

Dat Vo Dinh

Hanoi, Vietnam | +84-399-565-203 | www.dinhdat.com | github.com/datvodinh | vodinhdat03@gmail.com

EDUCATIONS

Hanoi University of Science and Technology

BSc in Data Science and Artificial Intelligence (DS-AI)

Hanoi, Vietnam

Aug 2021 – Aug 2025 (Expected)

TECHNICAL SKILLS

Programming Languages: Python, SQL, Java, C++.

Deep Learning: Pytorch, Lightning, OpenCV, Transformers, Unsloth, Ollama, LlamaIndex.

Deployment and MLOps: Docker, Gradio, WandB, AzureML, AWS.

Backend and Tools: MongoDB, Chroma, Pydantic, FastAPI, Git/Github, Github Actions, Postman, Figma, Jira.

EXPERIENCES

Freelance

Quantitative Researcher

Taiwan

Oct 2023 – Jan 2024

- **Technologies:** Pytorch, Lightning, WandB, Transformers, Polars, Amazon S3
- Design and improve the reward function for the RL based trading environment to mimic real-world scenarios.
- Developing an RL agent combine with current SOTA algorithms such as Transformers, surpass the performance of traditional ensemble models such as XGBoost, Gradient Boosting,... through backtest evaluation.

VIS (Startup)

Part Time AI Researcher

Hanoi, Vietnam

May 2022 – Jul 2023

- **Technologies:** Pytorch, Numpy, Numba, Gynasium, Selenium.
- Researched and implemented 2 deep reinforcement learning algorithms, successfully surpass 95% of the environments in the system by competing with other decision-making agents.
- Developed three gym-based multiplayer environments for researching algorithms, with optimization achieved using Numpy and Numba JIT, reaching the speed upto 1000 games per second.
- Research about Japanese stock market, crawl data from 3 different sources: Minkabu, MarketWatch and YahooJP and cross-check to ensure the credibility of sources.

PROJECTS

Open source contributions | Github

2024

- Ollama (60k stars): Add my Chatbot project to community integrations in the project's documents.
- LlamaIndex (30k stars): Fix a bug in fusion retriever module.

LLM finetuning | Transformers, Unsloth, PEFT, Trl

Apr 2024

- Finetune open sources LLM model such as Llama-3, Mistral to perform multiple task such as:
- Function calling: enable LLM to use external functions and tools provided to solve complex problems.
- Instruction tuned: helps the LLM follow instructions better to more reliably perform complex tasks and instructions.
- Structured output: LLM return structured output such as JSON, XML...based on user's input.

RAG chatbot | Ollama, LlamaIndex, ChromaDB, Gradio, Docker, FastAPI, Ngrok

Mar 2024

- Build a fully local Chatbot with advanced RAG pipeline to give precise answers base on multiple given documents.
- Pull and use any open sources model from Huggingface and Ollama with different quantization.
- Design a simple web interface so users can customize LLM models and input documents to use the chatbot.
- Containerize all the chatbot system and service using Docker for deployment and management.

Stable Diffusion from scratch | Pytorch, Lightning, WandB, Gradio

Feb 2024

- Implemented Latent Diffusion Model (Stable Diffusion) from scratch for education purpose.
- Implemented full training and tracking experiment pipeline, different sampling algorithms such as DDIM, PLMS...
- Build a simple UI with Gradio to visualize the inference process of the model with different sampling methods.