Le Khanh Duy

duyvt6663 | duy.le1201@hcmut.edu.vn | (+84) 901136210

Education

Ho Chi Minh City University of Technology **HCMC, VietNam**

Bachelor of Computer Science - *Senior year* - GPA: 3.8/4.0 Oct. 2020 - Present

Awards: 3rd Prize in *Vietnam Mathematical Olympiad* 2019; 1st Prize in *Blockchain Computing and Mathematics contest* VIASM 2022

Coursework: Data Mining, Algorithms and Data Structure, Computer Vision, Database System, Calculus, Linear Algebra, Statistics and Probability, Natural Language Processing, Machine Learning.

Professional Experience

KAIST INTERACTION LAB **Daejeon, Korea**

Visiting Student Researcher Jun. 2023 - Oct. 2023

- Proposed and implemented a multi-staged pipeline for data exploration focusing on text highlights, recommendations, and visualization using **Large Language Models** (LLM)
- Grounded LLM's visual reasoning on data using **Text-to-SQL** and **Semantic Similarity**. Recommendations were ranked with **pair-wise ranking prompting**
- Achieved **80.5%** match accuracy; **98%** relevance and **4.07/5** interestingness for recommendation

Publications

[p.1] DataDive: Supporting Readers' Contextualization of Statistical Statements with Data Exploration
Hyunwoo Kim, **Khanh Duy Le**, Gionnieve Lim, Dae Hyun Kim, Yoo Jin Hong, Juho Kim
IUI 2024: Annual Conference on Intelligent User Interfaces (to appear)

[p.2] Semi-supervised Semantic Segmentation using Redesigned Self-Training for White Blood Cell
Quoc-Vinh Luu, **Khanh-Duy Le**, et.al.
ICISN 2024: International Conference on Intelligent Systems & Networks (to appear)

Projects

Kalapa Medical QA | *contest* Oct. 2023 - Nov. 2023

- Developed a Vietnamese Medical QA system leveraging a medical corpus comprising 600 articles
- Integrated **Retrieval Augmented Generation** (RAG) by ensembling **SimeCSE_Vietnamese** and **BM25**, with a layer of keyword search using **Ahocorasick** on top
- Finetuned a 3B reader **Vietcunna** with **LoRA** on **GPT-4** generated data
- Achieved **59.6**% accuracy, **0.7** score, **Top 1** for 2 weeks

Hierarchical Stock Clustering | *research* Mar. 2023 - Apr. 2023

- Performed hierarchical clustering on S&P 500 data using **Hierarchical Agglomerative** Linkage algorithms and **Directed Bubble Hierarchical Tree** (DBHT), and benchmarked them against GICS
- Tailored DBHT with ward linkage to improve its Adjusted Rand Index by **20%**
- Constructed a portfolio out of S&P 500 by applying **Hierarchical Risk Parity** on the clusters

Technical Skills

English Proficiency	IELTS 7.5 Overall
Programming Languages	Python, C++, C, SQL, JavaScript
Tech/Tools	Langchain, OpenAI, Pandas, Numpy, Transformers, PyTorch, Django