스테이블 디퓨전

- **디퓨전 모델**
  - 디퓨전 모델은 물리학에서 물질의 확산 과정을 기술하는 데 사용되는 개념에서 착안하여 개발된 생성 모델
  - 이를 인공지능 및 기계 학습 분야에 도입함으로써, 데이터의 복잡한 구조와 패턴을 단계적으로 모델링하는 것이 가능
  - 물리학에서의 확산은 분자나 원자의 무작위 운동에 의해 발생하는 현상으로 이해
  - 물질의 확산은 대개 높은 농도에서 낮은 농도로 일어나며, 이러한 이동은 분자나 원자 간의 에너지 불균형 때문에 발생

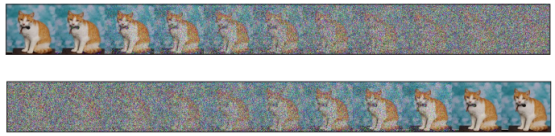
▼ **그림 1 잉크가 확산되고 있는 물컵**

1

3. 스테이블 디퓨전

- **디퓨전 모델**

  디퓨전 모델의 학습
  - 디퓨전 모델은 이미지 데이터에서 확산의 과정을 단계적으로 모방하면서, 각 단계에서의 노이즈를 통해 원본 데이터의 분포를 점진적으로 변화시킴
  - 점진적으로 변화시키는 과정의 반대 과정을 해낼 수 있도록 모델을 학습시키는 원리로 이미지를 생성
  - 이러한 각 과정은 정방향 변환(forward process)과 역방향 변환(reverse process)로 나눠서 생각해볼 수 있음



2

3. 스테이블 디퓨전

- **디퓨전 모델**

  정방향 변환
  - 정방향 변환은 이미지에 표준 정규 분포를 갖는 랜덤 노이즈를 순차적으로 더해가는 단계를 의미
  - 이는 앞서 이야기한 일종의 확산 과정으로써 노이즈를 더해감에 따라 점차 이미지의 형태를 알아볼 수 없게 됨
  - 여기서 노이즈를 더해가는 각 단계의 이미지를 인공지능 모델 학습에 사용
  - 노이즈를 더하는 과정을 총 T회 반복한다고 할 때, 다음과 같이 나타냄

$$x_0 \rightarrow x_1 \rightarrow \cdots \rightarrow x_{T-1} \rightarrow x_T$$

▼ **그림 2 고양이 이미지에 랜덤한 노이즈를 추가하는 정방향 변환**

3

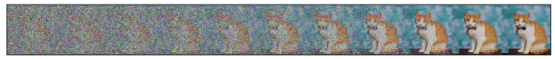
3. 스테이블 디퓨전

- **디퓨전 모델**

  역방향 변환
  - 이미지를 생성하는 주 과정은 해당 역방향 변환으로 이뤄지게 됨
  - 역방향 변환은 노이즈 제거(denoising) 과정이라고도 부르며, 정방향 변환의 반대 과정을 나타냄
  - 즉, 정방향 변환이 이미지의 본래 정보를 잃게 하는 확산 과정이었다면, 역방향 변환은 점차적으로 확산된 정보를 복구하는 역 확산 과정
  - 이때 목표는 각 단계에서 손실된 정보를 원래대로 복원하는 것

$$x_T \rightarrow x_{T-1} \rightarrow \cdots \rightarrow x_1 \rightarrow x_0$$

▼ **그림 6-20 손상된 이미지를 다시 고양이 이미지로 복원하는 역방향 변환**

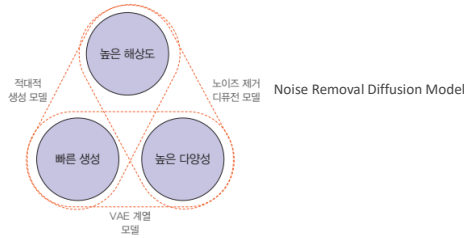


4

3. 스테이블 디퓨전

- **디퓨전 모델**
  - 디퓨전 모델의 핵심 아이디어는 단순한 초기 분포(ex 랜덤 가우시안 분포) 이미지를 입력으로 하여, 여러 단계의 역방향 변환 과정을 거쳐 원본 데이터의 복잡한 분포의 이미지로 만드는 것
  - 이러한 점진적인 접근 방식은 모델이 데이터의 내재된 구조를 더 효과적으로 생성하도록 도와줌

▼ **그림 6-21 이미지 생성 모델의 종류별 특성**



5

3. 스테이블 디퓨전

- **디퓨전 모델**

  텍스트와 이미지의 연결
  - 자연어와 이미지를 같이 처리할 수 있는 모델들이 제안되기 시작
  - 이는 다양한 종류의 데이터를 다룬다고 하여 '멀티 모달(multi modal)'이라 지칭

  - Mode??

6

3. 스테이블 디퓨전

- **디퓨전 모델**

  CLIP
  - 다양한 분야에서 멀티 모달 기법이 만들어지던 중에 OpenAI에서는 CLIP(Contrastive Language-Image Pre-Training)이라는 모델을 제안
  - CLIP의 특징은 다음과 같음
    - **언어와 이미지 연결**: CLIP은 이미지와 텍스트를 같이 학습하여 이미지와 텍스트의 관계를 이해
    - **대량의 데이터 학습**: 인터넷에서 수집한 대량(약 4억 개)의 이미지와 텍스트 데이터를 사용하여 훈련
    - **제로샷 학습**: CLIP은 언어와 이미지의 연관성을 파악하여 처음 보는 종류의 이미지에 대해서도 해당 이미지가 무엇인지 구분할 수 있는 능력을 갖추고 있음
    - **다양한 시각화 작업에 적용**: 이미지 분류, 객체 감지, 텍스트에서의 이미지 생성 등 다양한 시각화 작업에 CLIP 모델을 활용할 수 있음

  텍스트와 이미지를 함께 학습하여, 텍스트 설명과
  이미지 간의 관계를 이해할 수 있도록 설계된 모델

7

3. 스테이블 디퓨전

- **디퓨전 모델**

  CLIP의 학습 구조
  - CLIP 모델 학습의 핵심 원리는 그 이름(Contrastive Language-Image Pre-Training)에 맞게 '대조적 학습(Contrastive Learning)'에 기반함

▼ **그림 6-22 이미지와 이미지를 설명하는 레이블**

The man at bat readies to swing at the pitch while the umpire looks on.

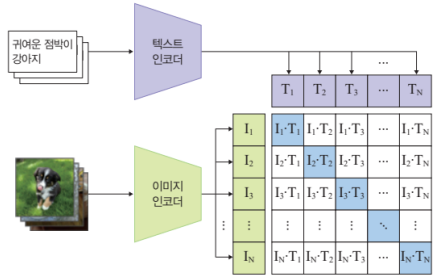
A large bus sitting next to a very tall building.

8

3. 스테이블 디퓨전

- **디퓨전 모델**
  - 학습 과정에서는 각각의 텍스트 인코더와 이미지 인코더가 자연어와 이미지의 임베딩 벡터를 연산해내는데, 여기서 긍정 레이블의 거리를 줄이고, 부정 레이블의 거리를 늘리는 방식으로 이미지와 텍스트의 관계를 학습

▼ **그림 6-23 CLIP 모델의 학습 방식**

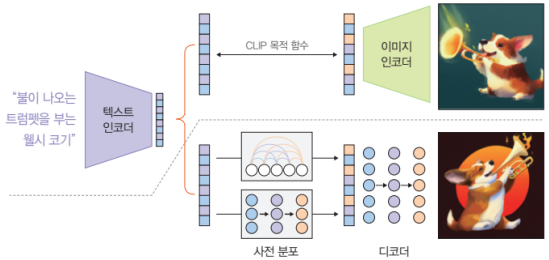


9

3. 스테이블 디퓨전

- **디퓨전 모델**
  - 이렇게 학습된 CLIP은 주어진 텍스트 설명에 가장 잘 맞는 이미지를 선택하거나, 이미지에 대한 텍스트 설명을 생성하는 데 유용하게 사용될 수 있음
  - OpenAI는 CLIP 모델을 사용하여, 텍스트를 입력으로 받아 원하는 이미지를 만들어내는 디퓨전 모델, DALL-E 2를 만들어냄

▼ **그림 6-24 DALL-E 2의 모델 구조**



10

3. 스테이블 디퓨전

- **디퓨전 모델**
  - 이미지 생성 모델은 계속해서 발전하고 있음
  - 현재 OpenAI에서는 사람이 원하는 다양한 이미지를 손쉽게 그려주는 서비스 DALL-E 3를 상용화하고 있음

▼ **그림 6-25 DALL-E 3 모델이 그려준 DALL-E 일러스트**

11

3. 스테이블 디퓨전

- **디퓨전 모델**

크로스 어텐션을 통한 멀티 모달 처리

- 멀티 모달 학습은 다양한 형태(Modality, 텍스트, 이미지, 음성 등)의 데이터를 동시에 처리하여 정보를 결합하고, 각 형태 간의 상관 관계를 학습
- 이때 각 데이터 형태의 메모리 용량이 상당히 달라지는 경우가 생김
- 크로스 어텐션(cross attention) 메커니즘은 이러한 멀티모달 학습에서 특히 중요한 역할
- 크로스 어텐션 멀티 모달 학습에서는 한 형태의 데이터(예 텍스트)를 쿼리(query)로 사용하고, 다른 모달의 데이터(예 이미지)를 키(key)와 값(value)으로 사용함으로써, 두 모달 간의 상호 관계와 상호 작용을 학습
  - **비디오 설명 생성**: 비디오 클립과 그에 해당하는 설명을 동시에 분석하여, 비디오의 내용을 자동으로 설명하는 텍스트를 생성
  - **이미지 설명**: 이미지를 입력으로 받아, 그 이미지의 내용을 설명하는 텍스트 캡션을 자동으로 생성

크로스 어텐션은 이미지의 각 부분과 관련된 텍스트 정보를 연결하는 데 중요한 역할

12

3. 스테이블 디퓨전

- **스테이블 디퓨전**
  - 스테이블 디퓨전은 뮌헨대학교 연구실, CompVis에서 고안한 디퓨전 기반의 텍스트-이미지 생성 인공지능 모델
  - 텍스트-이미지 생성 인공지능이 많은 관심을 받는 가운데, OpenAI의 DALL-E나 구글의 Imagen 같은 프로젝트는 많은 사람의 각광을 받았음
  - 이러한 모델들은 디퓨전 기반의 학습으로 매우 높은 컴퓨터 성능을 필요로 하여, 일반인이 접근하여 연구하기 어려운 면이 존재했음
  - 그들의 주요 목표는 기존의 방대한 컴퓨팅 자원을 필요로 하는 디퓨전 기반 모델의 한계를 극복하고, 더 많은 연구자와 개발자가 이러한 기술에 접근하고 활용할 수 있도록 하는 것
  - 스테이블 디퓨전 모델의 개발 과정에서 연구팀은 여러 최적화 기법과 알고리즘을 도입하여 모델의 학습 효율성을 크게 향상시켰음
  - 연구의 결과로 모델은 더 낮은 컴퓨팅 자원에서도 높은 성능의 이미지를 생성하는 능력을 보여줌

3. 스테이블 디퓨전

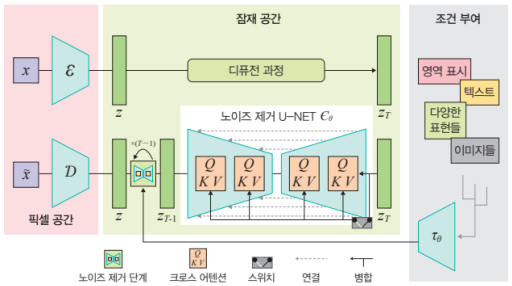
- **스테이블 디퓨전**

  잠재 공간의 활용
  - 스테이블 디퓨전 프로젝트의 중요한 측면 중 하나는 잠재 공간의 활용
  - 잠재 공간의 활용은 디퓨전 모델의 메모리 사이즈를 대폭 감소시켜주고, 효과적으로 학습할 수 있도록 도와줌
  - 스테이블 디퓨전이 등장할 당시의 디퓨전 모델들은 사용하려고 하는 이미지의 사이즈를 그대로 사용하여 노이즈 제거 과정을 거치기 때문에, 필요한 GPU의 용량이 매우 크다는 특징

3. 스테이블 디퓨전

- **스테이블 디퓨전**
  - 다음 이미지는 스테이블 디퓨전 모델의 모델 구조를 표현해주고 있음
  - 가장 좌측의 붉은 블록에서 이미지는 픽셀 공간에서 잠재 공간으로 변형하고, 다시 잠재 공간의 이미지를 픽셀 공간으로 변형

▼ **그림 6-26 스테이블 디퓨전의 학습 과정**



따라하기 실습 (동영상 Link)

- https://www.youtube.com/watch?v=_KAVdDrOGH0&list=PLKuQxQX8EZn3WK9uZQpdy8cdcF0IwHiGA

3. 스테이블 디퓨전

- **스테이블 디퓨전**

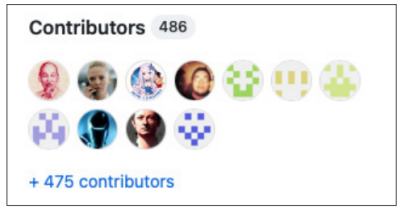
  스테이블 디퓨전과 WEB UI

  - 이렇게 만들어진 스테이블 디퓨전은 엄청난 파급력을 가져왔고 많은 사람이 사용하고 있음
  - 이러한 배경에는 해당 인공지능 모델의 코드가 무료로 공개되고, 라이선스로 'OpenRAIL-M'을 채택한 것이 크게 기여
  - 'OpenRAIL-M'은 '수정된 Open Responsible AI 라이선스'라는 의미
  - 연구, 상업적 또는 비상업적 목적으로 인공지능 모델의 파생물이 자유롭고 개방적인 접근, 재사용 및 용도에 맞는 변형 배포를 허용하도록 설계된 라이선스
  - 이에 대한 코드는 해당 모델을 발표한 저자들의 깃허브(https://github.com/CompVis/stable-diffusion)에서 자유롭게 확인 가능

17

3. 스테이블 디퓨전

▼ **그림 6-27 WEB UI 오픈 소스를 업데이트해주는 400명 이상의 기여자**

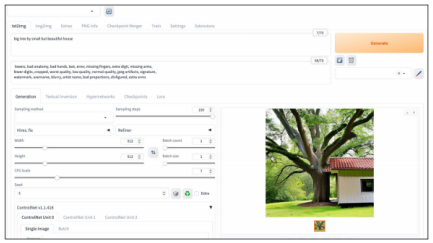


18

3. 스테이블 디퓨전

- **스테이블 디퓨전**
  - WEB UI란 사실 단순히 스테이블 디퓨전 등 이미지 생성 모델뿐만 아니라, 인터넷 웹 브라우저를 통해서 보이는 사용자 인터페이스를 지칭하는 용어
  - AUTOMATIC1111 저자는 해당 기술을 개발자들을 넘어서 일반인들에게도 확장하고 사용해볼 수 있게 개방시켰음

▼ **그림 6-28 스테이블 디퓨전 WEB UI 페이지, 큰 나무 옆에 작고 예쁜 집을 프롬프트로 이미지를 생성한 예시**



19

3. 스테이블 디퓨전

- **스테이블 디퓨전**

  WEB UI의 설치 방법

  - 해당 툴은 웹 브라우저 화면에서 사용할 수 있게 되어 있지만, 사실은 설치하여 사용하는 형태
  - 웹 사이트로 배포되는 무료 기능은 해당 웹 사이트 자체가 사라지면 더 이상 사용할 수 없는 반면에, WEB UI는 직접 설치해서 사용하기 때문에 웹 사이트가 사라져도 계속 무료로 해당 툴을 사용할 수 있음
  - 여기서 설치 방법은 환경에 따라 조금씩 다른데 크게는 운영 체제와 외장 GPU의 여부에 따라 달라짐
  - 자신의 PC에 설치하는 것도 좋지만, 우리가 앞서 사용한 코랩에서도 해당 기능을 써볼 수 있게 사람들이 다수 배포를 해놓았고, 통일된 환경인 만큼 사용도 단순함(현재 코랩 무료 버전에서는 해당 기능을 사용할 수 없고, 코랩 유료 사용자만 사용 가능)

20

3. 스테이블 디퓨전

- **스테이블 디퓨전**

  코랩에서의 WEB UI 사용법

  - 다음 사이트에서 접근하면 평소와 조금 달라보이는 코랩에 접근해볼 수 있음

▼ **그림 6-29 https://colab.research.google.com/github/Cobslab/stable-diffusion-webui/blob/master/fast_stable_diffusion_WEB_UI_maker.ipynb 페이지**

3. 스테이블 디퓨전

- **스테이블 디퓨전**
  - 코랩 무료 사용자들도 접속하면 동일한 페이지를 볼 수 있으나, 코드를 실행하면 잠시 후에 런타임이 끊기며 해당 내용을 진행할 수 없다는 안내 메시지가 나옴
  - 대신 가정에서 사용하는 게이밍 노트북이나 외장 그래픽 카드가 있는 컴퓨터가 있다면, 컴퓨터에 설치해서 진행
  - 그래픽 카드가 없을 경우에도 컴퓨터에 직접 설치하여 WEB UI를 사용해볼 수 있으나, 이미지를 생성하는 데 걸리는 시간이 30배 정도 더 오래 걸림
  - 대화형 양식은 사용자가 코드를 직접 입력하지 않아도 텍스트 박스, 스크롤 바 등으로 파이썬 코드의 입력 값을 조절할 수 있도록 도와줌

```
Shared_Drive = "" #@param {type:"string"}
```

▼ **그림 6-30 #@param 주석으로 만들어진 UI**

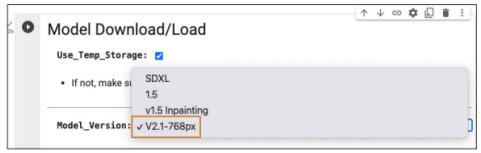
3. 스테이블 디퓨전

- **스테이블 디퓨전**
  - WEB UI는 다음 과정을 거쳐 동작
    1. **구글 드라이브에 접속**(Connect Google Drive)
       자신의 구글 드라이브와 연결
       이때 그림 6-30의 공유 드라이브(Shared_Drive)에 원하는 경로를 넣으면, 해당 경로로 코랩에 구글 드라이브를 마운트
       기입하지 않아도 괜찮으니, 그대로 실행시켜서 진행
    2. **Cobslab 저장소를 설치**(Install/Update Cobslab repo)
       블록을 실행시켜, WEB UI의 코드가 있는 깃허브 저장소를 불러옴
    3. **종속성 라이브러리를 설치**(Requirements)
       블록을 실행시켜, 종속성 패키지를 설치

3. 스테이블 디퓨전

- **스테이블 디퓨전**
  4. **모델을 다운로드**(Model Download/Load)
     원하는 모델을 다운로드
     블록을 실행하기 전, 다음 이미지와 같이 드롭 박스를 내려 모델 v2.1을 선택

▼ **그림 6-31 코랩에서 사용하는 WEB UI: 모델 선택**



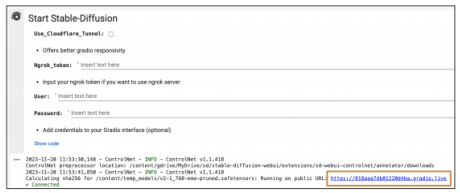
3. 스테이블 디퓨전

- **스테이블 디퓨전**

  5. **모델을 실행(Start Stable-Diffusion)**

  다음으로는 [Download LoRA]와 [ControlNet]이라고 되어 있는 부분을 건너뛰고,
  바로 아래 Start Stable-Diffusion 코드 블럭을 바로 실행
  이제 조금 기다리면 화면 우측 하단에 다음과 같은 링크가 생성

▼ **그림 6-32 Start Stable-Diffusion 블럭에 생성된 URL 링크**

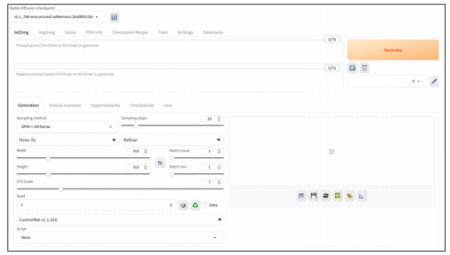


25

3. 스테이블 디퓨전

- **스테이블 디퓨전**
  - 해당 파란색 링크를 클릭하면 다음과 같은 WEB UI 화면을 볼 수 있음

▼ **그림 6-33 코랩에 설치된 WEB UI 화면**

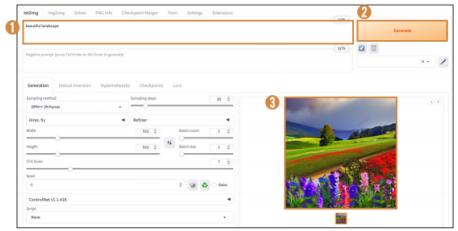


26

3. 스테이블 디퓨전

- **스테이블 디퓨전**
  - 이후 그림 6-34와 같이❶번 텍스트 상자에 원하는 내용을 써넣고,❷번 **제출**(Generate) 버튼을 누르면❸번 창에 생성 결과 이미지가 출력

▼ **그림 6-34 WEB UI로 생성된 이미지**



27

3. 스테이블 디퓨전

- **스테이블 디퓨전**

  WEB UI 구성
  - WEB UI를 성공적으로 구동했다면, 화면에 표시된 여러 낯선 개념들을 볼 수 있음
  - 무엇보다 먼저 이 요소들이 모델을 통한 이미지 생성 과정에서 어떻게 사용되는지 간략하게 소개
    - **Stable Diffusion Checkpoint**: 이미지를 생성할 모델
      서로 다른 모델을 사용할 경우 이미지의 화풍이 많이 달라짐
      예술 작품을 입력받아 학습된 모델은 그림을 예술 작품을 그려내듯 만들어 냄
      반면 현실의 데이터를 집중적으로 학습한 모델은 사실적인 묘사에 더욱 그 기능이 특화되어 있음
      스테이블 디퓨전이 갖는 강점 중 하나는 서로 다른 모델들을 병합하여 사용할 수 있다는 점
      여러 모델을 전략적으로 선별하여 섞을 경우 각 모델에서 취사 선택할 요소들을 적절히 배합하여 사용자가 희망하는 그림을 생성할 수 있음

28

3. 스테이블 디퓨전

- **스테이블 디퓨전**
  - **Txt2img**: 텍스트를 입력받아 이미지를 생성하는 태스크
    사용자가 그림에 대한 묘사를 언어로 제공하여 모델이 이를 바탕으로 그림을 그림
  - **Img2img**: 이미지를 입력받아 이미지를 생성하는 태스크
    이미지의 해상도를 높이고(super resolution), 마스크로 가려진 부분을 생성하여 복원하거나(inpaint) 혹은 이미지다른 스타일로 변환하는(image translation) 작업들이 이에 해당
  - **Prompt**: 사용자가 생성하고 싶은 그림을 텍스트 형태로 입력하는 부분
    모델은 텍스트의 지시사항에 맞춰 이미지를 생성
    반면 부정 프롬프트(negative prompt)는 그림에서 제외하고 싶은 특징이나 요소들이 모인 텍스트
    지시문을 입력하고, 우측에 있는 Generate 버튼을 클릭하면 그 아래 이미지들이 생성

29

3. 스테이블 디퓨전

- **스테이블 디퓨전**
  - **Sampling method**: 디퓨전의 역방향 변환에서 매 단계마다 노이즈를 제거하는 방식을 설정
    이 방식이 달라질 경우 이미지가 생성되는 속도, 품질, 스타일 등이 다소 변할 수 있음
  - **Sampling steps**: 역방향 변환을 통하여 초기 노이즈에서 이미지를 복원하는 과정을 몇 단계에 나누어서 수행할지 조절할 수 있음
    이 값이 높아질 수록 이미지를 생성하는 과정에서 단계 수가 늘어나므로 이미지 생성 시간 또한 증가
    다만, 단계 수가 이미지의 품질에 정비례하지는 않음
  - **Width & Height**: 생성될 이미지의 사이즈를 설정할 수 있음
  - **Batch count**: 모델에서 추론을 몇 번 반복할지 횟수를 정함

30

3. 스테이블 디퓨전

- **스테이블 디퓨전**
  - **Batch size**: 모델에서 추론을 한 번 진행할 때 생성할 이미지의 수를 의미
    이 값이 커질 경우 사용할 하드웨어 디바이스의 메모리에 부하를 줄 수 있음
  - **CFG scale**: Classifier-Free Guidance 비율의 약자로, 모델이 이미지를 생성할 때 사용자가 입력한 프롬프트를 얼마나 반영할지 강도를 조절하는 인수
    이 값이 커질수록 사용자의 명령에는 부합하지만 부자연스러운 이미지가 생성될 가능성이 높으며, 반대로 값이 작을수록 사용자의 의도를 충실하게 수행하지 않되 자연스러운 이미지가 생성될 가능성이 높음

31

3. 스테이블 디퓨전

- **스테이블 디퓨전**

  프롬프트 엔지니어링
  - 스테이블 디퓨전을 비롯한 이미지 및 텍스트 생성 모델에서의 프롬프트는 입력 데이터를 의미
  - 이 입력은 모델에게 이미지를 생성하도록 지시하는 역할
  - 이미지 생성 모델에서의 프롬프트는 일반적으로 이미지의 내용, 스타일, 분위기, 색상 등을 서술하는 텍스트로 구성
  - 모델은 이 프롬프트를 분석하여 사용자의 요구 사항에 맞는 이미지를 생성

32

3. 스테이블 디퓨전

- **스테이블 디퓨전**
  - 스테이블 디퓨전 모델에서 출력 이미지에 가장 직접적으로 관여하는 유일한 입력 데이터가 프롬프트인 셈이며, 이 텍스트 데이터의 주요한 특징을 세 가지로 꼽을 수 있음
    1. **서술성**: 프롬프트는 모델에 생성하고자 하는 이미지의 세부 사항을 전달
       예를 들어 '휴양지에서 석양을 바라보는 고양이'라는 지시문을 제공할 경우, 모델은 문장에서 열대 지방의 해안가, 붉게 물든 노을 등의 시각적인 묘사를 이해하고 표현할 수 있음
    2. **유도성**: 프롬프트는 모델이 특정 방향이나 스타일로 이미지를 생성하도록 유도할 수 있음
       프롬프트에 '인상파 스타일의' 또는 '중세 유럽 느낌의' 등 화풍이나 시대상 등 정보를 삽입할 경우 특정 예술적 스타일이나 분위기를 지정할 수 있음
    3. **창의적 조합**: 프롬프트를 통해 기존에 없던 창의적인 아이디어나 조합을 실험할 수 있음
       '고대 로마 시대의 로봇 기사'나 '목성을 배경으로 한 신화적 생물'과 같은 독특한 조합을 시도할 수 있음

33

3. 스테이블 디퓨전

- **스테이블 디퓨전**
  - 이러한 프롬프트의 특징을 잘 활용하여, 모델이 지시사항을 잘 이해하게끔 프롬프트를 가공하는 기술을 프롬프트 엔지니어링(prompt engineering)이라 함
  - 잘 정제된 프롬프트는 AI의 성능을 극대화하고, 원하는 결과를 더 정확하고 빠르게 얻을 수 있도록 도움을 줌
  - 다음은 프롬프트 엔지니어링 시 고려해야 할 요소들임
    1. 프롬프트가 최대한 구체적이며, 원하는 결과에 대한 명확한 지침을 포함할수록 사용자 의도에 부합한 결과물이 생성
       '자연에서 맞는 평화로운 아침'보다 '샘이 흐르는, 푸른 나무가 우거진 숲속 산 너머 떠오르는 아침 해. 구름 없는 하늘. 떼 지어 나는 새'가 훨씬 시각적으로 직접적인 묘사를 할 수 있기에 좋은 결과를 만들 수 있음

34

3. 스테이블 디퓨전

- **스테이블 디퓨전**
    2. 이미지에 등장할 객체뿐만 아니라 세부적인 묘사에 대하여 방향성을 제시할 수 있음
       이미지가 사진처럼 사실적으로 묘사되거나 그림처럼 묘사되게 지시할 수 있으며 화풍도 프롬프트에 포함될 수 있음
       또한 조명의 색상과 방향, 시점, 색상 등의 분위기를 표현할 수 있는 다양한 지시도 활용될 수 있음
    3. 스테이블 디퓨전의 언어 모델은 한국어보다 영어로 된 문장을 더 잘 이해하므로, 프롬프트를 영어로 제시하는 것이 훨씬 좋은 결과를 얻을 수 있음

35

3. 스테이블 디퓨전

- **스테이블 디퓨전**
    4. 부정 프롬프트를 적극 활용하여 제외하고 싶은 결과를 요소들을 제시하면 이미지 생성에 도움이 됨
       프롬프트에 부정적인 문장을 포함시키지 않아 길이를 줄일 수 있으며, 모델이 프롬프트를 과잉 해석하는 것을 막을 수 있음
       또한 스테이블 디퓨전 특성상 기형적인 이미지가 가끔 생성되는데, '저품질', '부자연스러운 신체 구조', '워터마크' 등 이미지 생성 시 당연하게 빠져야 할 지시들이 부정 프롬프트에 포함될 경우 불필요한 생성 절차를 간소화할 수 있음

36

3. 스테이블 디퓨전

- **스테이블 디퓨전**
  - 스테이블 디퓨전 모델에서 출력 이미지에 가장 직접적으로 관여하는 유일한 입력 데이터가 프롬프트인 셈이며, 이 텍스트 데이터의 주요한 특징을 세 가지로 꼽을 수 있음
    1. **서술성**: 프롬프트는 모델에 생성하고자 하는 이미지의 세부 사항을 전달
       예를 들어 '휴양지에서 석양을 바라보는 고양이'라는 지시문을 제공할 경우, 모델은 문장에서 열대 지방의 해안가, 붉게 물든 노을 등의 시각적인 묘사를 이해하고 표현할 수 있음
    2. **유도성**: 프롬프트는 모델이 특정 방향이나 스타일로 이미지를 생성하도록 유도할 수 있음
       프롬프트에 '인상파 스타일의' 또는 '중세 유럽 느낌의' 등 화풍이나 시대상 등 정보를 삽입할 경우 특정 예술적 스타일이나 분위기를 지정할 수 있음
    3. **창의적 조합**: 프롬프트를 통해 기존에 없던 창의적인 아이디어나 조합을 실험할 수 있음
       '고대 로마 시대의 로봇 기사'나 '목성을 배경으로 한 신화적 생물'과 같은 독특한 조합을 시도할 수 있음