

Ngo Hong Thong

DATA ENGINEERING

[✉ thonghongngo2004@gmail.com](mailto:thonghongngo2004@gmail.com) [☎ 0345223056](tel:0345223056) [🌐 Portfolio](#)
[🌐 LinkedIn](#) [🐙 GitHub](#)

About Me

- Looking for a data internship to apply my skills in data engineering. Highly self-motivated and adept at self-learning, aiming to contribute and develop in a dynamic setting.

Education

Industrial University of Ho Chi Minh City (IUH)

2022 – Now

Major: Data Science - 3.29/4.0

Scholarship: 100% (Year 2)

Experience

AI-Engineer intern

Dec 2024 – Now

Bluebolt-Software/Back-end

- Collecting data from multiple sources, including web scraping and API integration.
- Designing and optimizing APIs to improve website functionality, enhance data retrieval speed, and boost backend performance.
- Integrating tools like Kafka to build a scalable data collection and processing system.

Software Engineer Intern

Mar 2024 – Nov 2024

IX-Health/Front-end

- Developing and optimizing the user interface for a fitness studio application and landing page
- Collaborating on integrating AI models to enhance personalized user experiences.
- Contributing to the overall performance optimization and user experience, making the product more engaging and effective for end-users.
- Focused on building user-friendly features and ensuring compatibility across different platforms and devices.

Projects

Chatbot-Finacial-Langgraph

[Link](#)

- The core architecture using LangGraph to orchestrate a multi-step reasoning workflow.
- The hybrid RAG (Retrieval-Augmented Generation) system that combines: - FAISS vector store retrieval - SQL database integration - Web search capabilities. The intelligent decision-making process that determines the best information source
- Role: AI Engineer
- ****Tools Used****: LangGraph, OpenAI, Google Gemini, FAISS, SentenceTransformer, PostgreSQL, Pydantic, Tavily Search API, Docker

Social-Media-Sentiment-Pipeline

[Link](#)

- Engineered a fully Docker-composed data pipeline for real-time & batch sentiment analysis on social-media data (YouTube, TikTok, Facebook) using ETL & Streaming architectures.
- Orchestrated workflows in Apache Airflow; ingested streaming data via Kafka; performed cleaning & prediction with Spark Structured Streaming and predict with huggingface model; validated outputs with Great Expectations.
- Role: Data Engineer
- ****Tools Used****: Docker Compose, Apache Airflow, Kafka, Spark Structured Streaming, Huggingface, Great Expectations, MinIO, Hive Metastore, MariaDB, Trino, Superset

Tiki E-Commerce

[github](#) 

- Scraped data from Tiki using Apache Airflow.
- Performed ETL (Extract, Transform, Load) processes with Kafka to process the scraped data.
- Streamed processed data into PostgreSQL using Kafka.
- Configured and maintained PostgreSQL to ensure system stability and performance.
- ****Tools Used:**** Docker, Apache Airflow, Kafka, PostgreSQL

Streaming-kafka-spark-redshift

[Link](#) 

- A complete real-time data pipeline that streams data from Apache Kafka, processes it with PySpark, and stores it in Amazon Redshift.

CDC-Postgres-Kafka-ClickHouse

[Link](#) 

- Real-time + batch data platform: FastAPI API layer, PostgreSQL transactional DB, ClickHouse OLAP, Kafka + Debezium for CDC streaming, Airflow for scheduled ETL—all Docker-Compose powered.

Technologies

Database: Clickhouse, PostgreSQL, MySQL, MongoDB, SQLite, Hive

Frameworks/Libraries: Django, FastAPI, Express, ReactJs, Selenium, BeautifulSoup, HuggingFace

Technologies: TensorFlowJs, Firebase, Cloudflare, Socket, Redux, Axios, Tailwinds, Bootstrap, Sass

System: Docker, PySpark, Airflow, Kafka, MinIO, Hive, Trino, Clickhouse

Publications

Building curriculum management software for Industrial Uniniversity of Ho Chi Minh city	YSC-2024
Xay dung cac thuat toan toi uu dua vao quan the cho bai toan phan cong lich giang day	SRW-2024