

NGUYEN TRAN NGUYEN KHOI

INTERN AI ENGINEER

Luy Ban Bich, Hoa Thanh Ward, Tan Phu District, HCM City | +84 352 555 510
github.com/khoidubai | ntnkhoidubai@gmail.com

SUMMARY

AI-focused Data Science student with hands-on experience in computer vision, natural language processing, and speech systems including speaker diarization. Developed AI projects using PyTorch, with interest in scalable and production-ready systems. Seeking an AI Intern position to contribute to impactful projects and grow within a professional AI environment.

EDUCATION

Ho Chi Minh University of Foreign Languages - Information Technology Jul 2023 - Present
Data Science
Cumulative GPA: 3.78/4.0
TOEIC: ~665/990

EXPERIENCE

AI Engineer Intern — SEAL Code Team, VinCity CLB (Startup) Dec 2025 - Mar 2026

- Speech & Language Processing: Researched and deployed State-of-the-art architectures for Automatic Speech Recognition (ASR) and Named Entity Recognition (NER). Leveraged Librosa for audio feature extraction and Sherpa-onnx for optimizing edge-device inference.
- Evaluation & Optimization: Built an automated performance evaluation pipeline using Jiwer. Fine-tuned models to minimize Word Error Rate (WER), ensuring high accuracy for the MeetingTrace system across diverse Vietnamese accents.

KEY PROJECTS

Research Project: EREL - Entity Recognition and Linking Algorithm Jan 2025 - Dec 2025

- Core Research: Addressed complex Ambiguity Disambiguation challenges within Vietnamese text. Studied Coreference Resolution mechanisms to accurately map entity mentions to a global knowledge base.
- Advanced Data Pipeline: Engineered a large-scale data preprocessing pipeline for VLSP datasets. Conducted text normalization, tokenization, and feature engineering for entity recognition tasks (PER, LOC, ORG).
- Experimentation: Evaluated algorithm performance on complex datasets, focusing on F1-score, processing latency, and model scalability for real-world NLP applications.

"Ly Thai To" Election Chatbot (City-level Project) Mar 2026

Technologies: Python, Llama-cpp, Vector Search, SQLite

- LLM Optimization: Deployed Large Language Models (LLMs) using Quantization techniques via Llama-cpp, enabling high-performance inference on local hardware with minimal latency.
- Retrieval Architecture: Designed a candidate information retrieval system based on Vector Space Modeling (VSM). Integrated LLMs with structured SQLite databases to mitigate AI hallucination and ensure data integrity.

HUFLIT Information Retrieval System

Dec 2025 - Apr 2026

Technologies: Python, TF-IDF, Cosine Similarity, Rocchio Algorithm

- Search Algorithms: Developed a full-text search engine from scratch. Applied TF-IDF weighting schemes and utilized Cosine Similarity for high-precision document ranking.
- Relevance Feedback: Implemented the Rocchio Algorithm to automatically refine search queries based on user interactions, significantly improving the system's Precision and Recall metrics.

Real-time Delivery Tracking System

Dec 2024 - Apr 2025

Technologies: Apache Kafka, .NET Core, SQL Server, JavaScript

- Distributed Systems: Architected an event-driven Microservices system using Apache Kafka. Built high-throughput data streams between Producers and Consumers to handle thousands of order events per second in real-time.
- Asynchronous Processing: Optimized system responsiveness by implementing asynchronous event processing for status updates and notifications, maintaining data consistency across high-load environments.

SKILLS AND ABILITIES

Technical Skills

- **AI/NLP:** ASR, NER, IR, LLMs, Vector Space Models
- **Speech & Data:** Audio Processing (Librosa), Data Augmentation, Evaluation (WER, F1), Text Normalization
- **Programming:** Python, C# (.NET Core), SQL (SQLite, SQL Server)
- **Systems:** Apache Kafka
- **Frameworks:** PyTorch, Scikit-learn, Hugging Face, Llama-cpp, Sherpa-onnx
- **Tools:** Git, Conda, Jupyter, Pandas, NumPy

Core Strengths

- Research & experiment design (paper analysis, model evaluation)
- Model optimization for low-latency & edge deployment
- Building scalable AI systems from research to production

Key Soft Skills

- Analytical thinking & problem-solving
- Effective collaboration in technical teams

NOTABLE AWARDS

- **HUFLIT Talent Scholarship** (Top 1, Sophomore Year) – Awarded to the student with the highest academic achievement in the entire Information Technology department.
 - **HUFLIT Effort Scholarship** (Top 4, Freshman Year) – Awarded to students demonstrating outstanding academic improvement and determination.
-