

### 3.2 Data Overlap Analysis - 데이터 중복 분석

-모델의 제로샷 능력(특정 데이터셋이나 태스크에 대해 별도의 fine-tuning 없이도 즉시 동작 가능함)을 엄밀히 평가하기 위한 작업

-테스트로 사용하는 이미지가 학습 데이터와 완전히 같거나 매우 유사할 경우 모델의 성능이 과대평가될 가능성 o

-DALL·E는 2억 5천만 개의 텍스트-이미지 쌍을 인터넷에서 수집한 대규모 데이터셋으로 학습되어 이미 학습된 이미지가 테스트에 사용될 가능성이 있음

>대조 학습 기반 이미지 임베딩 모델 Contrastive Similarity Model을 사용하여 제거

### 3.3 Qualitative Findings - 정성적 발견

#### Abstraction & Concept Composition (고차 추상 개념 조합 능력)

모델이 언어적으로 추상화된 개념을 시각적으로 재해석할 수 있음

프롬프트: "a tapir made of accordion"

>생성 결과: 테이퍼의 몸이 아코디언처럼 되어 있거나, 아코디언의 키보드가 테이퍼의 주둥이처럼 묘사됨

-두 개념의 특징적 속성을 통합해 새로운 형태로 창조함, 모델이 단어가 아닌 의미 단위로 이미지를 생성하고 있다.

#### Variable Binding & Semantic Parsing (변수 결합 및 문장의 의미 구조 분석 능력)

복잡한 문장 구조 이해

프롬프트: "an illustration of a baby hedgehog in a christmas sweater walking a dog"  
(크리스마스 스웨터를 입은 아기 고슴도치가 개를 산책시키는 그림)

>일부 그림에서는 고슴도치와 개 모두 스웨터 착용 or 큰 고슴도치가 작은 고슴도치를 산책시키는 식의 해석 등장

-복잡한 구조를 가지는 문장, 언어 구문을 단순히 나열하는 수준을 넘어 구조적으로 해석하려는 경향을 보임 but 한계 존재

#### Zero-Shot Image-to-Image Translation (언어 제어 기반 이미지-투-이미지 전환)

시각 표현 조작 능력

프롬프트: "the same cat on the top as a sketch on the bottom" (고양이에 대한 실제 사진을 보여주고 그 고양이에 대한 스케치를 그리게 함)

>위쪽은 실제 사진, 아래쪽은 같은 고양이를 스케치 스타일로 재현

-이미지 내의 객체의 위치, 속성, 스타일 등을 어느정도 분할 및 추론하고 의미론적 조작을 수행 가능

