

Chenyu Wei

c63wei@uwaterloo.ca | 249-994-0088 | LinkedIn | GitHub | Portfolio

Education

University of Waterloo

Bachelor of Mathematics in Computer Science, Human-Computer Interaction Specialization

Waterloo, ON

Sep 2024 - Jul 2029

- Coursework: Data Structures, Operating Systems, Computer Networks, Databases, Software Engineering

Experience

Dameng Database Company

R&D Research Intern, Technology Research Department

May 2026 - Aug 2026

- Researched database-AI techniques for the technology research team, primarily reimplementing and evaluating a database-native function generation prototype for SQLite, PostgreSQL, and DuckDB.
- Extracted C/C++ declarations, registrations, types, and dependencies for model context; automated source updates, compilation, and tests so the team could evaluate generated functions with fewer manual steps.
- Worked with teammates on a Dify/MCP workflow that classified and summarized 378 database papers into JSON for internal research review; designed prompts and schema checks and revised outputs using reviewer feedback.
- Published three linked portfolio articles (click titles): [DBCooker/OpenCook](#), [Semantic Operator Estimation](#), and [Bob Metcalfe on Connectivity](#).
- Documented the prototype's setup, design, experiment results, and known limits; presented the work internally and handed the code and report to the R&D team for further evaluation.

Black Sesame Technologies

AI Application Intern

May 2025 - Aug 2025

- Designed and implemented the training-data pipeline for an AEB multimodal decision project and supported Qwen-VL fine-tuning; matched 50,000+ road images with radar data, vehicle signals, and JSON labels by frame ID.
- Automated prompt construction and internal model calls in Python, adding JSON field/type checks, bounded retries, error logs, and resume support to reduce manual repair and keep large batches consistent.
- Used OpenCV pixel-difference thresholds to select key frames from 20-frame sequences, reducing image input by about 80% while always retaining the final frame.
- Grouped samples into day/night and urban/highway sets using image brightness, timestamps, and road labels so the team could compare fine-tuning runs for different driving scenarios.

Projects

cydb - Persistent B+ Tree Database in C

Jul 2026 - Aug 2026

- Built a command-line database in C that stores fixed-schema records across restarts; implemented 4096-byte pages, a pager cache, serialization, B+ tree lookup and splitting, and ordered leaf traversal.
- Checked page capacity before insertions so an out-of-space error could not leave a partially modified tree; tested 1,200-row trees, reopen persistence, corrupt files, and on-disk invariants.

Text2SQL Agent - Validated Data Service

Jun 2026 - Aug 2026

- Built a Python service that turns questions into SQL after retrieving relevant schema; added SQLGlot AST rules and SQLite EXPLAIN checks before read-only execution.
- Recorded JSONL traces and created deterministic provider tests plus a 25-question regression set to compare changes and inspect failed queries.

ShellOps - Backend Delivery and CI Tooling

May 2026 - Jul 2026

- Automated repeatable Ubuntu setup with Bash for packages, Docker, SSH keys, environment files, and UFW; validation checks made reruns safe and exposed incomplete changes.
- Built a Git-hook CI flow that checks out the requested commit in an isolated workspace, runs Flask, Bats, and pytest checks, records durations, and cleans up temporary files.

paperBridge - Parallel Text Paper Translation Pipeline

Jan 2026 - May 2026

- Built a resumable workflow that parses research PDFs with MinerU, audits formulas, splits translation work into chunks, validates results, and merges bilingual Markdown.
- Stored file hashes and stage status in a manifest so interrupted runs could continue; tested missing chunks, math delimiters, code blocks, image paths, and bilingual markers.

Technical Skills

Languages: Python, SQL, JavaScript/TypeScript, Java, C/C++, Bash | **Backend/Data:** REST APIs, Flask, PostgreSQL
Engineering: SQLite, Linux, Docker, Git, CI/CD, JSON Schema, Unit/Integration Testing, Logging, Persistent Storage