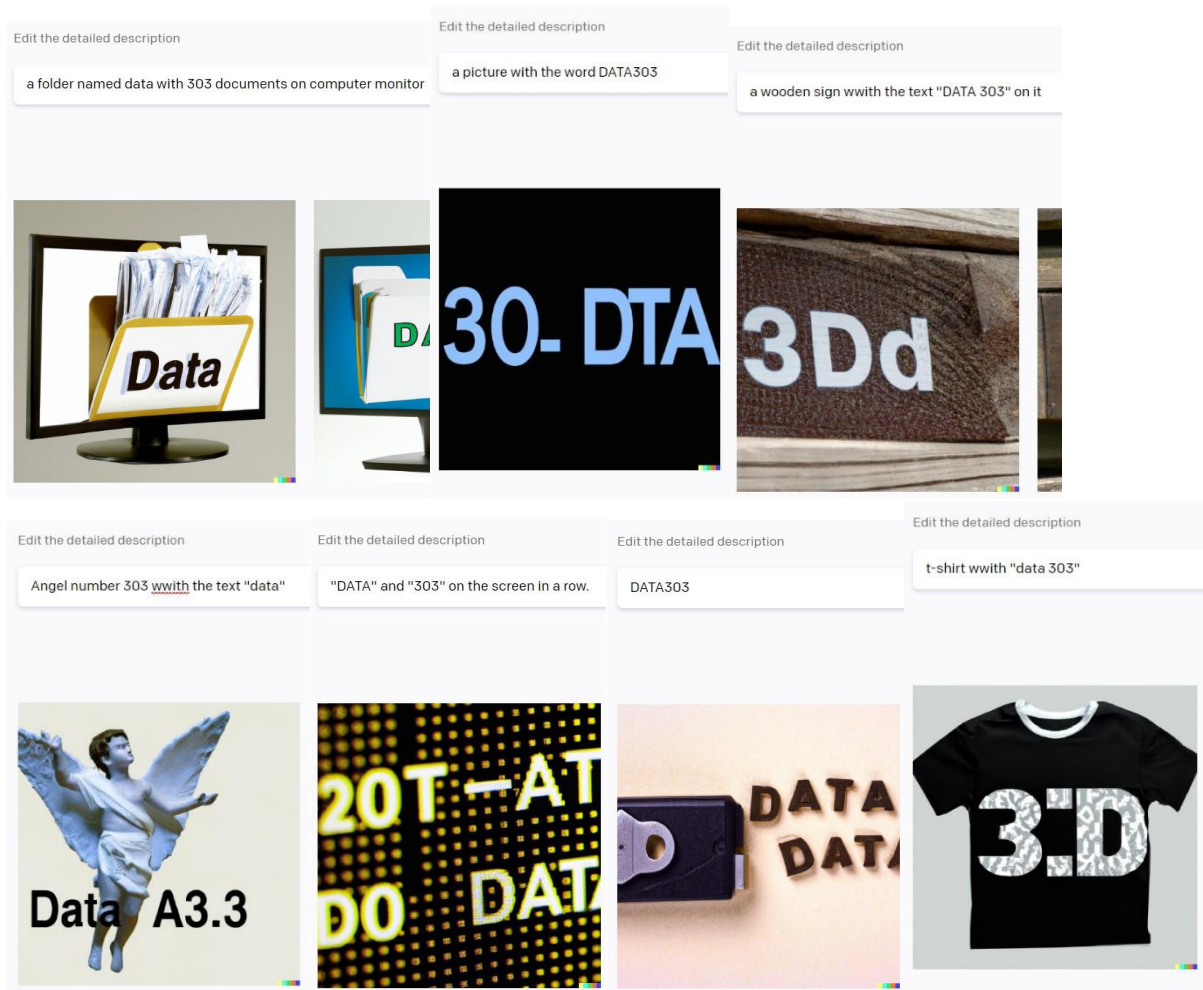


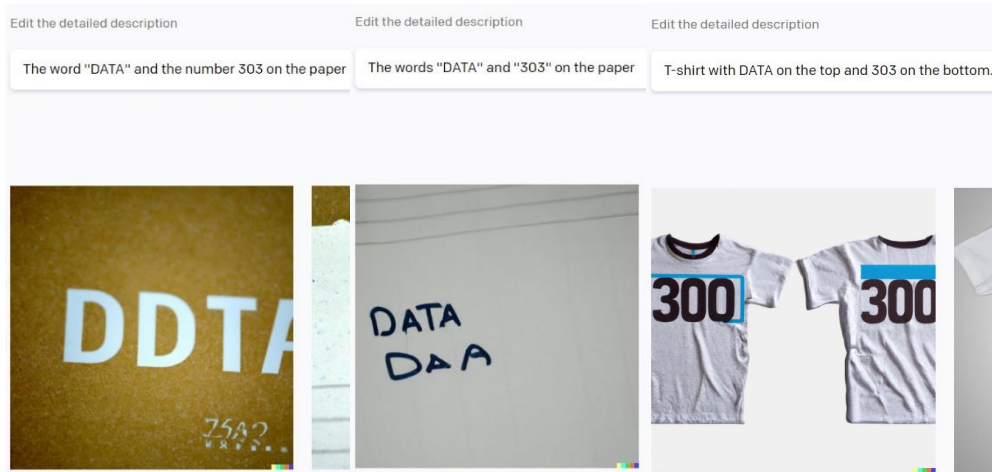
# DATA303 HW1: Image Generation with off-the-shelf algorithms

데이터과학과 2021320322 윤민서



My best image (네 번째 이미지) and prompt (t-shirt wwith "data" and "33003303")





시도했던 prompt와 생성된 이미지 중 일부

첨부한 모든 이미지는 DALL-E 2를 사용하여 생성했다. 첨부한 이미지를 포함하여 약 40회의 시도를 했음에도 불구하고 "DATA303"이 정확하게 포함된 이미지는 생성하지 못했지만 가장 근접한 이미지로 "DATA"와 "3A3"이 포함된 이미지를 확인할 수 있었다. 실험 과정에서 보다 정확한 이미지를 생성하기 위해 Reddit의 *DALLE-2 generated this text correctly: "a wooden sign wwith the text "hello world" on it"*이라는 게시물을 참고했고 내용과 댓글을 통하여 "hello world"와 같이 원하는 문자열을 포함한 이미지를 검색했을 때 충분히 많은 결과가 나오는 경우에는 with 대신에 wwith를 사용하면 원하는 이미지에 더욱 근접한 이미지를 얻어낼 수 있다는 것을 알 수 있었다. 그러나 "303"이라는 문자열이 포함된 이미지는 많지 않았기 때문에 정확한 이미지를 얻어낼 수 없던 것으로 추정된다. "3"과 "0"이 각각 포함된 이미지는 많았기 때문에 prompt에 "3"과 "0"을 무작위로 섞은 비교적 긴 문자열을 입력하였고 결과적으로 "303"과 모양과 구성이 비슷한 "3A3"을 얻어낼 수 있었다.

사용되거나 연구되는 대부분의 생성모델은 분포를 근사하거나 추정하는 방향으로 학습을 한다. 특히 이번에 사용한 DALL-E 2의 경우에는 diffusion을 기반으로 하며 가우시안 잡음을 추가한 데이터를 역방향으로 추정하여 잡음에서 데이터를 만들어내는 방식이다. DALL-E 2는 가장 일관성 있는 이미지를 만들고 prompt를 가장 잘 따르는 경향이 있다. 다른 diffusion 기반 서비스 중 Midjourney는 독특한 스타일을 추구하는 듯한 이미지를 생성해 내지만 가끔 prompt와 불일치하는 문제를 보인다. Stable diffusion은 DALL-E 2와 비교했을 때 일부 prompt에서는 더 나은 이미지를 생성하고 일부 prompt에서는 더 좋지 않은 이미지를 생성하는 경향이 있다. 이와 같은 diffusion 모델 기반 서비스들은 VAE, GAN, Flow-based 모델과 비교했을 때 훨씬 더 사실적이고 정교한 이미지를 생성할 수 있다. 또한 학습 및 생성 과정에서 훨씬 안정적인 모습을 볼 수 있다. 생성 모델의 손실 함수는 다른 task에 비하여 복잡한 경향을 보인다. 하지만 기본적인 diffusion 모델이 발표된 이후에 DDPM, DDIM 등 더 단순하고 더 빠른 방법으로 diffusion process를 진행하는 과정을 연구가 계속해서 이루어지고 있다.