

ĐẠI HỌC QUỐC GIA TP. HCM
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN



CSC14115 - KHOA HỌC DỮ LIỆU ỨNG DỤNG

PROJECT PROPOSAL

Đề tài:

TIKTOK SCRIPT WRITING ASSISTANT

Sinh viên thực hiện: Nhóm 5

MSSV	Họ và tên
21127038	Võ Phú Hãn
21127739	Vũ Minh Phát
21127742	Nguyễn Minh Hiếu

Giảng viên hướng dẫn:

Lê Ngọc Thành

Lê Nhựt Nam

Trần Quốc Huy

Ngày 15 tháng 02 năm 2025

MỤC LỤC

1. Team Information.....	1
2. Working Title	1
3. Background	2
4. Relevance/Significance	2
5. Research Objective/Question and Hypothesis.....	2
5.1. Những yếu tố nào ảnh hưởng đến mức độ phát triển của một tài khoản TikTok?	3
5.2. Những yếu tố nào trong video sẽ ảnh hưởng đến lượt xem và độ tương tác?.....	3
5.3. Những nội dung nào có khả năng viral cao nhất?	4
6. Research Methods.....	5
6.1. Thiết kế nghiên cứu và quy trình dữ liệu:	5
6.2. Đối tượng nghiên cứu & Phương pháp thu thập dữ liệu	5
6.3. Phương pháp phân tích dữ liệu	6
6.4. Mô hình hoá phân tích & Xây dựng tính năng ứng dụng.....	7
6.5. Bảo mật dữ liệu & Đạo đức nghiên cứu	7
6.6. Mục tiêu đầu ra & Ứng dụng thực tiễn	8
7. Timeline	8
8. Resources Required	13
9. Study Limitations.....	16
10. Back-up Plan	18
11. Presentation plan	20
12. References.....	22

1. Team Information

- Lớp: Khoa học dữ liệu ứng dụng - 21KHDL
- Nhóm: 5
- Danh sách thành viên trong nhóm:

STT	Vai trò	MSSV	Họ và tên
1	Leader, Product Owner, Data Engineer, Data Analyst, Machine Learning Engineer	21127038	Võ Phú Hân
2	Data Scientist, Data Engineer, Data Analyst, Machine Learning Engineer, Project Manager	21127739	Vũ Minh Phát
3	Data Scientist, Data Engineer, Data Analyst, Machine Learning Engineer, Team Manager	21127742	Nguyễn Minh Hiếu

2. Working Title

Working Title:

"Phát triển Hệ thống AI Hỗ trợ Viết Kịch Bản Video TikTok: Phương pháp Tiếp cận Dựa trên Dữ liệu và Ứng dụng Thực tiễn"

Output:

Mục tiêu của dự án là xây dựng một trợ lý thông minh hỗ trợ quá trình sáng tạo nội dung cho các video TikTok bằng cách tự động hóa quá trình viết kịch bản dựa trên các kỹ thuật học máy và phân tích dữ liệu. Báo cáo cho dự án sẽ được trình bày một cách có cấu trúc, bao gồm:

- **Giới thiệu tổng quan:** Mô tả bối cảnh và nhu cầu phát triển nội dung sáng tạo trên TikTok, cùng với những thách thức hiện tại trong việc lên ý tưởng và viết kịch bản.
- **Phương pháp khoa học dữ liệu:** Trình bày quy trình thu thập dữ liệu, khám phá dữ liệu, xử lý dữ liệu, phân tích dữ liệu, và áp dụng các mô hình ngôn ngữ hiện đại (ví dụ: transformer-based models) nhằm tạo ra các kịch bản phù hợp với xu hướng và thị hiếu của thị trường.

- **Demo ứng dụng:** Cung cấp một minh họa cụ thể (ví dụ: một web app) về cách thức hoạt động của trợ lý, từ việc phân tích dữ liệu đến việc tạo ra các kịch bản video một cách tự động, giúp người dùng dễ dàng tương tác và hiệu chỉnh nội dung.
- **Ý nghĩa và ứng dụng thực tiễn:** Đánh giá lợi ích mà giải pháp mang lại, từ việc tăng tốc quy trình sáng tạo đến khả năng cải thiện chất lượng nội dung và thu hút lượng khán giả trên nền tảng TikTok.

3. Background

Hiện nay, trên thị trường đã xuất hiện một số công cụ hỗ trợ tự động hóa việc sáng tạo kịch bản cho các video trên TikTok (như trang web ads.tiktok.com), trong đó các giải pháp dựa trên trí tuệ nhân tạo đã chứng minh được hiệu quả khi xử lý ngôn ngữ tiếng Anh và giúp giảm thời gian sản xuất nội dung. Tuy nhiên, những công cụ này còn có hạn chế đáng kể khi chỉ cho phép người dùng nhập mô tả và yêu cầu bằng tiếng Anh, chưa đáp ứng tốt nhu cầu của cộng đồng người dùng tiếng Việt với những nét đặc trưng về ngôn ngữ, văn hóa và phong cách sáng tạo riêng. Điều này mở ra một khoảng trống quan trọng, khi chưa có giải pháp nào tối ưu hóa khả năng tạo kịch bản tự động dựa trên tiếng Việt, từ đó tạo ra nội dung phù hợp và thu hút trên TikTok. Mục tiêu của dự án là phát triển một công cụ thông minh có khả năng xử lý và tạo ra kịch bản video chất lượng cao bằng tiếng Việt, giúp người dùng trong nước có thể sáng tạo nội dung dễ dàng hơn và phù hợp với đặc thù văn hóa bản địa.

4. Relevance/Significance

Dự án này quan trọng vì nó lấp đầy khoảng trống trong nghiên cứu và ứng dụng AI, khi hiện nay các công cụ sáng tạo kịch bản cho TikTok chủ yếu hỗ trợ tiếng Anh và không đáp ứng được nhu cầu của người dùng tiếng Việt. Bằng việc phát triển một công cụ tự động hóa có khả năng xử lý và tạo ra nội dung video bằng tiếng Việt, nghiên cứu này mở rộng ứng dụng của AI trong sáng tạo nội dung số và hỗ trợ tăng cường khả năng cạnh tranh cho các nhà sáng tạo nội dung trong nước.

5. Research Objective/Question and Hypothesis

Những câu hỏi này được thiết kế để thu thập những hiểu biết sâu sắc, những insight thú vị và có ý nghĩa, từ đó giúp chúng ta xây dựng một công cụ hỗ trợ đắc lực cho việc viết kịch bản video TikTok.

5.1. Những yếu tố nào ảnh hưởng đến mức độ phát triển của một tài khoản TikTok?

- **Giả thuyết 1:** Tài khoản có tần suất đăng video cao (ví dụ: 3-5 video/tuần) sẽ có tốc độ tăng follower nhanh hơn tài khoản đăng ít video.
- **Giả thuyết 2:** Tài khoản có video theo *series nội dung* (ví dụ: "Mỗi ngày một mẹo skincare") có tỷ lệ giữ chân follower cao hơn tài khoản không có series.
- **Giả thuyết 3:** Các tài khoản có *nội dung đa dạng* (nhiều chủ đề, nhiều định dạng video) có follower tăng trưởng nhanh hơn các tài khoản chỉ tập trung vào một loại nội dung.
- **Giả thuyết 4:** Những tài khoản có *tần suất đăng video ổn định* (ví dụ: đều đặn mỗi ngày) có tốc độ tăng trưởng tốt hơn tài khoản đăng video thất thường.
- **Giả thuyết 5:** Tài khoản đăng video theo *khung giờ "vàng"* (ví dụ: 19h-22h) có lượt tương tác cao hơn tài khoản đăng video vào giờ ngẫu nhiên.

5.2. Những yếu tố nào trong video sẽ ảnh hưởng đến lượt xem và độ tương tác?

Về thời lượng:

- **Giả thuyết 1:** Video có thời lượng ngắn (khoảng 10-30s) có xu hướng được xem nhiều hơn so với video dài.
- **Giả thuyết 2:** Video có độ dài tối ưu khác nhau tùy theo ngành hàng (ví dụ: mỹ phẩm thường ít hơn 15s, giáo dục/nấu ăn thường nhiều hơn 30s).
- **Giả thuyết 3:** Video dài hơn 30s-60s có tỷ lệ tương tác (like, comment) cao hơn do nội dung có chiều sâu hơn so với các video ngắn.

Về âm thanh & hiệu ứng:

- **Giả thuyết 4:** Video sử dụng âm thanh trending có cơ hội viral cao hơn so với video không dùng âm thanh phổ biến.
- **Giả thuyết 5:** Video có giọng nói (voice-over hoặc nói trực tiếp) có tỷ lệ tương tác cao hơn video chỉ có nhạc nền.

Về thời điểm đăng video:

- **Giả thuyết 6:** Video đăng vào *khung giờ tối (19h-22h)* có nhiều lượt xem hơn do đây là thời điểm người dùng hoạt động mạnh nhất.

- **Giả thuyết 7:** Video đăng vào *ngày cuối tuần (từ thứ 6 đến chủ nhật)* có lượt tương tác cao hơn so với các ngày khác trong tuần.

Về hashtag:

- **Giả thuyết 8:** Video có *hashtag phổ biến (#FYP, #trend, #viral, #xyzbca)* có khả năng lên xu hướng cao hơn.
- **Giả thuyết 9:** Video có *từ 3-5 hashtag* có hiệu suất cao hơn so với video có quá ít hoặc quá nhiều hashtag.

Về caption:

- **Giả thuyết 10:** Caption có *dạng câu hỏi hoặc CTA (Call-to-Action)* (ví dụ: “Bạn nghĩ sao về điều này?”) có tỷ lệ comment cao hơn.
- **Giả thuyết 11:** Caption *ngắn gọn* (dưới 10 từ) có tỷ lệ tương tác cao hơn caption dài.
- **Giả thuyết 12:** Caption có *yếu tố gây tò mò* (ví dụ: “Xem đến cuối để biết sự thật”) giúp tăng tỷ lệ giữ chân.

5.3. Những nội dung nào có khả năng viral cao nhất?

Về thể loại nội dung:

- **Giả thuyết 1:** Video có nội dung *hài hước, meme, thử thách (challenge)* có khả năng viral cao hơn.
- **Giả thuyết 2:** Nội dung *giáo dục/sức khỏe ngắn gọn (fact nhanh, mẹo vặt, kiến thức phổ thông)* có khả năng viral do dễ chia sẻ.
- **Giả thuyết 3:** Video liên quan đến *chuyện cá nhân (storytime, confession, tâm sự)* dễ tạo sự đồng cảm và có nhiều comment hơn.

Về yếu tố cảm xúc:

- **Giả thuyết 4:** Video có nội dung *tích cực, truyền cảm hứng* (câu chuyện thành công, động viên) có tỷ lệ share cao hơn.
- **Giả thuyết 5:** Video có nội dung *gây sốc, tò mò hoặc bí ẩn* sẽ có tỷ lệ giữ chân cao hơn.
- **Giả thuyết 6:** Video tạo *cảm xúc mạnh (cười, khóc, tức giận, đồng cảm)* có khả năng viral hơn so với nội dung trung lập.

Về yếu tố hình ảnh:

- **Giả thuyết 7:** Video có *giao diện bắt mắt, màu sắc tươi sáng* có tỷ lệ giữ chân cao hơn.
- **Giả thuyết 8:** Video có *subtitle hoặc text overlay* có tỷ lệ giữ chân người xem cao hơn video không có chữ.

Về cách dẫn dắt nội dung:

- **Giả thuyết 9:** Video có yếu tố *bắt ngờ ngay từ 3 giây đầu tiên* có tỷ lệ giữ chân cao hơn.
- **Giả thuyết 10:** Video có *câu hỏi mở ở cuối clip* khiến người xem muốn tương tác nhiều hơn.

6. Research Methods

6.1. Thiết kế nghiên cứu và quy trình dữ liệu:

Đề án này thực hiện nghiên cứu theo hướng *quan sát*, thu thập dữ liệu công khai từ TikTok để phân tích hiệu suất nội dung video. Dữ liệu sẽ được xử lý để xây dựng mô hình hỗ trợ tạo kịch bản cho video TikTok.

Quy trình dữ liệu gồm:

1. **Thu thập dữ liệu:** Crawl dữ liệu từ TikTok, bao gồm thông tin về tài khoản, video, lượt tương tác, và trích xuất nội dung video.
2. **Lưu trữ dữ liệu:** Tại local hoặc sử dụng các dịch vụ cloud (như MongoDB).
3. **Tìm hiểu dữ liệu kết hợp tiền xử lý:** Làm sạch dữ liệu, chuẩn hóa thông tin, loại bỏ dữ liệu nhiễu và giá trị ngoại lai. Đồng thời rút trích một số thông tin cơ bản về phân phối của dữ liệu.
4. **Phân tích dữ liệu:** Áp dụng các kỹ thuật phân tích dữ liệu và học máy để trích xuất insight theo phương pháp *hypothesis-driven*.
5. **Xây dựng mô hình:** Áp dụng các insight từ quá trình phân tích vào mô hình ngôn ngữ lớn để cho ra hệ thống hỗ trợ/gợi ý kịch bản cho video TikTok.
6. **Tích hợp vào ứng dụng web:** Deploy mô hình lên web và hiển thị thông tin qua giao diện trực quan để hỗ trợ người dùng tối ưu hóa nội dung TikTok.

6.2. Đối tượng nghiên cứu & Phương pháp thu thập dữ liệu

Vì dự án tập trung vào việc phân tích dữ liệu TikTok để phục vụ người sáng tạo nội dung, đối tượng nghiên cứu chính sẽ là: *các tài khoản TikTok nổi bật và video của họ*.

Phương pháp thu thập dữ liệu:

- Sử dụng *TikTokApi* (một API không chính thức được tạo ra bởi cộng đồng lập trình viên trên thế giới) để lấy dữ liệu từ TikTok. Dữ liệu thu thập được sẽ bao gồm: video, hashtag, caption, chỉ số tương tác và thống kê chung của các tài khoản.
- Dùng *web scraping* để lấy dữ liệu về danh sách thống kê các tài khoản TikTok nổi bật từ một số trang web phân tích influencer hay KOL (Key Opinion Leader).
- *Thu thập dữ liệu thủ công* thông qua hiểu biết, cũng như các trang báo mạng.

Tiêu chí thu thập dữ liệu:

- Các tài khoản và video công khai ở Việt Nam.
- Video ra mắt trong khoảng 6 tháng trở lại.
- Nội dung thuộc nhiều ngành hàng khác nhau (như: đồ ăn, công nghệ, du lịch, mỹ phẩm, v.v.).

Dữ liệu bị loại trừ:

- Video từ tài khoản riêng tư hoặc ở nước ngoài.
- Video thiếu thông tin tương tác.
- Các tài khoản không đăng tải video trong 1 tháng trở lại.

6.3. Phương pháp phân tích dữ liệu

Khái quát:

- Áp dụng phương pháp *hypothesis-driven*, đưa ra các câu hỏi và giả thuyết cần phân tích, sử dụng dữ liệu để đánh giá và rút trích các insight thú vị.
- Sử dụng phương pháp *thống kê định lượng*, kết hợp một số yếu tố *định tính* (ví dụ: phân loại nội dung).

Phương pháp phân tích thống kê và học máy:

- *Thống kê mô tả*: Hiểu về phân bố dữ liệu, giá trị trung bình, trung vị của các biến chính.
- *Phân tích tương quan*: Kiểm tra mối quan hệ giữa đặc điểm video và mức độ tương tác.
- *Phân tích chuỗi thời gian*: Nhận diện xu hướng thay đổi mức độ tương tác theo thời gian.
- *Phân cụm (K-Means)*: Phân nhóm các nhà sáng tạo nội dung và kiểu tương tác của khán giả.
- *Xử lý ngôn ngữ tự nhiên (NLP)*: Rút trích thông tin từ caption và hashtag.

- *Mô hình ngôn ngữ thị giác (VLM)*: Rút trích thông tin từ ảnh/video.
- *Mô hình ngôn ngữ lớn (LLM)*: Rút trích thông tin văn bản được chuyển thể từ âm thanh của video.

6.4. Mô hình hoá phân tích & Xây dựng tính năng ứng dụng

Dữ liệu sau khi thu thập sẽ được phân tích và ứng dụng vào các tính năng chính của hệ thống:

1. Phân tích nội dung video

- *Đặc điểm phân tích*: Hashtag, caption, thời lượng, đặc điểm thống kê, nội dung của video.
- *Phân cụm (Clustering)*: Xác định nhóm nội dung có hiệu suất cao.

2. Gợi ý tối ưu nội dung TikTok

- *Học máy (Machine Learning)*: Đề xuất hashtag, nhạc nền và phong cách nội dung phù hợp.
- *Xử lý ngôn ngữ tự nhiên (NLP)*: Phân tích caption để gợi ý cải thiện.
- *Theo dõi xu hướng*: Gợi ý các nội dung đang phổ biến.
- *Mô hình ngôn ngữ lớn (LLM)*: Áp dụng các insight từ quá trình phân tích dữ liệu để đề xuất cho người dùng tối ưu hoá nội dung TikTok.

3. Hiển thị dữ liệu trên giao diện web

- *Dashboard*: Cung cấp phân tích về hiệu suất video.
- *Biểu đồ trực quan*: So sánh các yếu tố tác động đến khả năng viral của video.
- *Công cụ hỗ trợ sáng tạo*: Gợi ý chiến lược đăng bài và nội dung video.

6.5. Bảo mật dữ liệu & Đạo đức nghiên cứu

Về dữ liệu thu thập:

- Dữ liệu được thu thập từ nguồn công khai, không lưu trữ thông tin cá nhân.
- Chỉ các thành viên phát triển đồ án mới có khả năng truy cập và sử dụng dữ liệu.

Về dữ liệu người dùng:

- Không chia sẻ hay tự ý thu thập thông tin của người dùng khi sử dụng ứng dụng.

6.6. Mục tiêu đầu ra & Ứng dụng thực tiễn

- **Sản phẩm cuối cùng:** Một ứng dụng web hỗ trợ người dùng sáng tạo nội dung cho video TikTok bằng tiếng Việt.
- **Lợi ích:** Cung cấp công cụ phân tích, dự đoán xu hướng và tối ưu nội dung.
- **Khả năng mở rộng:** Có thể áp dụng cho nhiều nền tảng mạng xã hội khác cũng như mở rộng quy mô dữ liệu.

7. Timeline

Giai đoạn	Các công việc chính	Thời gian dự kiến
Tuần 1-2	Topic Selection & Background Research <ul style="list-style-type: none">- Tìm hiểu các thách thức của người sáng tạo nội dung TikTok. Chẳng hạn như: xu hướng thay đổi nhanh, yêu cầu về tính sáng tạo và tương tác.- Khảo sát các công cụ hỗ trợ viết kịch bản hiện có và điểm yếu của chúng.- Xác định khoảng trống nghiên cứu và tiềm năng phát triển của dự án trong ngữ cảnh thị trường Việt Nam.- Nghiên cứu các tài liệu về xử lý ngôn ngữ tự nhiên trong việc tạo kịch bản cho video.- Tham khảo các nghiên cứu về phân tích xu hướng trên mạng xã hội.- Xác định các mục tiêu nghiên cứu cụ thể và phương pháp nghiên cứu.- Xác định phạm vi giới hạn của dự án và các yếu tố ảnh hưởng đến việc thực hiện dự án.- Lập danh sách các nguồn dữ liệu có sẵn hoặc cách thu thập dữ liệu từ TikTok.	2 tuần
Tuần 3-4	Proposal Drafting <ul style="list-style-type: none">- Điều chỉnh lại phần <i>bối cảnh, ý nghĩa, mục tiêu nghiên cứu,</i>	2 tuần

	<p>và câu hỏi nghiên cứu và giả thuyết dựa trên kết quả của giai đoạn nghiên cứu ban đầu.</p> <ul style="list-style-type: none"> - Xây dựng chi tiết phương pháp nghiên cứu, bao gồm cách thu thập, phân tích dữ liệu và các công cụ sử dụng. - Xác định các nguồn lực cần thiết cho dự án như: phần mềm, công cụ, dữ liệu, nhân sự, và chi phí (nếu có). - Xây dựng lịch trình dự án với các mốc thời gian quan trọng, đảm bảo các bước được thực hiện theo đúng kế hoạch. - Đánh giá các rủi ro tiềm ẩn trong quá trình thực hiện dự án và đề xuất các biện pháp khắc phục, cũng như các vấn đề liên quan đến đạo đức nghiên cứu. 	
Tuần 5-6	<p>Environment Setup</p> <ul style="list-style-type: none"> - Thiết lập môi trường Python cho việc xử lý dữ liệu và tính toán, cài đặt các thư viện cần thiết, để đảm bảo tính ổn định và tránh xung đột giữa các thư viện. <p>Data Collection</p> <ul style="list-style-type: none"> - Chọn lọc danh sách các nhà sáng tạo nội dung TikTok dựa trên mức độ tương tác của người dùng với nội dung của họ. Việc này giúp đảm bảo dữ liệu thu thập được có giá trị và phản ánh đúng xu hướng phổ biến trên nền tảng. - Thu thập dữ liệu từ TikTok về các nhà sáng tạo đã được chọn bằng cách sử dụng thư viện TikTokApi. Dữ liệu này bao gồm thông tin về bản thân nhà sáng tạo và các số liệu thống kê liên quan đến video của họ. <p>Data Understanding</p> <ul style="list-style-type: none"> - Tìm hiểu về tập dữ liệu đã thu thập được bằng cách trả lời một số câu hỏi cơ bản: "Dữ liệu có bao nhiêu hàng và bao nhiêu cột?", "Ý nghĩa của mỗi hàng là gì?", "Dữ liệu có các 	2 tuần

	<p>dòng bị lặp không?", "Ý nghĩa của mỗi cột là gì?", v.v.</p> <ul style="list-style-type: none"> - Thông qua việc trả lời câu hỏi này ta có thể phần nào kiểm tra được chất lượng của dữ liệu, xác định được các vấn đề cần giải quyết trong quá trình tiền xử lý dữ liệu. <p>Writing & Documentation</p> <ul style="list-style-type: none"> - Ghi chú lại các phát hiện và các kết quả đã đạt được vào tài liệu. 	
Tuần 7-8	<p>Data Exploration & Preprocessing</p> <ul style="list-style-type: none"> - Tìm hiểu sâu hơn về phân bố của các trường dữ liệu, kiểm tra giá trị thiếu, kiểm tra các giá trị ngoại lai, và xác định các mối quan hệ giữa các biến. - Tiền xử lý dữ liệu: chuẩn hóa, mã hóa, xử lý giá trị thiếu, và chọn lọc đặc trưng. - Áp dụng các mô hình học máy tiên tiến để trích xuất một số thông tin quan trọng từ video TikTok, như: nội dung chính, ngữ cảnh, và cảm xúc của video. <p>Exploratory Data Analysis (EDA)</p> <ul style="list-style-type: none"> - Thực hiện quy trình phân tích khám phá dữ liệu (EDA) để hiểu rõ hơn về đặc điểm của dữ liệu và xu hướng trên TikTok. - Thử nghiệm nhiều giả thuyết khác nhau để làm sáng tỏ câu trả lời cho các câu hỏi nghiên cứu. <p>Writing & Documentation</p> <ul style="list-style-type: none"> - Ghi chú lại các phát hiện và các kết quả đã đạt được vào tài liệu. 	2 tuần
Tuần 9-10	<p>Data Modelling & Model Deployment</p>	2 tuần

	<ul style="list-style-type: none"> - Xây dựng một web app đơn giản nhưng đầy đủ chức năng cần thiết để demo khả năng của trợ lý viết kịch bản cho video TikTok. - Sử dụng các mô hình ngôn ngữ lớn hiện đại để tạo ra các kịch bản phù hợp với xu hướng trên TikTok, dựa trên những thông tin mà người dùng cung cấp (như ngành nghề, mô tả, hoặc yêu cầu cụ thể). - Tận dụng các insights từ giai đoạn phân tích dữ liệu để tối ưu hóa quá trình prompting và tạo kịch bản. <p>Analysis & Validation</p> <ul style="list-style-type: none"> - Đánh giá hiệu suất của mô hình dựa trên các tiêu chí như: độ chính xác, độ phủ, và độ đa dạng của kịch bản. - So sánh kết quả của trợ lý với các công cụ tự động hóa kịch bản hiện có, đánh giá ưu nhược điểm của giải pháp đề xuất. - Xác định các hạn chế hiện tại của sản phẩm để tinh chỉnh quá trình viết prompt cho các mô hình ngôn ngữ lớn. <p>Writing & Documentation</p> <ul style="list-style-type: none"> - Ghi chép các kết quả và phát hiện thu được từ quá trình nghiên cứu và thử nghiệm. - Viết bản báo cáo dự án theo cấu trúc logic (Giới thiệu, Phương pháp, Kết quả, Thảo luận, Kết luận) để truyền đạt toàn diện các phát hiện của dự án. - Tạo các biểu đồ, đồ thị và hình ảnh minh họa giúp trực quan hóa số liệu và kết quả của đồ án. - Đảm bảo trích dẫn đầy đủ, chính xác các nguồn tài liệu, bài báo và báo cáo đã sử dụng trong quá trình làm việc. - Rà soát lại toàn bộ báo cáo để đảm bảo tính rõ ràng, mạch lạc, đồng nhất về ngôn ngữ và định dạng. 	
--	--	--

<p>Tuần 11-12</p>	<p>Presentation Preparation</p> <ul style="list-style-type: none"> - Tạo slide trình bày với nội dung mạch lạc, bao gồm cả hình ảnh, biểu đồ minh họa và video (nếu cần). - Nội dung của slide cần phản ánh toàn bộ quá trình làm việc và kết quả của dự án. Từ việc đặt vấn đề, thu thập dữ liệu, khám phá dữ liệu, xây dựng ứng dụng, đến đánh giá hiệu suất và kết luận. - Soạn sẵn bài nói kèm ghi chú (speaker notes) để đảm bảo các thành viên trình bày rõ ràng. - Luyện tập phát biểu để tự đánh giá và cải thiện thời gian trình bày. <p>Final Review & Submission</p> <ul style="list-style-type: none"> - Rà soát lại toàn bộ báo cáo, code, slide và các tài liệu liên quan để sửa lỗi chính tả, định dạng và logic nội dung. - Kiểm tra tính đầy đủ và nhất quán của tài liệu, đảm bảo các phần trích dẫn, số liệu, kết quả đều chính xác với thực tế. - Xác nhận các yêu cầu của giảng viên và hoàn thiện tài liệu theo đúng hướng dẫn. - Nộp báo cáo, mã nguồn và slide trình bày lên Moodle. <p>Final Presentation & Defense</p> <ul style="list-style-type: none"> - Trình bày kết quả thực hiện đồ án trước giảng viên, sử dụng slide đã chuẩn bị. - Chuẩn bị trả lời các câu hỏi có thể được đưa ra, dự đoán các thắc mắc và luyện tập trả lời trước. - Nhận phản hồi từ giảng viên, ghi chú các điểm cần cải thiện cho các dự án sau. - Thể hiện sự chuyên nghiệp, tự tin và sẵn sàng trao đổi ý kiến trong suốt buổi bảo vệ đồ án. 	<p>2 tuần</p>
-------------------	--	---------------

8. Resources Required

Nhân sự và thời gian làm việc:

– Vai trò và trách nhiệm của từng thành viên trong dự án:

- + Do nhóm 5 chỉ có 3 thành viên (ít hơn 1 thành viên so với hầu hết các nhóm khác), nên mỗi thành viên trong nhóm sẽ phải đảm nhận nhiều vai trò khác nhau và làm việc nhiều hơn để đảm bảo tiến độ dự án.
- + Tất cả thành viên đều cần tham gia vào mọi giai đoạn của quá trình khoa học dữ liệu, từ việc nghiên cứu để hiểu rõ bài toán cần giải quyết, thu thập dữ liệu, khám phá dữ liệu, tiền xử lý dữ liệu, phân tích dữ liệu, xây dựng ứng dụng, đánh giá hiệu suất, đến viết báo cáo và trình bày kết quả.
- + Như vậy, mỗi thành viên cần phải chủ động trong việc tự học, tự nghiên cứu và chia sẻ kiến thức với nhau để hoàn thành dự án một cách hiệu quả.

– Vai trò cụ thể:

Thành viên	Vai trò	Các trách nhiệm chính
Võ Phú Hân	Leader Product Owner Data Engineer Data Analyst ML Engineer	<ul style="list-style-type: none">- Định hướng dự án, xác định rõ mục tiêu và phạm vi nghiên cứu của dự án.- Đặt ra danh sách các câu hỏi lớn cần giải quyết, xác định các giả thuyết và phương pháp nghiên cứu.- Viết mã nguồn hoàn chỉnh để thu thập dữ liệu từ TikTok thông qua các thư viện mã nguồn mở.- Phân tích Khám phá dữ liệu (EDA), trả lời các câu hỏi nghiên cứu bằng cách sử dụng các phương pháp thống kê và trực quan hóa dữ liệu để rút ra những phát hiện thú vị.- Xây dựng ứng dụng demo ở phía backend, xử lý thông tin được truyền từ frontend để viết prompt cho mô hình ngôn ngữ.- Đánh giá chất lượng kịch bản, thu thập case studies từ thị trường TikTok.- Tham gia vào việc viết báo cáo và trình bày kết quả.
Vũ Minh Phát	Project Manager Data Scientist	<ul style="list-style-type: none">- Giám sát tiến độ dự án, tổ chức các cuộc họp để báo cáo tiến độ và phân công công việc cho các thành viên.

	Data Engineer Data Analyst ML Engineer	<ul style="list-style-type: none"> - Hỗ trợ thu thập dữ liệu từ TikTok. - Tổng hợp dữ liệu thô từ các thành viên khác để tạo thành tập dữ liệu hoàn chỉnh. - Thực hiện quy trình khám phá dữ liệu để trả lời một số câu hỏi cơ bản về kích thước, ý nghĩa và chất lượng của dữ liệu. - Thực hiện tiền xử lý dữ liệu, chuẩn hóa, biến đổi và trích xuất các đặc trưng cần thiết cho việc phân tích dữ liệu. - Phân tích Khám phá dữ liệu (EDA), trả lời các câu hỏi nghiên cứu bằng cách sử dụng các phương pháp thống kê và trực quan hóa dữ liệu để rút ra những phát hiện thú vị. - Xây dựng ứng dụng demo ở phía backend, xử lý thông tin được truyền từ frontend để viết prompt cho mô hình ngôn ngữ. - Tham gia vào việc viết báo cáo và trình bày kết quả.
Nguyễn Minh Hiếu	Data Scientist Data Engineer Data Analyst ML Engineer Team Manager	<ul style="list-style-type: none"> - Thiết lập môi trường lập trình, cài đặt các thư viện và công cụ cần thiết cho việc thu thập dữ liệu, phân tích dữ liệu và xây dựng ứng dụng. - Hỗ trợ thu thập dữ liệu từ TikTok. - Phân tích Khám phá dữ liệu (EDA), trả lời các câu hỏi nghiên cứu bằng cách sử dụng các phương pháp thống kê và trực quan hóa dữ liệu để rút ra những phát hiện thú vị. - Xây dựng ứng dụng demo ở phía frontend, thiết kế giao diện người dùng, xử lý thông tin được truyền từ backend để hiển thị kịch bản và tương tác với người dùng. - Tham gia vào việc viết báo cáo và trình bày kết quả.

– **Phân bổ thời gian:**

- + Mỗi thành viên dự kiến dành 10-15 giờ mỗi tuần cho việc nghiên cứu, phát triển và tổng hợp kết quả trong suốt quá trình dự án.

– **Lịch họp để báo cáo tiến độ:**

- + Nhóm sẽ tổ chức 2 cuộc họp cố định vào thứ 3 và thứ 6 hàng tuần để báo cáo tiến độ, thảo luận vấn đề phát sinh và phân công công việc cho giai đoạn tiếp theo.

- + Trong trường hợp cần thiết, nhóm sẽ tổ chức thêm cuộc họp linh hoạt để giải quyết các vấn đề khẩn cấp hoặc thảo luận vấn đề quan trọng, giúp đảm bảo tiến độ dự án.

Ngân sách (nếu có):

- Hiện tại, dự án không yêu cầu ngân sách cụ thể cho việc phát triển và triển khai sản phẩm, vì cả nhóm sẽ ưu tiên sử dụng các công cụ mã nguồn mở và dịch vụ miễn phí để thực hiện dự án.
- Tuy nhiên, nếu cần thiết, nhóm sẽ tự chịu chi phí cho việc thuê máy chủ/đám mây, đăng ký sử dụng các dịch vụ trực tuyến (ví dụ: API từ OpenAI), hoặc mua sách học thuật để hỗ trợ quá trình nghiên cứu. Dự kiến chi phí này sẽ không vượt quá 2 triệu đồng cho toàn bộ dự án.

Technical frameworks và công cụ cần thiết:

- **Ngôn ngữ lập trình:** Nhóm sẽ sử dụng *Python* làm ngôn ngữ chính để thực hiện dự án, vì Python hỗ trợ nhiều thư viện và công cụ hỗ trợ cho việc phân tích dữ liệu và học máy.
- **Công cụ thu thập dữ liệu:** Nhóm sẽ sử dụng thư viện *TikTokApi* để thu thập dữ liệu từ TikTok, bao gồm các thông tin về video và nhà sáng tạo nội dung.
- **Công cụ xử lý dữ liệu:** Nhóm sẽ sử dụng *Pandas*, *NumPy*, *Scikit-learn* để khám phá, tiền xử lý và phân tích dữ liệu.
- **Công cụ trực quan hóa dữ liệu:** Nhóm sẽ sử dụng *Matplotlib*, *Seaborn*, *Plotly* để tạo các biểu đồ, đồ thị giúp trực quan hóa dữ liệu và kết quả đạt được.
- **Công cụ để trích xuất audio từ video:** Nhóm sẽ sử dụng thư viện *MoviePy* để trích xuất audio từ video TikTok.
- **Công cụ để chuyển đổi audio sang văn bản:** Nhóm sẽ sử dụng thư viện *SpeechRecognition* để chuyển đổi audio sang văn bản.
- **Công cụ xây dựng web app:** Nhóm sẽ sử dụng thư viện *Streamlit* để xây dựng ứng dụng web demo cho trợ lý viết kịch bản.
- **Công cụ lập trình AI:** Nhóm sẽ sử dụng các API từ *HuggingFace* để truy cập các mô hình AI tiên tiến trong việc xử lý ngôn ngữ tự nhiên và thị giác máy tính để hỗ trợ việc trích xuất thông tin từ video TikTok và tạo kịch bản phù hợp.

Các công cụ và tài liệu hỗ trợ chuyên ngành:

- Truy cập các bài báo khoa học, sách giáo trình và tài liệu nghiên cứu về AI, xử lý ngôn ngữ tự nhiên (NLP) và tự động hóa sáng tạo nội dung.
- Case studies, ví dụ thực tiễn từ các ứng dụng hiện có nhằm đánh giá tính khả thi của dự án, và phân tích những hạn chế của các công cụ hiện nay.
- Các tài liệu hướng dẫn và mã nguồn mẫu từ các dự án tương tự trên GitHub hoặc các trang web chuyên ngành.
- Các công cụ hỗ trợ như Google Colab, Kaggle Notebooks để thực hiện các thử nghiệm mô hình trên nền tảng đám mây.

Một số công cụ tiện ích để quản lý dự án:

- **Zalo:** Chủ yếu được sử dụng để trao đổi thông tin nhanh chóng giữa các thành viên trong nhóm.
- **Google Meet:** Dùng để tổ chức các cuộc họp trực tuyến hàng tuần giữa các thành viên.
- **GitHub:** Dùng để quản lý mã nguồn, theo dõi tiến độ dự án.
- **Google Drive:** Dùng để lưu trữ và chia sẻ tài liệu, dữ liệu, báo cáo và slide trình bày.

9. Study Limitations

Hạn chế về kinh nghiệm khai thác dữ liệu video và văn bản:

- Các thành viên trong nhóm có thể chưa có đủ kiến thức và kinh nghiệm trong việc xử lý và phân tích các loại dữ liệu phức tạp như video và văn bản từ TikTok, điều này có thể ảnh hưởng đến hiệu quả thu thập và xử lý dữ liệu, từ đó tác động tiêu cực đến chất lượng sản phẩm và kết quả cuối cùng.
- **Biện pháp giảm thiểu:** Cả nhóm sẽ song song nghiên cứu và cập nhật kiến thức mới liên quan đến khai thác và xử lý dữ liệu video, văn bản; đồng thời thực hiện các bài thực hành và thảo luận nội bộ để chia sẻ kinh nghiệm, nhằm đảm bảo tiến độ và chất lượng của dự án.

Hạn chế về dữ liệu:

- Dữ liệu thu thập có thể thiếu thông tin hoặc không phản ánh đầy đủ hành vi người dùng, đặc biệt do thuật toán TikTok liên tục thay đổi.
- Các nguồn dữ liệu hiện có có thể không cập nhật kịp thời với xu hướng mới hoặc chứa các thông tin nhiễu, ảnh hưởng đến độ chính xác khi phân tích dữ liệu.

- **Biện pháp giảm thiểu:** Liên tục cập nhật và bổ sung dữ liệu mới, lên lịch thu thập dữ liệu định kỳ vào các thời điểm quan trọng, mỗi tuần hoặc mỗi tháng; kết hợp sử dụng các kỹ thuật tiền xử lý tiên tiến để lọc và chuẩn hóa dữ liệu.

Hạn chế về phân loại và gán nhãn nội dung:

- Việc phân loại các nội dung như "hài hước", "giáo dục", "lifestyle" có tính chủ quan và đòi hỏi gán nhãn cẩn thận, điều này có thể gây ra sai lệch trong quá trình sử dụng mô hình và ảnh hưởng đến chất lượng kịch bản cuối cùng.
- **Biện pháp giảm thiểu:** Áp dụng phương pháp gán nhãn kết hợp giữa tự động và kiểm duyệt thủ công; tổ chức các buổi họp và đánh giá nội bộ để đảm bảo tính nhất quán trong quá trình gán nhãn.

Hạn chế về mô hình ngôn ngữ:

- Mô hình ngôn ngữ có thể hoạt động tốt trên dữ liệu huấn luyện, nhưng không đảm bảo hiệu quả khi xử lý các nội dung hoàn toàn mới, đặc biệt là trong bối cảnh TikTok với nhiều xu hướng "chế" và ngôn ngữ không chính thống.
- Tiếng Việt với nhiều phương ngữ và từ lóng cũng là một thách thức khi mô hình chưa được huấn luyện đủ dữ liệu đa dạng.
- **Biện pháp giảm thiểu:** Áp dụng kỹ thuật fine-tuning, prompt engineering, và huấn luyện lại mô hình với bộ dữ liệu tiếng Việt phong phú; kết hợp các mô hình nhẹ hơn để tối ưu chi phí và hiệu suất khi cần thiết.

Thay đổi liên tục về ngữ cảnh và xu hướng:

- Ngữ cảnh, xu hướng TikTok thay đổi nhanh chóng và không đồng nhất theo thời gian, điều này có thể làm giảm tính khả dụng của kết quả nghiên cứu trong dài hạn.
- **Biện pháp giảm thiểu:** Xây dựng một pipeline cập nhật dữ liệu tự động và định kỳ đánh giá lại kết quả của mô hình; đưa ra các cơ chế thích ứng linh hoạt để điều chỉnh cách hoạt động của mô hình theo xu hướng mới.

Hạn chế về nguồn lực và công nghệ:

- Việc fine-tune hoặc huấn luyện mô hình ngôn ngữ lớn đòi hỏi kiến thức chuyên sâu và tài nguyên tính toán lớn, điều này có thể gây áp lực lên ngân sách và thời gian dự án.
- **Biện pháp giảm thiểu:** Sử dụng các mô hình ngôn ngữ nhẹ hơn khi nguồn lực hạn chế; tận dụng các nền tảng đám mây và dịch vụ thuê ngoài để đáp ứng nhu cầu tính toán; tổ

chức các buổi chia sẻ kiến thức nội bộ nhằm nâng cao khả năng tự học và áp dụng công nghệ của các thành viên.

10. Back-up Plan

Xử lý ngôn ngữ tiếng Việt:

- Nếu mô hình AI gặp khó khăn trong việc nắm bắt những đặc trưng ngôn ngữ và văn hóa tiếng Việt, nhóm sẽ bổ sung dữ liệu huấn luyện tiếng Việt, áp dụng kỹ thuật fine-tuning và prompt engineering nhằm cải thiện hiệu suất mô hình.

Học tập và chia nhỏ kiến thức mới:

- Trong quá trình làm đồ án sẽ có những kiến thức mới mà nhóm chưa từng tiếp xúc (như xử lý ngôn ngữ tự nhiên, mô hình ngôn ngữ lớn, phân tích dữ liệu trên TikTok). Khi này, nhóm sẽ vừa phải học kiến thức mới vừa phải thực hiện dự án, có thể gây ra tình trạng quá tải kiến thức.
- Do đó, nếu phải học và làm quen với những kiến thức chuyên sâu mới cùng lúc, nhóm sẽ phân chia kiến thức thành các phần nhỏ, mỗi thành viên đảm nhận một phần và tổ chức các buổi chia sẻ nội bộ để tổng hợp lại, giúp cả nhóm nắm bắt nhanh chóng những khái niệm phức tạp. Điều này sẽ giúp giảm áp lực và tăng hiệu quả làm việc.

Định hướng phạm vi dự án:

- Nếu phạm vi dự án chưa được xác định rõ ràng, nhóm sẽ tiến hành rà soát và điều chỉnh kế hoạch, đồng thời thực hiện phân tích nguyên nhân để điều chỉnh lịch trình và phạm vi công việc một cách phù hợp.

Phụ thuộc vào API của bên thứ ba:

- Nếu gặp vấn đề khi sử dụng thư viện TikTokApi, nhóm sẽ thu thập dữ liệu từ các nguồn thay thế như báo cáo, phân tích thị trường, diễn đàn chia sẻ kinh nghiệm, hoặc sử dụng dữ liệu mẫu/giả lập; đồng thời cân nhắc hợp tác với các bên thứ ba có sẵn dữ liệu.
- Nếu gặp vấn đề khi gọi API từ các mô hình state-of-the-art ở trên HuggingFace, nhóm sẽ chuyển sang sử dụng các mô hình nhỏ hơn hoặc các mô hình ngôn ngữ nhẹ hơn. Ngoài ra, nhóm cũng có thể cân nhắc việc chạy các mô hình nhẹ hơn trên máy cục bộ để tránh các rủi ro liên quan đến việc gọi API.

Thời gian hoàn thành dự án:

- Nếu thời gian dự kiến không đủ để hoàn thành các công việc được đề ra lúc ban đầu, thì nhóm sẽ linh hoạt điều chỉnh mốc thời gian, *ưu tiên phát triển các tính năng cốt lõi*, và các thành viên đã hoàn thành nhiệm vụ có thể hỗ trợ các thành viên khác để đảm bảo tiến độ chung.

Khả năng làm việc của thành viên nhóm:

- Nếu có thành viên gặp khó khăn do các vấn đề cá nhân như ốm đau hoặc bận việc gia đình, dự án sẽ được sắp xếp lại lịch làm việc, cho phép thành viên đó làm việc vào thời điểm thuận tiện hơn mà không ảnh hưởng đến tiến độ chung của dự án.

Chuyển giao nhân sự:

- Nếu có sự thay đổi đột xuất về thành viên chủ chốt (ví dụ như có thành viên rời nhóm), thì các thành viên còn lại sẽ tổ chức phiên chuyển giao kiến thức và ghi chép tài liệu chi tiết, sau đó phân chia lại công việc để giảm thiểu tác động tiêu cực đến dự án.

Giao diện người dùng (UI/UX) kém:

- Trong trường hợp gặp khó khăn về thiết kế giao diện và quỹ thời gian cho quá trình phát triển sản phẩm không đủ, nhóm sẽ tham khảo các sản phẩm cạnh tranh khác trên thị trường, sau đó chỉnh sửa và thiết kế lại giao diện nhằm tạo sự khác biệt và dành sự tập trung cho việc cải thiện trải nghiệm người dùng.

Chi phí cho dự án vượt quá ngân sách cho phép:

- Nếu chi phí phát sinh trong quá trình thực hiện dự án vượt quá ngân sách dự kiến ban đầu, nhóm sẽ lập danh sách các yếu tố quan trọng và ưu tiên để xác định các phần cần cắt giảm hoặc thay đổi để giảm chi phí. Đối với các thành phần không quá quan trọng, nhóm sẽ cố gắng tìm kiếm các giải pháp thay thế miễn phí hoặc rẻ hơn để giữ ngân sách dự án ổn định. Trong trường hợp không thể giảm chi phí, nhóm sẽ xem xét mở rộng ngân sách dự án nếu cần thiết.

Dữ liệu thu thập được có chất lượng không đạt yêu cầu:

- Nếu dữ liệu thu thập được thông qua TikTokApi không đủ chất lượng hoặc không đáp ứng yêu cầu của dự án, nhóm sẽ xem xét sử dụng phương pháp tạo dữ liệu giả lập (synthetic data) hoặc thu thập thêm dữ liệu từ các nguồn phụ như báo cáo, phân tích thị trường và diễn đàn chuyên ngành.

11. Presentation plan

1. Introduction:

- Giới thiệu tổng quan về hệ thống trợ lý viết kịch bản dựa trên AI cho video TikTok.
- Trình bày bối cảnh ra đời của dự án, những xu hướng mới trong sáng tạo nội dung số và nhu cầu của người dùng tiếng Việt.
- Nhấn mạnh tầm quan trọng của dự án trong bối cảnh hiện tại và lý do lựa chọn đề tài.

2. Research Problem & Objectives:

- Xác định vấn đề chính: Những yếu tố nào ảnh hưởng đến lượt xem, tương tác và khả năng viral của video TikTok?
- Trình bày các giả thuyết đã đề xuất về các yếu tố ảnh hưởng đến lượt xem, mức độ tương tác và khả năng viral của một video trên nền tảng TikTok.
- Nêu rõ mục tiêu nghiên cứu, bao gồm việc phát triển hệ thống hỗ trợ viết kịch bản tự động cho video TikTok bằng tiếng Việt.

3. Methodology:

- Mô tả quy trình thu thập dữ liệu: Các nguồn dữ liệu được sử dụng (case studies, dữ liệu từ TikTok, báo cáo thị trường, v.v.).
- Quá trình khám phá dữ liệu và tiền xử lý: Các bước làm sạch dữ liệu, và trích xuất các đặc trưng quan trọng.
- Phân tích Khám phá dữ liệu (EDA): Sử dụng dữ liệu để trả lời các câu hỏi nghiên cứu, tìm hiểu về xu hướng và đặc điểm của video TikTok, từ đó đưa ra các phát hiện thú vị.
- Trình bày chi tiết về phương pháp xây dựng ứng dụng demo và các kỹ thuật đánh giá hiệu suất.
- Đề cập đến các công cụ và công nghệ sử dụng trong quá trình phát triển (Python, TensorFlow/PyTorch, thư viện NLP, ...).

4. Demo Application:

- Trình bày cách thức hoạt động của trợ lý viết kịch bản: Từ bước phân tích dữ liệu đến tạo ra các kịch bản video tự động.
- Trình chiếu trực tiếp demo ứng dụng, minh họa quy trình tương tác giữa người dùng và hệ thống, bao gồm việc tùy chỉnh, hiệu chỉnh nội dung.

5. Results & Discussion:

- Trình bày kết quả thực nghiệm, đánh giá mức độ chính xác của mô hình khi đề xuất nội dung.
- So sánh giải pháp của dự án với các công cụ tự động hóa kịch bản hiện có, đánh giá ưu điểm và hạn chế.
- Phân tích và thảo luận về những phát hiện quan trọng, các vấn đề gặp phải trong quá trình triển khai và cách khắc phục.

6. Conclusion & Future Work:

- Tóm tắt những kết quả chính của dự án, đánh giá mức độ đạt được mục tiêu nghiên cứu.
- Đánh giá các hạn chế còn tồn tại của hệ thống và khả năng cải thiện.
- Đề xuất các hướng nghiên cứu và phát triển tiếp theo, như mở rộng phạm vi ứng dụng, tối ưu hoá mô hình, tích hợp thêm các tính năng mới hay cải thiện giao diện người dùng.

7. Q&A:

- Chuẩn bị các câu trả lời cho các câu hỏi thường gặp về dự án.
- Đảm bảo mọi thành viên trong nhóm đều chuẩn bị sẵn sàng để trả lời các câu hỏi một cách chính xác và logic.

12. References

- [1] SamurAIGPT. (n.d.). *GitHub - SamurAIGPT/AI-Youtube-Shorts-Generator: A python tool that uses GPT-4, FFmpeg, and OpenCV to automatically analyze videos, extract the most interesting sections, and crop them for an improved viewing experience.* GitHub. <https://github.com/SamurAIGPT/AI-Youtube-Shorts-Generator>
- [2] RayVentura. (n.d.-b). *GitHub - RayVentura/ShortGPT: ShortGPT - Experimental AI framework for youtube shorts / tiktok channel automation.* GitHub. <https://github.com/RayVentura/ShortGPT>
- [3] Davidteather. (n.d.). *GitHub - davidteather/TikTok-API: The Unofficial TikTok API Wrapper In Python.* GitHub. <https://github.com/davidteather/TikTok-API>
- [4] Ai, P. (n.d.). *GPT-4o is a YouTube scriptwriting assistant.* The Prohuman. <https://www.theprohuman.ai/p/gpt-4o-is-a-youtube-scriptwriting-assistant>
- [5] Zulko. (n.d.). *GitHub - Zulko/moviepy: Video editing with Python.* GitHub. <https://github.com/Zulko/moviepy>
- [6] Uberi. (n.d.). *GitHub - Uberi/speech_recognition: Speech recognition module for Python, supporting several engines and APIs, online and offline.* GitHub. https://github.com/Uberi/speech_recognition