# Truong Tien Anh

+(84) 982 619 731 | truongtienanh16@gmail.com | linkedin.com/trgtanhh | github.com/trgtanhh04

## Education

**University of Science - VNUHCM** Oct. 2022– Present
*Bachelor of Science in Data Science and Computer Science* *Current GPA: 3.7/4.0*

## Career objective

As a final-year Computer Science student, I have a strong interest in data science, machine learning, and big data. I am eager to find an internship or entry-level position as a Data Engineer or AI Engineer, where I can apply what I have learned to real-world challenges and continue developing my practical skills in a professional environment.

## Skills

**Programming Languages**: Python, SQL, JavaScript, C/C++
**Data Science & ML**: LangChain, TensorFlow, Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn
**Data Engineering**: Apache Hadoop, Spark, Airflow, Kafka, PostgreSQL
**Cloud & DevOps**: AWS (S3, Glue, Redshift, Athena), Docker, Git/GitHub, MongoDB Atlas

## Certification & Blog

**TOEIC**: 810/990
**My Portfolio**: portfolio.com/trgtanhh

## Projects

**End-to-End MovieDB** 2/2025 – 3/2025

- **Key Technologies**: Apache Airflow, Spark, Kafka, HDFS, PostgreSQL, Streamlit, Machine learning
- **Description:**
  - Developed a complete data pipeline to crawl, process, and analyze movie data for powering a recommendation web app.
  - Crawled movie information and transformed it into structured JSON format.
  - Stored raw data in HDFS; used Spark for data cleaning and preliminary analysis.
  - Integrated Kafka and Airflow to orchestrate and trigger ETL workflows.
  - Loaded processed data into PostgreSQL, deployed on Neon for cloud-based access.
  - Built a prediction model for movie pricing and a Streamlit app for personalized recommendations.
- **GitHub:** github.com/trgtanhh04/End-to-End-MovieDB-Data-Engineering

**Mobile AWS Pipeline Engineering** 3/2025 – 4/2025

- **Key Technologies**: AWS (S3, Glue, Athena, Redshift), Docker, Apache Kafka, Spark, PostgreSQL, Machine learning
- **Description:**
  - Built an AWS-based data pipeline to support a mobile recommendation system and price prediction model.
  - Crawled and ingested mobile phone data using Kafka, with event logs stored in PostgreSQL.
  - Processed streaming data in real time with Spark and stored outputs on S3.
  - Automated data cleaning via AWS Glue; queried processed data using Athena and Redshift.
  - Improved data processing speed by 30–50% through optimized infrastructure and parallel processing.
  - Designed dashboards to visualize key insights for decision-making.
- **GitHub:** github.com/trgtanhh04/Mobile-AWS-Pipeline-Engineering

**CV Analysis using LangChain** 5/2025 – 6/2025

- **Key Technologies**: FastAPI, LangChain, OpenAI API, Streamlit, PostgreSQL, FAISS
- **Description:**
  - Built a system to extract and embed structured data from CVs using LLMs.
  - Stored data in PostgreSQL and FAISS for semantic search.
  - Developed search APIs to match candidates by job titles and skills.

    - Deployed backend (FastAPI) and frontend (Streamlit) to the cloud.
- **GitHub:** https://github.com/trgtanhh04/CV-Analysis-using-Langchain.git

**TopDev LLM Recommendation** 5/2025 – 7/2025

- **Team size**: 2
- **Key Technologies**: FastAPI, VueJS, Mistral API, MongoDB, FAISS
- **Description:**
    - Developed a web application to compare user CVs with job descriptions (JDs) using LLM-based analysis.
    - Crawled job data from the TopDev.vn website and stored it in MongoDB for processing and display.
    - Extracted CV content from PDF files, generated vector embeddings, and performed semantic search via FAISS.
    - Used Mistral API to evaluate skill matching, missing skills, and recommend relevant learning paths.
    - Deployed backend (FastAPI) and frontend (VueJS) to Render and Railway for public access.
- **GitHub:** https://github.com/trgtanhh04/TopDev-LLM-Recommendation.git
- **Demo:** topdev-llm-recommendation-frontend.onrender.com