(+84) 972865664
Ho Chi Minh City, Vietnam
nvatuan3@gmail.com

Personal blog: tuanio.github.io
github.com/tuanio
linkedin.com/in/tuanio
kaggle.com/tuannguyenvananh

Nguyen Van Anh Tuan

AI Engineer

I have 1+ year of experience working on data science-related tasks like data processing, visualization, and modeling. I also have the skills to research Anomaly Detection (3+ months) and Automatic Speech Recognition (10+ months). Those things come with good programming skills, and I am seeking a job as a AI Engineer.

Skills

Programming	Python (PyTorch, Tensorflow, Numpy, Pandas), C/C++.
Communication	English (Strong reading, listening and Intermediate writing, speaking)
Misc.	Academic research, teaching, teaching assistant, training, consultation.

Technical Experience

Machine Learning Researcher / EduplaX　　　　　　　　　　　　　　　**Aug 2021 — Present**
WeAi　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　*Ho Chi Minh City, Vietnam*
- Successfully apply the research's results to build a new feature for the product.
- Research Computerized Adaptive Testing, Automatic Speech Recognition.

Projects

Automatic Pronunciation Error Detection (APED)　　　　　　　　　　**Aug 2022 — Dec 2022**
Graduation Thesis - Personal
- Successfully implemented and applied Semi-supervised Learning techniques Noisy Student Training for the APED task. (GitHub).
- Relative reduce of 26% from 17% PER (teacher) to 12% PER (student) on Phoneme Recognition task on Libri-Light and LibriSpeech dataset.
- **Achivments:** Knowledge about Speech Recognition, Semi-Supervised Learning, Self-supervised Learning, Noisy Student Training, wav2vec2.0, Conformer.

Image to Latex　　　　　　　　　　　　　　　　　　　　　　　　　**Aug 2022 — Oct 2022**
Computer Vision subject project - Leader
- Successfully built a codebase for experimenting with various Encoder-Decoder models for Image to Latex task (GitHub).
- Achieve BLEU score of 77% on im2latex-100k Dataset.

Vietnamese Automatic Speech Recognition (ASR)　　　　　　　　　　**Dec 2021 — Present**
Research project - Leader
- Successfully built a toolkit for experimenting with some state-of-the-art ASR models on English dataset to Vietnamese dataset (GitHub).
- Achieve WER 20% on Vivos Test Dataset, training on 115 hours of data (VLSP 2020 and Vivos Train Dataset).
- Have accepted paper to local university YSC - IUH 2022 (pp 231-239) and accepted as a short paper at FDSE - 2022.

Anomaly Detection　　　　　　　　　　　　　　　　　　　　　　　**Dec 2021 - Mar 2022**
Research project - Leader
- Successfully applied Hypothesis Testing on an Unsupervised Autoencoder for an Anomaly Detection task.
- Experiment on NSL-KDD Dataset. Achieved accuracy: 96%. GitHub.

Recommendation System　　　　　　　　　　　　　　　　　　　　**Oct 2021 - Dec 2021**
Application Development subject project - Leader
- Successfully apply Content-based Collaborative Filtering in the Android app, Book Recommendation app (Mobile source, API source).

Education

Bachelor of Science in Data Science, *Industrial University of Ho Chi Minh City, Vietnam*　　　　Aug 2019 — 2023

- GPA: 3.35/4.0
- Thesis title: *A Semi-Supervised Learning approach for Automatic Pronunciation Error Detection.* (Progress: *Done*)

Rewards & Achivements

Third Prize, Euréka at IUH.	2022
Top 20, startup contest Inno Greenlife IUH.	2021
Second Prize, TDMU - Entropy Data Analytics contest.	2021
Consolation Prize, Olympic Informatics for Vietnamese students (specialize board).	2020
Consolation Prize, ACM/ICPC Asia region.	

Activities

Administrator of Problem Based Learning, the Data Science playground for IUH students.	May 2021 — Aug 2021.
Teaching Assistant, including subjects: Programming Stats and Application, Programming basic, Programming Techniques, Introduction to Programming	2020, 2021