LE QUAN

benjamin.lequan@gmail.com | +65 98503192
linkedin.com/in/benjamin-le-quan | github.com/benjaminlq

WORK EXPERIENCE

SINGAPORE GENERAL HOSPITAL, Singapore – Data Scientist **Apr 2023 – Current**

Software & Tools: Pytorch, Langchain, LlamaIndex, HuggingFace, Scikit-learn, Unstructured

- Lead the technical dimensions for SGH and Singhealth's efforts in research and implementation of Large Language Models (LLMs) for clinical assistance and value-driven care.
- Setup evaluation pipelines for prompt engineering, hyperparameter tuning and error analysis for systematic development and optimization of LLM applications.
- Built a customized ingestion pipeline capable of handling several data modalities (texts, tables, flow charts and images) for downstream multimodal retrieval-augmented generation (RAG) on medical guidelines.
- Developed a multi-stage LLM copilot for recommending thyroid cancer treatments, utilizing function-calling tools and multimodal retrieval-augmented generation (RAG) agentic framework, achieving performance comparable to senior residents at over 90% accuracy.
- Implemented a multi-modal RAG audit system to identify improper MRI scan orders made by physicians with **88%** accuracy, reducing unnecessary scans expense by **15%** and saving radiologists' time spent on manual audits.
- Fine-tuned Llama-2 with distributed mixed precision LORA training to extract medical entities and Protected Health Information (PHI) from clinical free-texts to achieve over **98%** recall.
- Research on synthetic data generation for structured and unstructured clinical data to address the challenges of data sensitiveness and improve the representation of low-data regimes on certain medical domains.
- Optimized an XGBoost model to predict risks of post-operative events based on biomarker features extracted from patients' body CT scans and other clinical features.
- Conducted workshops at several AI Healthcare Conferences (IAIM, AI Health Summit) to raise awareness on safe and effective usage of Large Language Models in clinical research and operations.

DATATURE ANALYTICS, Singapore – Machine Learning Engineer Intern **Jan 2023 – Mar 2023**

Software & Tools: OpenCV, Pytorch, Tensorflow, ONNX

- Research, create prototypes and deploy into production cutting-edge Computer Vision algorithms such as object detection (YOLO and Faster-RCNN) and video tracking (DeepSORT, ByteTrack).
- Deployed a **Flask API** for polygon interpolation between frames to reduce users' effort video annotation by **80%**.
- Wrote technical articles on applying the latest CV technology to solve real-world industrial problems.

HEALTH PROMOTION BOARD, Singapore – Data Science and Analytics Intern **Apr 2022 – Oct 2022**

Software & Tools: Python, R, Scikit-learn, Spark

- Segmented 200K NSC users using unsupervised models on Fitbit/Apple Watch physical activity metrics (exercise volume, heart rates, walking patterns) for targeted customers nudging and program recommendation.
- Built regression models with Difference-In-Difference and Interrupted Time Series analysis to evaluate LumiHealth program impact on users' long-term exercise performance based on 500 days of exercise data.

GLOBALFOUNDRIES INC, Singapore – Senior Process Engineer **Jul 2017 - Jul 2021**

- Designed controlled experiments, collected data to qualify new processes, performed correlation analysis and hypothesis testing on inline data with JMP tools to assess impact and effectiveness of new process changes.

TECHNICAL SKILLS

- **Languages/Software/Platform:** Python, R, SQL, Google Cloud, Linux/Ubuntu, Git/Github, Docker
- **Libraries:** Scikit-learn, TensorFlow, PyTorch, PySpark, FastAPI, Langchain, LlamaIndex, HuggingFace, OpenCV
- **Technical areas:** Statistics, Machine Learning, Deep Learning, Data Science, Data Analytics, Natural Language Processing, Computer Vision, Large Language Models, Generative AI, Retrieval Augmented Generation

PERSONAL PROJECTS

Classification of pill images with multi-modal image (CLIP) and text (OpenAI) embeddings

Software & Tools: PyTorch, OpenAI, HuggingFace, LlamaIndex

- Implemented a hybrid retrieval system for identification of medical pill images using a combination of visual features (CLIP embeddings) and text feature embeddings (image captions) extracted using OpenAI GPT-Vision.

Handwriting image translation with PyTorch and FastAPI

Software & Tools: Pytorch, Docker, FastAPI

- Optimized a **98%** accuracy ResNet-RNN model trained on CTCLoss to translate handwriting images to text data.
- Containerized and deployed the trained model with **FastAPI** web framework for prediction serving.

Image Generation with Variational Autoencoder (VAE) and Generative Adversarial Network (GAN)

Software & Tools: Pytorch, Docker, Streamlit

- Optimized several VAE and GAN architectures to generate images from MNIST and CIFAR-10 datasets.
- Deploy trained generative models on web application with Docker and Streamlit.

End-to-end ML pipeline for telecom customers churn prediction on Google Cloud Platform

Software & Tools: Google Cloud Platform, Kubeflow, Docker, Scikit-learn

- Constructed an end-to-end Machine Learning pipeline on **Google Cloud Platform** and deployed trained classification models on Vertex AI to predict potential churn telecom customers.
- Containerized individual components and orchestrated training pipeline on Kubernetes clusters.

Prediction of survivability for patients with coronary artery disease

Software & Tools: Scikit-learn, Pandas

- Optimized and evaluated multiple **Scikit-Learn** classical Machine Learning models (Logistic Regression, Decision Tree, SVM, KNN) to predict survivability of patients with coronary artery disease with **>95%** accuracy.
- Improved accuracy by **15%** with polynomial predictors and recursive feature elimination (RFECV) algorithm.

EDUCATION

SINGAPORE MANAGEMENT UNIVERSITY **Aug 2021 - Dec 2022**

Master of IT in Business (Artificial Intelligence Track)

- GPA: **4.0 / 4.0**
- Awards: ASEAN Graduate Scholarship, SAS Institute Top MITB Student Award, Dean's List

NATIONAL UNIVERSITY OF SINGAPORE **Aug 2012 - Jun 2016**

Bachelor of Engineering in Chemical Engineering

- Awards: ASEAN Undergraduate Scholarship

OTHERS

- **Languages**: English (Fluent), Vietnamese (Native)
- **Work Authorization**: Requires working Visa (Singapore)