Dang Khoa Le (Liam)

Swinburne, Melbourne • ✉ binkhoa1812@gmail.com • 0450-445-373 • https://github.com/Lelekhoa1812 • https://huggingface.co/BinKhoaLe1812 • https://www.linkedin.com/in/dang-khoa-le-96a6332a8

EDUCATION

SWINBURNE UNIVERSITY OF TECHNOLOGY — Hawthorn, Victoria, Australia
Bachelor of Software Engineer (with Honours) — February 2022 – November 2025

EXPERIENCES

MedSwin — Hawthorn, Victoria, Australia
AI Researcher | Lab Leader — August 2025 – February 2026

Led research lab team, funded by Swinburne University of Technology to build a multi-agent system with domain-specialist agents (diagnostics, pharmacology, triage) coordinated by a reasoning agentic orchestrator, boosting IR rate and clinical guidance for specialists.

- Boosted diagnostic **accuracy & safety** for LLM agents on 500k+ curated SFT & synthetic cases, using Knowledge Distillation, LoRA/QLoRA (GRPO reasoning), merging techniques, and augmentation (QAC paraphrasing/chunking, self-consistency, counterfactuals, backtranslation, PHI removal).
- Built **agentic RAG** (Tree + HNSW + FAISS) over EMR/EHR + PubMed with retrieval safety rails, real-time update, and citations to mitigate hallucination. Finetuned 300M embedding and reranking model on over 5M curated.
- Trained biomedical reranker model on 8M selective medical corpus. Integrated in MAC (multi-agent conversation) architecture to boost IR rate.
- Established strict evaluation: MedMCQA / PubMedQA / MedQuAD (86% BertScore) benchmarking, semantic similarity audits with biomedical embeddings, early-stopping LR scheduling; trained reproducibly on HPC.

Mizco — Pert, WA, Australia
LLMOps Engineer | Contract — November 2025 – March 2026

Led a self-sourced team (contracted via Teqar) to build a client–server, plugin-style multi-agent assistant that ingests PDFs/DOCX/XLSX, normalizes to Markdown and JSON objects, performs semantic chunking + RAG, and auto-completes technical outcome docs with grounded, auditable facts.

- Cut manual document assembly from **~80 hours to minutes**, while maintaining source traceability, accuracy and customisability.
- Architected a privacy-first stack: **Llama-3.1** (orchestrator), **olmOCR** (parsing), **EmbeddingGemma** (embeddings), **FAISS/HNSW** (vector search) in multi-layered tree, FastAPI on private on-prem runtime.
- Delivered agentic tooling to CRUD documents on staff request (update tables, fill cells, rewrite sections) and incremental re-ingest on file changes; achieved **≥90% autofill coverage** with **<2% error** in audits**.**

SuperEdTech — Melbourne, Victoria, Australia
Cofounder | CTO — June 2025 - Present

Our team of 4 believing in domain-specific, reliable and affordable education, we support university students and graduates in both their studies and career, leveraged by AI and technology capability.

- **AI Tutor:** Developed a dynamic, curriculum-aligned tutoring web app that creates personalized learning roadmaps and automates assignment, grading using a multi-agent RAG and MCP tools integrations.
- **StudyBuddy:** End-to-end hierarchal TreeRAG, HNSW, and FAISS application with LLM and OCR technologies allows students to chat with their uploaded documents (PDF/DOCX) for precise, source-based learning.
- **IELTS Enhancer:** AI writing coach that provides personalized feedback, correctness, and guidelines to help users achieve their target IELTS band. Developed with CoT multi NLP, SLM/LLM hybrid architecture.
- **CVA:** Engineered career coaching AI, semantic provisional, crafts job-specific resumes and cover letters to maximize applicant success, leverage RAG.
- **Fretboard:** Interactive string instruments tutor app with AI assistants for chord, scale, vibe analyser and instructional feedbacks on progress.

AITC — Ho Chi Minh City, Vietnam
LLMOps Engineer | Team Leader — May 2025 – November 2025

Architected and led the development of a novel AI/LLM solution, integrating Large Language Models (LLMs), Natural Language Processing (NLP), OCR (Transformer), RAG and hybrid recall architecture, to transform unstructured PDF documents into actionable industrial AI assistance. Partnered with major hardware suppliers like Nvidia, IBM, HPE, Oracle, we aim to integrate advance AI solution to automate tasks, enhance productivity and efficiency.

Pythera — Ho Chi Minh City, Vietnam
AI Engineer | Part-time — June 2025 – September 2025

Contributor of CLARA, Vietnamese nationally recognized medical AI assistant (LLM and VLM), trained on 400k+ doctor-annotated cases, including 210k+ X-rays, MRI scans, and synthesis / distillations.

- Agentic RAG Development: Led the creation of an EMR-integrated RAG module to boost diagnostic accuracy and medication safety.
- Model Excellence: Engineered systems for a model achieving top-tier benchmarks, including >95% Recall and >92% Precision on labelled dataset.

Hai Au Academy — Ho Chi Minh City, Vietnam
Fullstack Developer | Contract — July 2025 – September 2025

Develop a comprehensive web-based student management platform designed for language schools to handle student enrollment, class management, placement testing, and administrative tasks. The system supports both Vietnamese and English interfaces and serves multiple user roles including students, staff, and administrators.

CPG Seedcom — Ho Chi Minh City, Vietnam
Data Scientist | Contract — June 2025 – August 2025

Leading a team of 6 to build a Vietnamese-language AI assistant, converts business questions into SQL queries across MySQL and Oracle, fusing multi-modal self-QA for Reinforcement Learning over failures and successes. Built with FastAPI, Gemini & Qwen3, MongoDB vectorbase, leveraging Chain-of-Thought and Agentic RAG for autonomous analytics in sales, supply chain, and marketing. The system **reduces manual workforces by up to 80%**, delivers instant, explainable insights, a unique plugin solution enables direct database-to-decision workflows without human intervention.

PopTech — Ho Chi Minh City, Vietnam

MLOps Engineer | Part-time — May 2025 – August 2025

Led a 4-member team to architect and build a scalable electricity forecasting and control application from IoT device. Integrated data cleaning pipeline, including KNNImputer for data backfiller and abnormalities detection, with LinearRegression imputer for timestamp fallbacks. Integrated ML/DL (TFT, XGBoost), Reinforcement Learning (DQN), and LLM-driven advisory to predict billing, optimize device usage, and automate energy-saving decisions. Enabled multimodal interaction via ASR, text, and voice synthesis for a fully interactive LLMOps experience.

Nam Á Bank — Ho Chi Minh City, Vietnam

Software Engineer - Data Science | Internship — December 2024 – March 2025

Participate in fintech digitalization projects with full-stack development (Vert.x Backend, Java Android) and AI integration (CV, OCR, LLM, Chatbot). Proposed strategic enhancements:

- **Explore remittance services** targeting 5M+ overseas Vietnamese. Taking role in participation at Australian market, research for deployment.
- **Redesign mobile app** concept for a faster access, shifting complex tasks to web-services, deliver a user-friendly UI for accessibility-focused users.
- **Implement passport OCR** to enable foreign customer onboarding with banking programs, overcome traditional limitation for identification with domestic ID. Implement synthetic data generation.

iNet Solution — Ho Chi Minh City, Vietnam

Software Engineer - Data Science | Internship — December 2024 – March 2025

Analysing data, document, article for Sentiment Analysis and Business Overview with LLM models, integration with cloud computing architecture. Leading team to leverage Obsei framework and Transformers for sentiment analysis, fact-checking, summarization, **deliver actionable business insights** for clients.

The Auburn Coffee — Hawthorn East, Victoria, Australia

Website Developer | Part-time — April 2024 – November 2025

Developed the official website for a local commuter café at Auburn Station. Features include interactive menu, real-time pre-ordering, loyalty program with personalize QR code, and staff dashboard for order management, rating and feedback, sale analysis. Built with VueJS, Node.js (Vercel), database storage on MongoDB, and Firebase. Revolutionizing the business with digitalisation procedure, skyrocketing sale on an instance by 30%.

VietDan JSC — Ho Chi Minh City, Vietnam

MLOps Engineer | Part-time — December 2024 – February 2025

Developed and deployed a real-time Water Meter OCR solution for a government-partnered client, leveraging Next.js, Node.js, and a Flask API. Independently curated and annotated over 10,000 training data samples, enabling precise object detection with YOLOv5xu and advanced CV/OCR for text recognition. Achieved a 35% reduction in manual labor by successfully deploying the solution on Jetson Nano and AWS EC2, exceeding project key performance indicators.

Alta Media — Ho Chi Minh City, Vietnam

DevOps Engineer | Internship — December 2022 – February 2023

Dev-Ops Engineer, deploying software architecture with Docker. Implement backend infrastructure for internal software systems and AWS deployment.

PROJECTS

MedViệtAI — Ho Chi Minh City, Vietnam

Founder – National Talent Program — June 2025 – Present

MedViệtAI is a strategic national initiative, non-profit organisation sponsored by Vietnam's National Data Association (NDA), to assert the nation's sovereignty in the age of Artificial Intelligence. The program delivers foundational, Vietnam-owned LLM/SLM, purpose-built for the Vietnamese language and context. This core technology will power a novel agentic AI architecture designed to tackle critical national challenges, beginning with a transformative focus on healthcare. By leveraging a secure, national-scale data acquisition strategy, MedViệtAI will enable advanced analysis of Electronic Medical Records (EMRs) for proactive health monitoring and personalized medical insights, placing Vietnam at the forefront of AI-driven healthcare innovation.

SkyLedge - Capstone Project — Hawthorn, Victoria, Australia

Team Leader – AI and System Architecturer — February 2025 – February 2026

Led a team of 4 to build an OBD-II-based predictive maintenance system for SkyLedge, targeting reduced fleet servicing costs. Implemented real-time data logging with Raspberry Pi, automated cloud data engineering (FastAPI, Docker), used KMean/BGMM PCA clustering for driver behaviour and cost-efficiency analysis. Prototyped fault prediction with XGBoost, LSTM, all models are opted in RLHF to adapt with new policies, design interactive dashboard

Triage Response: LLM-Driven Solution and Drone Medication Delivery — Hawthorn, Victoria, Australia

Team Leader – AI, RAG and System Architecturer — February 2025 – May 2025

Developed and deployed an AI-powered emergency triage system at Swinburne University of Technology, integrating React Native (mobile), React (web portal), FastAPI (backend), Whisper-v3, Gemini 2.5, Qwen-VL, FAISS, and MongoDB with MIMIC-IV triage-response data to enable real-time voice-triggered alerts, clinical decision (CPG) support, and secured-fast drone (Great Shark 330 PRO VTOL) medication delivery.

SKILLS

Programming Languages: Python, Java, Kotlin, SQL, C++, C#, JavaScript, TypeScript.

Communication Languages: English, Vietnamese, Mandarin

Frameworks: FastAPI, Flask, Node.js, SpringBoot, Vue, React, Angular, Next.js.

AI/ML: TensorFlow, PyTorch, YOLO, ResNet, CNN, LLM, VLM, NLP, RAG, Knowledge Distillation, ASR, Machine Learning, Deep Learning, Synthesis

Tools: Docker, Kubernetes, MongoDB, MySQL, AWS, Azure, Google Cloud Console, Oracle, Jenkin, Jira, Git, Firebase.