

Education

Ha Noi University of Science and Technology
Bachelor of Science in Computer Science (Excellent Degree)

GPA: 3.64/4.00

Relevant Coursework

- Data Structures
- Caculus
- Discrete Math
- Artificial Intelligence
- Algorithms Analysis
- Linear Algebra
- Numerical Optimization
- Computer Architecture

Publications

Adaptive Prompting for Continual Relation Extraction: Within-Task Variance Perspective

- **Authors:** Minh Le*, **Tien Ngoc Luu***, An Nguyen The*, Thanh-Thien Le, Trang Nguyen, Tung Thanh Nguyen, Linh Ngo Van, Thien Huu Nguyen
- **Conference:** 39th AAAI Conference on Artificial Intelligence (AAAI 2025) - **Oral Presentation**

Optimizing Answer Generator in Vietnamese Legal Question Answering Systems Using Language Models

- **Authors:** Huong Le*, **Tien Ngoc Luu***, Thanh Nguyen*, Tuan Dao, Sang Dinh
- **Journal:** ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP 2024)

* Equal contribution

Research Experience

Data Science Laboratory - SoICT | Undergrad Research Assistant

October 2022 – May 2024

- Advisor: **Dr. Linh Ngo Van**
- Main research topics: **Continual Learning, Natural Language Processing**
- Knowledge Gained: Foundations of Machine Learning, Deep Learning, LLM, Chatbot, Relation Extraction,...
- Improve performance Continual Relation Extraction.

Industrial Experience

DNSE Company | Data Science Intern

June 2023 – September 2023

- Developed a chatbot system that can answer users' questions in the field of stock.
- Learned about technologies and frameworks used for projects such as Langchain, Elastic Search,...
- Trained the **Bartpho** model for the problem of combining multiple summaries in the field of stock, with the aim of minimizing the repetition of too much information with the same semantics.

CMC ATI Company | NLP Research Engineer

March 2024 – October 2024

- Build an internal chatbot system to serve employees.
- Develop NER models to extract key information from official documents, such as citizen identification cards and driver's licenses.
- Developing a post-processing model for Automatic Speech Recognition (ASR) to improve accuracy in speech-to-text tasks.

PREP Company | AI Engineer

October 2024 – Present

- Developing and enhancing a comprehensive scoring system for English proficiency exams, such as IELTS, TOEIC,....
- Developing a virtual classroom system that allows students to practice and prepare for international English proficiency exams, providing them with an interactive and effective learning environment.”

Personal Projects

Legal Question Answering Model for Vietnamese Language | Python, Transformers

- Developed an advanced **Vietnamese Legal Chatbot**, enabling users to easily access and retrieve accurate legal information.
- Curated a high-quality legal question-answering (QA) dataset by collecting data from reputable sources such as *DichVuCong* and *ThuVienPhapLuat*.
- Leveraged two powerful models, **Bartpho** and **Llama2 7B**, pretraining them on a 5GB legal corpus and fine-tuning them with 430,000 instruction examples and 200,000 legal QA pairs for optimal performance.

Zero-shot Named Entity Recognition for Vietnamese General Task | *Python, Transformers*

- Developed a zero-shot Named Entity Recognition model for the Vietnamese language using the **GliNER** method.
- Tested and validated the model on real-world Vietnamese datasets, demonstrating its effectiveness in recognizing entities such as names, locations, organizations, and other relevant categories.

IT Job Analysis | *Scrapy, Python*

- Collected Vietnam IT job recruitment posts from websites such as LinkedIn, CareerBuilder, Devwork, Jobgo, ITviec, ...
- Conducted a detailed analysis of information within IT job postings to extract valuable insights.
- Built a model capable of providing recommendations for IT jobs based on individual competencies.

Build a model for Intent classification and NER | *Python*

- Train a model with the base model **PhoBert**, using the **JointIDSF** architecture to improve performance on the test set by an average of **94% accuracy**.

Testing Machine Learning Models | *Python, Scikit-learn, TensorFlow*

- Conducted a comprehensive assessment of model accuracy in handwritten digit recognition and Vietnamese news classification tasks.
- Experimented with various machine learning models, including but not limited to KMeans, KNN, and Support Vector Machines, to refine and optimize the performance of recognition and classification tasks.

Population management system in the district | *Java, IntelliJ, JavaFX*

- Developing a comprehensive system dedicated to the management of household registration books and contributions.
- Leveraging JavaFX to craft an intuitive Graphical User Interface (GUI) that facilitates and enhances user interactions.

NER extension | *Python, JavaScript, HTML, CSS*

- Created a **Chrome** extension utilizing the **DistilBert** model for named entity recognition.

Technical Skills

Programming: Python, C/C++

Developer Tools: VS Code, Git, Jupyter Notebooks, Docker

Frameworks: Pytorch, Hugging Face Transformers, Tensorflows, Scikit-learn

Version Control: Git, GitHub, GitLab

Operation System: Windows, Linux, MacOS

Writing: LaTeX

Achievements

Top 0.01% in Vietnam National Exam

2020

29.25 - A00 combination in Vietnam National Examination

Top 2 in the Leaderboard for the Spoken Language Understanding track at the SoICT Hackathon

2023