

AI를 활용한 차세대 추모 플랫폼 서비스 플랫폼 기억

2023 졸업작품

20180566 박성호

20180571 백승우





목차

01

프로젝트 개요

플랫폼 기억,
프로젝트에 대한 설명

02

수행 계획

플랫폼 기억,
수행 계획에 대한 설명

03

개발 과정

플랫폼 기억,
개발 과정에 대한 설명

04

시연 영상

플랫폼 기억,
시연 영상에 대한 설명



01

프로젝트 개요

문제 인식 / 해결 방안 / 프로젝트 소개



?

문제 인식



“어렸을 적 가족과 함께 했던 추억이
기억 속에만 존재하고 사진이나
영상으로 남아있지 않아요”



“보고싶지만 다신 보지 못하는 사람들이 있어요.
그들이 사무치게 그립지만, 그들의 모습이 담긴
사진을 그대로 보는 것이 아직 힘들어요”



“우리 가족의 이야기가 담긴
고유한 콘텐츠가 있었으면 좋겠어요”



“사랑하는 사람과의 추억을 따뜻하게 간직하고 싶다”

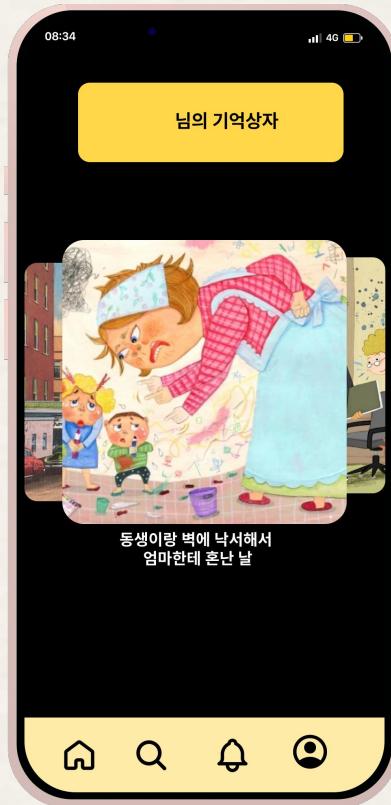
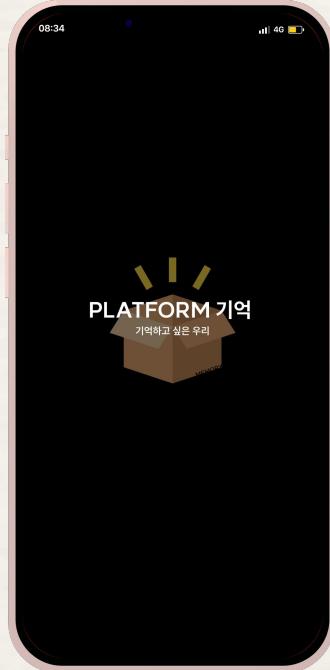




그들과의 추억이 담긴 동화책을 만들면 어떨까?

?

플랫폼 기억

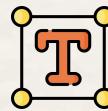


AI로 생성한
'기억 그림책'을
담은 플랫폼

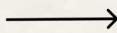


?

개발 목표



[INPUT]



[PROCESS]

AI 이미지
AI 음성
AI 텍스트



추상화
음성 모델
내용 생성



[OUTPUT]



02

수행 계획

수행 계획 및 역할 분담



수행 계획 및 역할 분담



| No | 내용 | 팀원 | 7월 | 8-9월 | 10월 |
|-----|---------|-----|----|------|-----|
| 기획 | 작품 구상 | 전원 | | | |
| 설계 | 구조도 설계 | 전원 | | | |
| | AI모델 선정 | 전원 | | | |
| 개발 | AI 이미지 | 박성호 | | | |
| | AI 음성 | 백승우 | | | |
| | 플랫폼 | 전원 | | | |
| 테스트 | 개선 및 고찰 | 전원 | | | |



03

개발 과정

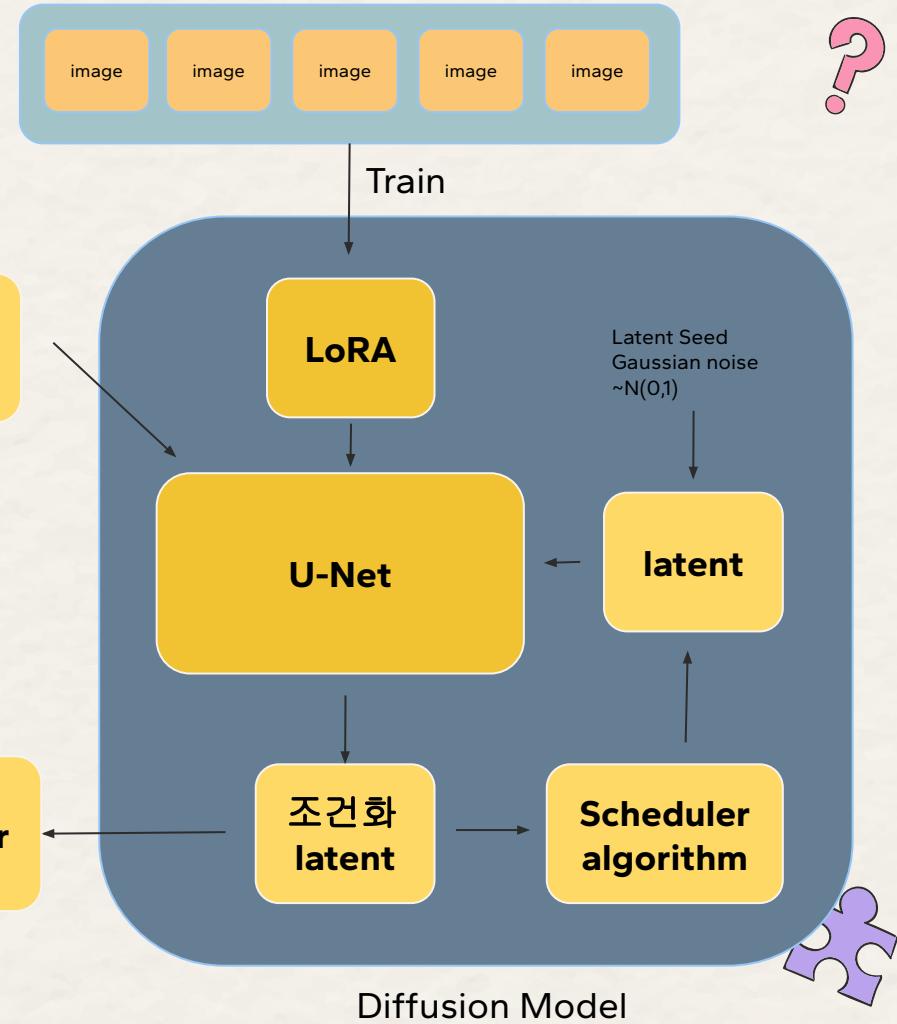
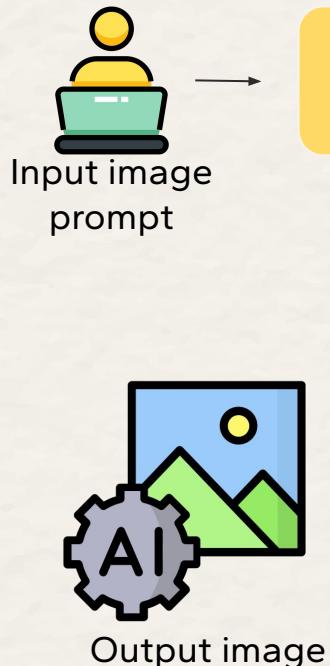
그림+음성+텍스트  플랫폼





그림 생성

생성 흐름도



Textual inversion



새로운 개념을 포착하는 단어 임베딩을 생성

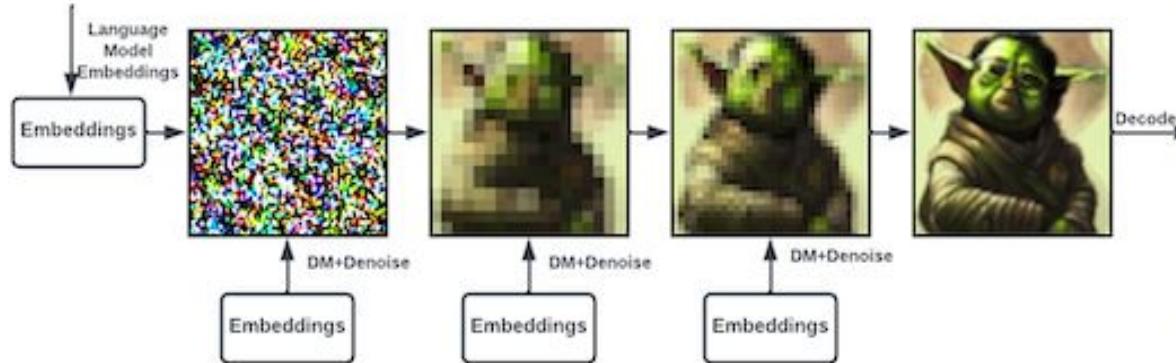
Textual inversion 훈련 방법은 기본 모델을 변경하지 않고 임베딩만 훈련하는 방식으로 작동

Noise 상태에서 임베딩이 추가되어,
원하는 화풍으로 유도

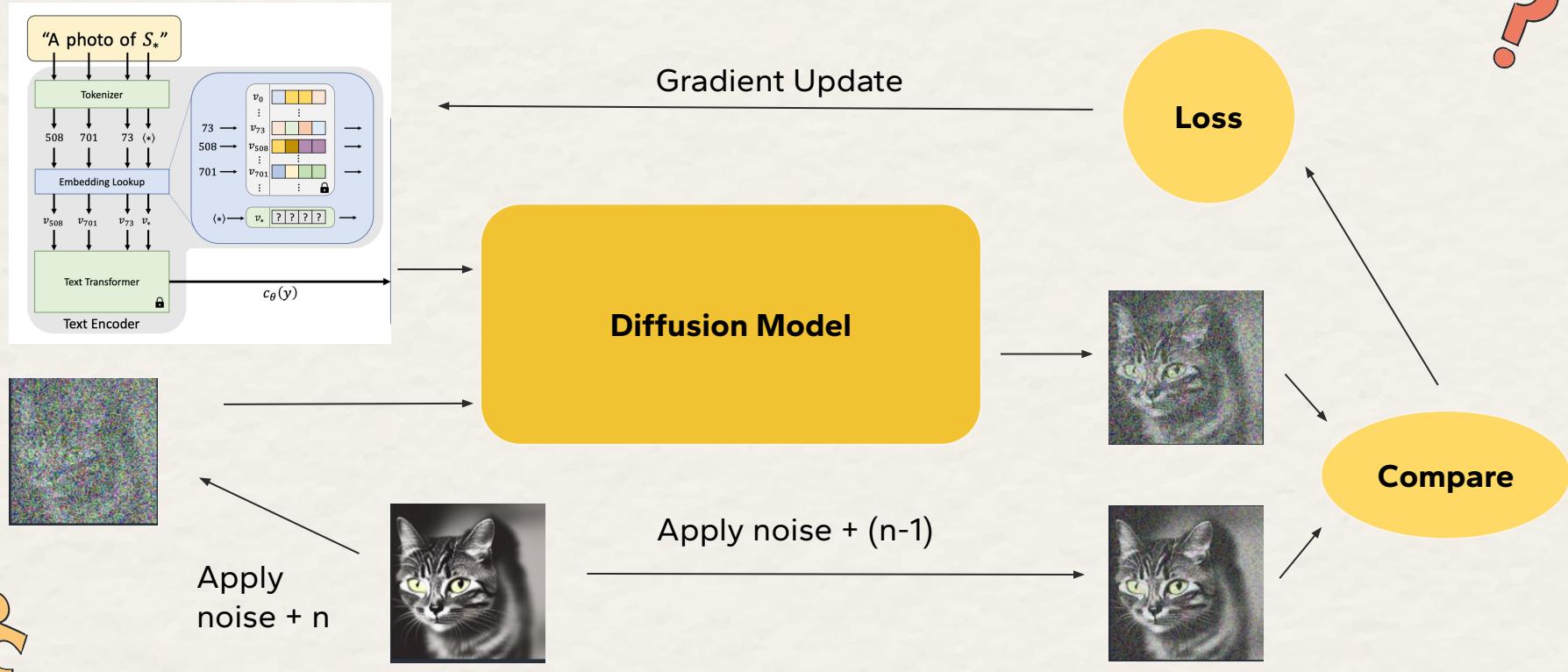
$$v_* = \arg \min_v \mathbb{E}_{z \sim \mathcal{E}(x), y, \epsilon \sim \mathcal{N}(0,1), t} [\|\epsilon - \epsilon_\theta(z_t, t, c_\theta(y))\|_2^2],$$

$$L_{LDIM} := \mathbb{E}_{z \sim \mathcal{E}(x), y, \epsilon \sim \mathcal{N}(0,1), t} [\|\epsilon - \epsilon_\theta(z_t, t, c_\theta(y))\|_2^2],$$

"(realbenny-t1 | yoda) person, star war, art by
artgerm and greg rutkowski and alphonse mucha"



Textual inversion



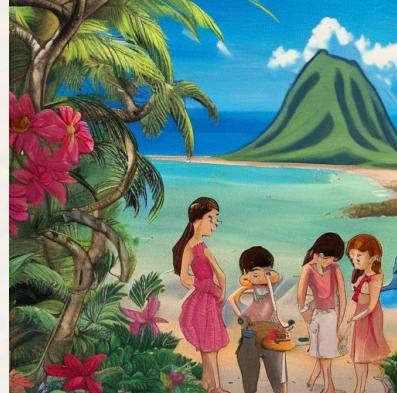
Textual inversion

원하는 화풍으로 임베딩을 추가하기 위해 지브리 스튜디오의 '마루밑
아리에티'의 3~5개의 데이터를 크롤링하여 준비했다.



Textual inversion을 하여 ckpt파일을
만들어 임베딩을 추가한 결과 원하는
화풍이 제대로 **나오지 않았다**.

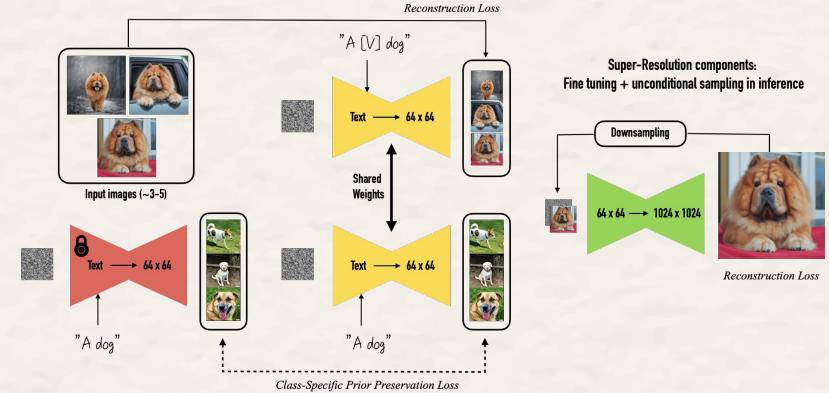
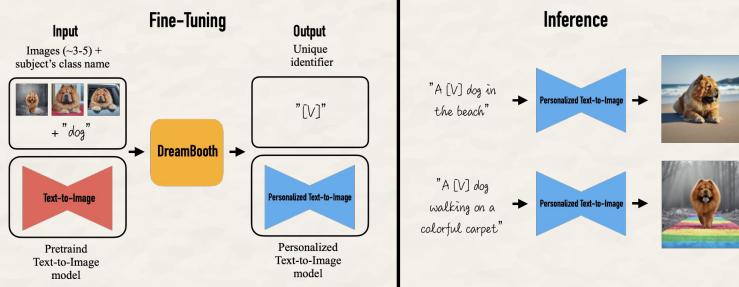
텍스트 반전의 결과로 임베딩이
나오고, 그 임베딩을 포함해 변경되지
않는 모델의 내부에서 최대한 비슷한
결과를 내는 방법으로 프로젝트와는
맞지 않았다.



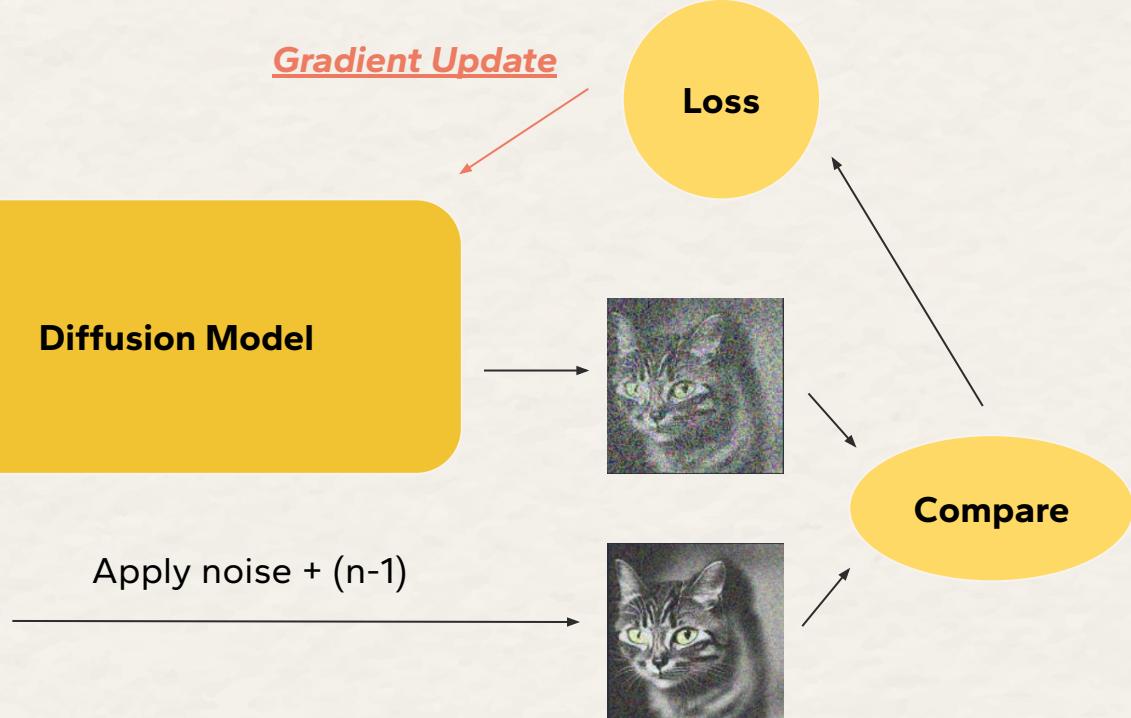
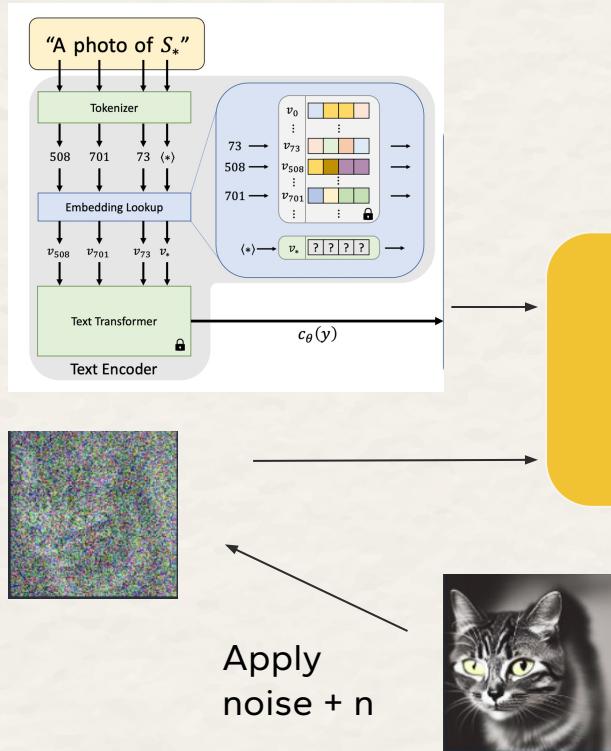
Dreambooth

주어진 사진의 시각적 특징에 대해 높은 유사성을
유지하고 새로운 맥락의 사진을 합성
몇 장의 이미지로 text to image diffusion 모델을
fine-tuning하면서도 기존 모델의 Semantic
knowledge를 보존

기존의 Fine-tuning과는 다르게 적은 수의 이미지로
모델의 오염없이 학습이 가능함
기존 모델이 가진 class name과 input image를 같이
학습
새로운 개념을 이해할 때까지 Diffusion model을
미세 조정하기 때문에 오래 걸리고 용량이 큼



Dreambooth



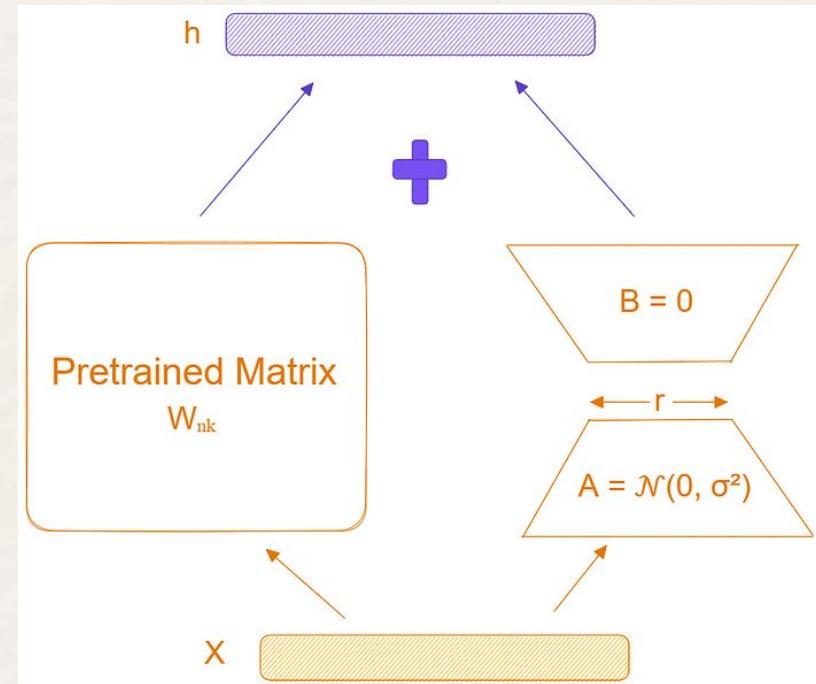
Dreambooth 기반 LoRA

LoRA(Low-Rank Adaptation)는 LLM(Large Language Models)을 특정 task에 fine-tuning하는 데 있어 cost가 너무 큰 단점을 해결하기 위해 나온 방법

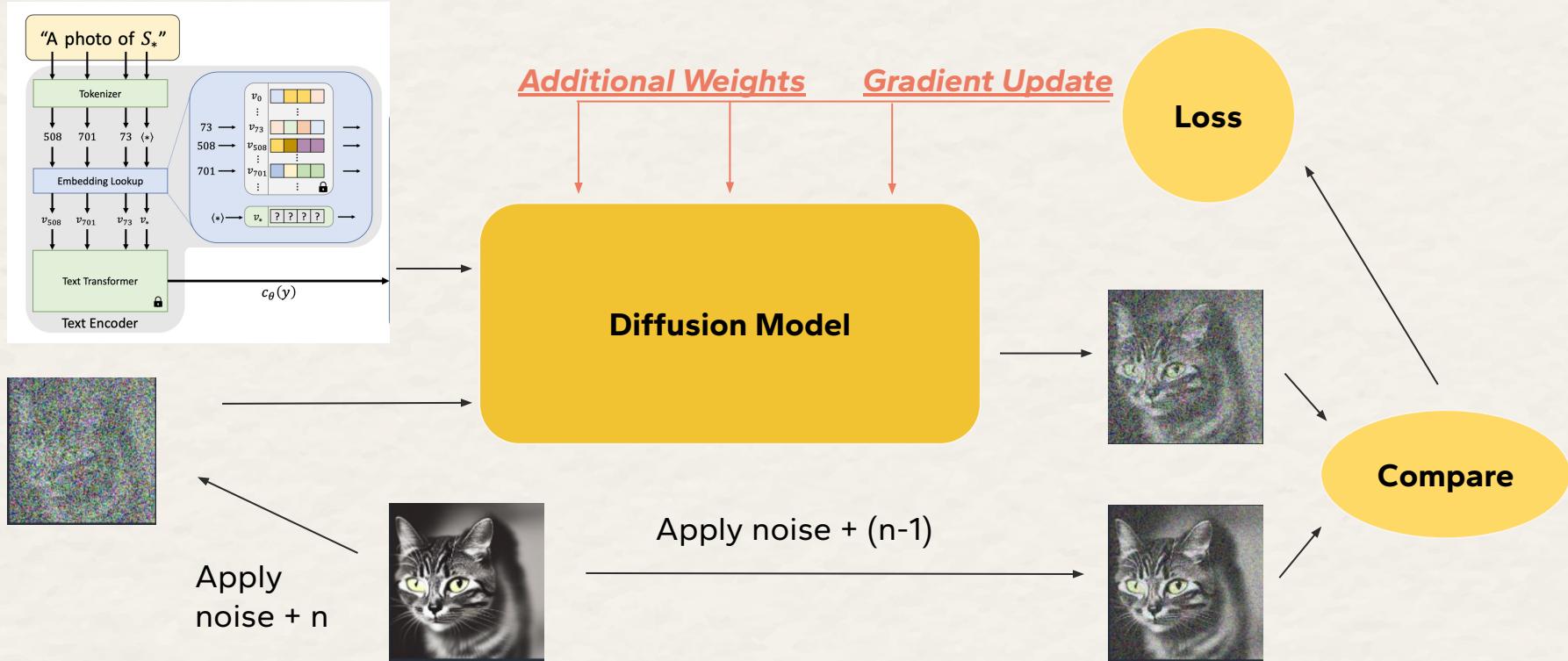
기존 pre-trained weights는 동결시키고 몇개의 dense layers만 학습하는데 weight를 low rank로 분해한 matrices만을 optimization

Diffusion model에 약간의 가중치를 더하고 수정된 모델이 개념을 이해할 때까지 훈련

빠르고 용량이 적음



Dreambooth 기반 LoRA



Dreambooth 기반 LoRA

크롤링한 데이터를 프롬프트화하여 200개 준비

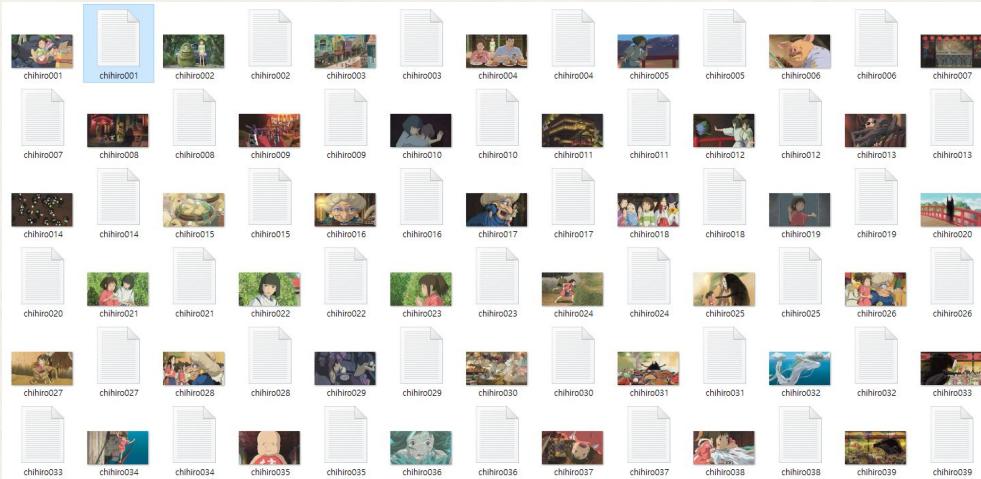
Train batch size:4

Learing rate:0.0001

Optimizer:AdamW8bit

Rank:8

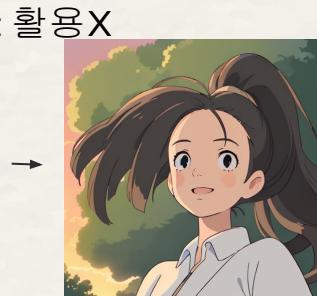
로 훈련시켜 LoRA파일을 얻음



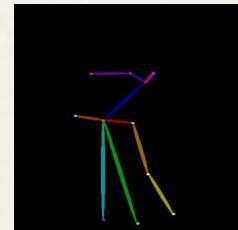
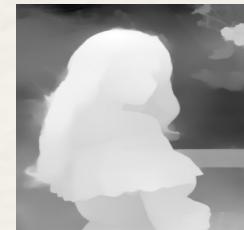
```
"LoRA_type": "Standard",
"adaptive_noise_scale": 0,
"additional_parameters": "",
"block_alpha": "",
"block_size": 0,
"block_zero_threshold": "",
"bucket_no_scale": true,
"bucket_res_stes": 64,
"cache_latents": true,
"cache_latents_to_disk": false,
"caption_dropout_every_n_epochs": 0.0,
"caption_dropout_rate": 0.0,
"caption_extension": "",
"clip_skip": 1,
"color_aug": false,
"conv_alpha": 1,
"conv_block_aliases": "",
"conv_dil": 1,
"decompose_both": false,
"dim_from_weights": false,
"dom_lr_weight": "",
"dump_b64": true,
"epoch": 1,
"factor": -1,
"flip_aug": false,
"full_lora": false,
"grad_norm": 1.0,
"gradient_accumulation_steps": "1",
"gradient_checkpointing": false,
"keep_tokens": "0",
"learning_rate": 0.0001,
"logging_dir": "C:/Users/ravra/crawled_lig/data/log",
"lr_scheduler": "cosine",
"lr_scheduler_args": "",
"lr_scheduler_warmup_steps": "",
"lr_scheduler_power": 1.0,
"max_bucket_reso": 2048,
"max_data_loader_n_workers": "0",
"max_resolution": "512,512",
"max_timestep": 1000,
"max_token_length": "75",
"max_train_epochs": "1000",
"mixed_precision": "fp16",
"new_eff_attn": false,
"mid_lr_weight": "",
"min_bucket_reso": 256,
"min_lr_weight": 0,
"n_timestep": 1000,
"mixed_precision": "fp16",
"model_list": "custom",
"module_dropout": 0,
"multi_res_noise_discount": 0,
"num_inference_iterations": 0,
"network_alpha": 1,
"network_dila": 8,
"network_dropout": 0,
"no_token_padding": false,
"noise_offset": 0.0,
"noise_type": "Original",
"num_col_threads_per_process": 2,
"optimizer": "AdamW8bit",
"optimizer_args": "",
"output_dir": "C:/Users/ravra/crawled_lig/data/model",
"output_name": "Mika.w2v",
"persistent_data_loader_workers": false,
```

Image 생성

결과



ControlNet 활용 O(depth,openpose)

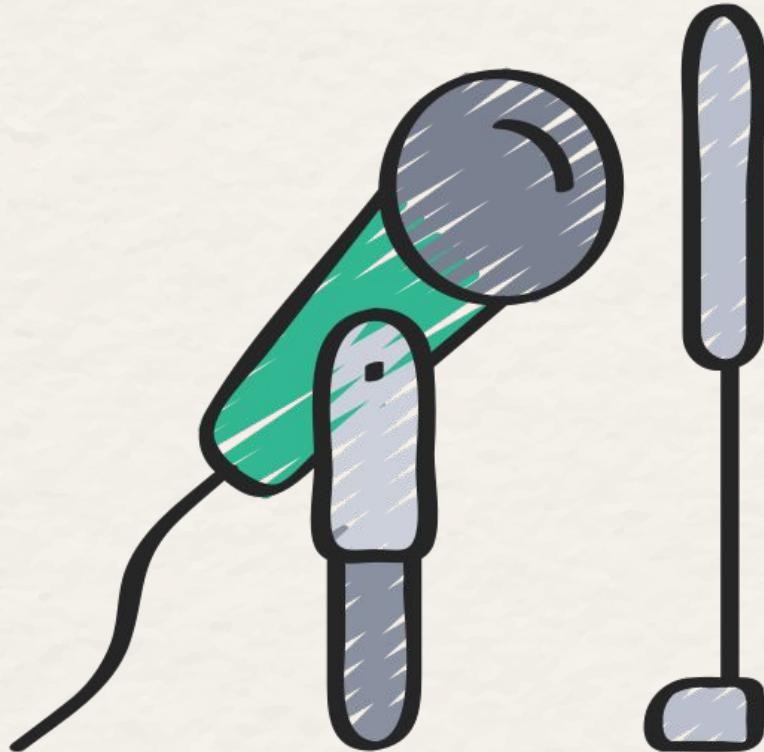


ControlNet 활용 O(openpose)



원했던 화풍이 제대로 나온 것을 확인

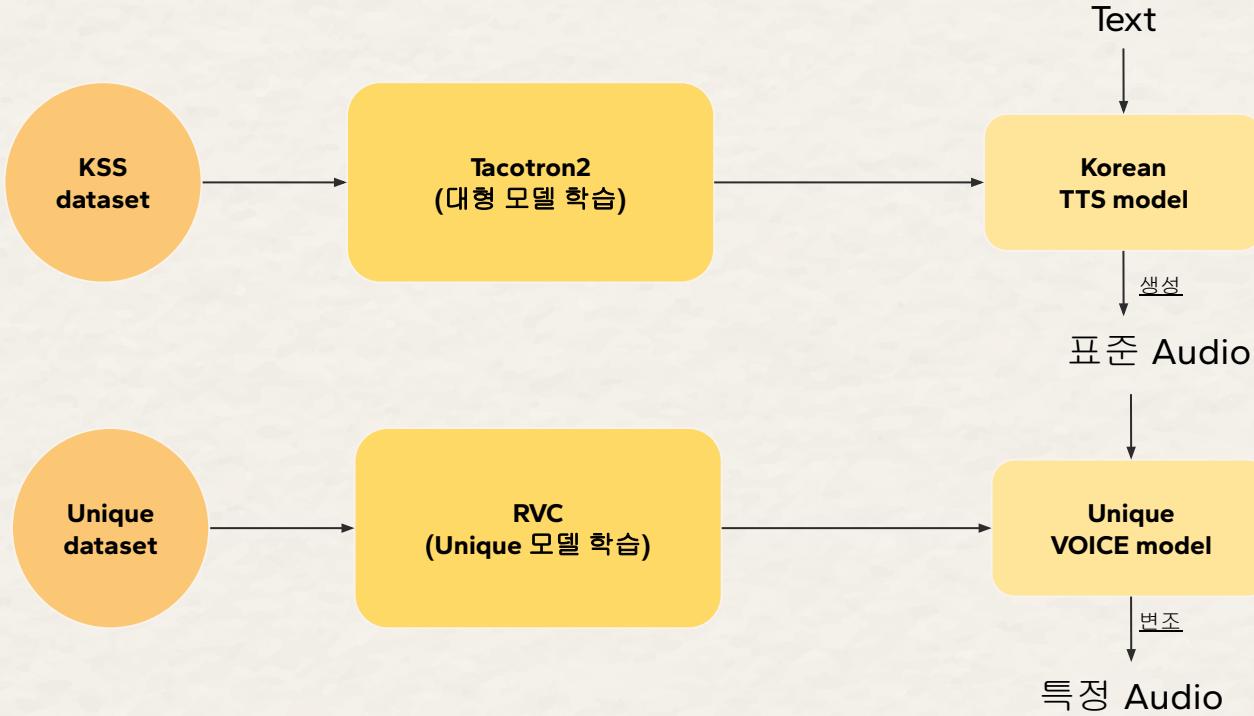
ControlNet을 사용하여 포즈와 형태를
잡음



음성 생성

?

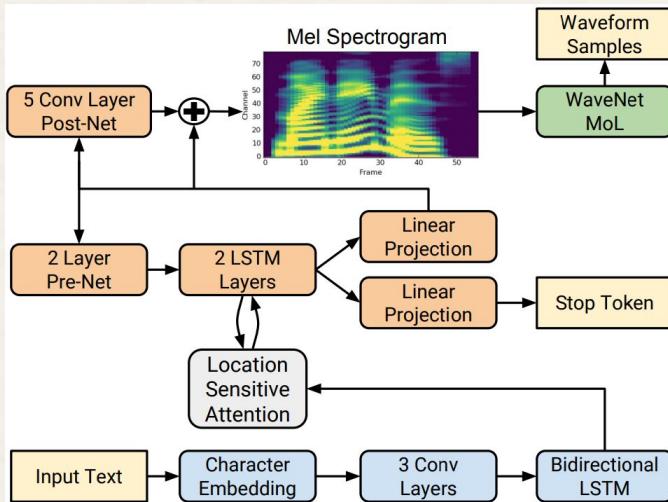
생성 흐름도



?

대형 음성 모델 생성

Tacotron2



좋은 정확도를 보여 주나,
너무 긴 학습시간 ...

=> pretrained 모델을 활용





‘특정’ 음성 모델 생성

| | Transfer-Learning | Diff-SVC | RVC |
|---------|-------------------|------------------|------------------|
| 기술 | 기존 모델에 전이 학습 | Stable-Diffusion | 음성 변조 |
| 정확도 | 높음 | 보통 | 보통 |
| 학습시간 | 매우 느림 | 빠름 | 매우 빠름 |
| 필요 데이터셋 | 고음질 많은 데이터 | 고음질 적당한 데이터 | 저음질 가능 작은 데이터 |

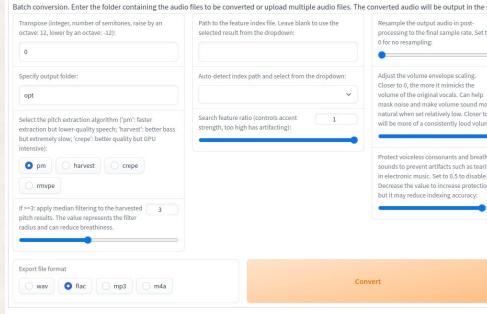


작은 데이터 양으로도, 빠르게 학습할 수 있는
‘RVC’를 사용해 음성 모델을 생성





‘특정’ 음성 파일 생성



Text를 입력하면
‘Korean TTS model’로
 표준 Audio를 생성

표준 Audio를 입력하면
‘Unique VOICE model’로
 특정 Audio로 변조

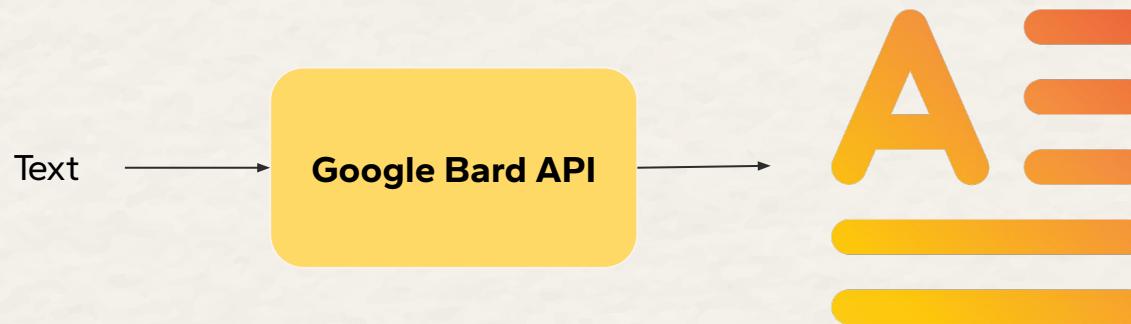


Ti

텍스트
생성



생성 흐름도





LLM 활용

추억 키워드
“눈 내리는 숲의 밤”



옛날 옛적에, 눈 내리는 숲에 사는 한 마리의 토끼가 있었습니다. 토끼는 숲에서 가장 빠른 동물이었지만, 매우 외로웠습니다. 토끼는 친구가 생기기를 바랐지만, 숲의 다른 동물들은 토끼를 무서워했습니다.

토끼는 숲의 가장 높은 나무 꼭대기에 올라가 밤하늘을 바라보았습니다. 별이 빛나는 밤하늘은 너무나 아름다웠고, 토끼는 행복했습니다. 토끼는 숲의 다른 동물들이 자신을 무서워하는 이유를 몰랐습니다. 토끼는 자신은 착한 동물이라고 생각했습니다.

토끼는 밤하늘을 바라보며 노래를 불렀습니다. 토끼의 노래는 매우 아름다웠고, 밤하늘의 별들은 토끼의 노래에 맞춰 춤을 추었습니다.

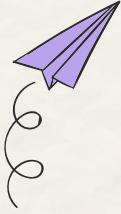
토끼의 노래를 들은 숲의 다른 동물들은 토끼가 무서운 동물이 아니라는 것을 알게 되었습니다. 동물들은 토끼에게 다가가 친구가 되었고, 토끼는 더 이상 외롭지 않았습니다.

토끼와 동물들은 밤하늘을 바라보며 노래를 불렀습니다. 토끼와 동물들은 모두 행복했습니다.



Text를 입력하면
LLM인 ‘**Google Bard**’를 통해
동화 내용으로 반환한다.





플랫폼

AI가 접목된 차세대 추모 플랫폼:
플랫폼 기억



기술 스택



Back

FastAPI



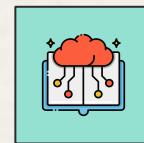
Front

HTML / CSS /
JavaScript



Database

MongoDB



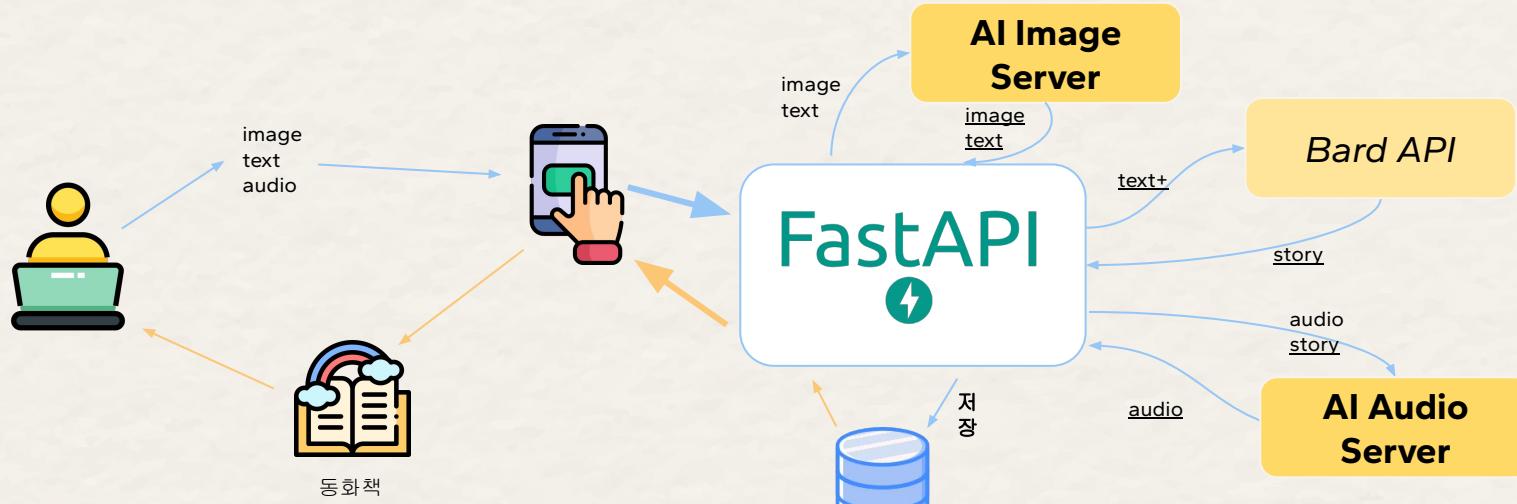
DeepLearning

StableDiffusion /
RVC / Bard





플랫폼 구조



04

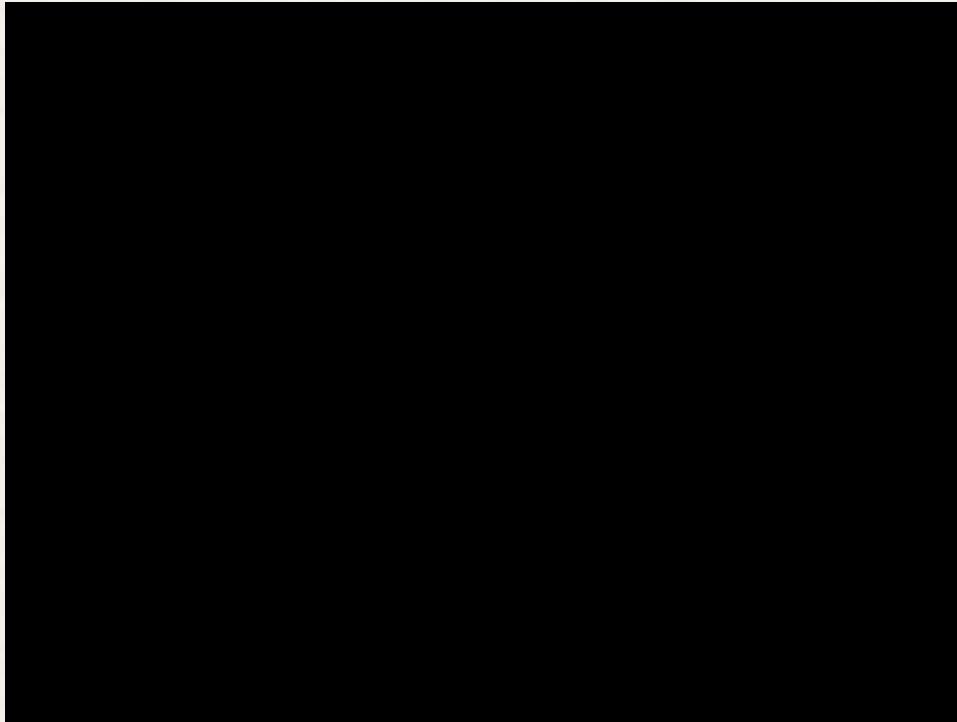
시연 영상

플랫폼 기억, 프로토타입 시연 영상



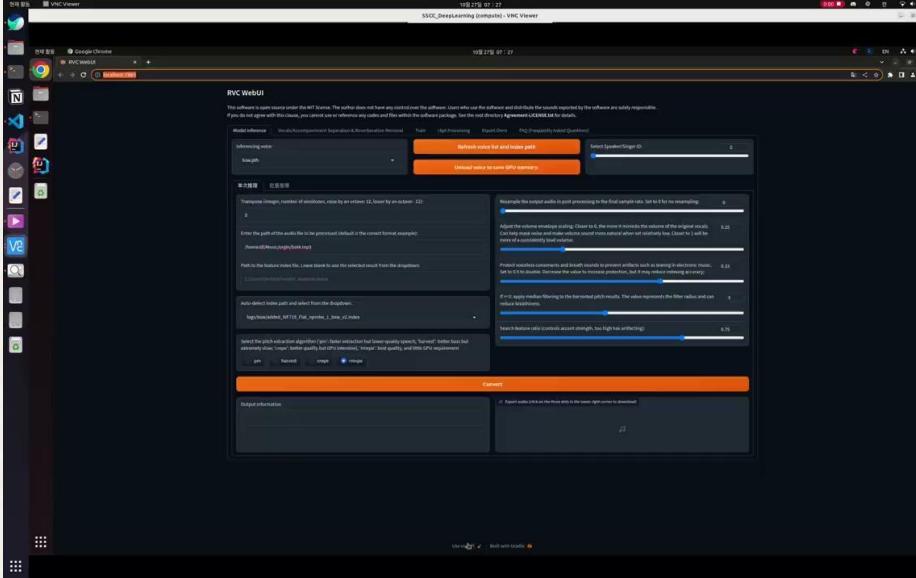


시연 영상



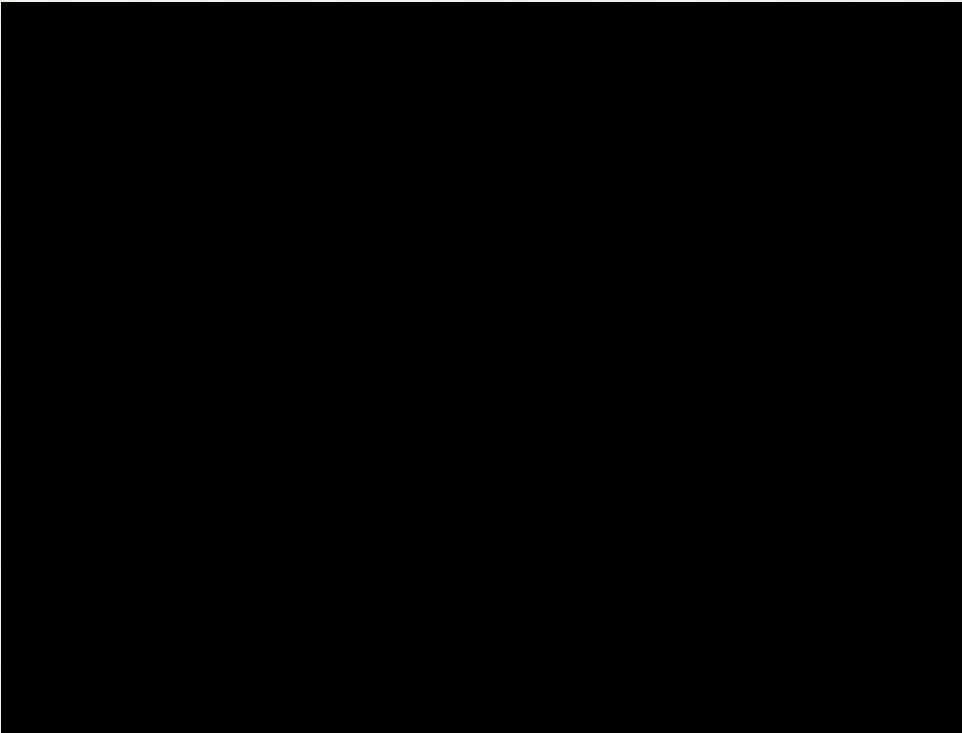


시연 영상





시연 영상





팀 소개



백승우

AI 음성 / 텍스트
플랫폼 개발



박성호

AI 이미지
서류 작성



참고 문헌

Image:

- <https://arxiv.org/abs/2002.05202>
- <https://arxiv.org/abs/1605.05396>
- <https://arxiv.org/abs/2106.09685>
- <https://huggingface.co/>



Sound:

- https://pytorch.org/hub/nvidia_deeplearningexamples_tacotron2/
- <https://github.com/RVC-Project/Retrieval-based-Voice-Conversion-WebUI>

Text:

- <http://bard.google.com>
- <https://aws.amazon.com/ko/what-is/large-language-model/>

Web:

- <https://fastapi.tiangolo.com/ko/>



감사합니다!

이어서 질의응답 진행하겠습니다



CREDITS: This presentation template was created by [Slidesgo](#), and includes icons by [Flaticon](#), and infographics & images by [Freepik](#)

Please keep this slide for attribution

