

# Benjamin Mah

+1 647-901-5309 | [benjaminmah.bm@gmail.com](mailto:benjaminmah.bm@gmail.com) | [benjaminmah.com](http://benjaminmah.com) | [linkedin.com/in/benjaminmahh](https://linkedin.com/in/benjaminmahh) | [github.com/benjaminmah](https://github.com/benjaminmah)

## EDUCATION

### University of Toronto

May 2026

*Bachelor of Applied Science in Engineering Science*

*Toronto, Canada*

Major in Machine Learning Engineering, Minor in Engineering Business

Relevant Courses: Data Structures & Algorithms, Neural Networks, Deep Learning, Operating Systems, Probabilistic Reasoning

## EXPERIENCE

### Asana

May 2025 – Present

*Data Science Intern*

*Vancouver, Canada*

- Leading data-driven analysis to identify key patterns in customer churn using clustering and predictive churn modeling, revealing a potential recovery of **\$2M+ ARR** through proactive and personalized retention techniques
- Initiated and executed opportunity sizing for automatic seat upgrades, analyzing upgrade-to-churn patterns and assisted improving customer communication frameworks to mitigate post-upgrade customer churn by **15%**
- Delivered ad-hoc analytical requests from product management teams using **Databricks** and **Redash**, building interactive dashboards and providing rapid data insights to support product decisions and A/B experimentation

### Mozilla

May 2024 – April 2025

*Machine Learning Engineer Intern*

*Toronto, Canada*

- Built and deployed ML models using **XGBoost** and **scikit-learn** for automated bug triaging through a bot deployed in **Bugzilla** to automate triage workflows and reduce manual processing by **50+ hours** monthly
- Co-authored a research paper on a GPT-based code review bot, collaborating with **Ubisoft**, **Queen's University**, and **Polytechnique Montréal**, used by over **200 engineers** and leading to a **30%** reduction in review time
- Fine-tuned pre-trained large language models using **Hugging Face Transformers**, **PyTorch**, and **PEFT** for generating code that addresses inline comments, using a dataset of **15,000+** annotated patch fixes and review comments, achieving an **8%** higher acceptance rate compared to baseline models like **GPT-4** and **Claude 3**

### Cohere

September 2023 – April 2024

*Data Specialist*

*Toronto, Canada*

- Audited and annotated machine learning datasets using **Scale AI**, identifying and correcting labeling errors across thousands of samples, enhancing model performance and saving over **100 hours** of processing time each month
- Conducted in-depth comparative analysis of LLM-generated code completions written in **Python**, **JavaScript**, and **SQL**, refining validation processes to align with coding standards, resulting in faster deployment cycles

### Royal Bank of Canada

May 2023 – August 2023

*Machine Learning Engineer Intern*

*Toronto, Canada*

- Developed a machine learning tool for financial advisors, using semantic search with **LangChain**, **Qdrant**, **Dask**, and **NumPy** to efficiently process user queries from financial databases, boosting response speed in information retrieval (IR) and extractive question answering (EQA) by **40%** through optimized query embeddings
- Designed and implemented a scalable multi-type file processing pipeline with efficient retrieval endpoints, using **Python** and **Amazon S3** to handle various complex textual data formats, reducing proof-of-concept (PoC) development time by **10 hours** per project and standardizing reusable code for future generative AI initiatives

## PROJECTS

### Audio Style Transfer | *PyTorch, Librosa, NoiseReduce*

- Developed a neural style transfer model to transform non-standard English accents into a standardized accent, validated using Google's ASR model to calculate Word Error Rate (WER) and improve transcription accuracy
- Implemented a robust data processing pipeline using **Librosa** and **NoiseReduce** for audio decoding, noise reduction, and spectrogram generation, supporting the training of the NST model built with convolutional layers

## SKILLS

**Languages:** Python, SQL, C/C++, JavaScript, HTML/CSS, MATLAB

**Frameworks/Tools:** PyTorch, TensorFlow, Databricks, Redash, LangChain, AWS, Flask, Git, VSCode, Jupyter

**Libraries/APIs:** pandas, NumPy, scikit-learn, XGBoost, Transformers, SciPy, Matplotlib, ChromaDB, Qdrant

**Interests:** Record Collecting, Studio Ghibli, Videography, Music Production, NBA, Formula 1