NORMAN BUI

Ottawa, ON | +1 613-983-3217 | normanbui23@gmail.com | linkedin.com/in/nbui23 | github.com/nbui23

EDUCATION

Carleton University | *Bachelor of Computer Science (Honours),* ***3.9/4.0 GPA*** Graduated Spring 2026

- Coursework: Statistical Modeling, Data Structures & Algorithms, Database Management Systems, Linear Algebra

EXPERIENCE

Electronic Arts (EA) | *Software Developer Intern* May 2025 - Aug 2025

- Improved CI/CD load-test reliability from **35% to 100%** by fixing auth failures and unstable workflows
- Evaluated LLM-powered agentic QA system for reliability, failure modes, and model behavior across load/soak workflows
- Documented failure taxonomy and evaluation findings to drive model-behavior improvements in automated playtesting pipeline

Autodesk | *Software Developer Intern* Jan 2025 - Apr 2025

- Built TypeScript/Java validation and repair infrastructure for Fusion manufacturing data, including an **O(n)** repair algorithm that improved throughput **8x**
- Designed deterministic validation logic for manufacturing data repair workflows
- Deployed GraphQL validation and repair endpoints with **100%** Mocha, Chai, and Sinon coverage

Public Health Agency of Canada | *Data Scientist Intern* Sep 2024 - Dec 2024

- Built LLM-assisted literature review and MDM workflows for global health data with Imperial College London and WHO
- Evaluated Ollama embeddings for RAG retrieval quality and baseline system-tuning metrics
- Prototyped GPT-4o paper collection for scalable PDF retrieval and metadata extraction

Bedarra Corporation | *Student Researcher* Aug 2024 - Oct 2024

- Researched and developed agentic RAG systems to answer queries over obscure programming language corpora (Factor, K)
- Benchmarked embedding models and chunking strategies for retrieval quality over low-resource language corpora
- Implemented LangChain tool-use and retrieval patterns for context-aware, grounded query responses

Health Canada | *Data Scientist Intern* Sep 2023 - Dec 2023

- Developed toxicity classifiers with pandas, NumPy, SciPy, and scikit-learn
- Built Neo4j knowledge-graph pipeline extracting chemicals, diseases, and literature relationships
- Architected TDD scraping framework with unittest, expanding data sources by **113%** with **100%** coverage

PROJECTS

Orbital Refueling Simulator | *Python, Streamlit, scikit-learn, Qwen LoRA, pytest* github.com/nbui23/orbital-refueling

- Modeled **410-second** refueling mission across **9** phases, **9** anomaly classes, and deterministic generation
- Separated hard-limit engineering rules from **9** phase-aware IsolationForest drift models
- Benchmarked deterministic rule alerts against **0-1** advisory ML drift scores across anomaly scenarios
- Fine-tuned Qwen LoRA explainer on synthetic telemetry instructions, improving held-out rubric score to **4.86/5**

TwinQuery | *Python, LangGraph, FastAPI, Streamlit, PostgreSQL/PostGIS, Ollama, RAG, Docker* github.com/nbui23/twinquery

- Built open-source digital-twin query assistant enabling non-technical users to ask plain-English questions over building-stock and retrofit datasets
- Implemented safe Text-to-SQL pipeline with deterministic guardrails, fallback query templates, and PostgreSQL/PostGIS execution over Ottawa building-footprint geometry
- Integrated local RAG over retrofit guidance with hybrid synthesis, grounding LLM answers in both database rows and retrieved document sources

SKILLS

Languages: Python, TypeScript, JavaScript, C++, Java, SQL
Frameworks & Runtimes: FastAPI, Flask, Streamlit, LangGraph, LangChain, Next.js, Node.js, Spring Boot
AI / ML / Data: LLMs, RAG, LangGraph, LangChain, Ollama, Qwen LoRA, scikit-learn, pandas, NumPy, PyTorch, FAISS
Databases & APIs: PostgreSQL/PostGIS, pgvector, MySQL, SQLite, MongoDB, Firebase, Neo4j, GraphQL
Tools & Practices: Git, Docker, Docker Compose, AWS, CI/CD, Ubuntu, TDD, pytest, unittest, Postman, Agile, SDLC
Spoken Languages: English, French, Vietnamese