

Khoi Nguyen

nguyentienkhoi01@gmail.com · (438) 938-0432 · Montreal, Canada

linkedin.com/in/khoi-tien-nguyen · github.com/khoitiennguyen · khoitiennguyen.github.io

Work Experience

Intact Insurance, Data Lab

Montreal, Canada

DATA SCIENTIST INTERN

May 2023 – Aug 2023

- Performed exploratory data analysis on client flood claims, historical floods events and satellite imagery, using **Python (Pandas, Matplotlib, NumPy)** and geospatial libraries in **R**.
- Utilized and improved weather claims classification Large Language Model (LLM) to effectively identify flood claims, revealing key patterns in claim distributions.
- Formulated and investigated hypotheses on flood model's efficacy in high-risk areas, contributing to improved understanding of flood risk patterns affecting 700,000+ customers.
- Consolidated disparate datasets to perform extensive data cleaning and merging using **SQL** and **R**, ensuring data consistency and reliability for model benchmarking.

McGill University, Distributed Digital Music Archives & Libraries Lab

Montreal, Canada

SOFTWARE DEVELOPER

May 2022 – Apr 2023

- Developed a workflow for user-based image layer labeling and optical recognition model training.
- Implemented image processing functionalities using **Python (NumPy, TensorFlow, OpenCV, Django)** to accelerate training data preparation efficiency, saving annotators 80+ hours per quarter.
- Designed a dynamic data loading scheduler for user-labeled datasets, increasing model training speed by 200%.
- Collaborated with researchers for iterative modeling changes, leading to significant experimental advancements.

Accreon

Fredericton, Canada

SOFTWARE DEVELOPER CO-OP

Sep 2018 – Jan 2019

- Refined **Java**-based **ETL** processes for improved data handling, introducing new features for enhanced flexibility and efficiency in data processing.
- Streamlined client data extraction into bilingual formats and implemented compression techniques, significantly improving data storage and accessibility for developer use in a web application.

Education

McGill University

Montreal, Canada

B.Sc. COMPUTER SCIENCE, GPA: 3.53/4.0

Sep 2019 – May 2023

- Awarded the J.W. McConnell Major Entrance Scholarship for outstanding academic achievement.
- Courses: Data Science, Applied Machine Learning, Deep Learning, Natural Language Processing, Artificial Intelligence.
- Extracurriculars: McHacks Hackathon (2020), McWiCS Hackathon (2022), McGill Orientation Leader (2020 – 2022).

Skills

Programming	Python · SQL · R · Java · C · C++ · JavaScript · HTML · CSS · Bash · Unix · OCaml · Assembly
Libraries	NumPy · Pandas · TensorFlow · PyTorch · Scikit-learn · SciPy · HuggingFace · XGBoost · Flask
Tools	Git · Jupyter · Docker · AWS · PostgreSQL · Redis · Spark · Snowflake · Kafka · Grafana · Jenkins

Projects

ML Reproducibility Challenge 2022: BERT Data Size Probing [Paper](#) [GitHub](#)

- Reproduced and expanded on experiments from an ACL 2022 research paper: arxiv.org/pdf/2203.09627v1.pdf
- Fine-tuned and evaluated **BERT** models on SuperGLUE tasks across data size intervals, leveraging **HuggingFace** and **PyTorch** to analyze semantic and syntactic knowledge and assess linguistic recoverability.

Automatic Music Transcription [GitHub](#)

- Engineered an automatic music transcription system using a Convolutional Recurrent Neural Network (**CRNN**) in **TensorFlow**, achieving over 90% accuracy with sheet music format conversion on four of seven instruments.
- Adapted the **CRNN** model from an [OCR paper](#) to accommodate time series audio data, preventing pooling on the time axis, integrating additional convolutional layers, and customizing tensor dimensions to support variable batch sizes.