

Song DALL-E

빅데이터 지능형 서비스 개발 11, 12회차

Index

Introduction

- 프로젝트 개요 - 역할 분담

Data Processing

- 데이터 수집

- 데이터 전처리

- EDA

Model Training

- 모델 선정

- 모델 학습

Model Evaluation

- 모델 평가

- Text-to-Image

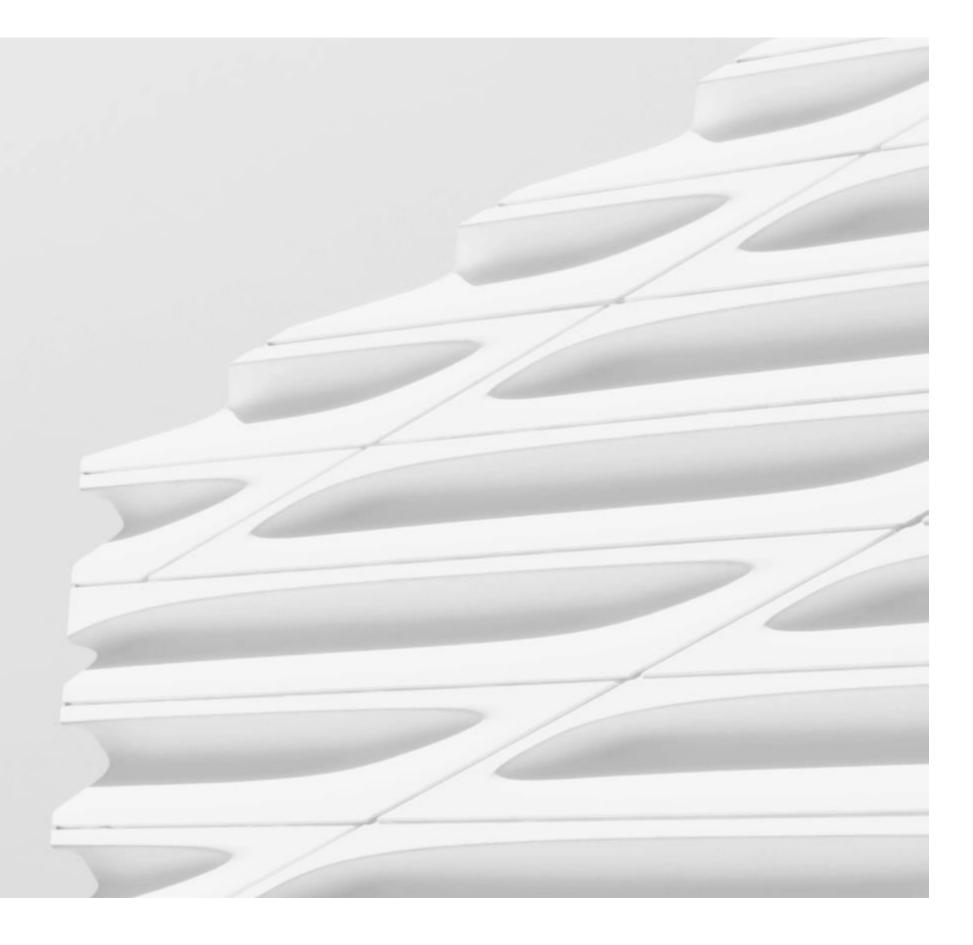
Service Development

- 서비스 청사진

- Django

Conclusion

- 기대효과 - 한계점

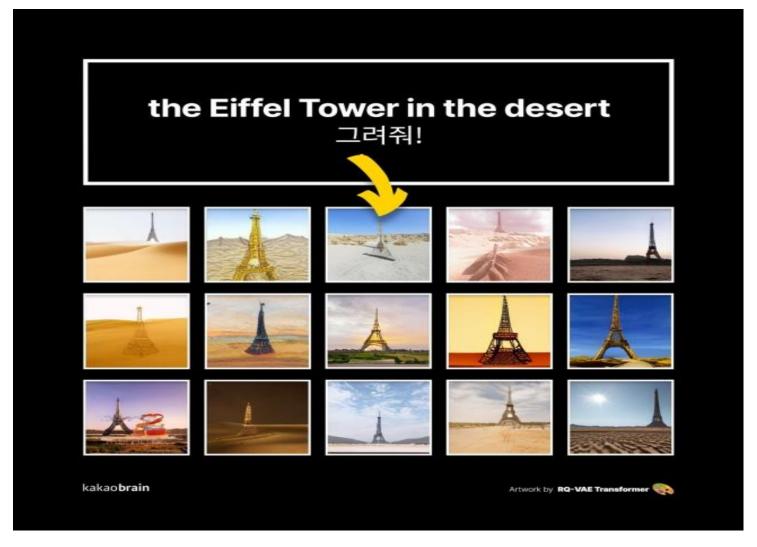


✓ 엔비디아, 당신의 상상력을 명작으로 그려내는 딥러닝 기술 공개



[사진1] 텍스트 '안개가 낀 날 키 큰 나무로 둘러싸인 평화로운 호수'에 의해 GauGAN2에서 생성된 이미지 출처: NVIDIA Vlog

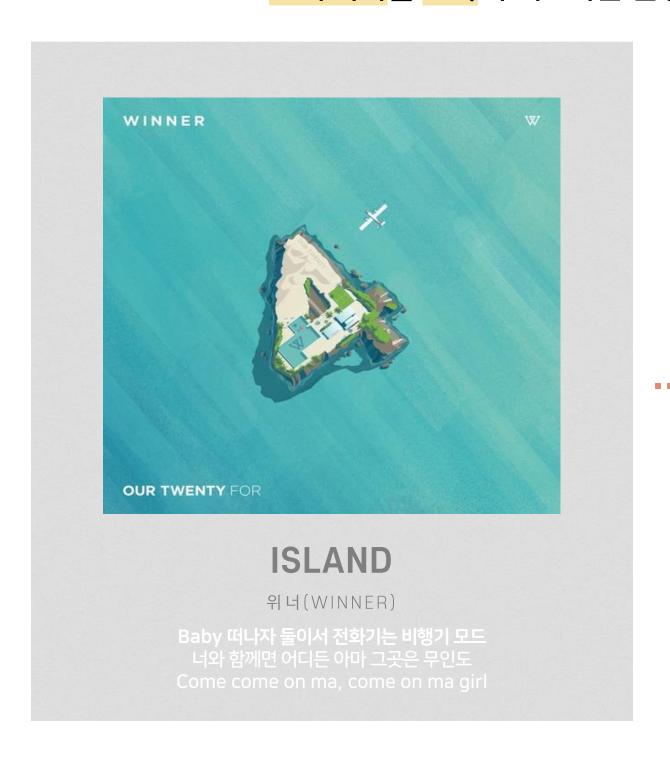
✓ [아무Tech]개·고양이 구분 못하던 인공지능이 어느새 '화가'로



[사진2] 이미지 생성 AI모델 'RQ-Transformer'에 '사막에 있는 에펠탑'을 입력하자 등장한 이미지 출처: 카카오브레인

이러한 Text-to-Image AI 기술을 <mark>노래 가사</mark>에 활용해보면 어떨까?

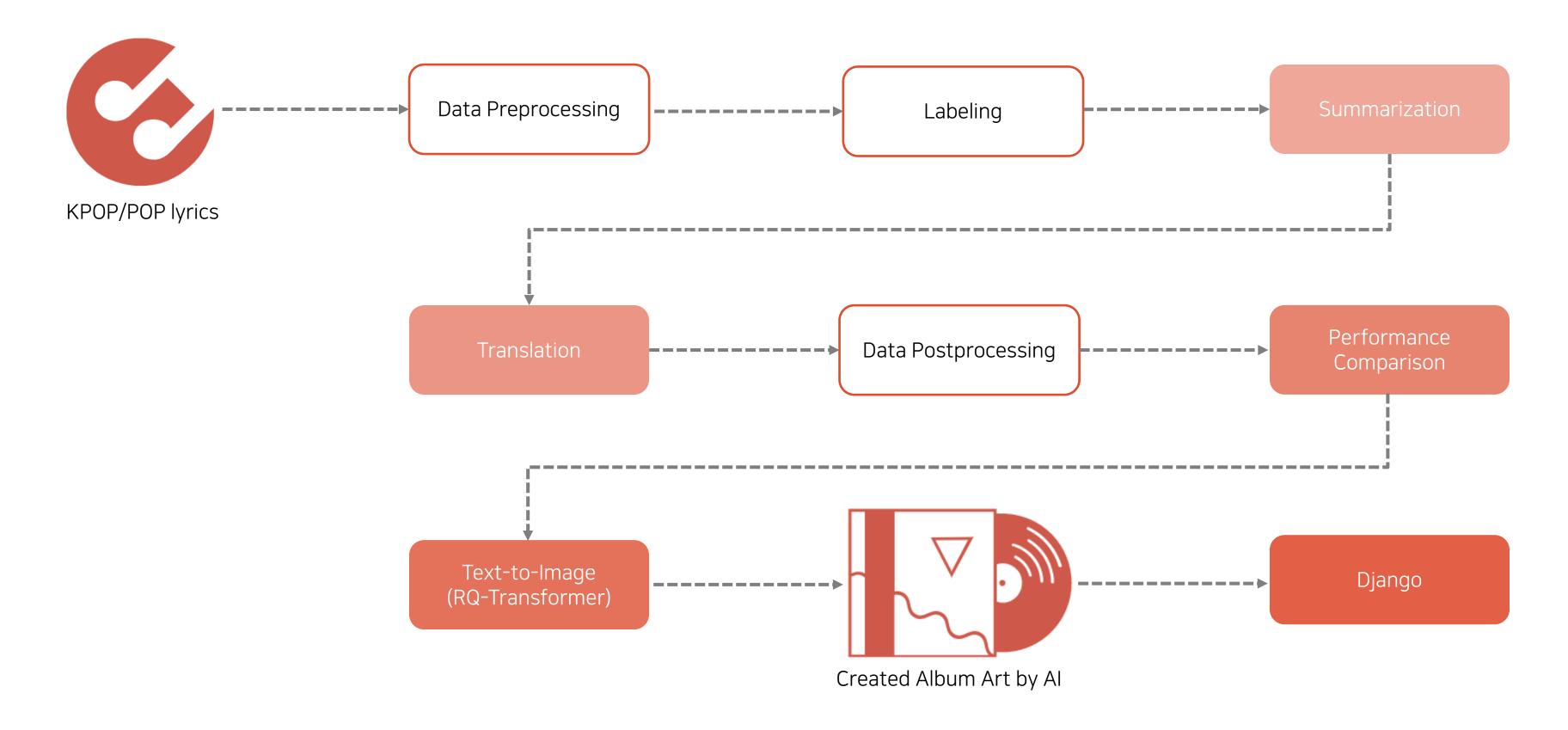
<mark>노래 가사</mark>를 <mark>요약</mark>해 대표하는 문장을 도출하고, 이를 통해 <mark>앨범 커버 이미지</mark>를 추천 해주는 시스템을 구축하자



Whenever I am with you, maybe it's a deserted island, let's leave.



[사진3] ISLAND 노래 가사 기반 RQ-Transformer 학습 결과



인원은 적지만 좋은 팀, *나는자연인이*조 입니다.



팀장 | 이세원

- 가사 크롤링
- 데이터 전처리
- 모델 취합
- 발표



부팀장 | 권미래

- 가사 크롤링
- 데이터 전처리
- 웹 구현
- 발표 자료 제작



팀원 | 박주영

- 가사 크롤링
- 데이터 전처리
- 모델링
- 발표 자료 제작



팀원 | 최준섭

- 가사 크롤링
- 데이터 전처리
- 모델링
- 웹 구현

✓ 멜론 크롤링

- 차트 파인더를 이용해 연도별 월간 차트를 크롤링
 - 1960년대부터 2020년대까지
 - 단, 가사가 없는 노래는 try, except 구문으로 공백을 집어넣음.
- **장르/스타일**: 국내종합(kor+eng) / 해외종합(eng)

✓ 데이터 수집 결과

- Train Data Set: 8,102곡 | 한국 노래(한글+영어 혼합)
- Validation Data Set: 2,099곡 | 외국 노래(영어)

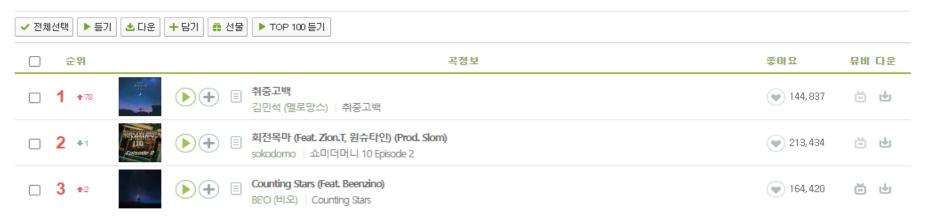
Melon

차트 파인더



검색

2022년 01월 국내종합 월간차트



[사진4] 멜론 차트 파인더

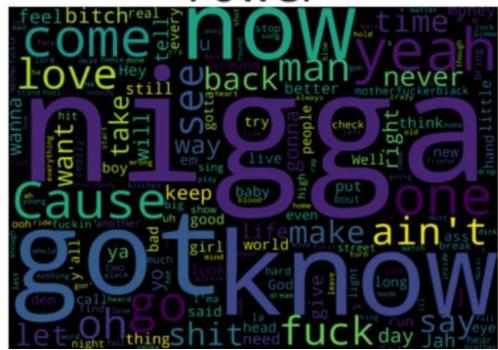
✓ Preprocessing

- 1) 개행 문자(₩n) **마침표(.)로 변환**
 - 한 여름 밤의 꿈<mark>₩n</mark>가을 타 겨울 내린 눈 → 한 여름 밤의 꿈<mark>.</mark> 가을 타 겨울 내린 눈
- 2) 정규식 표현("[^'A-Za-z0-9가-힣'.]", ")을 이용해 영어가 아닌 **제2외국어, 특수 문자 제거**
 - 소리쳐 (say ho) → 소리쳐 say ho
- 3) 띄어쓰기 처리(SoyNLP 사용)
 - <mark>스쳐가고아무</mark> 말 못한 채 → <mark>스쳐가고</mark> 아무 말 못한 채
- 4) 반복 문자 제거
 - <mark>아아아</mark> 어려워 → <mark>아</mark> 어려워
- 5) 영어의 가사의 경우, 소문자로 모두 통일
 - When I was young → when i was young

✓ Postprocessing

- 1) 불용어 처리: 비속어, 혐오단어 등 제거
- 2) 문장 길이 제어: 요약 생성문에서 전체 문장 길이 선정

Power



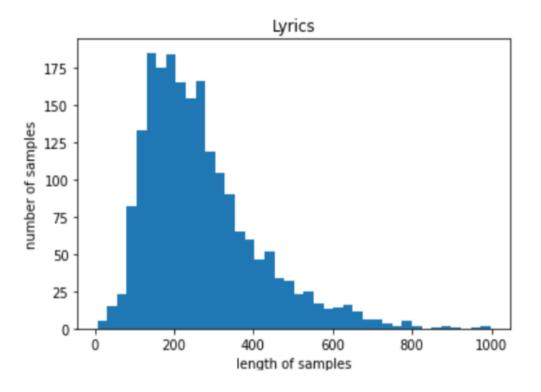
[사진5] 감성 분석 결과(비속어 도출 예시)

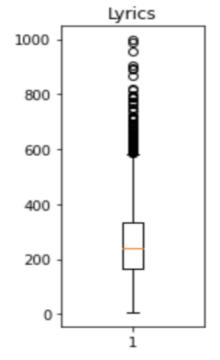
✓ Lyrics Length

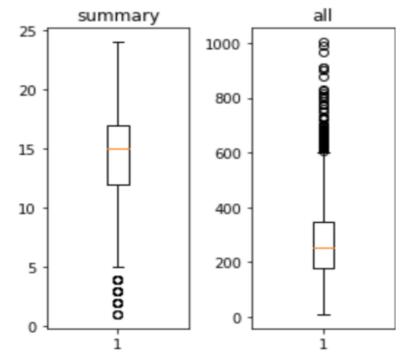
• Mean: 267

• Min/Max: 6 ~ 998

Max token: 512 1 token: 2~3글자수 Max Limit 1,500~2,000자





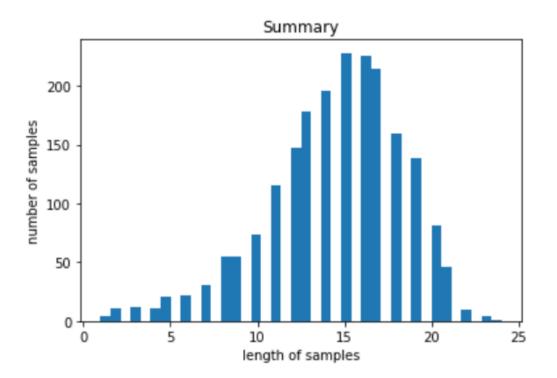


✓ Lyrics Summary

• Mean: 14

• Min/Max: 1 ~ 24

Mean: 14 1 token: 2~3글자수 Summary Min/Max 28~42



✓ Translation



- 1) 크롤링한 데이터 중, 특히 한국 노래에는 **한국어-영어의 혼합 가사가 존재**
 - → 원활한 학습을 위해 **단일 언어(영어)로 통일**이 필요
- 2) 요약 모델을 이용한 한국어 요약 문장 생성
- 3) 이미지 모델에 넣기 위해 papago API를 이용해 변역을 진행
 - → 영어로 된 문장을 넣었을 때 더 **정확한 이미지**가 나오기 때문

✓ 원본 가사(번역 이전)

Love love the stars Love love the moon

별 다를 것 없이 똑같은 공기 똑같은 침대에서 보이는 천장까지

왜 별 이유 없이 공허한 건지 그저 몇 시간 째 멍만 때리는 게 처음에는 배가 고픈가 했었는데 No way no way no way 요새는 칸예 새 앨범을 들어도 뻔해 뻔해 뻔해 TV를 틀어 놓은 채 채널만 돌려 보네 But bae 별 의미 없네

✓ 전처리된 가사(번역 이후)

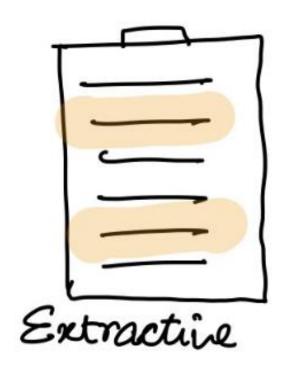
Love love the stars Love love the moon

Nothing's really different It's the same air It's the same bed Looking at the same ceiling

Why do I feel so empty
For no reason at all
It's been a few hours
Since I've been spacing out

At first
I thought I was just hungry
No way no way no way
These days
I listen to Kanye's new album
But it's just typical
I leave the TV on
But I'm just flipping the channels
But bae, there's not much meaning

✓ Extractive Text Summarization



입력 데이터 내에 존재하는 문장을 토대로 요약문 생성



나는 밥을 먹었다. <mark>순두부 찌개를 먹었</mark>는데 가성비도 좋고 맛있었다. ✓ Abstractive Text Summarization



입력 데이터를 토대로 요약문 예측 생성

<mark>나는 순두부</mark> 찌개를 맛있게 <mark>먹어</mark>서 배가 너무 부르다.

← Extractive Text Summarization



✓ Abstractive Text Summarization



입력 데이터를 토대로 요약문 예측 생성

앨범 이미지 추천은 **주관적이고 추상적인 영역**이라고 판단

따라서 보다 <mark>주관적 생성에 가까운 Abstractive 기반 Model을 사용</mark>

✓ BART (Bidirectional & Auto-Regressive Transformer)

Transformer (Self-attention)

"그 남자는 작은 보트로 큰 강을 건넜다."

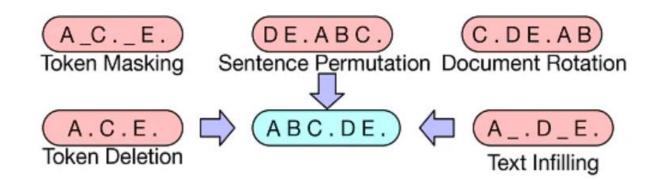
Auto-Regressive

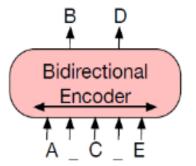
"그 남자는 작은 **'보트'**로 큰 강을 건넜다."

Bidirectional

"그 남자는 작은 [__]로 큰 강을 건넜다."

✓ Document corruption

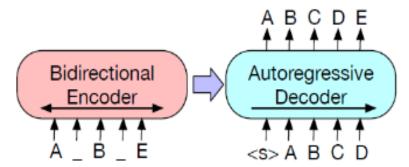




ABCDE Autoregressive Decoder <s> A B C D

(a) BERT: Random tokens are replaced with masks, and (b) GPT: Tokens are predicted auto-regressively, meaning the document is encoded bidirectionally. Missing tokens GPT can be used for generation. However words can only are predicted independently, so BERT cannot easily be condition on leftward context, so it cannot learn bidirecused for generation.

tional interactions.



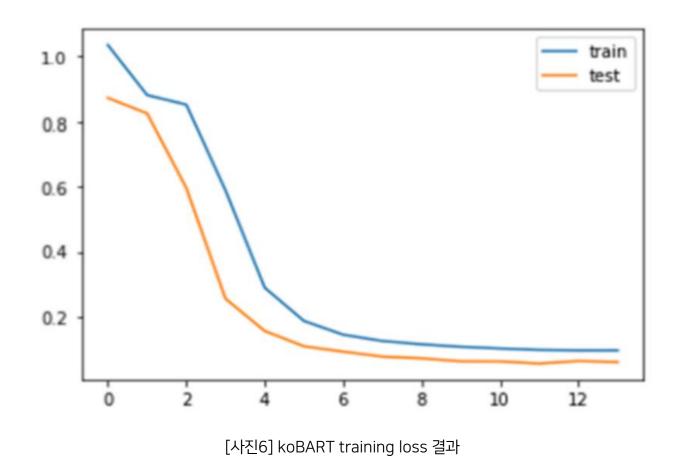
(c) BART: Inputs to the encoder need not be aligned with decoder outputs, allowing arbitary noise transformations. Here, a document has been corrupted by replacing spans of text with mask symbols. The corrupted document (left) is encoded with a bidirectional model, and then the likelihood of the original document (right) is calculated with an autoregressive decoder. For fine-tuning, an uncorrupted document is input to both the encoder and decoder, and we use representations from the final hidden state of the decoder.

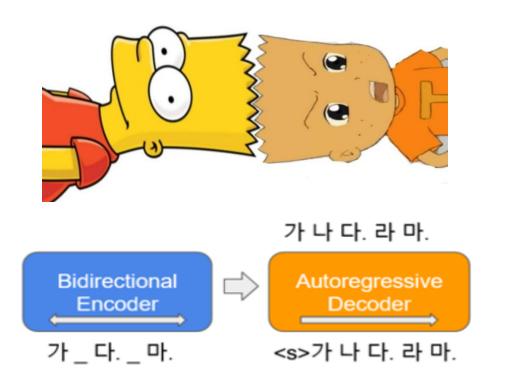
[사진?] Encoder - Decoder Structure



√ KoBART

- 문서 요약 테스크에 강한 한국어 BART(KoBART)
- Input text 종류로 한국어, 영어 둘 다 가능





Model	# of params	Туре	# of layers	# of heads	ffn_dim	hidden_dims
KoBART- base	124M	Encoder	6	16	3072	768
		Decoder	6	16	3072	768

[사진7] model parameter and layer

03 Model Training - 요약 모델 선정

나는자연인이조 🕍

✓ KoBART_Train

- 가사라는 Dataset 특징을 고려하여 Text Infilling 기법으로 학습
 - Text Infilling

여러 개의 text span 선택_ 하나의 [mask] Token화하는 것 Sentence Permutation의 일종

Sentence Permutation란?

문장 단위로 나눈 후에 Random Shuffle

Pretrained DataSet_ (KoBART) (CNN/DailyMail)

Train Data Set: 286,817

Validation Data Set: 13,368

(Kor Wiki. 외 Kr Text (약 40GB))

Crawled DataSet

Train Data Set: 8,102 한국 노래(한글+영어 혼합)

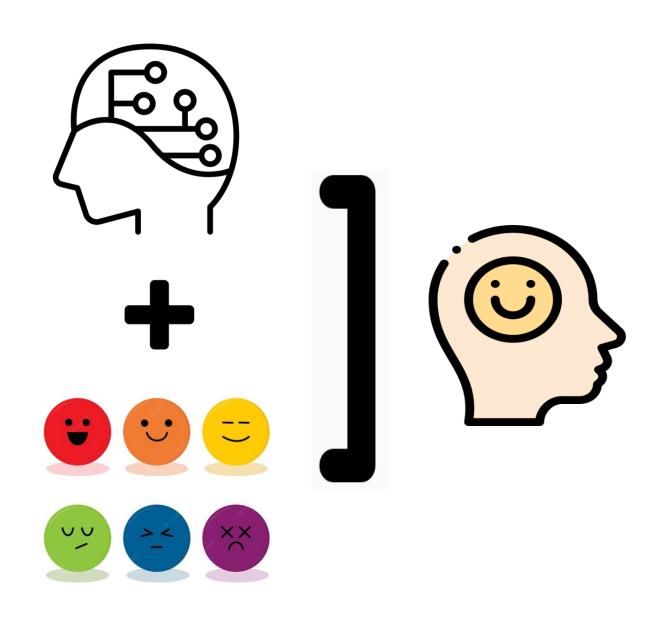
Validation Data Set: 2,099 외국 노래(영어)

✓ KoBART_Fine Tune

- Eos_token_id: 문장 마지막 token화 개수
- Max_length: 생성할 글자 수
- Num_beams: 토큰 앞뒤 참고할 토크 개수
- Length_penalty: 1.0(default)보다 클 경우 더 긴 문장 생성 유도

✓ KoBERT 감성 분석 모델

• 감정을 분석할 때, 다중 분류가 가능한 것이 강점



✓ Modeling

- Labeling한 data 토대로 5가지 감성 라벨 중, 가장 가까운 감성 출력
- Label: sentiment [Sadness, Power, Tension, Joyful, Calmness]
- Train: 약 3,000곡의 가사 Dataset에 Labeling 학습 진행
 (약 600곡으로 총 5가지 감성 Labelling)

✓ Setting parameters

```
max_len = 512

batch_size = 8

warmup_ratio = 0.1

num_epochs = 100

max_grad_norm = 1

log_interval = 200

learning_rate = 5e-5

class BERTClassifier(nn.Module):

def __init__(self,

bert,

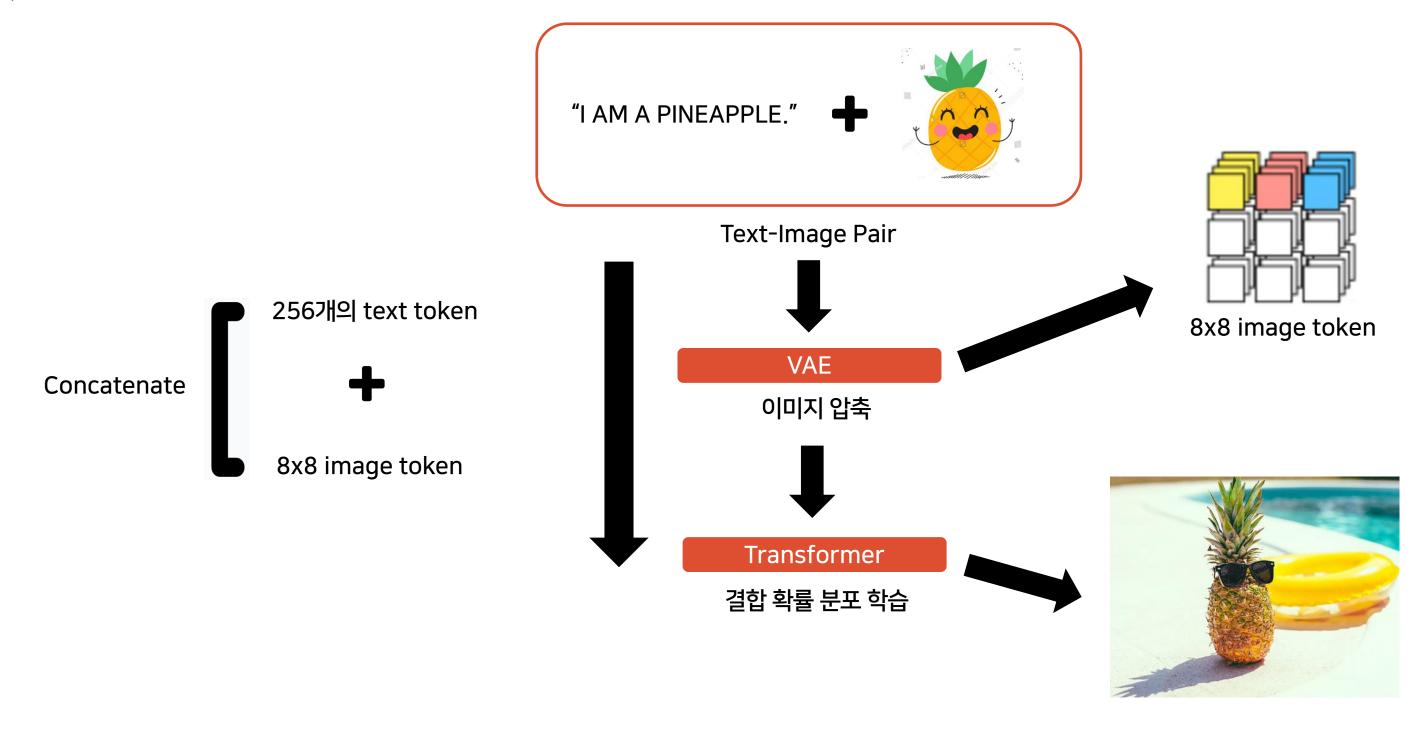
hidden_size = 768,

num_classes=5, # 클래스 수 조정

dr_rate=None,

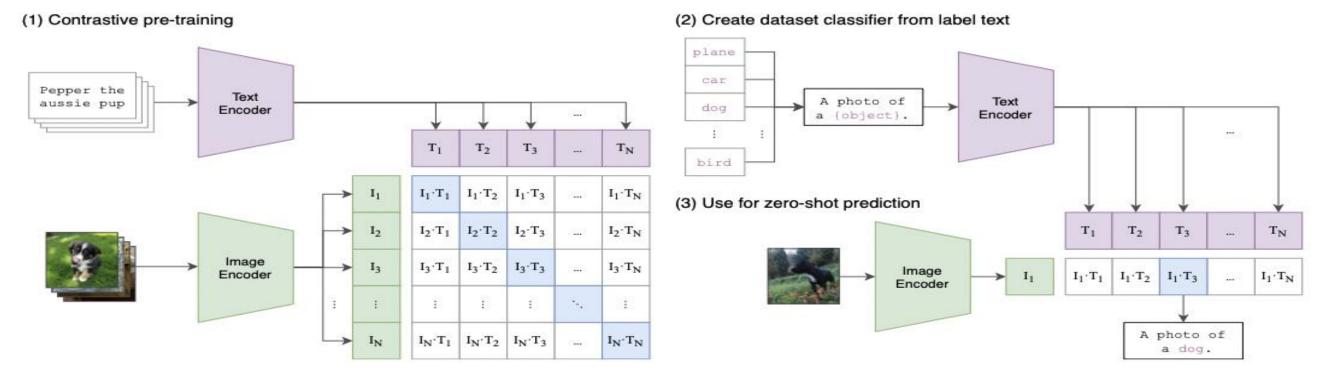
params=None):
```

✓ RQ-Transformer



✓ CLIP Score

• Text-Image 쌍의 상관 관계 평가Text(요약문)과 생성된 Image 간의 유사도를 평가할 수 있음.



[사진8] CLIP Score 산출 구

✓ Score Comparison

Song Info. 풍선 - 동방신기

Model	KoBART(Pretrained)	Song Dall-E	Natural Summarization
CLIP Score	0.2441	0.3174	0.2510

✓ Sentence Transformation

	접두어		
① Customize	A {} of		
② Sentiment	A mood of {}		

	문장 변형		
③ 문장 길이	1~3 Sentence		
④ 불용어 제거	반복 단어, 비속어		

√ Score Comparison

	1+3+4	2+3+4	1)+2+3+4
CLIP Score	0.2824	0.2613	0.3037

① a painting of

③, ④ in my childhood, I had a pretty dream of flying in a balloon

② A mood of joyful

③, ④ in my childhood, I had a pretty dream of flying in a balloon

① a painting of

② A mood of joyful

③, ④ in my childhood, I had a pretty dream of flying in a balloon

✓ KoBART 요약 모델 학습 전후 비교

- Song Info. 여행 볼빨간 사춘기
- 학습 전 요약 결과
 - ".. Let go. Please find me. Please find me. Please find me. It's s"
 - ".. 놔. 제발 저를 찾아주세요. 제발 저를 찾아주세요. 제발 저를 찾아주세요. s다"
- 학습 후 요약 결과

"I'm in love with this beautiful poem and I'm breathing"

"나는 이 아름다운 시에 사랑에 빠졌고 숨쉬고 있다"

Song Info. 풍선 - 동방신기

• 학습 전 이미지 결과



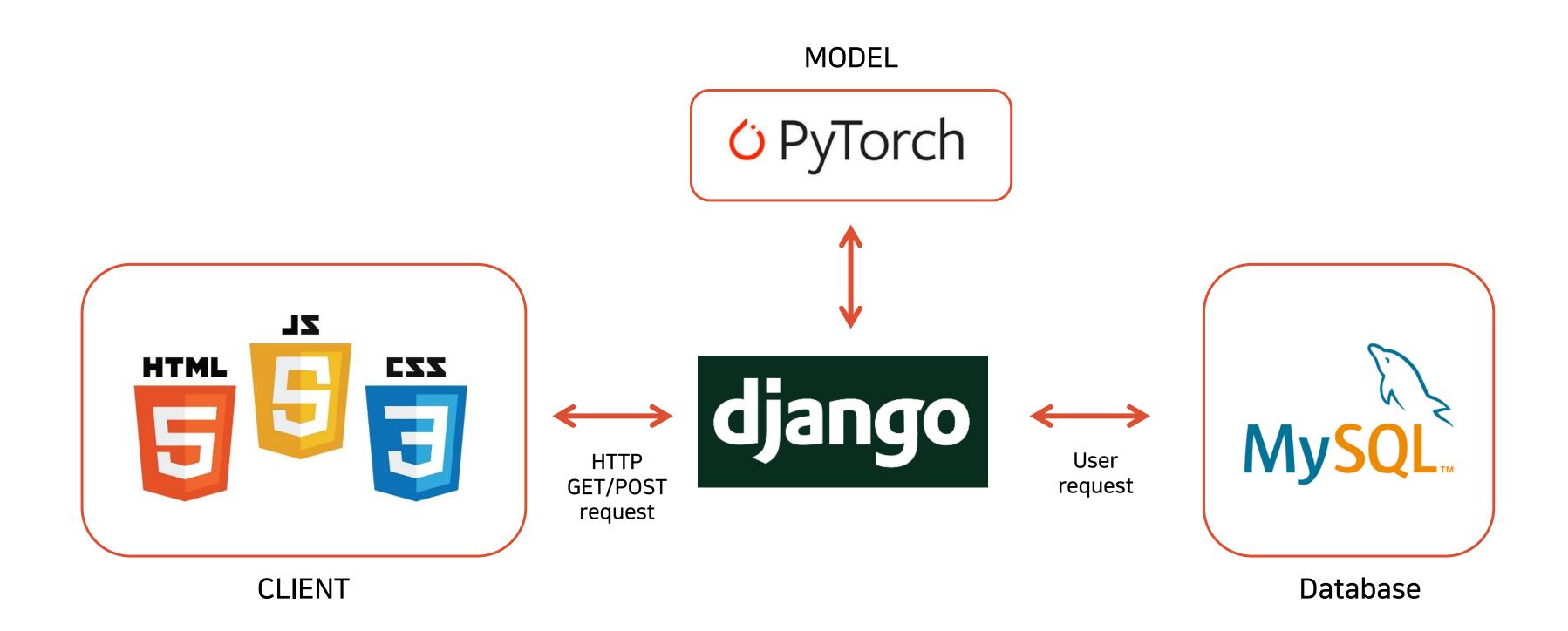


• 학습 후 이미지 결과





학습 이후, 요약과 이미지 학습 결과에서 확연히 차이 나는 것을 확인할 수 있었음.



```
song_dall_e
-- README.md
-- requirements.txt
 -- view.py
-- app.py
 -- view.py
 -- blog - post record
       +-- templates - html for frontend
      ├ -- styles - css
      -- js
       | <u>--</u> ...
       L—— ...
     single_pages - Modules required to loading
         -- templates - html for frontend
           L___ ...
         -- Summarization - model for text summarization(KoBART)
         ├-- rq-transformer - model for text-to-image (rq-transformer)
                     L—— CLIP (decoder for text-to-image)
                           L____
         ├-- sentiment - model for sentiment analysis (KoBERT)
                 L____
         -- preprocessing.py
         -- forms.py - formation for creation
         | -- forms2.py - formation for crawling lyrics data
         ├ -- papago.py - translation
         └-- postprocessing.py
```

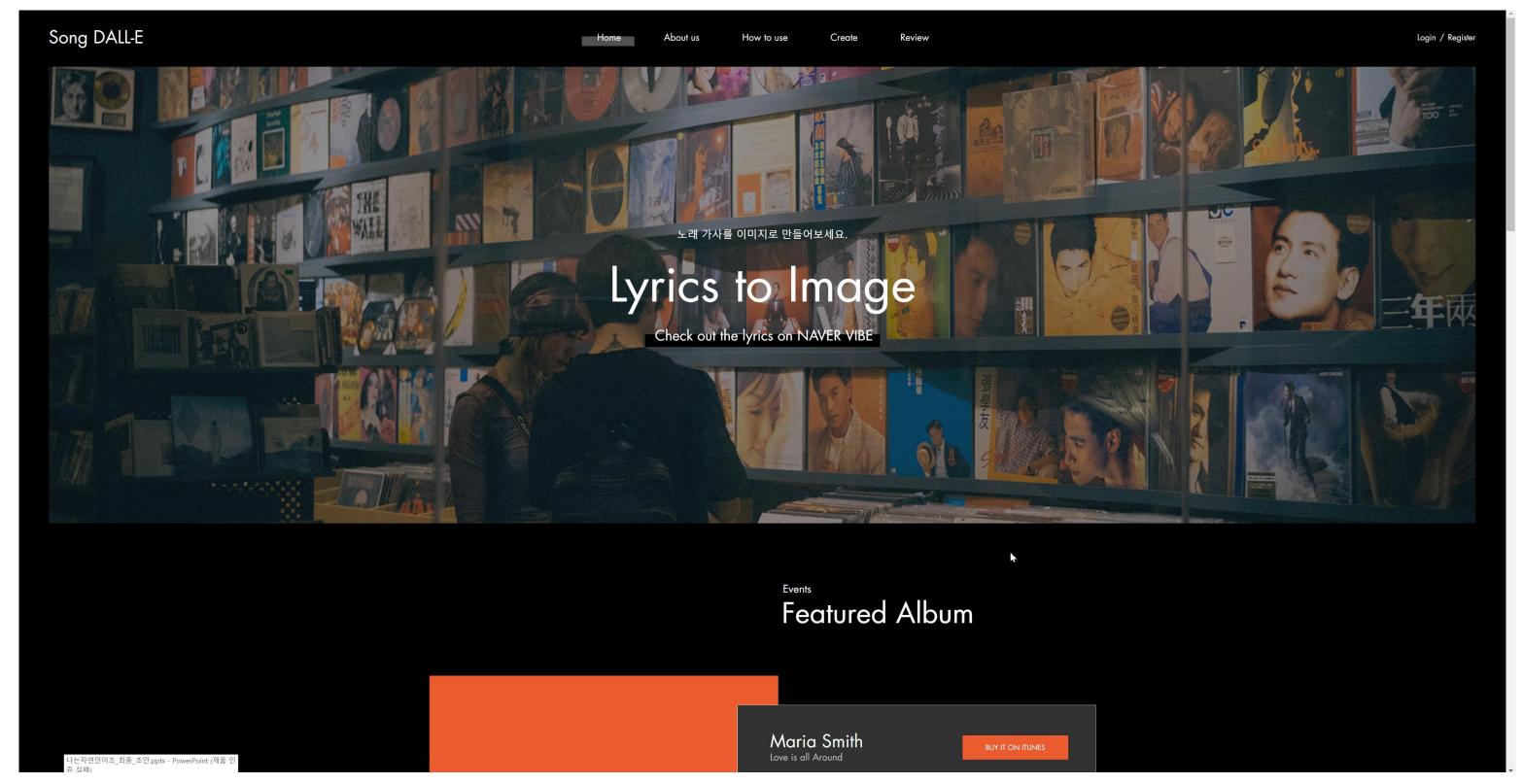
[사진9] Song DALL-E web **structure**

✓ 주요 기능

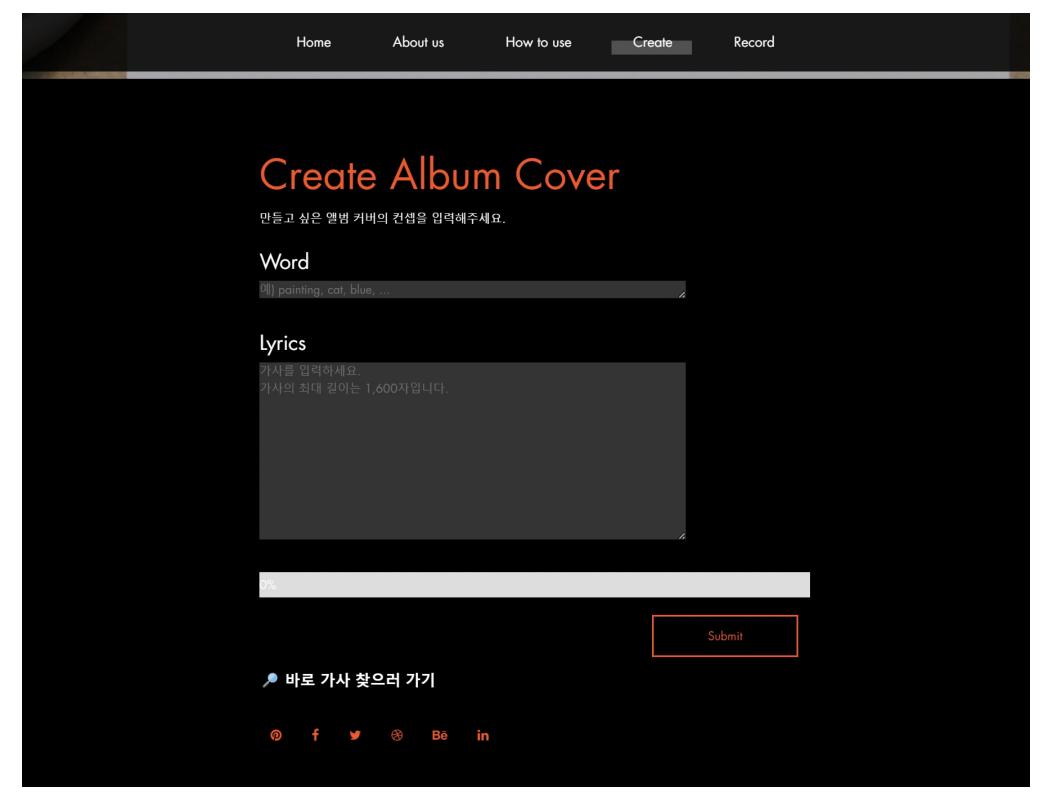
- 텍스트 전처리(preprocessing, postprocessing)
- 텍스트 요약(KoBART)
- 번역(papao API)
 - 정확한 이미지 생성을 위해 번역
- KoBERT를 이용한 감성 분석 결과 제시
- RQ-Transformer를 이용한 이미지 생성
 - 12개의 이미지 생성(약 30-40초 소요)
- 가사 검색(melon crawling using Beautifulsoup)
 - "제목" 혹은 "제목 가수"로 입력

id	predict_datetime	lyrics	summary	word	sentiment
111	2022-09-29 15:07:40.944072	하루에도 몇 번씩 널 , 보며 웃어 난, 수백번 말했잖아. You're the love of my life. 거짓된 세상 속 불안한 , 내 맘 속, 오직 나 믿는 건 너	a painting of i told you that you are the only on	painting	Joyful
112	2022-09-29 16:45:39.198785	1 2 3 4 5 6 7 8 9 10. 딱 10CM만, 더도 말고 멀도 말고 딱 그 정도만, 한번 해보자, 우리 멀어져 보자, 딱 반 발자국만, 더도 말고 멀도	a person of 1 2 3 4 5 6 7 8 9 10 , 1 2 3 3 4 , 4 ,	person	Sadness
113	2022-09-29 16:48:04.885429	흐르는 강물을, 거꾸로 거슬러, 오르는 연어들의, 도무지 알 수 없는, 그들만의 신비한 이유처럼, 그 언제서 부터인가, 걸어 걸어 걸어	a fish of this road you walk on , how much more	fish	Sadness
114	2022-09-29 16:49:44.110969	하나 둘 셋 넷. Left and right. Left and right. Left and right. Seventeen ha. 잊지 말아야 해 출발선에 설 때. 두 눈 부릅뜨고 고갤 들어	a picture of left and right , left and right , left a	picture	Calmness
115	2022-09-29 16:50:42.839504	믿을 수 있나요. 나의 꿈속에선. 너는 마법에 빠진 공주란걸. 언제나 너를 향한 몸짓엔, 수많은 어려움 뿐이지만. 그러나 언제나 굳은	a painting of when you cross the swamp past th	painting	Sadness
116	2022-09-29 16:51:49.860267	뭐하고 있었니 늦었지만, 잠시 나올래, 너의 집 골목에 있는, 놀이터에 앉아 있어, 친구들 만나서 오랜만에, 술을 좀 했는데, 자꾸만	a painting of i met my friends sitting in the playg	painting	Calmness
117	2022-09-29 16:52:59.367183	상처를 치료해풀 사람 어디 없나. 가만히 놔두다간 끊임 없이 덧나, 사랑도 사람도 너무나도 겁나. 혼자인 게 무서워. 난 잊혀 질까 두	a album art of it's not a jack sparrow who's got	album art	Sadness

[사진10] 데이터베이스 저장 결과



[동영상1] Song DALL-E 홈페이지 구조



[동영상2] Song DALL-E 시연

✓ 프로젝트 기대효과

- 누구나 간단하게 하나뿐인 창작물을 얻을 수 있다.
 - AI에 의해 창작된 그림이라는 것만 명시하면 자유롭게 사용 가능
- 창작자가 영감(가이드)을 제공받을 수 있다.
 - 예술의 영역을 기계가 침범하기보다는 상생하는 구조로 활용 가능
- Text-to-Image에 관한 모델들의 관심이 높아진다면, **더 좋은 성능의 모델**들이 나올 것이다.
- 데이터를 학습만 한다면, 여러 분야(시, 문학 작품 등)에서 활용 가능하다.
- 디자인 비용이 부담될 수 있는 스타트업, 1인 사업자 등 비즈니스에 도움을 줄 수 있다.
- 단순 이미지 생성에서 나아가서 교육용 콘텐츠 자료로 활용 가능

✓ 프로젝트 개발 후기 🎩



#CUDA out of memory #404 #transformer #Encoder #Decoder #이게왜안되죠 #런타임끊겼어요 #버전이안맞네요 #GPU다썼어요

- 자연어 처리는 모두가 처음이라, 데이터 학습과 모델링 과정에서 배운 것이 많았습니다.
- 먼 거리를 이동하며 참여해 체력적으로 많이 힘들었지만 이렇게 노력한 끝에 좋은 결과물이 나와서 기분이 좋습니다.
- 프로젝트 기간동안 힘들었지만 팀원분들과 함께해서 많이 배우고 버틸 수 있었습니다. 감사합니다.

✓ 프로젝트 한계점

- 가사의 주제가 명확하지 않다면 결과가 불만족스러울 수 있다.
- 서버 구축을 하지 못해 로컬에서만 구동이 가능했다.
- 감성분석 결과의 종류가 5가지로 다채롭지 못하다.
- 모델 학습시간이 길어 충분한 학습을 하지 못했다.

✓ OPENAI Policy (DALL-E)

Content policy

Updated July 20, 2022

Thank you for trying our generative AI tools!

In your usage, you must adhere to our Content Policy:

Do not attempt to create, upload, or share images that are not G-rated or that could cause harm.

- Hate: hateful symbols, negative stereotypes, comparing certain groups to animals/objects, or otherwise expressing or promoting hate based on identity.
- · Harassment: mocking, threatening, or bullying an individual.
- · Violence: violent acts and the suffering or humiliation of others.
- Self-harm: suicide, cutting, eating disorders, and other attempts at harming oneself.
- Sexual: nudity, sexual acts, sexual services, or content otherwise meant to arouse sexual excitement.
- Shocking: bodily fluids, obscene gestures, or other profane subjects that may shock or disgust.
- Illegal activity: drug use, theft, vandalism, and other illegal activities.
- Deception: major conspiracies or events related to major ongoing geopolitical events.
- Political: politicians, ballot-boxes, protests, or other content that may be used to influence the political process or to campaign.
- Public and personal health: the treatment, prevention, diagnosis, or transmission of diseases, or people experiencing health ailments.
- · Spam: unsolicited bulk content.

Disclose the role of AL

- You must clearly indicate that images are AI-generated or which portions of them are - by attributing to OpenAI when sharing, whether in public or private.
- You may post these images to social media. Please refer to our <u>Sharing and Publication Policy</u> for further details.

Respect the rights of others.

- . Do not upload images of people without their consent, including public figures.
- Do not upload images to which you do not hold appropriate usage rights.
- · Do not attempt to create images of public figures (including celebrities).
- To prevent deepfakes, we are currently prohibiting uploads of all realistic faces, even when the face belongs to you or if you have consent.

출처: https://labs.openai.com/policies/content-policy

1) 혐오, 증오 와 같은 부정적 영향을 끼칠 수 있는 이미지 생성 및 공유 시도 금지

- 이미지 생성 전 요약문 후처리 간 불용어(비속어 등) 처리

2) 이미지 출처가 AI에 의한 생성임을 밝힐 것

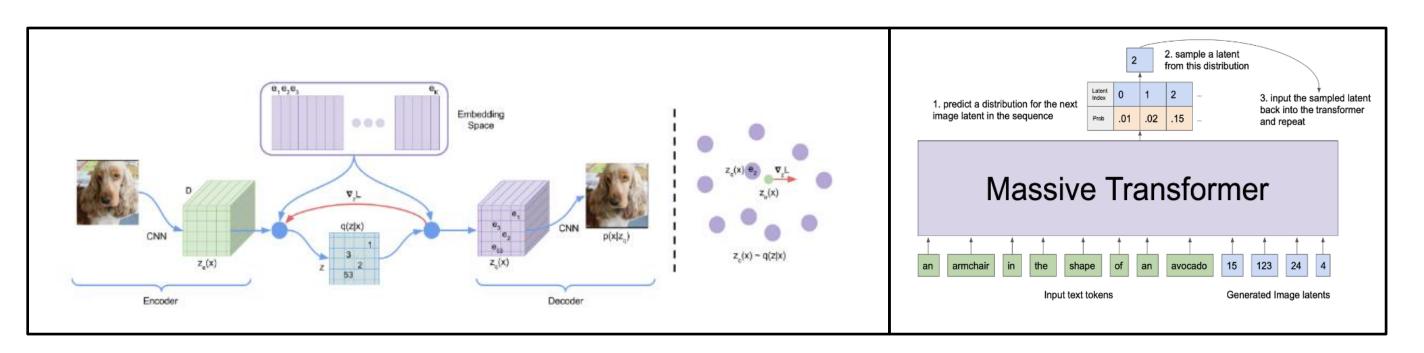
- 이미지 생성 전/후 저작권 명시



This is an image created by Al.

3) 초상권_유명인, 공인 포함 동의 없이 사람 얼굴의 이미지 업로드 및 생성 금지

✓ minDALL-E



✓ D(Half-moon) – 딘(DEAN) 가사 기반, Text-to-Image 모델 비교

minDALL-E



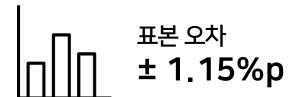
RQ-Transformer



- D(Half-moon) 딘(DENA)의 노래 뿐만 아니라 대체적으로 RQ-Transformer가 우수한 성능을 보여줌,
- 또한 이미지 생성 시간에 있어서도 minDALL-E보다 RQ-Transformer가 우수한 성능을 보여줌.

응답수 104건





1. 다음 이미지는 사람이 만든 이미지일까요? 인공지능이 만든 이미지일까요?

1-3. 위너 - 아일랜드

1-2. 딘(Dean) - D(Half Moon)*

1-2. 딘(Dean) - D(Half Moon)
104개 중 53개 맞춤

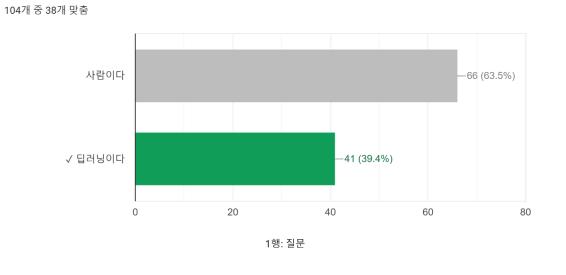
✓ 사람이다

□ 30 40 60

1행: 질문

정답: 실제 앨범 아트
정답률: 52.9%





정답: SongDalle로 만든 앨범 아트 정답률: 39.4%

2. 노래 가사와 이미지를 비교했을 때 어느 것이 실제 앨범 표지일까요?

위너 - 아일랜드

Baby 떠나자 둘이서 전화기는 비행기 모드 너와 함께면 어디든 아마 그곳은 무인도

Come come on ma, come on ma girl

자꾸 망설여진다면

Don't worry

#

네가 무서워 두려워 주저하는거 다 I'll make that go bang

B-b-bang bang

난 너만을 위한 수평선 위의 포물선

네가 내 옆에 있다면 거기가 보물섬

회색 빌딩 감옥 안에서 널 구해줄게

파란 하늘 모래 위에서 그냥 쉬어 가면 돼

떠나자 island 떠 떠 떠 떠나자 island

저 저기 저 파도와 함께 다가와 나에게 내 품에 올인

더 더 더 다가와 나에게 점 점 점 가까이 나에게

저 저기 저 태양과 함께 들어와 나에게 뜨겁게 골인

떠나자 island

우리의 island

Island

저 저기 저 야자수 아래 시원한 샴페인

I'll be your island

한여름의 뜨거운 태양보다 석양빛 너의 미소가 좋아

칵테일 shake it 하듯이 너와 소파 위에서 섞이고파

Shimmy shimmy ya

Shimmy shimmy ya hey

운전하고파 너의 매끈한 yeah

러브핸들 skrrr skrrr I'll make that go bang, b-b-bang bang

난 너만을 위한 수평선 위의 포물선

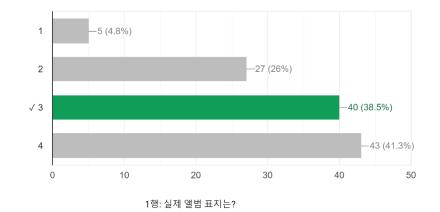
네가 내 옆에 있다면 거기가 보물섬

회색 빌딩 감옥 안에서 널 구해줄게

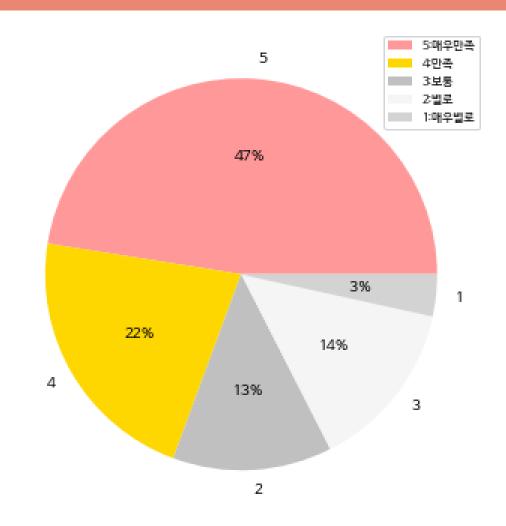
위너 - 아일랜드



위너 - 아일랜드 104개 중 36개 맞춤



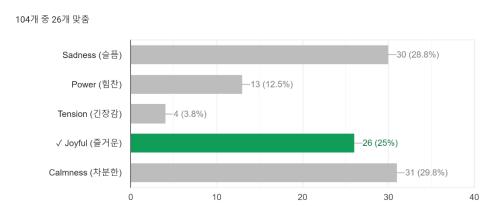
3. 다음 이미지는 해당 노래의 앨범 표지로 적당하다고 생각하시나요?



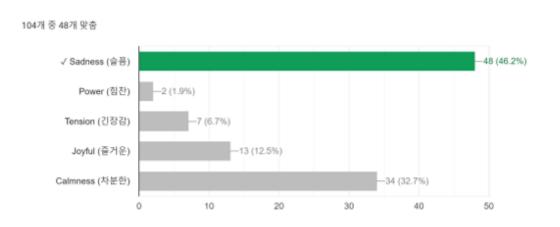
1. 만족도 평가 한 결과 '매우 만족'의 비율이 47%로 가장 높게 나옴. 2. 노래 가사와 이미지를 비교하여, 실제 앨범 표지를 맞추는 문항에서는 정답을 맞추지 않은 결과가 많았음.

4. 다음 이미지의 감정은 어떻다고 생각하시나요?



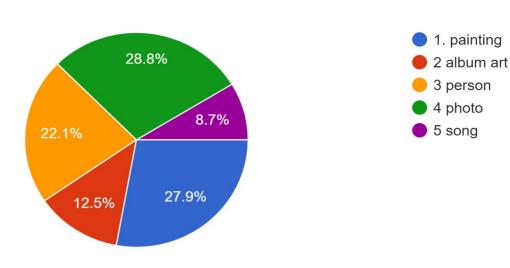






5. 어떤 Keyword가 마음에 드시나요?

'이문세 - 붉은 노을'을 여러 주제로 만든 이미지입니다. 어떤 주제가 마음에 드시나요? 응답 104개





#

5. 사용자 피드백

<긍정적>

- 좋은 사업 아이템인 것 같아요.
- 예술이 인공지능으로 대체될 수 있겠다는 생각을 했습니다.
- 허걱 모든게 인공지능이 만든 이미지였다고요? 믿을 수 없어요 정말 대단한 기술력이군요!
 - 아이디어가 참신하고 결과물이 재밌어요

<부정적>

- 사진에 대한 감정을 선택하는 문제에서 생각한 감정이 보기 중에 선택지에 없어서 아쉬웠습니다.
 - 더 다양한 이미지가 나왔으면 좋았을 텐데 아쉽습니다.

Thank You