

# Garreth Edderick Lee

778-840-2450

[garreth.edderick@gmail.com](mailto:garreth.edderick@gmail.com) | [linkedin.com/in/garrethlee](https://linkedin.com/in/garrethlee) | [github.com/garrethlee](https://github.com/garrethlee)

## EDUCATION

---

### University of British Columbia

*Bachelor of Science in Computer Science and Statistics*

Clubs: UBC Biztech, TEDxUBC, Data Science Club, Google Student Developer Club

Vancouver

*Sep 2021 – Present*

## EXPERIENCE

---

### Machine Learning Engineer Intern

*Hugging Face*

Vancouver

*Sept 2024 – Dec 2024*

- Worked on improving text extraction for future [Fineweb](#) datasets
- Conducted experiments to determine optimal tokenization strategies for mathematical reasoning in LLMs and wrote a [detailed report](#) on the Hugging Face platform
- Created and maintained internal tools to better visualize and inspect LLM pretraining data

### Data Science Intern

*Wealthsimple*

Toronto

*May 2024 – Aug 2024*

- Developed an ML-based in-house solution to route engineering on-call tickets, improving accuracy from 40% to 80% and saving thousands of manual labor hours annually
- Optimized and increased coverage on the PII (Personal Identifiable Information) scrubber used for the company's internal LLM gateway
- Developed a pipeline to automatically update an in-house vector database with dynamically changing metadata, reducing manual data management for engineering managers and ensuring real-time accuracy

### Member of Technical Staff

*Cohere*

Toronto

*Sept 2023 – Dec 2023*

- Built and optimized Python and Spark data pipelines processing terabytes of pretraining data for large language models
- Assembled an LLM-based pipeline for document quality evaluation, saving thousands of dollars in human annotator costs
- Integrated language identification models to existing HTML parsing pipelines, enhancing downstream textual data quality by 20%

## PROJECTS

---

### Email Generation Model Fine-tuning | *PEFT, Python, Ollama, Transformers*

May 2024

- Fine-tuned Mistral 7B to streamline the creation of initial email drafts for UBC Biztech's partnerships team
- Managed end-end preparation of the dataset for effective model training and ensuring high-quality input for the fine-tuning process.
- Deployed the fine-tuned model on a local machine using Ollama, optimizing the inference process and maintaining control over sensitive information.

## TECHNICAL SKILLS

---

**Languages:** Python, SQL, R, HTML/CSS, Java, Javascript, Bash

**Frameworks:** Django, React, Flask, JUnit, Streamlit

**Tools:** Git, Docker, Google Cloud Platform, Amazon Web Services, Apache Airflow, Terraform, Spark, dbt

**Libraries:** Pandas, Tensorflow, NumPy, Pytorch, Scikit-Learn, Ggplot, Seaborn