Guanli Liu

liuguanli22@gmail.com · LinkedIn · GitHub
Melbourne, Australia · Open to relocation and remote · UTC+10/+11 · Global time overlap

Profile

Software engineer and postdoctoral researcher with a strong background in scalable backend systems, data infrastructure, and performance optimization. I have built and studied large-scale systems for data ingestion, query processing, indexing, and retrieval, with a focus on reliability, scalability, debuggability, and efficient execution. My PhD and postdoctoral work involved designing and implementing high-performance data systems published in top database venues such as VLDB and ICDE. In both research and industry settings, I have delivered end-to-end systems involving backend services, data pipelines, benchmarking frameworks, and query optimization components. Earlier in my career, I worked as a Software Engineer and Data Scientist, gaining practical experience in software development, distributed processing, and production-oriented data workflows.

Research and Work Experience

PhD + Postdoc, **The University of Melbourne**, Australia June. 2019 – Present

Lead research and engineering projects on database benchmarking, indexing, and AI-driven query processing. Design system prototypes, supervise junior researchers, and collaborate with academic and industry partners.

- **Data Layout and Physical Design**
  - Built a **layout advisory** system for data lake-style datasets, supporting dataset and workload ingestion, SQL workload parsing, and layout recommendation. *[VLDB 2026, submitted]*
  - Proposed a file **layout method** for data lake systems and developed an $\mathcal{O}(1)$ cost model to efficiently measure layout quality. *[VLDB 2025]*
- **Benchmarking and Evaluation**
  - Built **DriftBench**, a framework for evaluating performance stability under workload and data drift across database tuning and benchmarking scenarios. *[VLDB 2026]*
  - Built a **benchmarking framework** for traditional and learned spatial indexes, analyzing latency, I/O, and workload sensitivity under controlled query workloads. *[ICDE 2026]*
- **Indexing, Spatial Indexes and Learned Indexes**
  - Proposed a learned index construction method based on model reuse and fine-tuning, improving the efficiency and adaptability of index learning. *[ICDEW 2023]*
  - Built learned spatial index structures for multi-dimensional data, enabling efficient spatial query processing with improved lookup performance. *[ICDE 2023]*
  - Developed recursive learned spatial index structures and studied their effectiveness for spatial query workloads. *[VLDB 2020]*
- **Other (Supervised Works)**
  - Supervised work on learned cardinality estimation models. *[ICDE 2026: CoLSE]*
  - Supervised work on LLM-assisted processing of complex spatial queries. *[ADC 2025]*
  - Supervised work on embedding-based spatial keyword query processing.*[ADC 2025]*

Data Scientist, **nftDb**, Australia Feb. 2023 – Feb. 2024

A platform for NFT market trend tracking and multimodal search over digital assets.

- Built Python-based ingestion pipelines for raw blockchain transaction data, supporting Kafka-based streaming and Airflow-orchestrated batch workflows for downstream analytics.
- Developed SQL and dbt workflows to clean, normalize, and model large-scale transaction data into analytics-ready tables for product insights and reporting.
- Queried and analyzed transaction-level datasets in BigQuery to uncover user behaviour, marketplace activity, and performance trends across digital asset platforms.

- Developed an internal RAG-based knowledge assistant for technical documentation retrieval, data asset discovery, and workflow support across engineering teams.
- Partnered with engineers to improve pipeline reliability, data quality, and metric consistency across the end-to-end analytics and reporting stack.

Software Engineer, **Baidu**, China Jul. 2015 – Aug. 2017

Worked on Baidu's IM platform serving employees and business partners.

- Designed messaging protocols and implemented message deduplication mechanisms to improve delivery reliability and consistency across large-scale communication workflows.
- Improved database performance through profiling, query optimization, and systematic tuning of backend data access workflows.

Engineering Skills

- **Programming:** Proficient in Python and Java, with working knowledge of C++ and SQL
- **Data Systems & Pipelines:** Experience with SparkSQL, Apache Hudi, REST APIs, data ingestion pipelines, and batch / incremental data processing
- **Databases:** Experience with PostgreSQL, PostGIS, pgvector, BigQuery, and spatial data management
- **AI / Retrieval Systems:** Experience with RAG pipelines, vector databases, embedding-based retrieval
- **Cloud & Infrastructure:** Experience with Google Cloud Platform and Docker
- **Machine Learning:** Hands-on experience with PyTorch, TorchLib, Scikit-learn, and common machine learning algorithms for classification, regression, clustering, and representation learning
- **Algorithmic Knowledge:** Solid understanding of fundamental data structures and algorithms

Education

PhD in Computer Science, *The University of Melbourne*, Australia 2019 – 2023

M.S. in Computer Technology, *Northeastern University*, China 2013 – 2015

B.Eng. in Software Engineering, *Northeastern University*, China 2009 – 2013

Teaching

- **COMP90018 – Android Application Development (The University of Melbourne):** Tutor from Aug. 2019 to 2023, responsible for tutorials, student support, and assessment marking.
- **COMP90041 – Programming and Software Development (The University of Melbourne):** Tutor from Aug. 2019 to 2023, responsible for tutorials and assessment marking

Research Service

- **Conference Reviewer: SIGMOD** 2026, **VLDB** 2027 (PC), 2026 (Shadow PC), 2025 (External), KDD 2026, 2025 (Excellent Reviewer)
- **Journal Reviewer:** TKDE, WWW, Transactions on Spatial Algorithms and Systems (TSAS)