



아나운서 준비생을 위한 맞춤형 AI 스피치 연습 애플리케이션, Loro(로로)

2024 CAPSTONE TEAM 8

20191579 김필모

20203095 안지원

20203090 신민경

20203110 윤하은

목차

- 문제 정의
- 프로젝트 목표

프로젝트 개요

- 개인화 AI 음성
- 스크립트 생성

아키텍처

향후 추진 계획

서비스 주요 기능

프로젝트 진행 및 관리

프로젝트 개요

문제 정의

비용적



고가의 등록비와 수업료

*기사출처: <https://www.mediaday.co.kr/news/articleView.html?idxno=212098>

“765만원 정도 쓴 거 같아요.”

하늘씨가 아나운서를 준비하며 2년간 나간 학원비는 700만원 이상이었다. 현직 아나운서는 “아나운서가 되기 전 1000만원은 기본으로 쓰고 시작한다”고 말했다. 6개월 정규반(2020년 기준)이 400만원에서 500만원 사이, 고급반은 100만원에서 300만원 사이다. 거기에 공채 시즌마다 프로필 사진과 아나운싱 영상을 찍으려면 기본 30만원 이상이 든다고 설명했다. 사진 3장과 10분 영상을 찍기 위해서는 전문 메이크업샵에서 헤어·메이크업을 받고 의상도 빌려야 하기 때문이다.

아카데미명	강의횟수	수강료	최대정원
투비앤아나운서아카데미학원	50	4,900,000	5
아이비스피치	50	6,000,000	3
봄온 아카데미	40	5,200,000	6
스포티비아나운서스피치아카데미	40	5,200,000	3
아나레슨	10	1,800,000	1
MBC 아카데미	10	1,350,000	1

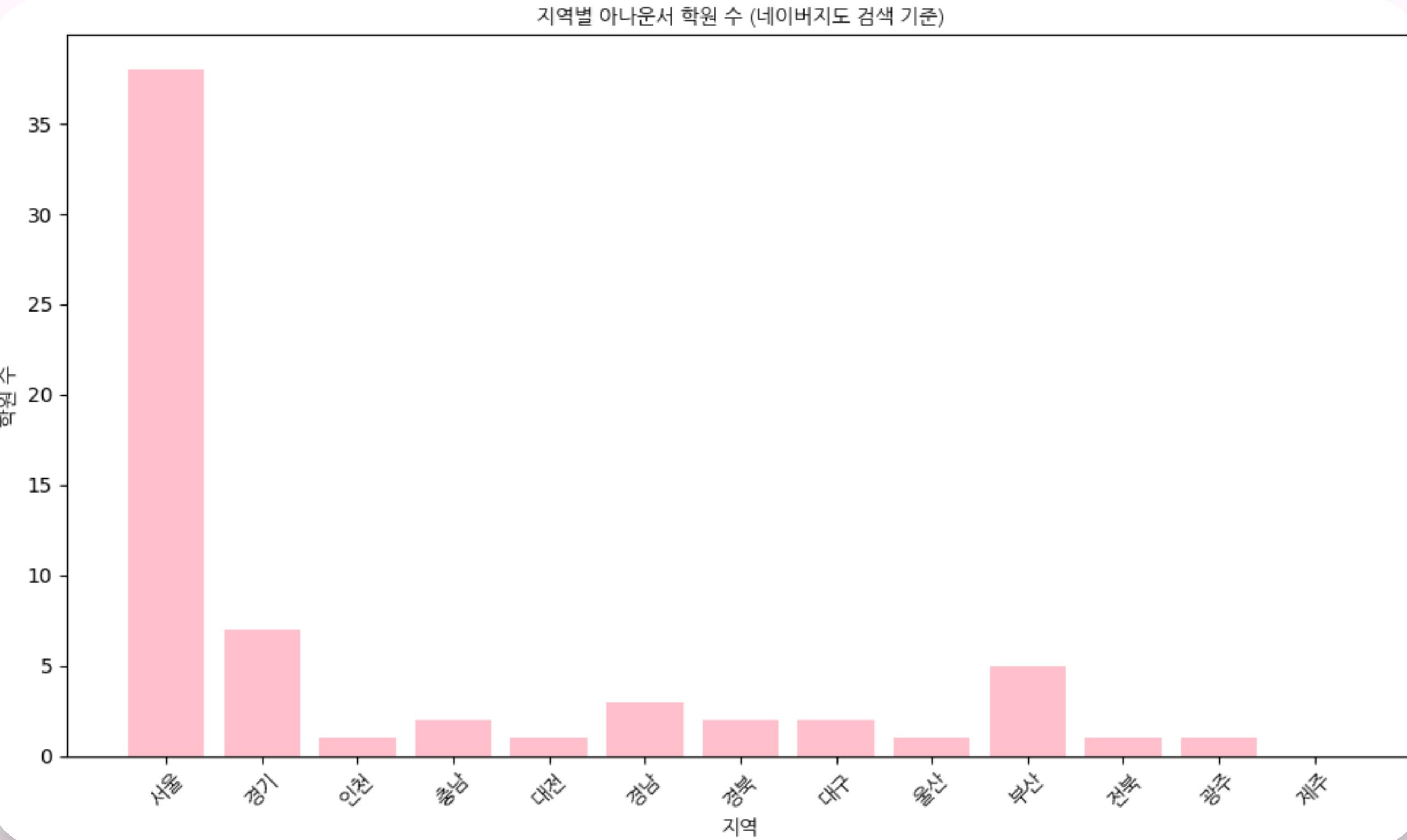
*직접 조사한 결과를 바탕으로 작성한 표(2024년도 3월 기준)

프로젝트 개요 문제 정의

지리적



대도시나 번화가에 위치하여
지방과는 떨어지는 접근성



*직접 조사한 결과를 바탕으로 작성한 그래프(네이버지도 검색 기준)

프로젝트 개요

유사 서비스 현황

차별점

- '아나운서' 직업적 특성
- 사용자 맞춤형 음성 가이드



*직접 조사한 결과를 바탕으로 비교 분석

프로젝트 개요

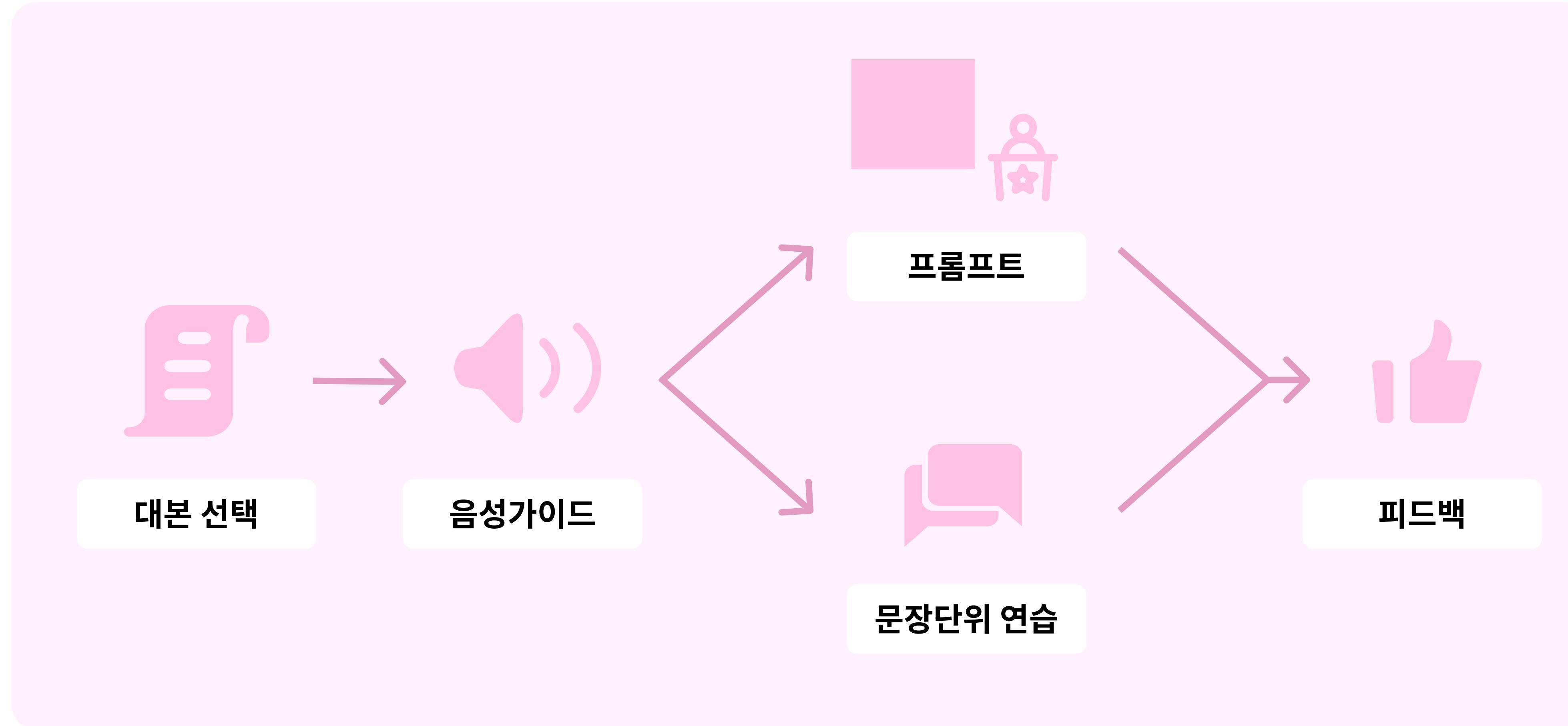
프로젝트 목적

Loro

비용적, 지리적 부담 없이
“Anytime Anywhere”

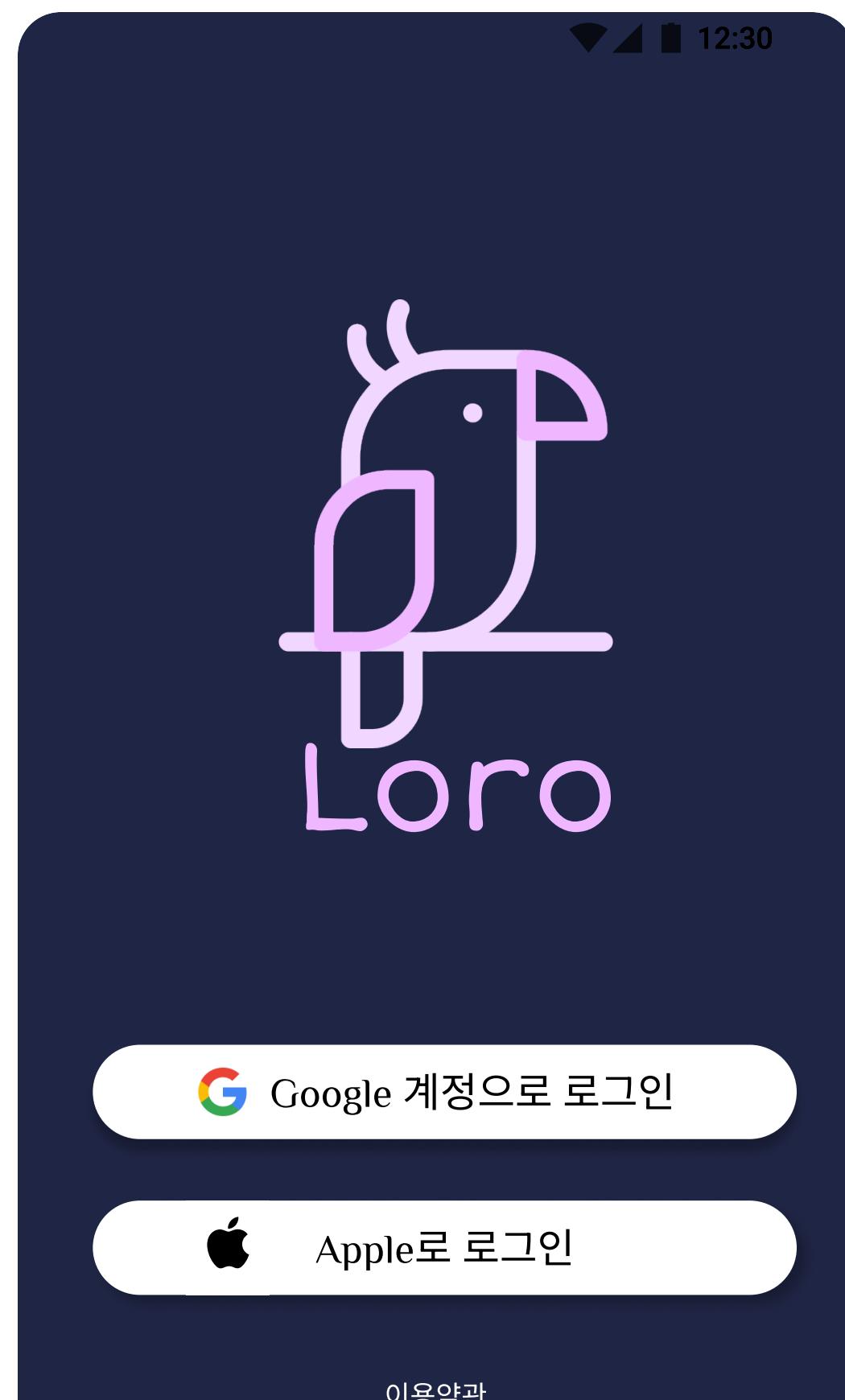
사용자마다 아나운서 억양 개인화 TTS 모델을 만들어 이를 바탕으로
실제 학원에서 배우는 것과 유사하게 발성을 연습하고 피드백 받을 수 있도록 !

프로젝트 주요기능



프로젝트 주요기능

회원가입/로그인



음성 정보 수집

1/3

오늘도 강한 별이 내리쬐는 하루였습니다.

사용자의 음성을 기반으로 더 정확한 피드백을 제공합니다.

다음

음성 정보 수집

2/3

오늘부터 오는 28일까지 한시적으로 사회적 거리두기가 완화됩니다. 수도권의 경우 식당이나 카페 등의 영업시간이 늘어나고 비수도권은 시간 제한이 완전히 풀립니다.

사용자의 음성을 기반으로 더 정확한 피드백을 제공합니다.

다음

음성 정보 수집

3/3

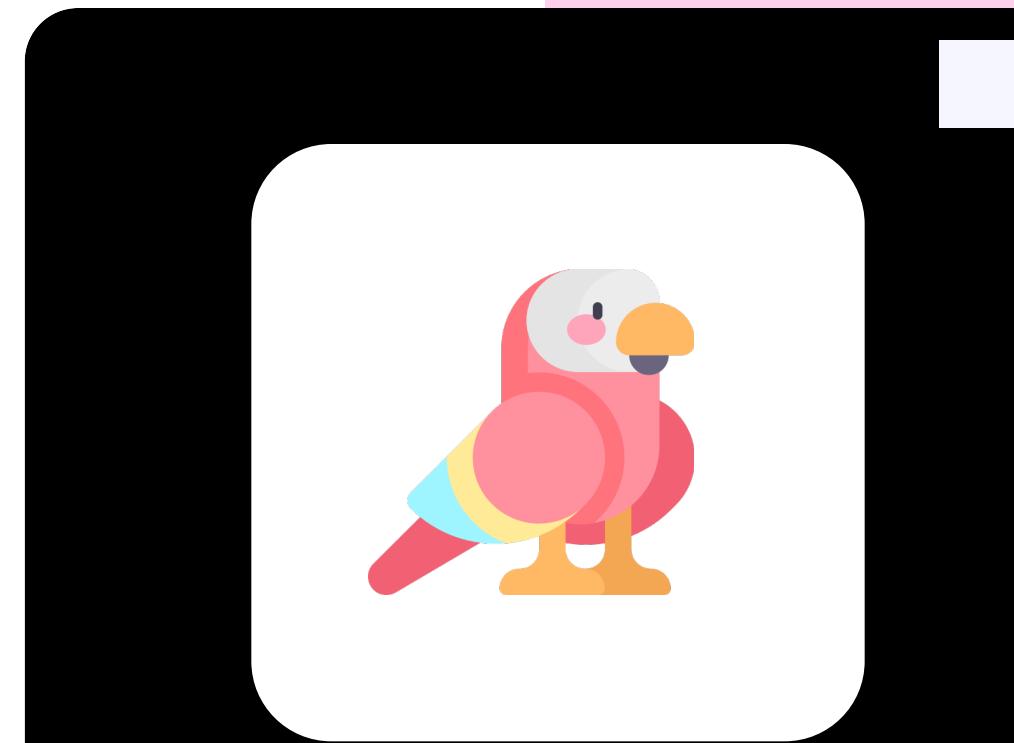
최근 국내 정치와 경제에 대한 관심이 높아지고 있는 가운데, 정책의 변화와 함께 시민들의 관심도 고조되고 있습니다. 특히 정치적으로는 국회에서 논의되고 있는 법안들이 시민들의 이목을 사로잡고 있으며, 정치인들의 발언과 행동 역시 화제를 모으고 있습니다. 경제적으로는 최근의 경기 호조와 함께 산업 부문에서의 변화와 혁신에 관한 보도들이 주목받고 있습니다.

사용자의 음성을 기반으로 더 정확한 피드백을 제공합니다.

완료

프로젝트 주요기능

홈



설정

마지막 연습 기록

마지막으로 연습한 대본

한국 선적 수송선,
일본 앞바다서 전복

(YTN, 재난)

+ 한국인 2명을 포함해 모두 11명이 승선했으며 현재 4명은...

연속 연습일 수

커비님은 12일 연속 연습 중

캐릭터

진행률

기록

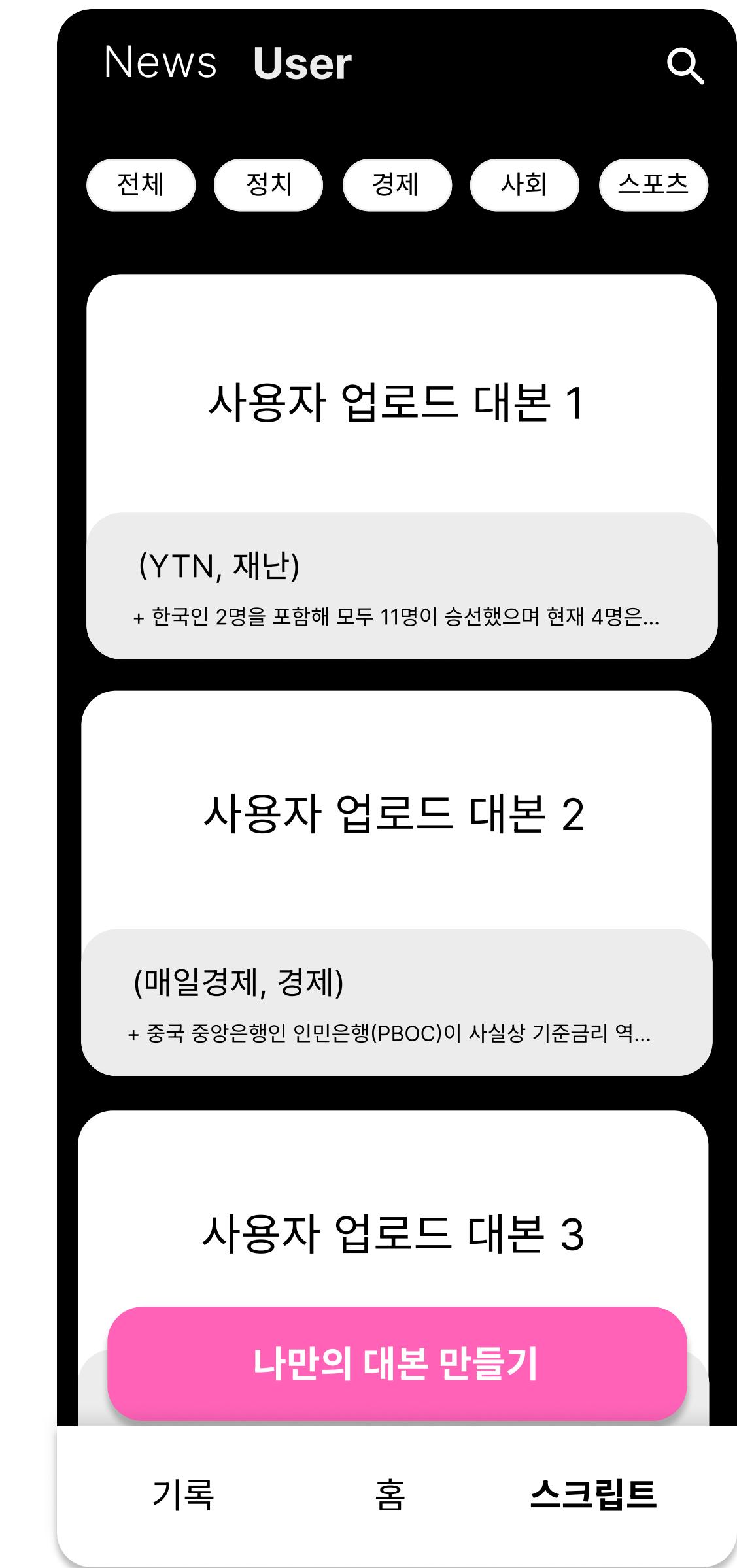
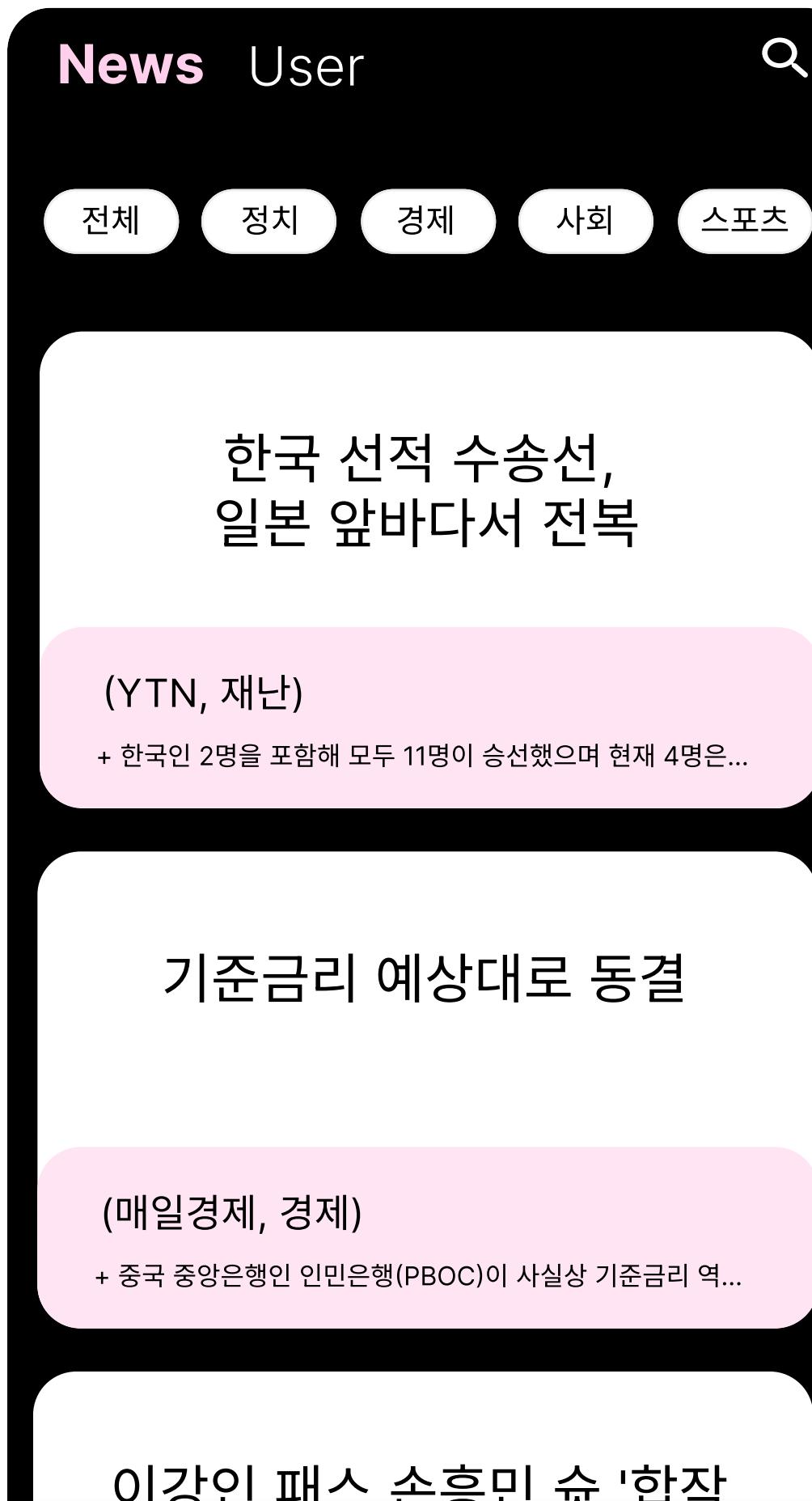
홈

스크립트

네비게이션 바

프로젝트 주요기능

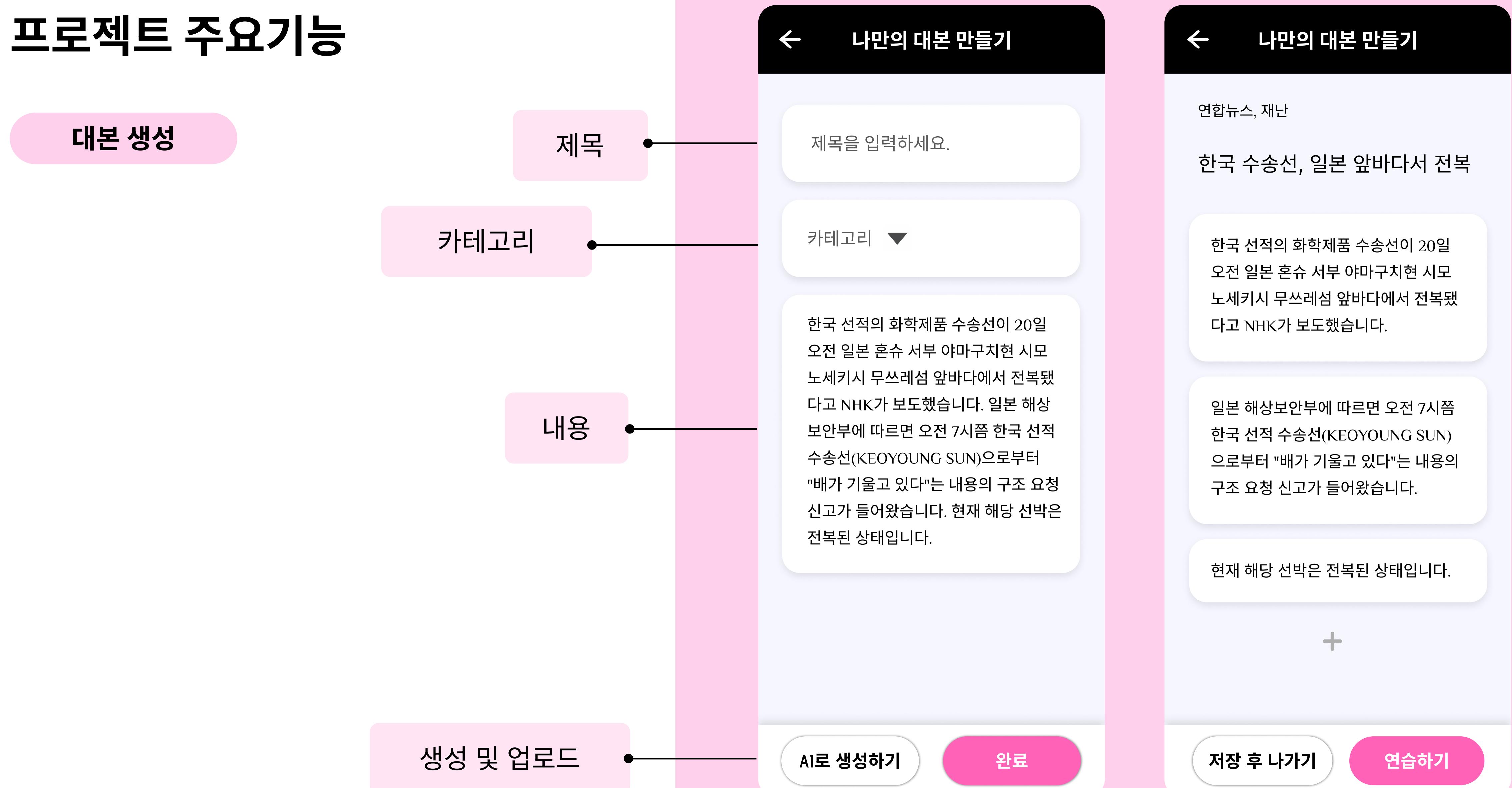
스크립트



카테고리

사용자 대본 생성

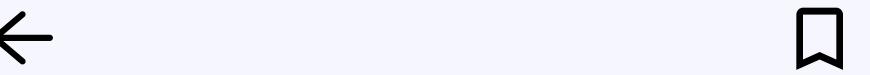
프로젝트 주요기능



프로젝트 주요기능

스크립트

대본 확인 및 연습 방법 선택



연합뉴스, 재난

한국 수송선, 일본 앞바다서 전복

한국 선적의 화학제품 수송선이 20일
오전 일본 혼슈 서부 야마구치현 시모
노세키시 무쓰레섬 앞바다에서 전복됐
다고 NHK가 보도했습니다.

일본 해상보안부에 따르면 오전 7시쯤
한국 선적 수송선(KEOYOUNG SUN)
으로부터 "배가 기울고 있다"는 내용의
구조 요청 신고가 들어왔습니다.

현재 해당 선박은 전복된 상태입니다.

연습하기



다시 연습하게 된다면
이전 음성 데이터 기록이 삭제 됩니다.
그래도 하시겠습니까?

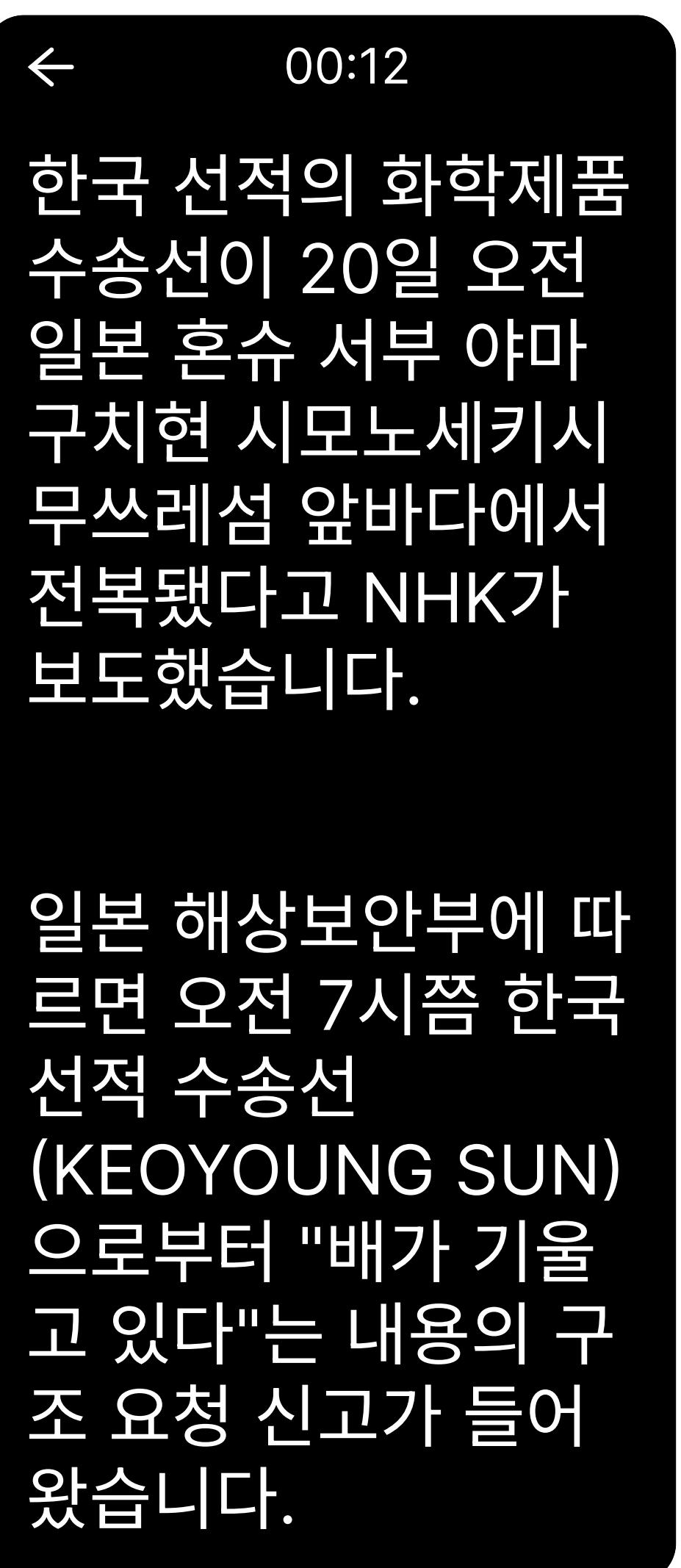
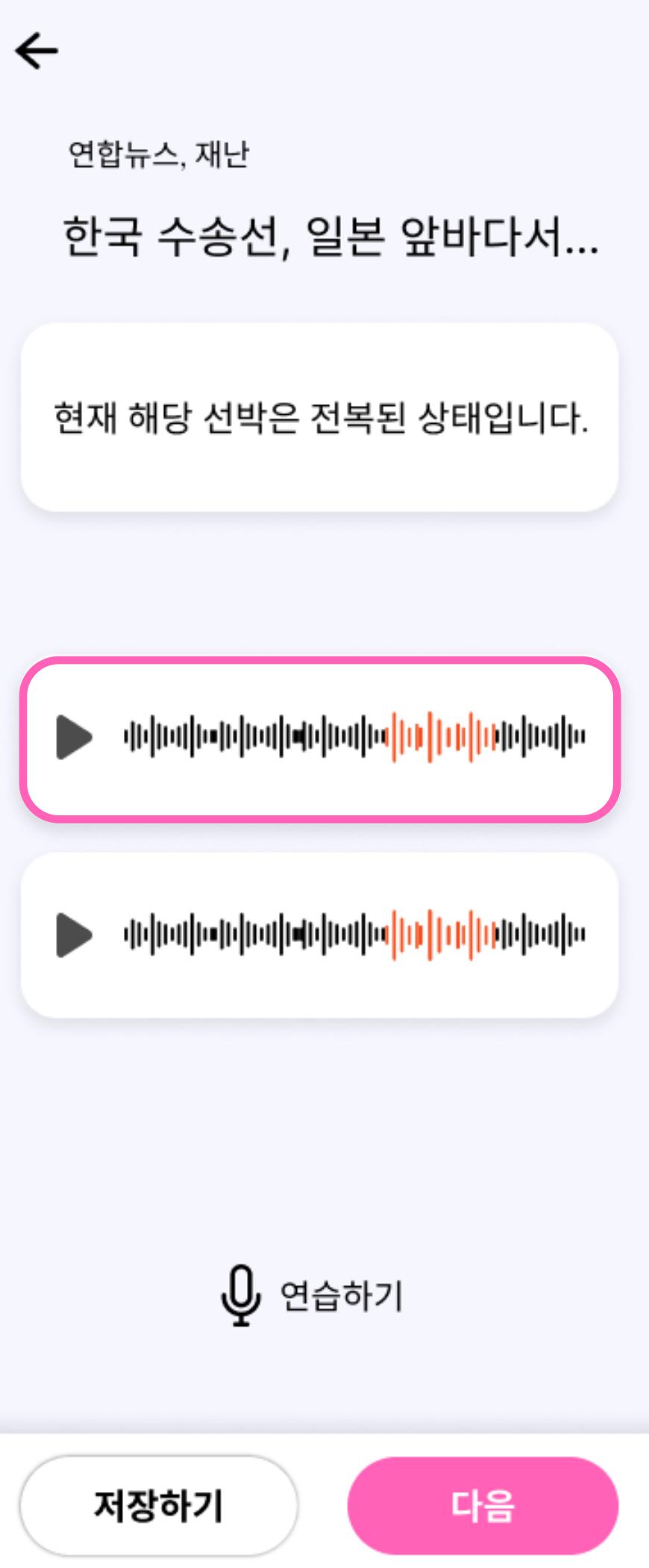
프롬프트

문장단위연습

프로젝트 주요기능

음성 가이드

개인화 TTS모델 이용해
‘사용자 목소리 + 아나운서 특성’ 반영 생성



프로젝트 주요기능

문장단위

한 문장씩 순차적으로 진행하면서
각 문장 집중 연습

가이드 음성



연합뉴스, 재난

한국 수송선, 일본 앞바다서...

현재 해당 선박은 전복된 상태입니다.



저장하기

다음

연습할 문장 표시

녹음된 사용자 음성

문장 단위로 저장

다음 문장으로 이동

프로젝트 주요기능

프롬프트

실제 방송 환경같은 프롬프트에서
대본 전문을 실전처럼 연습

← 00:12

한국 선적의 화학제품
수송선이 20일 오전 일
본 혼슈 서부 야마구치
현 시모노세키시 무쓰
레섬 앞바다에서 전복
됐다고 NHK가 보도했
습니다.

일본 해상보안부에 따
르면 오전 7시쯤 한국
선적 수송선

대본 전문

녹음된 사용자 음성



연합뉴스, 재난

한국 수송선, 일본 앞바다서...

한국 선적의 화학제품 수송선이 20일
오전 일본 혼슈 서부 야마구치현 시모
노세키시 무쓰레섬 앞바다에서 전복됐
다고 NHK가 보도했습니다. 일본 해상
보안부에 따르면 오전 7시쯤 한국 선적
수송선(KEOYOUNG SUN)으로부터
"배가 기울고 있다"는 내용의 구조 요청
신고가 들어왔습니다. 현재 해당 선박은
전복된 상태입니다.



저장하기

다시 연습하기

→ 가이드 음성

프로젝트 주요기능

피드백

음성 가이드와 유사도를 계산해
피드백 제공



연합뉴스, 재난

한국 수송선, 일본 앞바다서...

한국 선적의 화학제품 수송선이 20일
오전 일본 혼슈 서부 야마구치현 시모
노세키시 무쓰레섬 앞바다에서 전복됐
다고 NHK가 보도했습니다. 일본 해상
보안부에 따르면 오전 7시쯤 한국 선적
수송선(KEOYOUNG SUN)으로부터
"배가 기울고 있다"는 내용의 구조 요청
신고가 들어왔습니다. 현재 해당 선박은
전복된 상태입니다.

가이드 음성



사용자 음성



프로젝트 주요기능

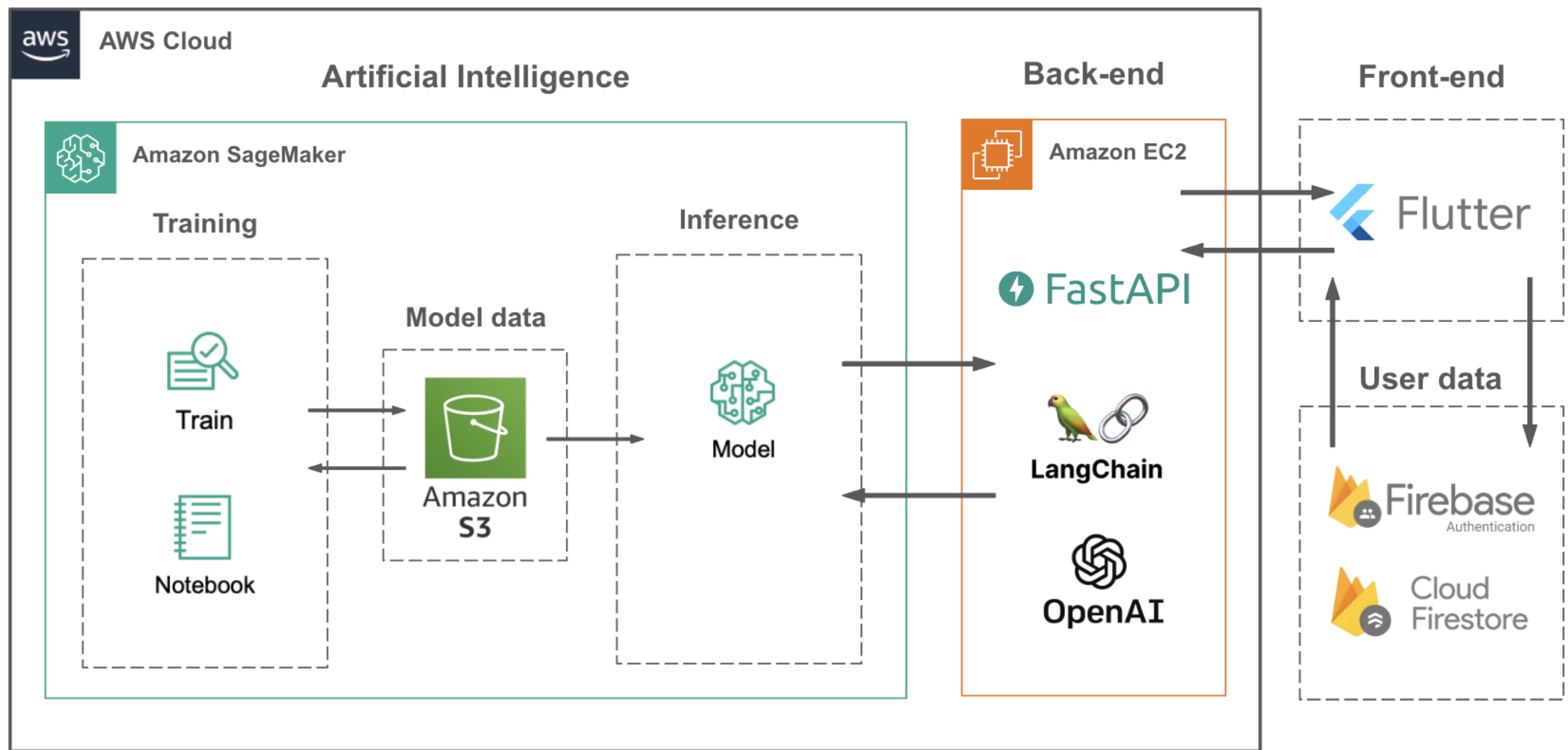
기록

연습한 대본의 점수 및 피드백 확인

저장한 대본 리스트



아키텍처



스크립트 생성

예시 대본 제공

공공데이터포털 데이터셋 사용
<한국언론진흥재단_뉴스빅데이터_메타데이터_언론>



*파일 업로드, 웹 크롤링, 텍스트 분할



*아나운서 대본에 적합하게 가공



생성형 AI로 실시간 대본 생성

사용자가 입력한 제목/카테고리

프롬프트 엔지니어링

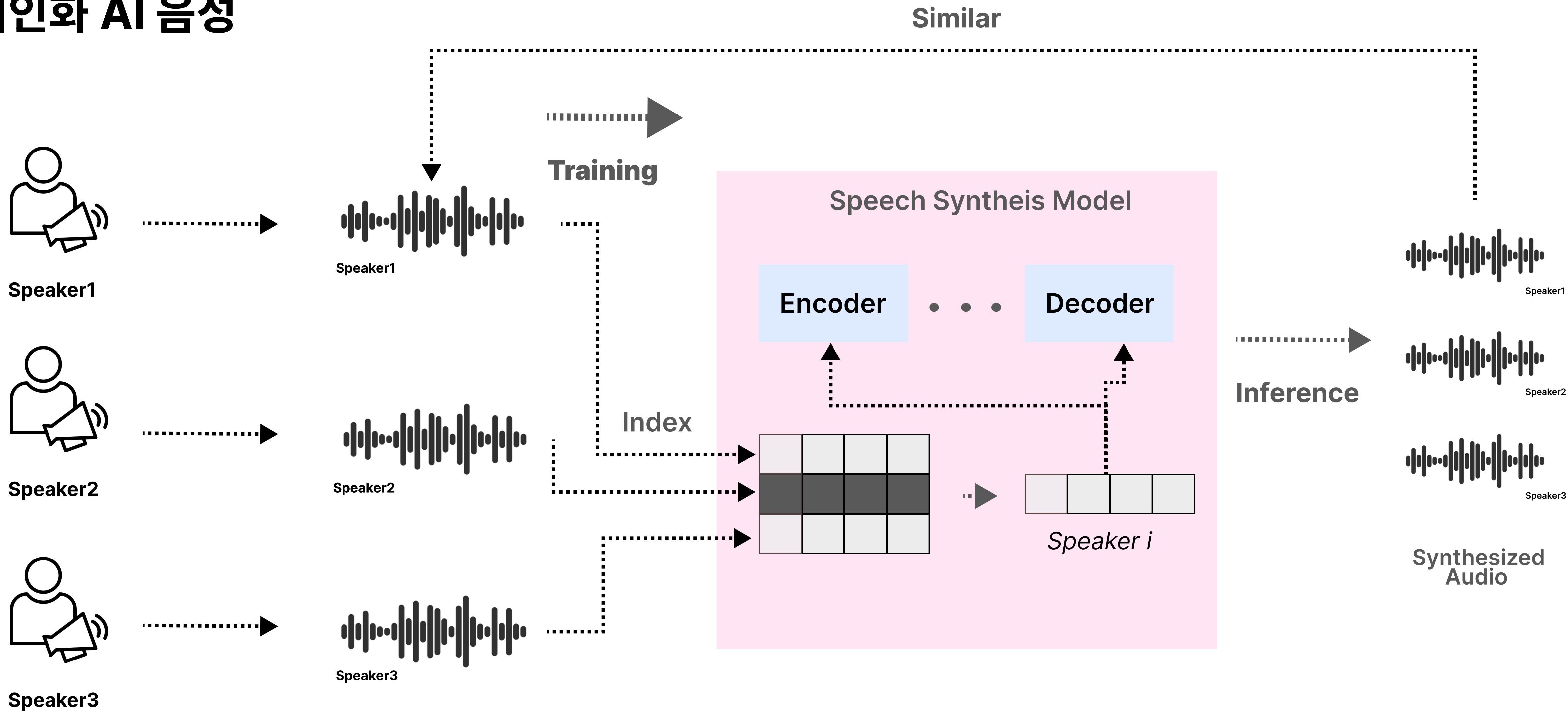


*프롬프트 템플릿 작성



*대본 생성

개인화 AI 음성



개인화 AI 음성

VITS

Matcha-TTs

OverFlow



부분적으로
부자연스러운 발음

AI-Hub 데이터셋 활용

#음성

NEW

뉴스 대본 및 앵커 음성 데이터

분야 한국어 유형 오디오

구축년도: 2022 갱신년월: 2023-11 조회수: 4,381 다운로드: 258 용량: 227.82 GB



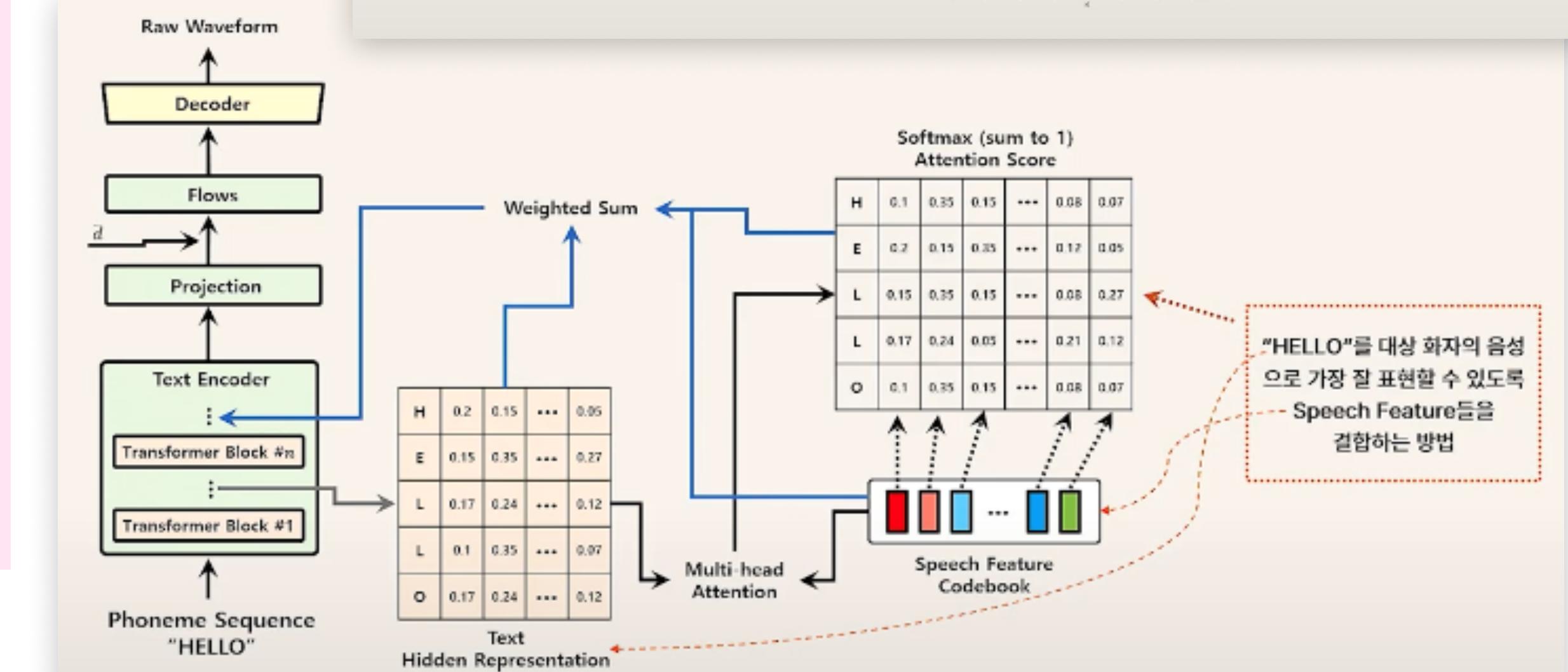
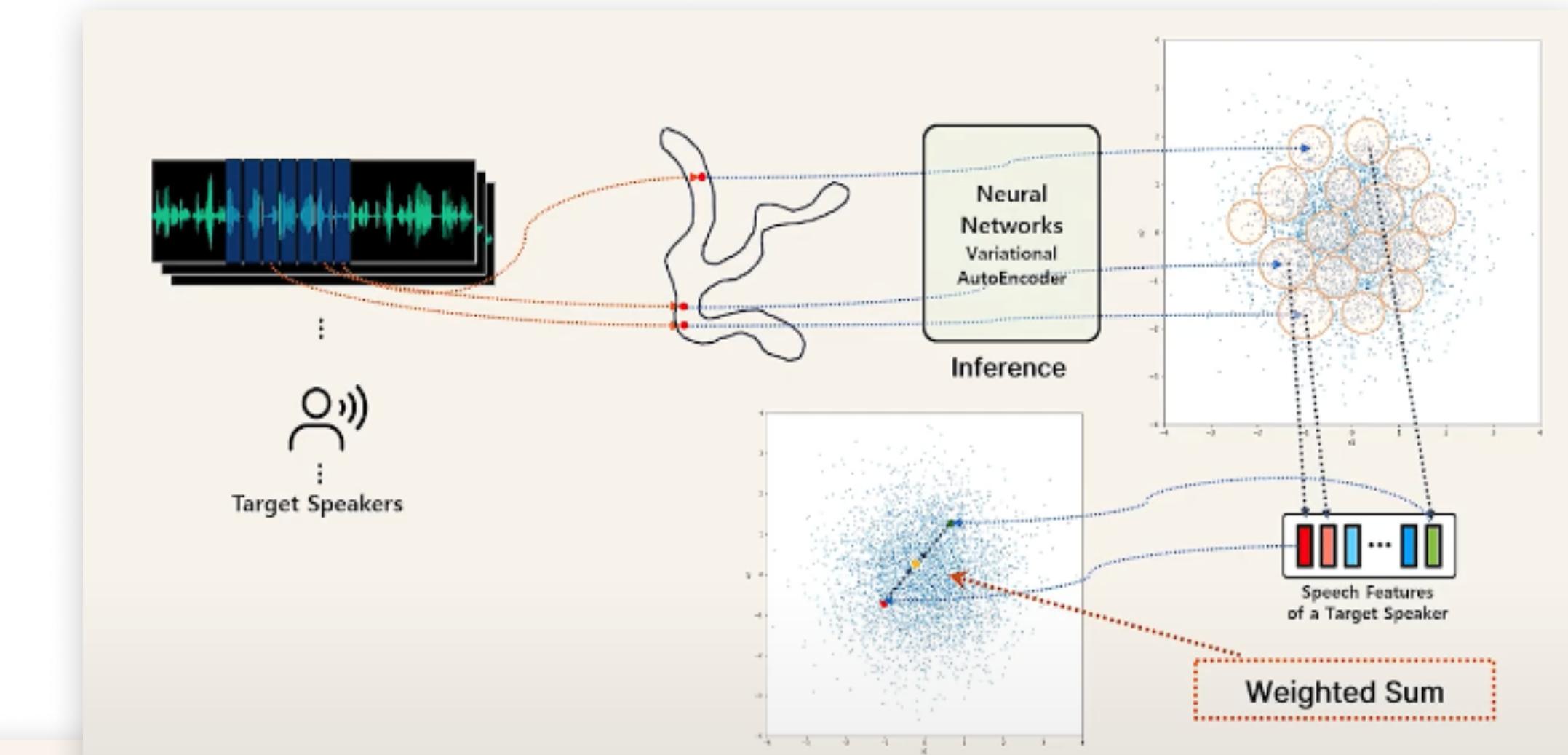
개인화된 음성
지원을 위해서는
추가적인 방법 고안이 필요

개인화 AI 음성

VITS2

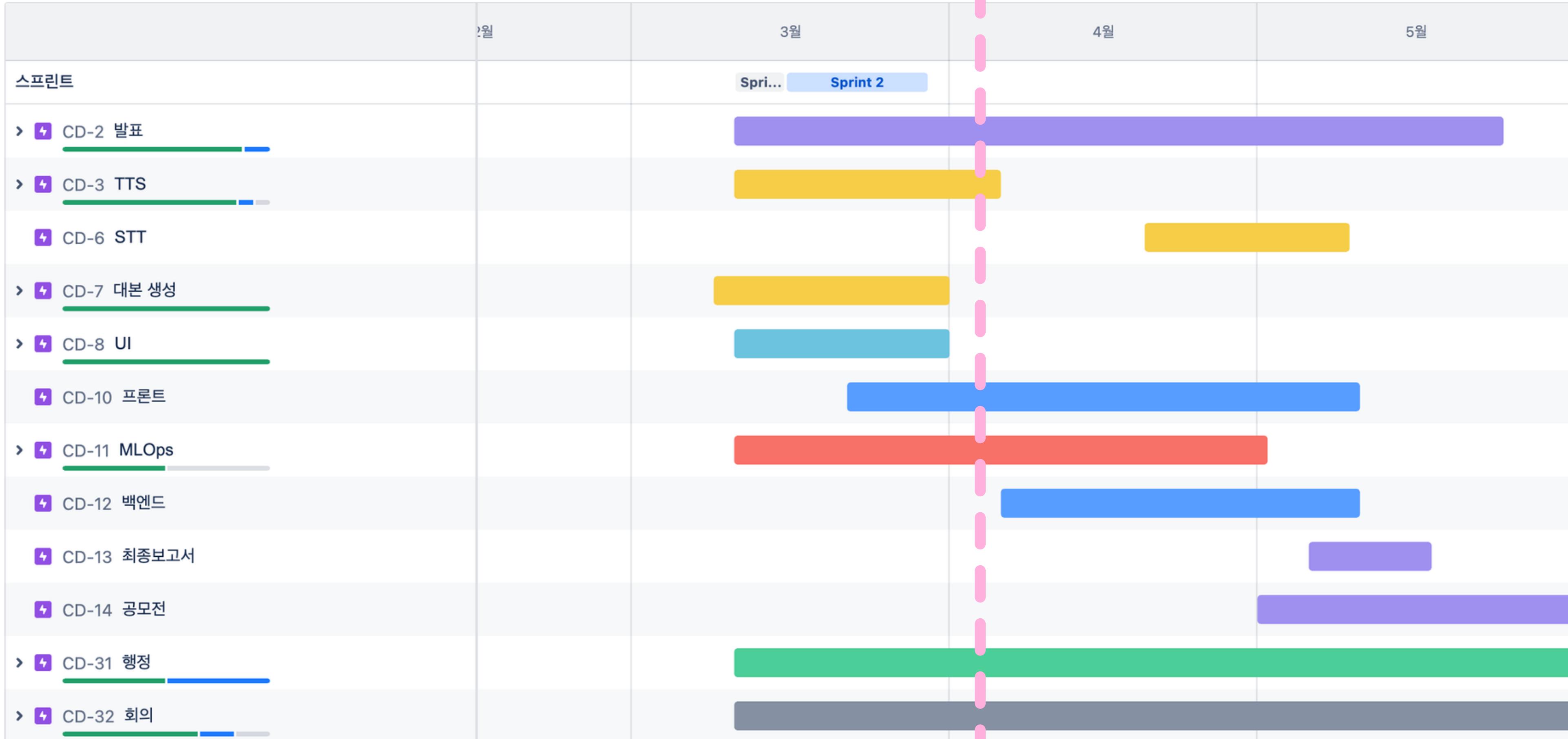
Conditioning Speech Features

- Speaker encoder에서 추출된 hidden representation을 클러스터링
- Attention의 weighted sum으로 Speech feature를 continuous space로 복원



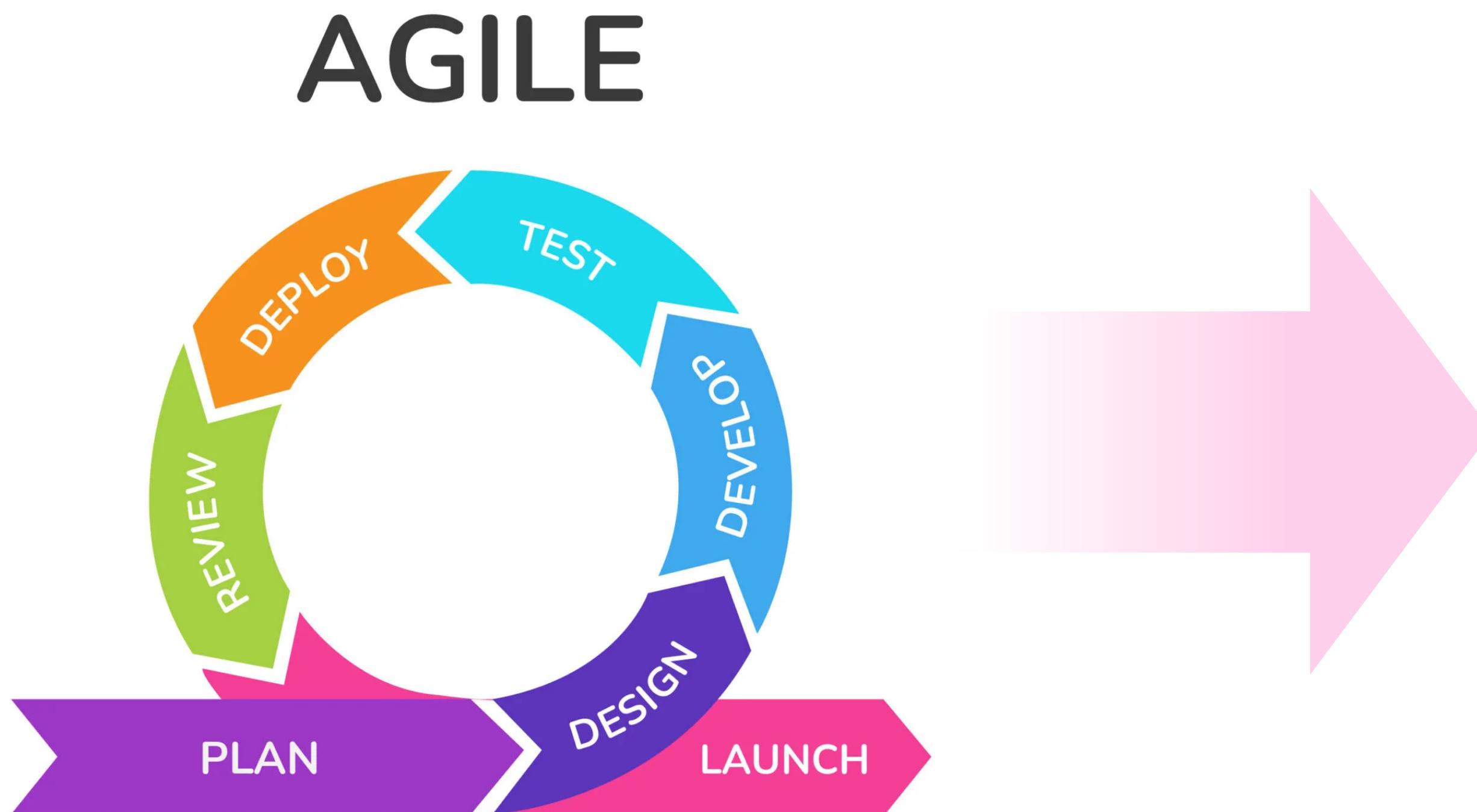
프로젝트 진행 및 관리 방법

일정



프로젝트 진행 및 관리 방법

개발 방법론 및 프로젝트 관리 도구



Jira Software

Confluence

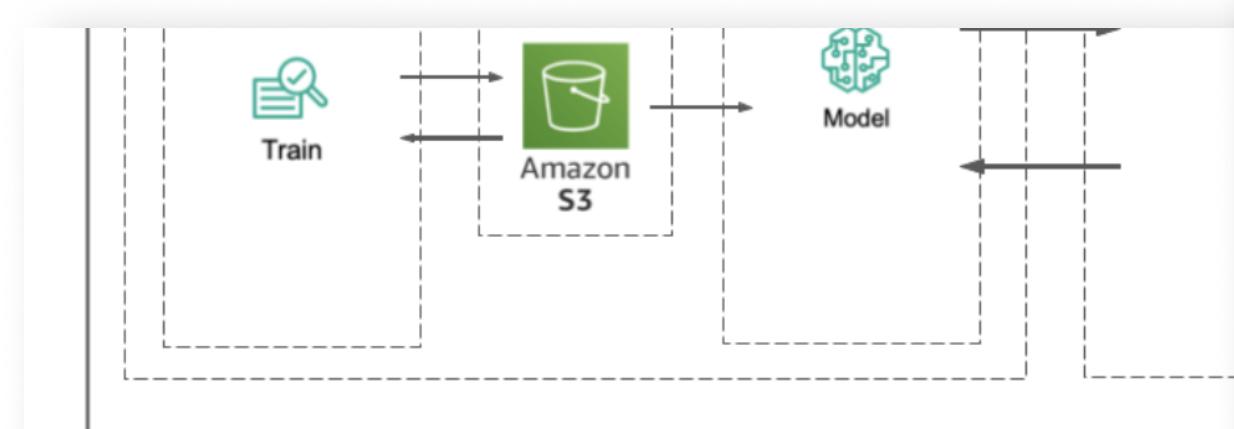
프로젝트 진행 및 관리 방법

개발 프로세스



프로젝트 진행 및 관리 방법

문서관리



하나의 모델 구현 방식으로 TTS 모델을 구현했을 때의 모습 구상도

문제점은?

- 사용자마다 학습 모델이 존재한다. 모델을 경량화 시켜서 최대 5mb로 줄인다. 동시에 접속하는 유저가 1000명이라고 치면 50GB의 모델이 동시에 추론 중이다.

해결방법은?

- RedisAI 사용 [MLOps] Multi-Model 서빙을 위한 RedisAI Cluster 구축
- 하지만 일반적으로 Redis를 운영할 때는 확장성을 위해 Redis Cluster를 구축하거나 ElasticCache for redis 혹은 Redis Enterprise와 같은 클라우드 서비스를 이용하는 경우 Redis Labs에 직접 문의해본 결과 RedisAI 모듈은 아직 위와 같은 클라우드 서비스를 받게 되었습니다.

위의 서비스를 이용할 수 없다면, 직접 Redis Cluster를 구축 후 RedisAI 모듈을 연동하는 것이 가능하리라 생각되었고, 결론적으로는 가능했습니다.

드롭아웃하여 조건부 및 무조건적인 목적으로 단일 Diffusion 모델

샘플링은 조정된 x -prediction $(z_t - \sigma \tilde{\epsilon}_\theta)/at$ 를 사용하여 수행된다

$$\tilde{\epsilon}_\theta(z_t, c) = w\epsilon_\theta(z_t, c) +$$

$$\epsilon_\theta := (z_t - \alpha$$

$\epsilon_\theta(z_t, c)$: conditional ϵ -prediction
 $\epsilon_\theta(z_t)$: unconditional ϵ -prediction

Imagen은 효과적인 텍스트 conditioning을 위해 classifier-free

w : guidance weight

- w = 1 : Classifier-free guidance 비활성화
- w > 1 : guidance 강화

Classifier guidance는 사전 훈련된 모델의 그래디언트

기술이며, Classifier-free guidance는 이러한 사전 훈련된 모델을 피하면서 효과적인 텍스트 조건을 위해 랜덤 드롭아웃을 사용하여 모델을 훈련하는 기술이다.

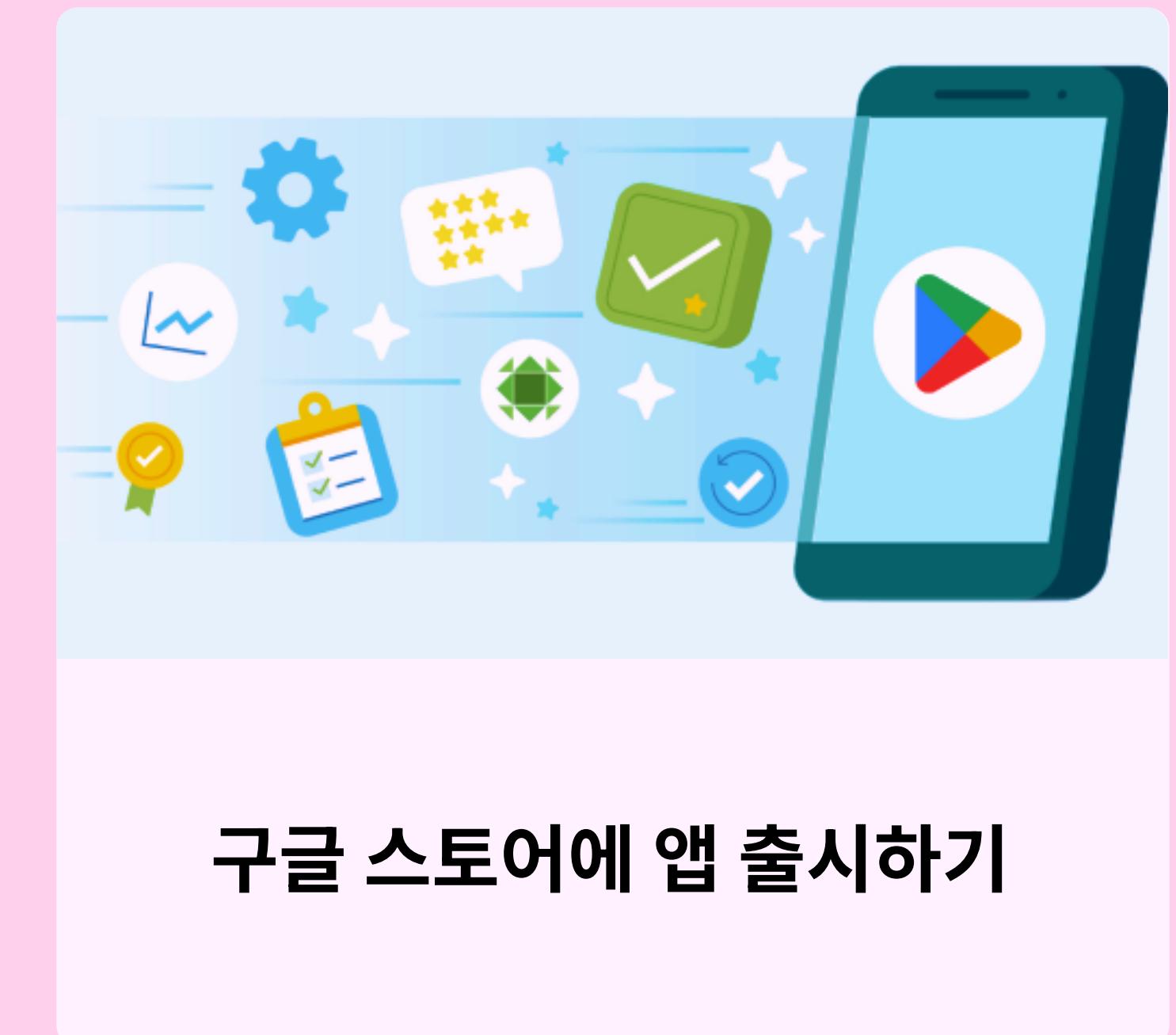
목표

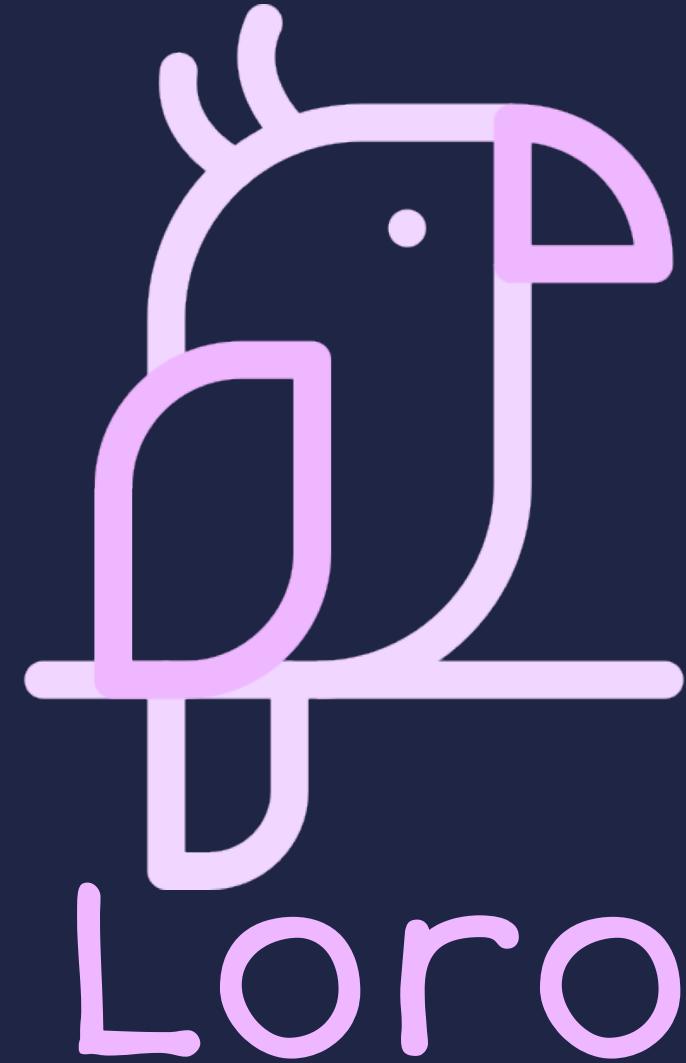
- 앱 이름 및 한 줄 소개
- 발표 자료 피피티 템플릿
- 자료 제작 역할 분담
- 대본 수정

토론 주제

항목	비고
대본 수정	<ul style="list-style-type: none"> • 프로젝트 목표 <ul style="list-style-type: none"> ◦ 아나운서 학원을 대체할 수 있다 (X) ◦ 아나운서 학원에 다니기 힘든 아나운서 지망생들에게 도움이 될 수 있다. or 아나운서 학원을 다니더라도 평상시에 연습할 수 있다. • 수행 내용 및 중간 결과 <ul style="list-style-type: none"> ◦ 회원가입 - 홈 - 스크립트 - 기록 - 설정 순으로 설명 ◦ 음성 정보 수집 <ul style="list-style-type: none"> ▪ 1분 정도의 분량 ▪ 감정이 드러나는 다른 종류의 글 ▪ 따옴표 등 학습에 도움이 되는 문장

향후 추진 계획





Thank you

Reference

- J. Kim, J. Kong, and J. Son, “**Conditional variational autoencoder with adversarial learning for end-to-end text-to-speech**”
- J. Kong et al. “**VITS2: Improving Quality and Efficiency of Single-Stage Text-to-Speech with Adversarial Learning and Architecture Design**”
- Shivam Mehta et al. “**OverFlow: Putting flows on top of neural transducers for better TTS**”
- Shivam Mehta et al. “**Matcha-TTS: A fast TTS architecture with conditional flow matching**”