TRAN PHUOC SANG

INTERN AI

Phone: 0347763291   **Email:** transhack09@gmail.com

PROFESSIONAL SUMMARY

As a fourth-year student majoring in Computer Science, I am seeking an AI Engineer internship to apply and expand my knowledge in machine learning, deep learning. My goal is to collaborate with experts, enhance my technical skills, and play an active role in developing impactful AI solutions that contribute to the growth and success of the organization.

EDUCATION

Bachelor of Computer Science| 2021-present | GPA: 8.1

Ton Duc Thang University

- Relevant Coursework: Data Structures And Algorithms, Introduction to Machine Learning, Deep Learning, Introduction to Natural Language Processing, Knowledge Discovery and Data Mining.

PROJECTS

Fine-tune Qwen2.5-3B for Legal Chatbot

- Description: Continual pre-training the Qwen2.5-3B model on an extensive legal text corpus and fine-tuned it using a curated QA dataset for legal question answering. Leveraged the unsloth library in combination with the LoRA technique, enabling efficient adaptation and enhanced performance for legal NLP applications..
- Key Contributions:
  - Dataset Preparation: Build a custom dataset comprising 20,000 legal documents and 20,000 QA pairs.
  - Optimized training process using LoRA, and the SFTTrainer method.
  - Applied advanced fine-tuning techniques - integrating continued pre-training and RAG to significantly boost performance, raising the F1 score from around 0.30 to nearly 0.38 while also improving ROUGE metrics by approximately 10–15%.
- Key Technologies: Python, PyTorch, Hugging Face, Transformers, Pandas.
- Huggingface: phuocsang/qwen2.5-3B-Finetune-Vietnamese-Law

Legal Chatbot with Advanced RAG

- Description: Developed an advanced legal chatbot system specifically tailored for Vietnamese law using Retrieval Augmented Generation (RAG) techniques. Implemented a hierarchical document retrieval mechanism (parent-child relationships) that efficiently processes and analyzes legal documents.
- Key Contributions:
  - Applied advanced RAG techniques.
  - Multi-stage RAG Architecture: Created a comprehensive pipeline that processes documents from ingestion through grading to final response generation.
  - Interactive Interface: Built a user-friendly Streamlit application with contextual source references and transparency in the reasoning process.
  - Key Technologies: Python, PostgreSQL, Qdrant, Streamlit, Hugging Face, LangChain.
- GitHub: PhuocSang07/legal-chatbot

Fine-tune Vistral-7B For Travel Chatbot

- Description: Loaded the Viet-Mistral-7B-Chat model in 4-bit precision and applied LoRA (Low-Rank Adaptation) technology to fine-tune it on a Vietnamese travel QA synthetic dataset, generated by Gemini-1.5-pro.
- Key Contributions:
  - Created a custom dataset with 4,000 rows for fine-tuning the model.
  - Developed the model to provide high-quality answers in the Vietnamese travel domain.
  - Fine-tuned the model using specific hyperparameters in Peft, LoRA, and the SFTTrainer method.
- Key Technologies: Python, PyTorch, Hugging Face, Transformers, Pandas.
- GitHub: PhuocSang07/Fine-tune Vistral Chatbot

ADDITIONAL SKILLS

- Certificates: Aptis B2, Agile development & Scrum framework course 2023 - TechBase (11/2023), PyTorch-Image - CODEPRESSO (8/2024).
- Programming Languages: Python, JavaScript.
- Database: Microsoft SQL Sever, MongoDB, PostgreSQL.
- Solf skills: Communication, Critical Thinking, Problem-Solving, Teamwork.