

Mingzhuo Yin

Xi'an Jiaotong University | B.Eng. in Computer Science | Class of 2026

✉ yinmingzhuo@gmail.com | 🌐 github.com/silver-ymz | ☎ +86 173-8208-4112 | 📞 ymz31415926535

Education

Xi'an Jiaotong University | B.Eng. in Computer Science & Technology (Honors Track) 2022.09 – 2026.07

- Admitted through the **Young Gifted Class**, a national program for early university admission, bypassing middle and high school entrance exams.
- Overall GPA:** 3.75/4.0 (86.27/100)
- Core Coursework:** Intro to Computer Systems (97), Formal Languages & Compilation (93), Operating Systems (90), Computer Networks (89), Linear Algebra (95), Discrete Mathematics (94), Optimization Methods (91), Probability Theory (89).

Work Experience

ByteDance | Database System R&D Intern (ByteHouse Storage Team) 2025.07 – Present

- Developing the cloud data warehouse, **ByteHouse**, a **cloud-native** platform with decoupled storage and compute, multi-tenancy, and elastic scalability.
- Leading the design and initial implementation of the Auto Partition feature:** Building an automatic partitioning and Min-Max index pruning scheme based on Snowflake's micro-partitioning concepts to enhance data management automation and query efficiency.
- Contributing to the development, testing, and maintenance of core modules including **Merge/Mutate** and **Transactions**.

Tensorchord | Software R&D Intern 2023.10 – 2025.06

- As a core member during the company's **seed stage**, played a key role in the 0-to-1 development of core products.
- Served as a core developer for **pgvecto.rs**, a PostgreSQL extension for vector similarity search, contributing 30% of the core codebase.
 - Conducted in-depth research on cutting-edge vector indexing algorithms like HNSW and IVF, implementing findings from academic papers into the system.
 - Independently designed and implemented support for indexing and retrieving **sparse vectors** and **binary vectors**, significantly broadening the plugin's applicability.
 - Developed a **SIMD-accelerated solution for sparse vectors**, a novel approach among known open-source products, which greatly boosted retrieval performance.
- Led the development of **VectorChord-bm25**, a BM25 ranking extension for PostgreSQL, from scratch, fully implementing the **Block-WAND** algorithm.
 - Benchmark tests show its geometric mean QPS is **2.26x** that of **Elasticsearch** in comparable scenarios.

Apache OpenDAL | PMC Member & Committer 2023.04 – Present

- As a core contributor to the Apache Top-Level Project **OpenDAL**, committed to its vision of "**One Layer, All Storage**" by providing a unified, seamless data access layer.
- Deeply involved in implementing and optimizing the **SFTP** and **GCS** backends, and led the development of **C++** and **Haskell** bindings, significantly expanding the project's cross-language ecosystem.
- Nominated as a **Project Management Committee (PMC) Member** in recognition of sustained and outstanding contributions.

Research Experience

Xi'an Jiaotong University, Institute of AI & Robotics | Research Assistant 2023.09 – 2025.06

- Advised by Prof. Pengju Ren and Prof. Tian Xia, focusing on **compiler prefetching optimization** and **side-channel attacks**.
- Publication:** Published as the second student author at **ISCA 2025 (CCF-A)**, a top-tier conference in computer architecture. Gelin Fu, Tian Xia, Mingzhuo Yin, Prashant J. Nair, Mieszko Lis, Pengju Ren*, "Magellan: A High-Performance Loop-Guided Prefetcher for Indirect Memory Access", ISCA 2025.
- Problem:** In applications like graph analytics and sparse linear algebra, irregular **Indirect Memory Accesses (IMAs)** cause traditional hardware prefetchers to fail, creating severe memory bottlenecks. Existing software prefetchers also struggle with complex IMA patterns in nested loops.
- Contribution:** Proposed **Magellan**, a high-performance software prefetcher. Its key innovations are:

- Constructing a **Loop Dependence Graph (LDG)** to accurately identify both local and global IMA patterns across nested loops for the first time.
- Capturing inner-outer loop semantics to issue prefetches for both current and future iterations, vastly expanding prefetching opportunities.
- **Implementation:** Independently developed an **LLVM Pass** to automate Magellan’s analysis and prefetching strategies, enabling seamless integration into existing compiler toolchains.
- **Evaluation:** Across 14 memory-intensive benchmarks, Magellan reduced cache misses by 25% and dynamic instructions by 14% on average compared to the state-of-the-art software prefetcher, achieving a 1.14x average speedup (up to 1.41x).

Projects & Coursework

- **Mini-LSM:** Implemented a persistent key-value store from scratch based on the **LSM-Tree** architecture.
 - Features include **Write-Ahead Logging (WAL)**, Manifest file management, multi-level SSTable compaction, and **MVCC**.
- **Bustub (CMU 15-445):** Independently completed four core modules of a relational database kernel.
 - Implemented the Buffer Pool Manager, **B+Tree** Index, Query Executor, and 2PL-based Concurrency Control.
- **SysY Compiler (PKU Compiler Lab):** Built a compiler for a C subset (SysY) targeting **RISC-V** assembly.
 - Implemented lexical analysis, parsing, semantic analysis, IR generation (LLVM IR), and linear scan register allocation.
- **xv6-labs (MIT 6.S081):** Enhanced the core functionalities of the xv6 teaching OS.
 - Implemented Copy-on-Write, multithreading, a network driver, locks, mmap, and other key system features.
- **TCP/IP Stack (Stanford CS144):** Built a complete TCP/IP stack from the ground up, passing all automated tests.
 - Implemented IP, ICMP, ARP, and TCP protocols, including reliable transport, flow control, and congestion control.

Extracurricular Activities

- 0w1 Cybersecurity Club | President2024.05 – 2025.05
- As President, I was responsible for the club’s overall operations, including recruitment, and organizing tech talks and training sessions.
 - Studied cryptography, gaining familiarity with modern techniques such as DES, AES, and RSA.

- ## Skills
-
- **Programming Languages:** Rust, C/C++, Python, Haskell, JavaScript, Java, C#
 - **Languages:** Chinese (Native), English (Proficient in reading technical documentation and academic papers, CET-4: 546, CET-6: 492)