

Re:cover

생성형 AI 기반 앨범아트 모션 생성

발표자: 고대영

CV-04조 Seevalue

고대영, 고재완, 양지훈, 최호준

INDEX

프로젝트 배경

문제 인식 및 아이디어

프로젝트 전체 구조

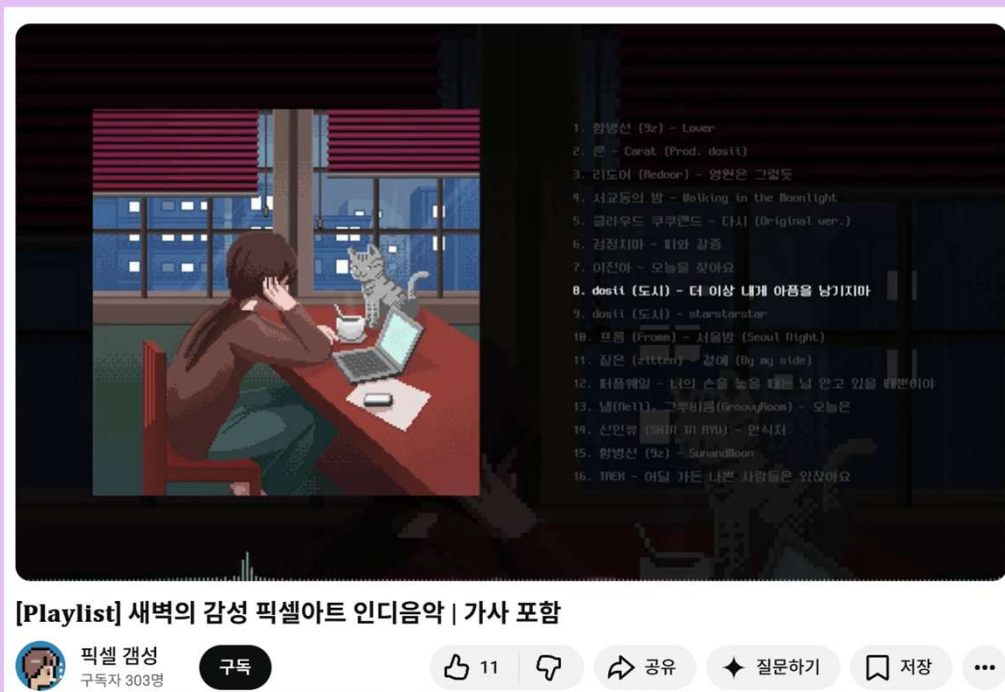
VLM-Judgement, Text layer separation, Image-To-Video

시연 영상

예시 앨범들의 Live2D 시각화

마무리 및 제언

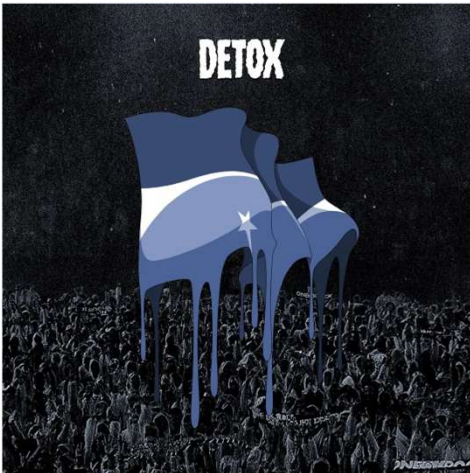
배경 아이디어 전개



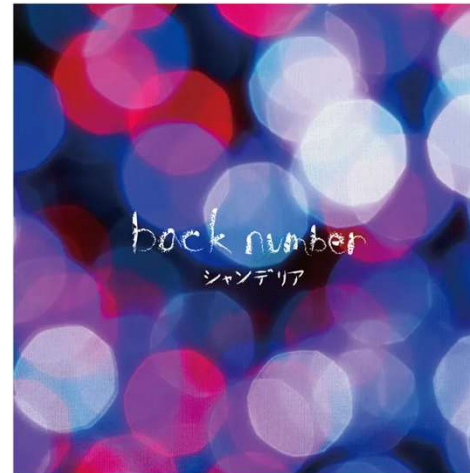
- Youtube에서 유행하는 pixel art frame loop 영상
- Spotify Canvas에서도 단순히 음악을 “듣는 것” 이 아닌, “보고 듣는” 경험을 제공함.

-> 아티스트의 앨범아트를, 살아있는 것처럼 움직이게 하자!

Stable Diffusion만을 적용할 경우

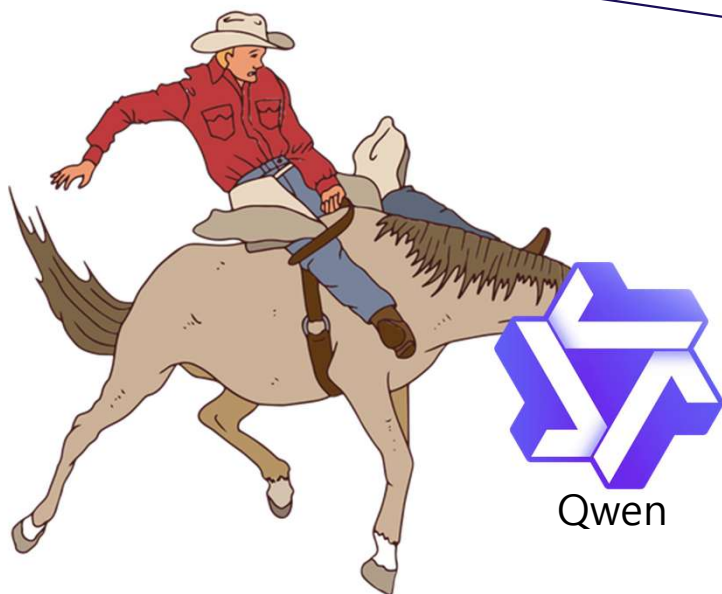


Distortion



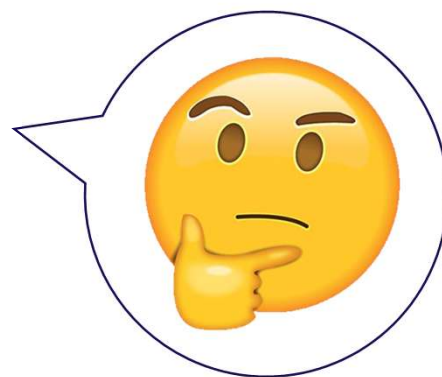
Data Bias

Text-Separate Judgement with VLM

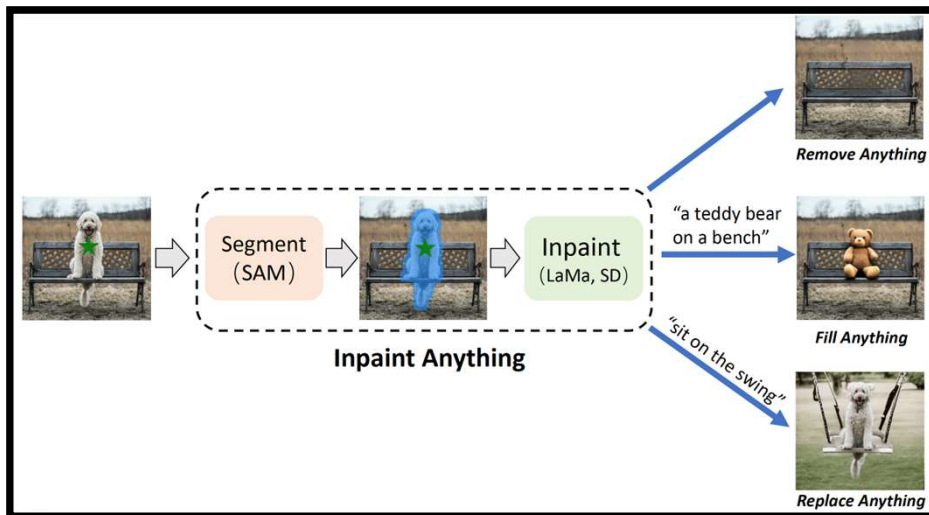


"이걸 좀 보거라.
텍스트를 떼 내어서
나중에 합성하는 게 낫겠냐?
아니면 놔두는 게 자연스럽겠느냐?"

*원문 일부발췌 후 의역



Text layer Separation



LaMa: Resolution-robust Large Mask Inpainting with Fourier Convolutions

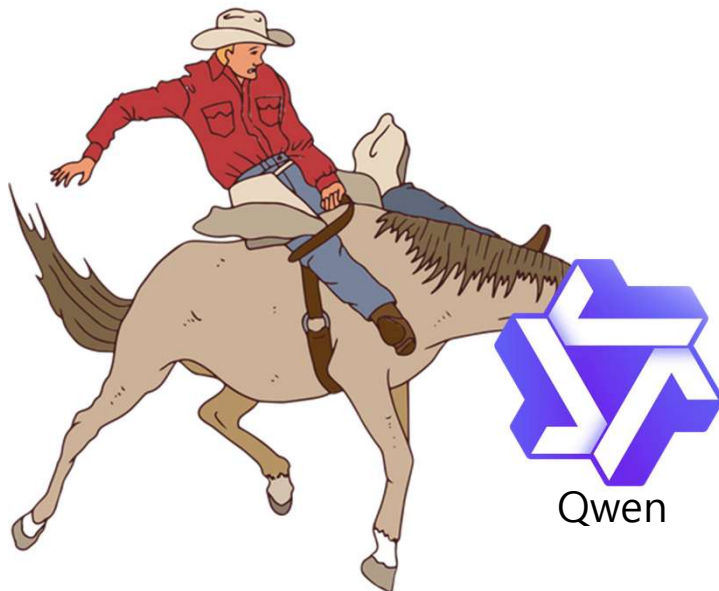


명도 중심 분석, 픽셀 주변의 local black과 대조하여 글자 탐지

Text layer Separation – Example results



Image Judgement with VLM



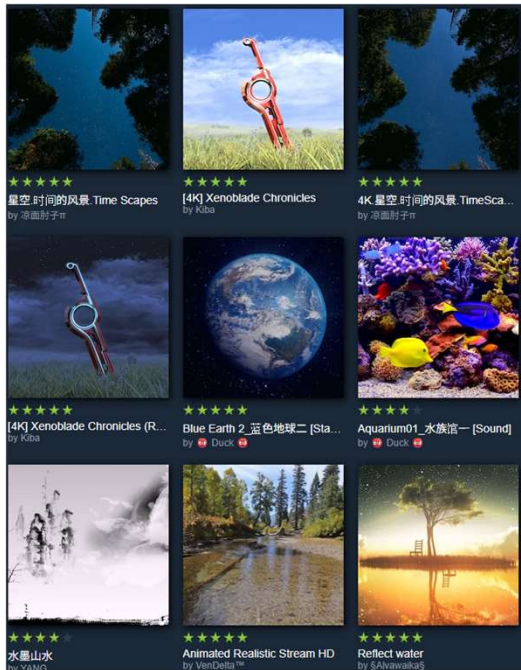
"Live2D 스타일 애니메이션을 위한 최고의 seamless loop 영상 프롬프트를 만들거라."

positive, negative prompt
각각 LoRA의 트리거 워드도 잊지말게나."

*원문 일부발췌 후 의역



Custom LoRA Training



1. 실사 배경, 애니메이션 스타일, 사람 형상 등 현실에 있을 법한 wallpaper dataset **약 40개 학습**

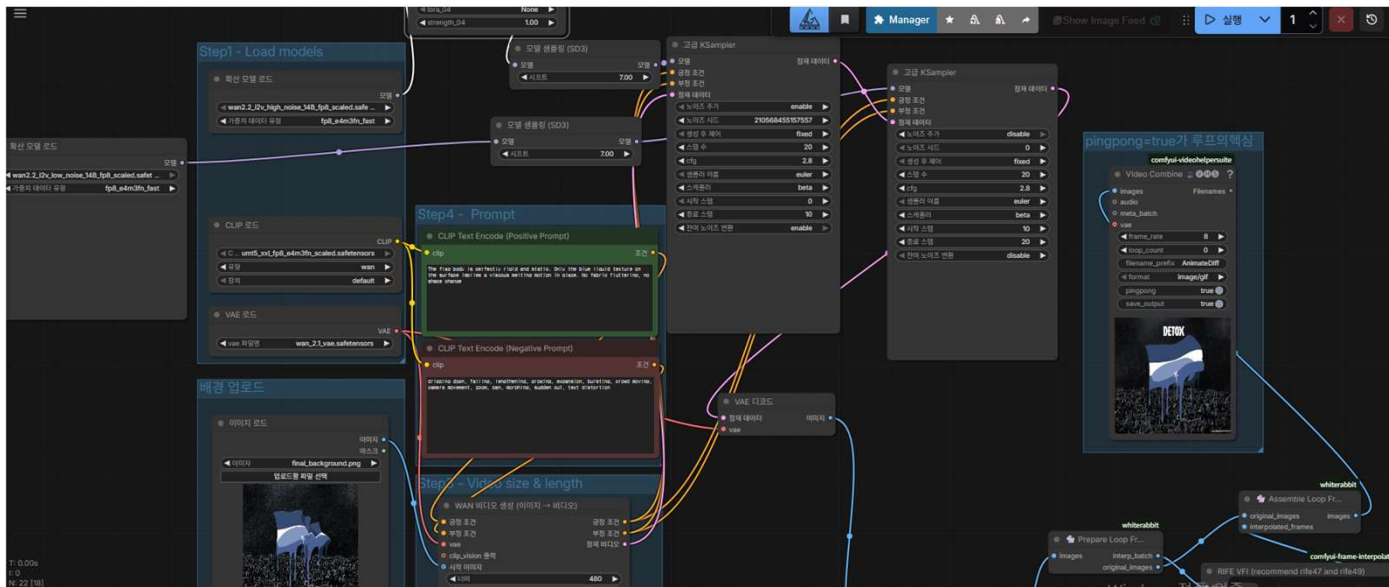
2. 1의 데이터를 해상도/프레임 조절 후 재학습

➔ 다른 버전의 custom LoRA 2개 학습 완료

➔ 기존의 오픈소스(live wallpaper LoRA) + custom 2개

총 3개의 LoRA를 활용하여 Image-To-Video 생성

ComfyUI Workflow with WhiteRabbit



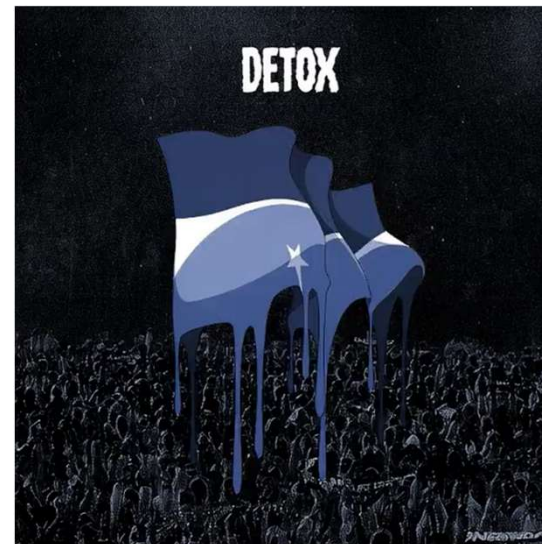
배경 첨부 – text layer 첨부 – prompt 입력 – LoRA 적용 – Sampler – Decode

ComfyUI Workflow with WhiteRabbit



Ping-pong Loop

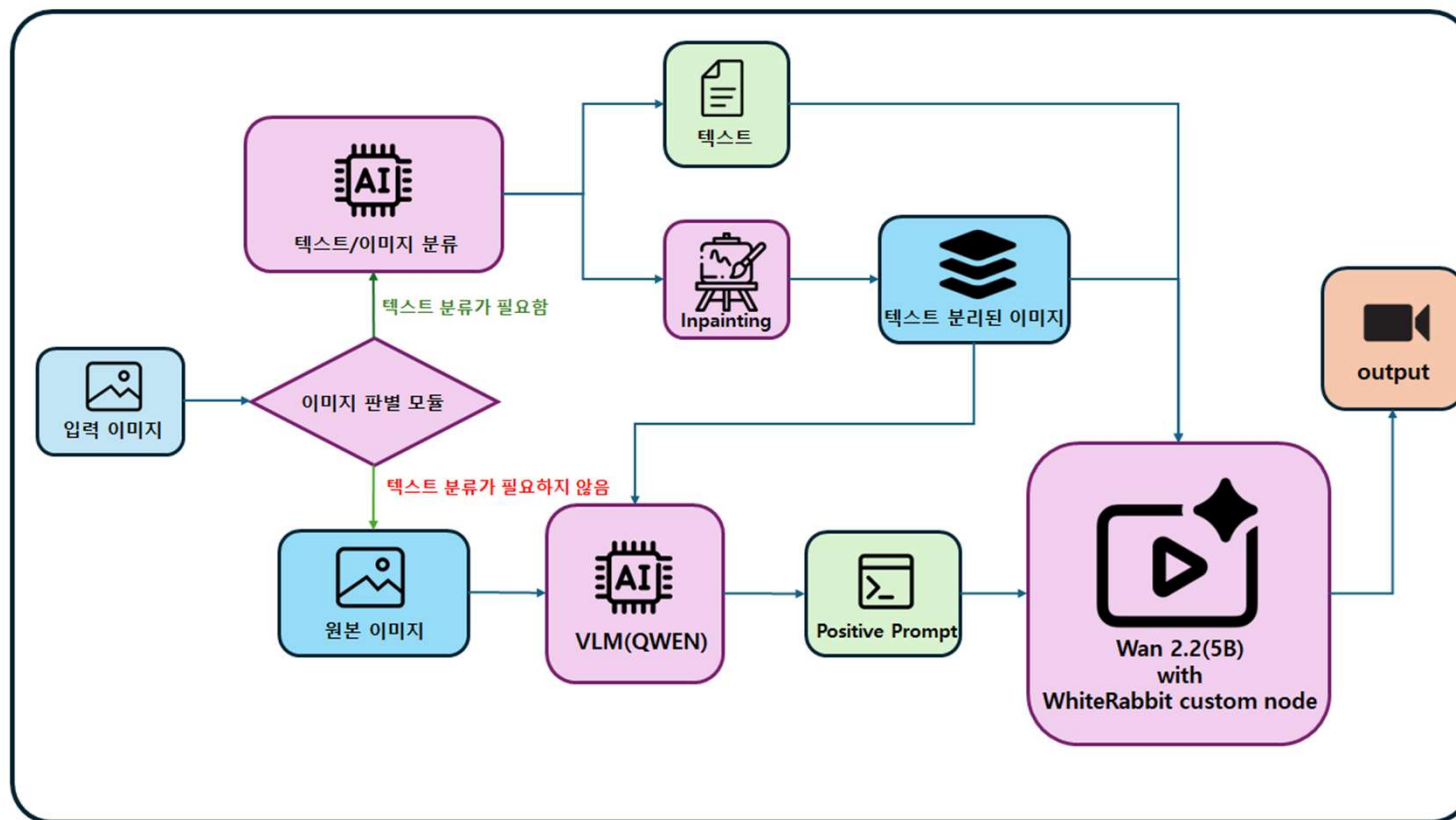
영상을 역재생하여 이어 붙임
-> 비가역적 움직임에서 부자연스러움



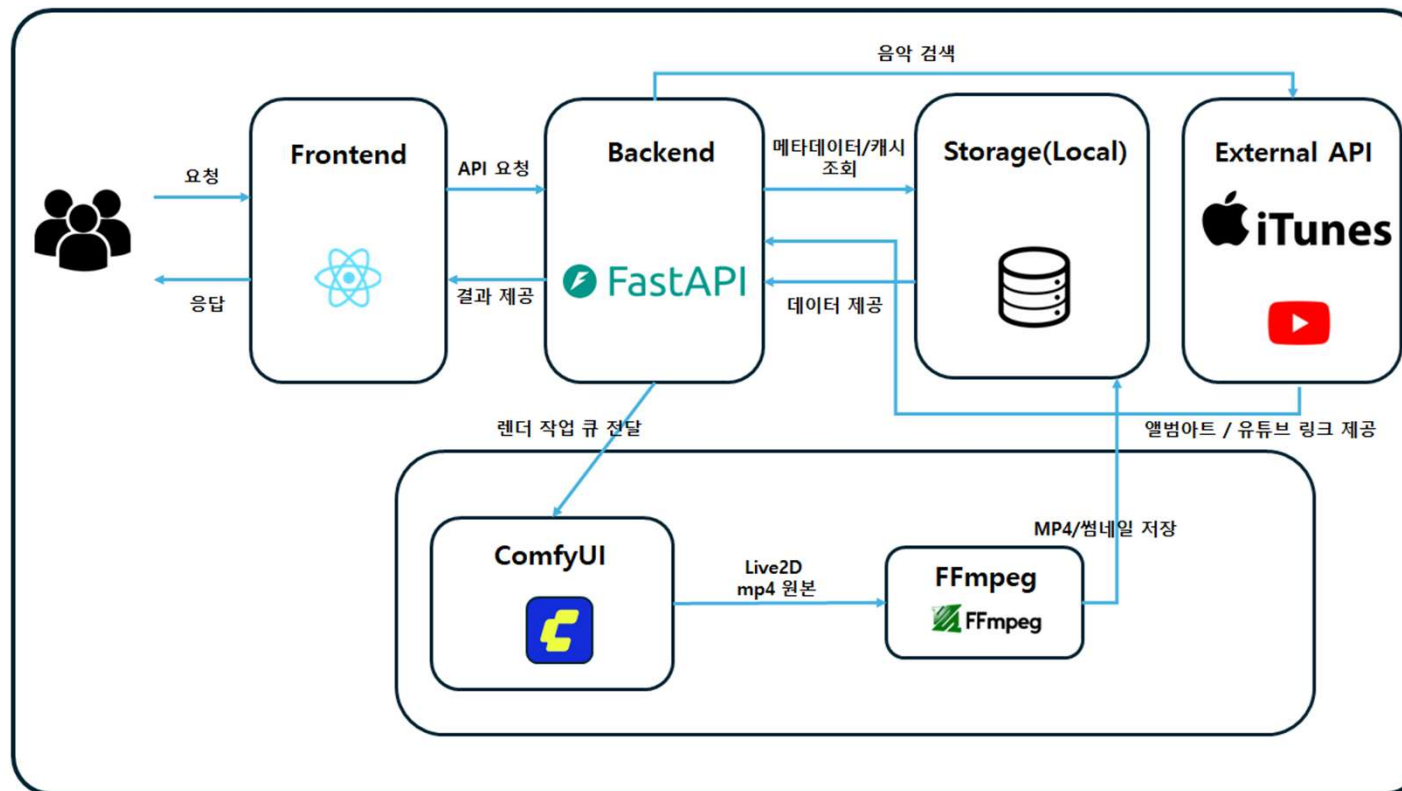
WhiteRabbit Loop

First Frame과 Last Frame 사이를 보강
보강된 프레임을 부드럽게 블렌딩

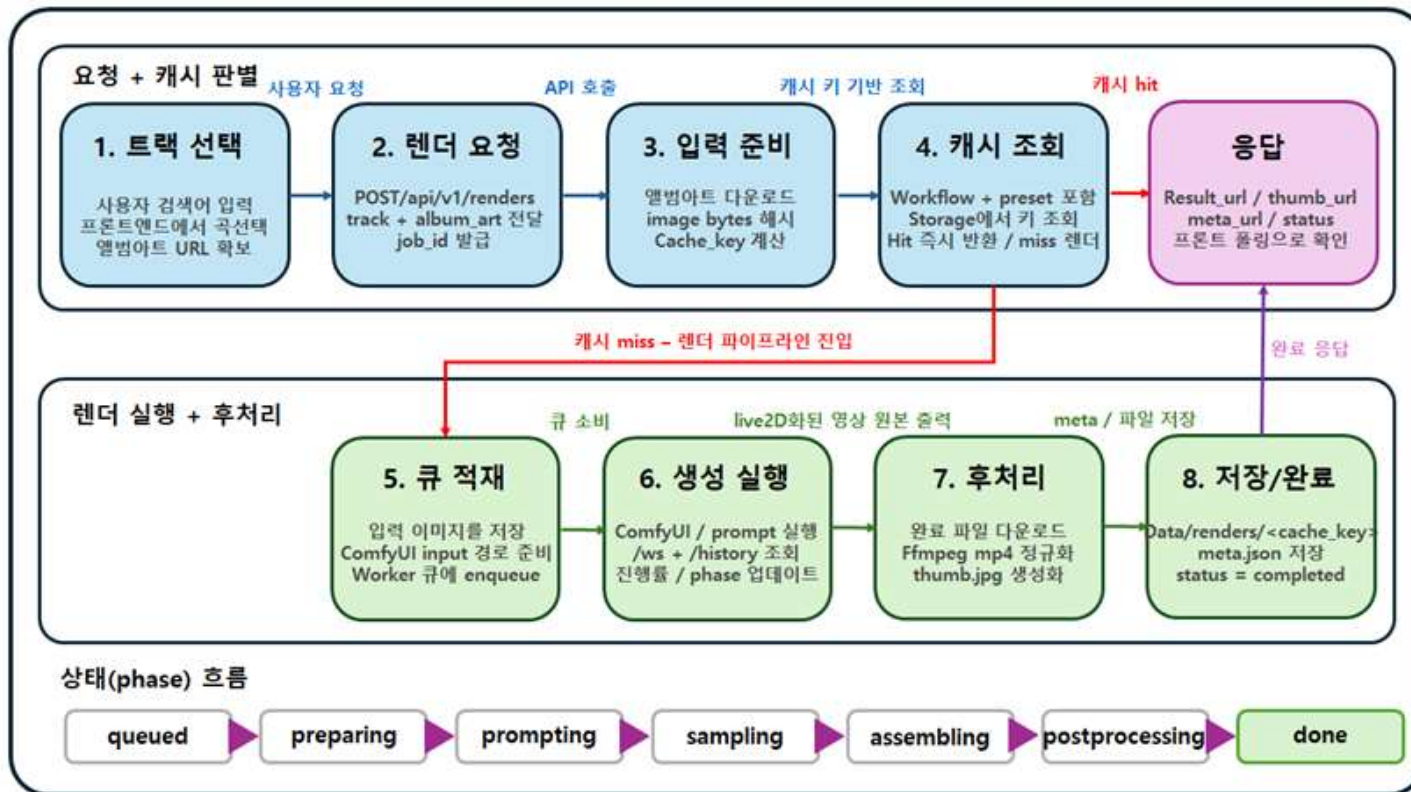
ComfyUI Workflow



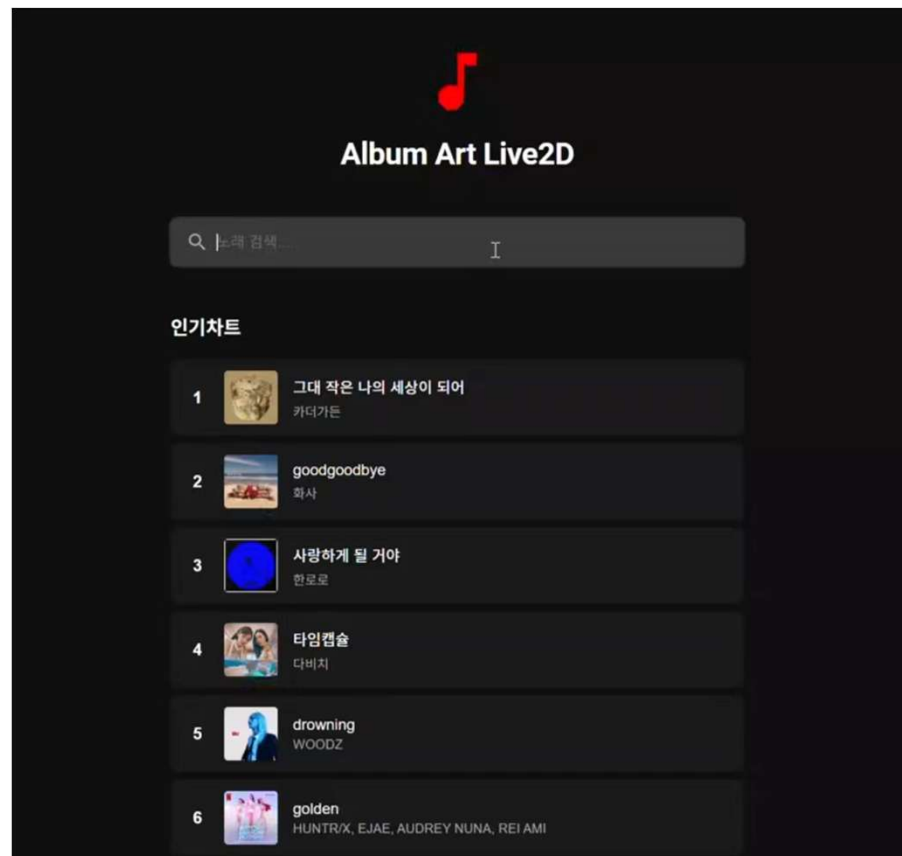
System Architecture & Tech Stack



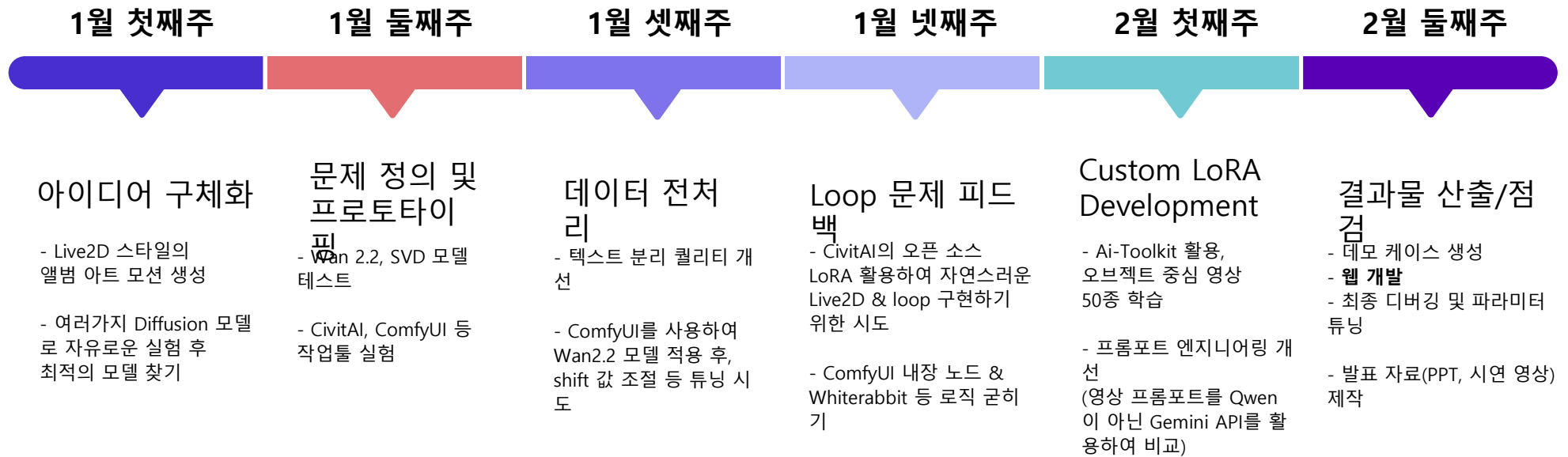
Rendering Pipeline



시연 영상



Process Timeline



Future Work

1. Multi-Modal Loop

- 앨범뿐 아니라 노래의 BPM, 장르, 가사 분위기 등 종합적인 요소까지 반영

2. Format Flexibility

- 1:1 비율의 앨범보다 확장하여 16:9 / 9:16의 디스플레이에 적용하도록 확장

3. Advanced Transition Effects

- human score을 통한 모델 성능 개선, 사용자 기반 맞춤형 출력물 생성

End of Document
Thank You.