



www.2023_track_project.com



2023 Track Project

도전!

저세계 아이돌

20 서양하

22 권주명

22 도운서

22 정재현

22 황효은

23 최진영



생성 모델을 이용한 노래 커버 영상 제작 서비스



목차

1. Introduction

2. 진행 상황

3. 앞으로 할 일

주제 선정 동기



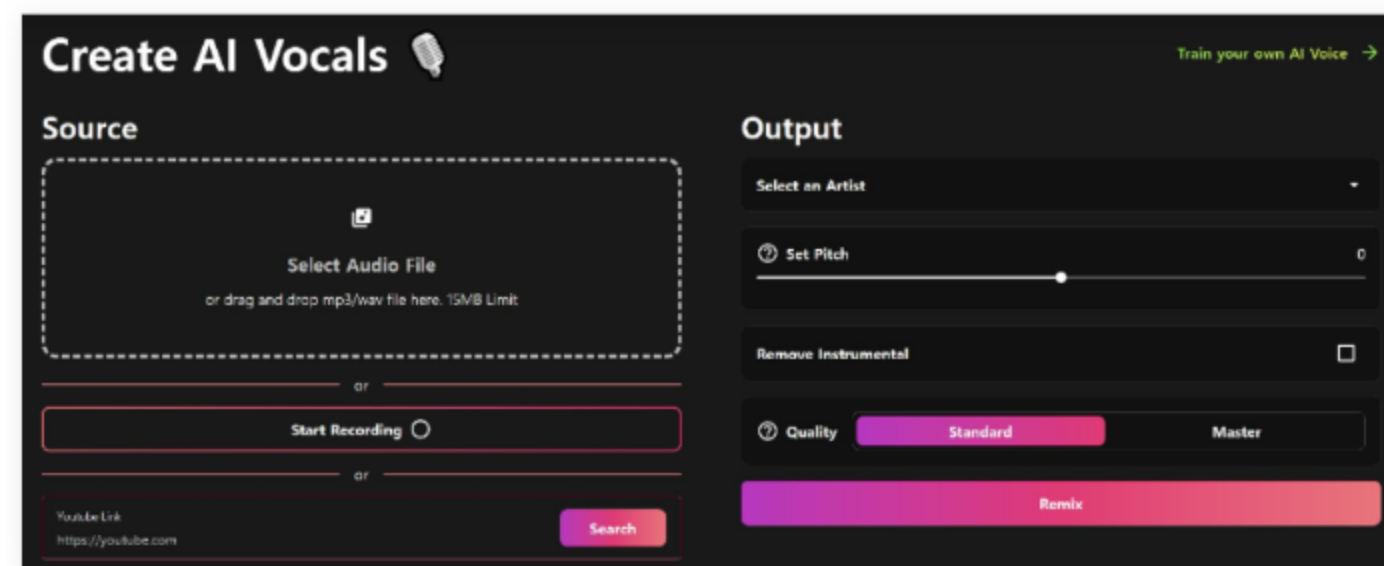
DEAN(딘) - Newjeans (Cover) AI

조회수 272만회 · 3개월 전



꼬깃꼬깃

dean #딘 #newjeans #뉴진스 #cover #ai.



AI 커버 노래 열풍

-> 생성 모델을 활용하여 커버 춤까지 제공해주는 서비스!(앱)



서비스명: 저세계 아이돌

의미: 현실세계에서는 아니지만 저세계(사이버월드)에서는 나도 아이돌!

서비스 한줄 설명: 생성 모델을 이용한 노래 커버 영상 제작 서비스

협업 툴 : Notion

📌 구성원

프론트	황효은, 최진영
백	정재현, 서양하
AI	도윤서, 권주명

📌 규칙

- 카톡 공지사항 읽으면 확인했다는 의미로 공감표시 달기
- 할일은 무조건 정한 마감기한 내에 끝내기, 사유가 있어 불가능할 것 같다면 단톡방에 미리 양해 구하기
- 칸반보드 페이지에 완료한 일 표시하기 + 단톡방에도 진행상황 보고
- 주마다 블로그 작성은 한명씩 돌아가며 노션 페이지에 정리
(도윤서 → 서양하 → 최진영 → 정재현 → 황효은 → 권주명)

회의록 정리

Aa	📅 날짜	⋮ 장소	👥 참여자	
📅 1차 회의	2023년 9월 25일 19:00	대면 칸나	도 도윤서 권 권주명	재
📅 AI팀 회의	2023년 9월 29일 20:00	비대면	도 도윤서 권 권주명	
📅 2차 회의	2023년 10월 2일 21:30	비대면	권 권주명 재 재현 정	👤
📅 3차 회의	2023년 10월 9일 21:30	비대면	권 권주명 재 재현 정	👤
📅 4차 회의	2023년 10월 16일 21:00	비대면	권 권주명 재 재현 정	👤

칸반 보드

📅 칸반 보드 📅 상세 칸반 보드 📊 표 보기 +

○ 할 일 9	🔄 진행 중 5	✅ 완료 4
▼ backend 4 ... +		
BLOB를 이용한 미디어 저장 양 양하 재 재현 정	회원가입/로그인 기능 구현 양 양하 재 재현 정	API 명세서 작성 양 양하 재 재현 정
데이터베이스 팀원 공유 방법 찾기 양 양하 재 재현 정	+ 새로 만들기	+ 새로 만들기

진행상황 - Front

- 언어: Dart, 프레임워크: Flutter
- 전체적인 디자인은 SNOW 앱 참고
- 마이페이지, 카메라 페이지, AI 제작 페이지,
제작 파일 제공 페이지 구현 완료

← 저세계 아이돌 🧑🏻

회원정보를 입력해주세요
*표시되어 있는 것은 필수로 작성해주세요

ID email? phone number? *
최소 5자 이상으로 입력해주세요
중복되는 아이디예요ㅠㅠ

PW ○ L O Q 2 L O 2 L O *
영어, 문자, 특수문자를 각각 한 개 이상을 사용하여 입력해주세요
8자 ~ 10자로 입력해주세요

PW L O Q 2 L O 2 L O 2 *
비밀번호 재확인을 위해 다시 한 번 입력해주세요
위의 작성한 비밀번호와 달라요ㅠㅠ

Phone Number

☐ 모두 확인하였으며 동의합니다.
☐ 개인정보 수집 및 이용 동의
☐ 개인정보 제3자 제공 동의
☐ L O Q 2 L O 2 L O 2 ㄱ쇼

동의하고 가입하기

← 저세계 아이돌 🧑🏻

회원정보를 입력해주세요
*표시되어 있는 것은 필수로 작성해주세요

ID email? phone number? *
최소 5자 이상으로 입력해주세요
중복되는 아이디예요ㅠㅠ

PW ○ L O Q 2 L O 2 L O *
영어, 문자, 특수문자를 각각 한 개 이상을 사용하여 입력해주세요
8자 ~ 10자로 입력해주세요

PW L O Q 2 L O 2 L O 2 *
비밀번호 재확인을 위해 다시 한 번 입력해주세요
위의 작성한 비밀번호와 달라요ㅠㅠ

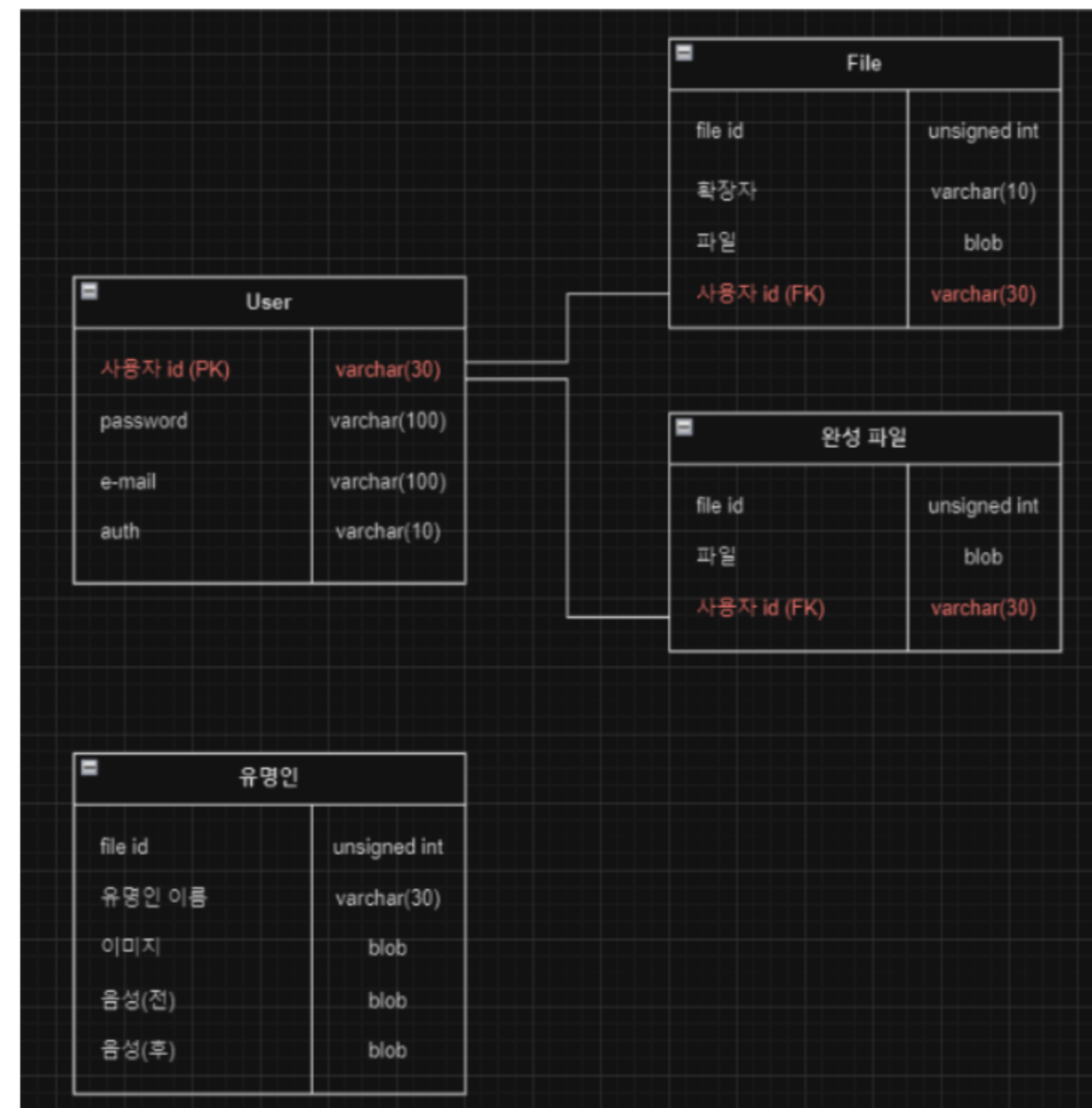
Phone Number

☒ 모두 확인하였으며 동의합니다.
☒ 개인정보 수집 및 이용 동의
☒ 개인정보 제3자 제공 동의
☒ L O Q 2 L O 2 L O 2 ㄱ쇼

동의하고 가입하기

진행 상황 - Back

- 언어: Java, 프레임워크: Spring Boot
- 데이터베이스: MySQL, Mybatis 사용
- 회원가입, 로그인/로그아웃 구현 완료



진행 상황 - AI

DiffSinger: Singing Voice Synthesis via Shallow Diffusion Mechanism

Jinglin Liu*, Chengxi Li*, Yi Ren*, Feiyang Chen, Zhou Zhao†

Zhejiang University
{jinglinliu,chengxili,rayeren,zhaozhou}@zju.edu.cn, chenfeiyangai@gmail.com

Abstract

Singing voice synthesis (SVS) systems are built to synthesize high-quality and expressive singing voice, in which the acoustic model generates the acoustic features (e.g., mel-spectrogram) given a text input. Previous SVS models adopt a simple generative adversarial network (GAN) to learn the distribution of singing voice features, while they suffer from blurry and over-smoothed training issues respectively. In this paper, we propose an acoustic model for SVS based on the diffusion probabilistic model. DiffSinger is a parameterized Markov chain that iteratively converts the noise into mel-spectrogram conditioned on the music score. By implicitly optimizing variational bound, DiffSinger can be stably trained and generate realistic outputs. To further improve the voice quality and speed up inference, we introduce a shallow diffusion mechanism to make better use of the prior knowledge learned by the simple loss. Specifically, DiffSinger starts generation at

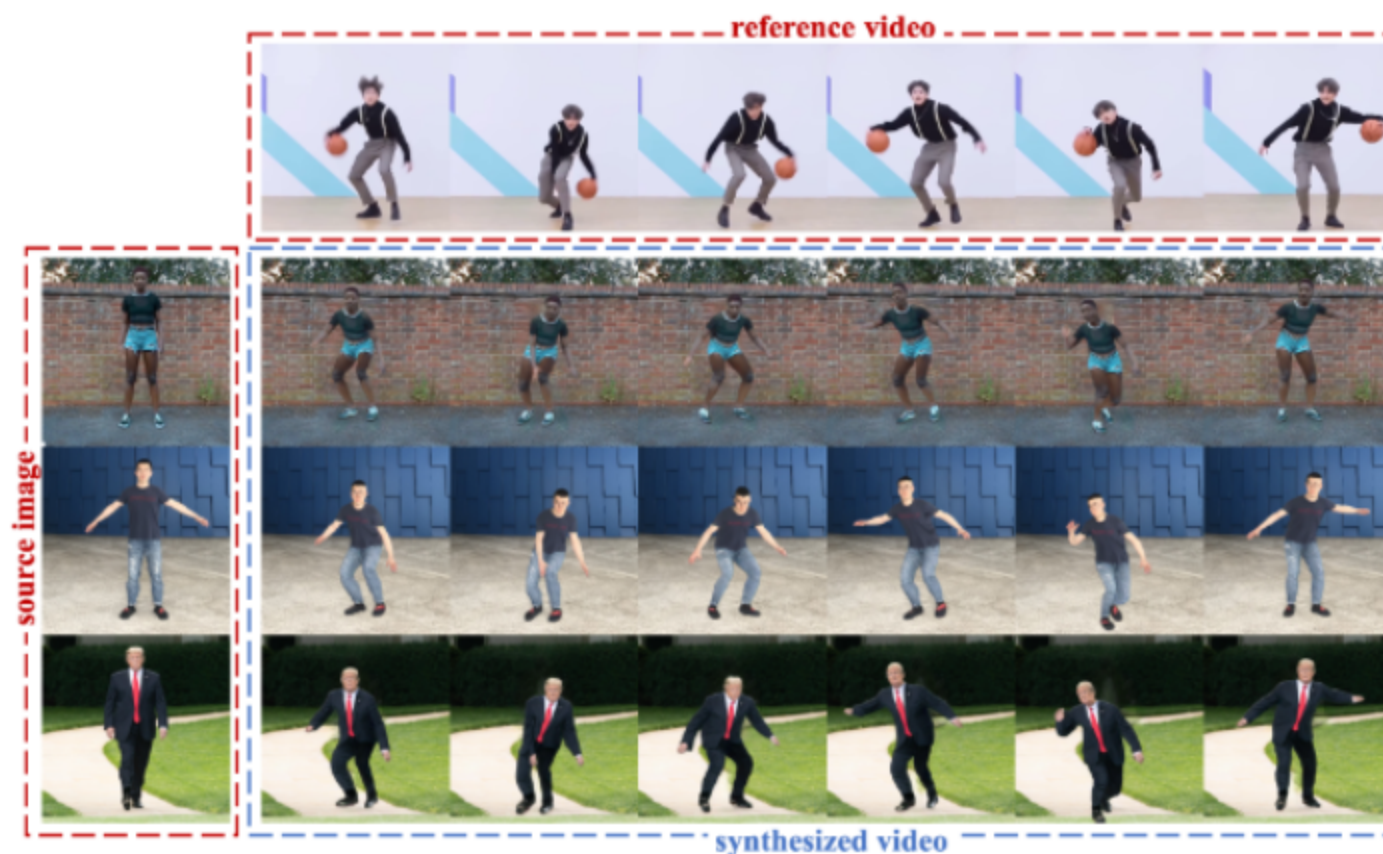
음성AI 만들 때
대충 예전의 GAN보다 낫다는 내용

2019; Lee et al. 2019; Blaauw and Bonada 2020; Ren et al. 2020; Chen et al. 2020)†.

Previous singing acoustic models mainly utilize simple the acoustic features. on the incorrect uniting to blurry and over-smoothed training issues respectively. In this paper, we propose an acoustic model for SVS based on the diffusion probabilistic model. DiffSinger is a parameterized Markov chain that iteratively converts the noise into mel-spectrogram conditioned on the music score. By implicitly optimizing variational bound, DiffSinger can be stably trained and generate realistic outputs. To further improve the voice quality and speed up inference, we introduce a shallow diffusion mechanism to make better use of the prior knowledge learned by the simple loss. Specifically, DiffSinger starts generation at

Recently, a highly flexible and tractable generative model, diffusion probabilistic model (a.k.a. diffusion model) (Sohl-Dickstein et al. 2015; Ho, Jain, and Abbeel 2020; Song, Meng, and Ermon 2021) emerges. Diffusion model consists of two processes: diffusion process and reverse process (also called denoising process). The diffusion process is a Markov

Diff - SVC



Liquid Warping GAN with Attention

진행 상황 - AI

- 언어: Python
- 데이터 전처리(묵음 구간 제거, 음성 자르기, zip파일 압축) 구현
- 음성 모델 개발중

전처리 코드

```
# 사용할 파일 로드
def load(cap_path, img_path, voi_path):
    # 동영상(커버할 영상, 노래) 파일 로드
    cap = cv2.VideoFileClip(cap_path)

    # 이미지(전신 사진) 파일 로드
    img = cv2.imread(img_path)

    # 음성(커버할 목소리) 파일 로드
    voi = AudioSegment.from_file(voi_path)

    return cap, img, voi

# 커버할 노래와 영상 분리
def cap_sep(PATH, cap):
    #1) 커버할 영상 소리 제거(mp4 -> mp4)
    os.makedirs(PATH + '/cover_video', exist_ok=True)
    new_clip = cap.without_audio()
    new_clip.write_videofile(PATH + '/cover_video/video_without_sound.mp4')

    #2) 커버할 영상에서 음성 추출(mp4 -> mp3)
    os.makedirs(PATH + '/cover_music')
    cap.audio.write_audiofile(PATH + '/cover_music/cover_audio.mp3')
```


앞으로 할 일

- Front

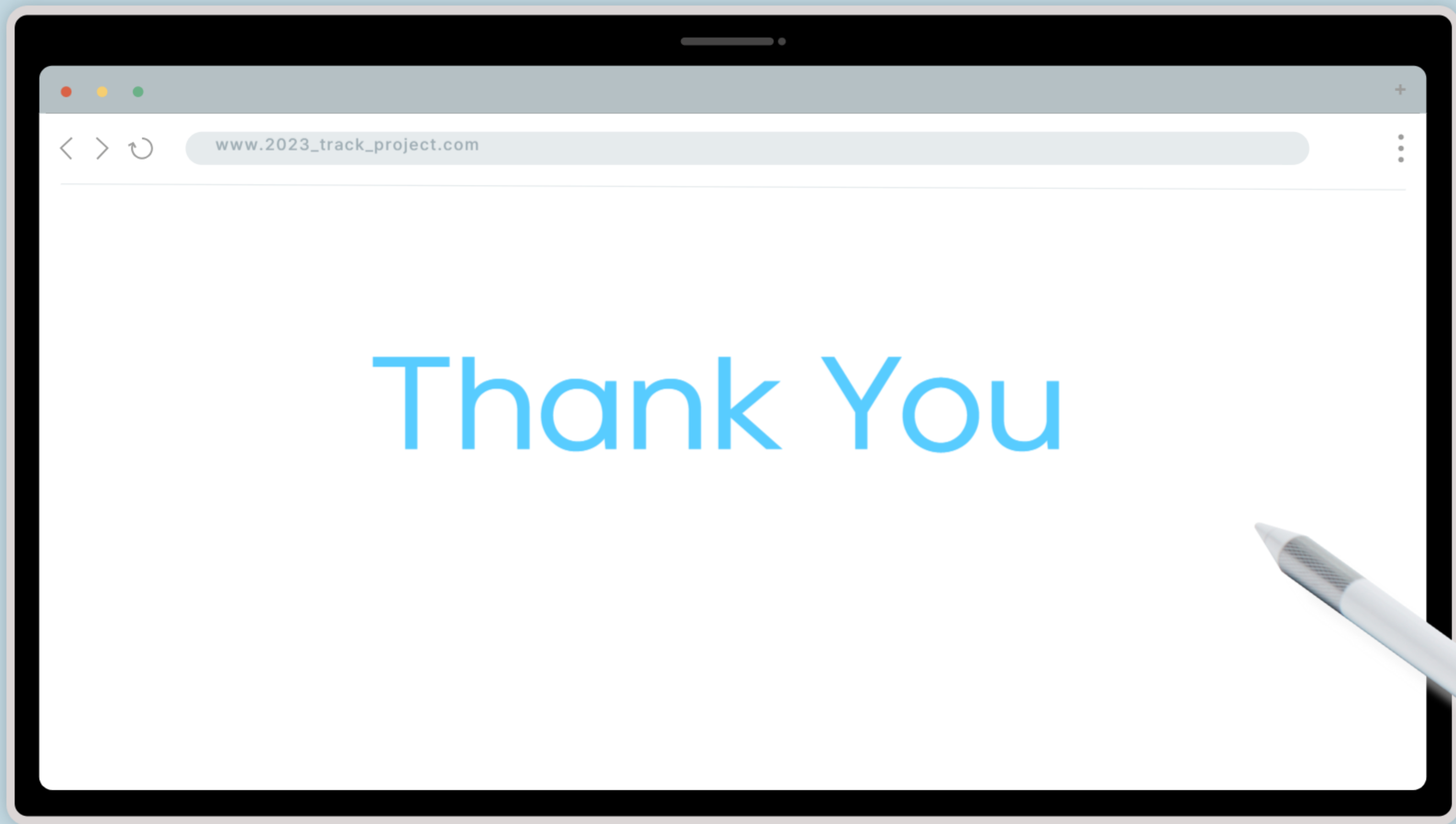
- 앱 기능 명세서 작성
- 주의사항 페이지 만들기
- 이미지/동영상/녹음 local 연결
- api 연결 모델

- Back

- 회원가입/로그인 기능 구현
- BLOB를 이용한 미디어 저장
- 데이터베이스 팀원 공유 방법 찾기

- AI

- Diff-SVC 음성 모델 구현
- MR 제거 코드 구현
- 영상 모델 구현



Thank You