



AI Brief 특집호

GPT-4 개요 및 특징

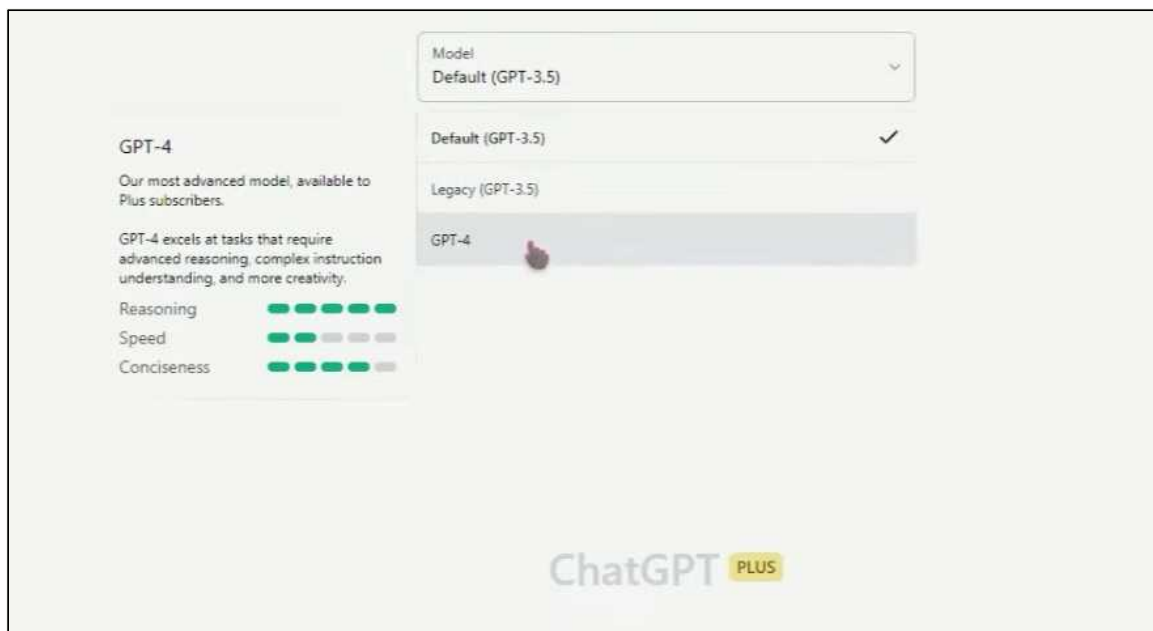
2023년 03월

KEY Contents

- OpenAI는 현재 가장 진보된 형태의 창의적이고, 지식 및 문제해결력을 갖춘 GPT-4를 출시(우리시간 `23.3.15.새벽)
- GPT-4는 기존 ChatGPT 대비 인식 및 추론능력이 향상되었으며, 이미지를 입력으로 인식하는 멀티모달 기능도 탑재
- GPT-4는 다방면에서 우수한 성능을 보였으나, 여전히 근본적인 한계점을 보유하고 있으며, 유료 서비스로 구체적인 스펙은 비공개

1. GPT-4 개요**□ OpenAI는 ChatGPT를 넘어서는 GPT-4를 출시(현지시간 '23.3.14.)**

- ▶ GPT-4는 현재 가장 창의적이며, 일반지식과 문제해결력을 갖춘 모델
 - 기존 모델(ChatGPT, GPT-3.5)대비 인식 및 추론능력의 향상, 이미지를 입력으로 인식하는 멀티모달*, 대화 메모리 향상 등의 기능을 탑재
- * MultiModal: 텍스트뿐만 아니라, 이미지, 음성을 비롯한 여러 형태의 입력을 인식하고 학습 및 추론 가능한 형태
- ▶ GPT-4는 ChatGPT PLUS(유료, \$20/월)로 업그레이드 하면 사용 가능
 - 기존 ChatGPT에서 유료 서비스인 ChatGPT PLUS로 업그레이드 하고, 화면의 구동엔진(모델)을 GPT-4로 선택하여 활용 가능

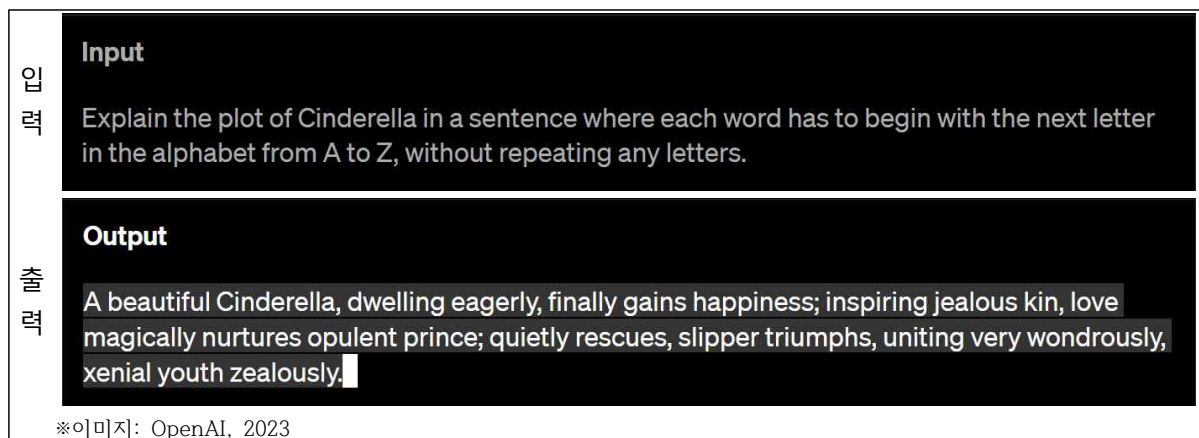


[그림 1] GPT-4 활용 방법 (OpenAI, 2023)

- ▶ GPT-4는 MS의 클라우드인 Azure 슈퍼컴퓨터에서 학습되었으며, 추후, MS의 Azure에서 최적화된 AI 인프라 서비스로도 제공할 예정

2. GPT-4의 특징 및 ChatGPT(GPT-3.5)와 차이점

- [창의력 향상] 요구한 질의에 대해 문제를 보다 창의적이고 정확하게 풀 수 있으며, 시나리오 작성 및 작문 등의 능력이 크게 향상
- ▶ (예시) 동화 신데렐라의 줄거리를 알파벳 A에서 Z까지 순서의 단어로 작성



[그림 2] GPT-4 작문 예시

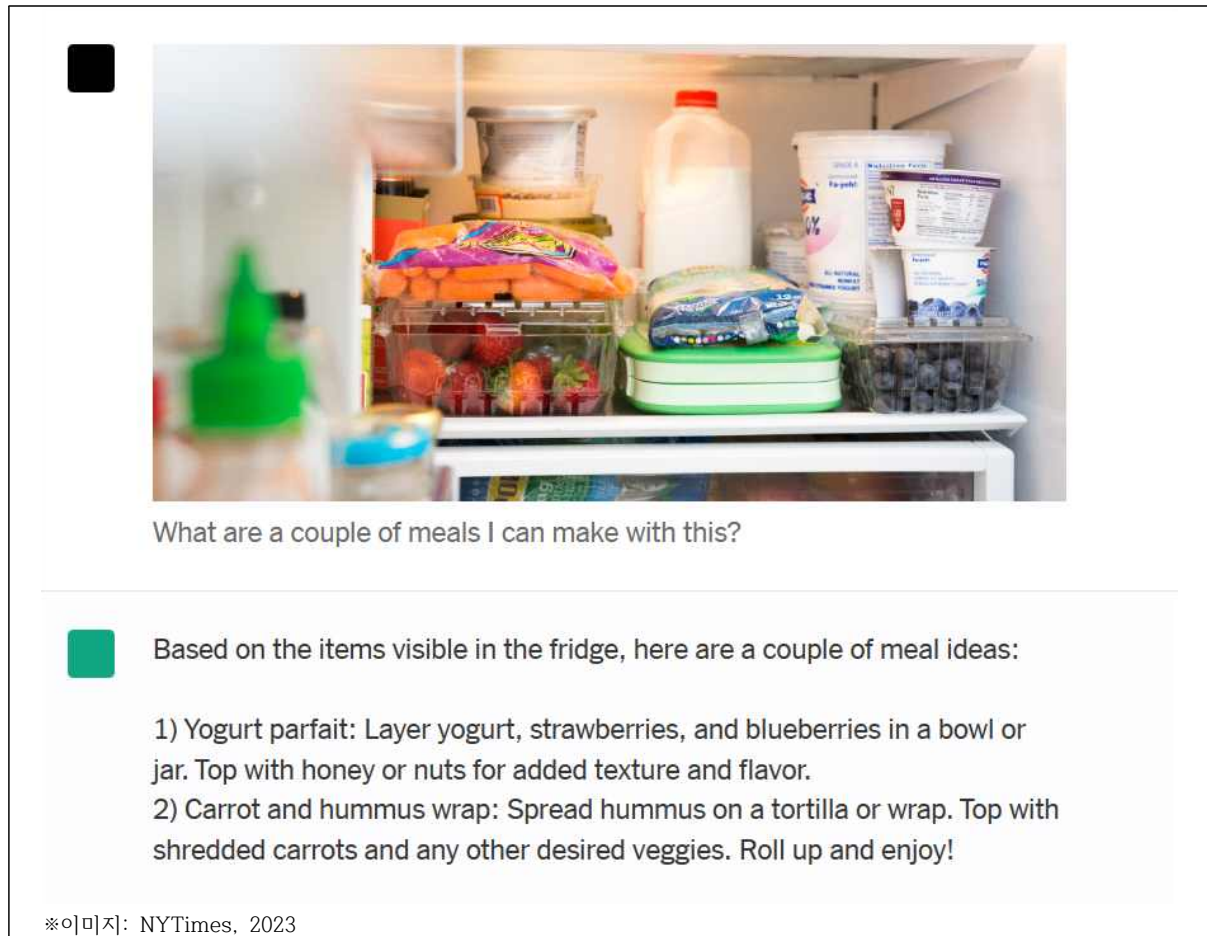
- [추론능력 향상] 고급 추론 기능에서 기존 ChatGPT의 성능을 능가하여, 여러 개의 사례들 사이에서 공통점을 추출하는 능력이 보다 진화
- ▶ (예시) 3명의 유희시간을 파악하여 미팅 시간을 추론하는 답변 비교

<표 1> GPT-4 추론 답변 비교 예시

구분	ChatGPT 답변	GPT-4 답변
답변	다음은 이용 가능 시간 목록입니다. - 앤드류: 오전11시부터 오후3시까지 - 조앤: 정오~오후2시, 오후3시30분~오후5시 - 한나: 정오~오후12시30분, 오후4시~오후6시 이용 가능 여부에 따라 세 사람 모두 이용 가능한 시간은 오후4시부터 오후4시30분까지 30분간입니다. 따라서 오후4시에 미팅을 예약할 수 있습니다.	앤드류: 오전11시~오후3시 조앤: 오후12시~오후2시, 오후3시30분~오후5시 한나: 오후12시~오후12시30분, 오후4시~오후6시 30분 회의가 가능한 시간: 오후12시~오후12시30분
비고	오답, 앤드류는 4시가 가능하지 않음	답변이 정확해짐

□ [멀티모달] 기존 텍스트로 국한된 입·출력에서 이미지를 입력으로 받아 추론할 수 있는 기능이 탑재

- ▶ (예시) 냉장고 내부의 음식을 촬영한 사진으로 가능한 요리를 추천



[그림 3] GPT-4 이미지 인식 및 추론

- 이 기능은 이미지 속 객체를 식별(인식), 객체 간의 관계 파악, 이를 바탕으로 추론, 적절한 단어를 조합하여 답을 만들어 내는 종합적 스킬이 필요

□ [대화 메모리 향상] 기존 대비 더 많은 입력(Input)을 수용 가능

- ▶ 기존 ChatGPT는 입력으로 약 8천 단어(4,096 토큰) 수준의 입력만 ‘기억’(추적)했던 반면, GPT-4는 약 6만4천 단어*의 입력을 저장 가능

* A4 약 50페이지 수준으로 희곡 시나리오나 단편소설 정도의 분량을 입력-저장 가능하며 토큰(말뭉치) 수로는 32,768개 (한 번에 프롬프트로 입력할 수 있는 단어는 ChatGPT 3천 개, GPT-4 2만5천 개)

- 기존에는 사용자와의 대화가 길어짐에 따라 ‘멀리 지나간’ 대화를 추적하지 못했던 반면, GPT-4는 이전 대화의 ‘기억’을 더 오래지속하며, 대화의 맥락을 유지

□ [신뢰성 향상] GPT-4는 지난 2년간 악성 질의에 대한 대응 훈련을 통해 신뢰성을 향상시켜 보다 윤리적인 측면의 답변을 하도록 진화

- ▶ 오늘날의 챗봇은 사용자의 악의적인 프롬프트에 의해 쉽게 오염(편향, 거짓 정보 학습 등)* 될 수 있으나, GPT-4는 기존 대비 면역력 강화

* 국내 AI 스타트업 스캐터랩의 챗봇 '이루다'는 답변에 성희롱혐오 등 윤리적 문제가 있어서 서비스 중단('21.1.)

□ [조종 가능성] 사용자와의 대화중에 말투나 어조 등 답변 스타일을 필요에 따라 더 적합한 형태로 변경하는 기능이 강화

- ▶ 고정적인 말투나 인터뷰에 응하는 사람과 같은 말투도 구현이 가능하며, 농담에 대한 이해 및 발현도 기존 대비 더 우수

GPT-3.5	GPT-4
 Tell me a novel joke about the singer Madonna.  Why did Madonna go to the bank? To get a Material Loan!	 Tell me a novel joke about the singer Madonna.  Why did Madonna study geometry? Because she wanted to learn how to strike a pose in every angle! 📐👤

※이미지: NYTimes, 2023

입력) 가수 마돈나에 대한 새로운 농담을 들려주기 바람.

답변) 마돈나는 왜 은행에 갔을까?
- 물질적 대출을 받으려고!

답변) 마돈나는 왜 기하학을 공부할까?
- 모든 각도에서 포즈를 취하는 방법을 배우려고! (삼각자,무늬,마이크 이모티콘)

해석) GPT-4는 가수 마돈나에 대한 이해를 바탕으로 농담 발현

[그림 4] ChatGPT(GPT3.5) 및 GPT-4의 농담 사례

□ [다국어 지원] 영어 및 한글을 포함한 총 26개국 언어를 지원하며, 높은 정확도로 답변할 수 있는 능력 보유

- ▶ 한글에 대한 지원도 우수하나, 학습 데이터가 처음부터 한국어 데이터가 아닌 영어를 기반으로 하며, 번역을 통한 학습 및 응답을 수행

□ [시험 능력] 표준화된 시험에서 기존 대비 우수한 성적을 달성하며, 진보된 인지 및 추론 능력을 보유

- ▶ GPT-4는 미국 변호사자격 시험을 상위 10% 성적으로 통과하며, 기존 ChatGPT(하위 10%) 대비 우수함을 입증
- ▶ 생물학 올림피아드에서는 GPT-4가 상위 1%의 성적을 거두어, 기존(ChatGPT 하위 31%)대비 우수한 성능을 보임

3. GPT-4의 활용(3rd Party)

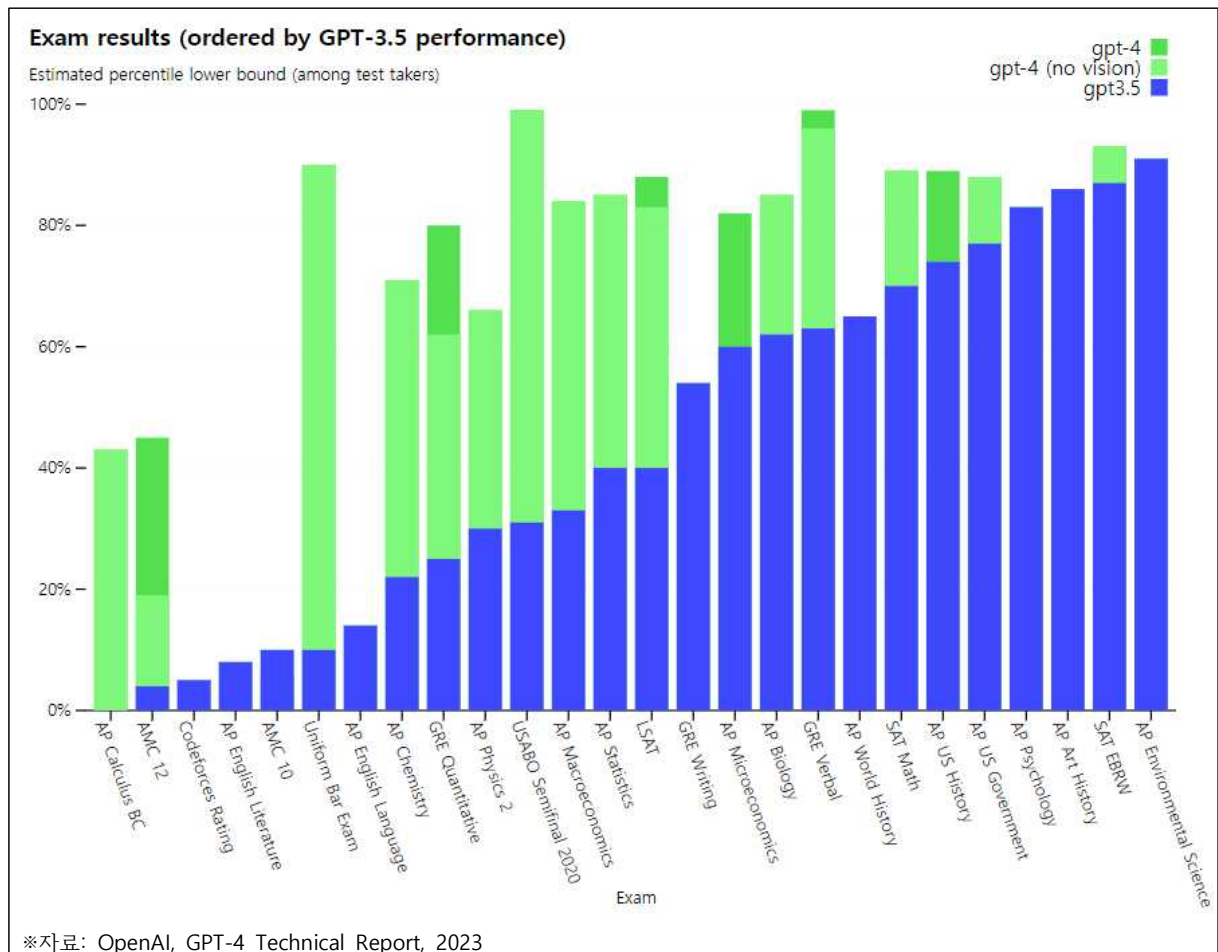
□ GPT-4를 활용하는 협력업체는 학습 플랫폼(듀오링고, 칸아카데미), 장애도우미(비마이아이즈), 금융(모건스탠리, 스트라이프), 정부(아이슬란드) 등 다양('23.3.14.자 발표)

- ▶ (Duolingo) 언어 학습 플랫폼으로, GPT-4를 영어테스트를 실행하는 프로그램에 활용하여 대화연습, 실수에 대한 상황별 피드백 등 제공
- ▶ (KhanAcademy) 교육 플랫폼에서 GPT-4를 활용한 가상 튜터 및 교사를 위한 교실 도우미(Khanmigo)를 구동할 예정
- ▶ (BeMyEyes) 맹인 저시력 커뮤니티를 위한 덴마크 기업으로, GPT-4 기반 인간 자원봉사자 수준의 컨텍스트 이해 플랫폼(Virtual Volunteer)을 개발
- ▶ (MorganStanley) 자산관리 및 투자전략, 시장조사 및 분석 등에 GPT-4 (기존 GPT-3)를 도입하여 콘텐츠 검색 및 분석을 수행
- ▶ (Stripe) e커머스 분야에서 소규모 및 대규모 비즈니스 결제 지원에 GPT-4를 도입, 정보의 자동 수집 및 요약을 바탕으로 서비스 품질 향상
- ▶ (Iceland Gov.) 아이슬란드 정부는 자국의 언어(고대 노르웨이어에 가까운)를 보존하기 위해, GPT-4 학습을 통한 정확한 아이슬란드어 번역 등에 활용

4. 경과 및 한계점

□ [FineTuning] 개발진들은 GPT-4를 파인튜닝하는데에 6개월의 시간을 소요했으며, 결과적으로 기존 대비 82% 더 (윤리적으로)안전하고, 40% 더 정확

- ▶ **(사람을 통한 피드백)** ChatGPT 사용자들이 제출한 피드백 및 50명 이상의 전문가 집단을 통해 AI 및 보안 영역의 피드백을 반영
- ▶ **(지속 개선)** ChatGPT의 실제 활용을 통해 얻은 결과 및 데이터를 GPT-4의 안전연구 및 모니터링에 적용했으며, 지속 업데이트를 진행 중
- ▶ **(개선 작업에 자체활용)** GPT-4의 고급 추론 및 지침 준수 기능을 기반으로 GPT-4의 미세조정을 위한 훈련데이터 생성·평가·모니터링에 활용



[그림 5] GPT-4의 성능 우수성

- **[한계점]** GPT-4는 여전히 사회적 편견, 환각(Hallucination), 악의적인 프롬프트에 대해 한계점을 보임(OpenAI 개발진이 홈페이지에 언급)
- ▶ 기존 ChatGPT 대비 입력값의 다양화(음성, 이미지 등) 및 성능 향상이 있으나, 미흡한 추론 및 부정확한(최신정보 결여) 답변을 생성

GPT-3.5	GPT-4
Imagine an infinitely wide doorway. What is more likely to fit through it, a military tank or a car? Since the doorway is infinitely wide, both the military tank and the car can fit through it without any issues. The width of the doorway is not a limiting factor, so any object of any size, no matter how big or small, can pass through it. However, the height and length of the objects could potentially be a limiting factor if the doorway has a finite height or length.	Imagine an infinitely wide doorway. What is more likely to fit through it, a military tank or a car? Both a military tank and a car would fit through an infinitely wide doorway since 'infinitely wide' means that the doorway has no limits to its width.

※이미지: NYTimes, 2023

입력) 너비가 무한한 출입구를 통과하기에 탱크와 자동차중 어느것이 더 나은가?

답변) 둘다 통과 가능할테지만, 출입구의 높이 및 깊이에 대한 고려가 필요

답변) 무한히 넓다는 것은 출입구의 폭 제한이 없으므로 둘다 통과 가능

해석) GPT-4가 ChatGPT보다 추론이 부족한 사례

[그림 6] GPT-4의 미흡한 추론 사례

□ [비공개] 한편, OpenAI는 GPT-4를 유료로 서비스하면서 GPT-4에 대한 구체적인 스펙을 공개하지 않을 것으로 전망

- ▶ OpenAI는 GPT-4의 파라미터 수를 밝히지 않기로 결정
- 기존 비영리에서 영리 및 MS와 합작하는 형태로 전환하면서, GPT-4가 기존 대비 얼마나 더 크고, 왜 더 나은지에 대해 언급하지 않을 것*

* MIT Technology Review, GPT-4 is bigger and better than ChatGPT—but OpenAI won't say why, 2023.3.14.¹⁾

1) <https://www.technologyreview.com/2023/03/14/1069823/gpt-4-is-bigger-and-better-chatgpt-openai/>