NGAN JI HIN

HP: 9012 5929 | Email: nganjihin@gmail.com | Website: https://njihin.github.io/

PERSONAL STATEMENT

A motivated candidate with hands-on exposure to LLM projects from a recent internship. Seeking a role to deepen practical AI engineering skills through real projects beyond coursework and self-study. Collaborative, reliable and quick to learn, with a track record of high-quality delivery.

SKILLS & ABILITIES

- **LangChain, ChromaDB, HuggingFace, BitsandBytes** – Created RAG through implementation of these components.
- **Python & Java** - created medium-scale software programs in order of thousands of lines of code and tens of classes using object-oriented design principles in Java.
- **PySpark, Apache Airflow & AWS Glue** – Developed multiple Machine Learning Pipelines for predictive analytics in Databricks. Developed basic Airflow DAGs for automation. Deployed AWS Glue Crawlers for ETL Pipelines.
- **RStudio & MySQL** - Understand basic techniques of data management, warehousing with relational database management systems, as well as data visualisation principles, techniques, and tools.
- **Excel** - Data analysis and reporting using pivot tables, lookups, conditional logic and charts; built sheets automate routine data entry and checks used in day-to-day operations during National Service.
- **Data Communication** - Adobe Illustrator and Photoshop for infographics, posters and report visuals; working knowledge of Premiere Pro and After Effects.
- **Administrative, planning and communication skills** - Handled all financial and administrative tasks as Team Manager in NUS Varsity Bouldering.

EDUCATION

National University of Singapore   **Apr 2023 - Apr 2027**
Undergraduate Bachelor of Science in Business Analytics (Hons)
- Overall GPA: 4.41

INTERNSHIP & PROJECTS

Local Retrieval Augmented Generation   **Aug 2025 - Present**
- Built and optimised a locally hosted RAG assistant for MATLAB-syntax Q&A on a single RTX 3060 Ti 8 GB setup.
- Implemented an end-to-end RAG pipeline with LangChain using quantised Hugging Face models via bitsandbytes; created embeddings and persisted vectors in ChromaDB
- Ingested PDFs with PyMuPDF and applied recursive chunking and metadata for traceable answers and source citations
- Tuned retrieval parameters and quantisation settings to keep inference responsive within 8 GB VRAM, fully offline for data privacy
- Documented configuration, constraints and trade-offs to enable reproducibility on commodity hardware
- Website and other projects: https://njihin.github.io/

Tech stack: LangChain, HuggingFace, Bitsandbytes, ChromaDB, PyMuPDF, Python

Data Engineering Intern, ABeam Analytics, Singapore   **Jun 2025 - Jul 2025**

- Leveraged Microsoft Fabrics to set up access controls and basic data governance across multiple Lakehouses and Notebooks.
- Built a bronze to silver medallion ETL pipeline across multiple Lakehouses, applying job and notebook orchestration in Microsft Fabrics.
- Contributed as a junior data architect, proposing Fabric components and standards for an ongoing project
- Documented workspace structure, permissions and data lineage to support handover

Tech Stack: Microsoft Fabric, Lakehouse, Notebooks, PySpark, Spark SQL

Card Fraud Analysis & Detection - Team of 5   **Jan 2025 - Apr 2025**

- Developed an end-to-end fraud detection pipeline in Databricks using PySpark, from feature engineering and model evaluation with clustering analysis
- Compared multiple classification models and tuned thresholds; evaluated with precision recall and ROC based metrics
- Explored unsupervised clustering to surface behavioural patterns in confirmed fraud cases
- Addressed severe class imbalance through resampling and class weight strategies while operating within Databricks Community Edition limits
- Delivered a written report and walkthrough for stakeholders
- Final report: https://tinyurl.com/4swb79vd

Tech stack: Databricks, PySpark, Python, SQL, MLlib, Git

HOBBIES

- Bouldering - Team Manager of NUS Varsity Climbing Team AY24/25