

Yikun Han

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

Education

University of Illinois

PhD in Information Sciences

2025/08 – 2030/05

Urbana-Champaign, United States

University of Michigan

Master in Data Science **GPA:** 3.97/4.00

2023/08 – 2025/05

Ann Arbor, United States

Sichuan University

Bachelor in Information Resources Management **GPA:** 3.81/4.00

2019/09 – 2023/06

Chengdu, China

Work Experience

Research Assistant, University of Michigan - with Advisor: [Ambuj Tewari](#)

Aug 2023 – Present

• Digital Olfaction and Molecular Analysis

- ▶ 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.

Research Intern, University of Texas at Austin — with Advisors: [Ying Ding](#) & [Jiliang Tang](#)

Feb 2024 – Present

• Uncertainty Calibration in Prompt-Sensitive LLMs [1]

- ▶ Introduced a semantic-concept perturbation framework: sampling across paraphrases of inputs to model prompt sensitivity as a form of generalization error.
- ▶ Developed a novel uncertainty decomposition metric that separates epistemic (meaning-based) and aleatoric (token-based) uncertainty for black-box LLMs.
- ▶ Demonstrated improved calibration on QA tasks when perturbation sampling is applied, and provided evidence that RLHF models are more prompt-sensitive than base models.

Research Intern – University of Notre Dame - with Advisor: [Nitesh V. Chawla](#)

Dec 2023 – Mar 2024

• Multi-teacher reasoning distillation [2]

- ▶ Designed TinyLLM, a multi-teacher knowledge-distillation framework that transfers both answers and chain-of-thought reasoning from multiple large LLMs into a compact student model.
- ▶ Developed an in-context example generator and teacher-forcing Chain-of-Thought strategy to ground rationale generation effectively.
- ▶ Achieved up to +23.4% improvement over teacher models and +15.7% over standard fine-tuning across six datasets and two reasoning benchmarks.

Research Intern - Chinese Academy of Sciences

Jun 2023 – Jan 2024

• Survey on VecDB-LLM technologies [3], [4]

- ▶ Co-authored a comprehensive survey examining how vector databases mitigate key LLM challenges—hallucinations, outdated knowledge, prohibitive API costs, and memory decay—by enabling scalable retrieval and external memory support. [3]
- ▶ Reviewed core VecDB techniques—hashing, tree-based, graph-based, quantization, and system architectures—and highlighted open issues in storage, indexing, retrieval, and LLM integration. [4]

• Graph-enhanced link prediction [5]

- ▶ Designed a community-detection + GNN pipeline for link prediction in scientific literature, yielding +4–10 pts AUC gains versus baseline GNNs.
- ▶ Applied this pipeline to zinc-battery literature, enhancing scientific relation graph construction.

Competition

DREAM Olfactory Mixtures Prediction Challenge [RSGDREAM 2024] [Slide]

1st Place (tied)

Awards

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

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 – 2025/05
----------------	-------------------

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

MDS ambassador, UMich	2024
-----------------------	------

Bibliography

- [1] K. Cox, J. Xu, Y. Han, R. Xu, T. Li, and et al., "Mapping from Meaning: Addressing the Miscalibration of Prompt-Sensitive Language Models," in *Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence*, 2025, pp. 23696–23703. doi: [10.1609/aaai.v39i22.34540](#).
- [2] Y. Tian*, Y. Han*, X. Chen*, W. Wang, and N. V. Chawla, "Beyond Answers: Transferring Reasoning Capabilities to Smaller LLMs Using Multi-Teacher Knowledge Distillation," in *Proceedings of the Eighteenth ACM International Conference on Web Search and Data Mining*, in WSDM '25. Hannover, Germany: Association for Computing Machinery, 2025, pp. 251–260. doi: [10.1145/3701551.3703577](#).
- [3] Z. Jing*, Y. Su*, and Y. Han*, "When Large Language Models Meet Vector Databases: A Survey," in *2025 Conference on Artificial Intelligence x Multimedia (AIXMM)*, 2025, pp. 7–13. doi: [10.1109/AIXMM62960.2025.00008](#).
- [4] L. Ma*, R. Zhang*, Y. Han*, S. Yu, Z. Wang, and et al., "A Comprehensive Survey on Vector Database: Storage and Retrieval Technique, Challenge." [Online]. Available: <https://arxiv.org/abs/2310.11703>
- [5] C. Liu*, Y. Han*, H. Xu, S. Yang, K. Wang, and Y. Su, "A Community Detection and Graph-Neural-Network-Based Link Prediction Approach for Scientific Literature," *Mathematics*, vol. 12, no. 3, 2024, doi: [10.3390/math12030369](#).