Daniel Tsiang

London, UK

✉ dan_tsiang@hotmail.co.uk | 🌐 https://danieltsiang.github.io

Experience

Starling Bank

London

Tech Lead – Machine Learning Software Engineering

Aug 2024 – Present

- Managing a team of 8 software engineers to delivery efficiencies across the business using cutting-edge Machine Learning and AI, and promoting best practices with MLOps & LLMOps.

Senior Software Engineer in Machine Learning

Jul 2023 – Jul 2024

- Deployed LLMs into production to optimise workflows extracting information from images and summarising transcripts.
- Integrated Google AI APIs directly into the Java banking platform to boost customer operations efficiencies and detect fraud.
- Engineered the backend, unit & integration tests for Python apps and packages serving Machine Learning (ML) models.
- Deployed ML models into production by running microservices in Docker containers and orchestrating them with Kubernetes.
- Implemented Prometheus metrics and created Grafana dashboards to monitor app and ML model performances, allowing detection of model drift and adhering to MLOps best practices.

Software Engineer

Aug 2021 – Jun 2023

- Created Java-based simulator to put ML services under constant load, resolving performance bottlenecks and validating accuracy of responses.
- Created serverless event-driven AWS Lambda to sync files between S3 and Cloud Storage using keyless cross-cloud auth.
- Automated data ingestion & processing pipelines by writing Bash CI/CD pipelines to create templates for GCP Dataflow jobs.
- Established automated CI workflows to scan code for security vulnerabilities, with GitHub and Slack integrations for alerts.

Projects

Deploying LLMs to enhance operational efficiency

Starling Bank - 2024

Automated Call Summarisation and Insight Generation

- Integrated Starling Bank's Java banking platform with AWS Transcribe API and Google Cloud's Gemini Pro LLM to automatically generate summaries and extract key insights from customer call transcripts. E.g. complains, advice given, required follow-up actions, potential customer vulnerability.
- This solution eliminated the need for manual summarisation by customer agents, reducing the write-up time from over 10 minutes to under 2 minutes on average, which led to an improvement in customer service response times.
- The vulnerability flagging feature led to an increase in the identification of vulnerable customers.

Automated Data Extraction from Documents:

- Developed an LLM-based solution to extract information from customer ID documents and cheques, automatically populating relevant fields in production databases.
- This automation replaced a previously manual, error-prone process, saving hundreds of hours of manual data entry per week and reducing data entry errors.

Optimised LLM Performance through Experimentation:

- Designed and implemented Kubeflow pipelines on GCP to orchestrate and manage LLM prompt engineering experiments.
- By tracking metrics (e.g. accuracy, precision, recall, inference speed), I iteratively improved model performance leading to more accurate data extraction and insightful summaries. I then deployed the optimised models to production.

Technologies: Python, Java, AWS Transcribe API, Google Cloud Platform (GCP), Gemini LLM, Kubeflow, PostgreSQL

Stock Simulator (demo)

Personal - Apr 2021

- Engineered a real-time stock portfolio simulation web app, leveraging live stock price data via API integration.
- RESTful web app built using Python with Flask's MVC framework, connected to a SQL database and served by Gunicorn.
- Designed UI to update data displays and validate data in real-time by making AJAX calls from JavaScript.

Skills

Programming	Python, Java, JavaScript, SQL, C, Bash
Python Libraries	Flask, TensorFlow, Scikit-learn, NumPy, Pandas, PyTest
Java Frameworks	Guice, Mockito, AssertJ
Cloud & DevOps	GCP, AWS, Docker, Terraform, Grafana, Prometheus, Kubernetes, Git
Data Engineering	Kubeflow, Dataflow (Apache Beam)
Databases	BigQuery, PostgreSQL, SQLite

Education

Harvard University

Remote

HarvardX CS50ai & CS50x: Computer Science with Artificial Intelligence using Python

Jan 2021 – Dec 2021

- Completed intensive courses covering fundamental computer science concepts and advanced AI techniques.
- Created AI programs to play Minesweeper, Tic-Tac-Toe and Nim games optimally.
- Developed AI programs to generate crossword puzzles, rank webpages by importance and solve logic puzzles.

Imperial College London

London

Master of Engineering in Chemical with Nuclear Engineering (First-Class)

Oct 2014 – Jun 2018

- Developed strong foundation in numerical computing and algorithm development for solving complex engineering problems.

Certifications

2022 **Google Cloud Big Data and Machine Learning**, Coursera (certificate)