Yunan Zhang

3307 Hopkins Dr, West Lafayette, IN 47906

312-241-0055 zhan4404@purdue.edu linkedin.com/in/yunan-zhang-641351183

Education

Purdue University – West Lafayette **Aug. 2023 - 2027**
Ph.D. Student in Computer Science (Advisor: Prof. Jianguo Wang) *West Lafayette, IN*

Purdue University – West Lafayette **June 2021 - May 2023**
M.S. in Computer Science *West Lafayette, IN*

RESEARCH & PROJECT

Filtered Vector Search on DiskANN | *Microsoft Research* **June 2025 – Now**

- Aiming to develop a search strategy adaptable to varying query selectivities and predicate cardinalities.
- Planning to publish a benchmark with new datasets for filtered vector search for the database community.

DiskV: Disk-based Vector Search on Commodity Storage | *Purdue University* **June 2024 – Now**

- The first disk-based vector index designed to achieve high search performance across different storage devices, including both low-cost storage (e.g., HDDs and cloud storage) and high-speed storage (e.g., SSDs).
- DiskV outperforms 5 SOTA disk-based vector databases with minimal index build time and index size.
- Submitted to **VLDB'26** as **first-author**.

Specialized vs. Generalized Vector Databases | *Purdue University* **Apr. 2022 – May 2023**

- Analyzed performance disparities between generalized and specialized vector databases using Alibaba PASE and Facebook Faiss in terms of index construction time, index size, and search time for three types of representative indexes (IVFFLAT, IVFPQ and HNSW).
- Provided insights and directions for building a future generalized vector database that can achieve comparable performance to a high-performance specialized vector database.
- Published in **ICDE'24** as **first-author**.

RELEVANT WORK EXPERIENCE

Research Scientist Internship | *Microsoft Research* **June 2025 – Aug. 2025**

- Interned on the Microsoft Azure DiskANN team.
- Created an internal dataset for filtered vector search.
- Built a new filtered vector search strategy for DiskANN, delivering a 2× search-performance speedup over the existing solution; planned for deployment to Azure Cosmos DB in 2026.

PUBLICATIONS

- **Yunan Zhang**, Yinghao He, Jiayi Liu, Jianguo Wang. "DiskV: Disk-based Vector Search Optimized for Low-Cost Storage". Submitted to **VLDB** 2026.
- **Yunan Zhang***, Shige Liu*, Jianguo Wang. "Are There Fundamental Limitations in Supporting Vector Data Management in Relational Databases? A Case Study of PostgreSQL". In **ICDE** 2024. https://www.cs.purdue.edu/homes/csjgwang/pubs/ICDE24_VecDB.pdf
- Jiayi Liu, **Yunan Zhang**, Chenzhe Jin, Aditya Gupta, Shige Liu, Jianguo Wang. "Fast Vector Search in PostgreSQL: A Decoupled Approach". In **CIDR** 2026.
- Chenzhe Jin, **Yunan Zhang**, Jiayi Liu, Jianguo Wang. "Efficient Vector Index Merging in Vector Databases". In **SIGMOD** 2026.
- Cheng Chen, Chenzhe Jin, **Yunan Zhang**, Sasha Podolsky, Chun Wu, Szu-Po Wang, Eric Hanson, Zhou Sun, Robert Walzer, Jianguo Wang. "SingleStore-V: An Integrated Vector Database System in SingleStore". In **VLDB** 2024. https://cs.purdue.edu/homes/csjgwang/SingleStoreVec.pdf