

more_research

Harito ID

<https://harito.id.vn/>

2025-11-25

1 Nội dung chính ở tuần 12 của học kỳ

Do tuần này không học lý thuyết nên các bạn sẽ chú ý các nội dung chính sau:

- Chuẩn hoá link report các lab từ 1 tới 6. Đây là nhiệm vụ quan trọng nhất. Mình xem bài làm của nhiều bạn không viết báo cáo, chỉ có code thuần và tổ chức code không khoa học nên khó theo dõi. Report là quan trọng nhất để người xem hiểu nỗ lực của các em trong làm bài lab.
- Hoàn thiện các bài lab trước đây chưa xong, đặc biệt là lab5 và lab6.

Cấu trúc gợi ý cho repository như sau:

```
./
|__src/: Chứa các code như .c .py .scala ...
|__report/: Chứa các báo cáo
|__notebook/: Chứa các notebook để code nhanh 1 chủ đề~
|__test/: Chứa code ghi test.
|__data/: Chứa nội dung data.
|__README.md: Giới thiệu tổng quan về`project.
|__gitignore: Loại bỏ các tệp không quan trọng hoặc quá lớn.
```

1.1 Ví dụ về ./report/

Bao gồm như: lab1_part1.md, lab2_part2.pdf, lab3_part4.txt, ... Để đảm bảo việc trích xuất thông tin không bị lỗi, khuyến khích để report ở định dạng có thể lấy nội dung 1 cách dễ dàng và không chứa các nội dung gây nhiễu.

1.2 Lưu ý về .gitignore

Ví dụ khi push lên github thì ./data/ chỉ chứa các mô tả về các cột, kiểu dữ liệu. Tuyệt đối không up toàn bộ dataset lớn lên vì làm quá tải của số token đầu vào của AI cũng như khiến model khó tập trung vào các nội dung chính cốt lõi.

Đối với những bạn đã xong 2 nhiệm vụ trên thì tham khảo phần dưới đây.

2 Nội dung thêm ở tuần 12 của học kỳ

2.1 Bối cảnh

Khả năng tự học là quan trọng, đặc biệt là khi các em học xong và không còn ai hướng dẫn. Với sự phát triển của các hệ thống tìm kiếm, truy vấn thông tin, các AI/Agent và Internet. Việc tự học đang dễ dàng hơn bao giờ hết trong lịch sử.

Trong phần nội dung thêm này, các em sẽ thử sức làm 1 nhà nghiên cứu để tìm hiểu về chủ đề bài toán Text To Speech (TTS).

Yêu cầu hiện tại cho tuần này chỉ là tìm hiểu tổng quan về tình hình nghiên cứu, các hướng phát triển hiện tại, các cách triển khai, ưu nhược điểm từng cách.

2.2 Bức tranh toàn cảnh

Mình tổng hợp trước bức tranh tổng quan để việc tìm hiểu của các em thuận lợi hơn.

Level 1. Khởi đầu của TTS là các luật âm tiết cơ bản, chạy nhanh, tốt cho đa dạng các ngôn ngữ. Tuy nhiên ít tính tự nhiên.

Level 2. Các model Deep Learning vào cuộc và giúp tạo âm thanh tự nhiên hơn. Tuy nhiên thách thức đảm bảo tính đa dạng ngôn ngữ khó hơn do yêu cầu về dữ liệu. Nhiều nghiên cứu đã được thực hiện để tạo pipeline toàn diện cho người dùng. Mỗi người dùng sẽ tự ghi âm và tinh chỉnh model TTS vs trọng số riêng cho 1 người. Cách này giúp model chạy ít tài nguyên hơn (so vs level 3) trong khi vẫn đảm bảo tính tự nhiên.

Level 3. Few-shot 1 vài giây âm thanh để tạo âm thanh vs đặc trưng giọng nói cho trước. Model phức tạp và tốn nhiều tài nguyên hơn.

Cả 3 hướng tiếp cận có ưu/nhược điểm riêng, phụ thuộc ở nhu cầu và tài nguyên nguồn lực mà người dùng có.

Các thách thức chung mà nghiên cứu trong vấn đề này hướng tới:

- Hiệu suất nhanh (eg: như level 1)
- Tốn ít tài nguyên tính toán
- Đảm bảo tính tự nhiên (eg: như level 2, 3)
- Đảm bảo tính đa ngôn ngữ
- Thêm được cảm xúc cho giọng nói được tạo ra
- Tốn ít công sức cho người dùng (eg: level 1 hay chỉ cần vài giây ghi âm - level 3)

Về đạo đức nghiên cứu trong chủ đề này:

- Nhúng watermark để đánh dấu mọi đầu ra tạo bởi model AI, tránh hiểm hoạ deepfake và thông tin sai lệch.

2.3 Yêu cầu cụ thể cho bài tập (tiêu chí chấm điểm)

Do chỉ là bài tập làm thêm nên tiêu chí cho bài tập này chỉ bao gồm:

- Tìm hiểu tổng quan về bài toán, tình hình nghiên cứu, các phương pháp triển khai.

- Với từng hướng triển khai kết luận được ưu/nhược điểm, phù hợp vs các trường hợp sử dụng cụ thể nào.
- Cách các nghiên cứu tạo pipeline để tối thiểu hoá nhược điểm cho từng hướng tiếp cận hiện tại, cũng như tối đa hoá ưu điểm của chúng.