Yikun Han

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

Education

University of Michgan   2023/08 – 2025/05
Master of Data Science   Ann Arbor, United States
- GPA 3.8/4.0

Sichuan University   2019/09 – 2023/06
Bachelor of Information Resources Management   Chengdu, China
- Rank 2/76 - WES GPA 3.87/4.0 - Major GPA 3.95/4.0

Peer-Reviewed Publications

Chunjiang Liu*, **Yikun Han***, Haiyan 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)

Papers Under Review

Yijun Tian*, **Yikun Han***, Xiusi Chen*, Wei Wang, Nitesh Chawla. *TinyLLM: Learning a Small Student from Multiple Large Language Models* [Link] [Under review at COLM 2024, Conference on Language Modelling]

Zhi jing*, Yongye Su*, **Yikun Han***, Bo Yuan, Haiyan Xu, Chunjiang Liu, Kehai Chen, Min Zhang. *When Large-Language Model Meets Vector Databases: A Survey* [Link] [Under review at COLM 2024, Conference on Language Modelling]

Yikun Han, Chunjiang Liu, Pengfei Wang. *A Comprehensive Survey on Vector Database: Storage and Retrieval Technique, Challenge* [Link] [Under review at TBD, IEEE Transactions on Big Data]

Preprints

Kaipeng Wang, Zhi Jing, Yongye Su, **Yikun Han**. *Large Language Models on Fine-grained Emotion Detection Dataset with Data Augmentation and Transfer Learning* [Link]

Research Experience

Cell Maps for AI Knowledge Graph   2023/02 – Now
Supervised by **Ying Ding**   AI Health Lab
- Aimed to explore GNN with LLM to identify similar scientists.

Digital Olfaction and Molecular Analysis   2023/08 – Now
Supervised by **Ambuj Tewari**   LSA Statistics
- Developed algorithms to decipher the relationship between molecular structures and their perceived odors.
- Implemented and fine-tuned a graph neural network using the pyrfume dataset, exploring transfer learning to enhance model adaptability across diverse chemical datasets.
- Evaluated and compared model performance across GCN, MPNN, ENN, and 3DCNN.

Innovations in Large Language Model Compression   2023/12 – 2024/02
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.

Development of a Knowledge Base for Retrieval Language Terms with Chinese Perspectives 2021/03 – 2022/01
Supervised by **Wei Fan**
- Constructed a comprehensive Chinese knowledge base aimed at enhancing information retrieval capabilities, seamlessly integrating it into the Linked Open Data Cloud to enrich the theoretical framework with Chinese insights and methodologies.
- Contributed to the **DCMI Virtual 2021 Student Forum**, where I presented ongoing research on Linked Data and the Semantic Web, highlighting the project's advancements and its potential impact on the field.

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**, deeply involved in projects like **llm-cookbook**, **llm-universe**, **d2l-ai-solutions-manual**, **whale-paper**.
- 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 – 2024/04
- Deeply involved in NLP project **OLHSA** (Oakland Livingston Human Service Agency).

Awards

Outstanding Graduate of Sichuan University 2023
Sichuan University Second Prize Scholarship 2022
Outstanding Student of Sichuan University 2021
Outstanding Student of Sichuan University 2020