

BENNY WU

✉ benny.wu.new@gmail.com
in benny-wu1 🌐 bennywu1210
🌐 bennywu.me

EDUCATION

University of Waterloo

Sept. 2022 - Apr. 2027

Bachelor's of Computer Science, Artificial Intelligence Specialization

DeAlmeida Award: One of two candidates with the **highest admission averages (98.5%)** of the BBA stream

Student Clubs: ML Researcher @ Wat.ai, Organizer @ 4 hackathons, UW Basketball Intramural Champion (S24)

TECHNICAL SKILLS

Languages: Python, Java, C, C++, Go, JavaScript, TypeScript, SQL

Frameworks/Libraries: PyTorch, TensorFlow, LangChain, React, Flask, Django, Scikit-learn

Tools: AWS, GCP, Docker, Kubernetes, Firebase, MongoDB, Jupyter, Redis, Kafka, Hugging Face

WORK EXPERIENCE

Twilio Inc.

May. 2025 – Aug. 2025

Software Engineer Intern

San Francisco, CA

- Building distributed microservices for Twilio's Voice stack, delivering high-availability Recording/Conference features for 150k+ live users using **Hazelcast** clusters and scaling real-time events with **Java** and **AWS**
- Reduce DB load by **70%** by rolling out a retriever + **RateLimiter/CircuitBreaker** to process queued tasks in outages
- Replacing a legacy audio caching layer with a **AWS CloudFront + Lambda@Edge + MediaConvert** flow
- Wrote custom **JMeter** scripts to replay peak voice flows in Jenkins, piping live metrics to **Datadog + BigQuery**

Royal Bank of Canada

Jan. 2025 – Apr. 2025

Software Engineer Intern

Toronto, Canada

- Shipped robust microservices using **Java Spring Boot** and **GraphQL** for RBC's flagship HomeX mortgage platform, releasing **5+ features** to streamline the loan renewal process for **90,000+** clients across Canada
- Rebuilt 2 legacy flows into fault-tolerant **Kafka + Camunda** pipelines, removing monthly downtime incidents
- Built shared error-handling modules for mortgage processing operations, slashing duplicated logic by **60%**
- Documented an AI resume classifier POC, leveraging **LangChain** and **Gemini** to enhance candidate screening

UW Artificial Intelligence

Sept. 2024 – Dec. 2024

ML Researcher

Waterloo, Canada

- Applied **LangChain** with **RAG** to analyze LLM agent's effectiveness in scenarios such as political reasoning
- Optimized retrieval efficiency through custom **ChromaDB** embedding pipeline and semantic chunking
- Integrated **Fast Mind, Slow Mind** (SOFAI) architecture to model agent decision-making using both rapid heuristics and deliberative reasoning, improving win rate by **50%** in simulated gamified environments

Ford Motor Company

Jan. 2024 – May 2024

Software Engineer Intern

Oakville, Canada

- Developed **5+ features** for a vehicle order placement system using **Spring Boot** and **Google Cloud Run**, expected to process **400,000+** annual orders and inquiries for commercial dealers across North America and Europe
- Wrote robust **JUnit** and **Cucumber** test suites, boosting coverage from **60%** to **92%** and reducing regression bugs
- Implemented a **Elasticsearch** semantic search layer to improve accuracy benchmark in *fordpro.com* by **20%**

RESEARCH & LEADERSHIP

Hack the North 📍 | Hackathon Organizer

Mar. 2025 – Present

- Organizing the **largest collegiate hackathon in Canada**, hosting 1500+ participants from 15+ countries annually
- Negotiated and secured **\$XXX,XXX** monetary sponsorships to execute project and event coordination

Geesehacks 📍 | Software Team Lead

May 2024 – Dec. 2024

- Led an **Agile** team of 7 to develop the software architecture for a hackathon backed by **Google** and **SunLife**
- Developed a product roadmap and built 3 full-stack applications with **Next.js**, **Vercel** and **PostgreSQL**

NSERC Canada 📍 | Undergraduate Research Assistant

May 2023 – Aug. 2023

- Collaborated with Dr. Navid Esfahani on a thorough analysis of a (28, 10, 5)-balanced incomplete block design
- Achieved a **97.4%** reduction in runtime for classifying linear codes through **compiler-level C optimizations**