

# AI 한국어 자동 화면 해설

start

#AI NLP

#구기현

#김성윤

#오정탁

#조병률

#황호성

#화면 해설

#나우유씨미

**1. 화면 해설이란?**

**5. 서비스 탑재 모델**

**2. Problem**

**6. 성능 평가**

**3. Solution**

**7. 서비스 시연 영상**

**4. Project Challenge**

**8. 향후 추가 기능**

---

---

**화면 해설이란?**

---

...

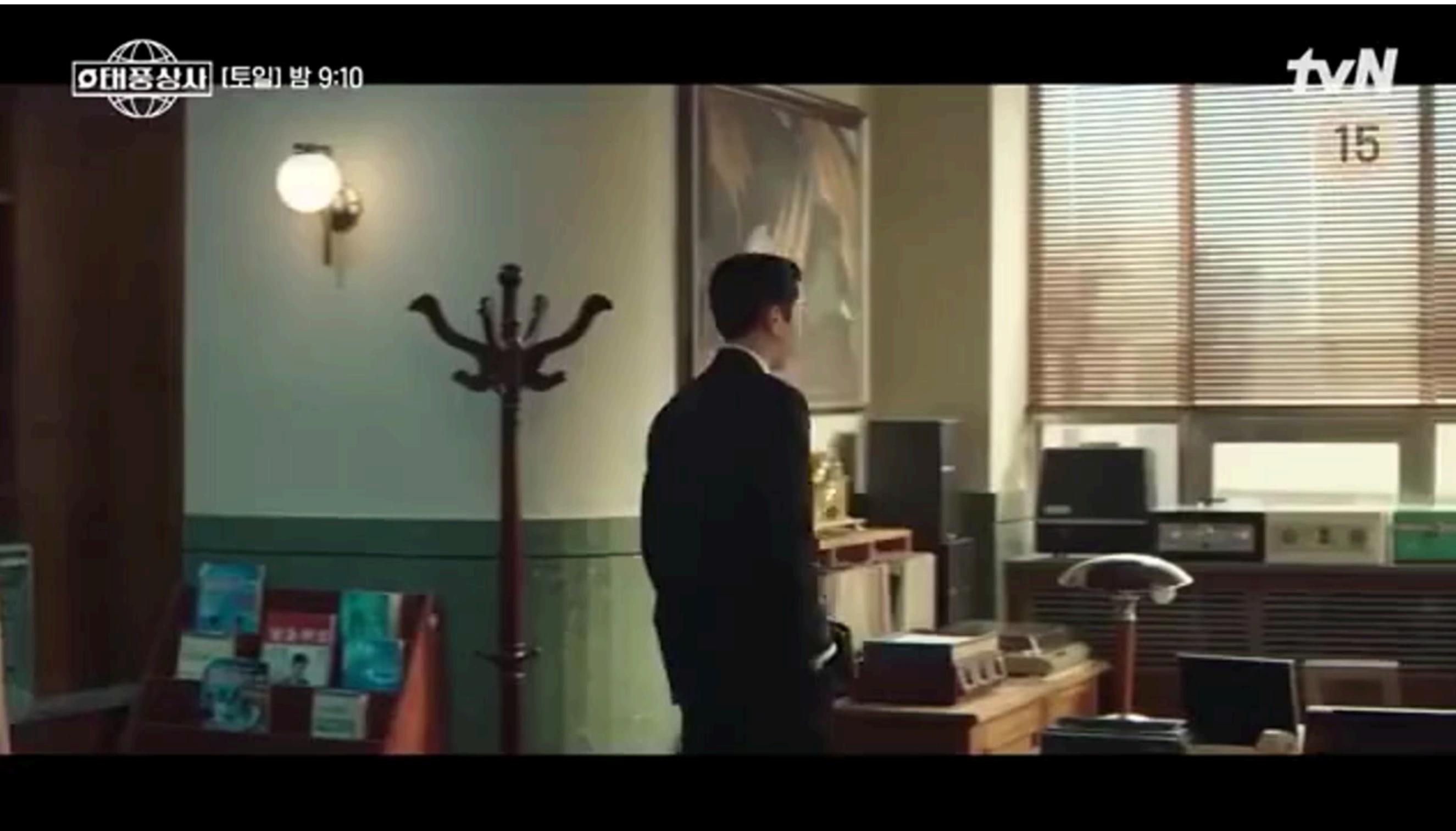
## **저시력자가 영상을 이해하기 위한 도구**

**“화면 해설(audio description)은 시각 장애인이 TV 프로그램을 이해할 수 있도록, 주요 시각 요소를 자연스러운 대사 공백에 삽입하여 들려주는 내레이션이다.”**

---

화면 해설이란?

...



---

# **Problem**

---

# Problem

•••

## 시장 현황

OTT	한국어 AD 지원 수준	특징
넷플릭스	△ (수십 편 수준)	가장 적극적이지만 비중 낮음
디즈니+	△ (영어 AD 중심, 최신 컨텐츠)	글로벌 AD 많으나 한국어 부족
쿠팡플레이	×	AD 지원 사실상 없음
웨이브(Wavve)	△ (약 20편, 확대 예정)	비중 매우 낮음
티빙(TVING)	△ (약 20편, 확대 예정)	비중 매우 낮음

## 국내 시장의 문제점

### 1. 기존 방식의 한계

- 비용&시간 ↑

### 2. 국내 법령 미비

- 화면 해설 의무 국외 O, 국내 X

# Problem

•••

## 기대효과

### 1. 양적 확대

- a. 영상 내 장면 정보, 대사, 무음 구간을 자동 분석하고 AD 초안 대량생산
- b. 이용자가 원하는 어떤 작품에도 AD 제공할수 있는 환경

### 2. 비용 시간 절감&자동화

- a. 편당 제작 단가를 크게 낮추고 리드타임 대폭 단축
- b. AD의 70~90%를 자동생성하고 전문작가는 검수, 편집, QA에 집중하도록 재설계

### 3. 화면 해설 일관성

- a. 사람이 쓸때마다 달라지는 스타일을 일정수준 이상으로 표준화
- b. 사용자 피드백을 학습루프로 투입함으로써, 콘텐츠가 늘어날수록 품질이 고도화되는 구조 설계

### 4. 규제, 정책 대응용 인프라

- a. 규제 준수를 적은 비용으로 달성할 수 있는 솔루션
- b. 법적 리스크관리 + 브랜드 ESG + 사회적 가치 실현

---

# **Solution**

---

# Solution

...

## 화면 해설 생성 기본 원칙

### 객관적 사실

현재에 일어난 일을  
객관적 사실 기술

### 앞뒤 맥락 유지

전 상황과 뒷 상황을  
고려해 맥락 파악

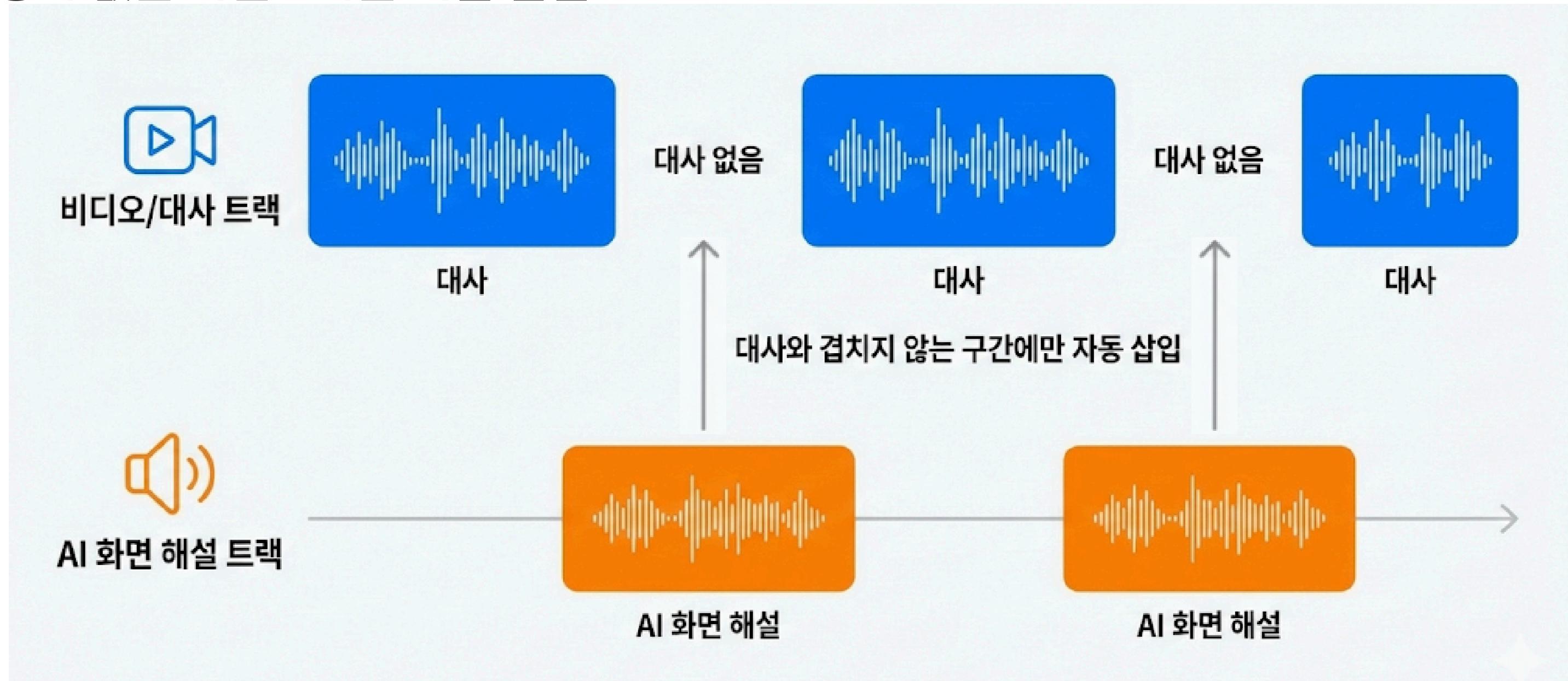
### 대사 중복 생략

대사와 중복되는 화면  
해설 생략

# Solution

•••

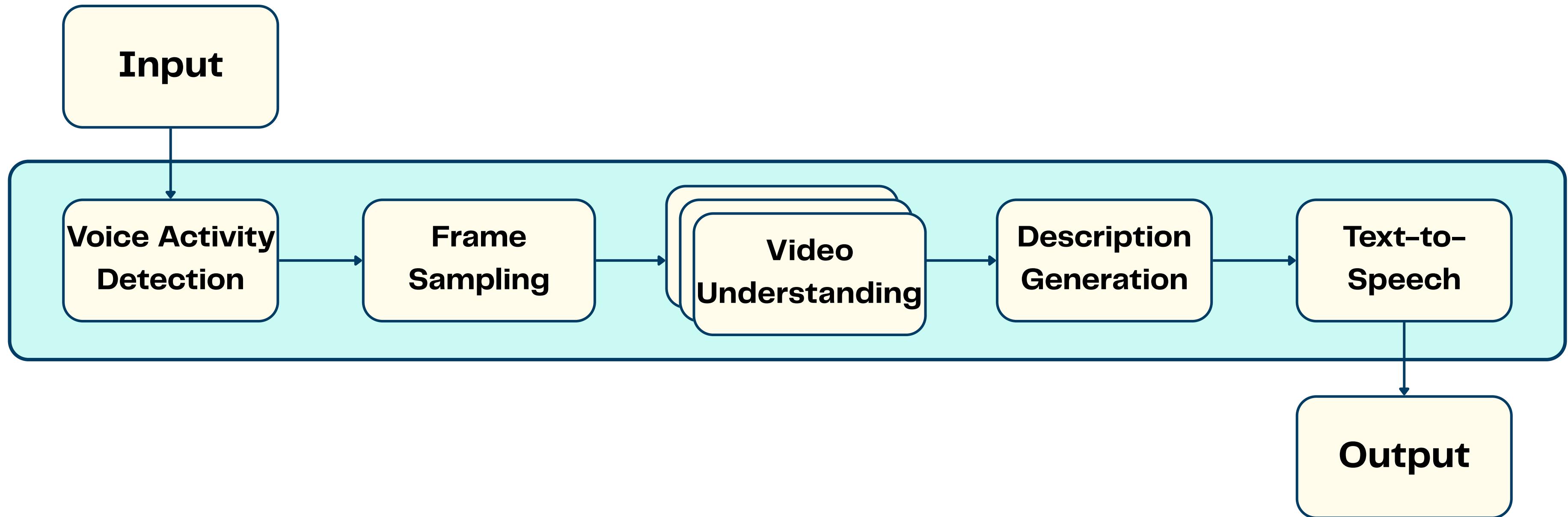
## 일시정지 없는 자연스러운 해설 삽입



# Solution

...

## < Workflow >



---

**Solution**

---

...

**우리 프로젝트는**

**“AI 기반 자동 화면 해설 생성 서비스”**

---

---

# **Project Challenge**

---

---

# Project Challenge

---

...

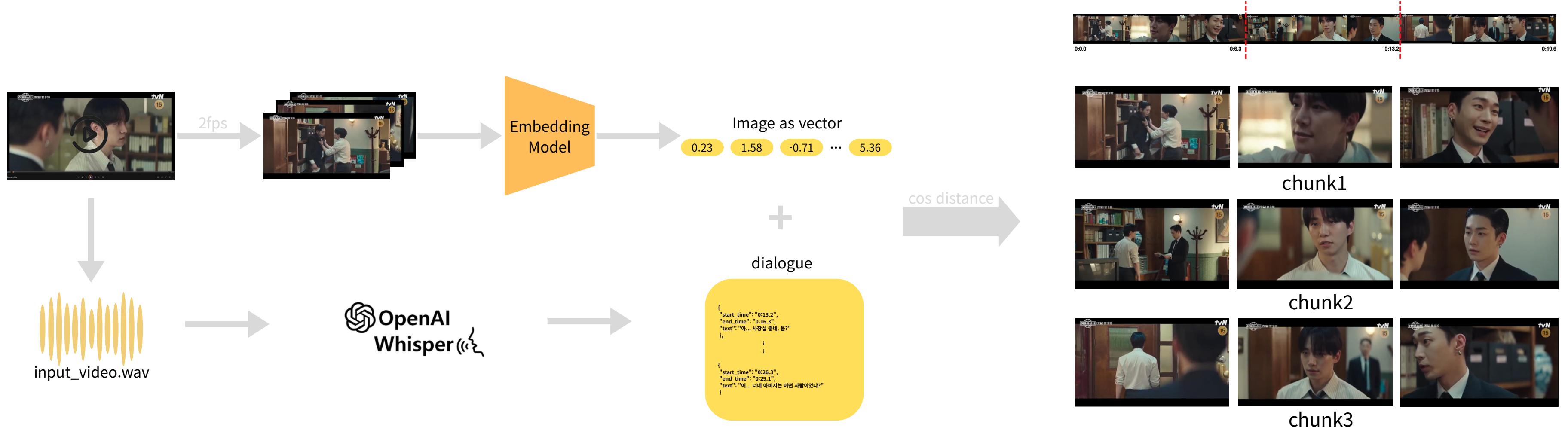
## 1. Chunk 단위로 영상 분할

- 화면 해설을 생성하는 최소 단위
  - 구간이 너무 짧으면 단편적 사실만 담길텐데...
  - 대사가 끊기면 어떻게하지?
-

# Project Challenge

•••

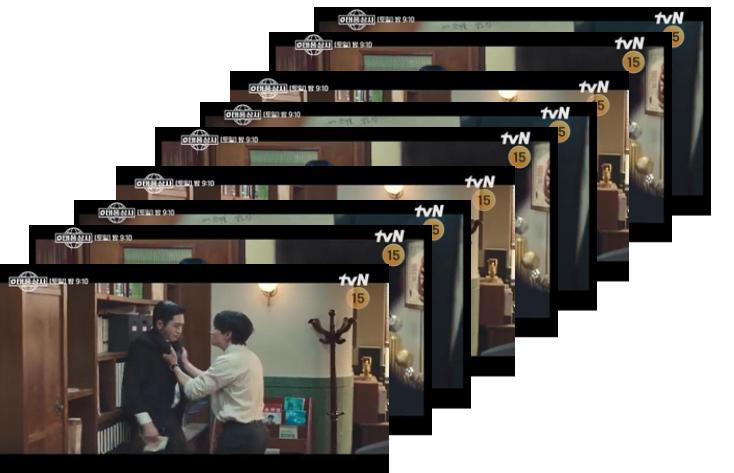
## 1. Chunk 단위로 영상 분할



# Project Challenge

...

## 2. Cut별 주요 이미지 선정



연산량 폭증



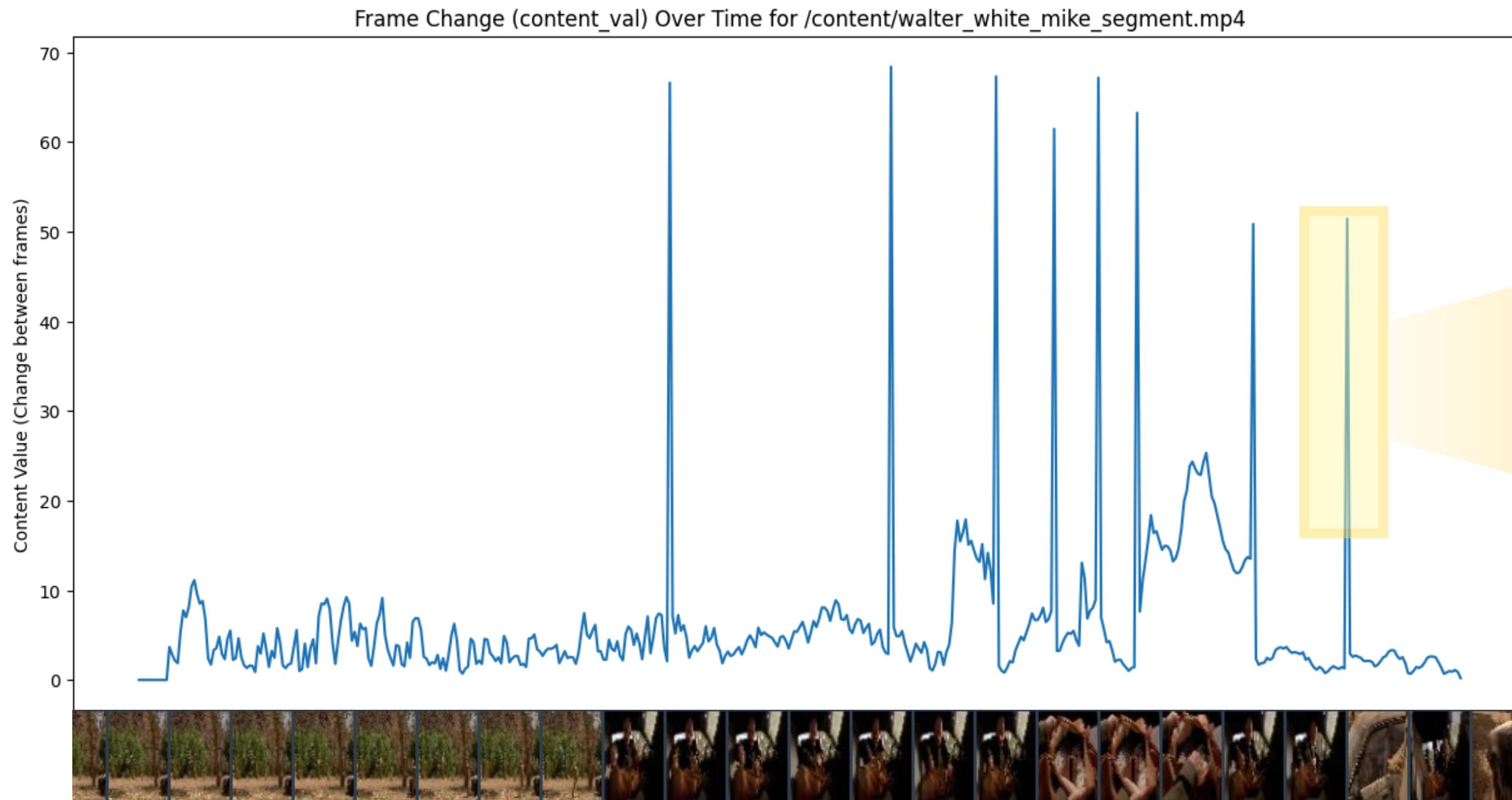
핵심 정보 손실



# Project Challenge

...

## 2. Cut별 주요 이미지 선정



# Project Challenge

•••

## 3. 전체 문맥 이해



**Scene #3 :** 액자 안의 **두 남자**가 있다.



**Scene #4 :** 액자를 보는 정장입은 남성

대사 : 너네 아버지는 어떤 사람이었냐

# Project Challenge

...

## 3. 전체 문맥 이해



Scene #3 : 액자 안의 두 남자가 있다.

맥락 반영



Scene #4 : 남자 2가 액자를 보고 있다.

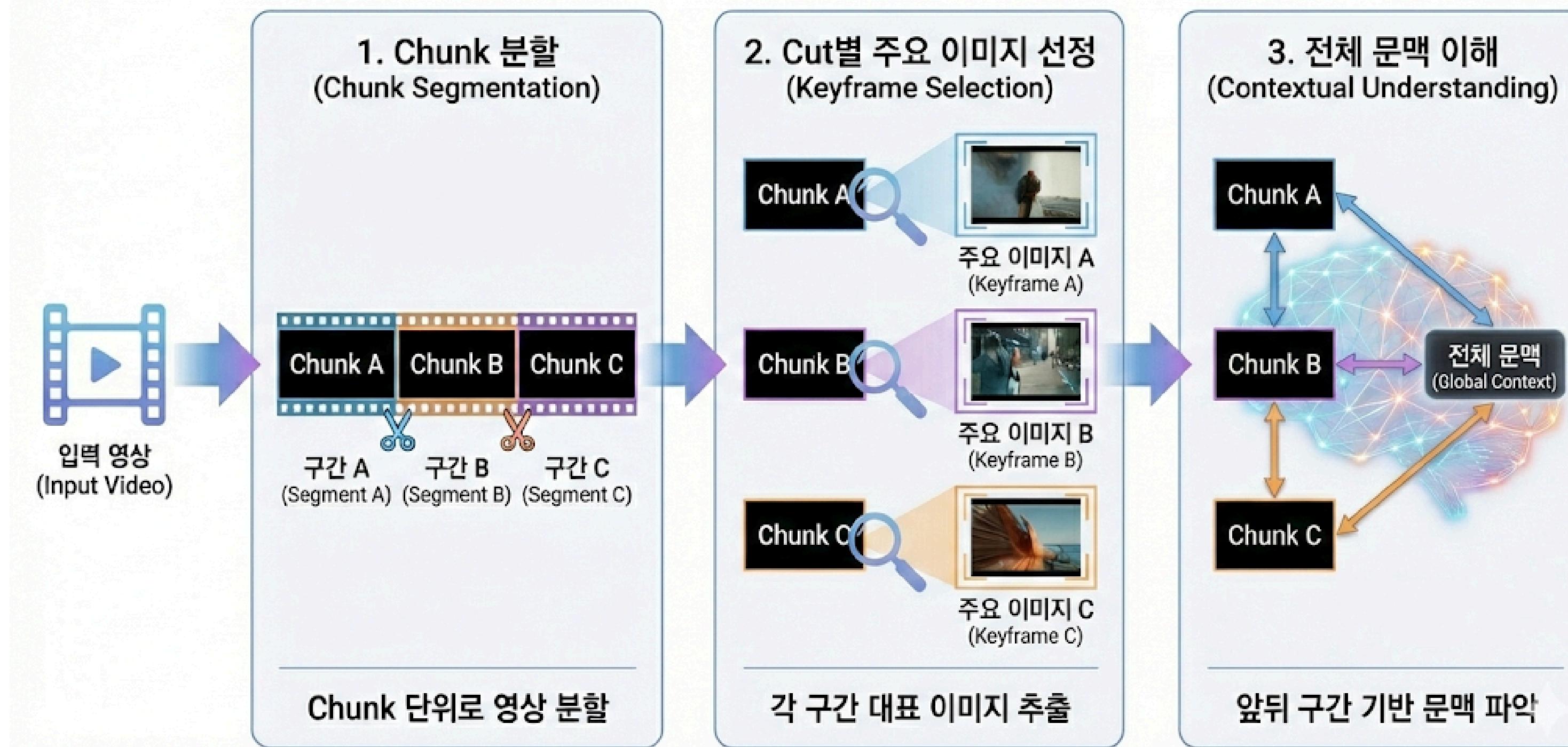
대사 : 너네 아버지는 어떤 사람이었냐

화면 해설: 사진 속에는 남자1과 아버지가 있다.

# Project Challenge

•••

## AI 모델의 영상 이해 단계 (AI Video Understanding Stages)

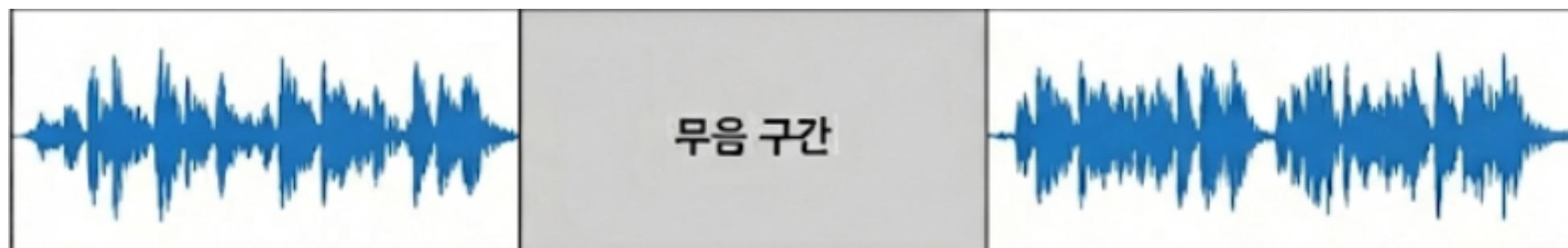


# Project Challenge

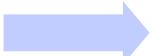
...

## 4. 화면 해설 길이

- 필수 제약: 무음 구간 내에 해설 삽입
  - 길이 초과 시, 대사 및 다음 해설과 소리 겹침 발생



"철수가 분노에 찬 얼굴로 책상 위에 놓인  
유리컵을 거칠게 바닥으로 내던진다."



오버랩

# Project Challenge

•••

## 4. 화면 해설 길이

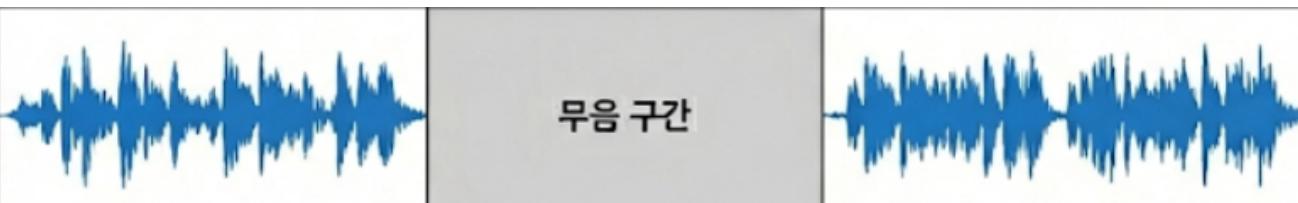
- 화면 해설 압축
  - 길이가 긴 화면 해설을 줄이는 방식

"철수가 분노에 찬 얼굴로 책상 위에 놓인  
유리컵을 거칠게 바닥으로 내던진다."



"화난 철수, 컵을 바닥에 던진다."

- TTS 속도 조절
  - TTS에 배속을 걸어서 속도를 무음 구간 안에 들어갈 수 있게 생성



# Project Challenge

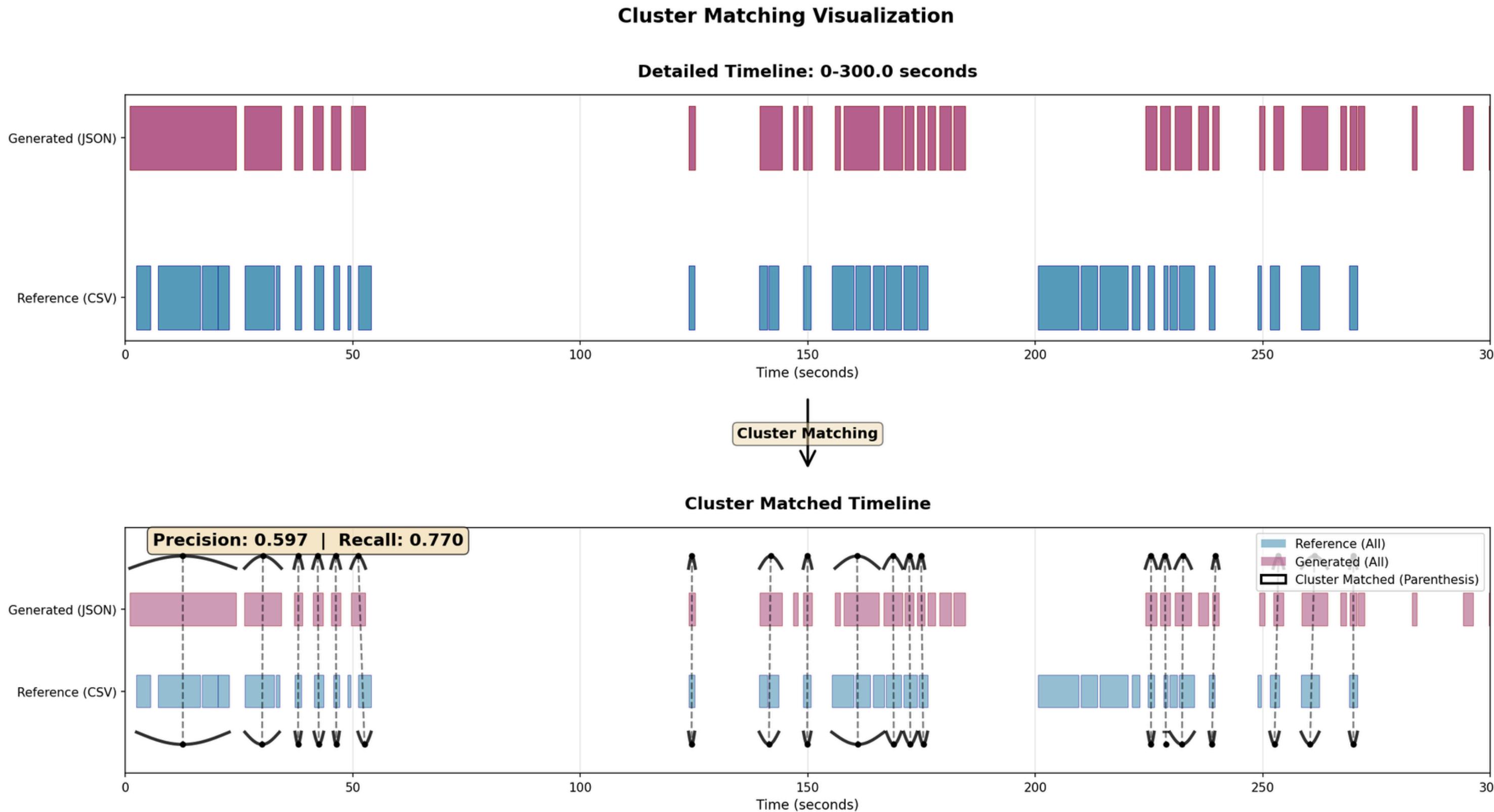
•••

## 5. 평가 지표

Metrics	기반/방식	설명
CIDEr	n-gram + TF-IDF	참조 문장들과 사용된 핵심 단어가 얼마나 비슷한지 측정하는 지표
METEOR	Token alignment	정확·어근·동의어까지 매칭해, 표현이 달라도 의미가 같으면 인정
BERTScore	임베딩 기반 (token-level)	BERT가 계산한 단어 의미 거리로 문장 간 의미 유사도 비교
CRITIC	엔티티·코어퍼런스 매칭	AD 문장에서 등장인물을 제대로 지칭했는지 평가하는 지표
LLM-Eval	LLM 기반 의미 평가	대형언어모델이 문맥·사실성·일관성을 직접 판단하는 평가 방식

# Project Challenge

•••



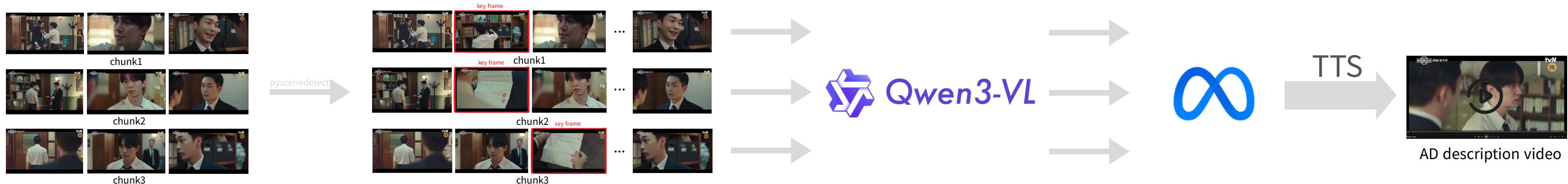
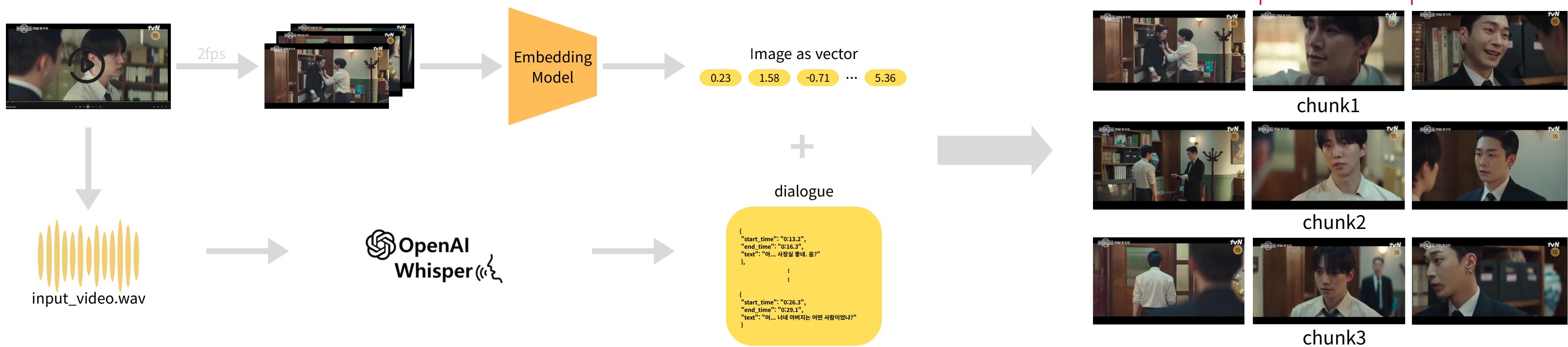
---

## **서비스 탑재 모델**

---

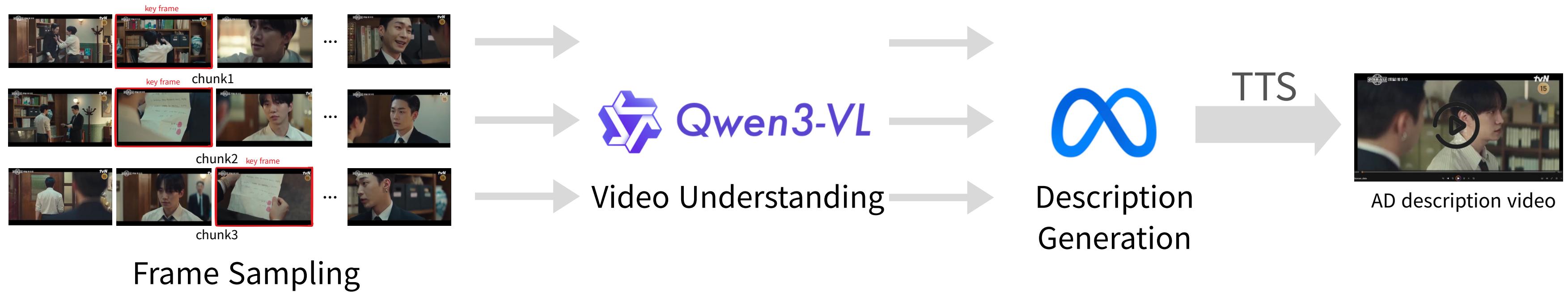
# Cookie 오픈소스 텍쳐

...



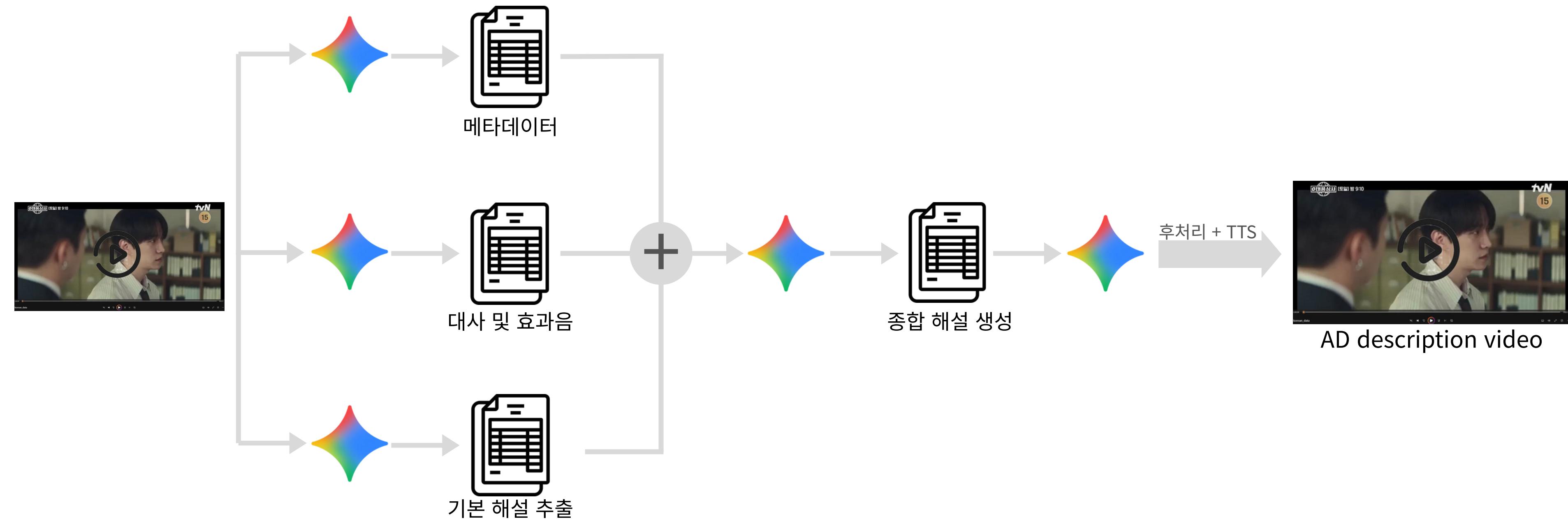
# Cookie 오픈소스

...



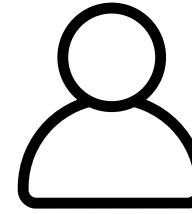
# Jack 아키텍쳐

...



# Gemini-api

•••



""""MISSION: 입력된 비디오를 정밀하게 분석하여 '대사가 없는 구간(침묵 구간)'에 삽입할 화면 해설(Audio Description) 스크립트를 작성하십시오. 모든 결과는 지정된 JSON 형식으로 반환해야 합니다.

:

```
{  
  "full_transcript": [  
    {  
      "time": "0:00.0",  
      "speaker": "[Sound]",  
      "text": "문이 열리고 봄싸움을 하는 소리"  
    },  
    :  
  ],  
  "audio_descriptions": [  
    {  
      "start_time": "0:19.0",  
      "end_time": "0:26.5",  
      "duration_sec": 7.5,  
      "description": "남자2가 사무실을 둘러보며 책상 위에 놓인 액자를 집어 든다."  
    },  
    :  
  ]  
}
```



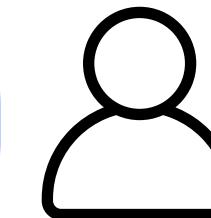
# GPT-api

•••



"""" 당신은 영상 분석 전문가입니다. 주어진 키프레임들과 전체 대본을 분석하여 영상의 전체적인 맥락을 파악해주세요.

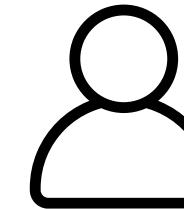
:



```
{  
  "video_info": {  
    "source_file": "96ea5378-4f78-483b-82eb-38388726f22f.mp4",  
    "fps_extracted": 2.0,  
    "language": "ko",  
    :  
  },  
  "video_context": {  
    "known_content": {  
      :  
    }  
  }  
}
```

"""당신은 시각장애인을 위한 화면 해설(Audio Description) 전문가입니다.

```
:  
{{  
    "known_content": {{  
        "is_known": true/false,  
        "title": "작품 제목 (알려진 경우)",  
        "season_episode": "시즌/에피소드 정보 (해당되는 경우)",  
        "description": "작품에 대한 간략한 설명 (알려진 경우)"  
    }},  
:  
}}
```



```
"full_transcript": [  
    {  
        "time": "6.3",  
        "speaker": "Speaker",  
        "text": "안maz"  
    },  
    :  
]  
"audio_descriptions": [  
    {  
        "id": 1,  
        "original_id": 1,  
        "start_time": 0.0,  
        "end_time": 6.306,  
        "duration_sec": 6.306,  
        "description": "남자1이 남자2를 향해 다가가며 말을 건다."  
    },  
    :  
]
```

# 서비스 탑재 모델

...

	 Cookie	 JACK	 Gemini	 ChatGPT
prob 1. 의미 구간 분할	SigLip을 통해 embedding 후 cos 유사도 계산	메타데이터 추출 후 반영	없음 (Native Video Input)	무음구간에 대하여 Hard Cutting
prob 2. 전체 문맥 이해	Llama를 통해 앞 뒤 맥락 파악	메타데이터 추출 후 반영	프롬프트 지시, 이전 내용 요약	영상 전체에 대한 요약 (캐릭터, 분위기, 내용)
prob 3. 키프레임 추출	cut별 대표 이미지 최소 2장 이상	비디오 전체 입력 (내부에서 1FPS 적용)	API 디폴트 설정	2FPS로 이미지 추출
prob 4. 화면해설 길이	Llama를 통해 압축	gemini api를 통한 압축	프롬프트 튜닝	프롬프트 지시사항

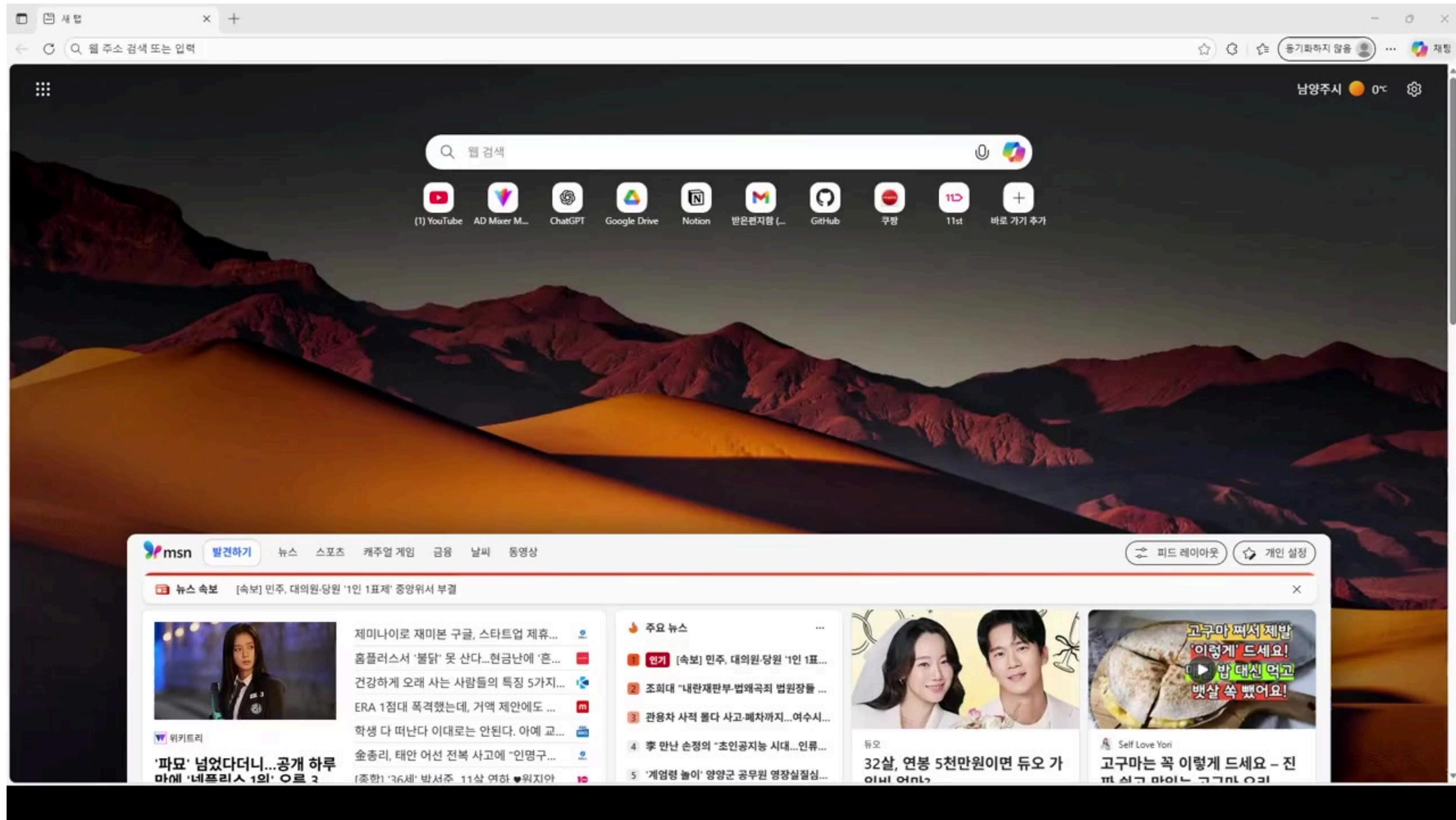
# 성능 평가 지표

...

	 Cookie	 JACK	 Gemini	 ChatGPT
<b>CIDEr (0 ~ 1)</b>	0.0000	<u>0.0713</u>	<b>0.3033</b>	0.0104
<b>METEOR (0 ~ 1)</b>	0.1976	0.1840	<b>0.2529</b>	<u>0.2333</u>
<b>BERTScore (0 ~ 1)</b>	0.8329	<u>0.8814</u>	<b>0.8864</b>	0.8675
<b>CRITIC (0 ~ 1)</b>	<u>0.3834</u>	<b>0.4872</b>	0.0909	0.3569
<b>LLM-Eval (0 ~ 5)</b>	<u>2.0000</u>	1.9231	<b>2.25</b>	1.9565

# 서비스 시연 영상

...



---

## **향후 추가 가능**

---

## 향후 추가 기능

---

...

RAG기반 인물 자동 트래킹

LLMOps를 통해 성능 개선

다양한 언어 추가 예정

유저 접근성 확대

---

• • •

# Q&A

...

**Thank you!**