

Trần Đức Trung

trungtdxz@gmail.com | [LinkedIn: niits](#)

Throughout my professional journey, I have held various roles such as Machine Learning Engineer, Data Engineer, and Data Analyst. This has provided me with a comprehensive understanding of data systems. My constant learning and task dedication make me ideal for high-performing teams.

EDUCATION

University of Engineering and Technology, VNU

Hanoi, Vietnam

Bachelor of Computer Science

June 2017 - July 2021

- **GPA:** 3.56 / 4.0
- **Honors:** Certificate of Excellent Thesis Defense

SKILLS

Knowledge: Deep Learning, Server-side Programming, Cloud Computing, CI/CD

Programming languages: Python, SQL, JavaScript

Tools and Technologies: PyTorch, ONNX, OpenVINO toolkit, Airflow, Docker, Kubernetes

Soft Skills: Communication, Teamwork, Time management, Problem-solving, Mentorship

WORK EXPERIENCE

Sun* Vietnam

Hà Nội, Việt Nam

Machine Learning Engineer, AI Research Department, R&D Unit

Feb 2021 – Present

- Participate in consulting, and implement AI-related solutions in both internal and customer projects
- Selected Project Experience:
 - **Migrating Sun*AI's PoC system**
 - Standardized the deployment process of Deep Learning models using BentoML
 - Developed a new model deployment system using Hasura and Knative in a serverless form to reduce the required resources by up to 80% compared to the original system
 - **OCR-related projects:** [Converting images to LaTeX - Extracting information from CV]
 - Participated in fine-tuning and training high-accuracy deep learning models
 - Optimized the models and chose a suitable deployment plan, reducing the model's latency up to 80% compared to the original version
 - **LLM-related projects:**
 - Implemented Retrieval-augmented generation systems for diverse domains
 - Participated in fine-tuning the LLM model using popular methods such as PEFT, Lora, and QLoRA
 - Survey solutions can be used to handle complex issues, such as personalized text generation
 - Applied quantization methods to operate LLMs in production using the Triton Inference Server
 - **Data platform projects:**
 - Consulted with clients and developed cloud-based data warehouses that were tailored to their specific requirements and budget constraints by using AWS Athena and GCP BigQuery
 - Participated in data modeling and analysis for retail, car rental, and enterprise systems
 - Built and maintained data pipelines to automate data collection and processing

CERTIFICATIONS

- **Microsoft Certified: Azure Data Scientist Associate**, issued by Microsoft in Feb 2024