# MINH N. TA
## Data & AI Research Engineer

*Last updated: February 22, 2025*

🌐 tnminh.com     ✉ minh@tnminh.com
📞 +84 942 942 968     ⊙ ngocminhta
📍 Hanoi, Vietnam     in 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, esspecially 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, NLP, LLM.

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

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

## EDUCATION

| | | |
|---|---|---|
| 2025 – 2027 | **Mohamed bin Zayed University of Artificial Intelligence (MBZUAI)**<br>Master of Science in Natural Language Processing | |
| 2021 – 2025 | **Hanoi University of Science and Technology**<br>Bachelor's in Data Science, specialized in Artificial Intelligence<br>**Thesis:** GenSpy: A Tool for Fine-Grained AI-Generated Code and Text Detection | **Grade: Excellent** |

## RELEVANT EXPERIENCE

| | | |
|---|---|---|
| May 2024 – Now | **Research Resident** | **Mohamed bin Zayed University of Artificial Intelligence** |

I joined the Research Internship program and currently working on some projects of *Machine-generated text detection* under the supervision of professor Preslav Nakov.

| | | |
|---|---|---|
| Aug. 2023 – Now | **Teaching Assistant** | **School of ICT, Hanoi University 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.

| | | |
|---|---|---|
| Feb. 2022 – Now | **Research Assistant** | **The International 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.

| | | |
|---|---|---|
| Apr. 2023 – Mar. 2025 *(on leave)* | **Data & AI Engineer** | **Viettel Group Headquarters** |

I joined Viettel Digital Talent program (selection rate: 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.
- 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 – Feb. 2025 | **Tutor** | **Datapot Analytics Group & Foreign Trade University** |

- Jointly contribute to building curriculum of Data Science Foundation course.
- Teach all practical parts of courses in Data Science in Economics.

## HONORS AND AWARDS

| | |
|---|---|
| 2024 | Certificate of Merit from Hanoi's Youth Union. |
| 2024 | Provincial-level prize for Student with Five Good Merits (Academic Performance, Ethics, Physical Fitness, Volunteerism, and Integration). (top 0.1% students in Hanoi) |
| 2024 | Best Presentation Award in MBZUAI UGRIP Program |
| 2024 | Gold Prize in HUST-SoICT 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. |
|---|---|
| 2023 | Certificate of recognition as an excellence mentor during the Digital Talent Development program of Google and Vietnam National Innovation Center. |
| 2022 | University's prize for Student with Five Good Merits. (top 1% students) |
| 2022 | Champion of Data Science Talent Competition 2022. |
| 2022 | Silver Prize in HUST-SoICT Creative Idea Challenge 2022. |
| 2022 | Bronze Prize in AOF Young Startup 2022. |

## CERTIFICATIONS

| 2024 | University of Illinois Urbana-Champaign's Certificate of Strategic Technology Management. |
|---|---|
| 2023 | IBM Data Engineering Professional Certificate. |
| 2023 | Google Data Analytics Professional Certificate. |
| 2020 | Intermediate Tensorflow in Google Cloud. |

## PUBLICATIONS

For an updated list, please refer to my Google Scholar profile.

| 2025 | **GenAI Content Detection Task 1: English and Multilingual Machine-Generated Text Detection: AI vs. Human**    **PDF**<br>Proceedings of the 1st Workshop on GenAI Content Detection, COLING 2025. |
|---|---|
| 2024 | **LLM-DetectAIve: a Tool for Fine-Grained Machine-Generated Text Detection.**    **PDF**<br>Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: System Demonstrations (*Co-first author*). |

## REMARKABLE PROJECTS

Oct. – Feb. 2024 **Sensitive data detection** Private project for Enterprise

| Project description | This project aims to create a tool to detect some sensitive data defined in Decree No. 13/2023/ND-CP of the Government on Personal Data Protection such as basic indentification information, and some specific sensitive data such as political or religious attitude, etc. in data streaming flow or big batch of data files in Viettel Lakehouse Platform. |
|---|---|
| Used technologies | PyTorch, LLM, Passkey retrieval, Kubernetes. |
| Team size | 1 |
| Results | • Generated a simulated dataset for sensitive data using the predefined contents in the decree.<br>• Developed a simple model to detect sensitive data based on patterns and handled some simple cases and a model with passkey retrieval techniques to identify some small sensitive data in a very long input.<br>• Deployed the model to Kubernetes and integrated it to a data lakehouse system. |

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 | • Completed building dataset and model for training process.<br>• 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** | • Complete training models for code generation and question-answering in English.<br>• Complete the extensions for Visual Studio Code and JetBrains'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** | • Sucessfully implement LongLoRA Techniques for LLM.<br>• Successfully proposed a method for collecting Vietnamese LongAlpaca Dataset, building in progress. | |

## EXTRACURRICULAR EXPERIENCES

Nov. 2023 – Now  **Deputy Secretary for Study, Scientific Research & Orientation**  Ho Chi Minh's 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 Group on Campus - Hanoi University of Science and Technology

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

## REFERENCES

**Prof. Preslav Nakov, PhD.**
*Head of Natural Language Processing department,*
*Professor of Natural Language Processing*
Mohamed bin Zayed University of AI
preslav.nakov@mbzuai.ac.ae

**Dr. Sang Viet Dinh**
*Vice Dean of School of ICT,*
*Director of The Intl. Research Center for AI*
Hanoi University of Science and Technology
sangdv@soict.hust.edu.vn

**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**
*Director of Innovation and Entrepreneurship Center,*
*Secretary of University's Youth Union*
Hanoi University of Science and Technology
tuan.truongcong@hust.edu.vn