

Yuchen Zhou

yuchenz8@andrew.cmu.edu • (786) 978-0535 • [linkedin.com/in/yuchen-zhou-325513317/](https://www.linkedin.com/in/yuchen-zhou-325513317/) • github.com/ErwinZhou

EDUCATION

Carnegie Mellon University

Master of Science in Information Networking | GPA: 3.86/4.0

Pittsburgh, PA

Aug. 2025 – Dec. 2026

Nankai University

Bachelor of Engineering in Information Security | GPA: 3.71/4.0

Tianjin, China

Sep. 2021 – Jul. 2025

SKILLS

Languages: C++, Python, Java, Go, C, Rust, SQL (PostgreSQL, MySQL), NoSQL (Neo4j, DynamoDB), Shell, HCL
System & Backend: Git, OOP, System Design, CI/CD, Linux/Unix, Redis, Kafka, Docker, Kubernetes, MapReduce, Hadoop, Terraform, REST APIs, FastAPI, Database Internals, Distributed Systems, Microservices, Query Optimization
AI Infra & Cloud: AWS (Bedrock, S3, ECS, DynamoDB, EventBridge, AOSS), GCP, LLM, NLP, MCP, PyTorch, Hugging Face, Transformers, ReAct, Multi-Agent System, CUDA, Speculative Decoding, Claude Code, Codex

PROFESSIONAL EXPERIENCE

ENFOS Inc.

Software Development Engineering Intern

Durham, NC

May 2026 – Aug. 2026

- Designed and implemented a scalable ReAct-based multi-agent architecture for 100+ tools via a tool-discovery layer, reducing average tool-calling latency by 18% while improving tool-selection relative accuracy by 40%.
- Refactored MCP agent context engineering with hybrid retrieval (BM25 + semantic search) internal skill over messages, answers, and attachments for long-term memory and verbatim window for short-term, decreasing average context token cost by 61% while preserving model answer quality across 100+ turn conversations.
- Architected and provisioned AWS-based agent infrastructure using EventBridge, SQS, Lambda, DynamoDB, Bedrock, and AOSS to support async attachment ingestion, vector retrieval, and long-term agentic memory.

Chinasoft International

Backend Software Development Intern

Beijing, China

May 2023 – Jul. 2023

- Drove backend Hive/HDFS-based MapReduce workflow over 4M+ records, boosting query performance by 32%.
- Delivered Spring Boot backend services and REST APIs for interactive data analytics and virtualization while coordinating closely with frontend engineers and non-technical designers to ship product features.
- Engineered MySQL data ingestion and streaming systems using Sqoop, Kafka, and ZooKeeper, ensuring reliable backend data flow and reducing end-to-end processing latency by 27%.

ACADEMIC PROJECTS

Relational Database Systems – BusTub DBMS ([GitHub](#))

Feb. 2026 – May 2026

- Implemented core DBMS internal components in C++, including a disk-oriented buffer pool manager, B+ tree indexing, transaction-aware executors for concurrency, batch-based query execution engine, and optimization rules.
- Extended BusTub's rule-based optimizer with 5+ plan rewrite rules, including filter-scan merging, SeqScan and OrderBy to IndexScan selection, NestedLoopJoin to HashJoin, Sort-Limit-to-TopN optimization.
- Built operators (e.g., external merge sort) improving benchmark by 56% and ranking in the top 5% leaderboard.

Distributed Scholarly Search Engine

Feb. 2026 – May 2026

- Implemented a distributed scholarly search engine on a 3-node GCP cluster microservices using Docker, Kafka, MapReduce, and Terraform, processing 10K+ IEEE abstracts parsed from 1500+ pages of Google Scholar results.
- Improved query execution efficiency versus provided baseline solution by 37.2% by conditionally skipping MapReduce jobs for search-only requests, while supporting asynchronous multi-user query workflows on GCP.

Distributed Bitcoin Miner

Sep. 2025 – Dec. 2025

- Developed a custom reliable transport layer (LSP) protocol with Go, ensuring robust (sliding window & retransmission), in-order UDP communication without data races with >99% delivery accuracy.
- Built a distributed and parallel systems with dynamic mining task reassignment atop LSP, decreasing re-computation overhead by 43% to increase system throughput and improving fault tolerance under node crashes.