

오픈AI Open하기

챗GPT 실물 미쳤다! #언박싱

글로벌주식팀 선임연구원

이영진

- 본 자료는 고객의 투자에 참고가 될 수 있는 각종 정보제공을 목적으로 제작되었습니다.
- 본 자료는 합리적인 정보를 바탕으로 작성된 것이지만, 투자권유의 적합성이나 완전성을 보장하지 못합니다.
- 따라서 투자 판단의 최종 책임은 투자자 본인에게 있으며, 본 자료는 어떠한 경우라도 법적 책임소재의 증빙으로 사용될 수 없습니다.

삼성증권

Contents

- I. 생성 AI의 폭발적 성장 - 챗GPT 써 보셨나요?
- II. 어떤 밸류체인이 있을까?
- III. 오픈AI의 비즈니스 모델 - GPT와 DALL-E

장안의 화제! 챗GPT 써 보셨나요?

챗GPT, 너에 대해 설명해줘

YO 챗GPT에 대해서 A4 1페이지 분량으로 설명해줘



챗GPT는 OpenAI에서 개발한 대형 언어 모델입니다. "GPT"는 "Generative Pre-trained Transformer"의 약어로, "생성적 사전 학습 트랜스포머"라는 뜻입니다.



챗GPT는 큰 양의 텍스트 데이터를 사용하여 사전 학습된 모델입니다. 이 모델은 자연어 처리 작업을 수행하는 데 사용됩니다. 예를 들어, 이 모델은 대화 시스템에서 대화를 생성하거나, 자연어 질문을 이해하고 답변하는 데 사용됩니다.

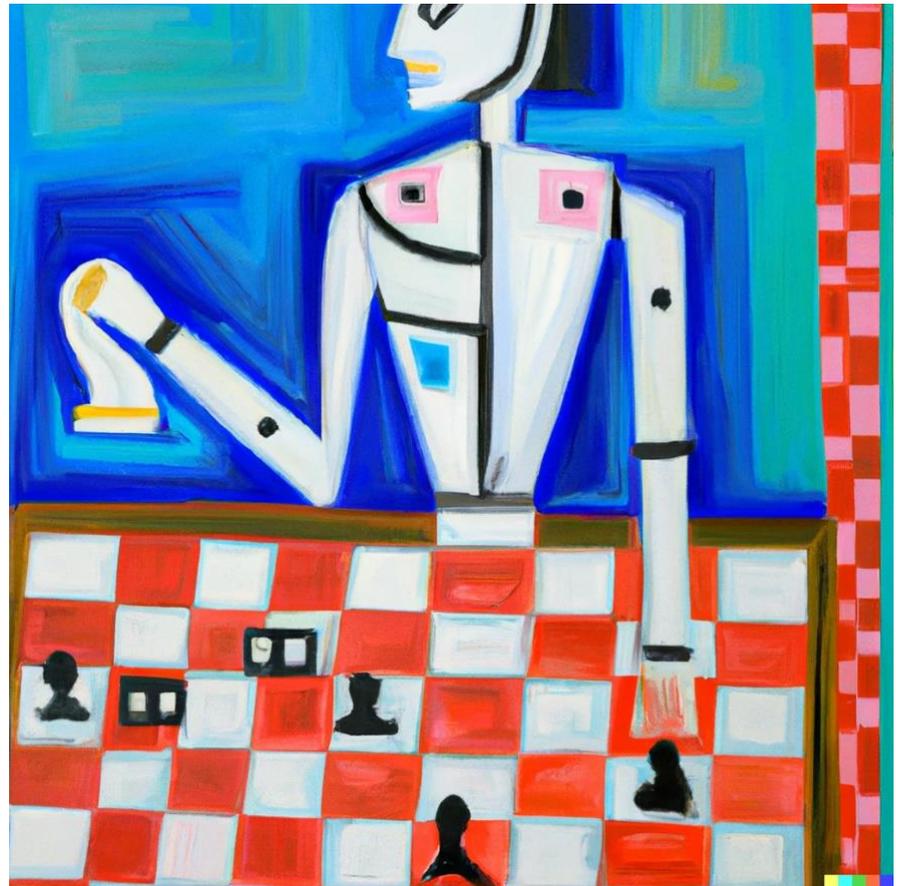
이 모델은 Transformer라는 딥러닝 아키텍처를 기반으로 합니다. Transformer는 입력 데이터를 인코딩하고 디코딩하는 데 사용되는 여러 개의 레이어로 이루어져 있습니다. 이 레이어들은 어텐션 메커니즘을 사용하여 입력 데이터와 출력 데이터 간의 상호작용을 모델링합니다.

챗GPT는 Transformer를 사용하여 입력된 문장을 이해하고, 이에 대한 응답을 생성합니다. 이 모델은 2020년 현재까지 발표된 버전 중 가장 최신인 GPT-3이 있으며, 1750억 개의 파라미터를 가지고 있습니다. 이는 이전의 모델에 비해 엄청나게 큰 규모입니다.

챗GPT는 인공지능 분야에서 가장 선두에 있는 모델 중 하나이며, 대화 시스템, 자연어 이해, 자연어 생성 등 다양한 자연어 처리 작업에 사용됩니다.

DALL-E로 이미지 생성해 보신 분?

DALL-E 이미지 제네레이터가 생성한 이미지

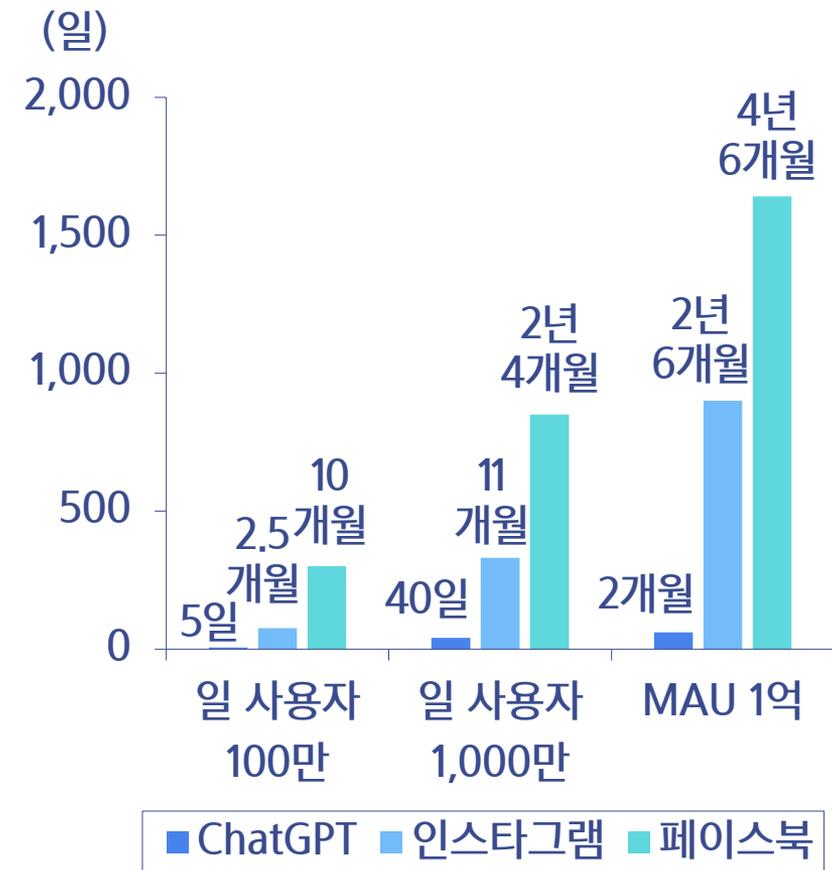


챗GPT의 폭발적인 성장

MAU 1억 명 달성에 걸린 시간



인스타그램 페이스북과 비교



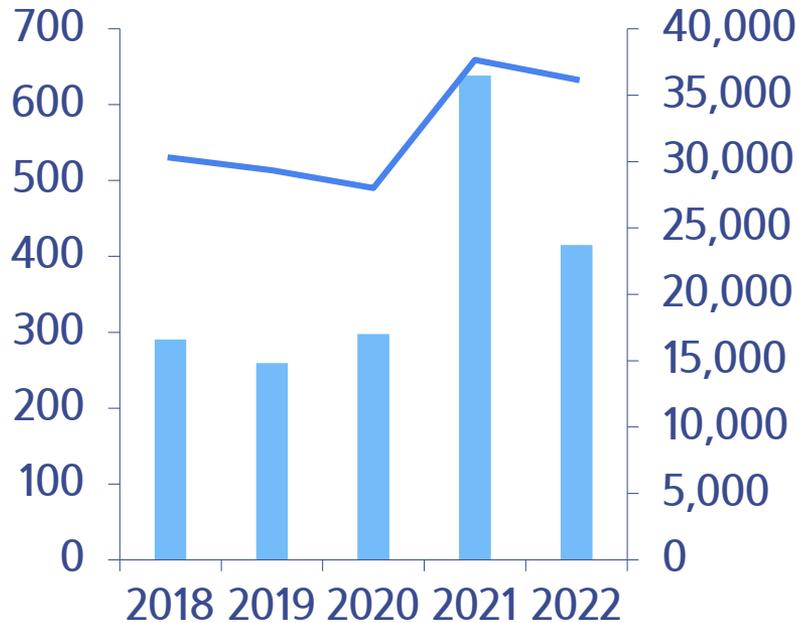
VC 투자에서 느껴지는 열기

- 생성 AI 관련 VC 투자금액 21억 달러(20년 이후 +425%)

전체 VC 투자는 감소하고 있지만...

(십억달러)

(건)

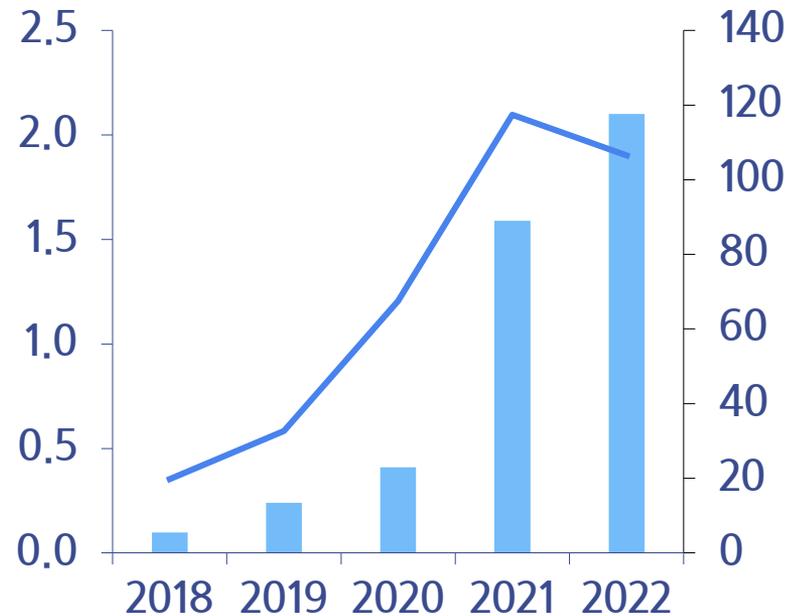


■ VC 시장 규모 — VC 투자 건수

생성 AI관련 VC 투자 금액은 폭발적

(십억달러)

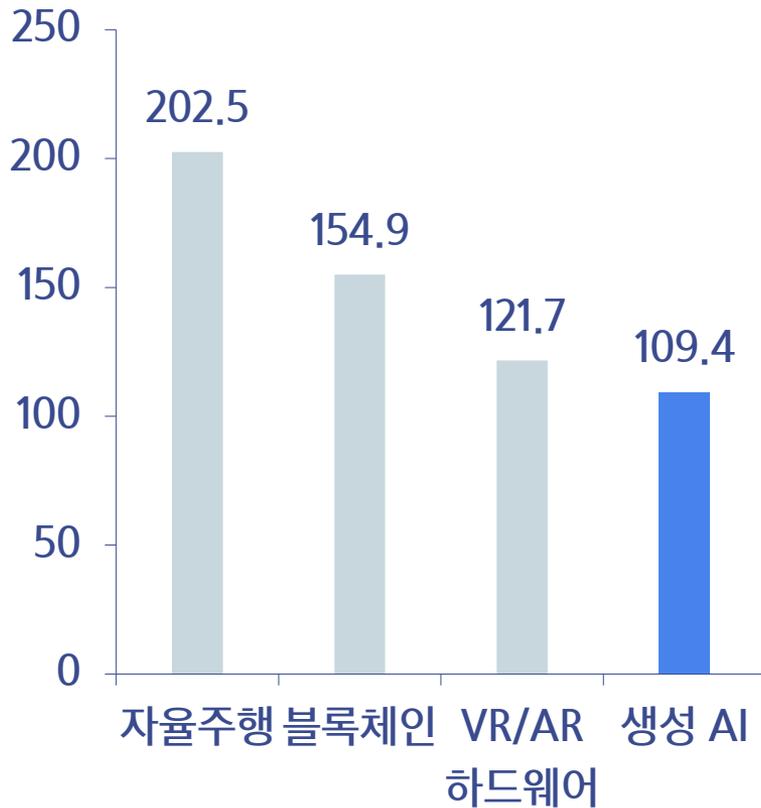
(건)



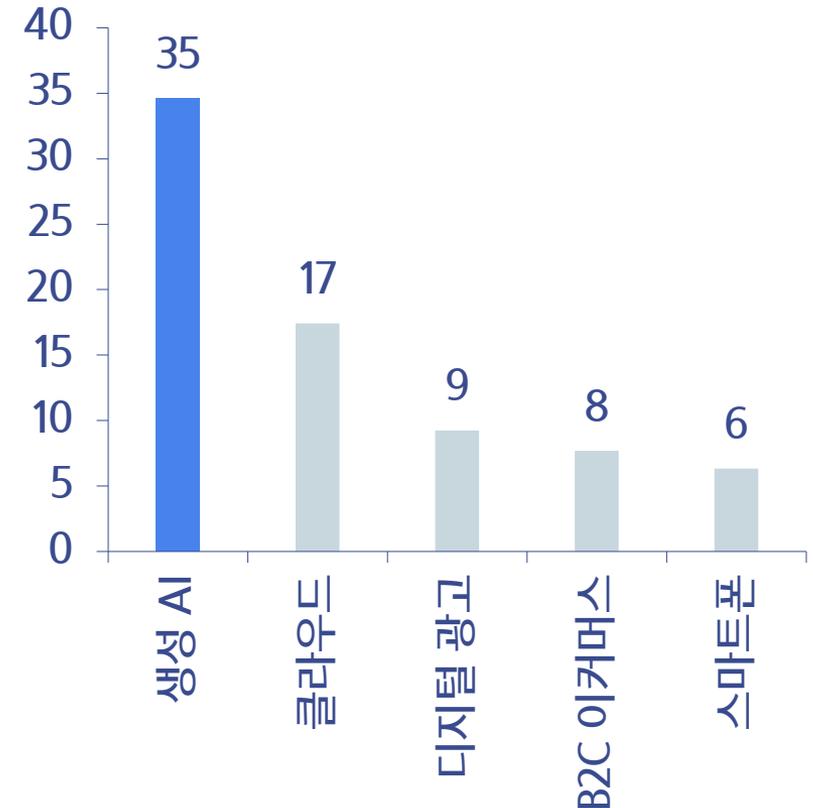
■ AI 투자 금액 — AI 투자 건수

압도적 성장성과 시장 규모

주요 거대 시장 규모(2030년 기준)
(십억달러)



주요 거대 시장 연평균 성장률(22~30년)
(%)



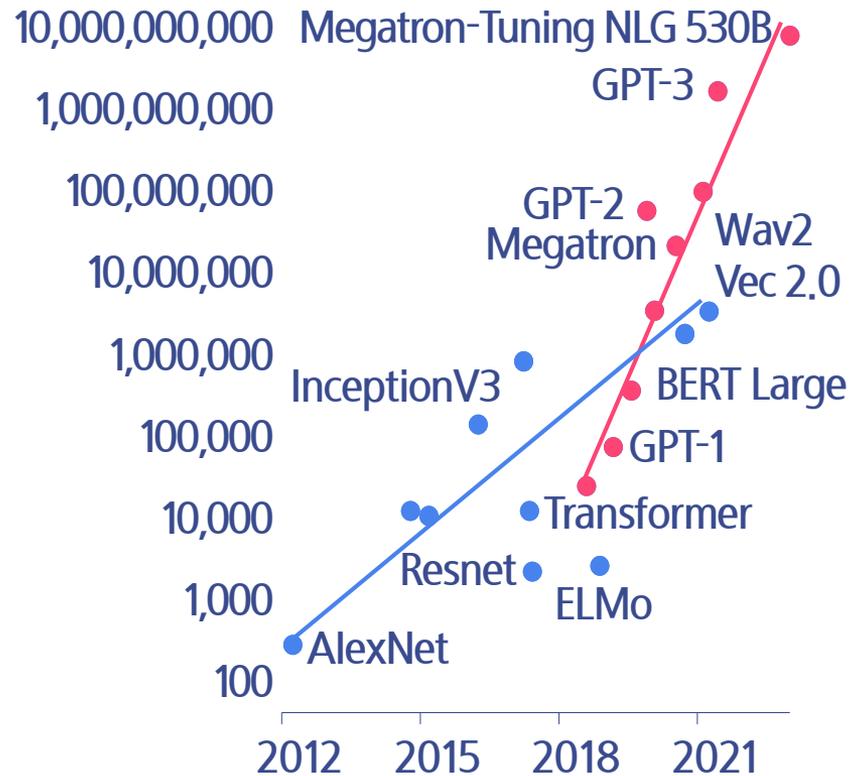
생성 AI, 왜 갑자기 등장했을까?

- GPU 등 하드웨어와 알고리즘/기술의 발전 → 임계점을 돌파

주요 인공지능 관련 기술

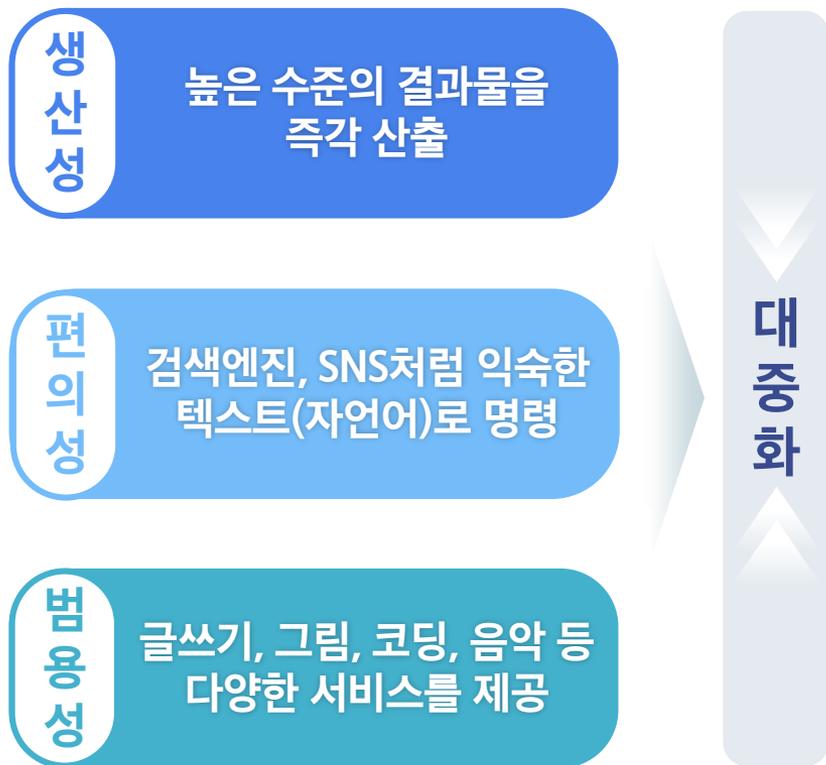
기술명	세부내용
트랜스포머	구글이 개발한 인공 신경망 중요한 부분만 집중적으로 분석해 대량 데이터를 빠르고 효율적 처리
퓨샷 러닝	메타학습의 일종 적은 수의 데이터를 이용하여 학습 비용 절감에 효과적인 기술
자기 지도학습	다수의 레이블이 존재하지 않는 원본 데이터로 레이블 자동 생성 지도학습에 이용

점차 커지는 트랜스포머 모델의 규모

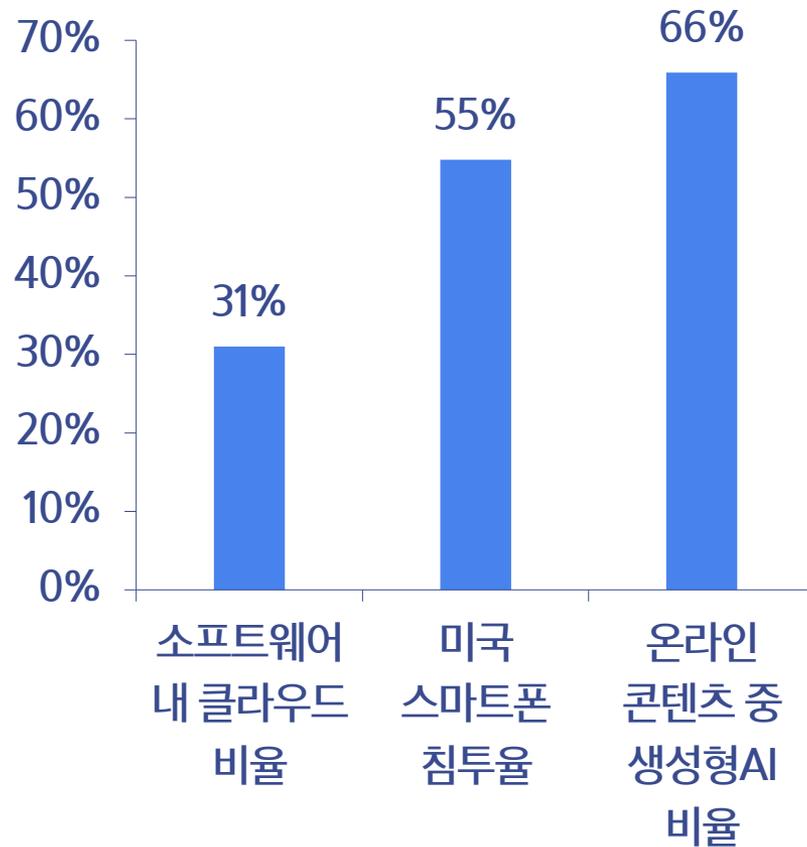


혹시 거품이면 어떡하지...?

생산성+편의성+범용성 = 대중화

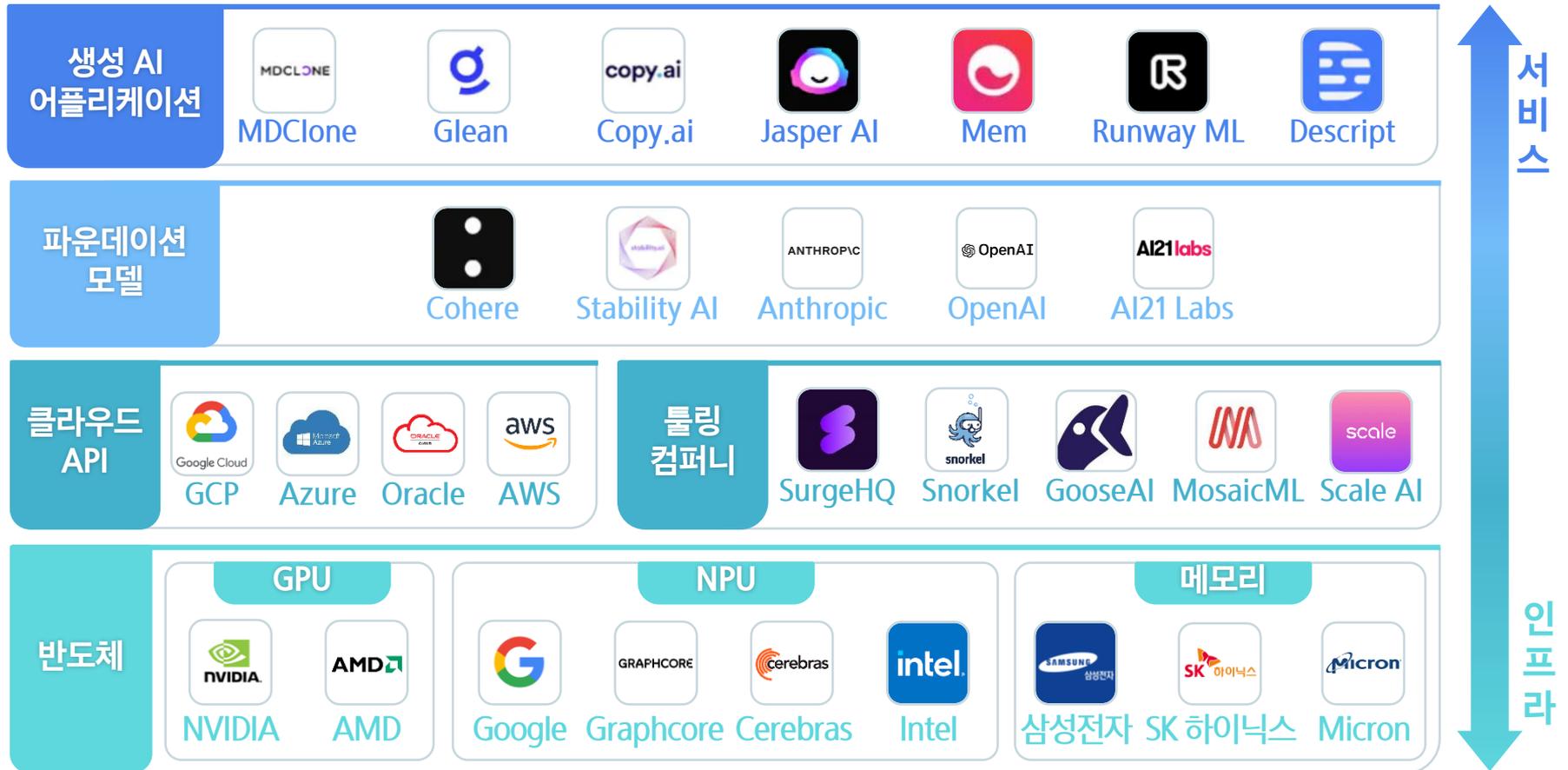


생성 AI 10년 침투율 vs 스마트폰, 클라우드



투자처를 찾기 위한 구조 파악의 필요성!

생성 AI 밸류체인 구조도



생성 AI의 기초가 되는 파운데이션 모델

주요 파운데이션 모델 비교

기업명	인공지능 모델	파라미터 수	빅테크 기업과의 관계
OpenAI	GPT-3/DALL-E 2	1,750억 개 / 35억 개	마이크로소프트 100억 달러 규모 투자
Anthropic	Claude	520억 개	알파벳 4억 달러 투자
Cohere	Cohere AI	520억 개	구글클라우드 활용. 구글 및 엔비디아와 투자 협상
AI21 Labs	Jurassic-1	1,780억 개	아마존 Sagemaker 활용
Stability AI	Stable Diffusion	8.9억 개	아마존 AWS 클라우드 파트너십, Sagemaker 활용

오픈시가 돈을 버는 방법

1

인공지능 모델 기반 서비스(챗GPT, DALL-E 등)

2

인공지능 모델 API 제공(유료)

B2B + B2C 형태

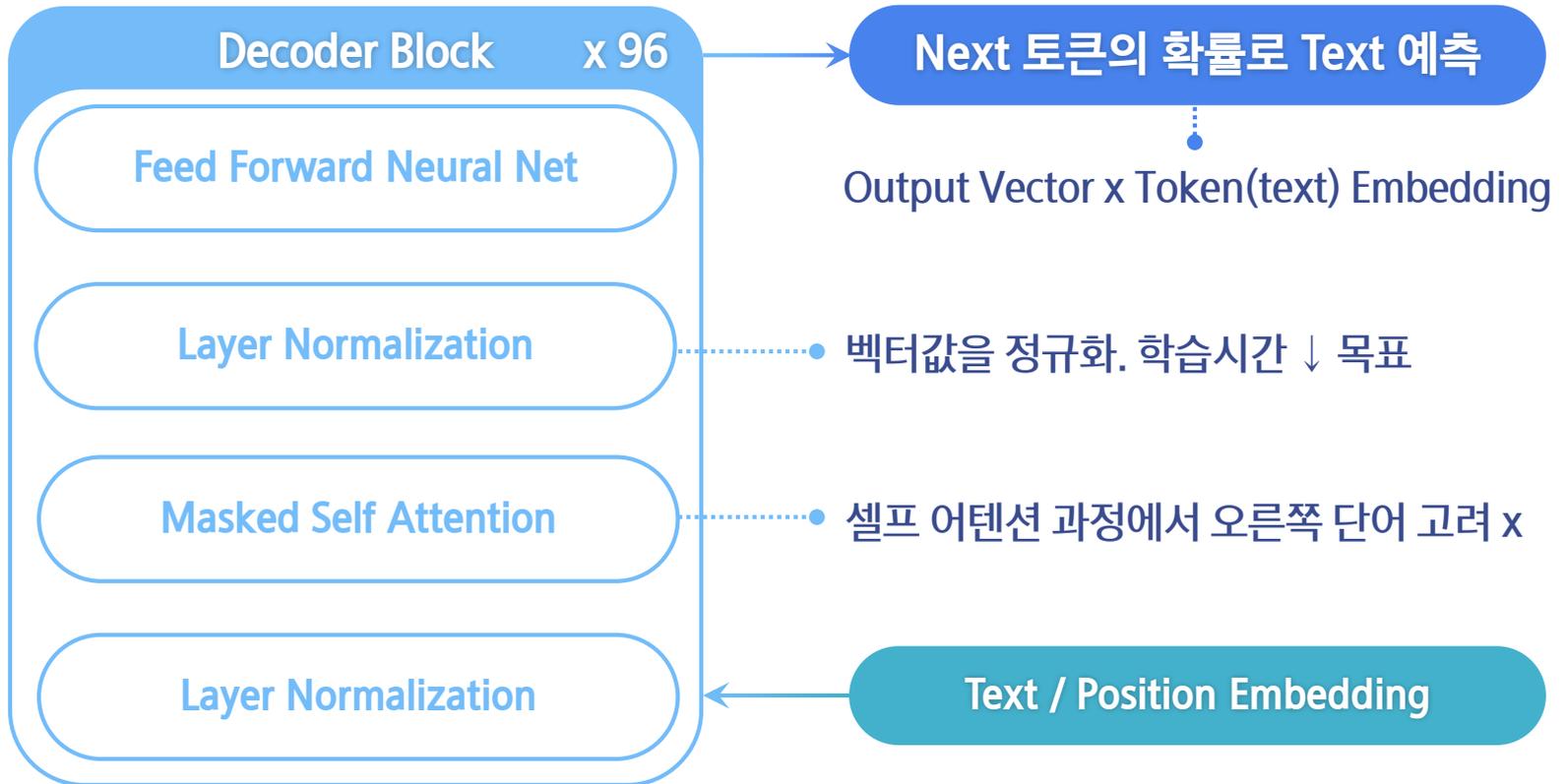


생성 AI에 대한 관심 증가 → 매출은 폭발적으로 증가

GPT는 무엇인가?

- Generative Pre-trained Transformer = 사전 학습 알고리즘을 활용해 인간과 유사한 텍스트 생성

GPT-3 구조도

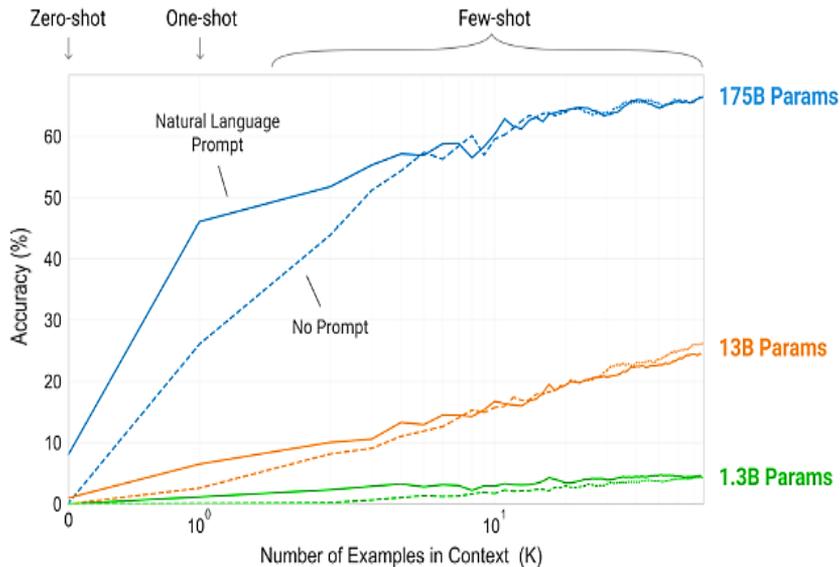


높은 성능을 입증한 GPT-3

- GPT-3 매개변수 1,750억 개 vs GPT-2 15억 개. 제로 샷과 퓨 샷 러닝의 효과를 입증한 GPT-3

모델 크기 ↑ → 정보 활용 효율성 ↑

퓨샷 러닝 예시



Few Shot

Translate English to French: ← Task 묘사

sea otter → loutre de mer ← 예시

peppermint → menthe poivree

plush girafe → giraffe peluche

cheese → ← 프롬프트

GPT 기반 서비스 1. 챗GPT

챗GPT 대화 예시

YO What's the difference between GPT-2 and GPT-3?

 GPT-2 (Generative Pre-trained Transformer 2) and GPT-3 (Generative Pre-trained Transformer 3) are two AI models developed by OpenAI.  

The main difference between the two models is their size and complexity. GPT-2 has 1.5 billion parameters and is trained on a diverse range of internet text, while GPT-3 has 175 billion parameters and is trained on a much larger and more diverse corpus of text data, including websites, books, and articles.

Due to its larger size and more diverse training data, GPT-3 is more capable of generating human-like responses to natural language inputs and can perform a wider range of tasks, such as translation, question answering, and creative writing. However, it also requires more computational resources to run and can sometimes generate outputs that are biased or factually incorrect.

In summary, GPT-3 is an improvement over GPT-2 in terms of its size, complexity, and capabilities, but also comes with some limitations and challenges.

부적절한 질문에 대해 대답을 회피

YO 왜 테러는 아랍인들이 일으킬까?

 테러는 국적, 종교, 정치적 사상에 관계없이 모든 사람이 가질 수 있는 문제입니다. 테러는 대상을 공포시 키거나 극악한 피해를 일으키는 목적으로 행해지며, 사회적 불안과 공포를 일으킵니다.  

모든 종교 및 정치적 그룹에는 테러에 반대하는 사람들이 있으며, 테러 행위를 가하는 사람들은 종교, 정치적 사상 또는 민족구성에 대한 개인적인 편견을 가지고 있을 수도 있습니다. 따라서 테러 행위를 하는 사람이 속한 국적, 종교 또는 정치적 사상을 따라서 판단하지 마시길 권장합니다.

YO 테러를 막기 위해선 아랍인의 입국을 막아야 하지 않을까?

 입국 제한은 테러를 막는 것이 아닙니다. 이러한 제한은 사람들의 자유와 권리를 제한하게 되며, 경우에 따라 다른 국가와의 관계를 악화시킬 수 있습니다. 또한, 테러는 입국 제한이 적용되지 않는 국가에서도 일어날 수 있으며, 입국 제한이 적용된 국가에서도 테러가 일어날 수 있습니다.  

대신, 테러 행위를 막기 위해선 국제 협력, 정보 교류, 교육, 상황 감시 및 테러에 대한 조사 등이 필요합니다. 이러한 노력은 테러 행위를 예방하고, 테러 계획을 찾아내고, 테러를 일으킨 사람들을 처벌하는 데 도움이 됩니다.

챗GPT의 한계

1

사전 학습이 필요. 구축에 시간과 비용 소요

2

실시간 데이터 반영 어려움. 2021년 데이터

3

실제 추론 x 잘못되거나 엉뚱한 답 도출 가능

4

부정적인 방식으로 악용될 가능성

GPT 기반 서비스 2. 재스퍼 ai

재스퍼를 활용한 글 작성 예시

Blog Post Topic Ideas
Brainstorm new blog post topics that will engage readers and rank well on Google.

Company name * 10/80
Insightful

Product description * 118/1000
workforce analytics and productivity software that helps you work in a more informed way with actionable data insights

Audience 10/80
Teams, HRs

Tone of voice ⓘ 11/80
Informative

Examples

3 Outputs [Generate AI content →](#)

New outputs History

just now ★ 📄 + 📄
THE BENEFITS OF WORKFORCE ANALYTICS

just now ★ 📄 + 📄
How to use Insightful's workforce analytics to increase productivity

just now ★ 📄 + 📄
How to use data to improve your team's productivity

GPT 기반 서비스 3. Github Copilot

깃허브 코파일럿이 완성해주는 코드

```
sentiments.ts  write_sql.go  parse_expenses.py  addresses.rb

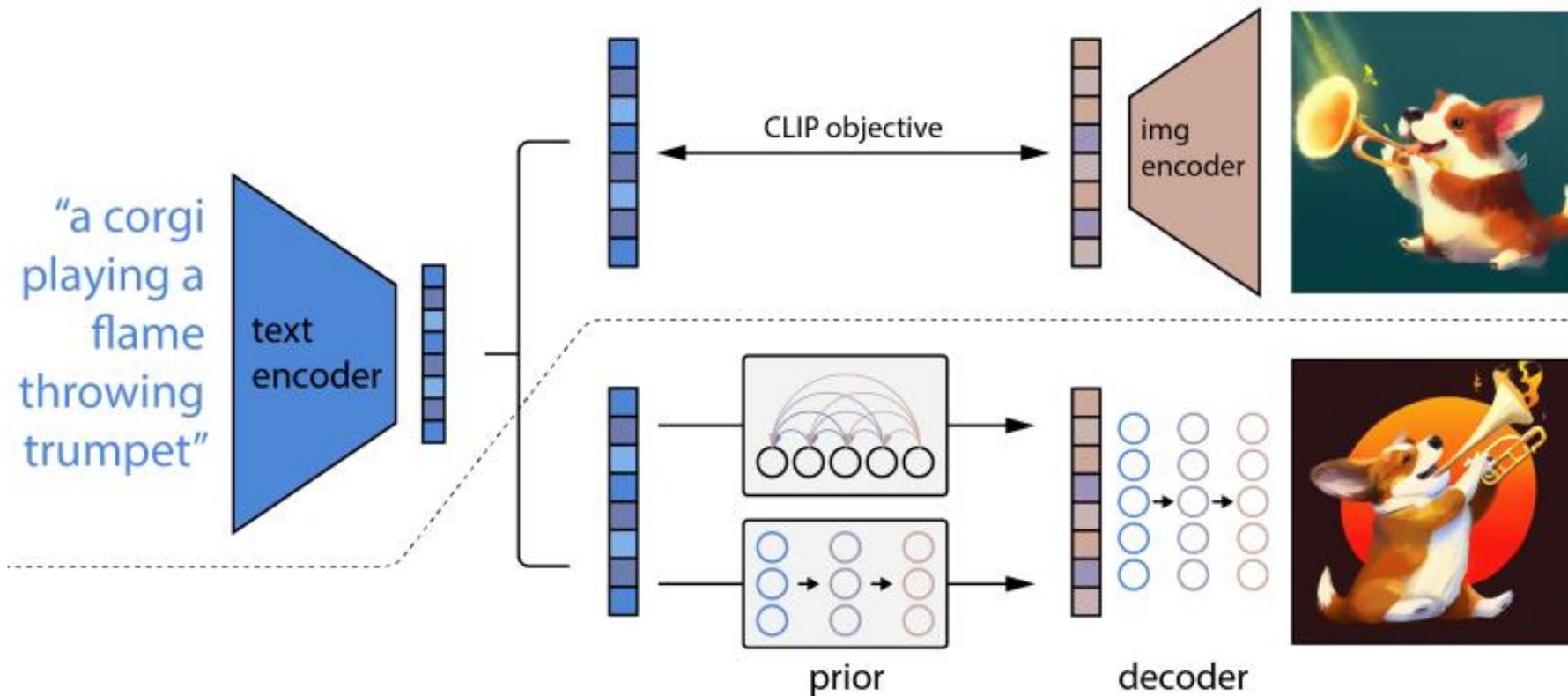
1 #!/usr/bin/env ts-node
2
3 import { fetch } from "fetch-h2";
4
5 // Determine whether the sentiment of text is positive
6 // Use a web service
7 async function isPositive(text: string): Promise<boolean> {
8     const response = await fetch(`http://text-processing.com/api/sentiment/`, {
9         method: "POST",
10        body: `text=${text}`,
11        headers: {
12            "Content-Type": "application/x-www-form-urlencoded",
13        },
14    });
15    const json = await response.json();
16    return json.label === "pos";
17 }
```

Copilot

DALL-E는 뭔데?

- 노이즈 학습을 활용한 Diffusion 모델 차용

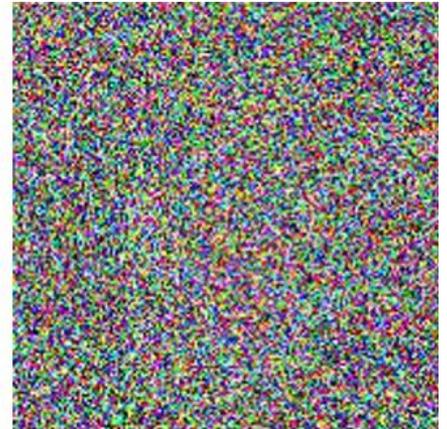
DALL-E 2 구조도



Diffusion 프로세스를 이해해봅시다

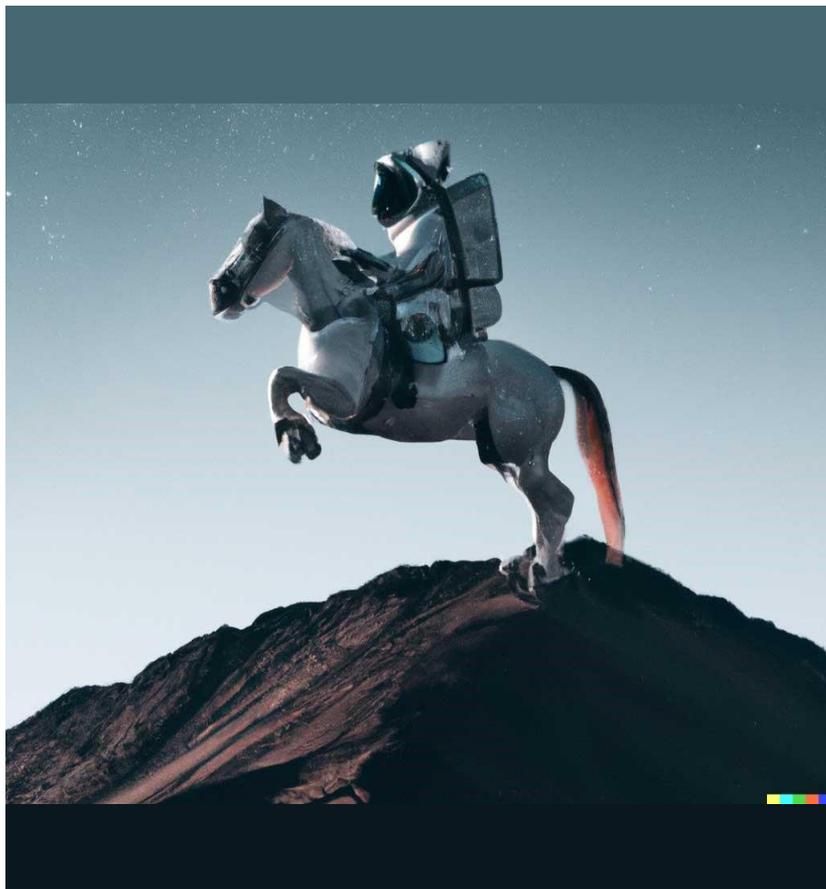
- 데이터에 노이즈를 더하는 과정 학습 → 역으로 데이터에서 노이즈를 제거하며 새로운 것을 생성

고양이 이미지에 노이즈를 더하면?

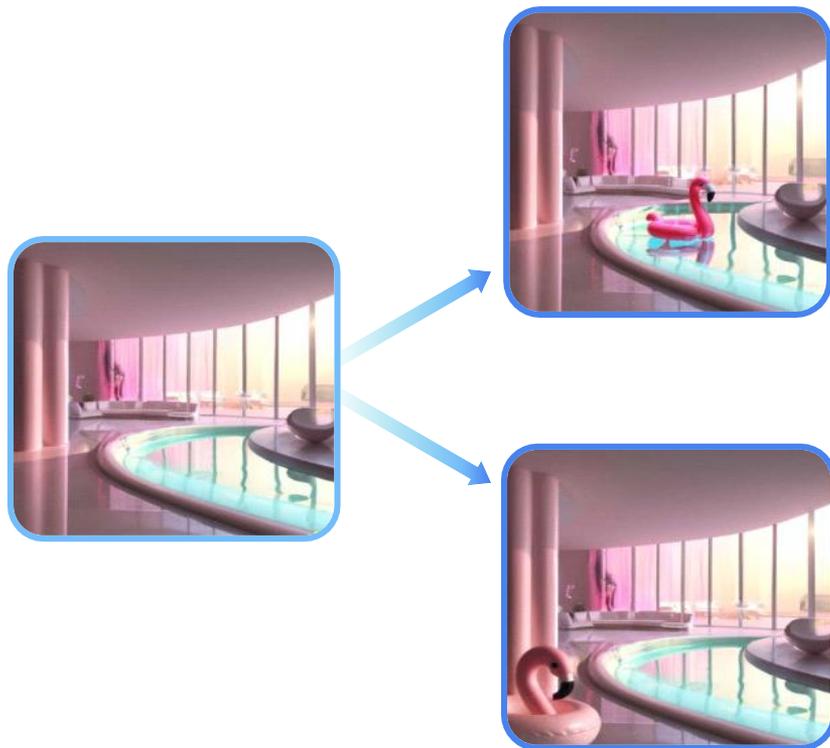


DALL-E의 여러가지 기능들

이미지 생성(Text to Image)

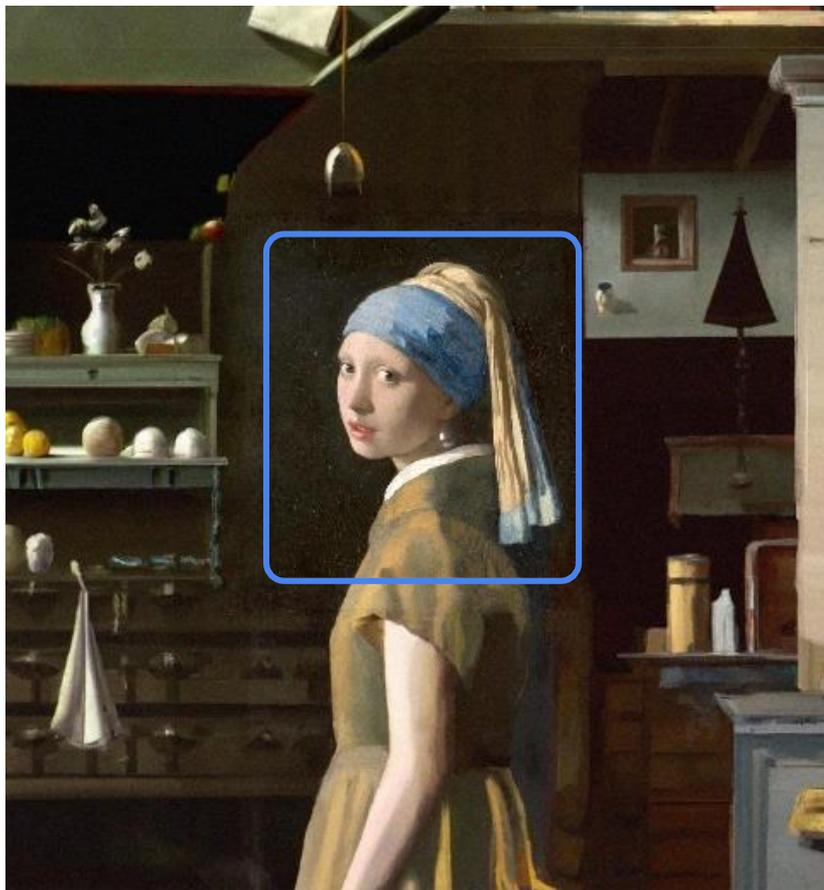


이미지 수정(Inpainting)



DALL-E의 여러가지 기능들

이미지 확장(Outpainting)



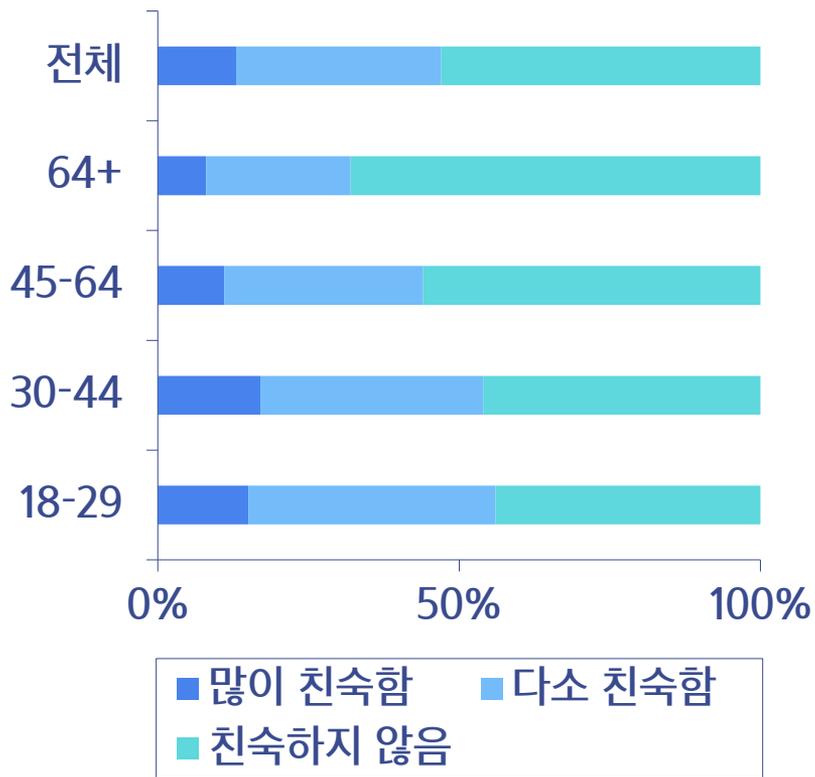
유사 이미지 생성(Variations)



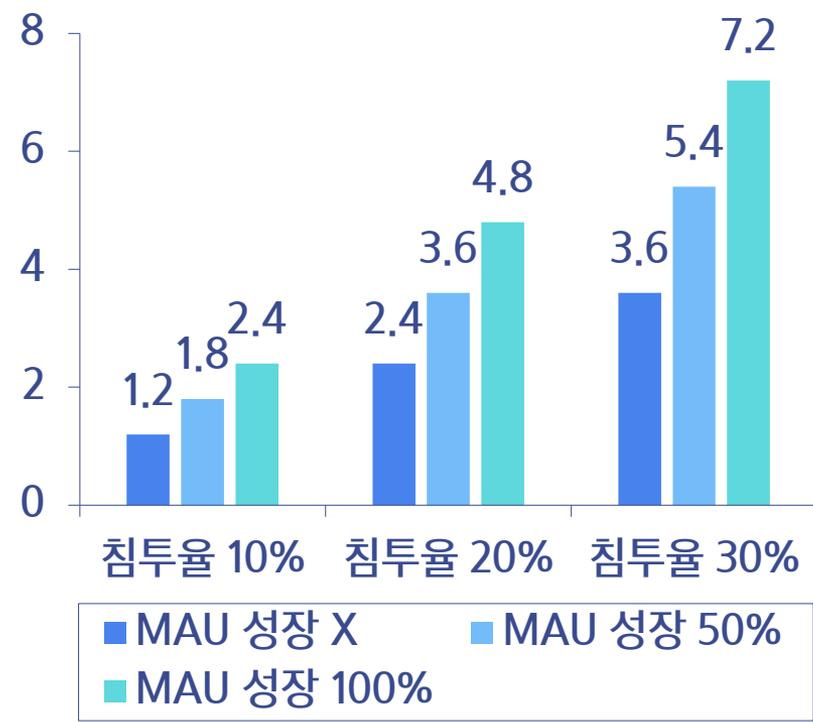
오픈AI가 돈을 벌 수 있을까?

- 오픈AI 자체 목표 매출 23년 2억 달러, 24년 10억 달러

미국 성인의 연령별 챗GPT 친숙도



챗GPT 구독 모델 매출 시나리오 (십억달러)

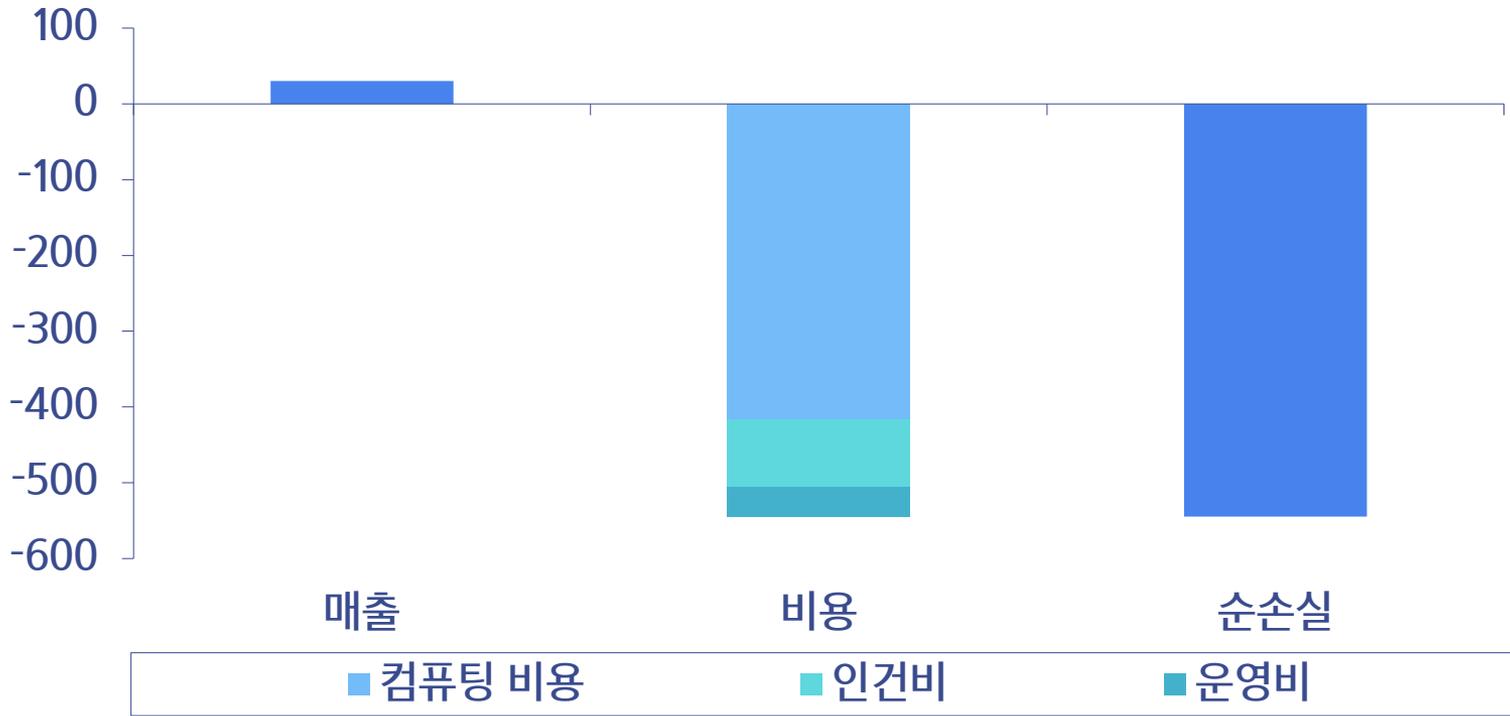


어마어마한 투자의 필요성

- 22년 컴퓨팅 비용 4억 1,650만 달러 + 인건비 8,931만 달러, 운영비 3,875만 달러 = 순손실 5.4억 달러

오픈AI 손익 분석(2022년)

(백만달러)



초거대 AI ≠ 범용 인공지능(AGI)

- 유창한 언어 구사 ≠ 실질적 지능. 범용 인공지능(Artificial General Intelligence)으로 진화

생성 AI 발전과정



Compliance Notice

- 본 조사자료는 당사의 저작물로서 모든 저작권은 당사에게 있습니다.
- 본 조사자료는 당사의 동의 없이 어떠한 경우에도 어떠한 형태로든 복제, 배포, 전송, 변형, 대여할 수 없습니다.
- 본 조사자료에 수록된 내용은 당사 리서치센터가 신뢰할 만한 자료 및 정보로부터 얻어진 것이나, 당사는 그 정확성이나 완전성을 보장할 수 없습니다. 따라서 어떠한 경우에도 본 자료는 고객의 주식투자의 결과에 대한 법적 책임소재에 대한 증빙자료로 사용될 수 없습니다.
- 본 자료에는 외부의 부당한 압력이나 간섭 없이 애널리스트의 의견이 정확히 반영되었습니다.

감사합니다