

PHONG DO NGUYEN THUAN

Phone: +84 788 927 163

Mail: phongdntvn@gmail.com

Location :Vietnam



I'm a researcher with a strong foundation in Natural Language Processing (NLP) and a passion for pushing the boundaries of Artificial Intelligence (AI) through cutting-edge machine learning and deep learning models. My experience lies in conducting research and experiments with state-of-the-art models, particularly focusing on machine reading comprehension, question answering, and information retrieval. Driven by a deep research background in these areas, I'm currently pursuing a PhD scholarship in AI, specifically focusing on the NLP and Large Language Model (LLM) fields.

PUBLICATIONS

I have experience in AI research on natural language processing, especially on the topic of question and answer and information retrieval. I have successfully published **nine papers** on AI at leading conferences around the world such as EACL, NAACL...

VLUE: A New Benchmark and Multi-task Knowledge Transfer Learning for Vietnamese Natural Language Understanding

Phong Nguyen-Thuan Do, Son Quoc Tran, Phu Gia Hoang, Kiet Van Nguyen, Ngan Luu-Thuy Nguyen, accepted by NAACL 2024 (Findings)

AGent: A Novel Pipeline for Automatically Creating Unanswerable Questions

Son Quoc Tran, Gia-Huy Do, **Phong Nguyen-Thuan Do**, Matt Kretchmar, Xinya Du

Multi-stage transfer learning with BERTology-based language models for question answering system in vietnamese

Kiet Van Nguyen, **Phong Nguyen-Thuan Do**, Nhat Duy Nguyen, Anh Gia-Tuan Nguyen, Ngan Luu-Thuy Nguyen, accepted by International Journal of Machine Learning and Cybernetics

Revealing Weaknesses of Vietnamese Language Models Through Unanswerable Questions in Machine Reading Comprehension

Son Quoc Tran, **Phong Nguyen-Thuan Do**, Kiet Van Nguyen, Ngan Luu-Thuy Nguyen, accepted by EACL 2023

The Impacts of Unanswerable Questions on the Robustness of Machine Reading Comprehension Models

Son Quoc Tran, **Phong Nguyen-Thuan Do**, Uyen Le, Matt Kretchmar, accepted by EACL 2023

ViMRC VLSP 2021: XLM-R Versus PhoBERT on Vietnamese Machine Reading Comprehension

Nguyen Duy Nhat, **Do Nguyen Thuan Phong**, accepted by VNU Journal of Science

XLMRQA: open-domain question answering on Vietnamese wikipedia-based textual knowledge source

Phong Nguyen-Thuan Do, Nhat Duy Nguyen, Kiet Van Nguyen, Tin Van Huynh, Anh Gia-Tuan Nguyen, Ngan Luu-Thuy Nguyen, accepted by ACIIDS 2022

Sentence extraction-based machine reading comprehension for Vietnamese

Phong Nguyen-Thuan Do, Nhat Duy Nguyen, Tin Van Huynh, Kiet Van Nguyen, Anh Gia-Tuan Nguyen, Ngan Luu-Thuy Nguyen, accepted by KSEM 2021

Vireader: A wikipedia-based vietnamese reading comprehension system using transfer learning

Kiet Van Nguyen, Nhat Duy Nguyen, **Phong Nguyen-Thuan Do**, Anh Gia-Tuan Nguyen, Ngan Luu-Thuy Nguyen, accepted by Journal of Intelligent & Fuzzy Systems

EDUCTIONS

BACHELOR OF COMPUTER SCIENCE

University of Information Technology, Vietnam National University, Ho Chi Minh city

2018 - 2022

GPA: 3.6/4

EXPERIENCE

RESEARCH ASSISTANT

The UIT NLP Group - University of Information Technology - <https://nlp.uit.edu.vn/home>

Mentored by Kiet Nguyen Van and Ngan Nguyen Luu Thuy

2020 - 2024

- Publishing a benchmark (VLUE) to evaluate language models in Vietnamese for natural language understanding tasks and a pre-trained state-of-the-art Vietnamese model (CafeBERT).
- Focus on in-depth research on machine reading comprehension models and □QA systems. Achieve SOTA results in Vietnamese and other languages.

AI ENGINEER

Zalo - VNG Corporation

2021 - 2024

- Developed machine learning models for **Kiki** virtual assistant on ZingMP3 app and in-car systems, reaching over **1 million users**.
- **Analyzed real-world data** from text and audio sources to **extract insights** and inform product development.
- Proven experience in **natural language processing** tasks, including text classification and entity detection; **speech processing** tasks, including voice activity detection, speaker diarization.
- **Collaborated** with cross-functional teams to gather requirements, receive data, provide data insights, and deliver machine learning models.

□AWARDS

Encouragement Scholarship	2018, 2019, 2020, 2021
The consolation prize of the 2016 National Youth Informatics Competition	2016
Bronze Medal in Informatics for 30/4 Olympic Exam	2017
Bronze Medal at Phuong Nam Summer Camp	2016
First prize in Informatics at the provincial competition for excellent students	2017