

NORMAN BUI

+1 613-983-3217 | normanbui@cmail.carleton.ca | linkedin.com/in/normanbui23 | github.com/nbui23

EDUCATION

- Carleton University** | *Bachelor of Computer Science (Honours), 3.9/4.0 (A) CGPA* Expected April 2026
- Specialization in Software Engineering and Minor in Forensic Psychology
 - Awards: Deans' Honour List, President's Scholars, Chalmers Jack Mackenzie, and Henry Marshall Tory Scholarships
 - Relevant Coursework: Data Structures & Algorithms, Discrete Math, Web Apps, Operating Systems, Object-Oriented Software Engineering, Statistics, Programming Paradigms, Human-Computer Interaction (UI/UX), Databases, QA

EXPERIENCE

- Junior Data Scientist** | *Public Health Agency of Canada - Python, LLM, MDM* Sep 2024 - Present
- Researching, implementing, and optimizing LLMs for automating literature reviews in epidemiological research, as part of a Master Data Management system with the World Health Organization and Imperial College London
- Software Engineer Intern** | *Kongsberg Geospatial - C++, Qt, Python, GIS* May 2024 - Aug 2024
- Designed and built GIS apps with the TerraLens SDK for ATC applications as part of the Solutions Engineering team
 - Aggregated various VFR and terrain data sources using Python scripting, reducing processing time **by 40%**
 - Implemented advanced map rendering (LiDAR, WMTS, S-57) and object tracking in 2D and 3D, performance visualizations (FPS, CPU, GPU, RAM), and modernized UI, contributing to the acquisition of funding and clients
 - Developed a Python script to process ATC software's logs, reducing processing time by **over 99%**
- Data Scientist Intern** | *Health Canada - Python, ML, NLP, ETL, Web Scraping* Sep 2023 - Dec 2023
- Developed Python solutions for chemical assessment, leading to improved risk data aggregation and analysis
 - Built random forest classifiers with pandas, NumPy, SciPy, and scikit-learn to accurately predict chemical toxicity
 - Rebuilt organization-wide federated chemical search system using BeautifulSoup4 and lxml, implementing session management, automatic retries, and dynamic rate limiting, restoring data source coverage from **47% to 100%**
 - Developed standardized API templates and established test-driven development practices using unittest framework, resulting in **100% test coverage** and streamlined integration of new chemical data sources
 - Developed automated workflow for extracting semantic triples from scientific articles using CoreNLP and Stanza, streamlining the collection, organization, browsing of qualitative data into a Neo4J knowledge graph
- Junior Data Scientist** | *Department of National Defence - Python, LLM, RAG* Jun 2023 - Nov 2023
- Built a RAG app with Flask and GPT-3.5 to simulate conversation with a database of **20+ tax documents**
 - Implemented a NLP pipeline with tokenization, context analysis, and named entity recognition using NLTK
 - Leveraged SQLite3 and FAISS for efficient storage and embedding retrieval, enabling fast response times
- Software Engineer Intern** | *NAV CANADA - C++, Qt, PowerShell, GIS* May 2023 - Aug 2023
- Developed C++ solutions for various ATC applications, resulting in improved performance and reliability
 - Optimized data management using C++ for internal data tool, resulting in a **60% decrease** in storage utilization and **30% boost** in automated data analysis performance
 - Optimized code and resolved multiple bugs in ATC software, achieving an overall graphical **improvement of 16%**
 - Automated network configurations using PowerShell, reducing manual work and setup time by **more than 95%**

RESEARCH & PROJECTS

- Obscure Language RAG Systems** | *Python, RAG, LangChain, FAISS, Sentence Transformers* [GitHub](#)
- Researched and developed RAG systems to answer obscure programming language queries @ Bedarra Corporation
- SocialSaplings** | *HTML/SCSS, JavaScript, Bootstrap, Node.js, Express.js, Firebase (Firestore, Auth)* [Devpost](#)
- Led the development of a reforestation web platform, securing **1st place overall** at KuriusHacks: March Edition
 - Implemented a tool using various APIs to recommend optimal tree species based on a user's geodata
 - Developed a tree planting visualization tool with Google Maps API to showcase global reforestation impact
- Pomodojo** | *HTML/CSS, JavaScript, Bootstrap, Node.js, Express.js, MongoDB, WebRTC* [Devpost](#)
- Developed a study platform to facilitate productive Pomodoro sessions with gamification elements for motivation
 - Enabled real-time group video calling with WebRTC and Agora APIs, ensuring seamless, low-latency interaction

SKILLS

Languages: C/C++, Python, Java, C#, SQL, HTML/CSS, JavaScript, PHP | English, French, Vietnamese
Frameworks & Databases: Flask, Express.js, MySQL, SQLite3, PostgreSQL, Firebase, MongoDB, Neo4J, JUnit, Selenium
Libraries: pandas, NumPy, matplotlib, scikit-learn, unittests, spaCy, Sentence Transformers, BS4, lxml, HuggingFace
Development Tools: Git, Node.js, Ubuntu, Qt Creator, Postman, Agile Methodology, SDLC, OOP, TDD, Cucumber