Lin (Bill) Qi

Research Engineer, AI & Machine Learning

Montreal, QC, Canada (Willing to relocate to the UK)

• +1 514-746-5515 • bill9555@hotmail.com

Right to Work: Youth Mobility Scheme visa (valid until March 2027)

PROFILE

Research Engineer with 8+ years' experience building complex AI systems. Deep expertise in designing advanced retrieval-augmented generation (RAG), multi-agent, and deep learning models. Proven ability to lead projects from concept to deployment, backed by a PhD and publications in top-tier journals.

EXPERIENCE

CGI | Montreal, QC

Generative AI Data Scientist | June 2024 – Present

(Client Project: Public Services and Procurement Canada (PSPC))

- ➔ Architected and led the development of an enterprise-scale conversational AI assistant for PSPC compensation advisers, achieving **91% benchmarked response accuracy** and significantly improving access to complex policy information using **Retrieval Augmented Generation (RAG)**; implemented using a combination of **LoRA fine-tuned LLaMa 3.1 8B and 70B** models served using **SGLang**.
- ➔ Architected a sophisticated RAG system featuring **fine-tuned embedding** and **re-ranking** models, which outperformed traditional retrieval methods on complex policy documents.
- ➔ Developed an LLM-powered pipeline using LangChain to automate the **transformation of semi-structured policy documents into structured JSON**, saving the team days of manual data processing.
- ➔ Developed a novel multi-agent OCR pipeline (using **OpenAI Agents** and Prompt Flow, deployed via **Azure AI Foundry**) combining **Azure Document Intelligence with OpenAI GPT-4o**, achieving **>99% accuracy** in extracting key data from scanned **PDF documents** (forms, letters) and outperforming standalone methods.

(Innovation Team Contributions)

- ➔ Led the design of enterprise-grade, zero-trust generative AI architectures adhering to stringent privacy and security standards.
- ➔ Automated cloud resource provisioning by implementing **Terraform (Infrastructure-as-Code)** pipelines.
- ➔ Accelerated AI solution deployment by creating reusable code templates and robust **CI/CD pipelines** for automated testing and delivery.
- ➔ Investigated and improved the speed and efficiency of RAG implementations.

McGill University, Department of Human Genetics | Montreal, QC

PhD Candidate & Researcher | September 2018 – February 2024

- ➔ PhD Thesis: **"Advancing the understanding and treatment of psychiatric and other complex diseases through machine learning and omics"**
  - ◆ https://escholarship.mcgill.ca/concern/theses/j3860d546?locale=en
- ➔ Built scalable **Python and Tensorflow data pipelines** using high-performance computing clusters to process, clean, and analyse terabyte-scale, semi-structured genomic datasets from ~500,000 individuals in the UK Biobank
- ➔ Published **8 peer-reviewed papers** (4 as lead author) in top journals, including *Biological Psychiatry*.

- ➔ Introduced a novel approach for machine learning model selection for high-dimensional genomics datasets to reduce risk of overfitting to small sample size.
- ➔ Engineered a novel **Graph Convolutional Neural Network** to integrate genomic and clinical data, building a predictive model for disease risk and medication optimisation (details described in thesis).

Ericsson Canada | Montreal, QC
Software Developer | July 2017 – September 2018

- ➔ Architected an ML-based engineer assignment system processing **10,000+ monthly tickets with 70% accuracy** (text classification task).
- ➔ Created a question-answering AI using NLP and information retrieval techniques on millions of documents, increasing engineer efficiency.
- ➔ Led the full-stack development of an application serving **1,000+ daily users** using AngularJS and Python.
- ➔ Reduced application load time by **60%** through frontend optimisation and implementing microservices with Docker and Kubernetes.

TECHNICAL SKILLS

- ➔ **Data Processing:** Pandas, Numpy, Jupyter
- ➔ **Generative AI & LLMs:** LangChain, Prompt Flow, OpenAI Agents, RAG, Vector Databases, SGLang, vLLM
- ➔ **AI & Machine Learning:** TensorFlow, PyTorch, Scikit-Learn, MLflow
- ➔ **Cloud & MLOps:**
  - ◆ **Azure** (AI Foundry, AI Search, Document Intelligence, Storage Accounts, Key Vault, Databricks),
  - ◆ **AWS** (EC2, Lambda, S3, EventBridge),
  - ◆ Terraform, Docker, Kubernetes, CI/CD, MLflow
- ➔ **Languages:** Python, JavaScript
- ➔ **Databases:** Azure (Cosmos DB, MongoDB, Gremlin Graph DB), PostgreSQL, SQLite, Elasticsearch, Pinecone
- ➔ **Software Development:** VSCode, Git, FastAPI, Flask, REST APIs, Linux/Unix, Websockets, Streamlit

---

EDUCATION

- ➔ **PhD in Human Genetics (Statistical & Machine Learning focus)** | *McGill University, Montreal, QC*
- ➔ **AI in Healthcare Nanodegree** | *Udacity*
- ➔ **Bachelor of Science, Microbiology & Immunology** | *McGill University, Montreal, QC*

AWARDS & CERTIFICATIONS

- Silver medal ranking (48/2893 teams) in **Ubiquant Market Prediction** competition: https://www.kaggle.com/certification/competitions/billqi/ubiquant-market-prediction
  - Designed and implemented a deep neural network solution for stock market prediction.
- Silver medal ranking (60/1687 teams) in **RSNA Screening Mammography Breast Cancer Detection** competition: https://www.kaggle.com/certification/competitions/billqi/rsna-breast-cancer-detection
  - Leveraged pretrained computer vision models, data augmentation, and Nvidia TensorRT for efficient breast cancer detection using GPUs.