Dang Khoa Le

103844421@student.swin.edu.au • 0450445373 • https://lelekhoa1812.github.io/My-Profile/ • https://www.linkedin.com/in/dang-khoa-le-96a6332a8

EDUCATION

Swinburne University of Technology

Software Engineer

SKILLS

Vietnamese, English, Mandarin

PROJECTS

Triage Response: LLM-Driven Solution and Drone Medication Delivery

* Engineered an AI-powered emergency triage system; developed and integrated LLM models (Whisper, Gemini, Qwen-VL) trained on MIMIC-IV data within a full-stack architecture. * Implemented real-time voice-triggered alerts, clinical decision support, and secure drone medication delivery capabilities, enhancing emergency response and patient care.

SkyLedge - Capstone Project

* Spearheaded SkyLedge, an OBD-II predictive maintenance system, optimizing fleet servicing costs by 15% through real-time data, cloud pipelines, and advanced fault prediction models. * Led a 4-member team, architecting the end-to-end solution from hardware integration and data processing to a comprehensive user monitoring dashboard.

AI Tutoring

* Solely engineered an AI tutoring web app, integrating multi-RAG LLMs (Gemini, Qwen) and custom KB embeddings to deliver personalized, curriculum-aligned instruction. * Designed and deployed a microservices architecture with Next.js, FastAPI, and MongoDB, powering features like dynamic tutoring, automated grading, and content crawling.

Medical Chatbot

* Developed a real-time AI medical chatbot utilizing Gemini Pro 2.5 and a NodeRAG architecture (FAISS-powered) for diagnosis and treatment assistance. * Integrated dynamic caching, model fine-tuning, and multi-language support for enhanced performance and global accessibility; deployed a full-stack system with FastAPI, NodeJS, and MongoDB.

EXPERIENCE

CPG Seedcom — Data Science | Team Leader

June 2025 - September 2025

Led a 6-member Data Science team to build a Vietnamese AI assistant utilizing FastAPI , Gemini , Qwen3 , MongoDB , Chain-of-Thought (CoT) , Agentic RAG , and Reinforcement Learning*. Automated natural language-to-SQL query generation (MySQL , Oracle) for autonomous analytics* in sales, supply chain, and marketing. Reduced manual workforce by 80%, delivering instant, explainable insights and enabling direct database-to-decision workflows*.

Pythera — AI Engineer | Part-time June 2025 - Now

Led the full-stack development and optimization of core AI components for Clara, a nationally recognized medical LLM+VLM assistant. This involved processing a massive dataset of 400k+ doctor-annotated cases, including 210k+ X-rays, and leveraging advanced data synthesis techniques to enhance model robustness and generalizability . Pioneered and led the development of an agentic Retrieval-Augmented Generation (RAG) module , integrating Electronic Medical Records (EMR) . This system improved diagnosis accuracy by an estimated 15% and medication safety by 20% by dynamically retrieving and synthesizing patient-specific information using semantic search and intelligent agent orchestration . Engineered and optimized large-scale model training pipelines, reducing training time by 25% , and conducted rigorous benchmarking to achieve state-of-the-art performance across critical clinical metrics. Designed and implemented an intelligent chat memory retrieval system , leveraging vector databases and advanced NLP , to ensure contextual continuity across patient interactions, improving user experience and clinical efficacy*.

Nam Á Bank — Software Engineering - Data Science | Intern December 2024 - March 2025

* Engineered full-stack solutions (Vert.x, Java Android) and integrated advanced AI (CV, OCR, LLM, Chatbot) to digitalize core fintech operations. * Spearheaded strategic product enhancements, proposing international remittance service expansion and designing a passport OCR system for foreign customer onboarding, leveraging synthetic data generation.