

Bui Nguyen Kim Hai

HF/wnkh | github/whynotkimhari | htlulem185@gmail.com | +36-70-40-168-48 | linkedin/in/wnkh

Education

BSc Computer Science – ELTE, Hungary (CGPA: 4.9/5.0) Sep 2023 – Expected Jul 2026
BSc Computer Science – UIT, VNUHCM, Vietnam (CGPA: 8.8/10) Sep 2022 – Jul 2023

Experience

Data Engineer, VietMed Lab – Remote July 2024 – Present

- Participate in a team of Vietnamese AI/ML researchers around the world in medical NLP topics such as **Intent Detection**, **Multilingual Translation**.
- Responsible for figuring out the suitable guild line for processing data by working with API of state-of-the-art LLM models. Using multiple prompting techniques like **chain-of-thoughts**, **few-shot prompting**, **contextual prompting**, **instruction prompting** to create datasets.
- Speed up the dataset creation process by leveraging the **paralleled querying** to LLM models.

Undergraduate Researcher, Modelling Lab – Hybrid Sep 2024 – Present

- Working unofficially under supervision of Dr. Gergő Bognár in Financial Prediction.
- Develop a novel neural network using Attention mechanism as the core to enhance the predicting results. Apply fine-tuning methods for hyperparameters optimization to achieve the high performance.

Projects

Transformers and Multi-features Time2Vec for Financial Prediction Mar 2024 – Present

Authors: Bui Nguyen Kim Hai, Nguyen Duy Chien

- Developed an innovative neural network model by integrating the Encoder of Transformer architecture with Time2Vec, inspired by traders' techniques of grouping high correlated stocks, to achieve high accuracy and competitive performance. Using GridSearch for hyperparameters tuning.

MultiMed Sep 2024 – Present

- Construct an extended version of a dataset (*leduckhai/MultiMed*) for the medical translation task. Working with Gemini API and using prompting techniques to create the initial translations.
- Build the pipeline to transform the initial dataset in the way that annotators can easily work on.

VietMed-ID Jul 2024 – Present

- Create a small instance of VietMed dataset (*leduckhai/VietMed*) specific for the Intent Detection task.
- Build the pipeline to transform the initial dataset into the result. Working with LLM APIs as a developer, using multiple prompting techniques to achieve the classification results. Working on processed dataset as an annotator to check the result from the LLM.

Technologies

Languages: Python, JS/HTML/CSS, Java, SQL, PHP, C/C++/C#, LaTeX, Clean

Software: Git, Hungging Face, Google Colab, Kaggle, Overleaf, Postman

Architecture: RESTful API

Awards

Third Prize, ELTE TDK 2024 Spring Conference April 2024

Stipendium Hungaricum Scholarship Jul 2023

Consolation Prize, Vietnamese National Student Mathematical Olympiad 2023 April 2023

Second Prize in Math - Vietnamese Excellent Student, City level April 2022

Certifications

AWS Educate:

- **Cloud:** Introduction to Cloud 101
- **Machine Learning (ML):** ML Foundations

Cisco: IoT Fundamentals: Big Data & Analytics

CS50: CS50X, CS50AI, CS50P

HackerRank:

- **SQL:** SQL Basic, SQL Intermediate, SQL Advanced
- **Problem Solving:** Problem Solving Basic
- **JavaScript (JS):** JS Basic, JS Intermediate
- **Java:** Java Basic
- **Role:** SWE Intern Test, SWE Test

Kaggle:

- **Course:** 5-Day Gen AI Intensive
- **Deep Learning (DL):** Intro to DL
- **Python:** Python, Pandas
- **Machine Learning (ML):** Intro to ML, Intermediate ML, ML Explainability
- **SQL:** Intro to SQL, Advanced SQL

DeepLearning.AI:

- **Short Course:** LLMs as Operating Systems: Agent Memory by Letta