

P: +84 907 198 648
E: dinhngoc123@gmail.com
W: github.com/dinhngoc267
S: linkedin.com/in/ngocledinh-267

NGOC LE DINH



AI ENGINEER

PROFILE

Year of birth: 1998
Work location: Ho Chi Minh / Ha Noi VietNam

SKILLS

PROFESSIONAL

- Python/C++/Javascript
- Machine Learning
- Deep Learning
- Pytorch/Keras
- NLP
- Computer Vision
- Data Mining
- SQL

LANGUAGE

- Vietnamese
- English: TOEIC 750

EXPERIENCE

December 2022 to current

NLP RESEARCHER - PARTIME | UNIVERSITY OF SCIENCE

- Problem: Entity Linking in Biomedical
- Journal: Frontiers
- Journal Impact: 3rd most-cited publisher
- Github: <https://github.com/dinhngoc267/Entity-Linking-in-Biomedical>

September 2020 to January 2023

PRACTICAL INSTRUCTOR, T.A | UNIVERSITY OF SCIENCE

- Practical instructor for general programming subjects in faculty of Information Technology such as programming introductions and data structures and algorithms in C/C++ programming language

June 2021 to September 2021

MACHINE LEARNING ENGINEER - REMOTE | QUBIT CAPITAL - INDIA

NLP tasks:

- Check if the answers are relevant to the question or not: Embedding text by pretrained model - BERT, word2vec, sen2vec. The challenge of this task is I don't have any dataset so I have to use pretrained model to hopfully embed sentences into meaning vectors then use cosine similarity to calculate the score of similarity and set a constant threshold to decide the answer is relevant or not
- Build an automatic response suggestion system for an email. This system will generate an appropriate response email for a user can use it directly or customize it before send to partner - using GPT 2

EDUCATION

2020-2023

UNIVERSITY OF SCIENCE
Master in Computer Science
GPA: 8.65/10

2016-2020

UNIVERSITY OF SCIENCE
Bachelor of Software Engineer
GPA: 8.2/10

March 2022 to September 2022

DATA SCIENCE INSTRUCTOR/MENTOR | CODERSCHOOL

- Teaching: Basic Python programming, SQL, Data Visualization and basic Machine Learning

June 2021 to February 2022

MACHINE LEARNING TEACHING ASSISTANT | CODERSCHOOL

- Support the students throughout the Machine Learning courses
- Guide the students to do their final project related to Machine Learning or Deep Learning, name a few:
 - VND banknote classification problem: Using deep neural network
 - Facial Recognition - By using Yolo3 and few shot learning technique to make the problem more realistic - learn to identify a person among a lot of people just by a few training images
 - Detect misleading information on Covid-19 - Using TF-IDF to represent text and using stack model to combine many machine learning classification algorithms like Logistic Regression, Decision Tree, Naive Bayes to improve the final performance

August 2019 to October 2019

WEB INTERNSHIP | TRI VIET

- Develop web application supports managing professional records - Using C# and javascript

March 2019 to July 2019

WEB DEVELOPMENT COLLABORATOR | Q.P.S HOUSE

- Develop web application for warehouse management - Using C# and Javascript

MY PROJECTS

LINKING NAMED ENTITIES IN MEDICAL DOCUMENTS TO CONCEPTS IN STANDARDISED MEDICAL TERMINOLOGIES

Context: In the medical field, the normalization of entities in the text plays an important role. Specifically in the patient's medical records, doctors can use different names for disease names, symptoms, and treatments. If we could standardize these different names to a set of standard vocabulary will be of great help to the above research patient's medical record

Method: I build a model which learns to represent entities and their variants through the BiLSTM Siamese network then I use cosine similarity to calculate the score similarities then apply K-nearest neighbors to standardize the query entities

OPEN DOMAIN LONG-FORM QUESTION ANSWERING

Context: Open domain Long-form QA is a fundamental challenge in natural language processing (NLP) that involves retrieving documents relevant to a given question and using them to generate an elaborate paragraph-length answer

Method: I divided the problem into 3 stages. Phase 1 uses Google api or Wiki api to search questions directly, then scrape these web pages and break them down into paragraphs of text. Phase 2, training the Paragraph Retriever model, this model will learn to represent text passages and questions, and then rank the passages that are likely to contain the answer to that question by cosine similarity. Stage 3 generates text, from the top k paragraphs taken in stage 2, I used the text summarization model - BART to generate the final response text.