

## Publications

- 1 Vu, Hieu M and Tan Minh Nguyen (2025). "Angular Steering: Behavior Control via Rotation in Activation Space". In: *Advances in Neural Information Processing Systems. Spotlight* (**top 3.5%**).
- 2 Son, Nguyen Hong, Hieu M Vu, Tuan-Anh D Nguyen, and Minh-Tien Nguyen (2022). "Jointly Learning Span Extraction and Sequence Labeling for Information Extraction from Business Documents". In: *2022 International Joint Conference on Neural Networks (IJCNN). Oral*. IEEE, pp. 1–8.
- 3 Nguyen, Tuan-Anh D, Hieu M Vu, Nguyen Hong Son, and Minh-Tien Nguyen (2021). "A Span Extraction Approach for Information Extraction on Visually-Rich Documents". In: *Document Analysis and Recognition-ICDAR 2021 Workshops: Lausanne, Switzerland, September 5–10, 2021, Proceedings, Part II 16. Best Paper Award*. Springer, pp. 353–363.

## Education

<b>Bachelor Degree, Computer Science (Honours Programme)</b> <i>UET - Vietnam National University, Hanoi</i>	<b>2020</b>
• Highest Ranking Graduate.	GPA: 3.83/4.00 (Rank: 1st/700+)

## Honours and Awards

<b>Best Paper Award</b> <i>ICDAR 2021, Workshop on Document Images and Language</i> Paper title: A Span Extraction Approach for Information Extraction on Visually-Rich Documents Accepted for oral presentation and awarded the Best Paper Award at Workshop on Document Images and Language, ICDAR 2021.	<b>2021</b>
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------

<b>Certificate of Highest Ranking Graduate</b> <i>UET - Vietnam National University</i> Awarded to students graduate with the highest GPA amongst the graduating class.	<b>2020</b>
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------

<b>Certificate of Merit for Excellent Graduation</b> <i>Vietnam National University</i> Awarded by the President of Vietnam National University to students with excellent academic performance and level of conduct during a 4-year undergraduate programme.	<b>2020</b>
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------

<b>Certificate of Excellent Thesis Defence</b> <i>UET - Vietnam National University</i> Awarded to the best thesis of the Undergraduate Thesis Defence Committee. Thesis title: A Layout-aware key-value relation predicting model for document images.	<b>2020</b>
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------

<b>Top 4 Zalo AI Challenge 2018 - Voice Track (Individual participant)</b> <i>Zalo, VNG Corporation</i> Finished at 4th place on the Private Leaderboard of the Voice Gender/Accent Classification challenge. Zalo AI Challenge is an annual Kaggle-like competition hosted by Zalo - one of the biggest tech companies in Vietnam. In 2018, the competition attracted over 700 teams competed in 3 challenges.	<b>2018</b>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------

## Professional Services

<b>Teaching/Mentoring</b> <i>Cinnamon AI Bootcamp 2020, 2022, 2023</i> • Mentored groups of 3-4 students. • Designed syllabus, prepared entrance tests, interviewed candidates. • Prepared materials and gave lectures on Language Modelling and Transformers.	
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--

<b>Reviewer</b> <i>ICLR 2026, MoFA @ ICML 2025, XLLM-Reason-Plan @ COLM 2025</i>	
-------------------------------------------------------------------------------------	--

## Experience

**AI Research Engineer**  
*Torilab*

July 2024 — Present

- **Building AI avatars.**

- Conducting research on controlling LLMs' behaviours without prompting (i.e. activation steering).
- LLMs fine-tuning (SFT, DPO, classification), prompt engineering, LLMs serving (vLLM, sglang, LitServe) with dynamic batching.
- Developing the core framework for the customizable chatbots with memory, tool use, dynamic prompts and multi-turn interactions.
- Other random tasks: User's engagement detection with head-pose tracking; Image Generation with Diffusion-based models and LoRAs; multi-speaker diartization.

**AI Research Engineer**  
*Cinnamon AI*

Nov 2018 — May 2024

- **Developed RAG-based applications.**

- Co-creator of **kotaemon** (24.5k+ ★): An open-source tool for local RAG application built for both end users and developers.
- Built demos for LLM-powered applications targeting the Insurance domain.

- **Researched, developed, and implemented AI solutions for Document Image Understanding.**

- Information Extraction and Cross-lingual adaptive pre-training for a low-resource language (Japanese).
- Created new technical assets by introducing new Information Extraction models that became the new standard for client projects. Increased the f1-score by 2% - 7%.
- Publish [papers](#) on Information Extraction at peer-reviewed conferences, 1 Best Paper Award at DIL-ICDAR'21.

- **Developed data-driven products and processes.**

- Implement synchronization and local version control for the internal data management system.
  - Initiated data standardization, defined and implemented processes regarding data life cycle and organization.
- Other tasks: Support cross-department alignment; Conduct/Facilitate technical sharing sessions; Training/Mentoring.

**Undergraduate Research Assistant**  
*IOT Lab, UET - Vietnam National University*

Aug 2017 — Sep 2019

- **Predictive models for wellbore data using machine learning and statistical methods.**

- Facies/rock type classification; Time-series Analysis; Permeability Regression; Integrated Prediction Error Filter Analysis (INPEFA) curve calculation; Cumulative and Federated Learning.

## Personal Projects

**Kotaemon** (24.5k+ ★) - An open-source tool for local RAG application.  
*Open-source project, Co-creator*

Jan 2024 — May 2024



- A local RAG-based tool for chatting with your documents. Built with both end users and developers in mind.
- For end users: A local Question Answering UI for RAG-based QA.
- For developers: A framework for building your own RAG-based QA pipeline.

**Data Utility Improvement Experiment for DECAF**  
*Personal research*

Oct 2022 — Nov 2022



- A personal research on Causal Inference, Algorithmic Fairness and specifically the paper [DECAF: Generating Fair Synthetic Data Using Causally-Aware Generative networks](#).
- Improved data utility of the DECAF method using alternating graph during synthesis while still achieving similar level of fairness.

**Gender/Accent Classification for Vietnamese short voice recordings**  
*Zalo AI Challenge 2018*

Aug 2018 — Sep 2018



- Classify the speaker's voice in a recording (typically under 3 seconds) by gender and regional accent.
- **4th place** on the Private Leaderboard, achieved **79.208% accuracy** in 10 days as an **individual participant**.