Nguyen Duy Minh

AI Engineer

- 0814249888
- ngduyminh2301@gmail.com
- Vinhomes Grand Park, 9 Districs ,TP Hồ Chí Minh

JOB OBJECTIVE

Aspiring AI Engineer with solid expertise in developing, fine-tuning, and deploying AI models, including **Large Language Models (LLMs).** With hands-on experience in both academic and professional settings, I have successfully contributed to AI projects involving model optimization, backend development, and chatbot implementation, eager to contribute my capabilities to innovative AI-driven projects.

CORE COMPETENCIES

Skills: LLMs, RAG, Git, Open-source models, Vector Database, Transformers, Deep Learning, Tool Calling, MLOPs, APIs.

Framework: Pytorch, Tensorflow, Langchain, Llama-index, Langgraph, HuggingFace, FastAPI, MongoDB.

WORK EXPERIENCE

Cyno Software Company

1/2024 -

AI ENGINEER

- **Customer Care Chatbot:** Contributed to the design **LLM system** and implementation of a RAG chatbot for company customer support.
- **Backend Development for POCY:** Led the development of a robust backend for POCY, an application designed to **detect scam/spam activities.**
- **Football Euro Prediction Model:** data collection and preprocessing efforts, using advanced data crawling techniques to build datasets for predicting outcomes in Euro football matches.

Amazing Technology and Solution Company Limited

9/2023 - 12/2023

AI ENGINEER INTERN (REMOTE)

- **Facial Recognition Module:** Engineered and implemented a facial tracking and tracing module, improving real-time monitoring capabilities.

PROJECTS

CUSTOMER CARE CHATBOT

5/2024 - 9/2024

Cyno Software

Team size: 3

ROLE: TEAM MEMBER

- **Contextual Enhancement:** Augmented model inputs by integrating access to external data sources, implemented vector-based storage for efficient retrieval of relevant information, improving the chatbot's contextual understanding and response accuracy.
- **System Optimization:** Implemented guardrails to safeguard the system and users, optimizing for reduced latency and cost efficiency through caching mechanisms.
- **Leveraged LangChain:** handling multi-step reasoning and dialogue management, ensuring seamless interaction flow.

TECHNOLOGIES:

- LangChain, PyTorch, Llama 3, HuggingFace, Faiss

POCY PROJECT

1/2024 - 6/2024

Cyno software

Team size: 8

ROLE: TEAM MEMBER

- **Data Preprocessing:** preprocessing of SMS and call data (remove unnecessary pattern and mormalize text) to prepare datasets for analysis and model training.
- **Custom architecture**: Reconstructed **Transformers from scratch**, ensuring enhanced model adaptability, hardware limitatation.
- **Deployment:** Integrated JWT into the base model.

TECHNOLOGIES:

- PyTorch, FastAPI, MongoDB, JWT