Michael Hoon Yong Hau

📞 +65 97769494 | ✉ michael_hoon@mymail.sutd.edu.sg | 🔗 linkedin.com/in/michaelhoon | ⦿ github.com/michael-hoon

EDUCATION

Singapore University of Technology and Design — May 2026 – Dec. 2026
Master of Engineering (Research), Artificial Intelligence — *(Expected)*

Singapore University of Technology and Design — Sep. 2022 – May 2026
Bachelor of Engineering, Engineering Systems and Design, Minor in Computer Science — *Current GPA: 4.83/5.0*

University of California, Berkeley — Jun. 2023 – Aug. 2023
Summer Exchange, Econometric Analysis — *Berkeley, CA*

EXPERIENCE

Natural Language Processing Engineer Intern — Mar. 2024 – Dec. 2024
DSO National Laboratories — *Singapore*

- Built production-ready Retrieval Augmented Generation (RAG) pipeline for ingesting of multiple internal unstructured documents, with local LLMs (Llama3) via Ollama and LangChain.
- Implemented Advanced RAG techniques for complex documents, experimented with multi-modal methods for image and table parsing capabilities. Evaluated models via LangSmith.
- Deployed and containerised models via Docker for internal document use

Data Analyst — Jan. 2024 – Apr. 2024
Poh Tiong Choon Logistics Ltd. — *Singapore*

- Conducted Exploratory Data Analysis with Python on internal data for delivery driver utilisation. Identified key areas for improvement in scheduling and boosted driver utilisation rates by 20%.
- Developed end-to-end automated pipeline solution for data cleaning and analysis with bash scripting, integrating existing database with Python and Power BI, significantly reducing working man-hours.
- Created PowerBI Dashboard for key metrics and patterns on route analysis. Wrote internal report discussing findings

Data Analyst — Sep. 2022 – Jan. 2023
Aviation Studies Institute — *Singapore*

- Extracted key business insights for industry partner, International Air Transport Association (IATA), via Exploratory Data Analysis in R, SQL, and Python. Conducted qualitative analysis of air cargo demand in Southeast Asia.
- Developed correlation models for predictive analysis of internal air cargo demand data, with open-sourced economic data. Evaluated hypothesis on key drivers of regional air cargo demand.
- Wrote an internal paper discussing findings, presented to IATA and the Civil Aviation Authority of Singapore.

PROJECTS

BrainHack 2024 | *Python, PyTorch, Docker* — May. 2024 – May. 2024

- Utilised transfer learning techniques for fine tuning of pre-trained Automatic Speech Recognition (ASR) and LLMs on a large corpus of audio and transcript data. Developed a model capable of understanding natural language commands
- Implemented Vision Transformers (ViT) for automated bounding-box generation of targets in image data based on natural language transcript provided by the LLM.
- Containerised and deployed models via Docker and FastAPI for evaluation of model accuracy on leaderboard ranking.

FinSight.AI | *Python, LangChain, Streamlit* — Mar. 2024 – Mar. 2024

- Industry project with Temasek Holdings (Financial Services Department): Large Language Model (LLM) enabled Financial Analysis Assistant via RAG, scalable with existing in-house data infrastructure.
- Summarizes multiple unstructured financial reports and key metrics with internal data, reducing man-hours by 80%. Able to retrieve metadata of reports and allow user to self-verify accuracy, and provide semantic score results.

Coursework / Certifications

- Machine Learning
- Optimization
- Natural Language Processing
- Applied Econometrics
- Data Science
- Data & Business Analytics

Skills

Languages: Python, R, SQL, Julia, LaTeX

Tools: Git/GitHub, Docker, Ollama, VS Code, Vim, DBeaver, Linux, WSL, Excel, Power BI

Libraries/Frameworks: LangChain, Streamlit, HuggingFace, PyTorch, Pandas, NumPy, Seaborn, scitkit-learn, Plotly, tidyverse (R)