

OWEN TAN KENG LENG

owentan2021@gmail.com | (+65) 9247 2807 | www.linkedin.com/in/owentankengleng | <https://github.com/OweT1>

EDUCATION

National University of Singapore

Aug 2022 - May 2026

Bachelor of Science in Data Science and Analytics (Honours) with 2nd Major in Computer Science

- Grade Point Average: 4.72/5.00 (Expected Honours with Highest Distinction)
- Attained Dean's List for AY2023/2024 Semester 1 & Semester 2 and Top Student for Data Structures and Algorithms
- Relevant Coursework: Natural Language Processing, Machine Learning, Big Data Systems

WORK EXPERIENCE

GIC, Quantitative Strategist Intern

Jan 2026 - Present

- Architected and optimised a high-fidelity data extraction pipeline converting PDFs to structured Markdown with extraction accuracy of more than 90% and reduced overall pipeline execution runtime by 50%
- Engineered 2 features for Linear Regression, Random Forest and XGBoost models to predict rent price of real estate assets in Korea

Temasek Holdings, Machine Learning / Data Engineer Intern

Aug 2025 - Dec 2025

- Designed and deployed 2 production LLM tools: a slide-generation tool and an investment tool that retrieves structured financial metrics (such as EV, EBITDA, P/E) from Snowflake via dynamic SQL querying
- Reduced metadata extraction latency by more than 50% (20s to less than 10s) through prompt decomposition and asynchronous processing
- Benchmarked semantic retrieval solutions (OpenAI File Search, Snowflake Cortex, Azure AI Search) across latency, document retrieval recall (Recall@5), and cost for production use case
- Built a Neo4j-based knowledge graph POC to model entity relationships and enable multi-hop document retrieval

DBS Bank, Data Science Intern

May 2025 - Aug 2025

- Designed LLM-based evaluation pipelines to audit call transcripts against regulatory guidelines, achieving more than 90% agreement with human reviewers
- Engineered 3 prompt workflows for Claude and Gemini, iterating on evaluation rubrics and negative sampling analysis
- Conducted multimodal Speech-to-Text experiments with Gemini, improving transcript quality for downstream NLP tasks

SATS Ltd, Data Scientist Intern

Aug 2024 - Dec 2024

- Fine-tuned a DeBERTa-based text classification model to categorise text into 10 classes, achieving 85% accuracy
- Deployed an end-to-end workflow and implemented Genetic Algorithm in Python to automate optimisation of baggage allocation under 10+ hard and soft constraints, reducing planning time by more than 50%
- Built Microsoft Azure based document digitalisation pipelines (Databricks, Blob Storage, Logic Apps), reducing paper usage to 0% and operational costs by 20%

PROJECT EXPERIENCE

Job Finder – Github Actions, OpenRouter, Postgres, Alembic

Dec 2025 - Present

- Built a fully automated job aggregation pipeline scraping and processing 100+ tech roles daily, with LLM-based classification to route job postings to appropriate Telegram channels
- Designed deduplication logic using indexed Postgres tables to prevent cross-workflow duplication
- Productionalised system with retries, logging, and CI checks, achieving more than 95% successful hourly runs

Resume Evaluator – LangGraph, Pydantic, Gradio

May 2025 - Jul 2025

- Designed a LangGraph-based multi-agent evaluation system to assess resumes against job descriptions and company context
- Evaluates and generates 5 actionable, rubric-based feedback across skills, experience alignment, and gaps

ESG Analysis Platform for Banking Sector – Express, JWT, Docker, Databases

Jan 2025 - Apr 2025

- Developed reliable backend service for ESG Platform using Express for API routes and JWT for Authentication
- Designed and deployed 5 MongoDB and PostgreSQL databases using Docker containers for data management

TECHNICAL SKILLS & CERTIFICATIONS

- **Cloud Platforms:** Google Cloud Platform, Microsoft Azure, Databricks, Cloudera, Snowflake
- **Programming Languages:** Python, SQL, Cypher, Go, Java, JavaScript, R
- **LLM Libraries:** DeepEval, LangGraph, LangChain, Pydantic, Transformers, HuggingFace, OpenRouter, LiteLLM, Neo4j, Graphiti
- **Other Libraries:** MLFlow, re, SpaCy, NLTK, Scikit-learn, Tensorflow, Keras, PyTorch, Pymoo, SQLAlchemy, Alembic, NumPy, Pandas, GeoPandas, PySpark
- **Containerisation & Orchestration:** Docker, Kubernetes
- **Certifications:** Google Cloud Certified Professional Machine Learning Engineer, Data Engineering with Databricks