

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 (CV Trực tuyến)

KỸ NĂNG CHUYÊN MÔN

Ngôn ngữ lập trình:

Python, JavaScript & TypeScript, C#, Java, C++, C

Frameworks & Thư viện:

LangChain, FastAPI, Flask, Streamlit, Spring Boot, Laravel, ExpressJS, React, React Native, Unity

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, Thiết kế RESTful API, Docker, Quản lý VPS (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 & Hợp tác:** Chủ động chia sẻ kiến thức.
- Điều phối nhóm:** Phân công nhiệm vụ, theo dõi tiến độ.
- Phân tích & Giải quyết vấn đề:** Giải pháp có hệ thống, sáng tạo.
- Giao tiếp:** Tương tác rõ ràng về kỹ thuật và phi kỹ thuật.
- Học hỏi liên tục:** Chủ động cập nhật kỹ năng về AI & lập trình.

TÓM TẮT

AI Engineer & Backend Developer (2+ năm kinh nghiệm backend, 1+ năm AI). Thành thạo thiết kế, xây dựng, triển khai các giải pháp AI tiên tiến (chatbot RAG, Agentic Systems, Tool Calling, ComfyUI image workflows). Chủ động nghiên cứu & áp dụng công nghệ mới để tối ưu hiệu suất, bảo mật, khả năng mở rộng.

- Mục tiêu ngắn hạn:** Vị trí AI Engineer, phát triển AI đột phá, xử lý dữ liệu, tối ưu hệ thống cho các dự án giá trị cao.
- Mục tiêu dài hạn:** Chuyên gia AI hàng đầu, dẫn dắt đổi mới, phát triển AI tiên phong mang lại lợi thế cho công ty & lợi ích cộng đồng.

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 Retrieval-Augmented Generation (RAG) cho chatbot, cải thiện độ chính xác truy xuất thông tin 15-20% & giảm thời gian phản hồi xuống <2s.
- Xây dựng Agentic Systems (Tool Calling, MCP) tự động hóa >8 quy trình nghiệp vụ phức tạp, tiết kiệm ~5-7 giờ/tuần cho mỗi thành viên.
- Áp dụng kỹ thuật MCP tối ưu luồng xử lý thông tin của Agent, đạt tốc độ thực thi nhanh hơn ~20% trong thử nghiệm.
- Tinh chỉnh Prompt Engineering, tăng tỷ lệ giải quyết vấn đề thành công của LLM thêm 20% & đảm bảo tính nhất quán nội dung.
- Phát triển chatbot AI chuyên biệt (Luật, Mạng xã hội, Xe cộ) sử dụng LangChain, LLMs (RAG, Agent), nhận phản hồi tích cực.

> Huấn luyện và Tối ưu Mô hình AI:

- Fine-tuning LLM trên tập dữ liệu để tiếp thu kiến thức mới và điều chỉnh phương pháp phản hồi phù hợp ngữ cảnh.
- Huấn luyện mô hình LoRA (AI-Toolkit, Flux) tạo ảnh chất lượng cao, giảm thời gian tạo ảnh thủ công 30%.

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

- Phát triển RESTful API hiệu năng cao (FastAPI) với thời gian phản hồi <200ms, hỗ trợ 800 reqs/phút.
- Triển khai & tối ưu Vector DB (Qdrant) & Graph DB (Neo4j) cho 2 dự án AI chủ chốt.
- Quản lý CSDL SQL (MySQL) & NoSQL (MongoDB), đảm bảo toàn vẹn dữ liệu & hiệu suất ổn định.

> Công nghệ Xử lý Ảnh AI (ComfyUI):

- Xây dựng & quản lý >20 ComfyUI workflows phức tạp (Tạo ảnh, FaceSwap, VTO, v.v.), xử lý 300-400 ảnh/ngày.
- Triển khai Web Sockets tích hợp ComfyUI workflow & tạo API cho frontend.

> Quản lý Hạ tầng và Tự động hóa:

- Quản trị VPS (Ubuntu, Linux) cho Local LLM, Training, Testing.
- Sử dụng N8N tự động hóa >5 workflows lặp lại (RAG, tích hợp ứng dụng, tác vụ dữ liệu).
- Quản lý hosting, domain (Hostinger), CDN (Cloudflare) cho subdomain, tốc độ & bảo mật.