Nguyen Quoc Anh — AI Engineer II

TheQuantScientist | andrewfinance | anhnguyen.aiml@gmail.com | +84 39-998-5047 | Portfolio

EDUCATION

Royal Melbourne Institute of Technology — Ho Chi Minh City, Vietnam
Bachelor of Business — Economics and Finance — *2021 — 2025*

INDUSTRY EXPERIENCE

Middle AI/ML Engineer, *(Hitachi Digital Services)* — *10/2024 — present*

– **Project Kasado: An Agentic Platform for Factory Workers** — *Deployment*

- Led 3 Engineers in developing automation workflows on Dify for standardized PDF document generation from unstructured input.
- Built Python-based plugins, enabled agentic communication via REST API and Workflows, successfully delivered apps to end user.
- Injected bilingual instruction prompts into HTML template for autonomous content mapping, engineered efficient system prompts.
- Developed a RAG pipeline for defects retrieval via Milvus vector database, set up PostgreSQL for query, wrote OpenAPI Schema.
- Performed weekly demos following Agile framework, mentored newcomers, handled 3 Detail Designs, managed WBS and bug list.

– **Project AI Tools: Small Language Models R&D for Q&A System** — *R&D*

- Research Japanese-focused models, benchmarked SLMs ($< 12B$) and LLMs on Japanese QnA datasets using cosine similarity.
- Conducted EDA reports on JaQuAD dataset and models' predicted output, designed auto-compiling pipeline for R&D documents.
- Maintained Docker containers for 4+ SLMs interaction for QnA services on UI, applied llm-as-a-judge evaluation via RAGAS lib.

– **Project Virtual Planner: Multi-agent System for High-speed Rail Management** — *POC*

- Developed an agentic system for train management through information retrieval with violation detection and penalty optimization.
- Built a Text-to-SQL agent using LangChain/LangGraph frameworks, managed ETL processes on PostgreSQL, designed FastAPI.
- Utilized RAG approach with FAISS index for relevance check and retrieval tasks, leveraging Llama3.3, Gwen2.5, and GPT-4o-mini.
- Engineered CoT-based system prompts, defined rules for data ingestion, enabling train reschedule with availability warning on UI.

– **Project Nestle HoloLens V2: Applied Computer Vision in Logistics Chain Installation** — *Deployment*

- Trained and fine-tuned YOLOv8-11n models on HoloLens for real-time object detection with Maya-3D logistics guidance for workers.
- Wrapped trained models in ONNX format, deployed Sentis models onto HoloLens for ITW testing across augmentation parameters.
- Engineered balance multilabel dev-test sets on CVAT for 11+ manufacturing steps, achieved Accuracy of 0.8+ at minimal overfitting.

– **Internal Responsibility**

- Served as a technical advisor for internal Text-to-SQL projects and core member of HFMT's GenAI team, attending VDC's meetings.
- Represented the company as an industry guest speaker, chosen as the AI talkshow champion by VP for internal podcast series.
- Deployed local LLMs for research, installed CUDA and cuDNN, managed Docker containers, monitored Ubuntu/Window servers.
- Drew high-level/detail agentic designs, prepared Powerpoint slides for client' meetings, updated daily MeMo, broke tasks in WBS.

Founder and Technical Lead, *(The Neurone Group)* — *10/2024 — present*

– **Small Language Models for Domain Text-to-SQL Tasks**

- Benchmarked 10 SLMs ($< 7B$) from Ollama versus gpt-oss-120B on 4 domain datasets (120 questions) for Text-to-SQL capability.
- Applied context engineering for domain knowledge integration, built SQL/Semantic validation metrics, studied GPU/CPU overhead.
- Built codebase skeleton, instructed 2 developers, set up PostgreSQL for E2E query run, designed behavioral and ablation studies.

Research Associate, *(Tech Mahindra)* — *09/2024 — 10/2024*

- Conducted user study (n=1000) on smart devices from Apple, led participants through 50+ mobile protocols in dynamic environments.
- Engaged in CRM training and daily F2F customer services, recorded daily outcome onto the database, in charged of tech-heavy duty.
- Run selected studies under NDA agreement, trained newcomer user research staff, recruited local participants, exceeding KPI by 50%.
- Established secure research environment following vendor specifications, ensured bulletproof data privacy, welcomed foreign vendors.

Collaborative AI/ML Researcher, *(RMIT Vietnam)* — *03/2023 — 09/2024*

- Developed SOTA deep Neural Network models on S&P 500 stocks for quant exchange, achieving 75% real-time swing-trading accuracy.
- Optimized models' metrics up to 85% by implementing Phase Space Reconstruction and Attention Mechanisms into their architectures.
- Managed ETL processes and 2M irregular observations from PhysioNet and YFinance using Python, increasing data credibility by 55%.
- Built cryptographic algorithms and Federated Environments for IoT cybersecurity experiments, peer-reviewed 43 PhD research papers.

RESEARCH PROJECTS

CLAM: A Hybrid Deep Learning Model for Weekly Stock Trend Forecasting *Project Link & Paper*

- Developed a synthetic model with stacked layers of Conv1D, LSTM, and Attention Mechanisms for multi-step stock trend forecasting.
- Fine-tuned CLAM through 48 hyperparameters at 90:10 split, optimized training with EarlyStopping and ReduceLROnPlateau callbacks.
- Improved MAE and RMSE by 90%, capturing 75% of out-sample stock trends with flash crashes thus outperforming LSTM and CNN.

CryptMAGE: An Intraday Crypto Price Movements Recognition System *Project Link & Paper*

- Converted hourly multivariate cryptocurrency time series into 12K arbitrary interval line-graph images for UP-DOWN binary classification.
- Engineered MA and RSI indicators into the OHLCV crypto data, reducing feature correlation by 53% to enable real-time categorization.
- Fine-tuned pre-trained Swin Transformer, achieved a 92% AUROC/AUPRC and 93.65% F1-Score, surpassing ViT, DeiT, and ResNet50.

PSR-NODE: A Neural ODE Model for Stock Price Forecasting *Project Link & Paper*

- Programmed a neural differential-equation-based model incorporating Taken's Embedding Theorem for financial time series prediction.
- Reconstructed data phase space into 3D-dimension with a delay of 1 and TimeSeriesSplit with 5-fold cross-validation for efficient training.
- Boosted training speed by 63% and reduced errors metric by 70%, surpassing SVR, LSTM, and TFT, hence mitigating investments' risk.

PUBLICATIONS

[1] **Anh Nguyen*** and Son Ha. "Transforming Stock Price Forecasting: Deep Learning Architectures and Strategic Feature Engineering". In: *MDAI 2024 Proceedings* (2024). **CORE B.** https://doi.org/10.1007/978-3-031-68208-720.

[2] **Anh Nguyen***, Son Ha, and Phien Nguyen. "LiteFormer: An Encoder-Only Multi-Head Attention Transformer for Financial Time Series Forecasting". In: *Engineering Applications of Artificial Intelligence* (2025). **Q1.** https://doi.org/10.210403.

[3] **Anh Nguyen***, Son Ha, and Hieu Thai. "Phase Space Reconstructed Neural Ordinary Differential Equations Model for Stock Price Forecasting". In: *PACIS 2024 Proceedings* (2024). **CORE A.** https://aisel.aisnet.org/pacis2024/ai-in-business.

[4] **Anh Nguyen***, Ngan Nguyen, and Hy Truong. "Small Language Models For Efficient Domain Data Retrieval Using Structured Query Language from Natural Language". In: *SSRN* (2025). https://dx.doi.org/10.2139/ssrn.5909763.

[5] **Anh Nguyen*** and Hy Truong. "CLAM: A Synergistic Deep Learning Model for Multi-Step Stock Trend Forecasting". In: *Intelligenza Artificiale* (2025). **Q2.** https://doi.org/10.1177/17248035251322877.

[6] Son Ha, **Anh Nguyen**, and Phat Tran. "SLIE: A Secure and Lightweight Cryptosystem for Data Sharing in IoT Healthcare Services". In: *ICSOC 2025 Proceedings* (2025). **CORE A.** https://arxiv.org/abs/2510.14708.

HONORS & ACCOLADES

Conference Presenter | *International Conference on Service-Oriented Computing* — Dec. 2025

Distinguished Alumni of RMIT Vietnam | *RMIT Alumni Impact Showcase x Australian Government* — July. 2025

Guest Speaker of Bloomberg | *Vietnam Investment Summit*, *Bloomberg Businessweek Vietnam* — December. 2024

Conference Presenter | *Vietnam Economist Annual Meeting* — December. 2024

Top 3% Global SSRN Researcher | *Social Science Research Network* — November. 2024

Breakthrough Lead Researcher in Deep Learning | *RMIT News*, *BaoMoi*, *StockBiz* — Aug. 2024

Conference Presenter | *Pacific Asia Conference on Information Systems* — July. 2024

Peer Reviewer of Q1 Journals | *ICPR*, *Elsevier*, *IEEE*, *Springer*, *Nature* — Jan. 2024

TECHNICAL SKILLS

Data Science: *TensorFlow, Keras, PyTorch, Pandas, Numpy, Scikit-learn, RAG, Semantic Search, Hugging Face Transformers, OpenAI API, LlamaIndex, YOLO, GPT, OCR, ONNX*

Full-Stack Development: *FastAPI, OpenAPI, Django, Flask, REST, GraphQL, OAuth2, HTML, CSS, React*

DevOps & Tools: *Docker, Git, GitLab, Azure, Postman, VS Code, LangChain, LangGraph, Dify, n8n*

Database: *PostgreSQL, MySQL, MongoDB, SQLAlchemy, FAISS, Chroma, Milvus*

Languages: *English (IELTS 7.0), Vietnamese (Native)*