PHAN VAN BANG

AI Engineer │ Backend Developer

CONTACT

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

TECHNICAL SKILLS

Programming Languages:
Python, JavaScript & TypeScript, C#, Java, C++, C

Frameworks & Libraries:
LangChain, FastAPI, Flask, Streamlit, Spring Boot, Laravel, ExpressJS, React, React Native, Unity

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

AI/ML & Image Processing:
Prompt Engineering, LLM Fine-tuning, RAG, Agentic Systems (Tool Calling, MCP), ComfyUI Workflows

Architecture & Systems:
Microservices, RESTful API Design, Docker, VPS Management (Ubuntu, Linux)

DevOps & CI/CD:
AWS (S3, EC2), Cloudflare, Vercel, GitHub Actions

Tools & Processes:
Git, N8N Automation, Data Crawling/Scraping, Unit Testing

EDUCATION

Engineer of Information Technology
VKU, Da Nang   *2020 - 2025*
Vietnam-Korea University of Information and Communication Technology

SOFT SKILLS

- **Teamwork & Collaboration:** Proactive knowledge sharing.
- **Team Coordination:** Task assignment, progress tracking.
- **Analytical & Problem-Solving:** Systematic, creative solutions.
- **Communication:** Clear technical and non-technical interaction.
- **Continuous Learning:** Proactive skill updating in AI & dev.

SUMMARY

AI Engineer & Backend Developer (2+ yrs backend, 1+ yr AI). Proficient in designing, building, deploying advanced AI solutions (RAG chatbots, Agentic Systems, Tool Calling, ComfyUI image workflows). Proactively researches & applies new tech for performance, security, scalability.

- **Short-term Goal:** AI Engineer role, develop breakthrough AI, process data, optimize systems for high-value projects.
- **Long-term Goal:** Leading AI expert, spearhead innovations, develop pioneering AI for company edge & community benefit.

WORK EXPERIENCE

AI Engineer │ Backend Developer   *06/2023 - Present*
MekongAI

> **Develop and Deploy AI Solutions:**
- Implemented RAG architecture for chatbots, improving info retrieval accuracy by **15-20%** & reducing response time to **<2s**.
- Built Agentic Systems (Tool Calling, MCP) automating **>8 complex business processes**, saving **~5-7 hrs/week** per team member.
- Applied MCP techniques to optimize Agent info processing, achieving **~20% faster** task execution in tests.
- Refined Prompt Engineering, boosting LLM problem-solving success by **20%** & ensuring content consistency.
- Developed specialized AI chatbots (Law, Social Media, Vehicle) using LangChain, LLMs (RAG, Agent), receiving positive feedback.

> **Train and Optimize AI Models:**
- Fine-tuned LLMs on datasets for new knowledge acquisition and context-aware response adaptation.
- Trained LoRA models (AI-Toolkit, Flux) for high-fidelity image generation, reducing manual creation time by **30%**.

> **Backend Development and Database Management:**
- Developed high-performance RESTful APIs (FastAPI) with **<200ms** response time, supporting **800 reqs/min**.
- Deployed & optimized Vector DBs (Qdrant) & Graph DBs (Neo4j) for **2 key AI projects**.
- Managed SQL (MySQL) & NoSQL (MongoDB) databases, ensuring data integrity & stable performance.

> **AI Image Processing Technology (ComfyUI):**
- Built & managed **>20 complex ComfyUI workflows** (Image Gen, FaceSwap, VTO, etc.), processing **300-400 images/day**.
- Implemented Web Sockets for ComfyUI workflow integration & API creation for frontend.

> **Infrastructure Management and Automation:**
- Administered VPS (Ubuntu, Linux) for Local LLM, Training, Testing.
- Used N8N to automate **>5 repetitive workflows** (RAG, app integrations, data tasks).
- Managed hosting, domains (Hostinger), CDN (Cloudflare) for subdomains, speed & security.