Bui Nguyen Kim Hai

huggingface/wnkh | github/whynotkimhari | buingkimhai@gmail.com | +36 704016848 | linkedin/wnkh

Experience

Backend Developer, Butopêa   Jan 2025 − Present
- Implementing unit and end-to-end tests to maintain module reliability. Developing features, integrating third parties such as OTPSimplePay, OpenCart, and performing error analysis to enhance workflow performance and optimize operations to manage **more than 20.000** products.

Research Assistant, VietMed Team   Jun 2024 − Present
- Designing and building a synthetic/augmented data generation and processing pipeline. Conducting validation and verification of generated datasets to ensure quality and reliability.
- Create the large medical multilingual speech-to-text dataset with **290,000 samples** across **5 languages** consist of Vietnamese, English, German, French, Traditional Chinese, and Simplified Chinese.
- Enrich the multilingual medical VLM CoT dataset of brain MRI images by generating sub-CoTs between each chain point, allow model to have a deeper understanding about the image and the task to increase the accuracy.

Research Intern, Modeling Lab   Sep 2024 − Present
- Developing a neural network combining Transformer Encoder architecture with Time2Vec embeddings.
- Proposing a novel approach for clustering highly correlated stocks to enhance financial prediction models. The result outperforms traditional methods by **approximately 15%**.

Research Intern, Neural Information Processing Group   Jan 2025 − May 2025
- Integrating state-of-the-art 6D pose estimation methods (MegaPose and GigaPose) for hand manipulation tasks.
- Topics: 3D Digital Twin & Human & Applications.

Education

S.B. Computer Science − ELTE, Hungary (CGPA: 4.9/5.0)   Sep 2023 − Jul 2026
S.B. Computer Science − UIT, VNUHCM, Vietnam (CGPA: 8.8/10)   Sep 2022 − Jul 2023

Publications & Preprints

Transformer Encoder and Multi-features Time2Vec for Financial Prediction   [EUSIPCO 2025]
*Authors: **Nguyen Kim Hai Bui**, N. D. Chien, P. Kovács, G. Bognár*

MultiMed-ST: Large-scale Many-to-many Multilingual Medical Speech Translation   [Preprint]
*Authors: K. Le-Duc, T. Tran, B. P. Tat, **Nguyen Kim Hai Bui**, Q. D. Anh, H.-P. Tran, T. T. Nguyen, L. Nguyen,*
T. M. Phan, T. T. P. Tran, C. Ngo, K. X. Nguyen, T. Nguyen-Tang

Model-Agnostic Capabilities of Large Audio Language Models   [Preprint]
*Authors: K. Le-Duc, T. Tran, B. P. Tat, **Nguyen Kim Hai Bui**, L. Phan, C. Ngo, K. X. Nguyen, T. Nguyen-Tang*

Activities

Giving oral lecture in Sequential & Neural Network Learning session at EUSIPCO 2025   Sep 2025.

Awards

University Excellence Fund for attending EUSIPCO 2025 conference   Sep 2025
Special Prize, Hungarian National Student Research Conference OTDK 37   Apr 2025
First Prize, ELTE TDK 2024 Fall Conference   Dec 2024
Third Prize, ELTE TDK 2024 Spring Conference   Apr 2024
Stipendium Hungaricum Scholarship   Jul 2023
Consolation Prize, Vietnamese National Undergraduate Mathematical Olympiad 2023   Apr 2023
Second Prize in Math − Vietnamese Excellent Student, City level   Apr 2022
Bronze Medal in Math − Vietnamese 30/4 Olympiad, Region level   Apr 2021

Technologies

Languages: Python, JS/HTML/CSS, Java, SQL, PL/SQL, PHP, C/C++/C#, LaTeX, Clean, R, GraphQL
Software: Git, Hungging Face, Google Colab, Kaggle, Overleaf, Jira, Postman, K3s
Architectures: REST API, MVC, ORM, Monolithic

Courses & Certifications

AWS Educate: Introduction to Cloud 101, Machine Learning Foundation
Cisco: IoT Fundamentals: Big Data & Analytics
CS50: CS50X, CS50AI, CS50P
DeepLearning.AI: LLMs as Operating Systems: Agent Memory by Letta
HackerRank: SQL Advanced, JavaScript Intermediate, SWE Test
HuggingFace: LLM Course
Kaggle: Intro to Deep Learning, Intermediate Machine Learning, Machine Learning Explainability, Advanced SQL