

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)

TECHNICAL SKILLS

Programming Languages:

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

Frameworks & Libraries:

LangChain, FastAPI, Flask, Streamlit

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,
CrewAI, 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 & Leadership:** Experienced in team coordination, task assignment, and guiding colleagues to achieve project goals.
- Continuous Learning:** Proactively learning new AI and software development technologies to enhance capabilities and contribute to team growth.

SUMMARY

AI Engineer with 2+ years experience in designing, building, and deploying advanced AI solutions. Deep expertise in intelligent chatbots utilizing RAG, Agentic Systems, Tool Calling, and complex image processing workflows with ComfyUI. Practical experience in LLM fine-tuning, Prompt Engineering optimization, and AI infrastructure management from vector databases to cloud deployment.

- Short-term Goal:** Take on AI Engineer position at forward-thinking enterprise, applying deep AI/ML expertise to develop breakthrough solutions and contribute to high-value products.
- Long-term Goal:** Develop into leading AI expert, spearheading pioneering technology initiatives and building strong AI teams to shape the future of artificial intelligence.

WORK EXPERIENCE

AI Engineer | Backend Developer

06/2023 - Present

MekongAI

- AI Solutions Development and Deployment:**
 - Implemented RAG architecture for chatbots, improving accuracy by 20-30% and reducing response time to under 2 seconds.
 - Built AI systems using Agentic Systems (Tool Calling, CrewAI, MCP)
 - Optimized Prompt Engineering, increasing LLM success rate by 30% and ensuring content consistency.
 - Developed specialized AI chatbots (Legal, Social Media, Vehicle) using LangChain, LLM, RAG, and Agent architectures.
- AI Model Training and Optimization:**
 - Fine-tuned LLMs on custom datasets for improved contextual responses.
 - Trained LoRA models using AI-Toolkit, Flux for high-fidelity product image generation.
- Backend Development and Database Management:**
 - Developed high-performance RESTful APIs using FastAPI, maintaining sub-200ms response times.
 - Deployed Vector DB (Qdrant) for embedding storage and Graph DB (Neo4j) for knowledge graphs.
 - Managed SQL (MySQL) and NoSQL (MongoDB) databases ensuring data integrity and stable performance.
- AI Image Processing (Diffusion, ComfyUI):**
 - Built 20+ complex image workflows using ComfyUI for Generation, FaceSwap, OutfitSwap, Background Replacement, VTO, etc.
 - Developed WebSocket API connecting workflows with ComfyUI for seamless Frontend integration.
- Infrastructure Management and Automation:**
 - Managed VPS (Ubuntu, Linux) for Local LLM, Training, Testing deployments.
 - Automated 5+ workflows using N8N (RAG, Google Calendar, Drive, Zalo, Discord, Telegram), reducing manual tasks by 40%.
 - Deployed cloud hosting with Cloudflare optimization for enhanced performance and security.
 - Managed AWS S3 for scalable data storage including documents, images, and static resources.