Khoi Nguyen

438-938-0432 | nguyentienkhoi01@gmail.com | Montreal, Canada
linkedin.com/in/khoi-tien-nguyen | github.com/KhoiTienNguyen | khoitiennguyen.github.io

Experience

Machine Learning Engineer June 2024 – Present
Groupe Dynamite *Montreal, Canada*

- Engineered and productionized an end-to-end demand forecasting system using Python (PySpark, PyTorch) and AWS (Glue, SageMaker), improving 4-week forecast accuracy from 52% to 74%.
- Led iterative improvements to ML models through rigorous experimentation, feature engineering, HPO, and evaluating diverse architectures (e.g., ARIMA, XGBoost, PatchTST)
- Developed interactive dashboards visualizing complex data analyses (Python, SQL), enabling stakeholders to extract actionable insights on customer behavior and sales drivers.
- Built and deployed a scalable, production-grade REST API for user recommendations (FastAPI, Docker, AWS ECS) with load-balancing, automated CI/CD (GitHub Actions), monitoring (CloudWatch), and IaC (Terraform).
- Prototyped and implemented novel solutions, including developing an LLM-based multimodal product tagger (AWS Bedrock) and finetuning Chronos transformer models.

Data Scientist Intern May 2023 – Aug 2023
Intact Insurance *Montreal, Canada*

- Evaluated a geospatial risk model by comparing predictions against historical flood events and internal claims data (Python, R, SQL), revealing findings that informed adjustments to internal risk assessment.
- Engineered critical features using geospatial analysis and leveraged LLMs for text classification, increasing relevant claim identification by 85% for benchmarking.
- Conducted exploratory data analysis across diverse datasets (claims text, satellite imagery) to define evaluation metrics and uncover key insights into model performance.
- Aggregated and prepared complex insurance datasets (claims, exposures, geospatial) for model evaluation, ensuring data consistency and readiness for analysis.

Software Developer May 2022 – Apr 2023
McGill University, DDMAL Lab *Montreal, Canada*

- Developed an end-to-end machine learning workflow (Python, Docker) enabling users to upload, partially annotate, and train custom models for automated image annotation.
- Implemented backend image processing services (Python, NumPy, OpenCV) to automate data preparation tasks like background removal and edge detection, reducing 80+ hours quarterly in manual annotation time.
- Engineered a data loading scheduler for ML training, optimizing RAM utilization through LRU eviction and caching strategies, resulting in a 3x (200%) increase in model training speed.
- Collaborated closely with PhD researchers to prototype and validate novel ML modeling techniques using TensorFlow, ensuring rigorous testing on staging servers prior to deployment into the production environment.

Software Developer Intern Sep 2018 – Jan 2019
Accreon *Fredericton, Canada*

- Developed ETL workflows using Java to extract user data, transform it into bilingual JSON formats, and package it into compressed archives for developer testing tools.

Education

McGill University Montreal, Canada
Bachelor of Science in Computer Science *Sep 2019 – May 2023*

- Awarded the J.W. McConnell Major Scholarship for outstanding academic achievement.

Technical Skills

Languages: Python, Java, SQL (Postgres), R, C, JavaScript, HTML/CSS, Bash
Libraries: PyTorch, TensorFlow, PySpark, XGBoost, Sklearn, Pandas, Numpy, Plotly, HuggingFace, Streamlit
Frameworks: FastAPI, Flask, PyTest, SQLAlchemy, Django
Tools: Git, Jupyter, Docker, Terraform, Github Actions, WandB, Mypy, Black, Ruff, UV, Poetry
AWS Services: SageMaker, Glue, Lambda, Step Functions, S3, ECR, Bedrock, EC2, RDS, ECS, CloudWatch