

# Nguyen Phuc Minh

Ho Chi Minh City, Vietnam  
☎ (+84) 34 9933 957  
✉ npminh@apcs.fitus.edu.vn  
🌐 demdecuong



## Education

2017–2021 **University of Science**,  
*Advanced Program of Computer Science, Bachelor.*  
GPA – 3.56/4.00 (8.13/10)

## Industrial Experience

- May 2022 – Present **Associate AI Engineer**, FPT Software AI Center.
- Apr 2022 – Present **Associate Chatbot Engineer**, Sani - chatbot giúp tìm quán cafe.
- Chat with Sani
- Sep 2021 – May 2022 **Junior NLP Applied Scientist**, VinBrain.
- **Healthcare Professional Recommendation System**
    - Build data provider pipeline for doctors/patients
  - **Smart Consultation**
    - Design and Implement triage-classification chatbot for common symptoms and COVID-19 with RASA.
    - Design and Build self-assessment for both narrow-based and deep-based.
    - Survey business applications and write Proposal document.
  - **Knowledge Base Construction**
    - Build Knowledge base question answering for Vietnamese healthcare.
    - Construct Healthcare Knowledge Base for Vietnamese
  - **DMEC**
    - Highlight registrant name, address, equipment in Technical Description Document.
    - Improve NER model to  $F1=0.98$  via adversarial learning in Medical's Document .
    - Synthesize dataset based on pattern analysis
  - **ViHealthBERT: Healthcare-related Language Models**
    - Build pipeline for Pretrained Language Model including: crawling, augmentation, pre-processing, pre-training, multi-task learning, fine-tuning.
    - Our Language Model outperforms PhoBERT in 3 downstream tasks span out 4 datasets (COVID19-NER, ViMQ, acrDrAid, FAQ Summarization).
- Apr 2021 – **Applied Scientist Intern**, VinBrain.
- Aug 2021
  - Achieve 49.04, 30.24, 42.31 in R-1, R-2, and R-L in Vietnamese's medical question summarization.
  - Improve stroke's sentence-matching top-1 accuracy to 76.66 with representation learning.
  - Release 2 model versions for stroke representation and vi-nlp-core library for appointment NER.
- May 2020 – **AI Engineer Intern**, GMO-Z RUNSYSTEM.
- Aug 2020
  - Error Correction for Japanese/Vietnamese
    - Use Masked Language Model concept to enhance model performance up to 80% sentence-AUC for Japanese correction.
    - Vietnamese correction model achieves 71.3% sentence-AUC.

## Research Experience

HCMUS - Knowledge Engineering Department

- Dec 2019 – **Abstractive Summarization**, *Improving Abstractive Summarization without Pretrained LM.*
- Aug 2021
  - Apply double keys and query-value interaction to improve Transformer model by (+0.57, +0.53, 0.58) in R-1, R-2 and R-L in CNN/DailyMail dataset
  - Propose the SEGMENT model which outperforms SOTA (+0.38, +0.5) in R-1 and R-2 metrics in CNN/DailyMail dataset.
  - **Thesis:** "Towards Robust abstractive text summarization via augmenting essential information" - Rate : Outstanding
- Advisor **MSc Tran T.Thao Nhi**, Associate Professor, Knowledge Engineering Department, University of Science (HCMUS), Vietnam ([Google Scholar](#)).

- Sep 2021 – **ViHealthBERT**, *Pretrained Language Models for Vietnamese in Health Text Mining*.
- Jan 2022
  - Crawling data and select domain-specific sentences from OSCAR dataset.
  - Achieve SOTA performance on 4 benchmarks.
  - Competitive baseline in industry tasks.
- Jan 2022 – **Knowledge Base Construction**.
- Feb 2022
  - Construct Healthcare Knowledge Base result in 696 disease, over 6200 disease attribute rich text, 710 symptoms, 2233 symptoms-disease relation, 511 disease-department relation.
  - Provide Cloud service for KB via Neo4j Aura.

## Achievements

- Apr 2022 **Top 4 VinBrain Hackathon 2022 : KnowLife**.
  - Achieved 9m VND stocks and ranked 4 among 11 teams for innovative and practical solutions
  - PageRank for related symptom suggestion
- Jan 2022 **VinBrain - Employee of 2021**.
  - Top 20% employee has best performance
- Nov 2021 – **Ranked 2 Post-Competition, Zalo AI Challenge**.
- Dec 2021
  - Build End-to-end Retrieval-Reranking System for Legal Text Retrieval
  - Achieve F2-score=70.61 on Public set & F2-score=56.41 on Private set (top-10 in overall).
- July 2020 – **Top 20 HAI challenge 2020 : X-Ray's Reporter**.
- Nov 2020
  - Awarded top 20 over 300 teams/start-ups during 6 months
  - Funded 200m VND for the second round
  - Build system (CV + NLP) that generate report from lung's xray image during 6 months
- Nov 2020 **Top 10 Online AI Hackathon reset101 : Plant's Dr**.
  - Ranked in Top 10 among over 60 global teams.
  - Build an AI platform that detect plant disease via mobile device.

## Publications

- 2021 **Minh-Phuc Nguyen** and Nhi-Thao Tran. "Improving Abstractive Summarization With Segment-Augmented And Position-Awareness". *Procedia Computer Science* 189 (2021). *AI in Computational Linguistics*, pp. 167–174.issn: 1877-0509
- 2021 Ta Duc Huy, Nguyen Anh Tu, Tran Hoang Vu, **Nguyen Phuc Minh**, Nguyen Phan, Bui Huu Trung, Steven Q.H. Truong. "ViMQ: A Vietnamese Medical Question Dataset for Healthcare Dialogue System Development". *Neural Information Processing*, pages 657–664, Cham. Springer International Publishing
- 2021 Vu Hoang, Nam Pham Hai, Trung Bui, Vu Tran Hoang, **Minh Nguyen Phuc**, Steven Q. H. Truong "CircleNLU: A tool for building Data-Driven Natural Language Understanding System". *Data-Centric AI Workshop (DCAI @ NeurIPS 2021)*
- 2021 Vu Hoang Tran, **Minh Phuc Nguyen**. "HN-BERT at VLSP 2021 - ViMRC Challenge: An empirical study of Vietnamese Machine Reading Comprehension with Unsupervised Context Selector and Adversarial Learning". *Proceedings of the 8th International Workshop on Vietnamese Language and Speech Processing - VLSP 2021* (Accepted)
- 2022 **Minh Phuc Nguyen**, Vu Hoang Tran, Hoang Vu, Huy Duc Ta, Trung H. Bui, Steven Q. H. Truong "ViHealthBERT: Pre-trained Language Models for Vietnamese in Health Text Mining". *13th International Conference on Language Resources and Evaluation* (Accepted)

## Certificates

- Mar 2022 Rasa Developer Certificate
- Nov 2021 HCMUS Science-Technology Student Award
- Oct 2019 Software Development With Scrum

## Skills

- Programming Languages Python, Cypher
- Frameworks Pytorch, Flask, Rasa
- Database Neo4j
- Tools & Technologies Huggingface, Git, Pypi, Docker, Strapi
- Search Engine Algolia, Elasticsearch