





MINH N. TA

Prospective MSc. Student

 tnminh.com
 +84 942 942 968
 Hanoi, Vietnam

 minh@tnminh.com
 ngocminhta
 ngocminh-ta

OBJECTIVES

Utilizing the skills and knowledge acquired during my university years and related work experiences, I want to pursue a career as a technopreneur specializing in AI, especially in Natural Language Processing. Pursuing higher education helps me get more knowledge and practical experience, which gives me an advancement in my career.

SKILLS

Languages: Vietnamese (Native), English (Advanced).

Technical: Statistics, Optimization & algorithms, Big Data Architecture, Machine Learning, Deep Learning, Natural Language Processing.

Tech-stack: Python, Java, Scala, SQL, Hadoop, Spark, Kubernetes, PyTorch, Langchain.

Interpersonal: Leadership, Logical & Critical thinking, Teamwork, Problem-solving.

EDUCATION

Sep. 2021 - Now	Hanoi University of Science and Technology Bachelor of Science in Data Science, specialized in Artificial Intelligence	3.52/4.0
May - Jun. 2024	Mohamed bin Zayed University of Artificial Intelligence (MBZUAI) Exchange research resident of NLP track with the acceptance rate of 4%	Exchange

RELEVANT EXPERIENCE

May 2024 - Now	Research internship resident	Mohamed bin Zayed University of Artificial Intelligence I joined the Research Internship program and currently working on some projects of <i>Machine-generated text detection</i> under the supervision of professor Preslav Nakov.
Aug. 2023 - Now	Teaching Assistant	School of ICT, Hanoi Univ. of Science and Technology I'm a teaching assistant in some undergraduate courses of Database, Deep learning, and NLP. I got certificates of recognition for outstanding performance in this activity.
Apr. 2023 - Now	Data & AI Engineer	Viettel Group Headquarters I joined Viettel Digital Talent program (selection rate of 10%), top 2 best projects of IT division. I'm the only selected intern in Data & AI engineering to be promoted to work at the headquarters as an official employee. <ul style="list-style-type: none">• Building new components and features for data science and data lakehouse project.• Investigate to develop AI components in new big data platforms and products.
Jun. 2023 - Now	Lecturer	Datapot Analytics Group & Foreign Trade University <ul style="list-style-type: none">• Jointly contribute to building curriculum of Data Science Foundation course.• Teach all practical parts of courses in Data Science in Economics.
Feb. 2022 - Now	Research Assistant	The Intl. Research Center for AI - Foundation Models Lab My research mainly focus on applying state-of-the-art LLM techniques to handle NLP tasks in Vietnamese.

HONORS AND AWARDS

2024	Best Presentation Award in MBZUAI UGRIP Program
2024	Gold Prize in HUST-SolCT Creative Idea Challenge 2024.

2023 – 2024	Certificates of Recognition from the dean for outstanding performance while serving as a teaching assistant.
2022 – 2024	Certificates of recognition from University's president for excellent academic performance and remarkable contributions to students' activities.
2022	University's prize for Student with Five Good Merits. (top 1% students)
2022	Champion of Data Science Talent Competition 2022.
2022	Bronze Prize in AOF Young Startup 2022.

CERTIFICATIONS

2023	IBM Data Engineering Professional Certificate.
2023	Google Data Analytics Professional Certificate.
2020	Intermediate Tensorflow in Google Cloud.

PUBLICATIONS

2024	LLM-DetectAlve: a Tool for Fine-Grained Machine-Generated Text Detection. PDF Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: System Demonstrations (<i>Co-first author</i>).
------	---

REMARKABLE PROJECTS

Sep. 2024 – Now **Sensitive data detection** **Private project for Enterprise**

Project description	This projects submitted a paper to EMNLP 2024 - System Demo. as a classification tool of an input text into one of the four types: human-written, machine-generated, machine-polished and machine-humanized text.
Used technologies	PyTorch, RoBERTa, Long-Alpaca, RAG.
Team size	1
Progress	<ul style="list-style-type: none"> Generating a simulated dataset for sensitive data. Research to implement in two ways: Detect as pattern-like and simple sensitive data, and detect as passkey retrieval.

May – Aug. 2024 **Machine-generated text detection** **Research project**

Project description	This projects submitted a paper to EMNLP 2024 - System Demo. as a classification tool of an input text into one of the four types: human-written, machine-generated, machine-polished and machine-humanized text.
Used technologies	PyTorch, Llama-3, RoBERTa, Doman-adversarial NN, GAN.
Team size	6
My involvement	<ul style="list-style-type: none"> Completed building dataset and model for training process. Research to implement in two ways: 1, Using RoBERTa with DANN architecture for simple detector; 2, Using Llama-3 with the GAN architecture for explainable detector.

Dec. 2023
– Mar. 2024

Integrated LLM Coding Partner for Data Lakehouse

Private project for Enterprise

Project description	This projects aim to create an assistant which can: i , give informations about data lakehouse and user manual; ii , generate SQL, Python and Scala code to extracting information from data lakehouse.
Used technologies	Python, Tensorflow, Transformer, QLoRA, SparkAI, RAG
Team size	1
Results	<ul style="list-style-type: none">• Complete training models for code generation and question-answering in English.• Complete the extensions for Visual Studio Code and Jet-Brains's IDE along with the library for LLM integration with connection to data lakehouse.

Oct. – Dec. 2023

LongLoRA Technique and Vietnamese LongAlpaca Dataset

Research Project

Project description	This projects investigates in the technique to make the LLM have the capacity of understanding a whole long book; along with create a new dataset for training using this technique
Used technologies	Python, Tensorflow, Transformer, LoRA
Team size	1
Results	<ul style="list-style-type: none">• Successfully implement LongLoRA Techniques for LLM.• Successfully proposed a method for collecting Vietnamese LongAlpaca Dataset, building in progress.

EXTRACURRICULAR EXPERIENCES

Nov. 2023 – Now

Deputy Secretary for Study, Scientific Research & Orientation

HCM Youth Union

I am the only student on the HUST-SOICT Ho Chi Minh Youth Union Executive Committee in charge of organizing student activities, particularly assisting students in engaging more in studying, scientific research, and acquiring more valuable skills for their future career paths.

Aug. 2022
– Aug. 2024

President & Leader

Google Developer Student Club - HUST

Sponsored by Google for Developers, GDSC-HUST aims to help students work with Google's latest technology. Besides, GDSC-HUST has been recognized as the most outstanding GDSC in South-East Asia.

REFERENCES

Dr. Sang Viet Dinh

Deputy Director, The Intl. Research Center for AI,
Head of Foundation Models Lab,
Hanoi University of Science and Technology
sangdv@soict.hust.edu.vn

Prof. Preslav Nakov, PhD

Head of Natural Language Processing department,
and Professor of Natural Language Processing
Mohamed bin Zayed University of AI
preslav.nakov@mbzuai.ac.ae

Mr. Thanh Chi Nguyen

Head of Customer 360 Department, VPBank
Former Head of Technology R&D Department,
IT Division, Viettel Group HQ
thanhnc29@vpbank.com.vn

Dr. Tuan Cong Truong

Faculty advisor
Director of Innovation and Entrepreneurship Center,
Hanoi University of Science and Technology
tuan.truongcong@hust.edu.vn