WILL LIM WEN LONG

Senior Data Scientist | IMDA SG Digital Scholar

wenlonglim96@hotmail.com +65 9843 7531 askwilli.dev linkedin.com/in/limwenlong github.com/wenlong96

PROFESSIONAL SUMMARY

Senior Data Scientist with 3+ years building production ML systems at PwC, specialising in LLMs, RAG, real-time data pipelines, and recommendation systems. Proven track record delivering US$1.2M in client revenue through Agentic Graph RAG solutions. Built WiLLi — a live RLHF fine-tuning system using Kafka streaming, DPO optimisation, and LoRA adapters — demonstrating end-to-end ML engineering from training to production deployment. AWS ML Specialty certified. Currently pursuing NUS Master of Computing (AI), cGPA 4.75/5.0.

WORK EXPERIENCE

PricewaterhouseCoopers
Senior Data Scientist | Singapore
Jan 2025 – Present

- Led architecture and delivery of Agentic Graph RAG solutions (LangChain, LlamaIndex, PyTesseract), directly generating US$1.2M in new client revenue.
- Engineered hyperparameter-tuned chunking and embedding strategies for vector DB retrieval, improving splitting accuracy by 20% over baseline and directly improving LLM answer quality in production.
- Custom-trained lightweight U-Net models for document image segmentation, improving extraction accuracy by 10% and reducing downstream OCR errors in client-facing pipelines.
- Drove end-to-end ML lifecycle across financial services and public sector clients: scoping, training, evaluation, deployment, and C-suite communication.

PricewaterhouseCoopers
Data Scientist | Singapore
Jun 2023 – Jan 2025

- Developed multi-variate time-series forecasting models (ARIMAX, XGBoost, LightGBM) for inventory demand planning, reducing MAPE by 20% across 15+ SKU categories.
- Re-architected RAG inference pipelines with asyncio and multiprocessing, achieving 80% runtime reduction — enabling real-time response latency for client-facing applications.
- Built automated web scraping and data mining pipelines (Selenium, SQL, factor analysis) for market intelligence and competitive whitespace identification across 3 countries.
- Collaborated with product managers, engineers, and business stakeholders to translate analytical findings into product decisions.

Diners Club International
Data Analyst | Singapore
Jan 2023 – Jun 2023

- Deployed automated data crawling, ingestion, and reporting pipelines, saving 30+ minutes of manual work daily and improving data freshness for leadership dashboards.
- Built and maintained complex PostgreSQL/MySQL queries and Cron Job workflows supporting real-time reporting for business stakeholders.

FEATURED PROJECT

WiLLi — Live RLHF-Powered AI Clone
Personal Project | askwilli.dev | github.com/wenlong96/streaming-rl-llm |
2026

- Built a production RLHF pipeline from scratch: fine-tuned Qwen 2.5 3B Instruct with LoRA adapters and DPO (Direct Preference Optimisation) — the same alignment technique used at OpenAI and Anthropic.
- Engineered real-time feedback streaming architecture with Redpanda (Kafka-compatible): admin corrections are published as DPO training pairs, consumed live by a training loop, and model weights update without server restart.
- RAG retrieval layer using ChromaDB, LangChain, and sentence-transformers; served via FastAPI with token-streaming responses and Streamlit frontend deployed at askwilli.dev via Cloudflare Tunnel.
- Demonstrates full-stack ML engineering: data ingestion, streaming infrastructure, online fine-tuning, LoRA adapter management, and deployment — fully built without managed ML services.

SKILLS & TECHNOLOGIES

Languages: Python, SQL, R
ML / Deep Learning: Transformers, LoRA/PEFT, DPO/RLHF, RAG, GNN, CNN, RNN/LSTM, Time Series, XGBoost, LightGBM, Bayesian Inference
LLM / AI Frameworks: PyTorch, TRL, HuggingFace Transformers, LangChain, LlamaIndex, ChromaDB, sentence-transformers
Data & Streaming: Kafka / Redpanda, asyncio, Selenium, PostgreSQL, MySQL, Pandas, NumPy
Deployment & Infra: FastAPI, Streamlit, Docker, AWS EC2, Cloudflare Tunnel, Git, Azure DevOps

Visualisation: PowerBI, Tableau, Plotly

EDUCATION

National University of Singapore *Jan 2025 – Jun 2027*
Master of Computing — Artificial Intelligence Specialisation | cGPA: 4.75 / 5.0 | Singapore

Singapore Management University *Aug 2019 – May 2023*
B.Sc. Economics, Data Science & Analytics | Magna Cum Laude | GPA: 3.71 / 4.0 | Singapore

Singapore Polytechnic *Apr 2013 – Apr 2016*
Diploma in Banking and Finance | GPA: 3.52 / 4.0 | Singapore

AWARDS & CERTIFICATIONS

• IMDA Singapore Digital Scholarship (Postgraduate) — Artificial Intelligence	*2025*
• AWS Certified Machine Learning — Specialty	*2024*
• PwC Spot Award — Top Performer in Division	*2024*
• SMU School of Economics Dean's List	*2024*