Khoi Nguyen

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

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

Work Experience

Groupe Dynamite   Montreal, Canada
MACHINE LEARNING ENGINEER   *June 2024 – Present*
- Adapted Chronos, a T5-based time series model, for supply chain demand forecasting across North American stores, fine-tuning it with proprietary data, achieving a 22% WAPE score using **HuggingFace** and **PyTorch**.
- Engineered end-to-end ML pipelines for automated model training, validation, tuning, and deployment with robust governance and versioning, utilizing **AWS SageMaker**, **Step Functions**, **Docker**, and **AutoGluon**.
- Architected and optimized AWS Glue and Lambda data pipelines using **PySpark**, implementing data partitioning and caching strategies, resulting in a 35% reduction in cloud computing costs and improved processing speeds.

Intact Insurance, Data Lab   Montreal, Canada
DATA SCIENTIST INTERN   *May 2023 – Aug 2023*
- Performed comprehensive exploratory data analysis on client flood claims, geospatial maps, and satellite imagery using **Python** (Pandas, Matplotlib, NumPy) and **R**, uncovering critical insights for flood risk assessment.
- Increased flood claim identification rate by 85%, leveraging weather claims classification **LLM** and feature engineering.
- Discovered key patterns in claims distribution affecting 700,000+ customers through rigorous hypothesis testing.
- Reduced data inconsistencies by 60% through extensive data preprocessing and wrangling using **SQL** and **Python,** ensuring data reliability and validity 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 labeling and optical recognition model training using **Python** and **Docker**.
- Implemented image processing functionalities with **NumPy** and **OpenCV** to accelerate training data preparation efficiency, saving annotators 80+ hours per quarter.
- Designed a data loading scheduler for user-labeled datasets, optimizing model training speed by 200%.
- Collaborated with researchers for modeling experiments using **TensorFlow**, leading to significant advancements.

Accreon   Fredericton, Canada
SOFTWARE DEVELOPER CO-OP   *Sep 2018 – Jan 2019*

Projects

Aircraft Trajectory Generation using Conditional Generative Adversarial Networks (cGAN)   Paper
- Developed a **cGAN** model to generate realistic aircraft trajectories using **PyTorch**, leveraging OpenSky database for dynamic behavior simulation, greatly enhancing turn and pitch rate accuracy through advanced data preprocessing.
- Achieved statistically significant improvements in trajectory accuracy and variability, demonstrating the cGAN's effectiveness over baseline models in airspace variable realism, enhancing airspace safety and efficiency simulations.

ML Reproducibility Challenge 2022: Large Language Model 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.

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.

Skills

Coding	Python · Java · SQL · R · C · JavaScript · HTML · CSS · Bash · Unix · Linux
Libraries	NumPy · Pandas · TensorFlow · PyTorch · HuggingFace · Transformers · Scikit-learn · SciPy · OpenCV
Tools	Git · Jupyter · Docker · AWS Suite · PostgreSQL · PySpark · CI/CD · GitHub Actions · Kafka · WandB · CUDA