

Yikun Han

yikunhan@umich.edu | [homepage](#) | [scholar](#) | [dblp](#) | [github](#) | [linkedin](#)

Education

University of Michigan

Master of Data Science

- GPA 3.894/4.0

2023/08 – 2025/05

Ann Arbor, United States

Sichuan University

Bachelor of Information Resources Management

- Rank 2/76 - 3.81/4.0 - Major GPA 3.95/4.0

2019/09 – 2023/06

Chengdu, China

Peer-Reviewed Publications

Yijun Tian*, **Yikun Han***, Xiushi Chen*, Wei Wang, Nitesh V. Chawla. *Beyond Answers: Transferring Reasoning Capabilities to Smaller LLMs Using Multi-Teacher Knowledge Distillation* [[Link](#)] [Accepted by WSDM 2025, ACM International Conference on Web Search and Data Mining] (Acceptance rate: 17.3%)

Chunjiang Liu*, **Yikun Han***, Haiyun Xu, Shihan Yang, Kaidi Wang, Yongye Su. *A Community Detection and Graph Neural Network Based Link Prediction Approach for Scientific Literature* [[Link](#)] [Mathematics] (JCR Q1)

Preprints

Kyle Cox, Jiawei Xu, **Yikun Han**, Abby Xu, Tianhao Li, Chi-Yang Hsu, Tianlong Chen, Walter Gerych, Ying Ding. *Mapping from Meaning: Addressing the Miscalibration of Prompt-Sensitive Language Models* [Under review at AAAI 2025, AAAI Conference on Artificial Intelligence]

Zhi Jing*, Yongye Su*, **Yikun Han***, Bo Yuan, Haiyun Xu, Chunjiang Liu, Kehai Chen, Min Zhang. *When Large-Language Model Meets Vector Databases: A Survey* [[Link](#)]

Yikun Han, Chunjiang Liu, Pengfei Wang. *A Comprehensive Survey on Vector Database: Storage and Retrieval Technique, Challenge* [[Link](#)]

Competition

DREAM Olfactory Mixtures Prediction Challenge [RSGDREAM 2024] [[Slide](#)]

1st Place (tied)

Research Experience

Cell Maps for AI Knowledge Graph

2024/02 – Now

Co-supervised by [Ying Ding](#) and [Jiliang Tang](#)

[AI Health Lab](#)

- Curated a comprehensive dataset to recommend potential collaborators for biomedical publications.
- Developed a benchmarking framework that integrates large language models, knowledge graphs, and retrieval-augmented generation to extract sub-graphs and deliver personalized recommendations.

Digital Olfaction and Molecular Analysis

2023/08 – Now

Supervised by [Ambuj Tewari](#)

[LSA Statistics](#)

- Applied graph neural networks to generate molecular embeddings, achieving state-of-the-art performance in predicting the distance of olfactory mixtures.
- Developed a novel framework integrating transfer learning and semi-supervised learning to predict odor descriptors for individual molecules.

Innovations in Large Language Model Compression

2023/12 – 2024/03

Supervised by [Nitesh Chawla](#)

[Lucy Family Institute for Data & Soc](#)

- Introduced TinyLLM, an advanced framework for knowledge distillation, aimed at compressing large language models into more efficient, smaller models without significant loss in reasoning capabilities.
- Demonstrated TinyLLM's effectiveness through rigorous testing across six diverse datasets and in two complex reasoning tasks, showcasing its potential in reducing computational resources while maintaining high performance.

Advancing AI for Scientific Knowledge Discovery

2023/06 – 2024/01

Co-supervised by [Chunjiang Liu](#) and [Kehai Chen](#)

[Chinese Academy of Sciences](#)

- Investigated the integration of retrieval-augmented generation techniques and traditional fine-tuning approaches on zinc battery research literature, assessing their impact on various NLP tasks.
- Enhanced graph neural network efficiency by incorporating community detection algorithms into the link prediction process, leading to significant performance improvements in knowledge base construction for scientific research.

Enhancement of Symmetric Matrix Function Solutions via Aasen's Algorithm

2022/05 – 2022/11

Supervised by **Gang Chen**

Tianyuan Mathematical Center

- Significantly improved the efficiency of LAPACK functions through threading, dichotomy, optimal matrix chunk size adjustments, and the implementation of OpenMP tasks.
- Achieved a substantial 40-fold increase in computational performance for large-scale matrices by parallelizing factorization and back substitution processes, along with strategic segmentation of Aasen's algorithm's critical steps.

Awards

RSGDREAM Travel Award	2024
Outstanding Graduate	2023
Second Prize Scholarship	2022
Outstanding Student	2021
Outstanding Student	2020

Professional Experience

Data Analyst

2022/07 – 2023/03

Tencent

Shenzhen, China

- Developed more than 5 interactive dashboards and implemented 60+ components with Javascript and SQL, letting users get information without writing queries thus reply about 400% faster.
- Wrote 30+ Python scripts to crawl websites and replace Excel for data pre-processing, increasing the speed by nearly 25 times.

Teaching

STATS 315 / DATASCI 315 Statistics & AI (course development)

Winter, 2024

Community Service & Volunteering

Datawhale

2022/07 – Now

- Led project [video-clip-extraction-by-description](#), [hugging-audio](#), deeply involved in projects like [llm-cookbook](#), [llm-universe](#), [d2l-ai-solutions-manual](#), [whale-paper](#), [what-is-vs](#).
- Wrote installation and implementation tutorials, prepared learning roadmaps, and organized relevant free courses as a teaching assistant for people who didn't have access to AI learning resources, such as [dive into deep learning](#).

STATCOM

2023/09 – Now

- Deeply involved in NLP project [OLHSA](#) and VIS project [MCYJ](#).

MDS ambassador, UMich

2024

Research Mentorship

Fengming (Stephen) Yang

BSCS@UMich