Nguyễn Đình Khải

+84 37838454 nguyendinhkhaiqt2005@gmail.com ngwkhai.com github.com/ngwkhai Hà Nội, Việt Nam

TÓM TẮT

Chuyên viên Ứng dụng AI tại VinSmart Future, phụ trách mảng dữ liệu AI và kiểm soát chất lượng dữ liệu; có kinh nghiệm nghiên cứu xử lý ngôn ngữ tự nhiên, tinh chỉnh mô hình ngôn ngữ lớn, cùng các sản phẩm AI tự xây dựng và triển khai trọn vẹn. Trong 3–5 năm tới, mục tiêu của tôi là đảm nhận vai trò dẫn dắt kỹ thuật hoặc quản lý dự án và đóng góp lâu dài cho sự phát triển của công ty.

HỌC VẤN

Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội 08/2023 - Hiện tại
Cử nhân Trí tuệ Nhân tạo GPA: **3.61/4**

KINH NGHIỆM

VinSmart Future

Chuyên viên Ứng dụng AI 11/2026 - Hiện tại
Đảm nhận mảng **dữ liệu AI chuyên sâu** và chịu trách nhiệm **kiểm soát chất lượng dữ liệu**, bảo đảm dữ liệu đạt chuẩn; hỗ trợ quản lý **đội ngũ 12 thành viên**.

Kỹ sư tập sự Ứng dụng AI 09/2026 - 11/2026
Thực hiện **kiểm soát chất lượng dữ liệu**, rà soát và đối chiếu với tiêu chuẩn trước khi đưa vào sử dụng; hỗ trợ quản lý **đội ngũ 6 thành viên**.

Thực tập sinh AI 05/2026 - 08/2026
Xây dựng **pipeline thu thập và kiểm soát chất lượng** dữ liệu du lịch: từ thu thập, nạp dữ liệu tới bước rà soát, xác minh.

Viettel Networks 09/2025 - 01/2026
Thực tập sinh AI
Nghiên cứu sinh **cấu trúc dự án** từ đặc tả ngôn ngữ tự nhiên, xử lý hiện tượng "mù cấu trúc" của LLM khi sinh kết quả phân cấp.
Xây dựng tập huấn luyện và bộ chỉ số đánh giá; tinh chỉnh các mô hình mở nhỏ gọn bằng **LoRA** với hàm mất mát đồng bộ với chỉ số đó.

DỰ ÁN

ResearchMap AI — *Một mình* (researchmap.ngwkhai.com) 2026
Tự xây toàn bộ sản phẩm: backend FastAPI, PostgreSQL, pgvector, frontend React, đăng nhập Google OAuth và TOTP, gói có hạn mức, trang quản trị.
Thiết kế pipeline phân tích: LLM lập kế hoạch truy vấn OpenAlex, phân cụm bài báo bằng KMeans, khoảng trống tính từ các chủ đề không có bài chung.
Xây dựng chat **agentic RAG** (7 công cụ, tối đa 4 vòng) chỉ trích dẫn bài đã truy xuất: **36/36** bài được xác minh có thật, precision@5 **0.867**.
Công nghệ: Python, FastAPI, PostgreSQL, pgvector, scikit-learn, OpenAI API, React, TypeScript, Cloud Run.

TaxMate AI — *Trưởng nhóm (4 người)* (taxmate.ngwkhai.com) 2026
Dẫn dắt nhóm bốn người xây dựng bốn core (Legal RAG, Risk, Graph, tối ưu lượng tử); chốt phạm vi, lịch phát hành và tích hợp thành một sản phẩm.
Làm lại truy xuất pháp luật bằng **BM25** trên cặp âm tiết tiếng Việt, chỉ trích dẫn chunk đã truy xuất: recall@5 tăng từ 32 lên **52/56** câu.
Xây dựng core Risk, xác thực đa tenant trên Supabase, triển khai dev và production trên Cloud Run qua GitHub Actions.
Công nghệ: Python, FastAPI, PostgreSQL, Supabase Auth, OpenAI API, React, TypeScript, Cloud Run.

Neural Machine Translation — *Trưởng nhóm (2 người)* (github.com/ngwkhai/Neural-Machine-Translation) 06/2025 - 07/2025
Xây dựng **bộ dữ liệu khôi phục dấu tiếng Việt quy mô lớn đầu tiên** (12,5 triệu cặp câu); cài đặt từ đầu **Seq2Seq** RNN, GRU, LSTM và Transformer.
Đạt **BLEU 81.31** và **ChrF++ 88.57** với Transformer, vượt trội đáng kể so với các mốc cơ sở RNN và GRU.
Công nghệ: Python, PyTorch, NumPy, Pandas, BLEU, ChrF++.

Optimization for AI Inference Engines on GPUs 2025
Tối ưu suy luận GPU cho ResNet50 và DistilBERT (FP16, INT8, torch.compile, ONNX); profiling CUDA xác định nút thắt ở kernel GEMM và ReLU.
FP16 tăng thông lượng DistilBERT từ 305 lên 1202 mẫu/s ở batch size 8 (**≈4×**); INT8 (TensorRT) tăng tốc **2–5×** nhưng cần hiệu chuẩn cẩn thận.
Công nghệ: PyTorch, TensorRT, ONNX Runtime, TorchScript, CUDA profiling.

KỸ NĂNG

Lập trình	Python, C/C++, Java, SQL, TypeScript
Học máy	PyTorch, Scikit-learn, TensorFlow, Keras, XGBoost, Transformer, Seq2Seq, YOLO, tinh chỉnh LoRA
LLM & Truy xuất	OpenAI API, RAG, agentic tool calling, BM25, pgvector, tìm kiếm lai (hybrid search)
Hệ thống & Suy luận	FastAPI, PostgreSQL, React, Docker, Cloud Run, GitHub Actions, Hadoop, PySpark, TensorRT, ONNX Runtime
Ngôn ngữ	Tiếng Việt: Bản ngữ, Tiếng Anh: Thành thạo

GIẢI THƯỞNG & THÀNH TÍCH

Giải Ba, Giải thưởng Nghiên cứu khoa học sinh viên — Trí tuệ Nhân tạo, Trường Đại học Công nghệ	2026
Giải Ba, Bảng Sinh viên — Humanitarian Logistics Hackathon	2026
Học bổng Khuyến khích học tập, Trường Đại học Công nghệ	2023-2024
Giải Nhì, Kỳ thi Học sinh giỏi cấp tỉnh môn Vật lý, tỉnh Thái Nguyên	2023
Giải Ba, Kỳ thi Học sinh giỏi cấp tỉnh môn Tin học, tỉnh Thái Nguyên	2021