

01. LLM 훑어보기

1.1 LLM 개념 ~

1.2 LLM 특징과 종류

발표 일시 : 2025.03.11(화)

발표자 : 권선옥

1.1 LLM 개념

LLM : 'Large Language Model' 의 약자
인간의 언어를 처리하는 모델

1.1 LLM 개념

1.1.1 언어 모델

언어 모델 종류

1) 개념 : 컴퓨터에게 어떻게 말을 하고 글을 쓰는지 가르치는 것

2) 언어 모델의 발전사 : 통계적 언어 모델 -> 신경망 언어 모델 -> 트랜스포머

- 통계적 언어 모델

- 통계적 방법 기반(컴퓨터가 문장이나 단어를 얼마나 자연스럽게 표현할지 수학적으로 계산)

- ex. "나는 오늘 점심에" 뒤에 "피자를/샐러드를 먹었다"가 오도록 완성

- 대표적 예시 : 'n-gram'

- 일련의 단어나 문자가 얼마나 자주 함께 등장하는지 계산, 'n'은 연속적으로 고려되는 단어의 수

- 종류 : 1-gram(유니그램), 2-gram(바이그램), 3-gram(트라이그램) 등

- "The cat sat on the mat."

- 1-gram(유니그램) : 'The' , 'cat', 'sat', 'on', 'the', 'mat'

- 2-gram(바이그램) : 'The cat', 'cat sat', 'sat on', 'on the', 'the mat'

- 3-gram(트라이그램) : 'The cat sat', 'cat sat on', 'sat on the', 'on the mat'

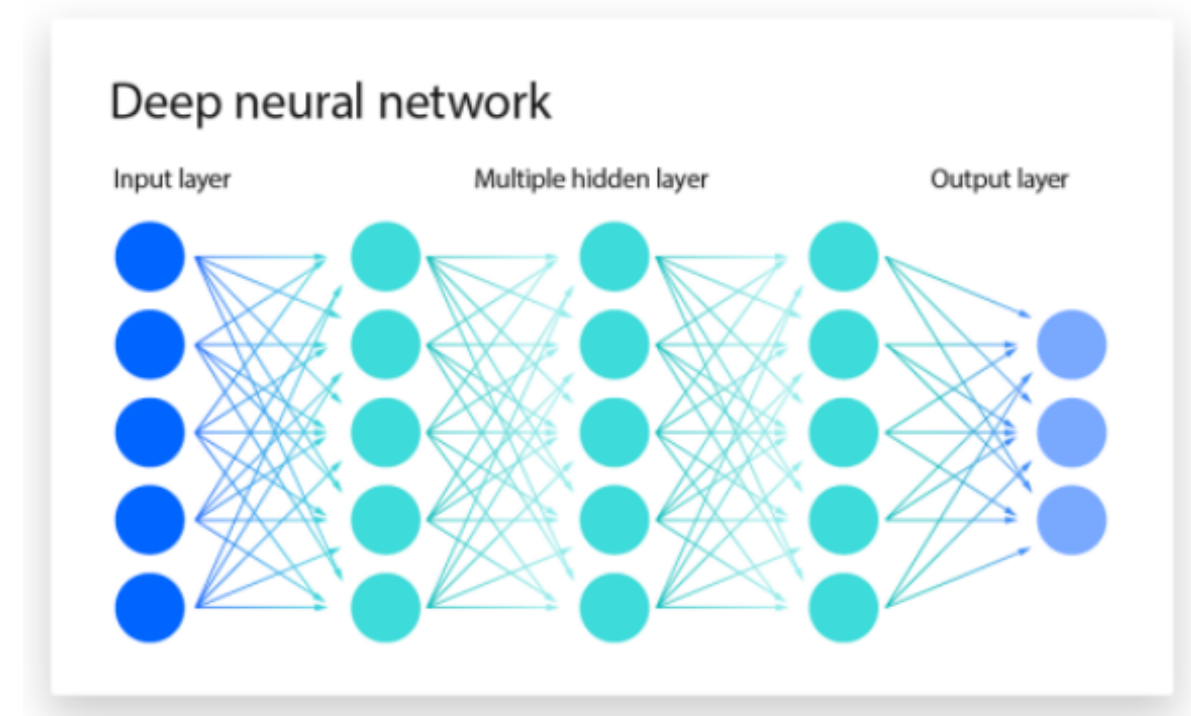
- 특징 : 언어의 통계적 패턴을 학습해서 문장을 이해하고 새로운 문장을 생성하는데 도움

- 한계 : 모든 n-gram을 DB에 저장해야함, 'n'이 커질수록 문맥을 제대로 이해하는데 어려움

1.1 LLM 개념

언어 모델 종류

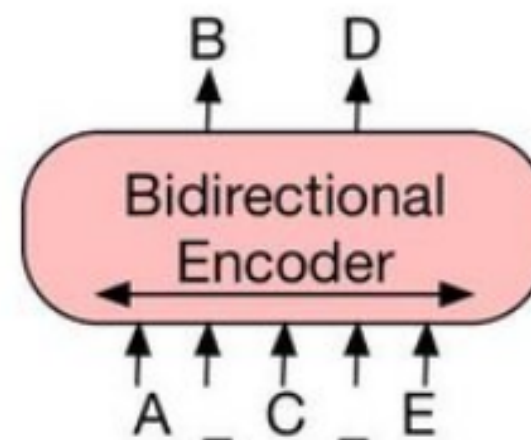
- 신경망 언어 모델
 - 머신러닝의 발전과 함께 등장, 뇌에 있는 신경세포(뉴런)가 서로 연결되어 정보를 처리하는 방식 이용
 - 데이터에서 복잡한 패턴을 학습해 문제를 해결하는 데 활용됨
 - 신경망 구성 : 입력층(input layer), 은닉층(hidden layer), 출력층(output layer)
 - 각 계층에는 여러 뉴런(node)이 서로 연결되어 있음
 - 대표적 예시
 - RNN
 - 시퀀스 데이터(순서가 중요) 처리에 적합
 - 시계열 데이터 분석 → 예측
 - 저장 공간 작음 → 긴 데이터 처리에 한계
 - LSTM
 - RNN 한계 극복을 위해 고안
 - 긴 시퀀스 정보를 기억(장기 기억 메모리)



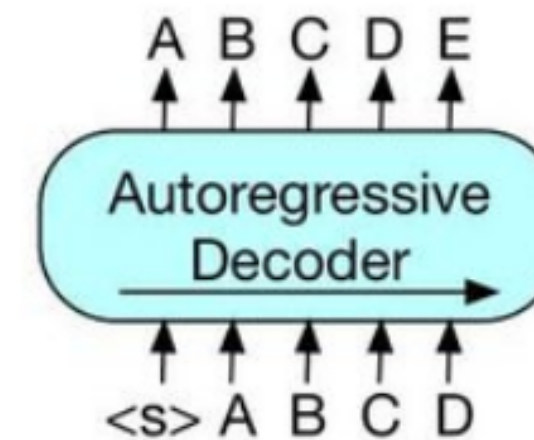
1.1 LLM 개념

언어 모델 종류

- 트랜스포머(Transformer)
 - 2017 구글 브레인 <Attention Is All You Need> 에서 도입
 - 기존 모델은 각 단어를 개별적으로 이해,처리 -> Transformer는 문장과 단락 전체 처리 가능(활용 범위 확대)
 - 대표적 예시
 - BERT : Google에 의해 개발된 텍스트를 양방향으로 맥락을 이해하는 언어 모델(Encoder)
 - GPT : OpenAI에 의해 개발된 인공지능 언어 모델, 다른 언어 모델에 비해 성능 높음(Decoder)
 - 현재 GPT-4.5 모델까지 출시된 상태



(a) BERT: Random tokens are replaced with masks, and the document is encoded bidirectionally. Missing tokens are predicted independently, so BERT cannot easily be used for generation.



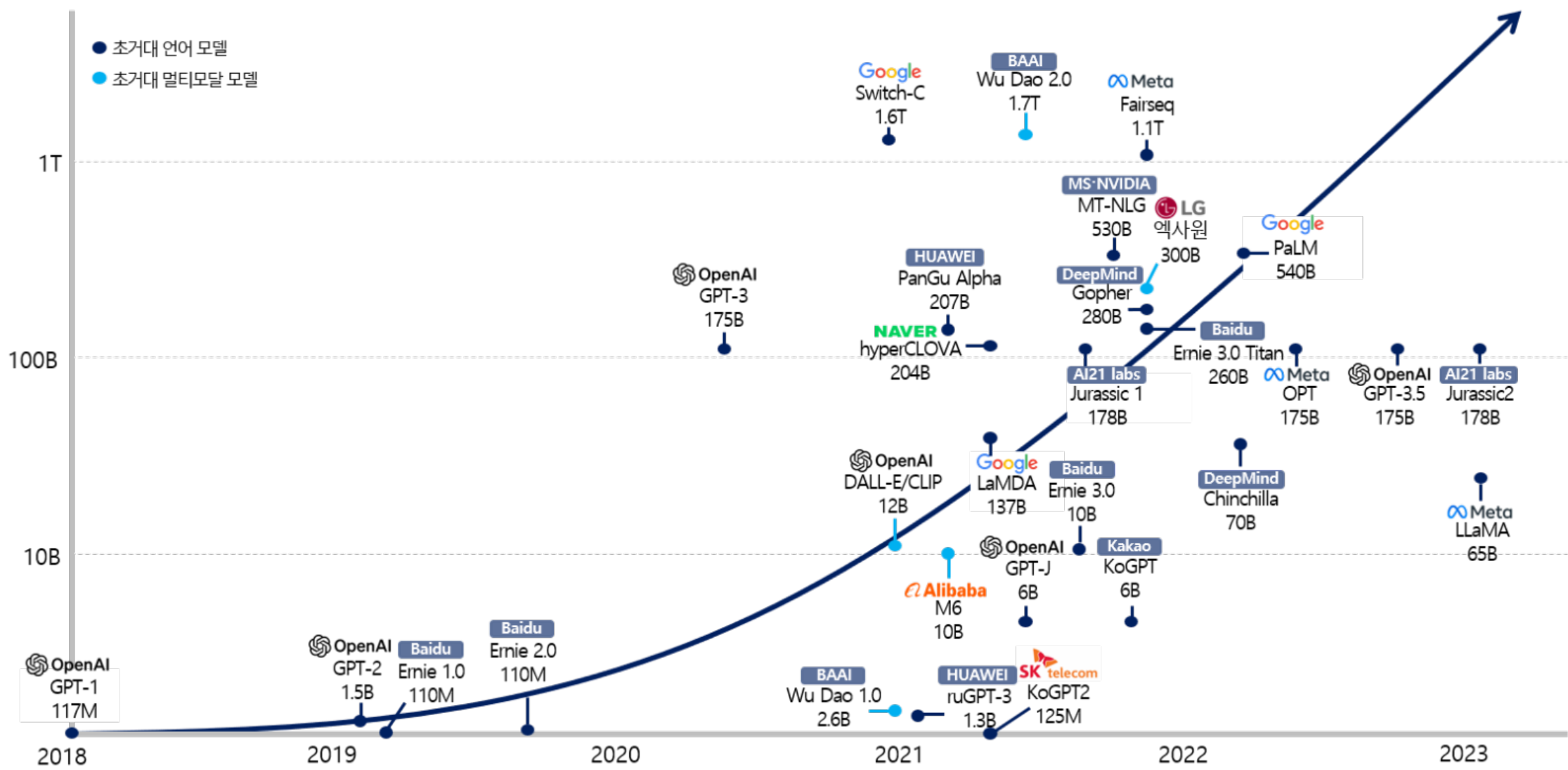
(b) GPT: Tokens are predicted auto-regressively, meaning GPT can be used for generation. However words can only condition on leftward context, so it cannot learn bidirectional interactions.

1.1 LLM 개념

1.1.2 거대 언어 모델

- 거대 언어 모델(LLM) : 대규모 데이터로 훈련된, 매우 큰 규모의 인공 지능 기반 언어 모델
 - 대규모 데이터 = GPT-3 모델 학습 시 약 45TB의 텍스트 데이터로 학습됨(≒ 고화질 비디오 수천 시간)
 - 모델의 크기 = 파라미터 수(AI 모델의 내부 변수), 모델이 진화할수록 파라미터 수도 증가

그림 - 초거대 AI 개발 동향(파라미터 수)

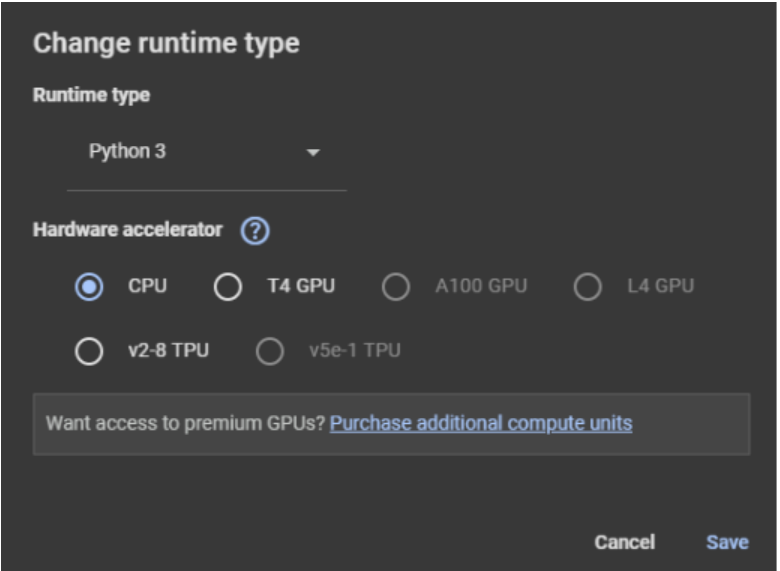


출처 : KDI 경제정보센터
자료연구팀 2023년 02호

1.2 LLM 특징, 종류

1.2.1 LLM 특징

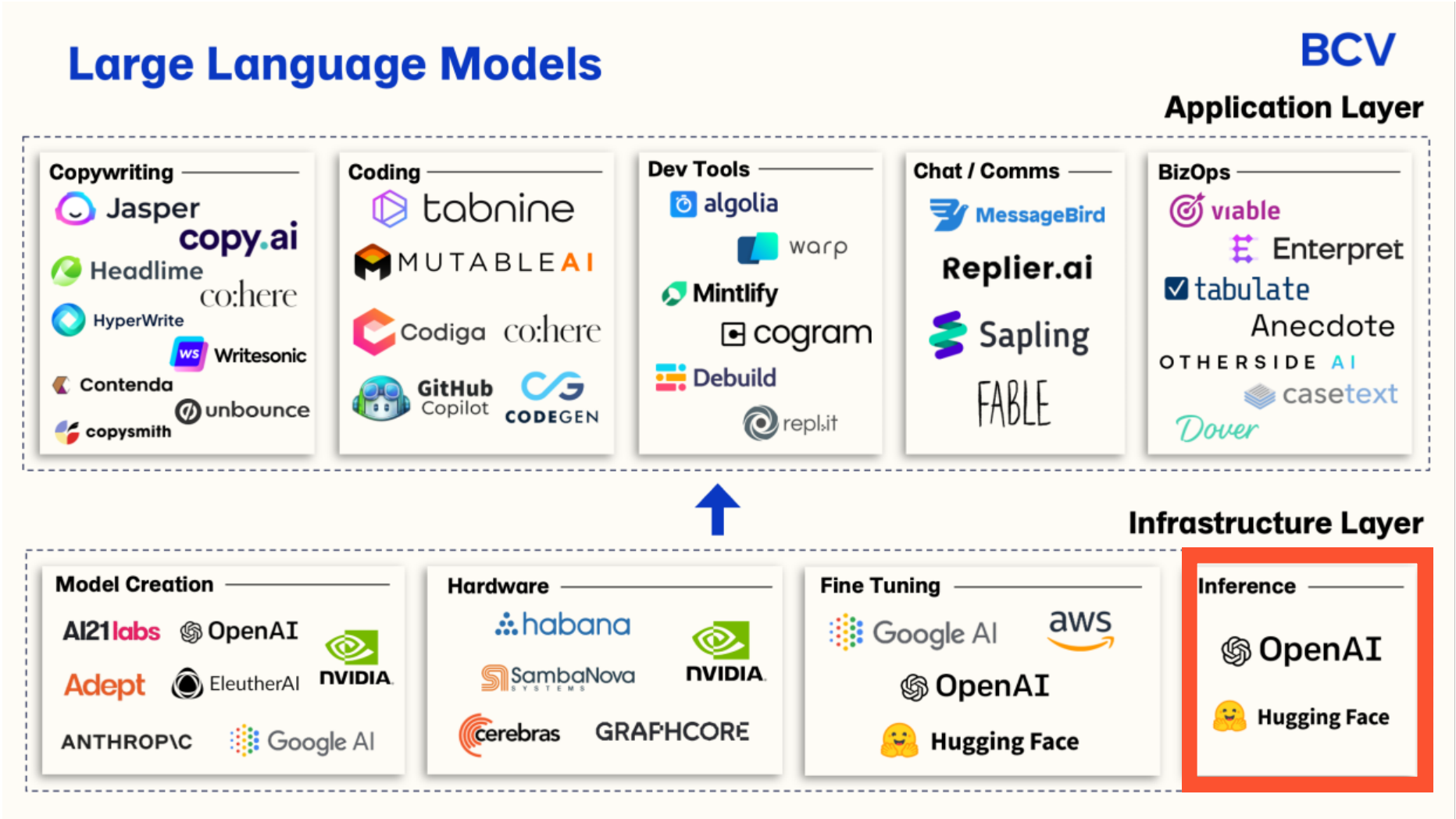
- 방대한 양의 텍스트 데이터로부터 학습됨
- 언어를 이해하고 생성하는데 특화됨(이해, 생성)
- 특정 작업을 위해 파인튜닝 가능
 - 파인튜닝 : 챗GPT와 같은 언어모델을 특정 데이터로 추가 학습을 시키는 과정
(특화된 분야에 더 정교하게 사용 가능)
- LLM 훈련을 위해 상당한 컴퓨팅 자원 필요
 - 컴퓨팅 자원 : GPU, TPU 같은 하드웨어
 - GPU - 대규모 데이터 학습, 복잡한 수학적 연산 빠르게 처리
 - TPU - Google이 개발한 머신러닝 및 딥러닝 작업에 최적화된 하드웨어(Google Colab에서 사용 가능)



1.2 LLM 특징, 종류

1.2.2 LLM 종류

- LLM 생태계
 - Application Layer - Framework*, BI** 툴 제공 회사 중심
 - *앱 개발 시 기본적인 것을 제공하는 소프트웨어 도구 모음
 - ** 데이터 분석 및 정보 제공을 위한 리포팅 툴
 - Infrastructure Layer - 하드웨어, 모델 관련 회사(ex. NVIDIA)

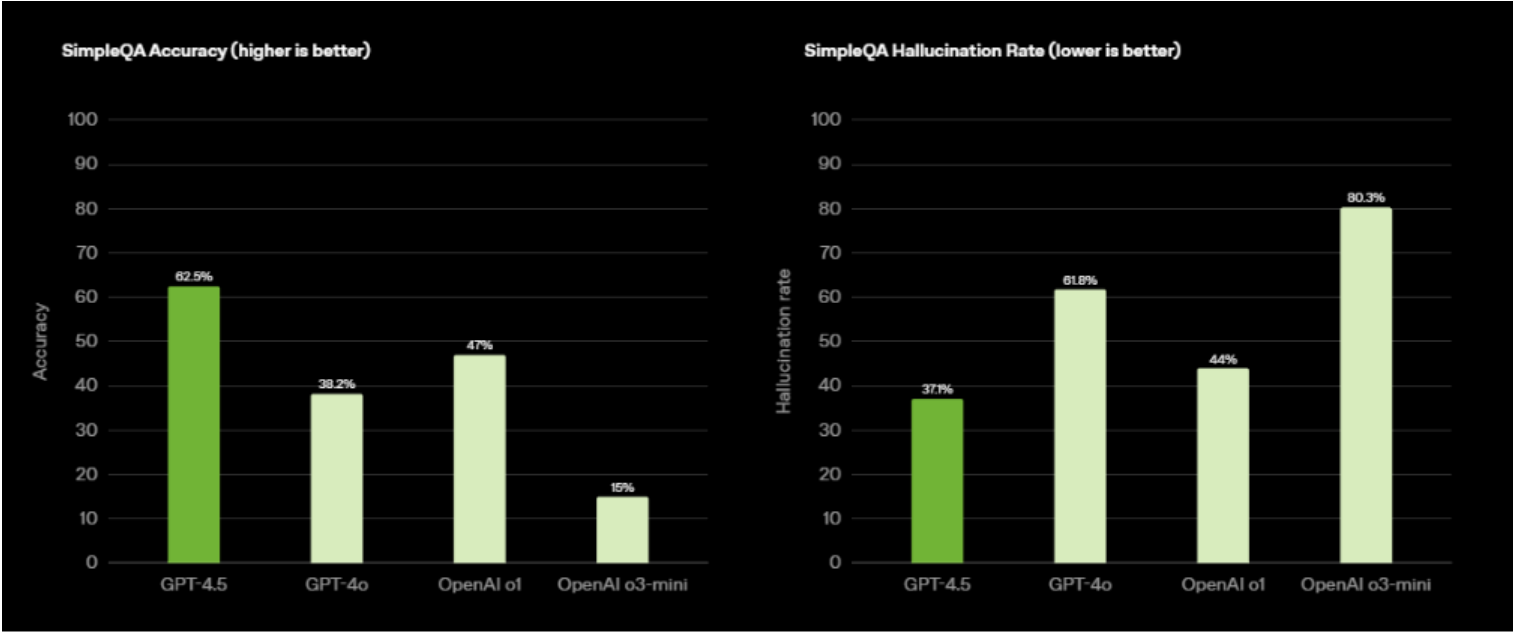


출처 : Labellerr blog

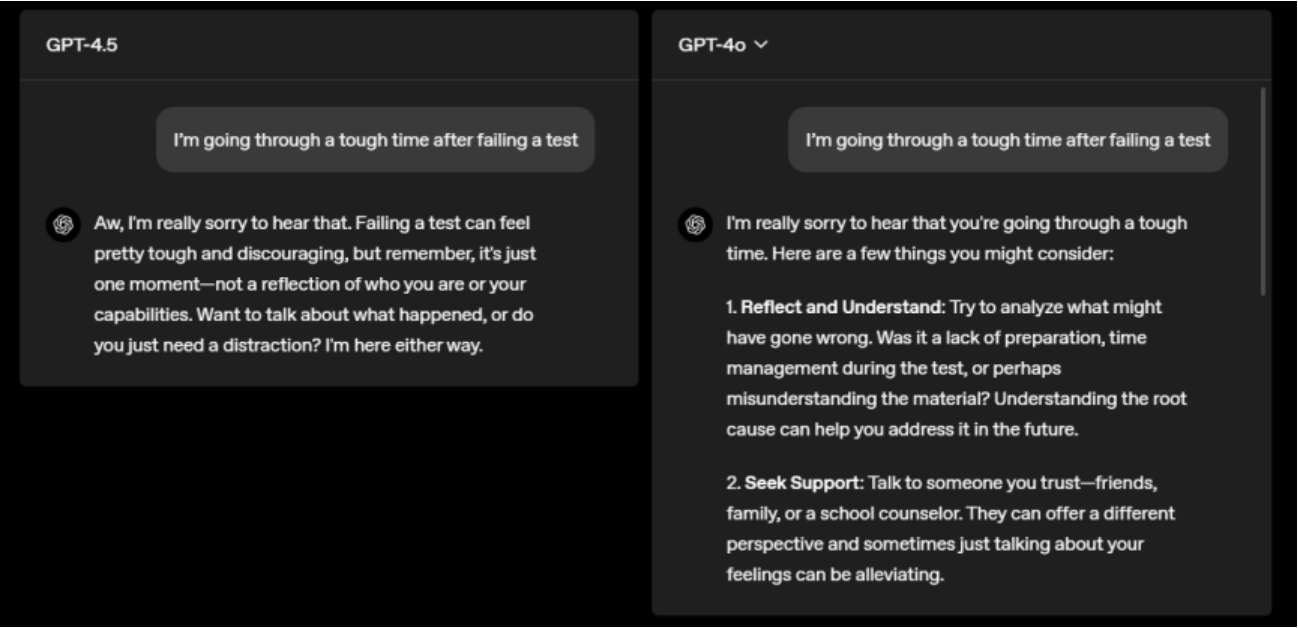
1.2 LLM 특징, 종류

1.2.2 LLM 종류

- GPT-4.5 참고 : OpenAI 공식 발표 자료
 - OpenAI 가 가장 최근에 발표한 모델(2025-02-27 발표)
 - multimodal* 기능 없음
 - *텍스트, 이미지, 오디오, 비디오 등 다양한 유형의 데이터(모달리티)를 함께 고려)
- 특징
 - GPT-4.5는 비지도 학습 확장으로 환각(잘못된 정보 생성)이 줄어들고 신뢰성이 향상됨
 - GPT-4.5는 감성 지능(EQ)을 갖추고 있다.



Accuracy 향상, Hallucination 감소

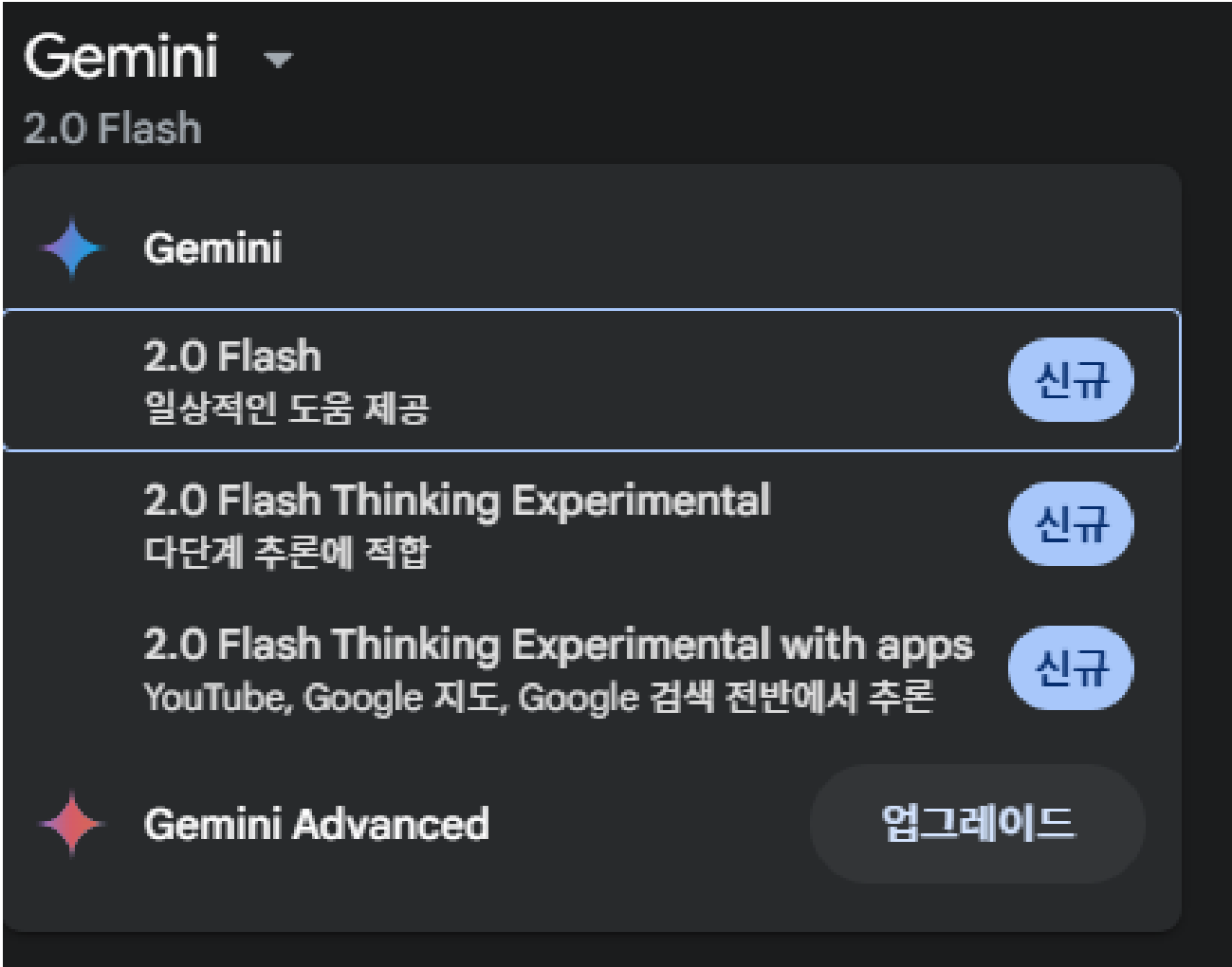


감성 지능(EQ)

1.2 LLM 특징, 종류

1.2.2 LLM 종류

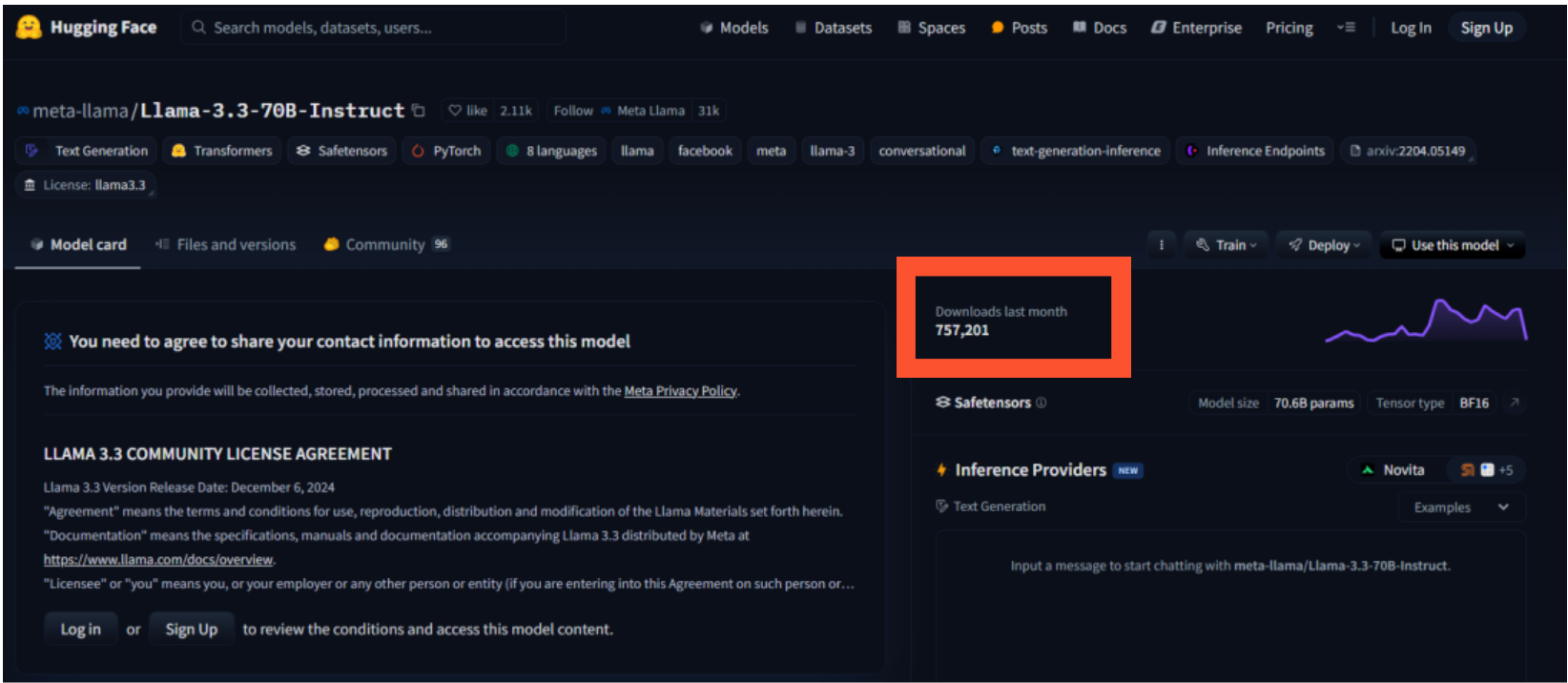
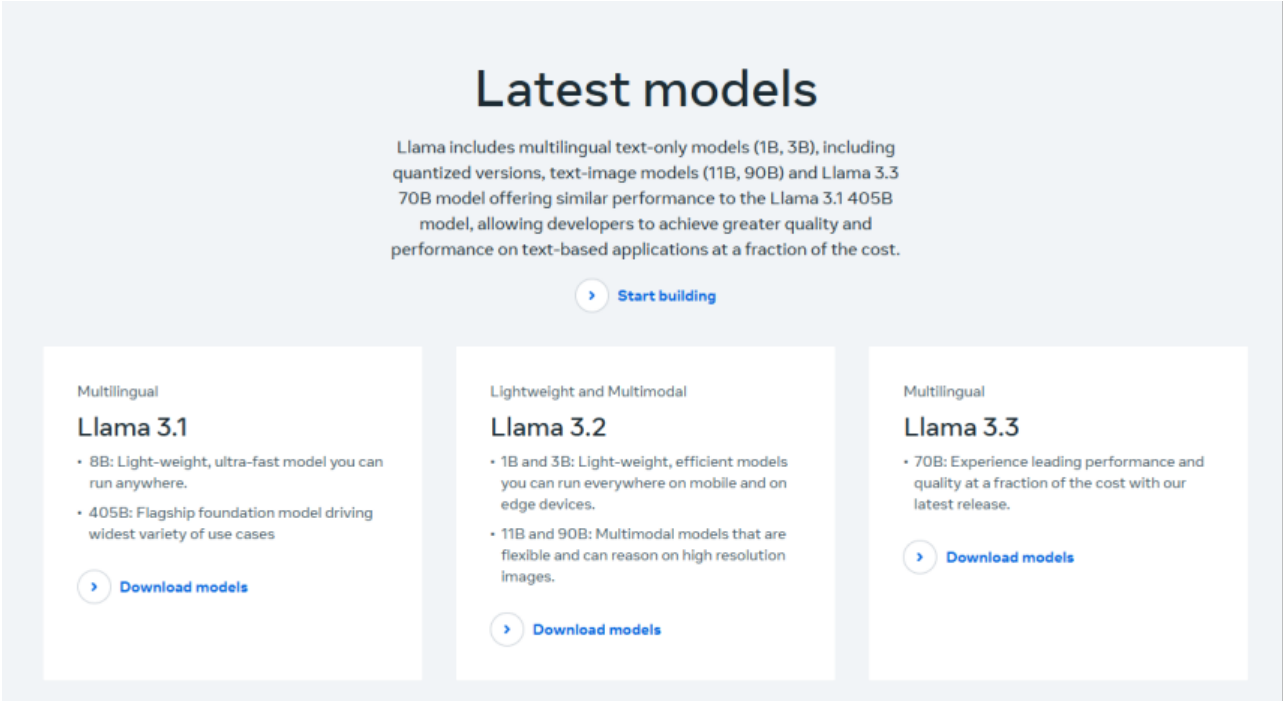
- Gemini 2.0(제미나이) 참고 : [Google 공식 발표 자료](#)
 - Google이 가장 최근에 발표한 모델(2025-02-05 발표)
 - 개발자들이 프로덕션 애플리케이션을 구축할 수 있도록 지원
 - 멀티모달 입력을 지원하며, 텍스트 및 이미지 생성, 음성 변환 기능도 추가될 예정
 - 뛰어난 성능과 빠른 응답 속도



1.2 LLM 특징, 종류

1.2.2 LLM 종류

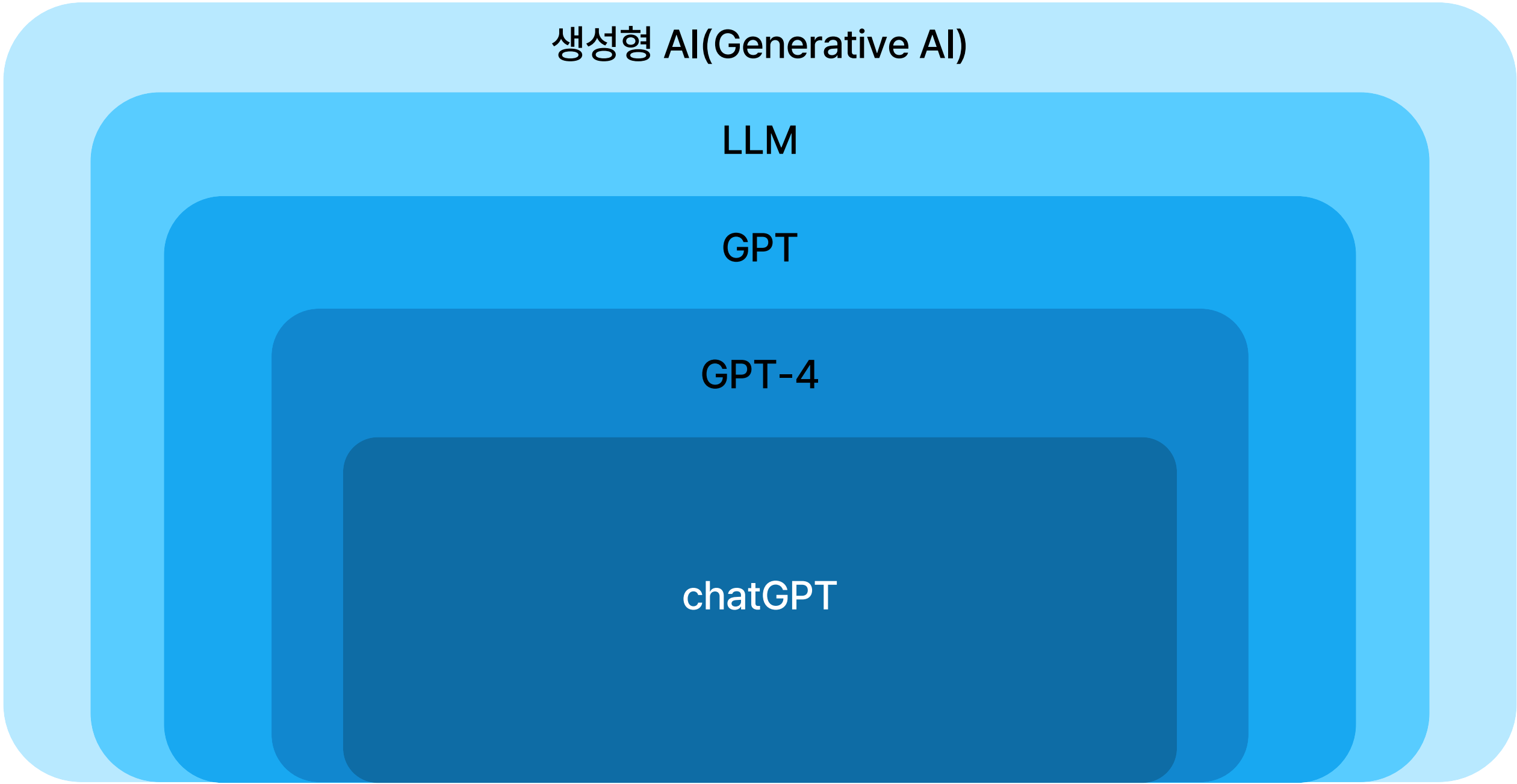
- Llama 3.3(라마) 참고 : Meta 공식 발표 자료
 - Meta에서 가장 최근에 발표한 모델(2024-12-06 발표)
 - Llama 3.1 405B 모델과 유사한 성능을 제공(더 낮은 비용으로 구현)
 - 다국어 대화 사용 사례에 최적화
 - Hugging-Face에 공개되었으며 약 75만 회 다운로드 됨(발표일 기준)



1.2 LLM 특징, 종류

1.2.3 LLM과 GAI,
SLM

- 생성형 AI 의 범위
 - $LLM \subset GAI$



1.2 LLM 특징, 종류

1.2.3 LLM과
GAI, SLM

● GAI vs. LLM

구분	GAI	LLM
콘텐츠 생성 범위	텍스트, 이미지, 음성, 코드 등 다양	텍스트에 국한
학습	대규모 데이터 의존, 이미지, 오디오 등 다양한 데이터 유형 학습	인터넷, 서적 및 기타 광범위한 텍스트 학습
출력	음악, 예술 작품 등 광범위한 출력	사용자 질문(prompt) 기반 텍스트 답변(completion) 생성
신경망(혹은 모델)	GAN, RNN 등 다양한 신경망 사용	RNN, Transformer
활용 분야	여러 창의적 분야에 사용 가능	언어 관련 작업에 특화

1.2 LLM 특징, 종류

1.2.3 LLM과
GAI, SLM

- SLM(Small Language Model) vs. LLM

구분	SLM	LLM
모델 크기 (파라미터 수)	1,500만개 미만	수천억 개
컴퓨팅	모바일 디바이스에서도 동작	수백 개의 GPU 필요
성능	단순 작업만 동작 (기후 데이터 분석 후 날씨 예측 등 특화된 언어 작업)	복잡한 작업 처리 가능 (자연어 처리, 텍스트 생성 등 다양한 언어 작업)
배포	쉬움	어려움 (상당한 인프라 필요 → 대체로 클라우드 사용)
학습	약 7일	수개월

핵심 요약

- LLM(Large Language Model) 은 인간의 언어를 처리하는 모델이다.
- 언어모델의 종류는 3가지로, 통계적 언어모델(n-gram), 신경망 언어모델(RNN, LSTM), Transformer(BERT, GPT) 가 있다.
- LLM은 방대한 양의 텍스트 데이터로부터 학습되고 파인튜닝(추가 학습)이 가능하며, 훈련을 위해 GPU 등 컴퓨팅 자원이 필요하다.
- LLM은 생성형 AI(GAI)에 포함되며 텍스트 학습에 특화되었다.
- LLM은 SLM에 비해 모델 크기(파라미터 수)가 수천억개로 매우 크기 때문에 다양한 언어 작업같은 복잡한 작업 처리에 유용하다.