang, Yu Meng, and Jiawei Han.
Edgeformers: Graph-Empowered Transformers for Representation Learning on Textual-Edge Networks.
The 11th International Conference on Learning Representations (ICLR 2023).

Bowen Jin, Chen Gao, Xiangnan He, Depeng Jin, and Yong Li.
Multi-behavior Recommendation with Graph Convolutional Networks.
The 43rd ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2020).

Bowen Jin, Wentao Zhang, Yu Zhang, Yu Meng, Han Zhao, and Jiawei Han.
Learning Multiplex Embeddings on Text-rich Networks with One Text Encoder.
(submitted to KDD 2025).

Other papers:

SeongKu Kang, **Bowen Jin**, Wonbin Kweon, Yu Zhang, Dongha Lee, Jiawei Han, Hwanjo Yu.
Improving Scientific Document Retrieval with Concept Coverage-based Query Set Generation.
The 18th ACM International Conference on Web Search and Data Mining (WSDM 2025).

Sizhe Zhou, Yu Meng, **Bowen Jin**, Jiawei Han.
Grasping the Essentials: Tailoring Large Language Models for Zero-Shot Relation Extraction.
The 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024).

Ran Xu, Wenqi Shi, Yue Yu, Yuchen Zhuang, **Bowen Jin**, May Dongmei Wang, Joyce C. Ho, Carl Yang
RAM-EHR: Retrieval Augmentation Meets Clinical Predictions on Electronic Health Records.
The 62nd Annual Meeting of the Association for Computational Linguistics (ACL 2024).

Tianxin Wei, **Bowen Jin**, Ruirui Li, Hansi Zeng, Zhengyang Wang, Jianhui Sun, Qingyu Yin, Hanqing Lu, Suhang Wang, Jingrui He, and Xianfeng Tang.
Towards Universal Multi-Modal Personalization: A Language Model Empowered Generative Paradigm.
The 12th International Conference on Learning Representations (ICLR 2024).

Hansi Zeng, Chen Luo, **Bowen Jin**, Sheikh Muhammad Sarwar, Tianxin Wei, Hamed Zamani.
Scalable and Effective Generative Information Retrieval.
The 2024 ACM Web Conference (WWW 2024).

SeongKu Kang, Shivam Agarwal, **Bowen Jin**, Dongha Lee, Hwanjo Yu, Jiawei Han.
Improving Retrieval in Theme-Specific Applications using a Corpus Topical Taxonomy.
The 2024 ACM Web Conference (WWW 2024).

Yu Zhang, **Bowen Jin**, Xiusi Chen, Yanzhen Shen, Yunyi Zhang, Yu Meng, and Jiawei Han.
Weakly-supervised Multi-label Classification of Full-Text Scientific Papers.
The 29th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD 2023).

Yu Zhang, **Bowen Jin**, Qi Zhu, Yu Meng, and Jiawei Han.
The Effect of Metadata on Scientific Literature Tagging: A Cross-Field Cross-Model Study.
The 2023 ACM Web Conference (WWW 2023).

Pengcheng Jiang, Shivam Agarwal, **Bowen Jin**, Xuan Wang, Jimeng Sun, and Jiawei Han.
Text-augmented Open Knowledge Graph Completion via Pretrained Language Models.
The 61st Annual Meeting of the Association for Computational Linguistics (ACL 2023 Findings).

Yu Zhang, Yanzhen Shen, SeongKu Kang, Xiusi Chen, **Bowen Jin**, Jiawei Han.
"Why Should I Review This Paper?" Unifying Semantic, Topic, and Citation Factors for Paper-Reviewer Matching.
(submitted to WWW 2025).

PROFESSIONAL SERVICES

Conference Program Committee Member
WSDM 2023; KDD 2023; NeurIPs 2023;
ICLR 2024; WWW 2024; SDM 2024; ICML 2024; ACL 2024; COLM 2024; NeurIPs 2024;
ICLR 2025; WWW 2025;

Journal Reviewer
IEEE Transactions on Knowledge and Data Engineering (TKDE);
ACM Transactions on Information System (TOIS);

Student Volunteer
KDD 2023

Guest Instructor
UIUC CS512 Fall 23

Lead TA
UIUC CS412 Spring 24

AWARDS

2024	NeurIPs Top Reviewer
2024	Yunni & Maxine Pao Memorial Fellowship
2024	Apple PhD Fellowship
2024	WSDM 2024 NSF Student Travel Grant
2023	KDD 2023 Student Travel Grant
2021	Outstanding Graduates, Tsinghua University (Top 1%)
2021	"Star of Tomorrow" Award, Microsoft Research
2020	National Scholarship (Top 1%)
2019	National Scholarship (Top 1%)
2019	Honorable Mention (top 15.35%), Mathematical Contest in Modeling
2018	National Scholarship (Top 1%)
2017	First Prize, National Olympiad in Mathematics in Provinces

TECHNICAL STRENGTHS

Skills	Machine Learning, Natural Language Processing, Language Modeling, Graph Mining, Weakly Supervised Learning, Unsupervised Learning, Information Retrieval
Programming Languages	Python, C/C++, MATLAB, R, Linux, Markdown, Shell, SQL
Machine Learning Packages	PyTorch, Keras, HuggingFace Transformers, Scikit-learn, PyTorch-Geometric, Deep Graph Library
Tools	Bash, LaTeX, Git