

Yinqi Sun | PhD Student

🌐 <https://billsun.dev>

✉ sunyinqi0508@gmail.com

🐙 github.com/sunyinqi0508

Education

Purdue University, Department of Computer Science

West Lafayette, IN

Ph.D. in Computer Science **Advisor: Prof. Ilia Sucholutsky**

2023 -

New York University, Courant Institute of Mathematical Sciences

New York, NY

M.S. in Computer Science, GPA: 3.84/4.0

2020 - 2022

Shandong University, Elite Class

Shandong, China

B.S. in Computer Science

Standardized Test Scores: TOEFL iBT: 113

Publications

- Y. Su, **Yinqi Sun**, Minjia Zhang, Jianguo Wang, “Vexless: A Serverless Vector Data Management System Using Cloud Functions” (SIGMOD’24)
- Yunhai Wang, Y. Wang, **Yinqi Sun**, L. Zhu, K. Lu, C. Fu, M. Sedlmair, O. Deussen, and B. Chen “Revisiting Stress Majorization as a Unified Framework for Interactive Constrained Graph Visualization”, in IEEE Transactions on Visualization and Computer Graphics (IEEE VIS), vol. 24, no. 1, pp. 489-499, Jan. 2018, doi: 10.1109/TVCG.2017.2745919.
- Yunhai Wang, Y. Wang, **Yinqi Sun***, H. Zhang*, C. Fu, M. Sedlmair, B. Chen, O. Deussen “Structure-aware Fisheye Views for Efficient Large Graph Exploration”, in IEEE Transactions on Visualization and Computer Graphics (IEEE VIS), vol. 25, no. 1, pp. 566-575, Jan. 2019, doi: 10.1109/TVCG.2018.2864911.
- Z. Yuan, **Yinqi Sun**, Dennis Shasha, “Forgetful Forests: Data Structures for Machine Learning on Streaming Data under Concept Drift”, in MDPI Algorithms, 2023; 16(6):278, doi: 10.3390/a16060278
- **Yinqi Sun**, “An Exact 56-Addition, Rank-23 Scheme for General 3×3 Matrix Multiplication”, arXiv Preprint, doi: 10.48550/arXiv.2604.27645

Research

Purdue University, Ph.D. Student

Aug 2023 –

Advisor: Prof. Ilia Sucholutsky

<https://ilia10000.github.io>

- Conducting research on AI Interaction with Human and Cognitive Science.
- Researching vector databases, machine learning systems, and serverless data processing.

New York University, Research Assistant

Jan 2022 – Aug 2023

Advisor: Prof. Dennis Shasha

<https://cs.nyu.edu/~shasha>

- Designed and implemented AQuery++, an in-memory column-store DBMS with compiled query execution, extensible UDF/plugin support, triggers, and time-series analytics.
- Built support for incremental decision-tree workloads on top of AQuery++, connecting database execution with online and streaming machine learning.
- AQuery++ is used as a teaching platform for NYU’s Advanced Database Systems course (CSCI-GA.2434).

Columbia University, Research Staff

Aug 2018 – Jan 2020

Advisor: Prof. Eugene Wu

<http://www.cs.columbia.edu/~ewu/>

- Conducted research related to Data Provenance, database architecture and compiled query execution.
- Proposed a physical-level lineage capturing scheme that reduces the lineage capture overhead and lineage size by:
 - taking advantage of patterns that each physical operators in the query plan reorganizes the data,
 - using lazy lineage materialization

Shandong University, Research Assistant

Sept 2016 – Aug 2018

Advisor: Prof. Yunhai Wang

<https://www.yunhaiwang.net>

- Researched graph visualization, dimensionality reduction and vector similarity search.

- Worked on improving over Stress Majorization (MDS) algorithm to allow for custom constraints (Vis'17) and a fisheye view algorithm that focuses on minimizing structural distortion (Vis'18).
- Worked on code optimization for lab projects and parallel algorithms on GPU (CUDA).
- Studied saliency detection algorithms based on kNN (LSH and randomized kd-tree forest) that are used in Vector field visualization and built an interactive vector field visualizer.

Work

eBay Inc., PhD Intern

May 2026 – Aug 2026

- Designed and implemented a Fast, Fresh, and Resilient Hybrid Indexing Pipeline for a Cloud Database Platform
- Built a multi-tier indexing pipeline with Real time cache (distributed kv store + vector), Near real-time ingestion based on Flink jobs, Bulk Rebuild with Spark.
- On billion scale data, with sharding and replication, 1000 qps mutation, millisecond hybrid query latency, > 0.95 recall, < 1s update visibility, Read Committed isolation guarantees.

eBay Inc., PhD Intern

May 2025 – Aug 2025

- Optimized components of an internal vector database for retrieval and analytics workloads.
- Built an end-to-end embedding and indexing pipeline that integrated embedding models with the vector database, reduced cross-system data movement, and simplified AI-powered data enrichment and analytics.

Teaching

Graduate Teaching Assistant for CS348, Information Systems

2023-2026 (7 Semesters)

- Developed automatic grading scripts on Gradescope for homework and lab exams.
- Developed a complete labexam arrangement/distribution/grading pipeline that include:
 - Session assignment system that uses a custom optimizer to find the optimal session assignment that maximizes student's satisfaction under the constraint that every student is available for the assigned session.
 - Automatic exam paper distribution based on a hashing algorithm that ensures equal distribution of exam versions.
- Assisted in homework/exams designs and gradings, held office hours and recitation sessions.

Projects (Pre-2022, no agentic coding)

Raytracing in GLSL

Feb - May 2021

- Real-time interactive ray tracing with refractions in a fragment shader of WebGL. Built from ground up without external libraries, more descriptions on demo page. see: <https://billsun.dev/graphics/hw4> Other weekly projects like this: <https://billsun.dev/graphics/>

AQuery++ Database

Jun - Nov 2022

- A database management system featuring Column-Store, Compiled Query Execution, Flexible User Defined Functions, builtin timeseries functions and much more. see: <https://github.com/sunyinqi0508/AQuery2>

Interactive Vector Field Visualization Tool

2017-2018

- A Vector Field Visualization tool written with QT (C++) that has kNN based saliency highlighting using KD tree and LSH implemented in CUDA kernels. see: <https://github.com/sunyinqi0508/cuLines>

ChocoPy Compiler

Feb - May 2021

- A compiler in Java that compiles ChocoPy (subset of Python) programs into RISC-V assembly. see: <https://git.billsun.dev/bill/Chocopy>

Skills

- Programming Languages:
 - Proficient in: C, C++, Python, JavaScript, Java, SQL, CUDA, GLSL.
 - Familiar with: Swift, C#, Matlab, Scala, Pascal, (x86, ARM, RISC-V) Assembly
- Professional Skills:
 - LaTeX, Bash, Git, Docker, PyTorch, Pandas, *nix, Windows, Adobe Illustrator, Photoshop
- Languages:
 - Native speaker of Chinese.
 - Fluent in English (TOEFL 113) and Japanese (JLPT N1).