

Terence Chow

Data Scientist/ Machine Learning Engineer/ Data Engineer



AI Developer, [Intact Data Lab](#), Toronto, ON Sep 2022 - Present

DBT/ETL/SAS/Snowflake/SQL/PowerBI/Jupyter/MLOps/MLflow/AWS/HuggingFace/Optuna/Databricks

- Design a scalable and extensible prototype of a DBT-based ETL pipeline for marketing PoC with BA and DEV teams, transferring data from the SAS data server to the insurance quoting system, and storing data in the Snowflake database.
- Build and review snowflake SQL views, integrating with Power BI dashboard to present model results to business partners, improve dashboard visualization continuously according to business requirements, and develop version control procedures.
- Primarily responsible for maintaining ML models in production, monitoring AWS-based preprocessing pipeline, detecting functional anatomy of multiple NLP models (NER & BERT), and reporting and resolving technical issues with the DEV team.

AI Developer Intern, [Intact Data Lab](#), Toronto, ON May 2022 - Sep 2022

JIRA/Confluence/MIRO/GitLab/CICD/Docker/K8S/Nexus/Cookiecutter/Airflow/AWS/LLM/Generative AI/GPT3

- Researched and tested generative LLM models to compare time and space complexity for multiple NLP use cases, prompting engineering to guide GPT3 models to create synthetic emails and perform QA tasks.
- Designed and developed in-house CLI commands for tracking usage among 1000+ users, enabling uploading data from the in-house platform to AWS S3 buckets for data retention.
- Designed and implemented extensible and scalable Airflow pipelines with Senior MLE, automated data extraction processes among various data warehouses, reducing 17% of working time for DS teams.

Software Engineer, [Arima](#), Toronto, ON Dec 2021 - May 2022

Trello/Keras/Matplotlib/Flask/NodeJS/GCP/Big Query/AWS/IAM/EC2/S3/CloudWatch/AlteryX/Pandas-gbq

- Responsible for marketing research to optimize the performance of generating data modeling, refactoring, and modularizing modeling scripts, and reducing model executing time by 30%.
- Extended Canadian company business to the United States, pre-processed 1M+ open source US census data, including noise removal, feature extraction and ranking, and classification by zip code.
- Implemented an automated AlteryX pipeline to process 100+Gb of marketing data, and classify data by geographical location and administrative division, such as state, county, city, and zip code.
- Developed a flask-based backend of the company webpage by creating API functions, integrating with AWS CloudWatch to track 100+ user's behavior, and prepared data logs to NodeJS-based frontend for building usage dashboards.
- Programmed automated bash scripts to generate synthetic data on GCP compute engine and upload data to Google Big Query.

Graduate Research Assistant, [Lakehead University](#), Thunder Bay, ON Sep 2019 - Aug 2021

TensorFlow/PyTorch/HuggingFace/NumPy/Pandas/OpenCV/Unsupervised/MATLAB/JAVA/SQL/C++

- Researched image segmentation, implemented ML framework from image preprocessing to model development, training, and evaluation, and increased 3.9% of mIoU on the VOC-2012 dataset. Paper submitted to Neural Computing, under peer-reviewed.
- Optimized unsupervised ML algorithm for feature learning using ELM by enabling multi-threading, reducing 100% of processing time. Built a semi-supervised CNN framework to resolve classification problems, and tested performances on 30+ image datasets.
- Collaborated with professors to prepare 8-course materials and live coding demonstrations, reviewed assignments and final exams, and provided grades on UG (Android APP Development, 100+ students) and PG courses (Machine Learning & Database System, 50+ students).

EDUCATION

- MSc. in Computer Science (Research-based, AI Specialization), [Lakehead University](#), Thunder Bay, ON Sep 2019 - Aug 2021
- BSc. in Electricity Engineering (Major in Information Engineering), [City University of Hong Kong](#), Hong Kong Sep 2015 - July 2019