

Wen Qing LIM's Resume

UK Graduate Visa holder until January 2027—full-time work permitted.

Email: lim.wenq@gmail.com | Personal site: <https://wengqinglim.github.io> | LinkedIn: <https://www.linkedin.com/in/wen-qing-l-67330771/> | GitHub: <https://github.com/wengqinglim> | Substack: <https://codingintune.substack.com> | Medium: <https://medium.com/@limwenqing>

Skills

Machine Learning & AI: Deep Learning & LLMs (PyTorch, HuggingFace, SentenceTransformers, LangChain) • Audio AI & Music Computing • Model Development & Implementation

Programming & Tools: Python, PySpark • AWS Sagemaker, GCP • MLOps: Docker, MLFlow, Airflow

Languages: English (Fluent), Chinese (Conversational)

Work History

Python Developer/Data Scientist @ SilentEight, Singapore/UK | Apr 2023 - Present

Tech stack: Python, PySpark, AWS Sagemaker, SentenceTransformers, PyTorch, LangChain

- Led Research and Development in ML applications:
 - Fine-tuned large language models (LLMs) with generated synthetic data to improve entity matching models' recall by 44%, resulting in greater efficiency.
 - Developed explainable solutions for rule generation
 - Optimised entity matching models using clustering techniques
 - Reduced PySpark processing runtime by 70% through optimisation
- Lead client projects
 - Build client-specific ETL pipeline to transform and extract data from raw text data
 - Create scoring system for identifying and evaluating common name tokens
 - Create Chinese character visual similarity model

Data Scientist @ Twitter, Singapore | May 2022 - Jan 2023

Tech stack: SQL, Python, Vertica, Google Cloud Big Query, Looker

- Sales Tools and Analytics: Build SQL queries for internal tools used by Sales teams to manage their portfolio and serve clients more effectively, saving potentially 50,000 hours of manual effort annually.
 - Collaborate with cross-functional teams to design and build technical solutions with the Engineering team
 - Write queries for Ads Performance metrics, standardise and consolidate metrics from multiple sources.
- Process Optimisation: Build logic for internal alerts tool that helped advertisers gain access to managed support 2 weeks sooner.
 - Designed success metrics for tool adoption, run regular analysis on the impact of the tool, and identify areas contributing to the delayed support.

Machine Learning Engineer @ foodpanda, Singapore | Jan 2021 - May 2022

Tech stack: Python, SQL, Airflow, Google Cloud Platform, Docker, MLFlow

- Built an entity matching algorithm to identify duplicate restaurants and products across multiple text data sources in 9 countries and 4 languages.
 - Served various functions: competitive intelligence, price governance, strategy, content optimisation.
 - Designed and implemented human-in-the-loop feedback to improve F1 score.

- Built a classifier to identify cuisines based on text data, reducing human labelling and review time by 1 week for every 1,000 cuisine tags. ([Article](#))
 - Experimented with NLP techniques (W2V, TF-IDF), rule-based approach, supervised learning, zero-shot learning, and distanced-based embedding.
 - Design online test experiments to measure the impact of the refined cuisine tags.
- Advisor for junior MLE/analysts working on ML projects including clicks estimation, zombie vendor prediction, menu item ranking, and fraud detection.

Data Scientist @ UCARE.AI, Singapore | Dec 2018 - Jan 2021

Tech stack: Python, PySpark, SQL, Airflow, Docker, MLFlow

- Worked on internal data science tools using Python and PySpark to streamline the data science workflow (data exploration/pre-processing, model iteration), reducing 2-3 man-days per data scientist for each project .
- Conducted end-to-end data science proof-of-concept projects:
 - Improved MAPE of hospital bill prediction by 33%, by improving model features and experimenting with boosted-tree models (XGB, LGBM).

Education

2023–2024 | **Queen Mary University of London**, School of Electronic Engineering and Computer Science

Master of Science in Sound and Music Computing — AI and Music Data Science Stream (Distinction)

2015–2019 | **National University of Singapore**, School of Computing

Bachelor of Science in Business Analytics (Second Upper Honours) • NUS Overseas Colleges Silicon Valley Program (2017–2018)

2018–2019 | **Stanford University**, SCPD • Management Science and Engineering Courses

Projects

Audio AI & Music Generation Research

MSc Thesis: Hierarchical Symbolic Music Generation | 2024

- Developed novel graph neural network architecture for long-form music generation ([ArXiv paper](#))
- Implemented 2-stage CNN/GCN Variational Auto-Encoder (VAE) system to capture both local and global musical structures
- Trained model on POP909 dataset to generate coherent musical compositions with structured sections
- Paper accepted and to be presented at upcoming 17th International Symposium on Computer Music Multidisciplinary Research (Nov 2025)

Deep Learning for Music Module Project: Multi-model pitch transcription | 2024

- Developed end-to-end pipeline for vocal source separation and pitch transcription ([GitHub](#), [Article](#))
- Implemented and fine-tuned ConvTasNet model for two-vocal source separation
- Built CNN-based singing pitch transcription system (CREPE architecture)
- Evaluated system performance on MIR-1K dataset

Music Information Retrieval Module Projects: Beat Tracking and Audio Fingerprinting | 2024

- Implement Ellis' (2007) Beat tracking algorithm and evaluate performance on a Ballroom dataset ([GitHub](#))
- Implement Phillips' audio identification algorithm and evaluate performance on GTZAN dataset

Computational Creativity Module Projects: | 2024

- Built multi-modal pipeline integrating text, music, and video generation, to explore how generative models can be used for creative expression ([GitHub](#) , [Article](#))
- Implemented mood/genre classification, with conditional music generation system responding to text and audio prompts

AI & Machine Learning Applications

AI Summit London Hackathon | 2024

- 1 of 3 winning teams ([LinkedIn](#))
- Prompt engineering using GPT4o to perform document Q&A to help insurance agents recommend policies to clients

Omdena [Data Science for Social Good - Volunteer] | 2020-2021

- Lead ML Engineer: Led a team to analyze and detect online child sexual abuse from chatroom datasets and research papers in 8 weeks. ([Article](#))
 - Data mining (BeautifulSoup, Selenium) and text analysis (TF-IDF, Bag of Words, Clustering, Market Basket Analysis)
- Singapore Chapter Lead: Co-led a 10-week project on uncovering the impact of Covid-19 on mental health in Singapore ([Article](#)).
 - Mentor and guide Junior collaborators at all project stages

Thanks for reading! 