

2024. 04.17

그림: Midjourney (고난의 테슬라... 그런데 xAI는?)

AI Weekly, 특이점이 온다 #12

W-135 to 

(AGI's ETA: 2026.11.21)

디지털리서치팀 한종목, 김은지

chongmok.han@miraeasset.com

02-3774-1498

eunji.kim.a@miraeasset.com

02-3774-8325



MIRAE ASSET
미래에셋증권

Highlight of the Week

▪ Google Cloud Next 2024

구글이 Cloud Next 2024 행사에서 생성 AI에 대한 높은 의지를 드러냄. Gemini를 클라우드 전반에 접목하고, AI에 최적화된 하드웨어와 소프트웨어 인프라를 제공하는 데 주력. 특히 '에이전트'라는 개념을 도입해 사용자의 업무 효율 향상을 위한 자동화 AI 구축에 방점. 하드웨어 측면에서는 자체 개발 칩인 TPU v5p와 엔비디아 H100 기반 VM, 그리고 Arm 기반 CPU 'Axion'을 선보였음. 한편, 순다르 피차이 CEO가 직접 나서 강조한 Gemini 1.5 Pro는 100만 토큰 처리가 가능해 긴 영상과 코드 입력이 가능한 수준으로 알려짐. 구글의 AI 전략에서 핵심은 결국, 반도체 인프라 및 Gemini의 '상대적 성능'이 될 전망.

▪ Don't bet against Elon Musk...?

테슬라 주가가 52주 신저가에 근접한 가운데, 1.5만 명 규모의 대규모 감원과 핵심 임원들의 이탈로 동요의 조짐이 있음. 이번 사태는 FSD 완성과 AI 역량 집중을 위한 조직 개편에 따른 불만이 주원인으로 보임. 테슬라는 자율주행과 AI 전쟁에 본격 참전하기 위해 FSD 요금 인하와 로보택시 서비스 준비에 박차. 한편, xAI는 설립 1년 만에 SOTA 모델과 어깨를 나란히 하는 Grok-1.5V를 출시. 이는 멀티모달 모델로 "실제 세계 QA 벤치마크"에서 타사와 차별화된 성과. xAI는 테슬라 및 X로부터 방대한 데이터를 수집할 수 있는 강점. 머스크의 AI 역량 집중이 성과를 거둘지, 구조조정이 부메랑이 될지는 좀 더 지켜봐야 할 것.

▪ GPT-4 최신 업데이트와 강력한 오픈소스들의 출시

OpenAI는 4월 12일, 유료 사용자를 위해 GPT-4 터보를 업그레이드하여 성능과 사용 경험을 향상. 특히 수학 및 논리 추론 능력이 크게 상승하여 클로드 3 Opus를 다시 능가. AI의 추론 능력은 에이전트와 AGI 개발에 매우 중요. OpenAI는 작년에 발표한 "Let's Verify Step by Step" 논문에서 Chain of Thought(CoT) 개념을 제안했고, 이번에 공개한 오픈소스 라이브러리를 통해 CoT 방식으로 언어모델의 추론 능력을 심화하려는 의도를 내비침. 이는 AI 성능 평가에 대한 '표준'을 주도하려는 것으로 해석. 한편, MS의 WizardLM-2와 Reka 등 강력한 성능의 새로운 언어모델들이 등장하여 GPT-4와 클로드 3 Opus에 필적하는 성과. 업계에서는 OpenAI가 연말까지 GPT-5 완성에 집중할지, 상반기 내 GPT-4.5를 내놓을지 주목하고 있음.

▪ Paper of the week – OpenEQA: Embodied Question Answering in the Era of Foundation Models (Meta)

AI 에이전트의 상식적 추론 능력을 평가하기 위한 벤치마크로 EQA를 제안. 과거 기억에 기반한 답변과 주변에서 정보를 수집해야 하는 답변으로 나누어 평가하며, 주요 대형 모델을 평가한 결과 GPT-4V가 가장 우수했으나 여전히 인간 수준을 크게 하회. 특히 공간 추론에서 부진한 모습 보여, 시각적 인지와 이를 기반으로 한 의미론적 이해에 대한 추가 연구 필요성 시사.

일	월	화	수	목	금	토
14 ICASSP 2024(~19)	15 (BIDU) 20214 바이두 크리에이트	16 Global AI Show(~17)	17 AI Tech Summit(~18) AI+ 2024(~18)	18 (TSMC) 실적발표	19 AI Rabbit Hole	20
21	22 Annual Development, Implementation & Management of ML Models(~24)	23 (TSLA) 실적발표 (GOOGL) 실적발표(예)	24 (NOW) 실적발표 (META) 실적발표 (IBM) 실적발표 (SK하이닉스) 실적발표	25 (SNAP) 실적발표 (MSFT) 실적발표 (INTC) 실적발표 (AMZN) 실적발표(예)	26	27 AI for Education Summit
28	29 (삼성전자) 실적발표	30	1 (QCOM) 실적발표(예) (EQIX) 실적발표(예)	2 (DDOG) 실적발표(예) (AAPL) 실적발표 (NET) 실적발표	3	4

AI issues incoming

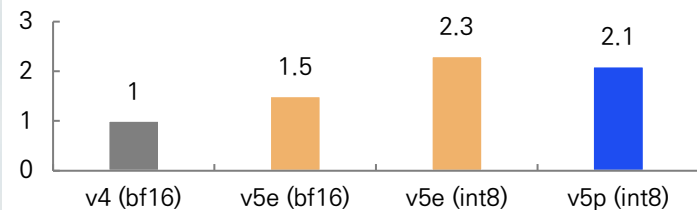
Google Cloud Next 2024

- 세계 3대 클라우드 사업자(CSP)인 구글 클라우드(GCP)의 주요 연례 행사 중 하나인 “Cloud Next 2024”가 지난주 라스베이거스에서 성황리에 종료. **2023년 행사는 불과 8개월 전인 8월에 열렸다는 점에서 생성 AI에 관한 구글의 절실함 내지는 조급함마저 느낄 수 있었음.** GCP의 CEO는 구글이 1,000개가 넘는 제품을 혁신했다고 자신 있게 밝혔음. 구글의 근간 전략은 AI 기반 언어모델 Gemini를 개발에서 운영에 이르는 클라우드 주영역에 집중하는 한편, 하드웨어와 소프트웨어 양 측면에서 AI에 최적화된 인프라를 제공하는 게 핵심. 키노트에서 특히 눈길을 끈 건 클라우드 활용의 새로운 패러다임으로 제시된 “에이전트”란 개념. 에이전트의 정의는 다양할 수 있겠지만 구글은 이를 “사용자의 목표 달성을 돕는 AI 개체”로 규정하고 있음. 다시 말해, 유저의 업무 효율을 올리기 위한 자동화 AI 구축에 방점을 찍은 것. 앞으로 구글의 에이전트는 Google Sheets/Docs부터 GCP의 아키텍처 전반을 아우르는 보편적 존재로 자리매김할 전망이다.

GCP 아키텍처를 들여다보면 AI 하이퍼 컴퓨터 / Gemini / Vertex AI라는 3개의 파운데이션 요소와 Gemini for GCP, Gemini for Workspace, Agent라는 3개의 애플리케이션 요소로 구성. 파운데이션 요소로서, AI 하이퍼 컴퓨터는 GCP의 AI 학습 및 추론 지원용 최적 인프라를 의미. **GCP 인프라는 업계 평균을 50% 이상 앞서는 에너지 효율성을 자랑하는데,** 이는 구글이 오랫동안 매달려온 자체 개발 칩인 TPU 덕분에 파악. **TPU 신제품인 v5p가 공개됐는데, 기존 v4 대비 FLOPS는 2배, HBM은 3배 늘어나며 유연성과 확장성이 대폭 개선.** 기존의 v5e와 함께 구색을 맞춰 고객들의 용도에 따라 달리 VM을 구성할 수 있음. TPU뿐 아니라 엔비디아의 H100으로 무장한 “A3 메가 VM”도 자랑하듯 공개했고 엔비디아의 차세대 아키텍처 기반 B200, GB200 NVL도 지원할 계획으로 밝혔음. **GCP 인프라의 “하마평”에 AMD와 인텔의 제품은 없었음.** 이와 더불어 **구글은 가속기 수준을 넘어 CPU까지 독자적으로 만들었다**며 공개. Arm 기반으로 만든 “Axion”으로, 현행 x86 기반 CPU 대비 50%의 성능 향상과 함께 60%의 전력효율 개선을 보인다고 주장. CPU를 독자적으로 만들어 GPU 내지는 AI 가속기와 붙이는 **엔비디아의 전략과 유사한데, 이러한 흐름은 Arm의 라이선스 사업부 수익에는 청신호, x86 진영에는 적신호를 의미.** 그리고, 구글을 10년 넘게 이끌어온 순다르 피차이 CEO가 직접 나서서 강조한 Gemini 1.5 Pro는 무려 100만 토큰을 동시에 처리할 수 있어 1시간 분량 영상은 물론 70만 단어, 3만 줄 길이의 코드까지 입력 가능한 수준. 마지막으로, **Vertex AI는 Llama 등 최신 오픈소스 LLM으로부터 커스텀 모델을 만들고 에이전트 구축환경까지 제공하는 올인원 AI 서비스 플랫폼. 엔비디아가 제시한 NeMo와 NIM과 가까운 서비스. 빅테크들은 미들웨어를 근래의 “노다지” 시장으로 보는 것.** **파운데이션 위에는 GCP용 Gemini, Workspace용 Gemini, Agent 등 애플리케이션 레이어가 얹혀 있음.** 예를 들어, 개발자들이 자연어로 코드 생성/디버깅을 하게 한다든가, 기업 고객들이 쓸 수준의 개인 보안성으로 이메일, 구글 닥스, 시트 등 업무 환경과, 창의적 영상 제작 도구에게까지 Gemini를 에이전트로 활용할 수 있게끔 하는 것. **구글의 AI 전략을 관통하는 6개의 층 중에서 로열층은 반도체 인프라와 언어모델 Gemini.** TPU+GPU+CPU 기반의 하이퍼컴퓨터가 TCO 관점에서 얼마나 효율적일지, Gemini가 SOTA모델로서 계속 우위에 있을지가 계속되는 관전포인트.

위: TPU별 스펙 / 아래: 달러당 AI 훈련 달러당 성능(가성비)

주요 스펙	TPU v4	TPU v5e	TPU v5p
Pod(4개의 TPU 캐비닛)당 칩 개수	4096	256	8960
Bf16 기준 TFLOPs	275	197	459
Int8 기준 TOPs	-	394	918
HBM 용량(GB)	32	16	95
HBM 대역폭 (GB/s)	1228	820	2765
칩간 통신속도 (기가비트/s)	2400	1600	4800



자료: Google, 미래에셋증권 디지털리서치팀
 주: v5e(bf16)은 2023년 8월 기준, v5e(int8)은 2023년 11월 기준

구글의 AI 전략을 관통하는 6개(파운데이션+어플리케이션)의 레이어

Gemini for GCP:
자연어로 코드 생성과 데이터 분석 가능. AI 개발 생산성 높임.

Gemini for Workspace:
맥락을 이해하여 문서 요약/번역 등 AI 기술을 업무용 앱에 적용

Your Agents:
사용자의 목표 달성을 보조하는 대화형 AI agent. 새로운 패러다임

Vertex AI: 최신 오픈소스 언어모델 기반 커스텀 모델 개발에서 MLOps에 이르는 원스톱 AI 플랫폼 서비스를 제공

Gemini Models: 성능과 비용, 용도에 맞게 다양한 버전의 Gemini 언어모델을 선택 활용할 수 있는 유연한 체계를 지님

AI Hypercomputer: GCP에 최적화된 AI 훈련 및 추론용 인프라로 에너지 효율과 성능이 업계 최고 수준

자료: Google, 미래에셋증권 디지털리서치팀

H100 GPU로 무장한 새로운 VM을 내놓았다고 자랑하는 구글



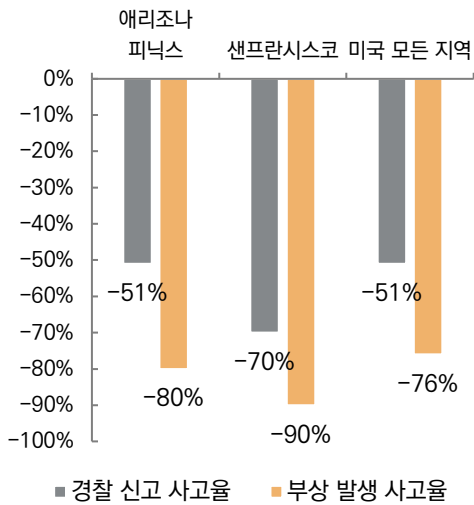
자료: Google, 미래에셋증권 디지털리서치팀
 주: VM은 Virtual Machine을 의미. 하드웨어 서버에 가상화 기술을 적용한 것.

AI issues incoming

Don't bet against Elon Musk...


- 테슬라는 52주 신저가에 거의 도달.** 고금리가 지속되며 차량판매에 부정적 영향을 받고, 차세대 보급형 모델 출시는 사실상 당분간 기대하기 어렵다는 평가 때문. 실제로 NV9로 알려진 이 저가모델 프로젝트를 위해 텍사스 공장 확장을 추진했다가 폐기된 것으로 파악. 게다가 테슬라에서는 **10% 인력 감원을 발표. 창사 이래 최대규모인 1.5만명에 이르는 정리해고는 생산인력 축소로도 이어질 것으로 보여, 테슬라에서도 올해 추가적인 수요 둔화에 대비한다고 시장은 해석한 것.** 더군다나 올초부터 성과급 축소나 연봉 조정, 스톡옵션 회수 등으로 동요의 조짐은 있었으나, 파워트레인 개발을 진두지휘하던 바글리노를 필두로 고위급 임원들이 테슬라를 떠나며 투자 심리는 극악으로 치달게 됨. 4680 및 양극재 프로젝트 주역들의 거취도 불투명. 다만, 이런 대규모 물갈이가 테슬라엔 그리 낯선 풍경도 아님. **2016년부터 2022년 사이에만 50여 명의 임원이 이탈한 전례.** 이 중에는 제롬 기엔(자동차 사업 총괄), 더그 필드(수석 부사장), 안드레 카파시(AI 총괄), JB 스트라우벨(CTO), 자크 커크혼(CFO) 등의 핵심 인재들을 포함. 이들의 사직 때마다 사내에 종말론적 분위기가 감돌곤 했고, 언론과 공매도 세력은 이를 톡톡히 악용했지만 **테슬라는 극복과 혁신을 거듭해 왔음.** 어쨌든 이번 사태는, TSLA 주가 하락에 따른 스톡옵션 가치절하로 사기가 떨어진 것도 있겠지만 그보다는 테슬라 내부에서 조직 개편 이후 새 보직에 불만이 있었기 때문으로 해석. **배터리 내재화 고도화나 보급형 모델 개발 보다는, FSD 완성을 위한 대규모 슈퍼컴퓨터 센터 등 AI역량에 더 집중하려는 것이 주요한 원인으로 보임.** 회사의 급성장기마다 '조직 재편과 슬림화'를 통해 다음의 국면을 열어젖혔다는 일론 머스크의 인사 철학으로 볼 때, AI / 자율주행 / 로봇공학 등에 사활을 걸겠다는 속내를 읽을 수 있음. 운송 중심 회사에서 벗어나 에너지와 AI 분야로 무게중심을 옮기려 했던 테슬라의 애초 청사진과도 동일. 특히 LA 진출을 선언한 구글 웨이모 또한 견제해야 하는 상황이기도 함. 테슬라가 올 8월 8일 로보택시 서비스 공개를 앞둔 가운데, **FSD 한 달 무료 정책을 시행하고, 월 구독 가격을 199달러에서 99달러로 파격적으로 낮춘 것도 자율주행/AI 전쟁에 본격적인 참전을 선포한 포석인 셈.** **"전시 상황이기 때문에 조직도 전시 체제가 돼야 한다"**라는 것. **앞으로 테슬라는 더욱 많은 GPU 구매와 Dojo 확장에 열중할 것으로 전망.**
- "전시 체제 CEO"가 AI 확장에 자원을 쏟아붓는 영역은 자율주행 뿐만은 아님. 머스크의 '순수 AI 기업'인 xAI는 지난 12일 Grok-1.5v를 출시.** 텍스트 처리에 더해, 문서, 다이어그램, 차트, 사진 등 다양한 시각 정보를 처리 가능한 멀티모달 모델. xAI는 **설립된 지 1년 밖에 되지 않았음에도, 언어론 능력에서 문서 이해력에 이르기까지 현존 SOTA 모델과 어깨를 나란히 하는 역량을 과시.** 1년 사이에 이 정도 성과를 올렸다면 내년엔 어떤 모습을 하고 있을지 기대할 수 밖에 없음. 비디오 등 다양한 양식을 지원도 있을 예정. Grok은 동급 모델을 압도하는 성과를 보인 벤치마크가 있는데, 이것은 **xAI가 새롭게 제안한 "실제 세계 QA 벤치마크". 실제 현실에서 AI가 3D 세상을 어떻게 이해하는지 그 능력을 평가하는 것.** 물리 공간 인지 능력의 고도화야말로 실사용을 위한 AI의 핵심 관건. **실제 세계의 복잡성을 이해하고 상호작용할 수 있어야만 유의미한 AI 에이전트를 이룰 수 있음.** Grok은 이에 대해 타사와 다른 차별점을 보유. 바로 **일론 머스크의 존재 덕분에 테슬라로부터 이미지/영상 데이터, X로부터 실시간 텍스트/이미지/영상 데이터를 수집할 수 있다는 점.** 그는 자본력, AI 인재, 데이터, 모델을 테스트해볼 사용자와 플랫폼까지 다 갖췄고, 없었던 것은 오직 파운데이션 모델뿐이었음. AI 역량 집중의 성과로 범테슬라 진영의 가치가 급부상할지, 구조조정으로 인한 조직력 약화가 부메랑이 될지는 조금 더 지켜봐야... **지난 10년간 AI의 발전 사항을 돌아켜보면 "Don't bet against Elon Musk"를 떠올리게 됨.**

인간 운전자 대비 웨이모의 사고율 감소폭



자료: Google, 미래에셋증권 디지털리서치팀
주: 부상발생율은 100만 마일당 0.41건(vs 인간 2.78건)

한살배기 xAI의 멀티모달 모델 Grok-1.5v(vision)의 주요 하이라이트



2024년 4월 12일

Grok-1.5 비전 미리보기

최초의 다중 모드 모델로 디지털 세계와 실제 세계를 연결합니다.

- Vision 기능이 추가되어 이미지, 다이어그램, 차트 등 시각정보 처리가 가능
- 다양한 벤치마크 테스트에서 GPT-4, 클로드 3 등 SOTA와 비등한 성능
- **테슬라로부터 수집한 실제 데이터를 활용**해 현실 세계에 대한 이해도 높음
- 트래픽이 급증 중인 **X(트위터)의 실시간 데이터도 xAI가 가진 해자**
- "실제 세계 QA" 라는 벤치마크와 이를 위한 데이터셋을 오픈소스로 공개
- 앞으로 이미지, 오디오, 특히 비디오 등 다양한 분야로 확장 예정

자료: xAI, 미래에셋증권 디지털리서치팀

Grok-1.5v와 최첨단 언어모델들간의 멀티모달 관련 벤치마크 점수 비교

벤치마크	Grok-1.5 Vision	GPT-4 Vision	클로드 3 Sonnet	클로드 3 Opus	Gemini Pro 1.5
MMMU 다중언어추론	53.6	56.8	53.1	59.4	58.5
Mathvista 수학	52.8	49.9	47.9	50.5	52.1
AI2D 다이어그램	88.3	78.2	88.7	88.1	80.3
TextVQA 글 읽기	78.1	78	-	-	73.5
ChartQA 차트 해석	76.1	78.5	81.1	80.8	81.3
DocVQA 문서 해독	85.6	88.4	89.5	89.3	86.5
RealWorldQA 실제 세계 이해	68.7	61.4	51.9	49.8	67.5

자료: xAI, 미래에셋증권 디지털리서치팀

AI issues incoming

GPT-4 최신 업데이트와 강력한 오픈소스들의 출시

- 지난 4월 12일, OpenAI는 유료 사용자들을 위해 한층 업그레이드된 GPT-4 터보를 선보임. 샘 알트만의 말마따나 이번 버전은 사용 경험 면에서 성능이 향상. ChatGPT 활용 시 답변이 한결 간결해지고 자연스러워진 것. 실제로 업데이트된 최신 GPT-4 버전은, **클로드 3 Opus에게 잠시 빼앗겼던 왕좌를 탈환**. 이번 버전은 **특히 수학 및 논리 추론 능력이 크게 상승한 점이 중요. AI가 자율적으로 reasoning(추론) 능력을 인간 수준으로 발휘해야만 ‘에이전트’가 유의미한 성능을 달성할 수 있고, 더 나아가 AGI에 닿을 수 있기 때문**. OpenAI에서는 작년 5월 “Let’s Verify Step by Step”이라는 논문을 낸 적이 있음. 수학 문제를 풀 때 답만 내놓는 것이 아닌 그 답을 내기 위한 단계별 풀이 과정을 내놓게 하고, 각 단계마다 옳게 된 풀이인지 평가하게 하는 방식을 제안한 것. 이는 Chain of Thought(CoT) 라는 개념과 닿아 있음. 그런데, OpenAI가 **이번에 GPT-4 업데이트뿐만 아니라 오픈소스로 공개한 라이브러리를 볼 때, OpenAI에서는 CoT 방식으로 언어모델의 추론 능력을 더욱 심화할 계산**으로 생각함. 이 라이브러리는 언어모델의 성능을 제대로 평가하기 위한 가이드라인 내지는 모델 간 성능비교를 위한 공통 기준을 마련하려 한 것. 이 기준에서는 다양한 프롬프트 기법 중 CoT가 핵심. OpenAI에서는 언어모델의 추론 능력을 따질 때는 CoT로 평가해야 객관성을 갖출 수 있다고 판단하고 있는 것. 다시 말해, **CoT로 LLM을 고도화 중인 본인들의 생각을 전파하여 AI 성능의 헤드를 계속 차지하려는 심산**. 테슬라와 메타가 각각 새로운 벤치마크를 들고 나온 것, OpenAI가 벤치마크 평가를 위한 라이브러리를 공개한 것, 이 모든 행위들은 **AI 성능에 관한 “표준 싸움”이 벌어지고 있는 것으로 생각하면 됨. ‘제각기 다른 벤치마크로 누가 낫다고 하지 말고, 객관성 있는(물론 내가 만든) 벤치마크로 경쟁해보자’인 셈**.
- 한편, 예상치 못한 강력한 성능의 언어모델들의 출현이 AI 업계를 또 다시 뜨겁게 달구고 있음. 우선 **마이크로소프트 AI팀이 내놓은 'WizardLM-2'가 'MT-벤치'에서 GPT-4에 버금가는 위력을 뽐냄. 미스트랄의 'Mixtral 8x22B' 모델을 기반으로 추가학습을 거친 이 모델은**, 오픈소스 모델 중에는 처음으로 MT-Bench 기준 9.12를 기록. 최고 수준인 **클로드 3 Opus(9.43), GPT-4-1106(9.32)와 엇비슷한 정도**. WizardLM의 가중치는 허깅페이스를 통해 곧 공유될 예정. WizardLM-2는 훈련 데이터를 사용하는 과정에 있어서, 가상의 데이터를 활용한 것이 부각. AI를 학습하는데 있어 데이터 고갈 이슈에 있어서 크게 우려할 필요가 없을 거란 희망적 전망까지 나온 상황. **또 다른 강자로 등장한 건 폐쇄형 멀티모달 LLM인 “Reka”. 불과 20명이 수천 대의 H100으로 빚어낸 결과물인데, GPT-4/Opus 레벨에 도달한 결과(Human Evaluation Multimodal 기준)를 내보임**. 심지어 아직 초기버전이라는 점과, 앞으로 공개될 기능이 ‘동영상 이해’라는 게 놀라운 부분. 우리가 올 초 예상한대로 GPT-4 수준의 모델들 다수가 수면 위로 올라온 상황. **OpenAI가 연말까지 인내하며 GPT-5 완성에 집중할지 상반기 내 GPT-4.5를 내놓지가 초미의 관심사**.

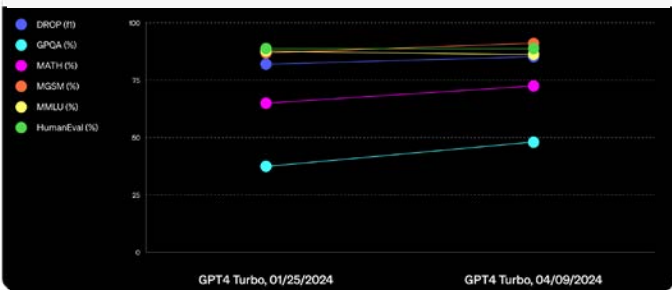
4월 12일 GPT-4 최신 업데이트에 대한 샘 알트만의 트윗



Sam Altman
@sama

gpt-4 now significantly smarter and more pleasant to use

GPT-4 최신 업데이트버전 유료 사용자에게 오픈 글쓰기, 수학, 논리적 추론, 코딩에서 성능 향상



9:19 AM · Apr 12, 2024 · 1.3M Views

자료: X(@sama, OpenAI), 미래에셋증권 디지털리서치팀

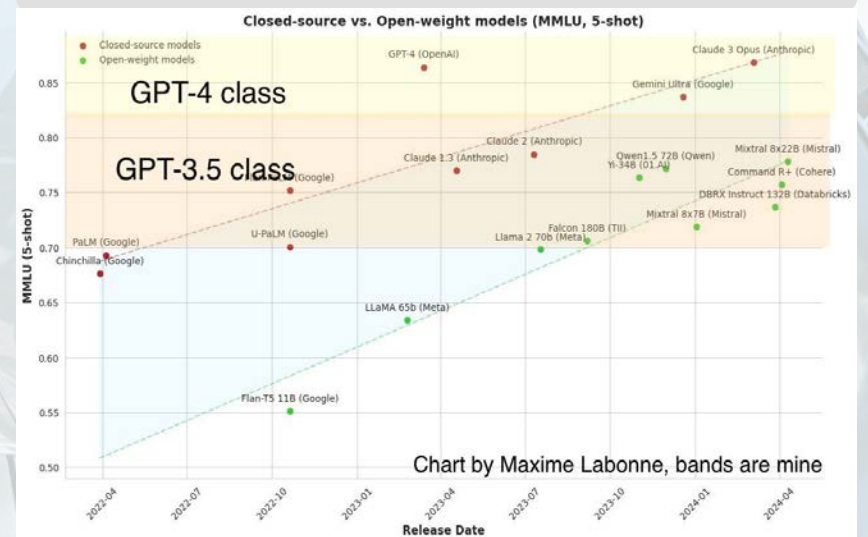
업데이트된 최신 GPT-4 버전이 다시 왕좌를 탈환 (Elo 기준)

모델	Elo 점수	기관	라이선스	지식 컷오프
GPT-4-Turbo-2024-04-09	1260	OpenAI	폐쇄소스	Dec-23
Claude 3 Opus	1255	Anthropic	폐쇄소스	Aug-23
GPT-4-1106-preview	1254	OpenAI	폐쇄소스	Apr-23
GPT-4-0325-preview	1250	OpenAI	폐쇄소스	Dec-23
Gemini Pro	1209	Google	폐쇄소스	실시간
Claude 3 Sonnet	1203	Anthropic	폐쇄소스	Aug-23
Command R+	1193	Cohere	오픈소스	Mar-24
GPT-4-0314	1189	OpenAI	폐쇄소스	Sep-21
Claude 3 Haiku	1182	Anthropic	폐쇄소스	Aug-23
GPT-4-0613	1164	OpenAI	폐쇄소스	Sep-21
Mistral Large 2402	1158	Mistral	폐쇄소스	모름

자료: LMSYS Chatbot Arena, 미래에셋증권 디지털리서치팀

주: 지식 컷오프란? 챗봇이 인지하는 데이터의 시점. 이후의 정보는 모른다는 뜻

폐쇄소스 모델과 오픈소스 모델간 성능 비교, ‘오픈소스의 개발속도 ↑’



자료: X(@maximelabonne)

주: X축은 각 언어모델들의 출시 날짜, Y축은 5-shot 기준 MMLU 점수

Paper of the Week

OpenEQA benchmark - Meta

✓ Core Contents

- AI 에이전트는 스마트폰, 스마트 글래스, 로봇 등을 통해 물리적 세상으로 나아가고 있어, 자연어 인터페이스와 인간적인 '상식'을 갖출 필요가 있음.
- 이런 맥락에서 Embodied Question Answering(EQA)이 에이전트의 세상에 대한 이해 능력을 평가하는 수단으로 제안됨.
- OpenEQA, 세계 최초의 개방형 어휘 EQA 벤치마크**로서, 현실 세계에서 수집된 1,600개 이상의 질문으로 구성. 과거 기억을 기반으로 답변해야 하는 Episodic Memory EQA(EM-EQA)와 직접 이동하면서 정보를 수집해 답변해야 하는 Active EQA(A-EQA)로 구분.

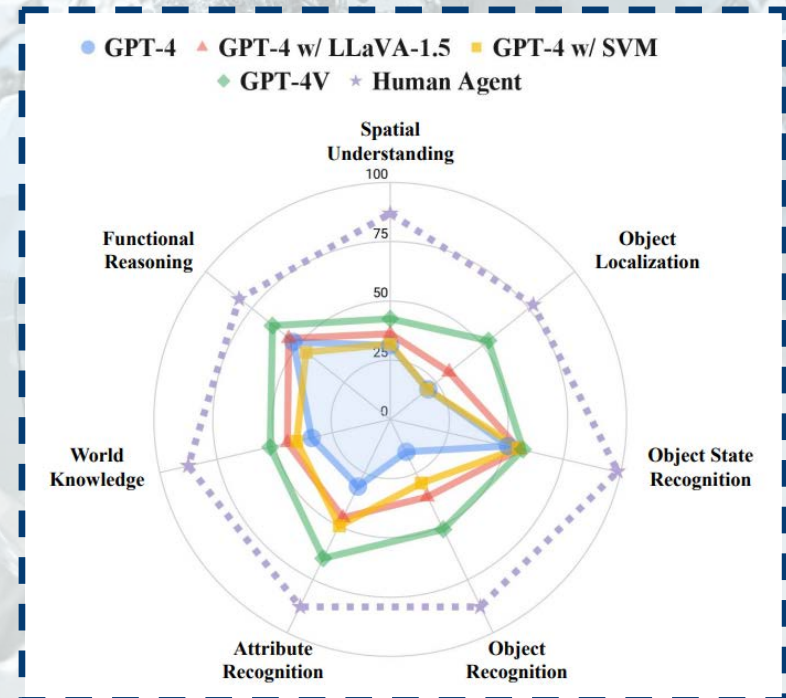
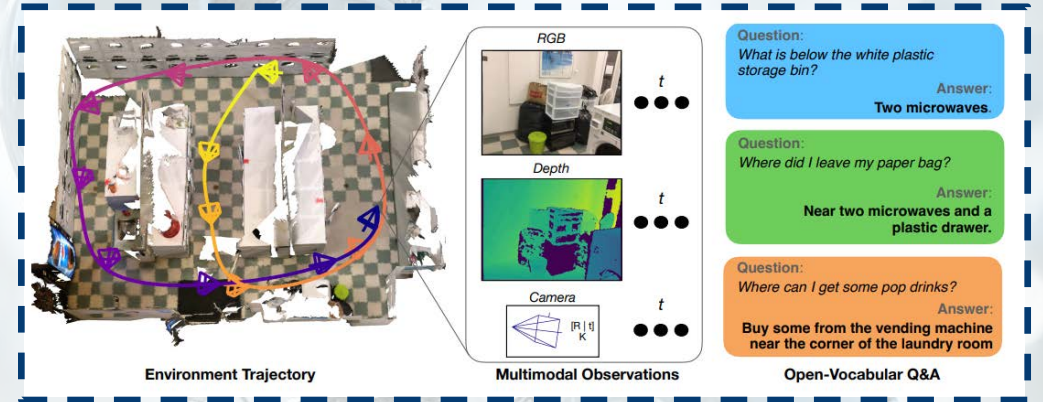
**개방형 어휘(Open-Vocabulary) 시스템: 자연어 처리나 질의응답 등에서 미리 정의되지 않은, 심지어 훈련 데이터에 등장하지 않은 단어나 표현도 처리할 수 있는 시스템*

- 개방형 어휘 특성상 정답 평가가 어려워, LLM 기반 **자동 평가 프로토콜을 제안.** 사람이 채점한 것과 모델의 답변을 주고 LLM이 유사도를 평가하게 하면, LLM의 평가가 사람의 평가와 유사해져 사람의 채점 없이 답변 적절성을 판별할 수 있음.
- GPT-4V, Claude-3, Gemini-1.5 등 최신 대형 멀티모달 언어모델을 포함한 다양한 베이스라인을 제공하며, **GPT-4V가 48.5%로 가장 우수한 성능을 보였으나 사람 수준(85.9%)에는 크게 미치지 못함.**




✓ Insights

- OpenEQA는 "Embodied AI 에이전트"에게 요구되는 자연어 이해와 상식 추론 능력을 평가할 수 있는 현실적 벤치마크. 테슬라가 제시한 벤치마크와 궤를 같이 함.
- 향후 Embodied AI, 대화형 에이전트, 세계 모델 등의 연구 발전을 촉진할 것으로 기대됨. 이 과정 속에서 어떠한 벤치마크가 헤게모니를 잡을 지가 관건.
- 공간적 추론이 필요한 질문에서 현재의 시각-언어 모델(VLM)들이 텍스트만 사용하는 LLM과 비슷한 저조한 성능을 보였음. 이는 인지 능력 향상을 위한 **시각적 인식의 의미론적 연결에 대한 추가 연구 필요성을 시사.**

자료: "OpenEQA: Embodied Question Answering in the Era of Foundation Models", Meta




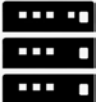

AI Peer Table (1)

밸류체인	종목명	티커	주가 (현시동화)	시가총액 (조원)	기간별 수익률(%)						엔비디아와의 상관관계(30일)	실적발표 예정일	AI 밸류체인 관련 한 줄 평
					1W	1M	3M	6M	1Y	YTD			
 최종 소비자 (온디바이스, 앱 등)	애플	AAPL US	169.38	3,642.8	-0.2	-1.9	-7.6	-5.0	3.1	-11.9	0.59	05-02 온디바이스 AI 위한 자체 경량화 언어모델 개발 중(Ajax GPT)	
	테슬라	TSLA US	157.11	696.9	-11.2	-3.9	-28.6	-38.1	-15.1	-36.8	0.49	04-23 자율주행 소프트웨어 FSD의 싱글스택 신경망 v12 공공출시 임박	
	삼성전자	005930 KS	80,000	477.6	-4.3	10.7	12.7	15.8	24.3	1.9	0.12	04-30 '24년부터는 온디바이스 AI에도 중점(Galaxy AI)	
	퀄컴	QCOM US	168.58	262.0	-4.0	0.8	20.1	54.3	43.9	17.2	0.62	05-01 생성 AI 처리 강화용 온디바이스 AP 및 CPU 설계(X Elite)	
	스냅	SNAP US	10.41	23.9	-7.3	-7.0	-33.0	7.1	-2.6	-38.5	0.36	04-25 이미지 생성 모델을 탑재한 증강현실 서비스(AR 필터, AR 렌즈)	
	로블록스	RBLX US	36.64	32.6	-6.3	-6.6	-6.9	20.5	-19.8	-19.9	0.44	05-09 생성 AI 탑재해 무료로 'Low code' 지원(캐릭터, 아이템 생성 위한 AI)	
 AI 비즈니스 솔루션	어도비	ADBE US	476.22	297.1	-3.3	-3.3	-20.3	-13.5	25.5	-20.2	0.59	06-13 생성 AI 기반 미디어 편집 툴(Firefly)	
	세일스포스	CRM US	277.41	374.8	-8.3	-5.7	3.2	33.2	42.7	5.6	0.54	05-31 세계 최고의 CRM 업체로 생성 AI를 탑재해 사용성 강화(Einstein GPT)	
	IBM	IBM US	183.75	234.6	-2.9	-3.8	11.0	34.7	50.1	13.4	0.29	04-24 기업 고객용 AI 어시스턴트(상담사 챗봇 구축, 업무 자동화, 코딩 지원)	
	서비스나우	NOW US	744.78	213.0	-4.8	0.1	2.4	35.0	60.8	5.4	0.61	04-24 워크플로우 자동화 기업으로 생성 AI를 탑재해 사용성 강화(Now Assist)	
	클라우드스트라이크	CRWD US	299.15	100.8	-3.7	-5.2	4.5	59.4	120.3	17.2	0.54	05-31 기업 대상 생성 AI 챗봇 기반 클라우드 및 엔드포인트 보안(Charlotte AI)	
	데이터독	DDOG US	126.95	58.7	1.2	5.4	2.6	40.2	88.6	4.6	0.53	05-03 생성 AI 기반 클라우드 인프라 및 보안 모니터링(Bits AI)	
	다이나트레이스	DT US	44.52	18.4	-6.7	-2.4	-20.2	-7.9	5.3	-18.6	0.51	05-17 생성 AI 기반 어플리케이션 성능 관리 및 보안 플랫폼(Davis)	
	유니티	U US	23.94	12.9	-11.6	-8.2	-29.8	-15.6	-18.7	-41.5	0.48	05-09 올해 출시 예정인 차세대 게임엔진에 AI 기능 대거 탑재(Sentis, Muse)	
	센티널원	S US	20.89	9.0	-6.8	-5.4	-15.8	20.5	16.9	-23.9	0.46	05-31 기업 고객 대상 생성 AI 챗봇 기반 클라우드 및 엔드포인트 보안(Purple AI)	
	루닛	328130 KS	51,500	1.5	-5.7	-20.2	-26.5	-30.9	73.8	-37.1	0.09	05-10 AI 기반 암 검출/진단 보조 솔루션(인사이트)	
 기반 모델	알파벳	GOOGL US	154.40	2,687.2	-1.4	9.4	8.4	11.0	41.8	10.5	0.57	04-25 OpenAI와의 AGI 경쟁에 있어 대적할 수 있는 거의 유일한 업체(Gemini)	
	메타 플랫폼스	META US	499.76	1,774.5	-3.3	3.2	36.1	55.8	125.9	41.3	0.52	04-24 '거의 완전한' 오픈소스 LLM 개발 및 배포를 선도(Llama)	
	바이두	9888 HK	94.15	47.0	-7.7	-7.1	-2.7	-21.1	-28.1	-18.9	0.12	05-16 중국 최초/최고 성능의 LLM 보유(Erne Bot)	
	네이버	035420 KS	179,500	29.2	-5.0	-3.5	-18.0	-4.4	-8.4	-19.9	0.13	05-08 자체 개발한 한국형 LLM 개발 및 AI 서비스 출시 예정(HyperCLOVA)	

자료: Bloomberg, 미래에셋증권 디지털리서치팀

주: 실적발표 예정일은 수시로 변동될 수 있음

AI Peer Table (2)

밸류체인	종목명	티커	주가 (현지통화)	시가총액 (조원)	기간별 수익률(%)						엔비디아와의 상관관계 (30일)	실적발표 예정일	AI 밸류체인 관련 한 줄 평
					1W	1M	3M	6M	1Y	YTD			
 AIOps	스노우플레이크	SNOW US	150.30	70.0	-3.4	-4.2	-20.1	-7.3	5.7	-24.5	0.44	05-24 AI 모델 구축, 배포, 관리를 위한 빅데이터 워크로드 플랫폼(Cortex)	
	몽고DB	MDB US	349.80	35.5	-3.1	-1.6	-13.7	-5.0	56.8	-14.4	0.55	05-31 AI 모델 구축 시 필수 인프라로 떠오른 '벡터 데이터베이스' 기능 보유(Atlas)	
	클라우드플레이어	NET US	91.04	42.8	-6.1	-0.6	16.0	43.3	40.8	9.3	0.55	05-02 트래픽 처리(CDN) 업체로, 실시간 AI 추론 작업에서의 역량 대두(Workers AI)	
	팔란티어	PLTR US	21.88	67.4	-4.1	-6.9	32.4	26.0	148.4	27.4	0.50	05-06 실제 조직 운영시의 의사결정 지원 및 자동화를 위한 AI 개발 플랫폼(AIP)	
	포스 패러다임	6682 HK	48.90	4.1	-9.9	-32.8	-0.8	-12.4	-	-2.6	0.14	08-19 실제 조직 운영시의 의사결정 지원 및 자동화를 위한 AI 개발 플랫폼(Sage)	
 데이터센터	마이크로소프트	MSFT US	414.58	4,290.3	-2.7	-0.4	6.4	25.1	46.1	10.5	0.66	04-25 OpenAI의 AI 모델 라이선스에 대한 독점권을 갖고 있는 세계 2위의 CSP	
	아마존 닷컴	AMZN US	183.32	2,652.1	-1.3	5.1	19.7	38.3	78.8	20.7	0.59	04-30 최고의 CSP로써 작년 9월 Anthropic에 40억 달러 투자하며 AGI 경쟁 참여	
	오라클	ORCL US	120.62	461.7	-2.1	-3.6	13.6	11.7	27.8	15.2	0.47	06-12 기업들의 자체 인프라 수요에 대응한 '온프레미스 서버' DB 관리 시스템의 강자	
	에퀴닉스	EQIX US	742.14	97.8	-7.4	-12.7	-9.0	-0.2	9.9	-7.4	0.46	05-03 물리적 공간의 데이터센터를 임대/위탁 운영하는 업계 최대 리츠 업체	
 반도체 인프라	엔비디아	NVDA US	874.15	3,043.7	2.4	-0.5	55.0	89.7	226.8	76.5	1.00	05-22 AI 훈련은 GPU가 필수, GPU 현존 최강자는 엔비디아	
	브로드컴	AVGO US	1,329.06	857.8	-0.4	8.0	19.7	48.5	118.8	19.6	0.68	06-12 VMware 인수로 클라우드 가상화 역량도 확보한 네트워킹 반도체의 최강자	
	TSMC	2330 TT	788.00	875.8	-3.3	5.1	36.2	44.4	54.6	33.5	0.16	04-18 압도적인 미세공정 점유율을 가진 대만의 반도체 파운드리	
	AMD	AMD US	163.46	367.8	-4.3	-14.4	3.0	53.5	78.2	10.9	0.78	05-02 HPC CPU에서 업계리딩, GPU에서 엔비디아의 적수로 꼽히는 패스트 팔로워	
	인텔	INTC US	36.26	215.0	-5.4	-15.0	-22.7	-0.2	15.3	-27.6	0.45	04-25 GPU 중심 서버 설계에는 후발주자. 파운드리 기업으로서 지정학적 가치 부상	
	마이크론	MU US	121.77	187.8	-0.7	30.7	44.1	76.7	95.6	42.8	0.62	06-28 HBM에서 SK하이닉스를 추격하는 미국의 메모리 반도체 업체	
	SK하이닉스	000660 KS	179,100	130.4	-2.1	11.1	36.7	38.1	103.9	26.6	0.21	04-25 GPU 병목의 키를 쥐고 있는 HBM에 있어 글로벌 선두주자	
	시놉시스	SNPS US	541.09	115.0	-4.9	-1.6	6.2	10.0	41.8	5.1	0.72	05-17 반도체 EDA 업체로, 3D 엔지니어링 시뮬레이터 업체 앤시스와 M&A 논의 중	
	아리스타 네트워크	ANET US	262.76	114.4	-10.9	-5.4	3.8	33.5	60.6	11.6	0.60	05-07 이더넷 중심 고속 네트워킹을 위한 스위치 설계	
	Arm	ARM US	122.22	174.8	-3.6	-3.7	78.5	134.6	-	62.6	0.21	05-08 AI 가속기 설계 위한 다수의 IP 블록 소유한 업체로 설계 라이선스/로열티 사업	
	마벨 테크놀로지스	MRVL US	68.28	82.4	-6.8	2.2	4.3	27.5	70.2	13.4	0.71	05-24 데이터 인프라용 네트워킹 및 스토리지 ASIC 설계	
	SMIC	981 HK	14.02	30.5	-4.6	-16.8	-11.8	-31.8	-43.1	-29.4	0.03	05-09 최근 화웨이와 밀접히 협력 중인 중국 유일의 미세공정 가능 파운드리	
	Alchip	3661 TT	2,830.00	9.6	-6.3	-21.5	-20.6	7.6	112.3	-13.6	0.16	05-09 대만에 있는 세계에서 AI 반도체 관련 매출 비중이 가장 높은 ASIC 업체	

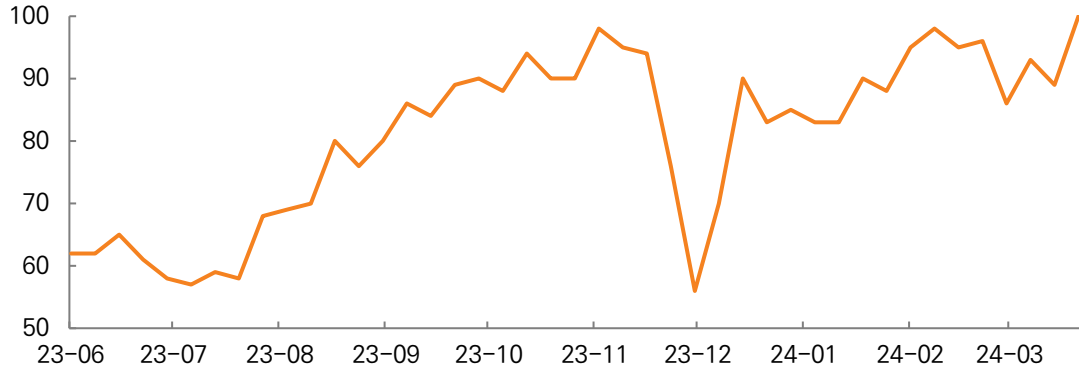
자료: Bloomberg, 미래에셋증권 디지털리서치팀
 주: 실적발표 예정일은 수시로 변동될 수 있음

Appendix: Charts

ChatGPT 출시일 이후, S&P 500에서 Magnificent Seven이 차지하는 비중 (시가총액 기준)



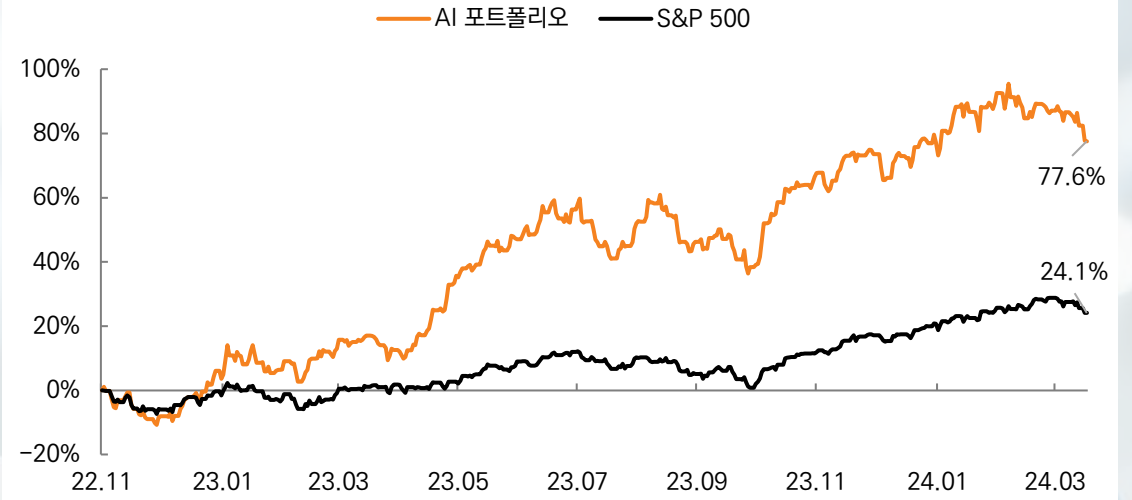
ChatGPT 키워드를 검색한 Google Trends 추이 (지난 90일 평균치 기준)



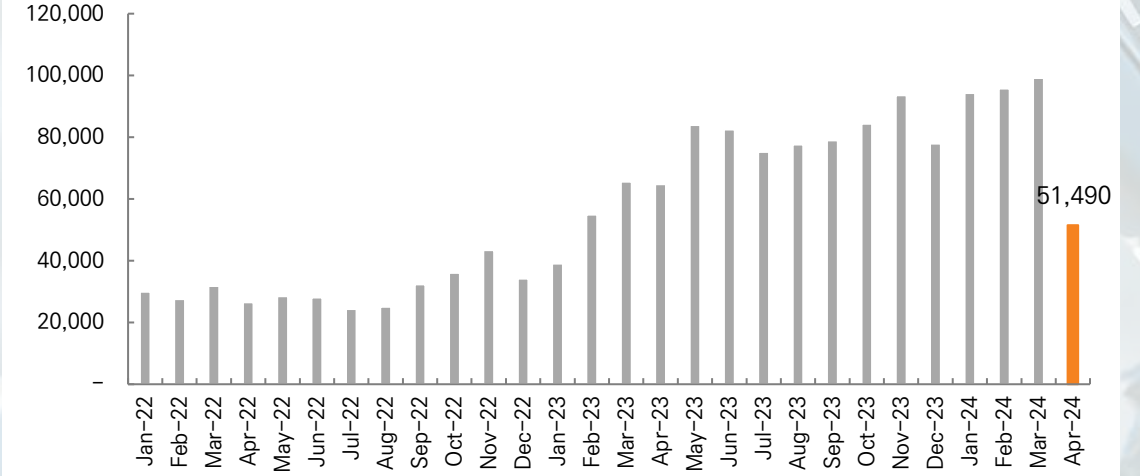
ChatGPT 관련, 지난 한 달간 주요 "급상승 키워드"

Claude

ChatGPT 출시일 이후, AI 밸류체인에 속한 유니버스 종목들의 평균 수익률 (동일가중 방식)



월간 기준, 글로벌 뉴스 기사에서 AI라는 키워드가 언급된 횟수 (해당월은 여전히 진행 중)



Compliance Notice

- 당사는 본 자료를 제3자에게 사전 제공한 사실이 없습니다.
- 본 자료는 외부의 부당한 압력이나 간섭없이 애널리스트의 의견이 정확하게 반영되었음을 확인합니다.

본 조사분석자료는 미래에셋증권이 신뢰할 수 있는 자료 및 정보로부터 얻은 것이나, 당사가 그 정확성이나 완전성을 보장할 수 없으므로 투자자 자신의 판단과 책임하에 종목 선택이나 투자시기에 대한 최종 결정을 하시기 바랍니다. 따라서 본 조사분석자료는 어떠한 경우에도 고객의 증권투자 결과에 대한 법적 책임소재의 증빙자료로 사용될 수 없습니다. 본 조사분석자료의 지적재산권은 당사에 있으므로 당사의 허락 없이 무단 복제 및 배포할 수 없습니다.