Wan Mohd Fahmi Bin Fauzi

Sungai Petani, Kedah
fahmifauzi109@gmail.com
LinkedIn Profile: linkedin.com/in/wan-fahmifauzi/

Summary

Self-taught Generative AI Developer who successfully transitioned from medical background, specializing in building RESTful APIs and AI applications using FastAPI framework. Strong expertise in LLM development, prompt engineering, and AI strategy with demonstrated ability to implement production-ready solutions. Experienced in developing secure APIs that integrate various LLM models, showcasing rapid technical learning and adaptability.

Professional Experience

Generative AI Developer | Mesolitica | April 2024 - Present

- Design and implement RESTful APIs using FastAPI for AI service integration and deployment
- Architect and develop AI solutions using LangChain and various LLM models
- Build and optimize RAG pipelines for enhanced chatbot performance through API endpoints
- Lead prompt engineering initiatives and implement them via standardized API interfaces
- Create scalable API architectures for AI model deployment and evaluation across multiple languages

Generative AI Developer (Contract) | Mesolitica | January 2024 - March 2024

- Developed AI solutions focusing on language model applications
- Implemented prompt engineering techniques for improved model performance
- Collaborated on AI strategy and productivity optimization
- Build and optimize a simple RAG pipeline for AI application

Projects

Biar Kami Masak - Fun Recipe Recommender | Production Deployment | November 2024
Generative AI Developer

- Architected and developed a secure FastAPI backend for an AI-powered recipe recommendation system
- Integrated OpenAI GPT-4o-mini model for intelligent recipe suggestions and meal planning
- Implemented comprehensive API authentication system to ensure secure access
- Deployed the application on Railway platform with custom domain configuration
- Collaborated in a team environment, managing backend development while coordinating with frontend developer
- Technologies: FastAPI, OpenAI GPT-4 mini, API Authentication, Railway

Project Link: https://github.com/fahmifauzi00/biarkamimasak

Claude Search Chatbot | AWS Bedrock Implementation |August 2024
Generative AI Developer

- Architected and developed a production-ready chatbot using the Claude model on AWS Bedrock
- Implemented real-time web search capabilities to enhance response accuracy with current information
- Integrated advanced prompt engineering techniques for improved context handling
- Technologies used: AWS Bedrock, Claude LLM, Python, SERP APIs

Project Link: https://github.com/fahmifauzi00/claude-search.git

Professional Development

- Self-taught in Generative AI Development through online resources, documentation, and hands-on projects
- Rapidly acquired expertise in LLM implementation, API development, and cloud deployment
- Actively participate in AI development communities and stay current with latest developments in the field

Education & Certifications

Data Analytics Bootcamp   August 2023 - September 2023
Excelerate Asia

Bachelor of General Medicine   September 2014 - July 2020
Volgograd State Medical University   Volgograd, Russia

Coursera
- Google Data Analyst Specialization
- Generative AI for Everyone

Skills

- **Backend Development:** FastAPI, RESTful APIs, API Authentication, API Documentation, Middleware Development
- **Deployment & Infrastructure:** Railway, Domain Configuration, Production Deployment
- **AI & Cloud Platforms:** AWS Bedrock, LLM Integration, Prompt Engineering, LangChain, OpenAI GPT-4, Claude, Gemini, ChatGPT
- **LLM Development:** Prompt Engineering, Model Evaluation, Context Optimization
- **Programming & Tools:** Python, FastAPI, Web Search Integration, API Testing
- **Soft Skills/Core Competencies:** Problem-Solving, Teamwork, Communication, Self-Directed Learning, API Architecture Design, Content Evaluation & Quality Assessment, AI Strategy Development, Critical Analysis, Cross-lingual Content Understanding
- **Languages**: Malay (Native), English (Professional Working Proficiency), Russian (Limited Working Proficiency)