Leo Leung

github.com/zcemycl | +447543781301 | lyc010197@gmail.com | linkedin.com/in/leo-leung-48524b134

Work Experience

Machine Learning Engineer   *Mar 2022 - Now*
Faculty AI   London, UK

— Enhanced the MobileNet YOLO object detection architecture in Tensorflow, achieving optimized model size and faster inference times.
— Boosted book sales through the implementation of a recommendation system based on book embeddings, leveraging Google Cloud Platform, ANNOY Indexing, and BigQuery.
— Created a comprehensive application that demonstrates the question-answering capabilities of the OpenAI Large Language Model. The development utilized AWS, Python FastAPI, Terraform, Chroma Vector Database, Guardrails-AI, and Postgresql.
— Led the rollout of feature updates for data visualization improvements in the web application, incorporating modifications through React Typescript, Python FastAPI, OpenAPI Tool, and Alembic for database migration. Automated end-to-end testing of the React web application, secured by the Amazon Cognito Authentication System, streamlining testing processes with Typescript, Cypress, and Gmail API.

Machine Learning Engineer   *Jun 2020 – Nov 2021*
RE5Q   London, UK

— Constructed an ETL pipeline utilizing Python Selenium to extract data from broker websites. Transformed the extracted data using various PyTorch NLP and computer vision models to generate meaningful signals for the real estate market.
— Deployed web applications with Flask and React to illustrate the improved functionalities of retrained AI models in advancing real estate predictions. This includes Named Entity Recognition (NER) using spaCy, summarization utilizing AllenNLP, and YOLOv5 integration with PyTorch.
— Managed hybrid infrastructures to sustain company operations, overseeing components like DNS servers, VPN servers, virtualization servers, and AWS services.

AI Consultant   *Jan 2020 – Jun 2020*
Dreams-AI   Cambridge, UK

— Applied 2D Homography and Camera Calibration to synchronise robotic arm with camera to pick up objects.
— Devised an Optical Flow algorithm to monitor money wheel rotating motion.
— Retrained an open source YOLOv3 model for custom labels.
— Managed AWS (EC2 and S3) pipelines to handle big data for training and deploying face recognition system.

Deep Learning Software Engineering Intern   *Sep 2019 – Dec 2019*
The Mathworks   Cambridge, UK

— Developed a prototype for the imageLIME function, inspired by the paper "Why should I trust you?", to enhance AI explainability within the MATLAB Deep Learning Toolbox.
— Accumulated more than 2000 downloads on Matlab File Exchange, received 4 citations on Google Scholar, and garnered 80 forks and 152 stars on a Github project focused on reproducing diverse GAN models—including InfoGAN, Pix2pix, and CycleGAN—utilizing the MATLAB Deep Learning Toolbox.

Projects

systemDeploy

— Built cloud infrastructures for a 3 tier system with Terraform and AWS, such as Fargate, Elastic Container Registry, Cloudmap, Route53, ACM, Application Load Balancer, Cloudwatch, and more.
— Integrated FastAPI RestAPI with Grafana and Prometheus to report service performance.
— Built local docker apps with Kubernetes and Helm.

TF2DeepFloorplan

— Led a github project to replicate 'Deep Floor Plan Recognition using a Multi-task Network with Room-boundary-Guided Attention' with Tensorflow2 to identify room types and boundaries in the floorplan.
— Streamlined a MLOps process to train the models with Tensorboard, optimise network weight via TFLite and deploy with Docker and Flask.
— Developed a raycasting pygame to transform 2d floorplan to 3d simulation.
— Achieved 52 forks and 143 stars in Github, and 2 citations in Google Scholar.

Education

University of Cambridge   *Sep 2019*
MPhil Machine Learning and Machine Intelligence (2nd upper)   Cambridge, UK

University College London   *Sep 2018*
BSc Physics with Medical Physics (1st)   London, UK

Skills

Programming: Python, Javascript, Terraform, C++, React, Typescript, Tensorflow, PyTorch, Spark
Databases: PostgreSQL, ChromaDB, DynamoDB, Neo4j, Opensearch
Cloud Tools: AWS, Azure, GCP