

# Generative YouTube

누구나 손쉽게 YouTube 영상 제작하기

진지훈(Jasper Jin)

# 작업물 정보

## 1. Github 주소

[https://github.com/4ix/ai\\_project\\_final](https://github.com/4ix/ai_project_final)

## 2. 프로젝트 기간

23.03.17(금) ~ 23.04.06(목)

## 3. 참여 범위

- OpenAI(ChatGPT, DALL-E) API 활용 및 프롬프트 적용
- 음성 합성 및 최종 동영상 제작
- Flask 서버 및 페이지 구축



진지훈

Jasper Jin

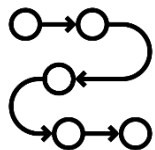
## 추진 배경



### 이젠 검색도 유튜브로 한다

문자보다 이해 쉬워 이용자 급증

2021 인터넷 이용자 조사(NPR)  
네이버(88.1%), 유튜브(57.4%), 구글(48.6%), 다음(25.4%)



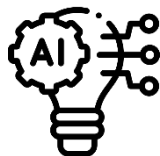
### 영상 제작에 거쳐야 하는 과정들

평균적으로 2천자 분량(대략 5분 길이 영상)의 스크립트를 만들기 위해  
4시간 이상 소요



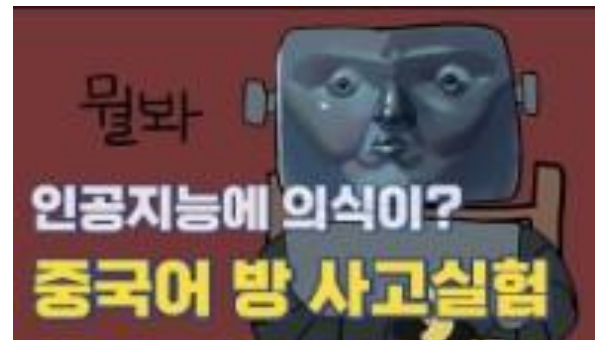
### 쏟아지는 크리에이터 영상

5분 내외의 이미지, 음성 및 자막으로  
이루어진 정보전달형 유튜브 영상



### Generative AI의 등장

문장 자동 생성 및 이미지 생성 기능  
등장



대본 생성에 Generative AI를 활용하면,  
나도 영상을 손쉽게 만들 수 있지 않을까?

## 목적

### 목적

#### **Generative AI를 활용하여 누구나 손쉽게 동영상을 만들 수 있는 서비스 모델 제시**

- 1) 사용자에게 동영상으로 만들고자 하는 주제를 입력 받아 동영상 대본을 생성하는 서비스 인터페이스 구현
- 2) 생성된 대본으로 이미지 생성 및 음성을 합성하는 기능 구현
- 3) 대본, 이미지 및 음성을 동영상 포맷(MP4)으로 저장 후 다운받을 수 있는 서비스 인터페이스 구현

### 배경지식 및 약어

**Generative AI:** 사용자가 AI에게 어떤 것을 만들어 달라고 요구하면, 그 요구에 맞춰서 결과를 만들어내는 인공지능.

**ChatGPT:** OpenAI가 개발한 프로토타입 대화형 인공지능 챗봇. ChatGPT는 대형 언어 모델 GPT-3의 개선판인 GPT-3.5를 기반으로 함.

**GPT:** Generative Pre-trained Transformer. 사전에 훈련된 생성형 자연어 처리 시스템.

**DALL-E:** OpenAI에서 제작한 생성형 이미지 인공지능.

# 개발 범위

## 사용자 주제 입력

Flask 활용 웹 페이지에서 사용자  
주제 입력 및 서버 호출



## 리소스 생성

영상 제작에 필요한 음성 및 이미지  
(Dall-E) 생성



## 생성된 동영상 다운로드

Flask 활용 웹 페이지에서 생성된 동영상을  
다운로드 받을 수 있게 구현



## 동영상 대본 생성

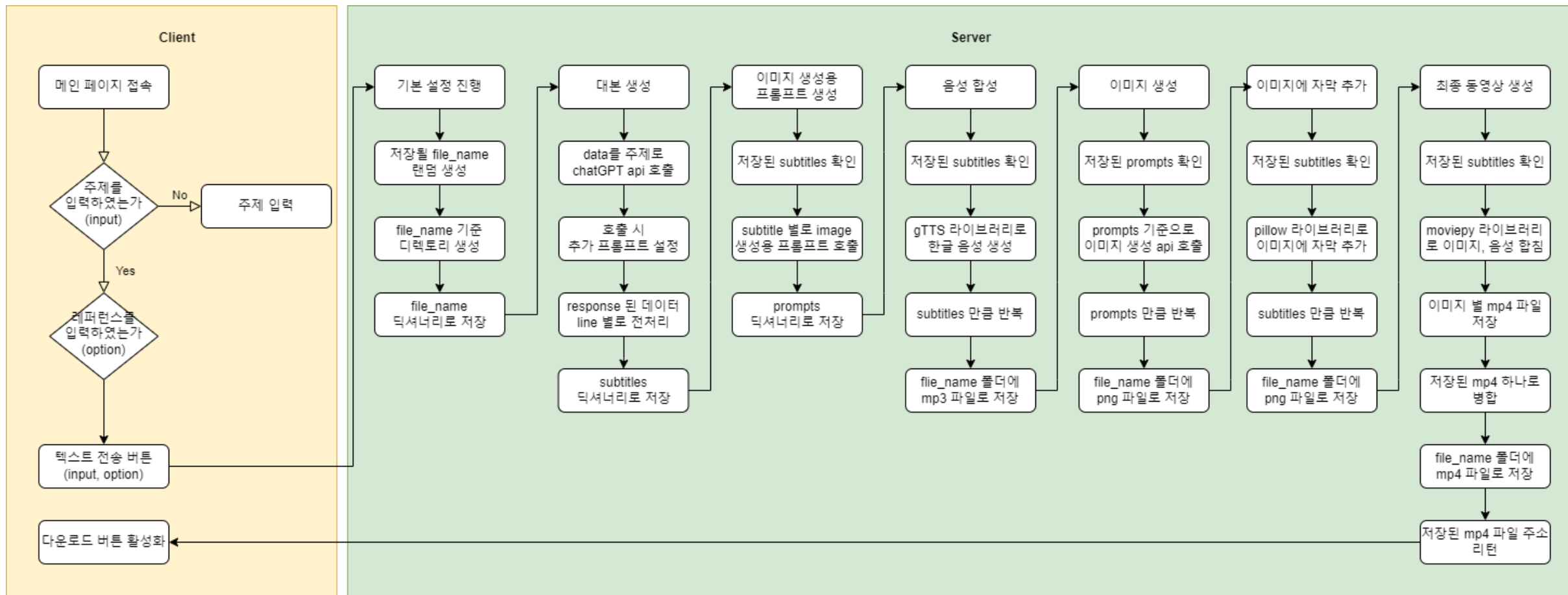
주제를 정보전달용 동영상 대본에 맞게  
프롬프트 적용 후 ChatGPT API 통신



## 동영상 생성

대본을 토대로 이미지에 자막 추가 후 하나의  
유튜브 포맷의 동영상(MP4)으로 생성

# 플로우 차트



\_\_\_\_\_

진지훈	황상진
시작일	03월 17일(금)
종료일	04월 06일(목)

[illegible]

## 중간 결과물 확인

Generative Youtube

127.0.0.1:5000/result?subject=라면이+건강에+끼치는+영향

Generative Youtube

축하합니다!

'라면이 건강에 끼치는 영향'에 대한 영상이 생성되었습니다.

파일 다운로드

- 1. 라면은 너무 많은 나트륨을 포함하고 있습니다.
- 2. 매일 라면을 먹는 것은 건강에 유해 합니다.
- 3. 라면은 인공 조미료를 사용하여 만들어 졌습니다.
- 4. 라면은 고열량이므로 지나친 섭취는 비만을 유발할 수 있습니다.
- 5. 건강한 식습관을 위해서는 라면을 자주 먹지 않는 것이 좋습니다.

기본 설정 시작

기본 설정 완료

한글 자막 생성 시작

한글 자막 생성 완료

영어 번역 시작

영어 번역 완료

음성 파일 생성 시작

음성 파일 생성 완료

이미지 생성 시작 1/5

이미지 생성 시작 2/5

이미지 생성 시작 3/5

이미지 생성 시작 4/5

이미지 생성 시작 5/5

이미지 생성 완료

자막 입히기 시작

자막 입히기 완료

최종 마무리 작업중입니다...

영상 생성 완료!





# 문제점

\*우선순위 [1: 프로젝트 기간 내 해결] [2: 추가 개발 범위]

구분	서비스	문제점	해결 방안	우선순위
사용자	주제 입력 (WEB)	주제 입력 후 진행 상황 확인 불가	페이지 내 로그 추가	1
		일관된 형식으로 대본을 만들고 싶음	레퍼런스 입력 기능 추가	2
		동일한 페이지 접속 시 접속한 사용자 모두 같은 화면이 나옴	세션 기능 추가	2
		생성 버튼 클릭을 계속할 수 있음	한번 클릭하면 disabled 처리	2
	동영상 다운로드	내가 생성한 동영상 내역을 확인하고 싶음	회원관리 기능 개발 필요	2

# 문제점

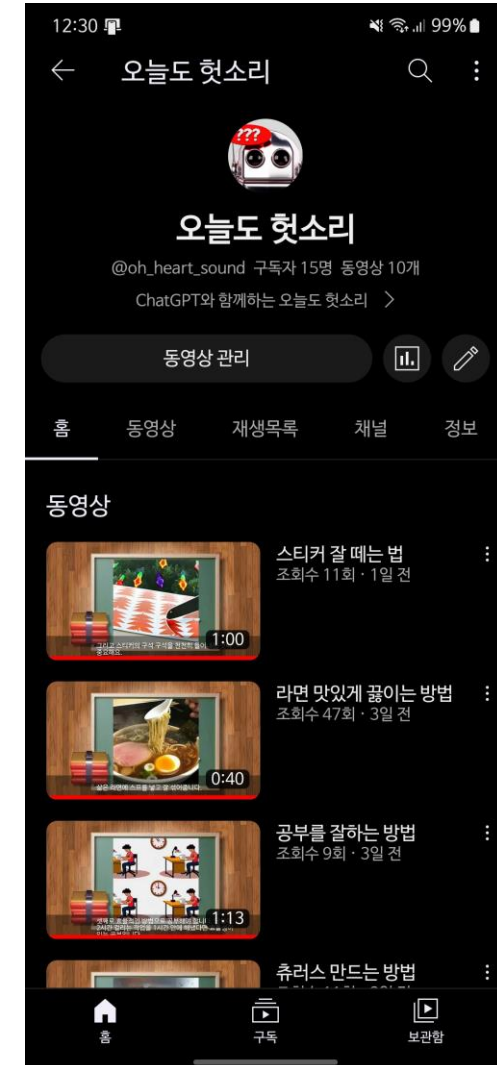
\*우선순위 [1: 프로젝트 기간 내 해결] [2: 추가 개발 범위]

구분	서비스	문제점	해결 방안	우선순위
서버	대본 생성	주제에 따른 일관된 포맷이 없음	추가 프롬프트 설정	1
	음성 생성	음성이 단조로움	추가 음성 합성 모델 확인	2
	이미지 생성	이미지의 퀄리티가 낮음	추가 프롬프트 설정	1
		주제와 전혀 상관 없는 이미지가 나옴	주제도 이미지 생성시 프롬프트에 적용	1
		선정적이거나 폭력적인 주제어는 생성이 안됨	해결 불가	2
	동영상 생성	유튜브 영상 같지가 않음	동영상 해상도 및 동영상 구성 변경	1
		음성 끝 부분이 반복됨	FPS 조정	1

## 최종 결과물



<https://youtu.be/1WOYalyz62c>



# IT '데이터 해적질' 논란부터 할루시네이션까지, 챗GPT 열풍이 부른 논란

챗GPT와 같은 생성 AI가 '만능'이라고 믿어서는 안 된다. AI가 오류가 있는 데이터를 학습할 경우 틀린 답변을 정답처럼 말하는 '할루시네이션(환각) 현상'이 나타날 수 있기 때문이다. 이처럼 AI는 인간의 업무를 돕는 편리하고 유용한 도구가 될 수도 있지만, 편견과 가짜 뉴스, 성적·인종적 편견 등을 포함한 잘못된 정보를 확대·재생산하는 도구로 악용될 가능성도 있는 셈이다. AI가 제공하는 정보에 대해 경각심을 갖고 검증하는 것은 어디까지나 인간의 몫이다.

## GPT-3.5

GPT-3.5 models can understand and generate natural language or code. Our most capable and cost effective model in the GPT-3.5 family is `gpt-3.5-turbo` which has been optimized for chat but works well for traditional completions tasks as well.

LATEST MODEL	DESCRIPTION	MAX TOKENS	TRAINING DATA
gpt-3.5-turbo	Most capable GPT-3.5 model and optimized for chat at 1/10th the cost of text-davinci-003. Will be updated with our latest model iteration.	4,096 tokens	<u>Up to Sep 2021</u>

## 프로젝트 후기

### 새로운 기술 및 API 적용의 성취감

단순히 ChatGPT나 Dall-E를 서비스 제공 페이지에서 사용하는 것을 넘어, 직접 API를 구현해서 서비스로 연결하는 것에 대한 성취감을 느낌

### 개발 환경 통일에 대한 아쉬움

프로젝트 시작 전, Anaconda 가상 환경 설정을 통해 개발 환경 통일을 목표로 했으나, 기본적인 OS가 통일되지 않았음

### 설계에 대한 중요성 인식

최초 기획 시 개발 설계를 제대로 진행하지 않아, 프로젝트 진행 중 구현 방향을 수시로 바뀌어야 했음

### 프로젝트 기간 내 목표 달성?

프로젝트 기간 내 최초 개발 범위를 달성한 것에 대해 다행이라 생각하지만, 오히려 개발 범위를 너무 작게 설정한 것에 대한 아쉬움

## 최종 기대효과



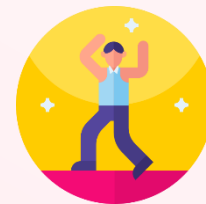
누구나 손쉽게  
빠르게

4시간 이상 걸리던 작업을  
수분 이내로 진행



테크놀로지  
적용

생성 모델 API를 활용한  
새로운 서비스 제안



흥미를 통한  
사용자 유인

짧은 주제어 만으로  
영상을 제작할 수 있다는  
흥미 유발