

# Dat Vo Dinh

Hanoi, Vietnam | +84-xxxxxxxx | dinhdat.com | github.com/datvodinh | vodinhdat03@gmail.com

## Technical Skills

---

**Programming Languages:** Python, Javascript, Java, C++.

**Deep Learning:** Pytorch, Lightning, Ray, Ultralytics, Transformers, OpenCV, Spacy.

**MLOps and Deployment:** MLFlow, Kuberaay, WandB, Dagster, Docker, Kubernetes, Helm, FluxCD.

**Fullstack Development:** FastAPI, Jinja, Gradio, NextJS, TailwindCSS, HTML, CSS.

**Developer Tools:** Git, Github Actions, Jenkins, Pydantic, Poetry, Ruff, Prettier.

## Work Experience

---

### AI Engineer

Contracting

Apr 2024 – Aug 2024

Australia

- **Technologies:** Pytorch, Transformers, Ray, Dagster, Qdrant, FastAPI, Docker, Kubernetes, Kuberaay, FluxCD.
- Build a distributed and scalable data pipeline and orchestration using Dagster, capable of ingesting thousands of documents, images, and videos concurrently from S3 storage.
- Develop and serve a robust and autoscale model orchestration with Ray Serve and Kuberaay on Kubernetes cluster, enabling seamless inference across multiple models to process data.
- Design and implement a vector retrieval search engine and GUI with FastAPI and Gradio, enhancing data retrieval with hybrid search and integrating visualization tools to display search results.

### Quantitative Researcher

Contracting

Sep 2023 – Jan 2024

Taiwan

- **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.

### Part Time AI Researcher

VIS Startup

May 2022 – Jul 2023

Hanoi, Vietnam

- **Technologies:** Pytorch, Numpy, Numba, Gynasium, Selenium.
- Research and implement 2 deep reinforcement learning algorithms, successfully surpass 95% of the environments in the system by competing with other decision-making agents.
- Develop 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.

## Personal Projects

---

**Open Source contributions:** Fix bug and contribute integration to Ollama (85k stars) and LlamaIndex (35k stars).

**RAG Chatbot:** Build and serve a chatbot application with advanced RAG pipeline and fully local LLM models.

**Personal Website:** A landing page with personal information and personal technical blog about AI/Deep Learning.

**Stable Diffusion from scratch:** Implemented Latent Diffusion Model from scratch for education purpose.

**LLM finetuning:** Finetune open sources LLM to perform downstream tasks such as function calling, summarization...

## Education

---

### Hanoi University of Science and Technology

BSc in Data Science and Artificial Intelligence - GPA 3.2/4.0

Aug 2021 – Aug 2025 (Expected)

Hanoi, Vietnam