

PHAN VĂN BẰNG

AI Engineer & Backend Developer

LIÊN HỆ

+84 96 419 6652
pvbang23092002@gmail.com
github.com/pvbang
linkedin.com/in/pvbang
pvbang.github.io/cv (online)

KỸ NĂNG CHUYÊN MÔN

Ngôn ngữ lập trình:

Python, C#, Java, C++, C

Frameworks & Thư viện:

LangChain, FastAPI, Flask, Streamlit

Cơ sở dữ liệu:

Vector DB (Qdrant, Faiss, Chroma),
SQL (MySQL, SQLite), NoSQL
(MongoDB, Firebase), Graph DB
(Neo4j)

AI/ML & Xử lý ảnh:

Prompt Engineering, LLM Fine-tuning,
RAG, Agentic Systems (Tool Calling,
CrewAI, MCP), ComfyUI Workflows

Kiến trúc & Hệ thống:

Microservices, RESTful API Design,
Docker, VPS Management (Ubuntu,
Linux)

DevOps & CI/CD:

AWS (S3, EC2), Cloudflare, Vercel,
GitHub Actions

Công cụ & Quy trình:

Git, N8N Automation, Data
Crawling/Scraping, Unit Testing

HỌC VẤN

Kỹ sư Công nghệ Thông tin

VKU, Đà Nẵng 2020 - 2025

Trường Đại học Công nghệ Thông tin
và Truyền thông Việt - Hàn

KỸ NĂNG MỀM

- Làm việc nhóm & Lãnh đạo:** Kinh nghiệm điều phối nhóm, phân công nhiệm vụ, và hướng dẫn đồng nghiệp để đạt được mục tiêu dự án.
- Học hỏi liên tục:** Chủ động học tập các công nghệ AI và phát triển phần mềm mới để nâng cao năng lực và đóng góp vào sự phát triển của nhóm.

TÓM TẮT

AI Engineer với 2+ năm kinh nghiệm trong thiết kế, xây dựng và triển khai các giải pháp AI tiên tiến. Chuyên sâu về chatbot thông minh sử dụng RAG, Agentic Systems, Tool Calling, và quy trình xử lý ảnh phức tạp với ComfyUI. Kinh nghiệm thực tế trong LLM fine-tuning, tối ưu Prompt Engineering, và quản lý hạ tầng AI từ vector database đến triển khai cloud.

- Mục tiêu ngắn hạn:** Đảm nhận vị trí AI Engineer tại doanh nghiệp có tầm nhìn xa, ứng dụng chuyên môn AI/ML sâu rộng để phát triển giải pháp đột phá và đóng góp vào các sản phẩm có giá trị cao.
- Mục tiêu dài hạn:** Phát triển thành chuyên gia AI hàng đầu, dẫn dắt các sáng kiến công nghệ tiên phong và xây dựng đội ngũ AI mạnh mẽ để định hình tương lai của trí tuệ nhân tạo.

KINH NGHIỆM LÀM VIỆC

AI Engineer | Backend Developer

06/2023 - Hiện tại

MekongAI

> Phát triển và Triển khai Giải pháp AI:

- Triển khai kiến trúc RAG cho chatbot, cải thiện độ chính xác 20-30% và giảm thời gian phản hồi xuống dưới 2 giây.
- Xây dựng hệ thống AI sử dụng Agentic Systems (Tool Calling, CrewAI, MCP)
- Tối ưu Prompt Engineering, tăng tỷ lệ thành công của LLM 30% và đảm bảo tính nhất quán nội dung.
- Phát triển chatbot AI chuyên biệt (Pháp lý, Mạng xã hội, Xe cộ) sử dụng LangChain, LLM, RAG, và kiến trúc Agent.

> Huấn luyện và Tối ưu Mô hình AI:

- Fine-tuning LLM trên dataset tùy chỉnh để cải thiện phản hồi theo ngữ cảnh.
- Huấn luyện mô hình LoRA sử dụng AI-Toolkit, Flux để tạo ảnh sản phẩm chất lượng cao.

> Phát triển Backend và Quản lý Cơ sở dữ liệu:

- Phát triển RESTful API hiệu năng cao sử dụng FastAPI, duy trì thời gian phản hồi dưới 200ms.
- Triển khai Vector DB (Qdrant) để lưu trữ embedding và Graph DB (Neo4j) cho knowledge graph.
- Quản lý cơ sở dữ liệu SQL (MySQL) và NoSQL (MongoDB) đảm bảo tính toàn vẹn dữ liệu và hiệu suất ổn định.

> Xử lý Ảnh AI (Diffusion, ComfyUI):

- Xây dựng 20+ workflow ảnh phức tạp sử dụng ComfyUI cho Generation, FaceSwap, OutfitSwap, Background Replacement, VTO, v.v.
- Phát triển WebSocket API kết nối workflow với ComfyUI để tích hợp Frontend liền mạch.

> Quản lý Hạ tầng và Tự động hóa:

- Quản lý VPS (Ubuntu, Linux) cho triển khai Local LLM, Training, Testing.
- Tự động hóa 5+ workflow sử dụng N8N (RAG, Google Calendar, Drive, Zalo, Discord, Telegram), giảm 40% tác vụ thủ công.
- Triển khai cloud hosting với tối ưu Cloudflare để nâng cao hiệu suất và bảo mật.
- Quản lý AWS S3 cho lưu trữ dữ liệu có thể mở rộng bao gồm tài liệu, ảnh và tài nguyên tĩnh.