**Dự án Giữa kỳ**

**Môn: Học sâu (Deep learning)**

Bài toán: hãy xây dựng mô hình giải quyết bài toán sau:

**Câu hỏi + bức ảnh (image) -> câu trả lời**

1. Hãy xây dựng dữ liệu, hạn chế trên một miền dữ liệu và một bài toán hẹp nào đó để dữ liệu được tập trung (tránh hiện tượng dữ liệu thưa), tức là:
   * Hạn chế về loại ảnh: ví dụ chỉ ảnh về trái cây
   * Hạn chế về loại câu hỏi: ví dụ chỉ hỏi về số lượng và nhận dạng

Có thể dùng LLMs để sinh dữ liệu, ví dụ:

* + Thu thập các ảnh
  + Dùng LLMs để tạo câu hỏi và câu trả lời

1. Xây dựng mô hình giải quyết bài toán này với các yêu cầu:
   * Sử dụng CNN để học đặc trưng của ảnh
   * Sử dụng LSTM để sinh output
   * Lưu ý: kết nối CNN và LSTM thành mô hình thống nhất
   * Thử nghiệm và so sánh các tiếp cận: **dùng pretrain models và train model từ đầu; sử dụng LSTM sinh sequence có attention và không có attention**

**Yêu cầu:**

* + Hiểu lý thuyết các mô hình
  + Hiểu code, hiểu cách làm

**Nộp bài:**

1. Tất cả các thành viên trong nhóm phải hiểu tất cả các bài (tránh tình trạng phân công mỗi người làm 1 bài và người này không biết người kia làm gì)
2. Tất cả các thành viên trong nhóm đều phải nộp bài; ai không nộp bài sẽ bị 0 điểm
3. Bài nộp bao gồm file pdf chứa lý thuyết, thuyết minh bài toán và cách làm; và các file code;