

Nguyen Anh Tu

Ha Dong, Ha Noi, Vietnam
☎ (+84) 964692796
✉ anhtunguyen446@gmail.com
🌐 My Webpage
🐙 Github in LinkedIn 🗣 Skype

Education

2016–2021: **Bachelor Engineer Degree**, *Posts and Telecommunications Institute of Technology*, Hanoi, Vietnam.

Major: Information Technology

Programing, Data structures and algorithms, Analysis & design of information systems

Thesis (9.2/10): Information extraction from legal questions

Research Experience

Posts and Telecommunications Institute of Technology, Hanoi, Vietnam

2019–present **Legal Text Processing**, *Research and develop neural network architecture to extract information from legal text and spoken language processing.*

Supervisor : **Dr Ngo Xuan Bach**, *Associate Professor, Vice Dean of Faculty of Information Technology, Posts and Telecommunications Institute of Technology (PTIT), Hanoi, Vietnam.* ([Personal Web-page](#))

Job Experience

10/2020– **Technology Specialist**, *Vingroup Big Data Institute.*

1/2021 ○ **Vietnamese Text Sentiment Classification**

- Developed classification models based on feature engineering and deep learning.
- Classification model reach top 1 private testset of AIVIVN challenging.

○ **Vietnamese Accent Restorer**

- Developed accent recovering models based on sequence labeling and sequence to sequence models.

2/2021– **Applied Scientist**, *Vinbrain JSC.*

3/2022 ○ **Automatic Radiology Report Editing Through Voice**

- Developed the NLP controller model based on JointBERT model for detecting intent and extracting content command of doctor

- Deployed NLP controller and integrate with ASR service

○ **Medical equipment registration classification system for ministry of health**

- Developed machine reading comprehension model based on XLM-R model to extract information from document (support both English and Vietnamese)

- Optimized machine reading comprehension model using knowledge distillation, quantization and transform pre-trained multi language model to bi language model

- Integrated OCR model, NLP model to classify the power of attorney

○ **Automatic speech recognition system**

- Developed a joint learning model to recover capitalization and punctuation

- Developed a spoken norm model to transform spoken language to written language

4/2022- **Research Engineer**, Samsung SDS R&D Center, Vietnam.

- present
- **Vietnamese Question Answering System**
 - Developed Machine Reading Comprehension model reach top 1 on private testset of VLSP 2021 challenging
 - Optimized model architecture and runtime for MRC model using knowledge distillation and quantization
 - **Vietnamese Korean machine translation**
 - Built large scale high quality Vietnamese-Korean machine translation dataset
 - Developed machine translation model based on mBART

Computer skills

Programming Languages Python, JAVA, C/C++

Frameworks and libraries Pytorch, Scikit-learn, Numpy, Pandas, Matplotlib, Flask, Huggingface, TensorRT, Triton server, ONNX, FastAPI

Software Development Programming Paradigms, GIT, Docker, Linux

Languages **Vietnamese:** Native , **English:** IELTS 6.5

Honors and Awards

- Bronze Medal, Google AI4Code – Understand Code in Python Notebooks, Kaggle Competitions
- Consolation prize in ACM/ICPC PTIT 2019 programming contest
- The first prize in Machine translation challenge hosted by VLSP2022

Publications

Journal Articles

- 2022 Oanh Thi Tran, Thang Van Nguyen, **Tu Anh Nguyen**, and Ngo Xuan Bach. Learning student intents and named entities in the education domain. *International Journal on Artificial Intelligence Tools*. accepted, SCIE Q3, 2022, (**Impact Factor:1.208**).

In Conference Proceedings

- 2023 **Nguyen Anh Tu**, Duong Xuan Hieu, Tu Minh Phuong, and Ngo Xuan Bach. A bidirectional joint model for spoken language understanding. In *under review ICASSP*, 2023.
- 2022 Hoang Thi Thu Uyen, **Nguyen Anh Tu**, and Ta Duc Huy. Vietnamese capitalization and punctuation recovery models. In *Interspeech*, 2022.
- 2022 Tran Ngoc Son*, **Nguyen Anh Tu***, and Nguyen Minh Tri. An efficient approach for machine translation on low-resource languages: A case study in vietnamese-chinese. In *Proceedings of the 9th International Workshop on Vietnamese Language and Speech Processing (to appear)*, 2022.
- 2021 **Nguyen Anh Tu**, Hoang Thi Thu Uyen, Tu Minh Phuong, and Ngo Xuan Bach. Analyzing vietnamese legal questions using deep neural networks with biaffine classifiers. In *International Conference on Neural Information Processing*, pages 513–525. Springer, 2021.
- 2021 Manh Hung Nguyen, Vu Hoang, **Tu Anh Nguyen**, and Trung H. Bui. Automatic Radiology Report Editing Through Voice. In *Proc. Interspeech 2021*, pages 4862–4863, 2021.
- 2021 Ta Duc Huy, **Nguyen Anh Tu**, Tran Hoang Vu, Nguyen Phuc Minh, Nguyen Phan, Trung H Bui, and Steven QH Truong. Vimq: A vietnamese medical question dataset for healthcare dialogue system development. In *International Conference on Neural Information Processing*, pages 657–664. Springer, 2021.