

## Đồ Án Cuối Kỳ

### I. Tổng quan

- Trong thời gian gần đây, ChatGPT với khả năng xử lý ngôn ngữ tự nhiên mạnh mẽ, đã trở thành công cụ vô cùng phổ biến và hữu ích, được áp dụng rộng rãi trong nhiều lĩnh vực khác nhau. Nền tảng của ChatGPT chính là kiến trúc Transformer – một kiến trúc đem đến sự đột phá trong lĩnh vực Xử lý Ngôn ngữ Tự nhiên trong những năm trở lại đây. Cơ chế self-attention được sử dụng trong Transformer giúp mô hình có được hiệu năng (cả độ chính xác lẫn tốc độ) hoàn toàn vượt trội so với các kiến trúc trước đây.
- Trong đồ án này, sinh viên sẽ tìm hiểu và xây dựng một ứng dụng thực tế từ một mô hình có nền tảng là kiến trúc Transformer.

### II. Yêu cầu

- Trước tiên, sinh viên sẽ tự chọn một bài toán trong lĩnh vực Xử lý Ngôn ngữ Tự nhiên để giải quyết. Gợi ý một số bài toán có thể chọn là:
  - Part of Speech (PoS) Tagging.
  - Text Classification.
  - Named Entity Recognition.
  - Machine Translation.
  - Text Summarization.
  - Question Answering.
  - ...
- Sau đó, sinh viên tiếp tục chọn một mô hình có nền tảng là kiến trúc Transformer để giải quyết bài toán được đặt ra và lần lượt thực hiện 2 yêu cầu sau:
  - 1) (7 điểm) Tìm một dataset phù hợp và huấn luyện lại mô hình trên bộ data đó. Sau đó, sinh viên cần thực hiện đánh giá mô hình bằng các độ đo phù hợp. Trong báo cáo, phải có mô tả chi tiết về tập dữ liệu, mô hình được sử dụng, cách phân

chia dữ liệu (train/validation/test) và kết quả đánh giá mô hình bằng các độ đo phù hợp.

- 2) (3 điểm) Xây dựng một ứng dụng có giao diện web từ mô hình vừa được huấn luyện.

### **III. Quy định nộp bài**

- Sinh viên cần nộp lại toàn bộ mã nguồn, bộ weight, dữ liệu huấn luyện và báo cáo.
- Trong trường hợp dữ liệu cần nộp có kích thước quá lớn, sinh viên có thể upload dữ liệu lên drive và nộp đường dẫn tương ứng.