



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 목적

프로젝트 개요



유사 서비스 현황



프로젝트 진행 및 관리



서비스 주요 기능

아키텍처

- 스크립트 생성
- 개인화 AI 음성
- 정확도 피드백

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

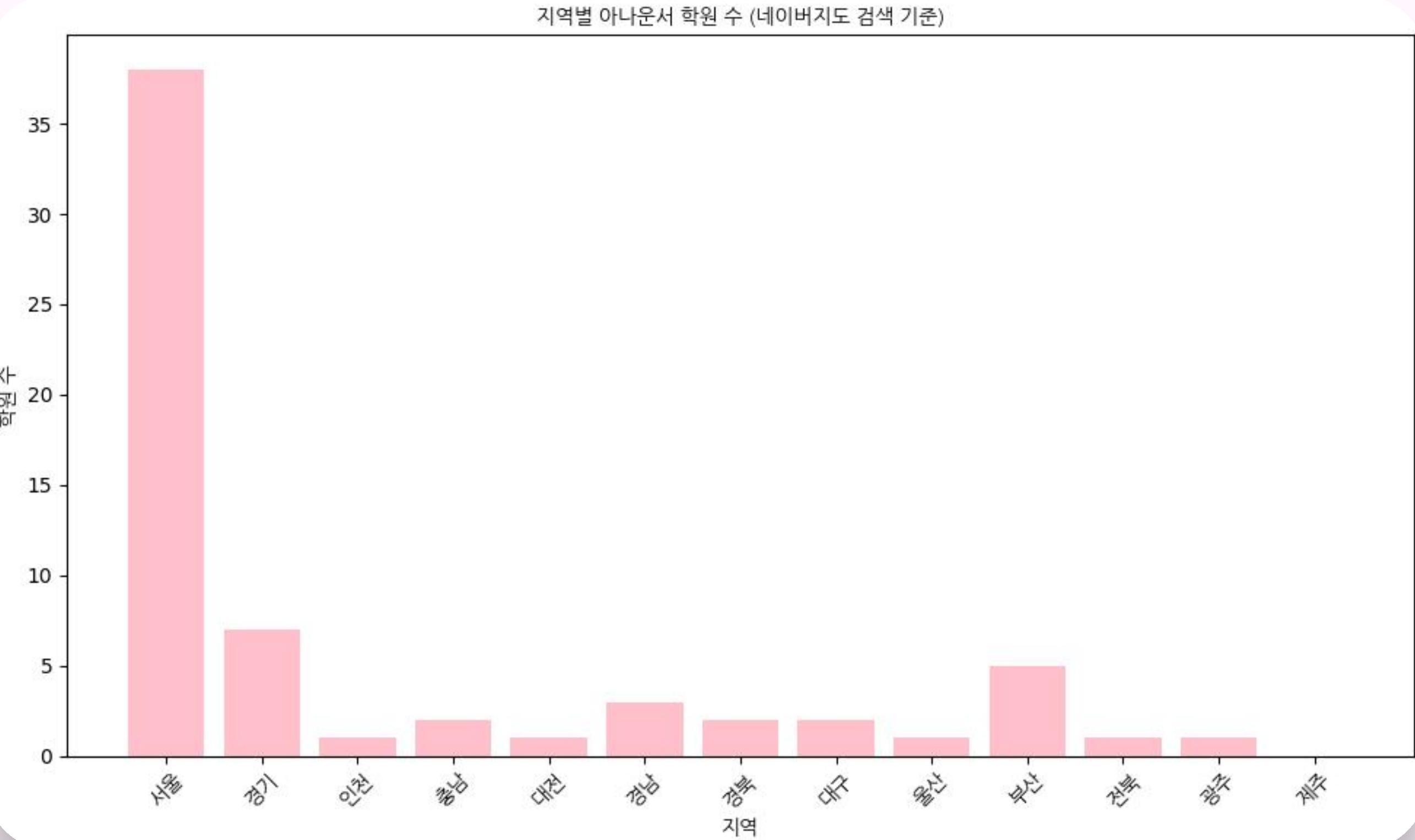
*직접 조사한 결과를 바탕으로 작성한 표(2024년도 3월 기준)

프로젝트 개요 문제 정의

지리적



대도시나 번화가에 위치하여
지방과는 떨어지는 접근성



*직접 조사한 결과를 바탕으로 작성한 그래프(네이버지도 검색 기준)

프로젝트 개요

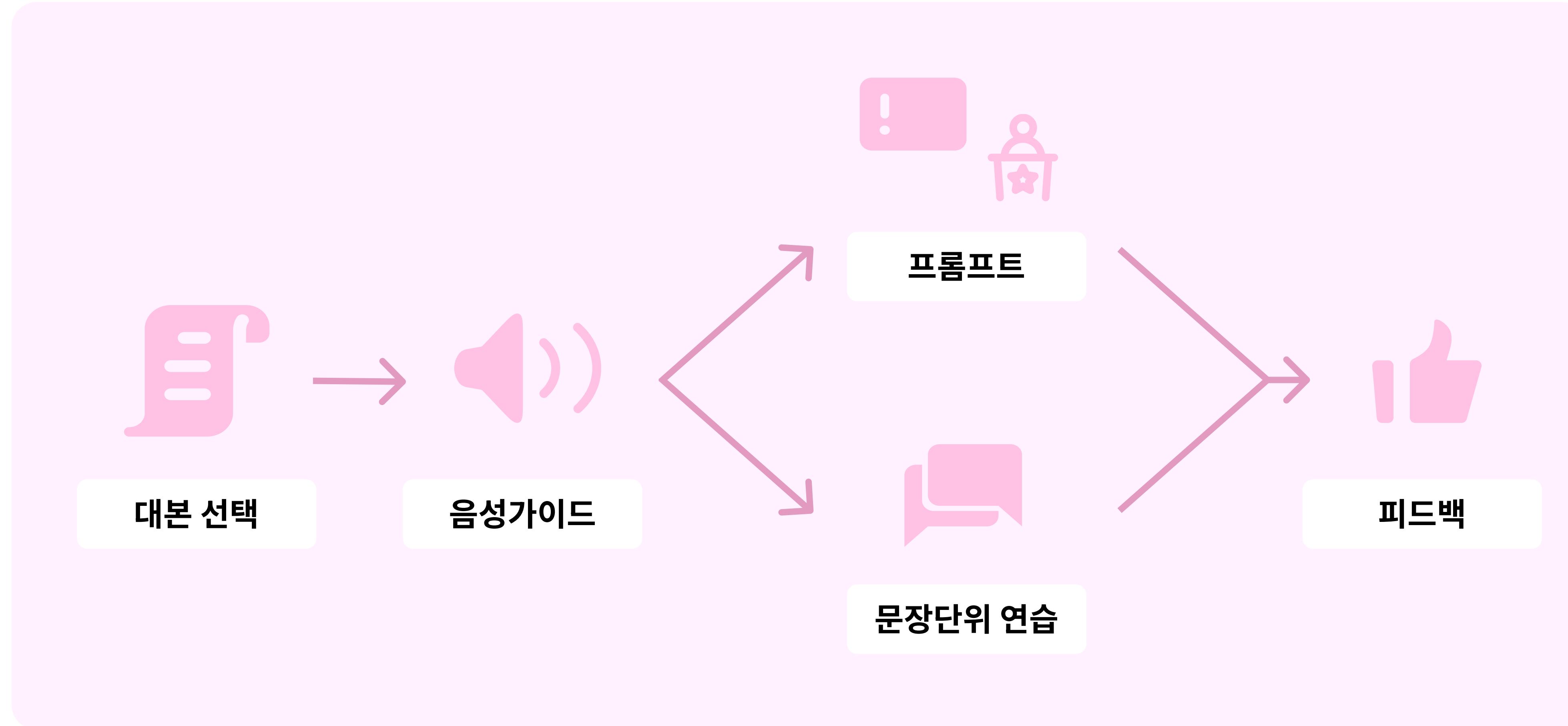
프로젝트 목적

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 개요

유사 서비스 현황

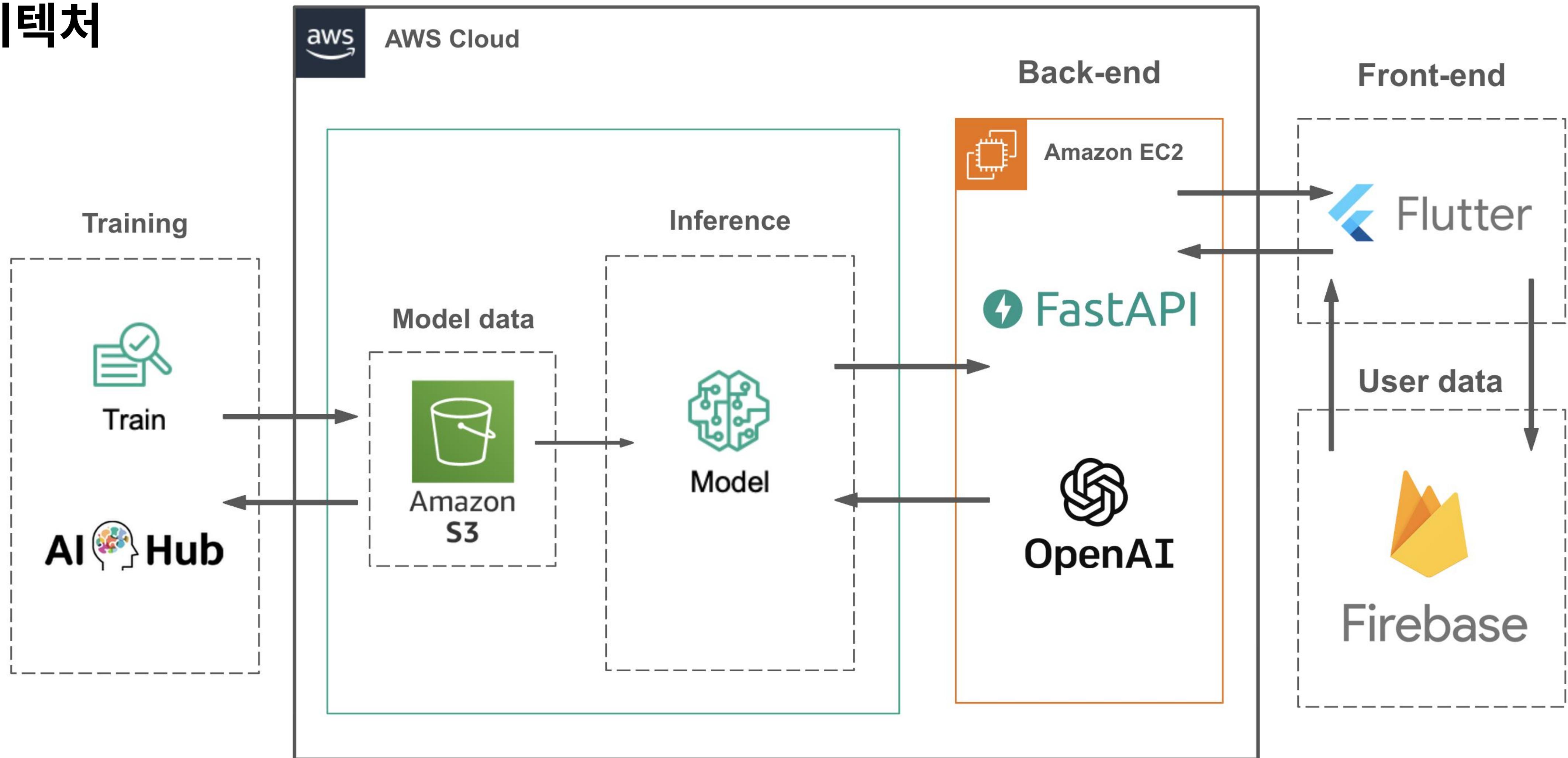
차별점

- '아나운서' 직업적 특성 반영
- 사용자 맞춤형 음성 가이드



*직접 조사한 결과를 바탕으로 비교 분석

아키텍처



스크립트 생성

예시 대본 제공

공공데이터포털 데이터셋 사용
<한국언론진흥재단_뉴스빅데이터_메타데이터_언론>



*파일 데이터 추출, 프롬프트 템플릿 작성



*아나운서 대본에 적합하게 가공



생성형 AI로 실시간 대본 생성

사용자가 입력한 제목/카테고리

프롬프트 엔지니어링



*프롬프트 템플릿 작성



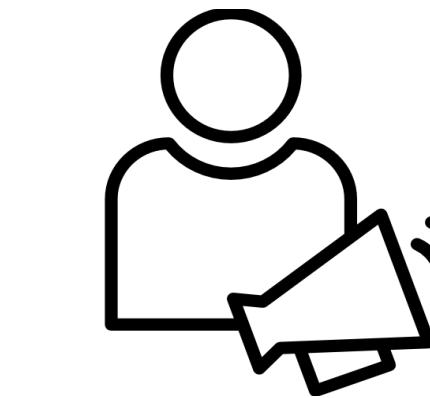
*대본 생성

가이드 음성 생성

개인화 TTS 모델

AI-Hub 데이터셋 활용

맞춤형 가이드 음성



Speaker 1



Speaker 2



Speaker 3

아나운서 억양

#음성

NEW

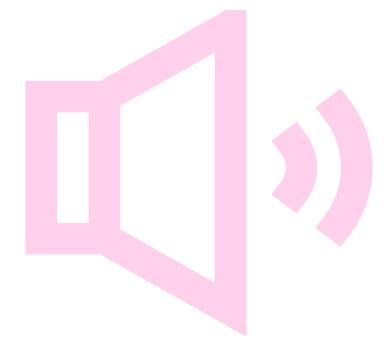
뉴스 대본 및 앵커 음성 데이터

분야 한국어 유형 오디오

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

개인화 AI 음성

VITS



부분적으로
부자연스러운 발음

Matcha-TTs

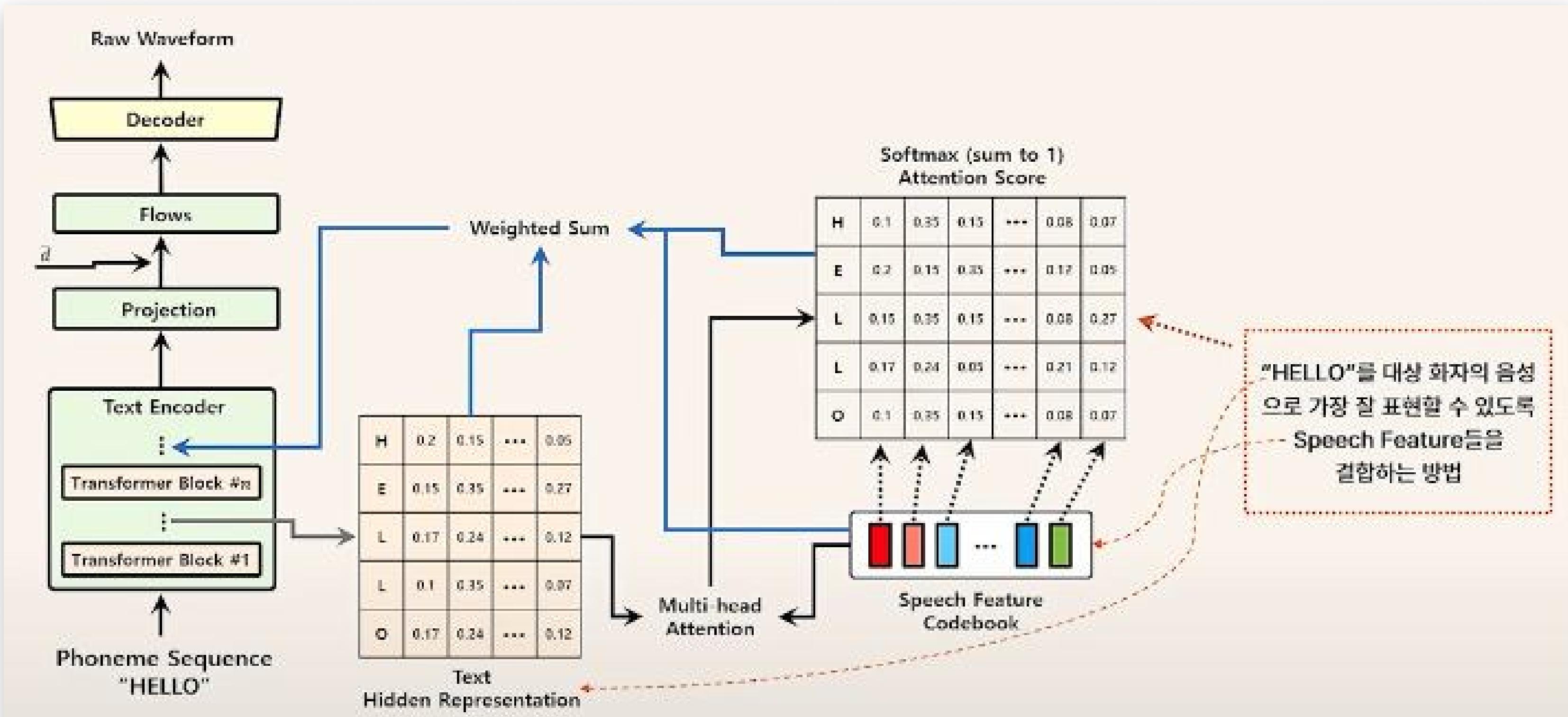
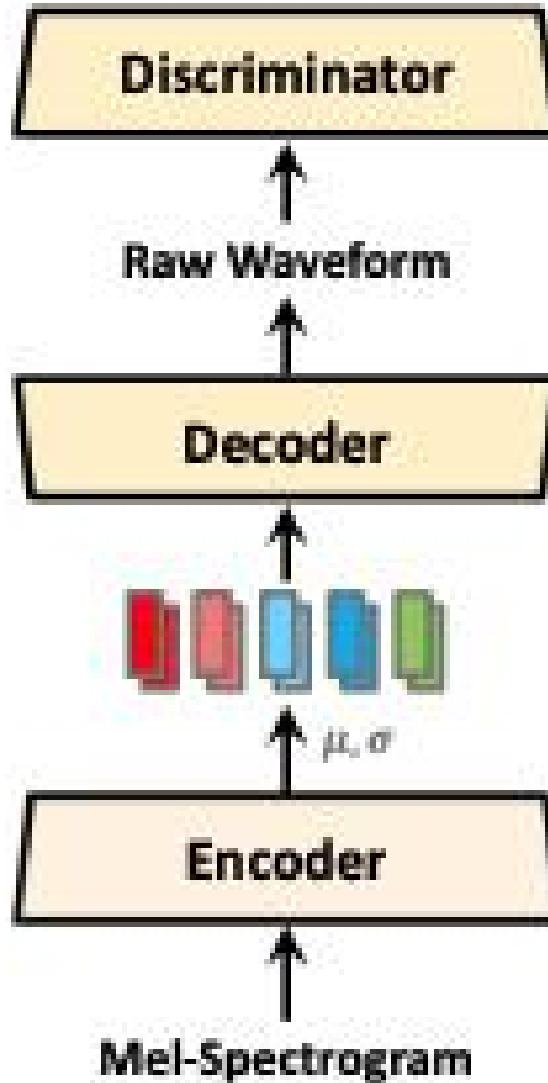


개인화된 음성
지원을 위해서는
추가적인 방법 고안이 필요

OverFlow

개인화 AI 음성

VITS2



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

정확도 피드백

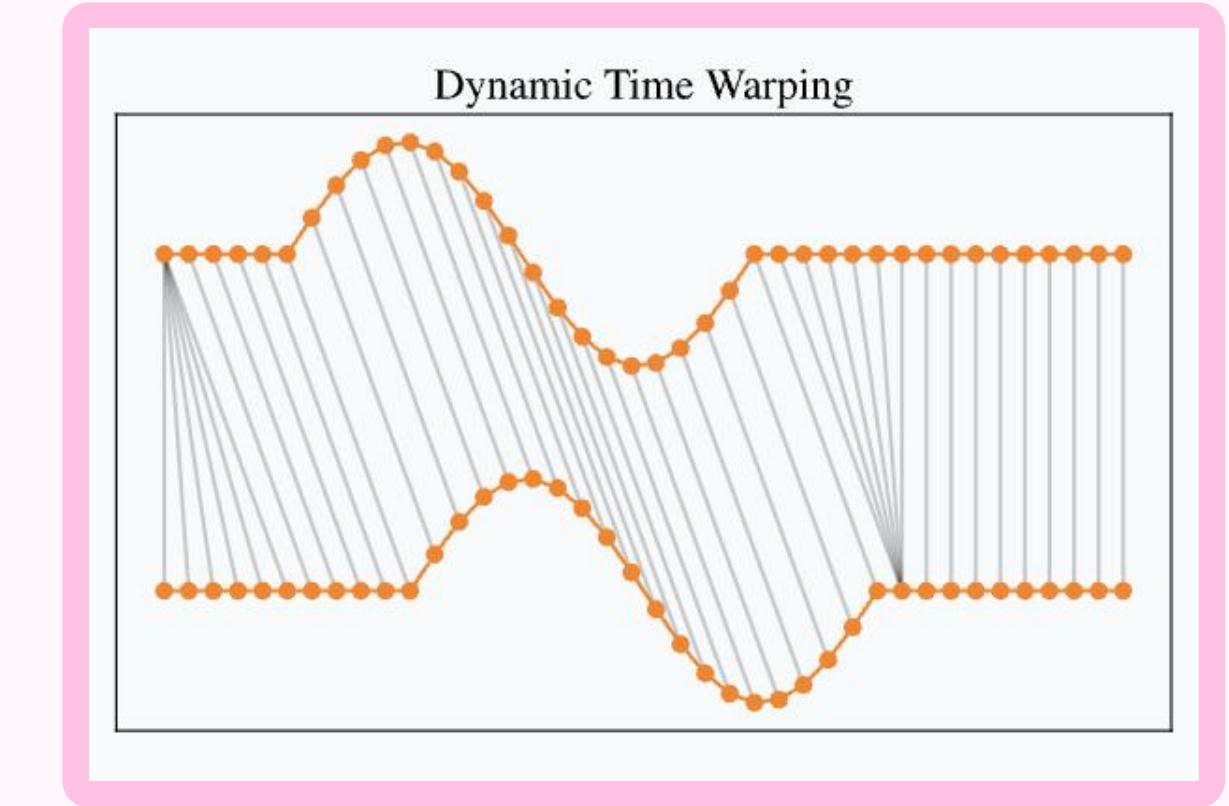


피드백

음성 간 유사도

mfcc 추출

DTW → 시간 축 매핑



텍스트 간 유사도

나는 집에 갔다

IPA 간 비교

가이드

n'ən̩wnts'ib̩e g'æd̩tə

사용자

n'ən̩wnts'ib̩e g'ætə

프로젝트 진행 및 관리 방법

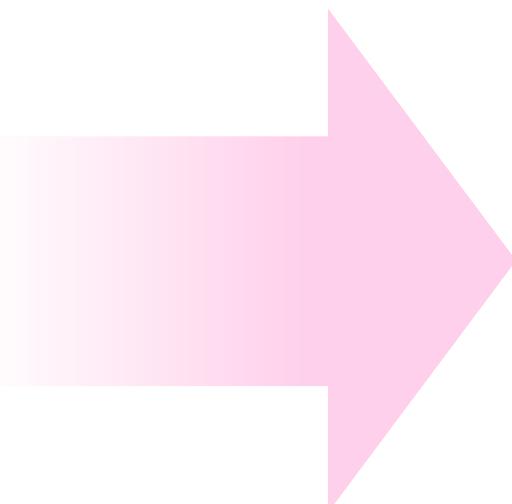
개발 프로세스



프로젝트 진행 및 관리 방법

스프린트 플래닝 및 진행

CD-223	[프론트] 음성 가이드 혹은 바로 연습 선택 화면 제작		안자
CD-222	[프론트] 가이드 음성 들려주면서 대본 자동 스크롤 화면 제작		안자
CD-221	Use Whisper-Jax for speed		김필
CD-220	[프론트] 캐릭터 다시 정하기		신민
CD-219	[프론트] SelectPractice argument 설정		신민
CD-218	[프론트] 대본 생성 후 바로 연습한 경우 record 없을 때 예외처리		안자
CD-217	[발표] 포스터 제작		신민
CD-216	[프론트] 가이드 음성 가져오기 테스트		안자
CD-215	[백엔드] EC2 서버에서 대본 생성 확인 테스트		안자



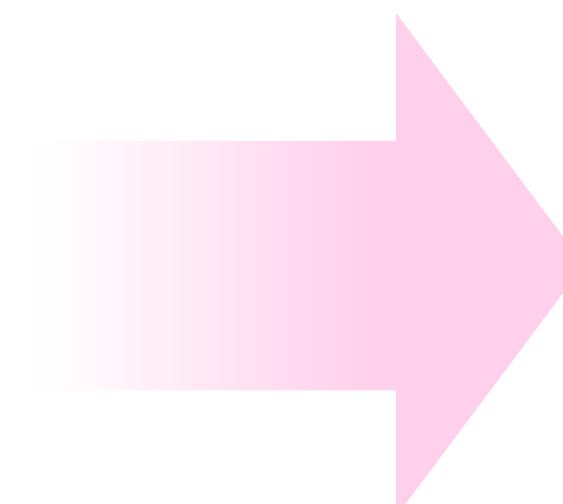
Project Management Board (Jira) showing tasks and their status:

- 개발 (Development):**
 - 1개 브랜치 (1 branch): [CD-103][CD-197] Feat: 연습 상세 화면 제작 및 프론트-백엔드 음성 데이터 주고 받기 테스트 진행 (enhancement)
 - 5개 커밋 (5 commits): Chore: 사용자 대본 내용 조정 페이지 버그 수정 & 텍스트 입력 unfocus 추가 (enhancement)
 - 2개 폴리퀘스트 (2 pull requests): [CD-137] Vits2 custom dataset 추가 (enhancement), Hotfix: prevent a negative similarity (bug)
- 보고자 (Reporter):**
 - 폴리퀘스트 (Pull Request): [CD-190][CD-219] Chore: 회원가입 화면 제작 (Merged, 2 days ago)
 - 만들자 업데이트 (Created Update): 최근 업데이트 2일 전
 - 해결됨 (Resolved): +그 외 1개 폴리퀘스트
- 할 일 9 (Work 9):**
 - [TTS] vits2 추가 개발 사항 확인 (TTS): CD-141 (프론트)
 - [프론트] 프롬프트 연습 상세 화면 제작 (프론트): CD-143 (백엔드)
 - [TTS] 성능 확인 후 Convolution layer를 사용해 VAE를 학습 (TTS): CD-192 (프론트)
 - [프론트] 디자인 수정 (프론트): CD-200 (백엔드)
 - [발표] 최종보고서 제작 (발표): CD-181 (백엔드)
- 신행 중 10 (In Progress 10):**
 - [TTS] vits2 speaker condition을 text encoder에 전달 (TTS): CD-162 (프론트)
 - [백엔드] 만들어진 API 가져와서 음성 데이터 주고 받고 피드백 결과 Flutter 단에서 확인 (백엔드): CD-167 (프론트)
 - [프론트] 구글 로그인 구현 (프론트): CD-189 (백엔드)
- 완료 21 (Completed 21):**
 - [TTS] vits2 speaker condition을 text encoder에 전달 (TTS): CD-162 (프론트)
 - [백엔드] 만들어진 API 가져와서 음성 데이터 주고 받고 피드백 결과 Flutter 단에서 확인 (백엔드): CD-167 (프론트)
 - [프론트] 구글 로그인 구현 (프론트): CD-189 (백엔드)

프로젝트 진행 및 관리 방법

Daily Scrum

- ▼ Daily Scrum
 - 3월의 Daily Scrum
 - 4월의 Daily Scrum
 - ▼ 5월의 Daily Scrum
 - 05/01 Daily Scrum
 - 05/03 Daily Scrum
 - 05/06 Daily Scrum
 - 05/08 Daily Scrum
 - 05/10 Daily Scrum
 - 05/13 Daily Scrum



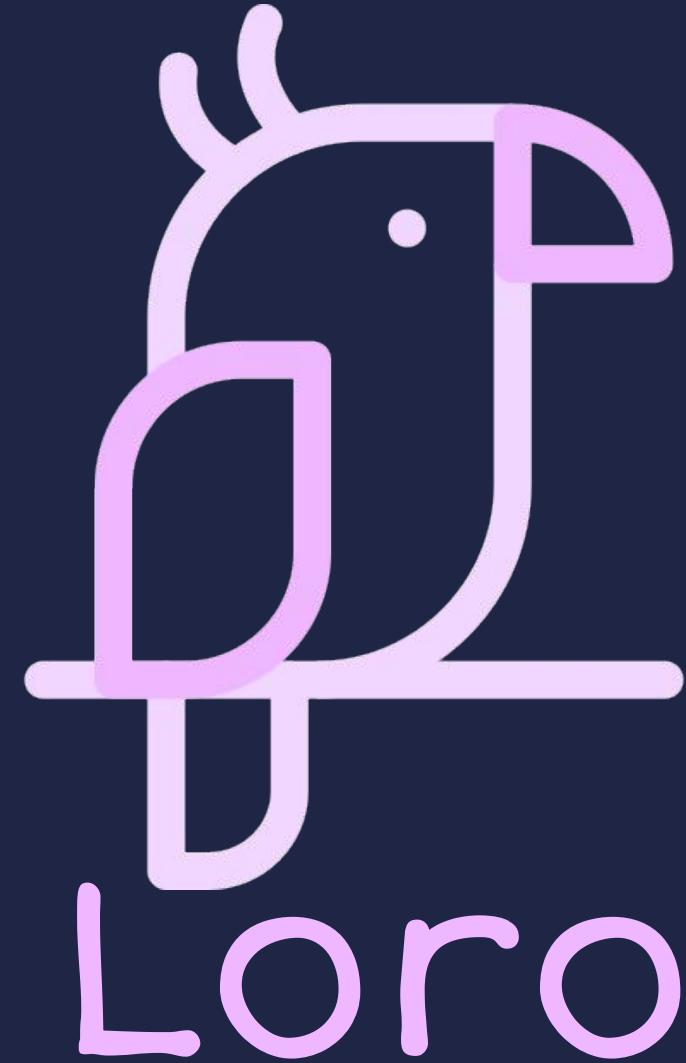
이름	오늘 한 일	내일 할 일	현재 맞닥뜨린 사항 및 문제점
@윤하은	<ul style="list-style-type: none"> <input checked="" type="checkbox"/> GitHub - p0p4k/vits2_pytorch: unofficial vits2-TTS implementation in pytorch <input checked="" type="checkbox"/> VITS2 모델 코드 파악하기 <input checked="" type="checkbox"/> VITS2 수정 필요한 부분 파악하기 	<ul style="list-style-type: none"> • vits2 구현해둔 두개 리포에서 공통으로 부족한 부분 정리 • 다른 리포에서 가져올 수 있는 부분 파악 <ul style="list-style-type: none"> ◦ GitHub - fishaudio/Bert-VITS2: vits2 backbone with multilingual-bert 	
@신민경	<ul style="list-style-type: none"> <input checked="" type="checkbox"/> 폰트 적용 <input checked="" type="checkbox"/> 사용자 대본리스트 화면 제작 <input checked="" type="checkbox"/> 대본 상세 화면 제작 	<ul style="list-style-type: none"> • 대본리스트 화면 완성 • 대본 상세 화면 완성 • 사용자 대본 생성 화면 제작 • 연습 방법 선택 화면 제작 • 반응형 공부 및 라이브러리 서치 	
@안지원	<ul style="list-style-type: none"> <input checked="" type="checkbox"/> Loro와 비슷한 Flutter project 서치 <input checked="" type="checkbox"/> Flutter에서 음성 데이터 입력 받기 <input checked="" type="checkbox"/> GetX 복습하기 	<ul style="list-style-type: none"> • 홈 화면 UI 제작 • 파일베이스 유저 데이터 디비 구조 설계 	
@김필모	<input checked="" type="checkbox"/> whisper 한국어 발화 인식 colab 코드 작성	<ul style="list-style-type: none"> • time step 마다의 stt 추가 예정 	stt에 4초 정도 걸림

프로젝트 진행 및 관리 방법

KPT 회고

- Sprint Retrospective Template
- Sprint 1 Retrospective (03.15)
- Sprint 2 Retrospective (03.29)
- Sprint 3 Retrospective (04.05)
- Sprint 4 Retrospective (04.17)
- Sprint 5 Retrospective (05.03)
- Sprint 6 Retrospective (05.08)

담당자	Keep		Problem	Try
	<ul style="list-style-type: none"> • 이번 스프린트에서 계속 유지했으면 좋은 점 • 좋아더 저 팀워크에게 고마워더 저 		<ul style="list-style-type: none"> • 이번 스프린트에서 문제가 되었던 점 • 개선할 점 	<ul style="list-style-type: none"> • 다음 스프린트에서 시도해보고 싶은 것 • 팀원들에게 하고 싶은 것 또는 말, EX) 우리 회사 하버 하시다
담당자	Keep		Problem	Try
	<ul style="list-style-type: none"> • 이번 스프린트에서 계속 유지했으면 좋은 점 • 좋아더 저 팀워크에게 고마워더 저 		<ul style="list-style-type: none"> • 이번 스프린트에서 문제가 되었던 점 • 개선할 점 	<ul style="list-style-type: none"> • 다음 스프린트에서 시도해보고 싶은 것 • 팀원들에게 하고 싶은 것 또는 말, EX) 우리 회사 하버 하시다
담당자	Keep		Problem	Try
@윤하은	<ul style="list-style-type: none"> • 중간 발표를 무사히 마쳤다. • 프론트 개발을 시작했다. 		<ul style="list-style-type: none"> • 플래닝 시에 프로젝트 환경 설정에 소요되는 시간을 고려하지 못했다. • 발표 준비로 개발에 많은 시간을 투자하지 못했다. • 의사소통이 잘 안된다. • 데일리스크럼으로 서로의 진행 상황이 공유가 잘 안되는 것 같다. 	<ul style="list-style-type: none"> • 다음 스프린트에서 시도해보고 싶은 것 • 팀원들에게 하고 싶은 것 또는 말, EX) 우리 회식 한번 합시다.
@김필모	<ul style="list-style-type: none"> • 중간 발표 준비에 모든 팀원이 시간을 많이 투자해주었다. • 개발에 들어감 • 중간 발표를 잘 마쳤고 우리 세션 한정 기획/설계가 다른 조에 비해 잘되었다고 느낌 		<ul style="list-style-type: none"> • 시간을 중간 발표 준비 및 설계/디자인에 쓴느라 늦어진 개발 • 서로 간의 개발 진행상황이 너무 공유가 안된다 • pr에 reviewer 설정이 안되어 빠른 pr 리뷰가 안되었음 • 현재 마스터와 하위 브랜치 간의 차이가 너무 큼. • 의사소통이 원활히 되지 않음 • ai를 잘 아는 교수님에게 프로젝트에 관한 피드백을 받은 적이 없음 	<ul style="list-style-type: none"> • 상대방의 말을 끝까지 듣고 말하기 • 커밋 컨벤션 지키기 • pr마다 assignee, reviewer 설정해 주었으면 좋겠음 • 나중에 pr 필터링할 때 편하게 auto-labeling 해주는 actions 추가 고려 • 스크럼에 보다 구체적인 내용을 적어 서로 간 진행상황을 알 수 있도록 하면 좋겠음. • 백로그에 개발에 필요한 이슈들을 미리 옮렸으면 좋겠음. 지금 백로그에 이번 스프린트 때, 당장 할 것만 있는데 앞으로 어떤 것이 더 필요한지 각 에피마다 태스크를 추가하면 방향성을 확인하는 데 좋을 것 같다. • 포크를 하고 pr을 옮려다 보니 코드 공유가 잘 안되는 것 같음. 코드를 다 완성하고 pr을 하려다 보니 생기는 문제라 생각함 • → 그러다 보니 코드양이 많아지면 리뷰하기 힘들어지고 완성 전까지 팀원이 알 수도 없음 • → (방법) 그 pr의 하위 태스크 중 하나만 작성했더라도 무조건 pr 올리고 설명란에 어떤 것에 대한 pr인데 그것 중 어떤게 됐고 아직 안됐는지 적으면 공유도 되고 리뷰도 순차적으로 할 수 있음(pr 머지 전까지는 추가 커밋 시 pr에 추가됨) • 위 내용을 위해 pr 컨벤션을 정해도 좋을 듯 함. 아주 간단하게 • 빠른 시일 내에 ai를 잘 아는 교수님이나 멘토님에게 상담/피드백을 받아보면 좋겠음 • 맛있는거 먹어야 우리
@안지원	<ul style="list-style-type: none"> • 최종 발표 막보기인 중간 발표를 성공적으로 완료. 만족스러운 피드백은 받지 못했지 		<ul style="list-style-type: none"> • 중간 발표 준비로 인한 부진한 개발 속도 • 중간 발표에 쏟은 시간에 비해 프로젝트 전반 	<ul style="list-style-type: none"> • 이제 오로지 개발에 집중하기 • 다음 스프린트 회고 때 각 파트별 머지 진행하기



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**”