Zikun Fu

Oshawa, Ontario Zikun.Fu@ontariotechu.net (437) 922-1214 LinkedIn GitHub

Summary

MSc Computer Science graduate (Ontario Tech, 2026) specializing in ML/NLP, AI automation, and data systems. Thesis formalized Database Entity Recognition, reaching 93.2% F_1 with a T5 + MiniLM pipeline (published at IEEE IRI 2025). Recently built grounded LLM systems: an analytics assistant measured at 100% grounding / 0% hallucination, and automated LLM pipelines with cost gating and human-review workflows. Strong Python, PyTorch/Transformers, SQL, Docker.

Key Achievements

Research & Publication: Formalized the DB-ER task; 80.8% F_1 open-world tagger, 93.2% F_1 with closed-world verification; published at IEEE IRI 2025.
Grounded LLM Analytics: Read-only, provenance-carrying analytics chatbot; 100% grounding accuracy and 0% hallucination on an oracle-backed harness; 59 unit tests.
LLM Automation: Graded ∼70 scanned handwritten finals via Claude for $10.34 total (∼$0.15/student) with confidence flags and human review.
Full-Stack Migration: Dockerized LimeSurvey 6.x (MySQL + Apache proxy); Python ETL via REST API importing 3,400+ multilingual records across 60+ research groups; DigitalOcean deployment with automated TLS.

Skills

Languages: Python, SQL, Kotlin, PHP, JavaScript, Bash
ML/NLP: PyTorch, Hugging Face Transformers (T5, BERT, MiniLM), scikit-learn; NER/sequence labeling, embeddings, semantic similarity, fine-tuning, evaluation harness design
LLM Engineering: grounded/tool-using agents, SQL validation gates (sqlglot), prompt design & caching, structured outputs, hallucination measurement, cost tracking; Claude API, OpenRouter, OpenAI-compatible APIs
Data & Pipelines: DuckDB, pandas, NumPy, ETL, SQL parsing (sqlglot, pglast), OR-Tools CP-SAT; MySQL, PostgreSQL, SQLite, MongoDB
Backend/Infra: REST API integration, Docker, Docker Compose, Apache reverse proxy, Linux, DigitalOcean, Git/GitHub, automated TLS (Let's Encrypt), pytest

Experience

Research Assistant *May 2026 - Present*
Ontario Tech University

- Building grounded LLM systems over structured business data: an analytics assistant with hard anti-hallucination guarantees (read-only DuckDB, sqlglot SQL allowlist, per-answer provenance; 100% grounding / 0% hallucination on an oracle-backed harness) and an automated web-presence audit pipeline with cost and human-review gates.

Teaching Assistant — CSCI 4020U Compilers *Jan 2024 - May 2026*
Ontario Tech University

- TA for the 4th-year compilers course (Kotlin), including lab support and exam marking.
- Built an LLM grading pipeline for scanned handwritten finals (Claude via OpenRouter): per-question image grading against model solutions, confidence/legibility flags, human-review workflow; ∼70 students for $10.34 total.

GREx Redevelopment Project Manager *Dec 2024 - Apr 2025*
Frazer Faculty of Education, Ontario Tech University

- Led migration from a legacy Symfony/Laravel stack to Dockerized LimeSurvey 6.x + MySQL; Python ETL via REST API importing 3,400+ multilingual records across 60+ research groups.
- Built PHP plugin modules and Chart.js dashboards; deployed to production with automated TLS.

Projects

Hermes Clinic Analytics — github.com/ZikunFu/hermes-clinic-analytics *2026*

- Grounded AI analytics chatbot for clinic business data: read-only DuckDB engine, sqlglot SQL allowlist, charts built from query results, exact SQL provenance on every answer; 100% grounding / 0% hallucination on an oracle-backed test harness; 59 unit tests + live bypass-safety suite.

Web Presence Report — automated local-SEO audit pipeline *2026 - ongoing*

- Automated pipeline for a local optometry clinic turning Google Business Profile, reviews, PageSpeed, geo-grid rankings, and backlinks into a graded client-ready report with a 30/60/90-day action plan; cost gates before paid API calls and human-review gates for health-content compliance.

Research

MSc Thesis: Database Entity Recognition using Language Models *Defended Apr 2026*

- Formalized DB-ER as sequence labeling over {O, Table, Column, Value}; released a 1,000-example human benchmark (Spider/BIRD) and a 15,000-example synthetic pipeline via SQL AST parsing and integer linear programming; T5-Large reaches 80.8% F_1, closed-world MiniLM verification raises it to 93.2%.

Publications

- **Z. Fu**, C. Yang, K. Davoudi, K. Q. Pu. "Database Entity Recognition with Data Augmentation and Deep Learning." *IEEE IRI 2025*, San Jose, CA. DOI: 10.1109/IRI66576.2025.00071
- Ontario Database Day 2024: Transforming Text-to-SQL Datasets into Closed-Domain NER Benchmarks

Education

Ontario Tech University — MSc, Computer Science *Jan 2024 – Apr 2026* — GPA: 4.24/4.3
Ontario Tech University — BSc (Hons) Computer Science (Data Science) *2019 – 2023* — GPA: 3.86/4.3