MAI CHI BAO

+84 327 620 475
dev.to/mrzaizai2k
Bảo Chí | LinkedIn
chibao24.12.1999@gmail.com
github.com/mrzaizai2k

EDUCATION

2024 – Now — **Ho Chi Minh City University of Technology** (*Master degree*)
Major: Computer science

2017 – 2022 — **Ho Chi Minh City University of Technology** (*Bachelor of Engineering*)
Major: Electronics – Telecommunications Engineering (7.6/10)

WORK EXPERIENCE

6/2025 – 10/2025 — **JOBLOGIC** — *Senior LLM Engineer*

- Built an end-to-end LLM evaluation platform using ReactJS, Azure Blob Storage, MongoDB, Azure AI Search, Kafka, LangChain for up to 30 concurrent processes.
- Fine-tuned and benchmarked LLaMA 3.1, Gemma 3, Phi 4 for RAG applications using both local and Azure resources, achieving a 23% performance improvement
- Developed an MCP-based agentic system for consumer consultation, integrating LLM reasoning and external API orchestration.

2/2023 – 6/2025 — **NAM A BANK** — *Team Leader - Senior Machine Learning Engineer*

- Fine-tuned a VLM (Qwen2VL) model using LoRA method for golf scorecard extraction, achieving 83% higher accuracy than traditional OCR with just 70 samples, while significantly reducing GPU VRAM usage.
- Developed an advanced RAG-based Chat Bot with LLM, FAISS vector database, leveraging model routing to reduce latency and cost by up to 100x
- Building a Facial Recognition system with Anti-spoofing functionality using PyTorch, YOLO, ONNX for 1.8M+ customers, aligning with Decision No. 2345 of the Government.
- Lead on Customer churn prediction with LightGBM, Scikit-learn, and Pandas offers over 78% accuracy while predicting 28.5% of the bank's overall customers.
- Keynote Speaker at International Conference on Economics 2023 about Big Data

2/2022 – 1/2023 — **LONG Centre for Educational Development** — *Algorithm Engineer*

- Researching on Point Set Registration with Deep Learning approaches
- Compare techniques of Graph-based, Convolutional-based, and Pointwise MLP methods

RELATED PROJECTS

8/2024 – 10/2024 — **Multilanguage Invoice OCR**

- Optimized handwritten invoice extraction for a German company using Vision LLM – Qwen2VL and OpenAI to meet domain-specific requirements.
- Developed and deployed a full-stack AI application using Python, ReactJS, MongoDB, and Nginx on AWS EC2, containerized with Docker Compose for scalable deployment.
- Automated workflows through DevOps/MLOps pipelines, CI/CD using GitHub Actions

AWARDS/ ACHIEVEMENTS

2024 — Received a promotion to the position of Senior Application Development at Nam A Bank.

2023 — Award Effective Contribution in Nam A Bank (3 consecutive quarters)

2022 — 7.5 / 9.0 (IELTS)