

# PHÁT TRIỂN HỆ THỐNG Q-A VỀ LUẬT VI PHẠM LIÊN QUAN ĐẾN XE MÁY TẠI VIỆT NAM

Nguyễn Quang Huy

Trường Đại học Công nghệ thông tin  
Đại học quốc gia thành phố Hồ Chí Minh

## What ?

Đề xuất phát triển ứng dụng Q-A để mọi người hỏi đáp về các luật liên quan đến xe máy, các vi phạm và xử lý

- Xây dựng dataset hoàn chỉnh, đầy đủ và chính xác về văn bản luật giao thông, các tài liệu hướng dẫn và các giải thích rõ ràng về các vi phạm giao thông liên quan đến xe máy
- Fine-tune mô hình PhoBERT cho tác vụ cụ thể này về tiếng Việt

## Why ?

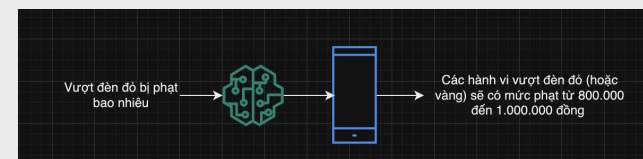
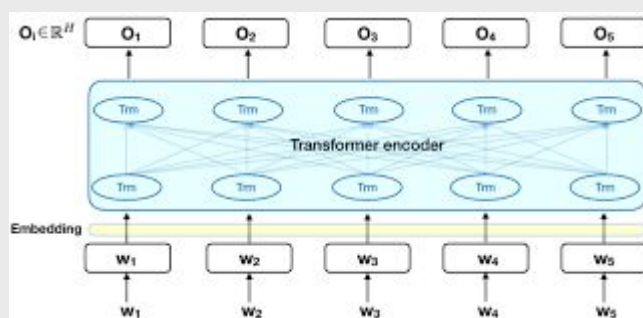
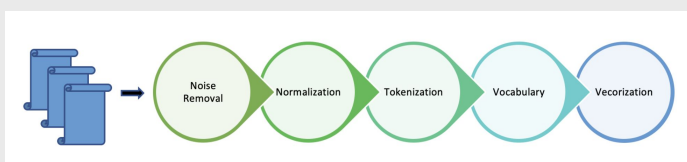
- Xe máy là phương tiện giao thông chính tại Việt Nam, mọi người khi tham gia giao thông phải tuân thủ luật. Tuy nhiên nội dung văn bản về luật giao thông liên quan đến xe máy lại quá dài dòng, cần có ứng dụng để hỏi đáp nhanh và thân thiện nhất với người dùng
- Với sự phát triển của deep learning (hiện nay), tận dụng mô hình PhoBERT được pretrain từ mô hình BERT nổi tiếng để xử lý các tác vụ liên quan đến tiếng Việt sẽ rất thực tiễn

## Overview

Thu thập dữ liệu

Fine-tune mô hình PhoBERT

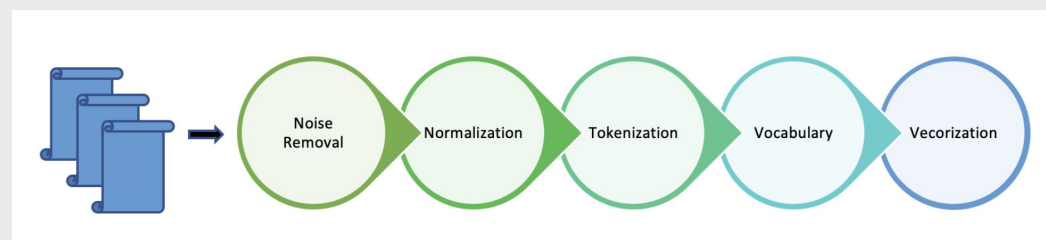
Deploy ứng dụng



## Description

### 1. Thu thập dữ liệu

- Dữ liệu được thu thập qua các văn bản luật giao thông, các tài liệu hướng dẫn và các giải thích rõ ràng trên các trang mạng chính thống về quy định trong luật liên quan đến xe máy.
- Tiến hành tiền xử lý dữ liệu: khử nhiễu, tokenize, split dataset
- Lưu dữ liệu thu thập được ở trong các dịch vụ cloud: Google Firebase



### 3. Deploy ứng dụng

- Sau khi mô hình test đạt yêu cầu, lưu mô hình sau khi fine-tune ở định dạng phù hợp: TorchScript hoặc ONNX
- Import mô hình vào trong ứng dụng di động

### 2. Fine-tune PhoBERT

- Import mô hình PhoBERT để fine-tune theo tác vụ: Q-A văn bản luật giao thông liên quan đến xe máy
- Sử dụng PhoBERT-large và thêm lớp Attention ở cuối để mô hình có thể tập trung vào tác vụ của câu hỏi về văn bản luật giao thông liên quan đến xe máy làm cho mô hình cải thiện được độ chính xác
- Đánh giá mô hình bằng phương pháp EM (Exact Match), F1 Score

