

## 9. 데이터 분석의 혁신



# Generative AI

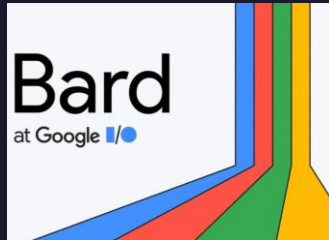
생성형 AI는 인공지능망을 이용하여 새로운 데이터를 생성해내는 기술로  
명령어(Prompt)를 통해 사용자의 의도를 스스로 이해하고  
주어진 데이터로 학습, 활용하여  
텍스트, 이미지, 오디오, 비디오 등 새로운 콘텐츠를 생성해내는 인공지능이다.

ChatGPT



<https://chat.openai.com/>

Bard



<https://bard.google.com/>

DALL-E 2



<https://openai.com/product/dall-e-2>

Stable  
Diffusion



<https://stablediffusionweb.com/>

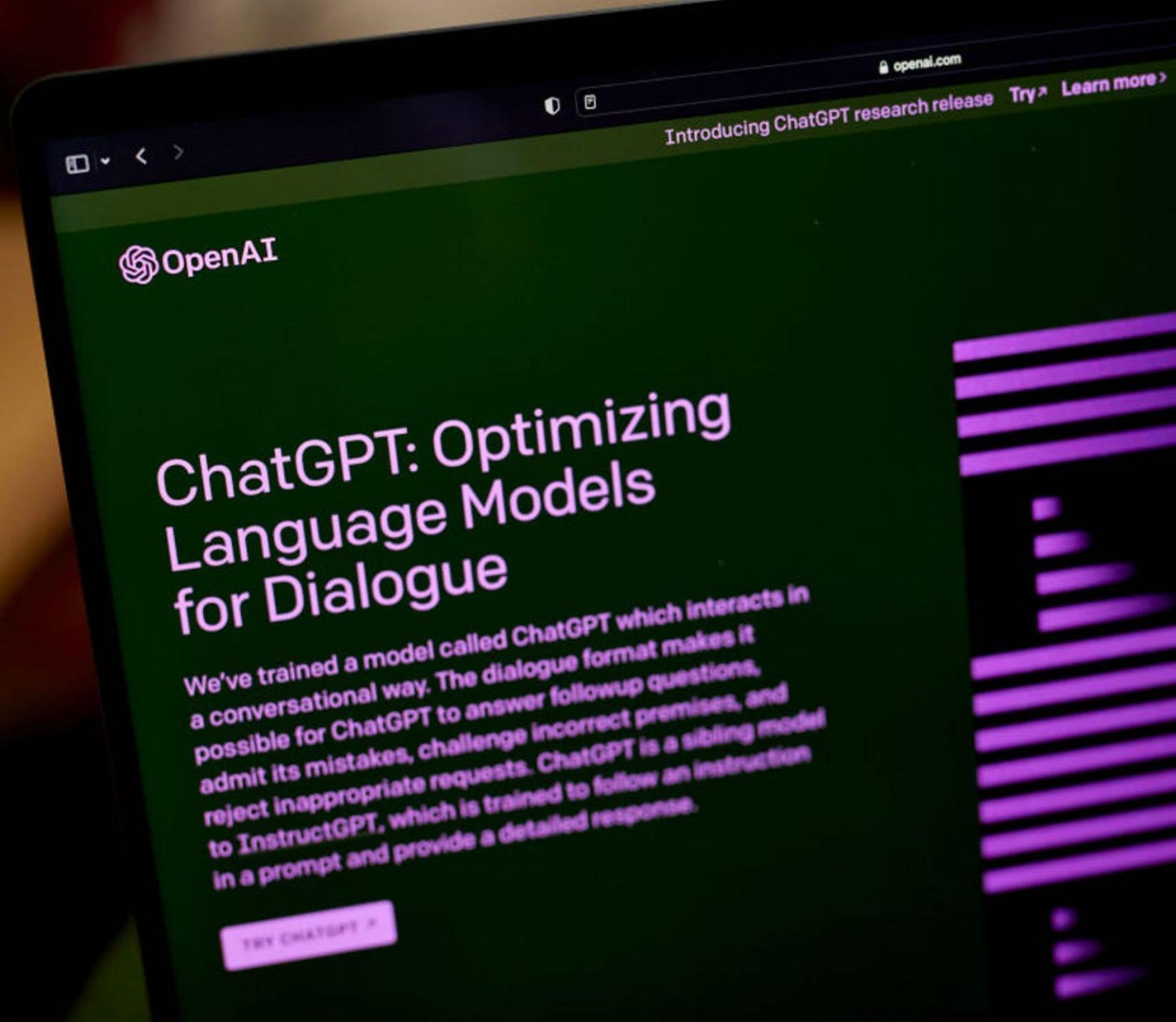
Midjourney



<https://www.midjourney.com/>

# ChatGPT

<https://chat.openai.com/>






# ChatGPT

텍스트 이해  
질문에 대한 답변  
콘텐츠 생성  
리스트 생성  
코드 작성/디버깅  
장단점 비교 제시  
교육 지원  
창의적인 글쓰기  
아이디어 기획  
번역  
단계별 지침 제공  
가상 비서 역할

[Reset Thread](#)  
[Dark Mode](#)  
[OpenAI Discord](#)  
[Learn More](#)  
[Log out](#)

## ChatGPT

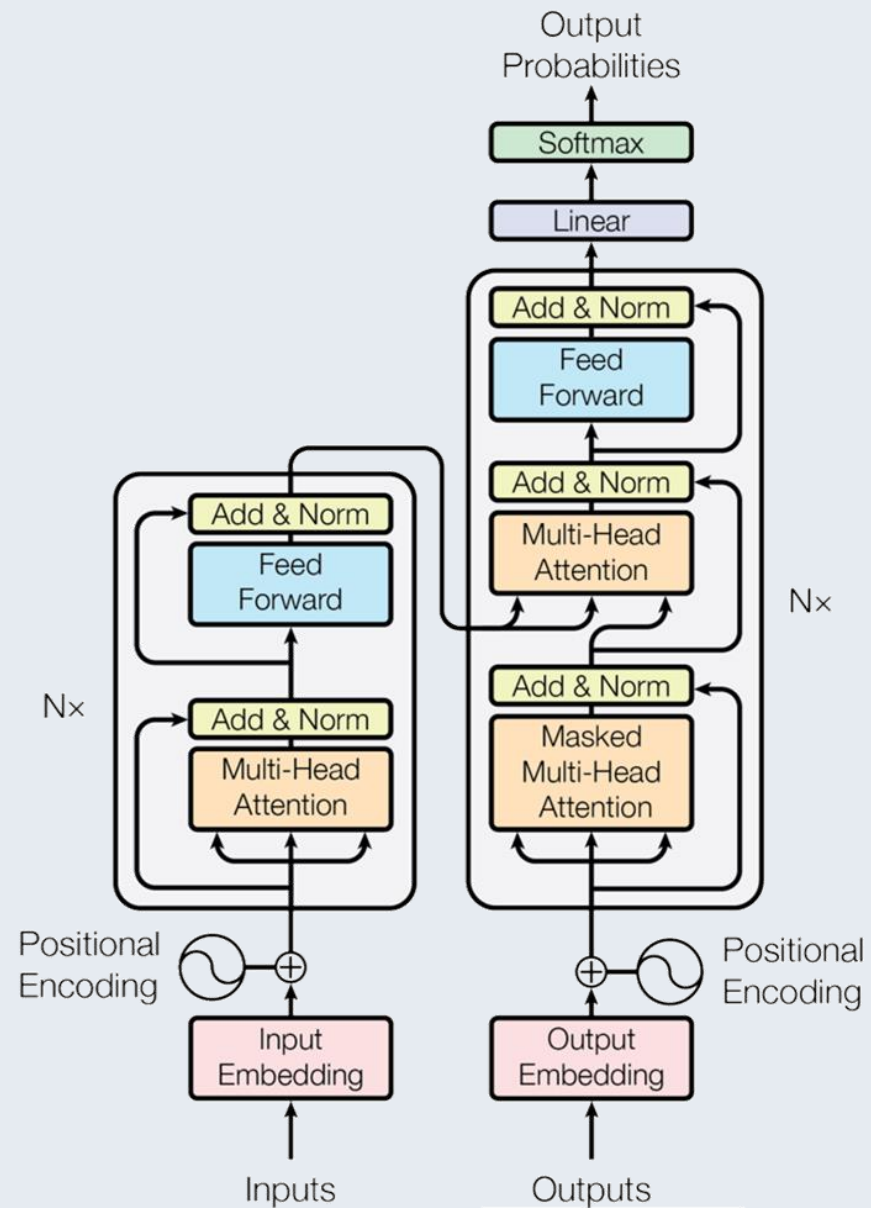
 Examples	 Capabilities	 Limitations
"Explain quantum computing in simple terms"	Remembers what user said earlier in the conversation	May occasionally generate incorrect information
"Got any creative ideas for a 10 year old's birthday?"	Allows user to provide follow-up corrections	May occasionally produce harmful instructions or biased content
"How do I make an HTTP request in Javascript?"	Trained to decline inappropriate requests	Limited knowledge of world and events after 2021

Free Research Preview: ChatGPT is optimized for dialogue. Our goal is to make AI systems more natural to interact with, and your feedback will help us improve our systems and make them safer.

# GPT

**GPT** = Generative  
Pre-trained  
Transformer

## Transformer





# 임베딩(Embedding)

## 원핫 인코딩(one hot encoding)

Human-Readable

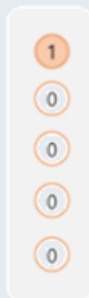
Pet
Cat
Dog
Turtle
Fish
Cat

Machine-Readable

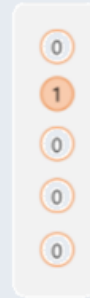
Cat	Dog	Turtle	Fish
1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	1
1	0	0	0

## 희소표현(sparse representation)

고양이

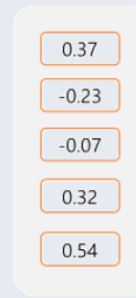


강아지

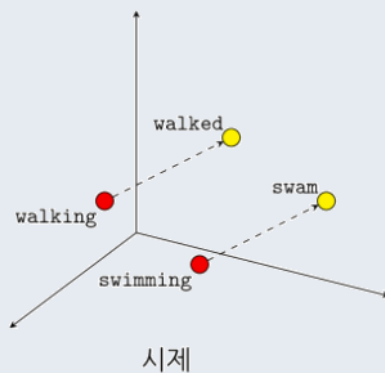
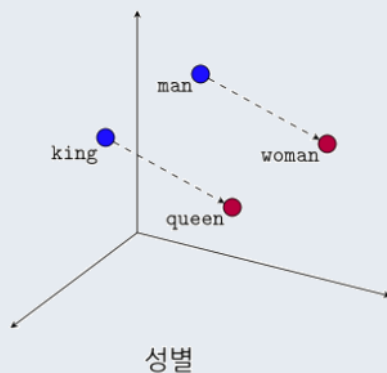
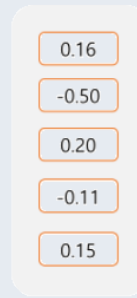


## 밀집표현(distributed representation)

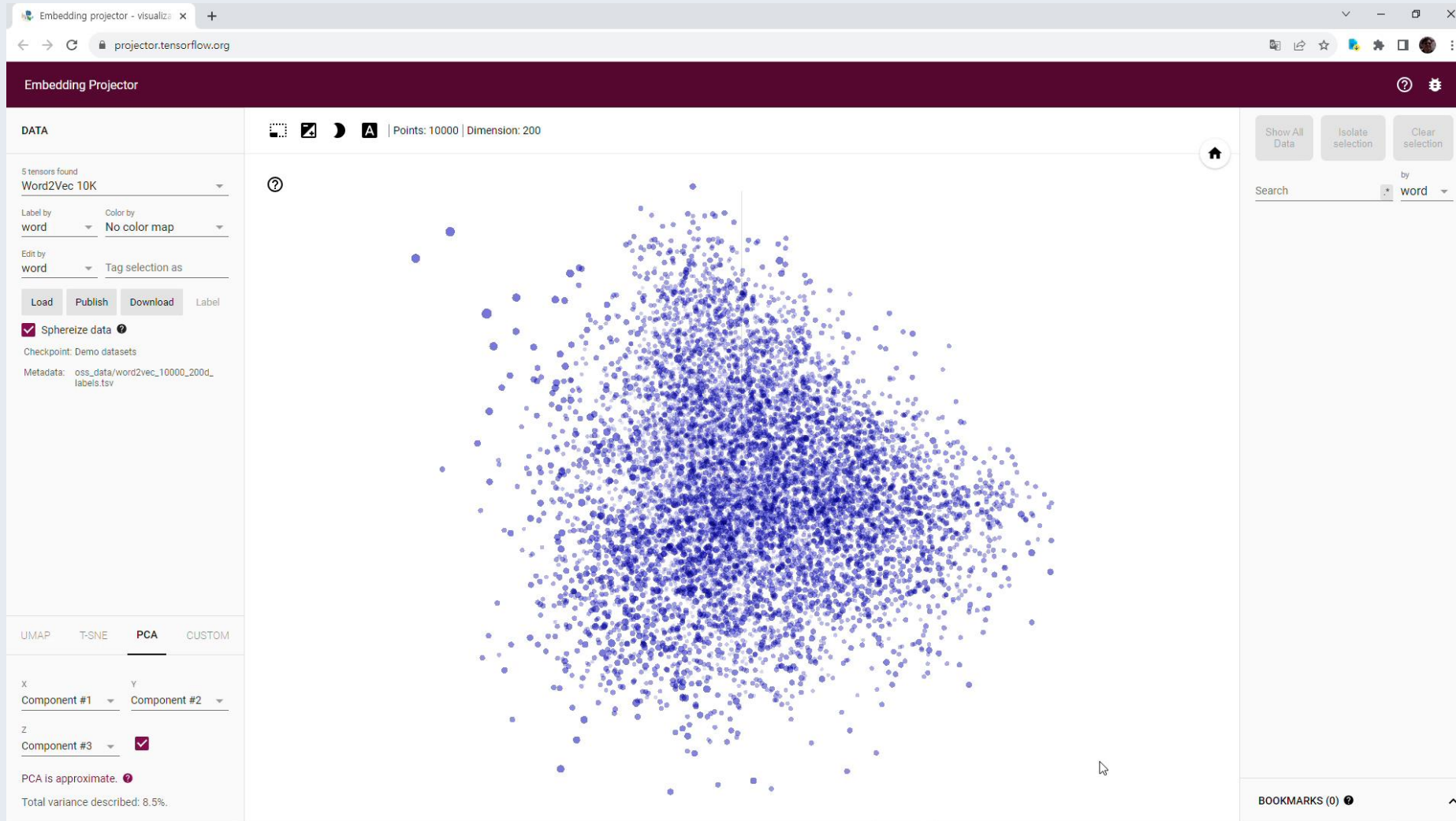
고양이



강아지



# 임베딩(Embedding)

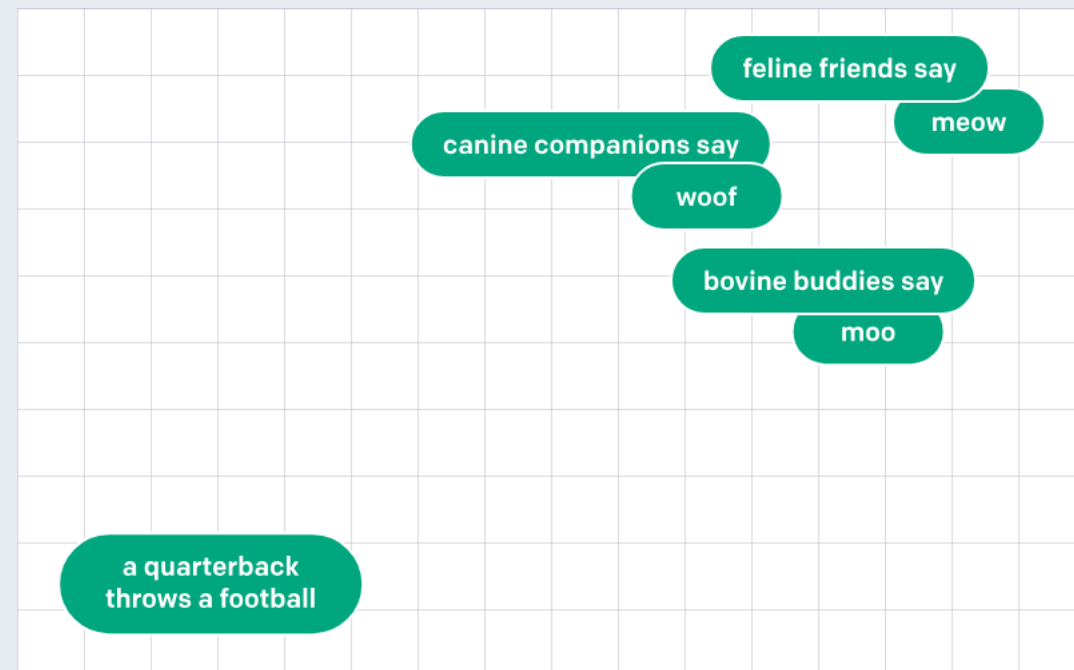
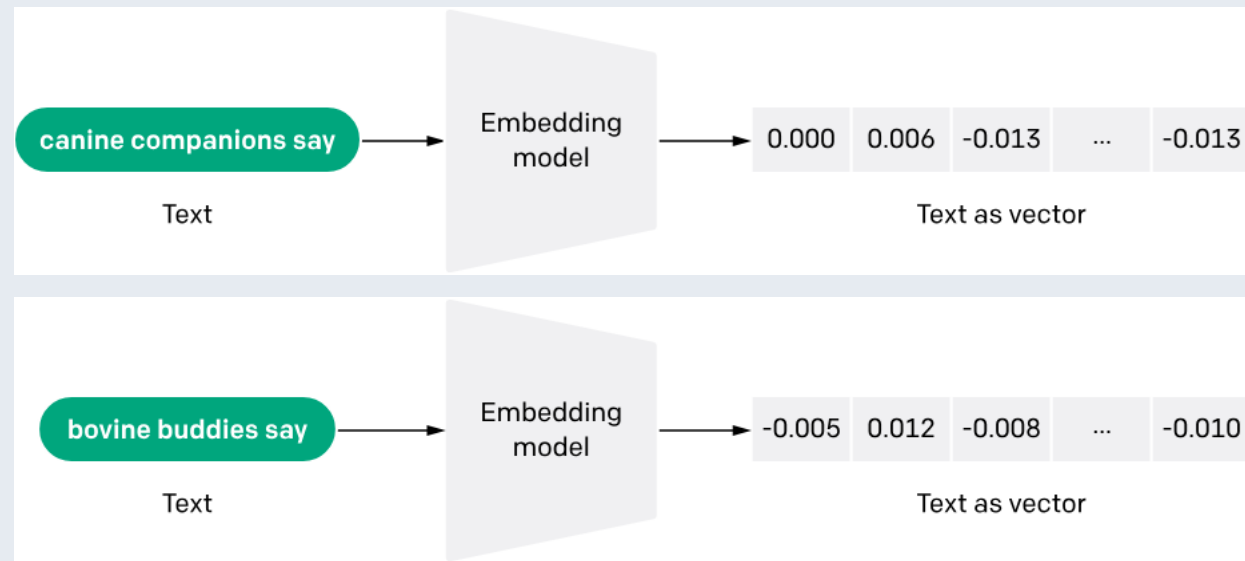


<https://projector.tensorflow.org/>

# 임베딩(Embedding)

```
import openai
response = openai.Embedding.create(
    input="canine companions say",
    engine="text-similarity-davinci-001")

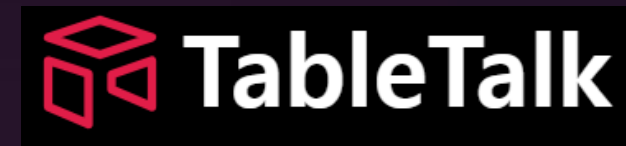
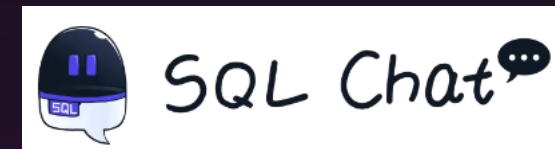
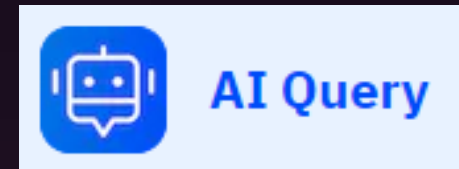
print(response)
{
  "data": [
    {
      "embedding": [
        0.000108064,
        0.005860855,
        -0.012656143,
        ...,
        -0.006642727,
        0.002583989,
        -0.012567150
      ],
      "index": 0,
      "object": "embedding"
    }
  ],
  "model": "text-similarity-davinci-001",
  "object": "list"
}
```



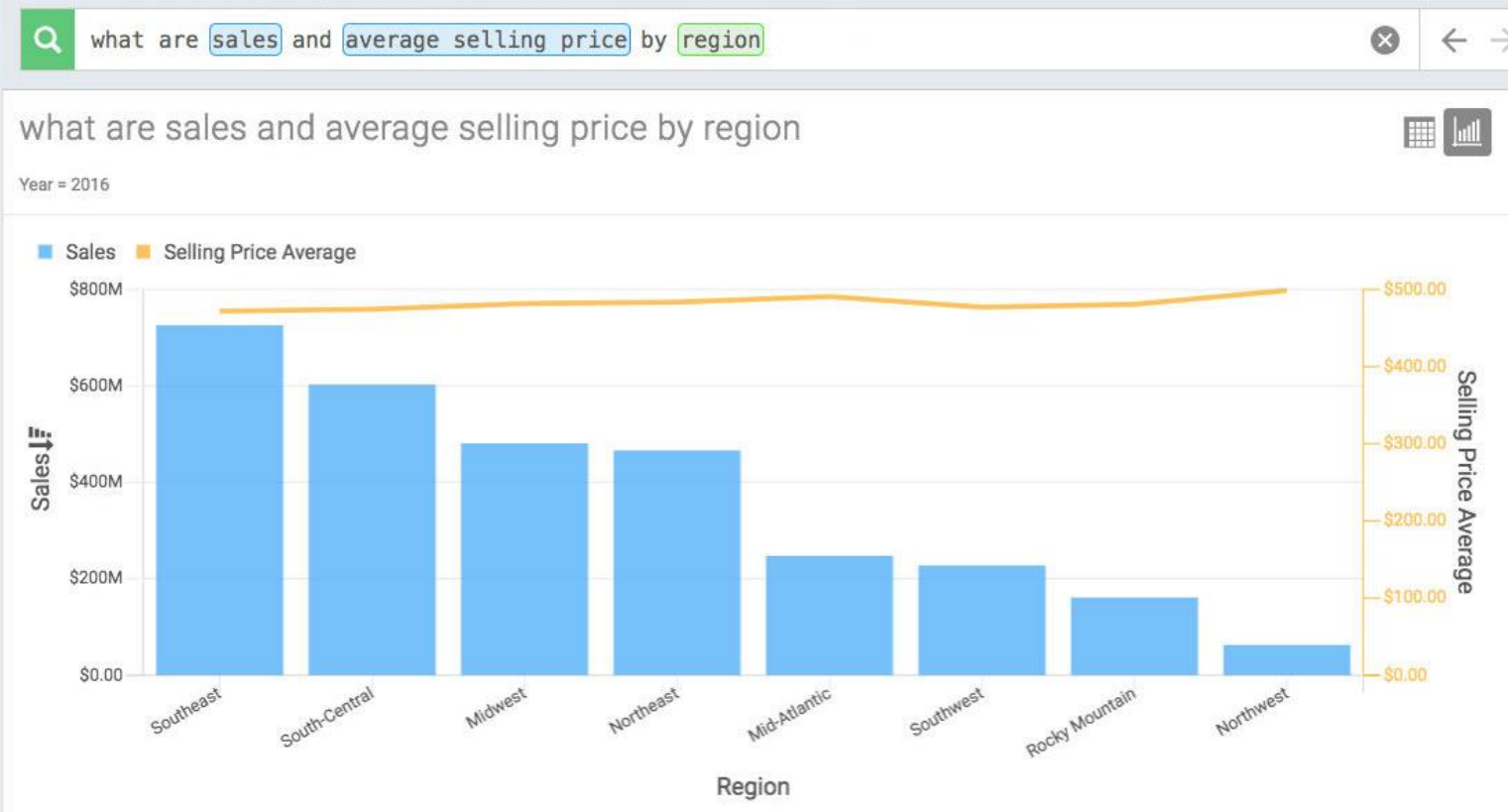


# Generative AI와 데이터 분석

TEXT2SQL.AI  
Generate SQL with  
AI in seconds!



# 데이터 분석 혁신



## Natural Language Query



## Natural Language Generation

### INSIGHTS

#### General

**Explanation** – Chart shows Sales broken out by Manufacturer from Jan 2015 to Sep 2018.

**Summary** – From Jan 2015 to Sep 2018, Manufacturer (Asteroid Belt Beers) had the largest Sales increase, rising by **1.62M** for a gain of **↑1.22%** share. By the same measure, Manufacturer (Dr Parsec) fell the most, losing **↓2.13%** in share despite growing by **1.34M**.

**Growth** – Comparing Jan 2015 to Sep 2018, 5 of 5 (100%) Manufacturers grew by a total of **9.71M** (**↑60.83%**), and none shrank.

**Pareto** – Top 1 of 5 (20%) Manufacturer (Rocket Cola) accounted for **356M** (38%) of all Sales, while the other 4 of 5 (80%) made up **593M** (62%).

**Outliers** – Extraordinary Sales spikes occurred for Jul 2018, Rocket Cola, Aug 2018, Rocket Cola, and Jun 2018, Rocket Cola.

# Sketch

<https://github.com/approximatelabs/sketch>

```
In [1]: import sketch  
import pandas as pd
```

```
In [2]: sales_data = pd.read_csv("https://gist.githubusercontent.com/bluecoconut/9ce2135aafb5c6ab2dc1d60ac595646e/raw/c93c3500a
```

```
In [ ]:
```

```
In [ ]:
```

# The Future of Data Analysis



오늘날 경쟁이 치열한 글로벌 비즈니스 환경에서  
데이터와 데이터 분석의 힘이 드러났습니다.

Generative AI는 데이터 분석이 수행되는  
방식을 빠르게 변화시키고 있습니다.

데이터 분석 기술과 AI 기술로  
데이터의 Power를 키우시길 바랍니다.



감사합니다.

