

# 尹铭卓

西安交通大学 | 计算机科学与技术 | 2026 届本科

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

## 教育背景

西安交通大学 计算机科学与技术 (试验班)

2022.09 – 2026.07

- 经由 **少年班** 全国选拔, 免中高考录取, 提前开始本科学习。
- 综合 GPA: 3.75/4.0 (86.27/100)
- 核心课程: 计算机系统导论(97), 形式语言与编译(93), 操作系统(90), 计算机网络(89), 线性代数(95), 离散数学(94), 优化方法(91), 概率论(89)

## 工作经历

字节跳动 数据库开发实习生 (ByteHouse 存储团队)

2025.07 – 至今

- 深度参与云原生数据仓库 **ByteHouse** 的研发。ByteHouse 基于 ClickHouse 修改, 作为 **云原生** 数据仓库, 实现了存算分离、多租户资源隔离与弹性伸缩等特性。
- 独立负责 Auto Partition 功能的前期设计与实现**: 目前正基于 Snowflake 微分区思想, 进行自动分区与 Min-Max 索引裁剪方案的开发, 旨在提升数据管理的自动化程度与查询效率。
- 参与并负责 **Merge/Mutate** 及 **Transaction** 核心模块的开发、测试与维护工作。

Tensorchord 开发实习生

2023.10 – 2025.06

- 作为公司 **种子轮阶段** 的极早期核心成员, 深度参与了核心产品从 0 到 1 的研发过程。
- 作为 PostgreSQL 向量搜索插件 **pgvecto.rs** 的核心开发者, 贡献了约 30% 的核心代码。
  - 深入调研了 **HNSW, IVF** 等前沿向量索引算法, 通过阅读大量学术论文, 将最新研究成果应用并实现在系统中。
  - 独立设计并实现了对 **稀疏向量** 和 **二值向量** 的索引与检索支持, 显著扩展了插件的应用场景。
  - 研究了 **稀疏向量 (Sparse Vector) 的 SIMD 加速** 方案, 在所有已知开源产品中尚属首创, 大幅提升了检索性能。
- 从零到一主导开发了 PostgreSQL 的 BM25 插件 **VectorChord-bm25**, 完整实现了 **Block-WAND** 算法。
  - 经 benchmark 测试, 其 QPS 几何平均值达到同等场景下 **Elasticsearch** 的 **2.26 倍**。

Apache OpenDAL PMC Member & Committer

2023.04 – 至今

- 作为 Apache 顶级项目 **OpenDAL** 的核心贡献者, 致力于实现其 “**One Layer, All Storage**” 的愿景, 为多种存储服务提供统一、无缝的数据访问层。
- 深度参与了 **SFTP** 和 **GCS** 后端的实现与优化, 并主动承担完成了 **C++** 与 **Haskell** 的语言绑定开发, 极大地扩展了项目的跨语言生态。
- 因持续的杰出贡献, 被社区提名为项目的 **PMC (项目管理委员会) 成员**。

## 科研经历

西安交通大学人工智能与机器人研究所 研究助理

2023.09 – 2025.06

- 师从任鹏举教授与夏天教授, 从事 **编译器预取优化** 与 **侧信道攻击** 领域的前沿研究。
- 研究成果**: 以学生第二作者身份在计算机体系结构领域的顶级学术会议 **ISCA 2025 (CCF-A 类)** 发表论文。  
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.
- 研究问题**: 在图计算、稀疏线性代数等应用中, **间接内存访问 (Indirect Memory Access, IMA)** 模式因其不规则性导致传统硬件预取器失效, 造成严重的访存瓶颈。现有软件预取器又难以处理嵌套循环中的复杂 IMA 模式。
- 核心贡献**: 提出并设计了名为 **Magellan** 的高性能软件预取器。Magellan 的创新点在于:
  - 构建 **循环依赖图 (Loop Dependence Graph, LDG)**, 首次实现了跨越嵌套循环的全局与局部 IMA 模式的精确识别。
  - 捕捉内外层循环的语义, 为当前和未来的循环迭代生成预取指令, 极大拓展了预取机会。
- 工程实现**: 独立开发了 **LLVM Pass**, 将 Magellan 的分析与预取策略自动化, 可无缝集成到现有编译流程中。
- 性能测试**: 经在 14 个内存密集型基准测试上验证, Magellan 相比最先进的软件预取器, 平均将缓存未命中率降低 25%, 动态指令数减少 14%, 取得了 1.14 倍的平均加速, 最高可达 1.41 倍。

个人项目与课程实验

- **Mini-LSM**: 从零实现了一个基于 **LSM-Tree** 的持久化键值存储引擎。
  - 实现了 **WAL** (预写日志)、Manifest 文件管理、SSTable 的多级 Compaction 以及 **MVCC** 并发控制。
- **Bustub (CMU 15-445)**: 独立完成了关系型数据库内核的四个核心模块。
  - 实现了缓冲池管理器 (Buffer Pool Manager)、**B+Tree** 索引、查询执行器以及基于 2PL 的并发控制。
- **SysY Compiler (PKU Compiler Lab)**: 实现了一个将 C 语言子集 (SysY) 编译到 **RISC-V** 汇编的编译器。
  - 完成了词法分析、语法分析、语义分析、中间代码生成 (LLVM IR) 和基于线性扫描的寄存器分配算法。
- **xv6-labs (MIT 6.S081)**: 完善了 xv6 教学操作系统的核心功能。
  - 实现了写时复制 (Copy-on-Write)、多线程调度、网络驱动、锁机制、mmap 等关键系统特性。
- **TCP/IP 协议栈 (Stanford CS144)**: 从零构建了一个功能完备的 TCP/IP 协议栈，通过了全部的自动化测试。
  - 实现了 IP, ICMP, ARP, TCP 协议，包括可靠传输、流量控制和拥塞控制。

校园经历

0w1 网络安全俱乐部 社长

2024.05 – 2025.05

- 担任社团社长，全面负责社团的日常运营、招新、技术分享和培训活动的策划与组织。
- 学习了密码学方向的相关知识, 熟悉 DES、AES、RSA 等现代密码学技术

技能清单

- **编程语言**: **Rust, C/C++**, Python, Haskell, JavaScript, Java, C#
- **语言能力**: CET-4 (546), CET-6 (492), 能够流利阅读英文技术文档与论文