

2024. 03.07

그림: Midjourney (큰 힘에는 큰 돈이 따른다)

AI Weekly, 특이점이 온다 #9

W-130 to AGI

(AGI's ETA: 2026.09.08)

디지털리서치팀 한종목

chongmok.han@miraeasset.com

02-3774-1498

MIRAE ASSET
미래에셋증권

Highlight of the Week

■ 구글의 AI 개발 수장, 데미스 허사비스의 인터뷰

지식 간 전이가 자연스러운 뇌의 존재는 AGI가 가능하다는 것을 증명. AI 초지능의 안전성과 사이버보안에 대해 UN 수준의 논의 필요. 한편 언어 데이터를 기반으로 대규모 모델의 개념화 및 추상화를 위해 '스케일링'이 가능한 알고리즘 사용 중요. 규모를 확장하는 과정에서 상당한 계산 자원이 요구되지만, 구글은 올해까지 어떤 연구소보다도 훨씬 많은 계산을 가지게 될 것. 다만, 스케일링만으로는 부족. 새로운 접근 방식이 필요. 1)월드모델을 만들어 planning 능력을 키워야 함. 이를 위해 게임을 활용해 시뮬레이션 및 자가플레이를 진행 2)월드모델 구축을 위해 다양한 유형의 데이터 처리가 가능한 멀티모달이 필요

■ AI를 향한 머니게임에 빅테크는 계속 더블다운

Anthropic의 Claude 3는 많은 벤치마크에서 GPT-4를 능가한다고 주장. 성능과 비용 면에서도 좋은 대안. 최고급 모델인 Opus는 전문가 수준의 이해력, 메타인지력, 큰 컨텍스트 윈도우를 제공. 성능 향상의 비결은 지속적 데이터 사전학습과 지시에 따른 미세조정. 이는 모델의 사전학습 시작시기와 데이터 품질/양이 중요함을 시사. 아마존과의 협력은 AI 발전에서 컴퓨팅 능력과 전력 인프라가 중요함을 강조.

■ 법적 분쟁 불구, 경쟁자 출현했으니 GPT 후속 나올까?

OpenAI의 공동설립자였던 일론 머스크가 조직의 설립 사명('오픈된 AI 형태의 개발')을 위반했다는 등의 이유로 OpenAI와 샘 알트만 CEO를 고소. OpenAI 측은 머스크 의견을 전면 반박했지만 법적 분쟁이 해결되기 전까진 신모델 출시가 지연될 가능성 존재. 다만 샘 알트만 CEO는 해당 이슈가 팀의 펀더멘탈에는 영향이 없다고 자신. 한편 클로드 3의 경쟁 압력 측면에서 GPT-4.5 출시 가능성에 대한 기대감 상승.

■ 구글 Gemini 사태는 더 근본적인 문제를 함의한다

구글 Gemini의 역차별 사태는 더욱 더 근본적인 문제인 기업문화, 리더십 부족을 함의하고 있음. 검색엔진 비즈니스로 레거시가 된 구글이 OpenAI 등 LLM 개발 경쟁에서 패할 가능성이 높아 보이는 이유.

■ Paper of the week: The Era of 1-bit LLMs: All Large Language Models are in 1.58 Bits

기존 모델 파라미터로 쓰였던 부동소수점을 -1, 0, 1의 세 정수로 대체해 추론에 필요한 행렬 계산을 단순화함. 이를 통해 레이턴시 및 메모리 사용량을 기존 모델 대비 크게 개선. 모델 경량화로 CPU를 활용한 추론이 가능해 온-디바이스 AI를 위해 사용될 수 있을 것으로 기대. 다만 모델 개량 방법의 일환이므로 기존 모델의 성능을 뛰어넘지는 못한다는 점에 유의.

일	월	화	수	목	금	토
3	4	5 (CRWD) 실적발표	6	7 (MRVL) 실적발표 (MDB) 실적발표 (AVGO) 실적발표	8 (Alchip) 2월 실적발표 (TSMC) 2월 실적발표	9
10	11 (ORCL) 실적발표 ACM/IEEE 인간-로봇 상호작용 컨퍼런스 2024(~14)	12	13 (S) 실적발표	14 (ADBE) 실적발표	15	16
17 EACL 2024(~22)	18 GTC 2024(~21) IUI 2024(~21) 3DV 2024(~21)	19 Gen AI for Business Apres-Cyber Slopes Summit(~21) Data Science Salon(~21)	20 (SNPS) 인베스터 데이	21	22	23

AI issues incoming

구글의 AI 개발 수장, 데미스 허사비스의 인터뷰 (1)

지난 2월 29일 AGI를 둘러싼 여러 예측과 계획에 관해 구글의 데미스 허사비스의 인터뷰 내용을 요약 정리

AGI가 언제쯤 도래할까?

- 애초에 AGI의 출현과 관련된 안전에 대한 우려로 2010년에 DeepMind를 시작했음. 설립 때 딥마인드를 20년 프로젝트로 생각했음. 실제로 AGI의 타임라인은 기존의 생각대로 진행 중. 현재 인류의 컴퓨팅 능력이 뇌의 시냅스 수와 비슷한 크기로 상승 중. **2030년 내 AGI 시스템이 출현해도 놀라지 않을 것.** 그러나 AGI가 현실화되기까지 마냥 기다리기보다 현재의 AI 기술 발전과 솔루션 개발에 집중해야 함. **개인적인 열정은 과학과 헬스케어를 위한 AI.** 구글이 내놓은 “AlphaFold” 프로젝트도 같은 맥락.

뇌 과학과 AGI, 그리고 ‘초지능’

- 우리의 뇌가 세계를 이해하고 처리하는 방식에 어떤 공통적인 원칙이나 알고리즘이 있다고 생각. 인간의 뇌에서는 인간이 학습하는 과정에서 얻게 되는 지식간 전이가 자연스러움. 이는 AI에서도 마찬가지. **대형언어모델(LLM)이 코딩 능력을 개선하면, 그 과정에서 얻은 논리적 사고나 문제해결 능력으로 다른 영역에서도 성능을 높일 수 있음.** 코딩과 수학을 잘 하게 되면 일반적인 인지 능력을 향상시킬 수 있다는 희망. AI의 발전을 위해 뇌 신경과학이 많은 것을 더했다고 생각. 특히 attention 개념 등 많은 흥미로운 방향성 단서 또한 인간 뇌의 복잡한 작동 방식을 이해하려는 노력으로 찾게 되었음. **뇌는 AGI가 가능하다는 것을 증명하는 존재, 가능성을 확인하면 남은 것은 스케일링.**
- 더 나아가, **AI 초지능은 사회 전체에 영향을 미칠 수 있는 중대한 기술.** AI 모델들을 ‘어떤 목적으로 배포하고 싶은지’를 판단하는 데 사회 전체가 관여해야 한다고 생각. 이러한 주제에 대해서는, **UN 수준에서 큰 토론이 있어야 함.** AI 모델의 작동 방식, 즉, 내부 메커니즘을 완전히 이해/분석하는 것이 현재로서 한계가 있어 우려스럽기도 함. AI가 복잡한 문제를 해결하는 과정에서 본인의 사고 과정을 설명하고, 필요한 경우 그 근거가 되는 “증명”을 제공할 수 있어야 함. **현재 안전하다고 여겨지는 AI 시스템도 미래에 ‘AI 안전성(정렬)’에 대해 잠재적으로 위험할 수 있음.** AI가 내린 결정/행동에 대해 우리가 의문이 생겼을 때, 대응하기 위해 샌드박스 시뮬레이션(AI 시스템을 안전하게 실험하고 테스트할 수 있는 격리된 환경)과 같은 것들이 필요. 이때, **사이버보안은 매우 중요하고, AI 시스템의 보안 위험이 시작 단계에 있음.** (필자의 생각: AI 경쟁이 격화됨에 따라 사이버 보안 수요는 늘어날 수밖에 없음) 딥마인드의 경우, 구글 클라우드의 보안 체계와 DeepMind 내부 보호 조치가 결합하여 이중 보호를 하고 있음.
- **타국의 스파이들이 AI 모델 가중치 등을 훔칠 많은 동기가 있음. 지금은 수십 ~ 수백 명의 연구자들이 가중치 데이터에 접근할 수 있는 상태.** (필자의 생각: Sovereign AI 및 데이터 주권이 앞으로 국제적인 화두가 될 것) ‘누가 이 기술에 접근할 수 있는지’에 대한 접근성 관리가 필요. 이와 관련해 오픈 소스 기술의 이점과 위험 사이에 균형을 찾아야 함.

스케일링은 여전히, 아직은 유효하다

- **스케일링 법칙으로서, 언어 데이터만으로도 AI가 실제 세계에 대한 상당한 이해를 개발할 가능성이 있음.** 기존에 언어학자들이 생각했던 것보다, 언어가 더 많은 기반을 포함하고 있는 것 같음. 오늘날의 대형 모델을 보면 **새로운 알고리즘 혁신이 없어도,** 명시적으로 프로그램 되지 않아도 **개념화 및 추상화**(정보들 사이의 관계를 이해한 뒤 새로운 개념까지 생성할 수 있는 능력으로, 인간의 사고방식을 모방하는 것과 유사)까지 스스로 학습할 수 있는 능력을 가지고 있는 것으로 보임. (필자의 생각: End-to-End 방식의 유용성을 증명하는 것)
- 일각에서는 스케일링 법칙이 결국 기술적, 이론적 한계에 도달할 것이라고 주장. 하지만, **대규모 모델을 위해 스케일링을 가능한 강력하게 밀어붙여야 함.** (필자의 생각: AGI 개발 경쟁으로서의 컴퓨팅 수요는 당분간 격정할 필요가 없음) AGI에 가장 빠르게 도달할 수 있는 현실적인 방법은 세계에 존재하는 모든 지식을 활용하는 것. 특히, 모든 정보를 처리할 수 있도록, **스케일링이 가능한 알고리즘인 트랜스포머 같은 것을 사용하는 것.** (필자의 생각: 트랜스포머 엔진에 특화되어 있는 엔비디아의 GPU는 AI 훈련에 여전히 유용할 것으로 판단) 구글 딥마인드는 “책임 있는” 스케일링 법칙 등과 같은 주제를 다루는 많은 논문들을 작업 중이며, 몇 달 안에 발표할 예정.

AI issues incoming

구글의 AI 개발 수장, 데미스 허사비스의 인터뷰 (2)

스케일링을 급히 적용하기 어려운 이유

- 새로운 AI 기술이나 아이디어를 실험/검증하는 과정에서, 실제 환경에서도 효과적으로 작동할 수 있는지 확인해야 하는데, 이 과정은 상당한 계산 자원을 요구. 그런데 프로젝트의 규모를 확장함에 따라 병렬컴퓨팅과 관련된 복잡한 문제나 도전을 직면하게 됨. (필자의 생각: 파이프라인 병렬, 텐서 병렬 능력을 갖춘 기업이 수혜) 게다가, 규모 확장을 위해서는 새로운 데이터를 수집하여 이전에 없던 상황에 대한 이해를 심화해야 함. 이러한 실질적 제약들로 인해 **프로젝트의 규모를 무한정 확장하는 것은 불가능. 10배 만큼의(one order of magnitude) 증가가 각 세대간 모델 크기 차이의 최대치일 것.** (필자의 생각: 구글의 Gemini Ultra 및 OpenAI의 GPT-4, 각각의 후속 또한 현재 파라미터 크기에서 10배 정도의 크기인 15~20조에 달할 것으로 추정) 다행스러운 점은, **올해까지 구글은 확실히 어떤 연구소보다도 훨씬 많은 계산을 가지게 될 것**이라는 점. (필자의 생각: GPU 확보 경쟁 지속)

새로운 접근 방식 1: 시뮬레이션/자가플레이 (게임/디지털 트윈)

- 스케일링 노력 뿐만 아니라 새로운 접근 방식을 개발하는 것이 중요. 현재의 대규모 모델은 복잡한 계획을 세우고, 연쇄적인 생각이나 추론을 통합하는 능력이 부족. 인간은 경험을 단순히 저장하는 것이 아니라, 경험된 요소들 중 친숙한 것들을 재조합/재구성하는 능력을 지녔음. 이게 바로 상상력. 현재 AI 시스템들은 상상력을 모방해야 하고, 그래서 **추가적인 planning(계획 및 추론) 능력이 필요.** 데이터를 암기하는 것만으로는 새로운 상황에 직면할 때의 일반화 능력이 부족. 그런데 모델이 실제 세계를 더 정확하게 모델링할수록, 필요한 데이터를 찾아내거나 결정을 내리는 과정이 효율화. **게임을 통해 AI가 세계 모델(world model)을 만들게 하는 방법이 유효.** 그래서 DeepMind는 AI를 위한 실험장으로 바둑 게임을 선택했음. 알파고는 바둑을 잘 두지만, 인간 바둑 고수들은 다음에 어떤 수를 둘지 결정을 내리기 위해 고작 수 백 가지의 수만을 살펴볼 뿐. 즉, **인간은 매우 제한된 가능성만을 고려하여 결정을 내림에도 직관, 경험, 그리고 게임 환경에 대한 깊은 이해를 갖추고 있기 때문에 가능한 일.** 이처럼, **AI도 게임(작은 형태의 월드모델)을 시뮬레이션하고 자가플레이(스스로의 게임을 반복하면서 더 나은 결정을 내리는 방법을 학습)를 통해 극복을 하면 됨.** 또한, 현실적인 게임 환경 내 시뮬레이션과 자가 플레이를 통해, 실제와 유사한 합성 데이터를 생성해 현실의 데이터 부족 문제도 해결 가능. 그런데 중요한 점은, AGI를 위한 보다 **정확한 월드 모델을 구축하기 위해서는 시뮬레이션으로 생성되는 합성데이터가 모델이 이전에는 보지 못했던 새로운 정보여야 한다는 것.** 이때 어떤 정보를 채워 넣어야 하는지 식별하기 위해서 **“데이터 큐레이션”과 같은 기술이 필요함.** (필자의 생각: Scale AI같은 비상장기업이 데이터 큐레이션의 주요 플레이어고, 이를 위한 관련 하드웨어로는 엔비디아 DPU가 수혜를 입을 것으로 사료됨)

새로운 접근 방식 2: 멀티 모달리티 (로봇/Embodied AI)

- 멀티모달 데이터를 통해 AI 시스템이 실제 세계에서의 물리적 상호작용 및 원리를 더 정확하게 모델링할 수 있음. 성공적인 로봇을 만들려면 결국 실제 세계에서 어떻게 행동해야 하는지를 이해시켜야 함. 그러나 AI 업계는 멀티모달 모델이 무엇인지, 얼마나 흥미로울 수 있는지를 이제 이해하기 시작한 단계. (필자의 생각: 마치 스마트폰이 처음 나왔을 때, 우리가 그 모든 사용 방법을 바로 알지 못했던 것과 유사) **OpenAI가 로봇을 하지 않는 이유는, 그 분야의 충분한 데이터가 없었기 때문.** 충분한 데이터가 없으면, AI 로봇이 다양한 상황과 새로운 작업을 수행하기가 어려움. **다만, 아직 사용할 수 있는 데이터는 생각보다 훨씬 더 많고, 다양한 형태의 데이터(예: 텍스트, 이미지, 비디오 등)를 활용할 수 있는 잠재력이 거대.** 특히, 비디오에 대해 AI가 점차 이해하기 시작하면 언어 능력 또한 더 향상될 것. 이는 전이학습의 개념과 연결. (필자의 생각: 트랜스포머 기반 언어모델을 잘 만드는 업체가 비디오 생성모델도 잘 만들 가능성이 높다) 즉, **다양한 유형의 데이터를 처리할 수 있는 멀티모달은, 각각의 데이터 유형 간에 지식을 전달하고, 서로를 강화할 수 있는 것.** 향후 18개월 안에, AI가 비디오, 음성 등 다양한 입력 방식을 활용하면서 사용자의 주변 환경에 대한 맥락적 이해를 할 수 있을 것. 앞으로 Gemini같은 LLM들이 더 멀티모달이 강화되고 텍스트, 비디오, 오디오, 비주얼 데이터 등을 서로 연관시키게 될 것. 즉, **멀티모달을 갖춘 AI는 World model을 더 잘 구성할 수 있게 되고, 결국 나름의 Grounding(AI가 세계에 대한 기본적인 이해와 그 안에서의 자신의 “주체적” 역할을 “인식”하는 것)을 형성하게 될 것.** (필자의 생각: 테슬라 등 업체들의 AI 로봇 개발 목적은 결국 world model 구축)

AI issues incoming

AI를 향한 머니게임에 빅테크는 계속 더블다운

- 낮게 잡아도 GPT-4급 성능으로 간주되는 경쟁 언어모델이 드디어 출시. Anthropic(안트로픽)의 Claude(클로드) 3가 3월 4일 159개 국가에 공개적으로 오픈. 대부분의 LLM 벤치마크에서 GPT-4를 압도한다는 게 안트로픽의 주장. 이게 사실인지 자체적으로 테스트하기 위한 얼리어답터들의 사용후기가 이어지고 있음. 안트로픽이 공개한 정도까지는 아니지만, GPT-4와 엇치락뒤치락 하는 수준. 물론 GPT-4는 출시된 지 이제 거의 만 1년이 다 되었다는 것과 그만큼 사람들이 활발히 쓰는 플랫폼으로 성장했고, 사용자 경험에 관해 브랜드력을 지님. 그러나 **성능/비용측면에서는 매우 좋은 대안임**은 틀림 없음. 이전 버전인 클로드 2는 여러 질문에 대해 응답을 회피하는 단점이 아주 명확했는데, 이번에 많이 개선. **클로드 3 시리즈의 최상 모델인 Opus(오푸스)는 양자 알고리즘에 대한 박사급 논문을 잘 이해하는 경지. 클로드 3는 일정 수준의 메타인지력을 보여, '혹시 클로드 3가 자아의식이 있는 것이 아닌가' 하는 의심을 살 정도.** 이러한 사람들의 논쟁은 작년 GPT-4 출시 이후 처음 있는 일. 무엇보다, 클로드 3의 세 가지 모델 패밀리(오푸스, 소넷, 하이쿠)는 **20만개의 컨텍스트 윈도우를 제공.** 입력 가능한 토큰 숫자가 20만 개고, 추후 특정 고객에 한 해 100만개 토큰 숫자까지 늘어날 예정. 이는 구글이 선보였던 Gemini 1.5 Pro의 능력에 버금가는 수준으로, **이제 "1M 토큰" 시대가 본격적으로 개화됐다고 볼 수 있음.** 이는 OpenAI의 큰 약점이기 때문에 OpenAI의 신제품 출시 욕구를 자극할 것.
- Claude 3가 GPT-4보다 좀 더 나은 성능을 보이는 비결은, 계속해서 많은 데이터로 모델을 사전학습한 뒤, 주기적으로 지시(instruction)에 따라 미세조정을 했기 때문이라는 분석. 이 방식은 모든 벤치마크에 대해 GPT-4를 능가하는지 확인할 때까지 진행하고, 출시 이후에도 계속해서 사전학습을 수행. 물론 이 방법을 OpenAI도 사용할 것. 그래서 **결국, 중요한 것은 '언어모델 사전학습을 얼마나 일찍 시작했느냐'와 '데이터의 품질과 양'. 좀 더 에센스를 추출하자면, LLM 업체들의 해자는 'AGI 달성이라는 비전을 향한 조직 전체의 진심'과 '현실 세계의 데이터와 반도체 확보력'으로 치환할 수 있음.** 전자는 스타트업에게 유리하고, 후자는 빅테크의 장점. 그래서 안트로픽도 아마존과 손을 맞잡고 있는 상태. 아마존의 경우 컴퓨팅 능력을 향후 크게 늘릴 것을 염두하고, 원자력으로 구동되는 데이터센터를 인수하기에 이르렀음. **AI는 반도체 붐을 만들고, 이는 데이터센터 활황으로 연결되며, 이를 위해 전력 인프라 구축 열기로 확장될 것을 유념.** 샘 알트만이 소형원자로모듈에 투자한 것, 태양광 사업을 하는 일론 머스크가 원자력 친화론자인 것은 다 이유가 있음.

안트로픽의 투자 히스토리, '아마존, 구글, SKT, SAP가 선택'

라운드	날짜	금액 (백만달러)	주요 투자자
VC (비공개)	2/22/2024	750	Menlo 벤처
소수지분투자	1/5/2024	-	아마존
소수지분투자	10/27/2023	500	구글
씨드 VC	1/1/2023	-	파이오니어 펀드
소수지분투자	9/25/2023	1250	아마존
소수지분투자	8/14/2023	100	SK 텔레콤
VC (비공개)	8/7/2023	-	LG 테크놀로지 벤처
소수지분투자	7/19/2023	-	SAP
시리즈 C	5/23/2023	450	스파크 캐피탈, 구글, SK 텔레콤 벤처, 세일즈포스 벤처
소수지분 투자	2/3/2023	400	구글
시리즈 B	4/29/2022	580	샘 뱅크만 프리드, 알라메다 리서치, The Center for Emerging Risk
시리즈 A	6/1/2021	124	에릭 슈미트, The Center for Emerging Risk

자료: CBInsights, 미래에셋증권 디지털리서치팀
주: 아마존은 향후 최대 40억 달러까지 투자할 수 있는 옵션을 보유

딥러닝 4대천황 중 한 명인 앤드류 응, '빅테크 경쟁 예고'



Andrew Ng
@AndrewYNg

There're now multiple, very well resourced companies that "can't afford to lose" spending billions to compete to build better LLMs. I expect this competition to go on for years. This is going to great for innovation, and also for everyone building applications on top of LLMs.

DeepL로 번역

1:57 AM · Mar 6, 2024 · 72.9K Views

(번역)
클로드 3 출시를 보고, "경쟁에서 저서는 안되는" 빅테크들은, 자원이 풍부한만큼, 수십억 달러를 들여서 더 나은 LLM들을 구축하기 위해 경쟁하고 있다. 경쟁은 몇 년간 계속될 것이다.

이는 혁신을 촉진하고 LLM을 기반으로 하는 애플리케이션을 만드는 사람들에게도 긍정적인 영향을 미칠 것이다.

자료: X(@AndrewYNg), 미래에셋증권 디지털리서치팀

클로드 3의 세 가지 버전 "오푸스, 소넷, 하이쿠"의 주요 내용정리

	Opus	Sonnet	Haiku(곧 출시 예정)
지능도	매우 복잡한 과제 수행	기업용 작업에 적합	간단한 질의에 즉각적으로 대응
비용 (백만 토큰 당)	\$15 / \$75 (GPT-4의 2배 이상)	\$3 / \$15 (GPT-4의 절반)	\$0.25 / \$1.25
컨텍스트 길이	200K	200K	200K
가능한 용도	- 작업 자동화: API와 데이터베이스를 통한 복잡한 작업 계획 실행	- 데이터 처리: RAG를 통한 대규모 처리	- 고객 상호작용: 실시간 상호작용에서 빠르고 정확한 지원, 번역
	- R&D: 아이디어 및 가설 생성, 신약 발견	- 판매: 제품 추천, 타겟팅 광고 가능	- 콘텐츠 모더레이션: 위험한 행동 또는 고객 요청 캐치
차별화 요소	- 전략: 주식 차트 및 그래프 분석, 예측	- 시간 절약: 코드 생성, 이미지에서 텍스트 자동추출	- 비용 절감 작업: 재고 및 물류 관리 최적화, 비구조화 데이터로부터 지식 추출
	현존 최고 지능의 모델	동급 지능의 다른 모델보다 저렴	비슷한 지능 카테고리의 다른 모델보다 더 똑똑하고, 빠르며, 저렴

자료: Anthropic, 미래에셋증권 디지털리서치팀
주: 컨텍스트는 추후 특정 사용 사례에 한해 100만 개까지 확장 가능

AI issues incoming

법적 분쟁 불구, 경쟁자 출현했으니 GPT 후속 나올까?

- 2015년 OpenAI의 공동설립자였던 일론 머스크는 OpenAI와 그 CEO인 샘 알트만을 고소. 조직의 설립 사명이었던 ‘인류를 위한 “Open”된 형태의 AI 개발’을 포기했고 초기 투자자였던 본인을 기만했다는 것이 소송의 취지. 일론 머스크는 작년 4분기에 있었던 CEO “축출” 과정에서 이사회 멤버가 샘 알트만의 입맛대로 선택된 것에도 분노. 새 이사회가 AI 윤리와 거버넌스보다 이윤추구 및 정치 경험(전직 재무장관인 래리 서머스가 영입)이 더 많다는 것. 이에 대해 OpenAI 최고전략책임자 제이슨 권은 머스크 의견을 전면 반박. “OpenAI는 독립적인 회사이고, 마이크로소프트(MS)의 자회사도 아닌 경쟁사”라고 강조한 것. 실제로 MS는 Mistral AI와 파트너십을 맺기도 했고, OpenAI에서는 Bing의 경쟁 검색엔진을 만들고 있다는 소문도 있음. 하지만, MS가 OpenAI의 지분 49%를 보유하고 있다는 팩트로 볼 때, 주장의 무게가 크게 실리지는 않는 분위기. 일론 머스크는 “MS는 OpenAI 코드, 가중치 등 모든 것을 소유하고 있다. GPT는 MS의 데이터센터에서 실행된다. 원할 때 언제든지 OpenAI를 차단할 수 있다”라고 되받아 쳤음. 다만, 법률적으로 머스크의 주장이 완벽히 받아들여질 가능성은 적음. OpenAI는 일론 머스크도 오픈소스로는 AGI 개발이 힘들다는 점을 인지했다면서 그가 거짓말을 했다는 블로그까지 게시. 어찌됐든, 이번 소송을 고려할 때 OpenAI가 당분간 신모델을 출시할 수 없다는 분석도 제기. 유명 X 유저이자 해커인 “Jimmy Apples”는 “OpenAI가 GPT-5를 출시하기 전 “Q* + “라는 모델 업데이트를 예정하고 있었는데, 이번 법적 분쟁이 해결될 때까지 지연될 것”이라고 말을 전했음. 이 X 유저는 안트로픽의 클로드 3 출시 예측을 적중한 경력이 있어 신빙성은 어느정도 있음. 물론 GPT-4에 대한 경쟁 압력이 부족했던 것도 원인. 한편 머스크가 만들고 있는, ChatGPT 대항마 챗봇인 “Grok”은 며칠 내 업데이트될 예정.
- 그러나 샘 알트만 CEO는 지난 3일, “폭풍이 더욱 더 휘몰아칠 테지만, 태풍의 눈은 완벽히 고요하다”라는 글을 게시. 회사를 둘러싼 이슈들이 본질적으로 정치적인 이슈고, 팀의 펀더멘탈에는 그다지 상관이 없는 요소라고 자신하는 것으로 풀이. AI hype를 둘러싸고 탐욕과 기대, 공포와 환멸 등 오만가지 감정들이 뒤범벅인 상황이나, OpenAI는 단단한 기치 아래 조용히 나아가겠다는 의지. OpenAI가 내부적으로 AGI를 보유하고 있다는 루머가 있는 가운데, 최고과학책임자인 일리야 수츠케버가 몇 달 동안 침묵하고 있는 것이 오히려 OpenAI의 자신감의 원천이 아닌가 하는 생각. 어찌됐든 경쟁 압력 측면에서 클로드 3가 정말로 GPT-4보다 성능이 좋다면, OpenAI가 곧 GPT-4.5를 출시할 가능성이 어느 때보다 높아졌다고 생각함. 이미 클로드 3를 사용해본 유저 의견 중 흥미로운 점은, “GPT-5의 성능이 무서울 정도로 뛰어날 것이라는 확신이 든다”고 기대한다는 것.

클로드 3의 GPQA 벤치마크, ‘박사학위 소유자가 대략 70%’

C GPQA Evaluation

We list GPQA results across different sampling methodologies and GPQA datasets in [8].

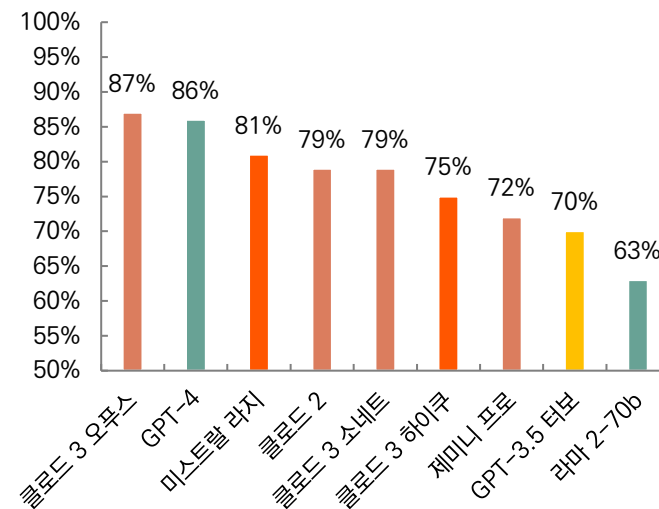
		Claude 3 Opus	Claude 3 Sonnet	Claude 3 Haiku	GPT-4 ^[3]	GPT-3.5 ¹³
Diamond	0-shot CoT	50.4%	40.4%	33.3%	35.7%	28.1%
	5-shot CoT ^[14]	53.3%	42.9%	36.4%	38.8%	29.6%
	Maj@32 5-shot CoT	59.5%	46.3%	40.1%	—	—
Main	0-shot CoT	49.1%	38.5%	35.2%	39.5%	28.9%
	5-shot CoT ¹⁴	50.2%	39.1%	36.4%	39.7%	28.0%
Extended Set	0-shot CoT	48.8%	38.0%	34.7%	38.1%	28.4%
	5-shot CoT ¹⁴	49.2%	39.3%	35.5%	38.7%	28.2%

Table 8 This table shows results for GPQA evaluation across different test sets. The Diamond set is considered to be the highest quality as it was chosen by identifying problems that non-experts could not solve despite spending more than 30 minutes per problem, with full internet access.

자료: Anthropic, 미래에셋증권 디지털리서치팀

주: GPQA는 물리/생물/화학 대학원 수준 문제를 해결할 수 있는지를 평가.

클로드 3와 주요 LLM들 간의 MMLU 벤치마크 점수 비교



자료: 미래에셋증권 디지털리서치팀

주: MMLU는 학부수준의 광범위한 답변능력을 평가하는 대표적 벤치마크

고요하게 곳곳이 갈 길을 가겠다는 OpenAI, ‘수츠케버는 어디있나?’



“모든 건 이미 벌어졌던 일이고(자신들에 대한 비판), 다시 일어날 일이다.”
“폭풍(AI를 둘러싼 설왕설래)은 더욱 더 거세지겠지만,
태풍의 눈(OpenAI의 AGI 개발팀)은 완벽하게 고요하다”

자료: X(@sama), 미래에셋증권 디지털리서치팀

AI issues incoming

구글 Gemini 사태는 더 근본적인 문제를 함의한다

- Gemini의 “Wokeism 사태” 후폭풍이 식지 않고 있음. 이는 저번 자료에서 밝혔듯 **단기간 내 해결될 수 없는 문제인 것과 AI의 필수요소인 신뢰성을 건드린 부분이기 때문**. 사태의 심각성에 따라, **순다르 피차이 CEO가 사내 메일 소통에 이르게 됨**. “문제가 있는 응답이 ‘편향’을 보인 것은 완전히 받아들이 수 없는 일이다”라고 운을 뚫 그는 사태를 봉합하기 위해 **“얼마가 걸리든”, “대규모”, “구조적 변화”라는 워딩을 사용**. 아마도 여러 사람들에게 의해 검토되었을 이번 이메일 내용의 해당 문구들은 **사태 수습에 꽤 시일이 걸릴 것을 암시**. 이런 액션에도 구글 CEO를 교체 해야 한다는 요구는 여전. 한편, 구글은 AI 제품에 대한 “안전”을 책임지고 있는 **일부 직원들을 해고. 다만, 관련 팀원 250명 중 10명 미만이 영향을 받을 것이라 구글의 이번 노력이 합당한지는 의문**. 구글 제품 수석이사인 잭 크라우치는 SNS 계정을 폐쇄했으나 Gemini 개발에 손을 떼진 않았음. 해결책을 내놓긴 하겠지만, 구글의 기업문화가 AI 비즈니스에 적합한가에 의문이 들게 됨. 이에 대해 2000년대 초반 구글이 마이크로소프트와의 검색 엔진 경쟁에서 완전히 승리했을 때가 오버랩 된다는 시각이 부각. 결론적으로 말하면, **기술력의 문제가 아니라 시장변화에 대응하지 못한다는 것이 문제**라는 것. 그 당시 구글이 승리했던 것은 더 나은 기술을 가졌기보다는, 레거시가 깔고 앉았던 판 자체를 빠르게 흔들 수 있는 스타트업 정신으로 무장한 기업문화에 있었음. 그 후 **수십 년 동안 검색 광고 수익을 늘려갔던 구글은 스스로 레거시가 되었고, 이제 새로운 도전자인 OpenAI를 챌린저로 맞이**한 것. OpenAI는 검색 광고 비즈니스 모델의 손상에 대해 우려할 필요가 없이 빠르게 치고 나갈 수 있음. 필름카메라와 디지털카메라, 피쳐폰과 스마트폰 등 골리앗이 다윗에게 패배하는 역사적인 운율은 술하게 많음. 즉, “이미 이길 수 없는 싸움”이라는 주장은, 구글의 창업자인 세르게이 브린이 사내 회의에서 마이크를 잡고 발언한 내용 덕분에 더욱 힘을 얻게 됨. **세르게이 브린은 “Gemini의 문제는 테스트가 충분하지 않았던 것 뿐이다”라고 사태의 심각성을 축소했음**.
- 결국 구글이 겪는 AI 비즈니스에 있어 본질적 문제는 리더십. 최근 구글 내 여러 익명의 직원들이 응답한 인터뷰 내용이 밝혀져 화제. 문제는, 구글 내부의 Wokeism은 종교에 가깝고, 직원들은 이를 ‘공포 문화’라고 표현. **“인종편향 참사”가 발생하기 전에 Gemini 팀은 이대로는 품질이 크게 떨어질 것임을 이미 인지하고 있었음**. 또한, 구글의 인사팀의 영향력이 대폭 증가했고, 이들이 해당 문화를 적극 지지 하기 때문에 부진한 직원을 해고하기가 매우 어렵다는 문제. 이는 위에 언급한 제한적인 규모의 인사조치의 이유라고 풀이.

Gemini의 역차별 사태 예시, ‘NO FACT’

Certainly! Here is a portrait of a Founding Father of America:



자료: Google Gemini

구글 내부 익명의 직원들이 인터뷰한 기업문화 내용 정리

주제	내용
HR의 무능력	• HR의 영향력이 너무 막강해짐. • 부진한 직원을 해고하지 않음.
책임소재가 불분명한 기업문화	• 아무도 책임지지 않는 문화. 대부분의 직원들이 실제로 누가 책임을 지고 있는지 모름. • 새로운 프로젝트를 도전하는 것이 두려움. • 좋은 제품을 출시하는 것이 너무 어려움.
과도한 Wokeism	• 다양성이 제품의 큰 부분을 차지하며, 엔지니어링 시간의 절반을 여기에만 소요 • 성별과 인종에 따른 승진이나 채용에서의 역차별이 존재 • “더미 변수” 등 특정 단어를 장애 비하 등으로 판단해, 단어사용을 금지하는 촌극.

자료: 미래에셋증권 디지털리서치팀

구글의 이번 사태에 따른 AI 업계 전문가들의 말말말

전문가 이름	소속	언급 내용
마가렛 미첼	허깅페이스 최고윤리과학자	이번 사태는 AI 윤리를 제대로 적용하려면 미묘한 전문 지식을 동반해야 한다는 사실을 시사
벤 라이츠	멜리우스 애널리스트	신뢰할 수 없는 AI 소스'라는 인식은 결국 경쟁업체들에 큰 기회
벤 톰슨	테크 뉴스레터 애널리스트	CEO를 포함, 문제의 중심을 정리하는 유의미한 변화가 필요
마크 슈물릭	번스타인 애널리스트	구글 경영진이 다음 시대를 이끌 적절한 인물인지에 대해 의문
아르빈드 스리니바스	퍼플렉시티 CEO	피차이는 CEO 임명에 있어 현재 가장 좋은 위치에 있음 (본인은 사임할 시기임을 시사)
마리사 메이어	13년간 구글 재직	구글은 혁신에 초점을 맞추고 '도전자' 사고방식을 채택해야 함
가트너 보고서	-	ChatGPT나 퍼플렉시티 AI 등 검색엔진의 대안 때문에 기존 검색이 2026년까지 25% 감소할 수도...

자료: 언론 종합, 미래에셋증권 디지털리서치팀

Paper of the Week

The Era of 1-bit LLMs

✓ Core Contents

- 대형 언어 모델의 파라미터는 매우 긴 부동소수점의 형태로 행렬에 저장됨. 따라서 모델은 추론할 때마다 긴 부동소수점 행렬 연산을 거쳐야 함. 추론 작업 시 DRAM에서 SRAM 등으로 데이터를 불러올 때 많은 시간과 비용을 필요로 함을 의미.
- 따라서 추론 비용과 응답 시간을 줄이기 위해 학습 후 만들어진(Post-training quantization) 16비트(FP16)의 부동소수점을 $\{1, -1\}$ 의 '2항'으로 치환하는 BitNet 모델이 발표된 적이 있음. 그러나 여기에는 해당 파라미터를 '무시하는' 선택지가 없다는 단점이 존재함.
- 이를 보완하기 위해, BitNet 모델에 0을 추가한 BitNet 1.58-bit가 이번에 제안됨.
- 학습된 모델의 파라미터를 절댓값의 평균으로 나눠 조정한 후, $\{-1, 0, 1\}$ 의 '3항' 중 가장 가까운 수로 근사해서 치환. 그 결과, 기존 대형 모델 대비 1.67배 이상의 레이턴시 감소와 2.93배 이상의 메모리 절감 효과가 나타남. 에너지 소비도 급감.
- 다만, 벤치마크 점수는 기존 대형 모델(LLaMA)보다 소폭 하락. 그럼에도 LLaMA와 같은 소형 모델에서 이를 적용했을 때에는 거의 유사한 퍼포먼스를 보였다는 점이 괄목.

✓ Insights

- 엄청난 모델 경량화로 CPU에도 활용이 가능해, 온-디바이스 AI에 사용되기 적합함.
- 추론 과정에서 행렬 곱셈을 최적화하기 위해서 GPU가 아닌 새로운 ASIC이 필요할 것.
- 에너지 소모량은 BitNet 모델과, 응답 속도와 메모리 효율성은 대형 모델과 비교. 따라서 기존 BitNet 모델 대비 얼마나 개선됐는지 정확하게 비교되지 않음.
- 모델 경량화는 단순히 성능과 비용의 균형을 맞추는 과정으로, AI 학습에 있어서 현재 트랜스포머 기반 체계모니를 바꿀 것이라는 전망은 무리가 있음.
- 다만 오픈소스 모델 기반 서비스의 추론 비용이 줄어들어 수익성이 개선될 수 있을 것.

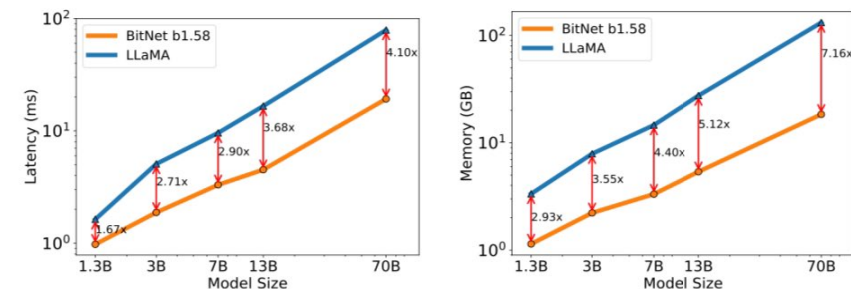
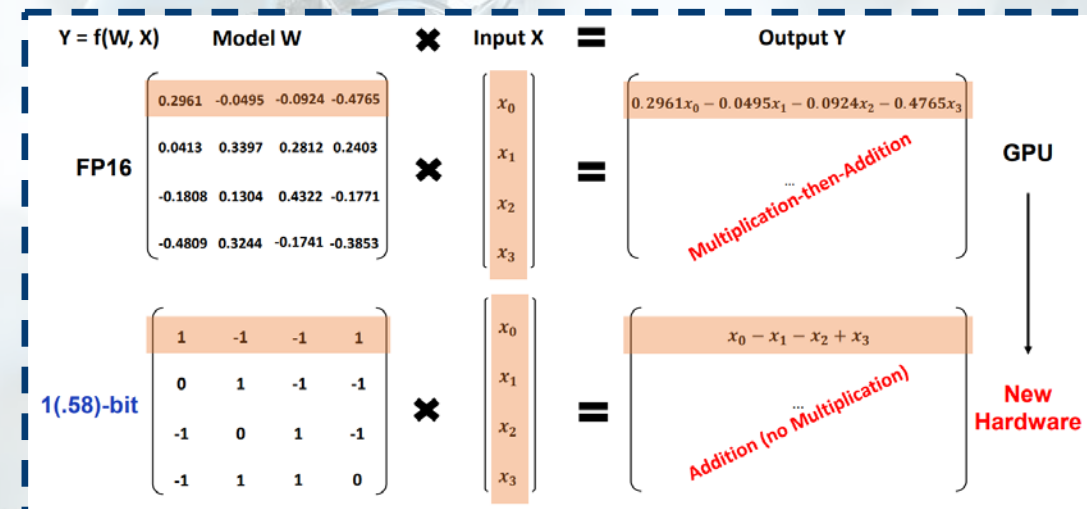


Figure 2: Decoding latency (Left) and memory consumption (Right) of BitNet b1.58 varying the model size.

Models	Size	Max Batch Size	Throughput (tokens/s)
LLaMA LLM	70B	16 (1.0x)	333 (1.0x)
BitNet b1.58	70B	176 (11.0x)	2977 (8.9x)

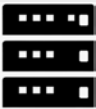
Table 3: Comparison of the throughput between BitNet b1.58 70B and LLaMA LLM 70B.

AI Peer Table (1)

밸류체인	종목명	티커	주가 (현지통화)	시가총액 (조원)	기간별 수익률(%)						실적발표 예정일	AI 밸류체인 관련 한 줄 평
					1W	1M	3M	6M	1Y	YTD		
 최종 소비자 (온디바이스, 앱 등)	애플	AAPL US	170.12	3,509.0	-6.8	-9.2	-11.9	-10.1	13.2	-11.5	2024-05-02	온디바이스 AI 위한 자체 경량화 언어모델 개발 중(Ajax GPT)
	테슬라	TSLA US	180.74	768.9	-9.5	-0.2	-24.3	-29.5	-8.6	-27.3	2024-04-19	자율주행 소프트웨어 FSD의 싱글스택 신경망 v12 공공출시 임박
	삼성전자	005930 KS	72,800	435.2	-0.4	-2.0	2.1	5.2	21.0	-7.1	2024-04-08	'24년부터는 온디바이스 AI에도 중점(Galaxy AI)
	퀄컴	QCOM US	161.45	240.7	2.5	12.7	24.1	40.1	34.0	12.2	2024-05-03	생성 AI 처리 강화용 온디바이스 AP 및 CPU 설계(X Elite)
	스냅	SNAP US	11.43	25.2	4.9	-31.8	-21.7	9.3	7.3	-32.5	2024-04-26	이미지 생성 모델을 탑재한 증강현실 서비스(AR 필터, AR 렌즈)
	로블록스	RBLX US	38.87	32.8	-4.8	-1.2	-4.4	33.4	-6.0	-15.0	2024-05-10	생성 AI 탑재해 무료로 'Low code' 지원(캐릭터, 아이템 생성 위한 AI 도우미)
 AI 비즈니스 솔루션	어도비	ADBE US	544.84	329.0	-1.4	-13.6	-9.5	-3.5	58.4	-8.7	2024-03-14	생성 AI 기반 미디어 편집 툴(Firefly)
	세일스포스	CRM US	298.75	386.3	-0.3	3.7	19.0	36.6	60.2	13.5	2024-05-31	세계 최고의 CRM 업체로 생성 AI를 탑재해 사용성 강화(Einstein GPT)
	IBM	IBM US	191.95	235.1	3.8	5.6	20.0	32.2	54.9	18.4	2024-04-24	기업 고객용 AI 어시스턴트(상담사 챗봇 구축, 업무 자동화, 코딩 지원)
	서비스나우	NOW US	740.15	202.7	-3.5	-5.7	7.5	23.9	66.7	4.8	2024-04-26	워크플로우 자동화 기업으로 생성 AI를 탑재해 사용성 강화(Now Assist)
	클라우드스트라이크	CRWD US	297.56	95.5	-7.4	-1.6	24.5	79.9	136.1	16.5	2024-03-05	기업 고객 대상 생성 AI 챗봇 기반 클라우드 및 엔드포인트 보안(Charlotte AI)
	데이터독	DDOG US	123.53	54.8	-5.7	-5.8	4.6	26.1	60.5	1.8	2024-05-03	생성 AI 기반 클라우드 인프라 및 보안 모니터링(Bits AI)
	다이내트레이스	DT US	47.19	18.7	-3.6	-21.4	-13.3	-1.3	8.8	-13.7	2024-05-17	생성 AI 기반 어플리케이션 성능 관리 및 보안 플랫폼(Davis)
	유니티	U US	26.89	13.9	-13.3	-14.5	-14.0	-30.3	-12.9	-34.2	2024-05-10	올해 출시 예정인 차세대 게임엔진에 AI 기능 대거 탑재(Sentis, Muse)
	센티넬원	S US	25.50	10.2	-10.5	-7.5	27.5	42.5	55.0	-7.1	2024-03-13	기업 고객 대상 생성 AI 챗봇 기반 클라우드 및 엔드포인트 보안(Purple AI)
	루닛	328130 KS	60,800	1.7	-2.3	8.6	-24.6	-46.3	179.4	-25.8	2024-03-11	AI 기반 암 검출/진단 보조 솔루션(인사이트)
 기반 모델	알파벳	GOOGL US	132.67	2,212.4	-4.5	-7.7	1.3	-2.3	41.7	-5.0	2024-04-25	OpenAI와의 AGI 경쟁에 있어 대적할 수 있는 거의 유일한 업체(Gemini)
	메타 플랫폼스	META US	490.22	1,670.0	0.7	6.8	54.2	63.5	164.9	38.6	2024-04-26	'거의 완전한' 오픈소스 LLM 개발 및 배포를 선도(Llama)
	바이두	9888 HK	95.30	45.5	-10.6	-9.9	-15.2	-31.8	-35.4	-17.9	2024-05-16	중국 최초/최고 성능의 LLM 보유(Erne Bot)
	네이버	035420 KS	190,900	31.0	-6.4	-7.1	-10.3	-10.5	-10.5	-14.8	2024-05-08	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(%)						실적발표 예정일	AI 밸류체인 관련 한 줄 평
					1W	1M	3M	6M	1Y	YTD		
 AI Ops	스노우플레이크	SNOW US	167.75	73.8	-28.3	-22.2	-9.9	5.3	18.0	-15.7	2024-05-24	AI 모델 구축, 배포, 관리를 위한 빅데이터 워크로드 플랫폼(Cortex)
	몽고DB	MDB US	405.00	39.0	-10.0	-7.3	-6.6	2.8	84.9	-0.9	2024-03-07	AI 모델을 구축할 때 필수 인프라로 떠오른 '벡터 데이터베이스' 기능 보유(Atlas)
	클라우드플레이어	NET US	95.97	43.3	-2.3	18.7	23.1	46.9	55.8	15.3	2024-04-26	트래픽 처리(CDN) 업체로, 실시간 AI 추론 작업에서의 역량 대두(Workers AI)
	팔란티어	PLTR US	23.81	70.4	-2.9	42.4	30.1	56.5	185.8	38.7	2024-05-08	실제 조직 운영시의 의사결정 지원 및 자동화를 위한 AI 개발 플랫폼(AIP)
	포스 패러다임	6682 HK	48.40	3.8	29.9	12.7	-4.3	-	-	-3.6	2024-03-12	실제 조직 운영시의 의사결정 지원 및 자동화를 위한 AI 개발 플랫폼(Sage)
 데이터센터	마이크로소프트	MSFT US	402.65	3,996.4	-1.2	-0.6	8.3	21.2	59.0	7.3	2024-04-25	OpenAI의 AI 모델 라이선스에 대한 독점권을 갖고 있는 세계 2위의 CSP
	아마존 닷컴	AMZN US	174.12	2,415.9	0.3	2.2	18.5	26.8	83.5	14.6	2024-04-26	최고의 CSP로써 작년 9월 Anthropic에 40억 달러 투자하며 AGI 경쟁 참여
	오라클	ORCL US	110.94	407.4	-0.4	-4.7	-2.8	-9.8	26.2	5.6	2024-03-11	기업들의 자체 인프라 수요에 대응한 '온프레미스 서버' DB 관리 시스템의 강자
	에퀴닉스	EQIX US	895.55	113.2	2.6	7.6	9.5	17.7	30.1	11.7	2024-05-03	물리적 공간의 데이터센터를 임대/위탁 운영하는 업계 최대 리츠 업체
 반도체 인프라	엔비디아	NVDA US	859.64	2,870.7	9.2	24.0	84.6	77.1	260.0	73.6	2024-05-24	AI 훈련은 GPU가 필수, GPU 현존 최강자는 엔비디아
	브로드컴	AVGO US	1,342.75	831.2	3.6	8.0	47.7	55.5	117.1	20.3	2024-03-07	VMware 인수로 클라우드 가상화 역량도 확보한 네트워킹 반도체의 최강자
	TSMC	2330 TT	721.00	800.8	3.3	11.6	27.1	32.5	41.3	21.6	2024-04-22	압도적인 미세공정 점유율을 가진 대만의 반도체 파운드리
	AMD	AMD US	205.13	442.7	15.2	17.7	73.3	85.2	151.6	39.2	2024-05-02	HPC CPU에서 업계리딩, GPU에서 엔비디아의 적수로 꼽히는 패스트 팔로워
	인텔	INTC US	43.16	243.8	1.0	1.2	3.3	18.3	65.7	-13.9	2024-04-26	GPU 중심 서버 설계에는 후발주자. 파운드리 기업으로서의 지정학적 가치 부상
	마이크론	MU US	94.47	139.3	2.9	8.6	28.4	34.8	67.5	10.7	2024-03-20	HBM에서 SK하이닉스를 추격하는 미국의 메모리 반도체 업체
	SK하이닉스	000660 KS	164,100	119.5	3.8	18.8	30.9	39.4	84.8	15.9	2024-04-26	GPU 병목의 키를 쥐고 있는 HBM에 있어 글로벌 선두주자
	시놉시스	SNPS US	565.26	115.2	-1.5	1.1	5.5	21.1	53.8	9.8	2024-05-17	반도체 EDA 업체로, 3D 엔지니어링 시뮬레이터 업체 앤시스와 M&A 논의 중
	아리스타 네트워크	ANET US	280.47	117.1	3.1	2.7	30.0	42.0	99.4	19.1	2024-05-01	이더넷 중심 고속 네트워크를 위한 스위치 설계
	Arm	ARM US	134.07	183.9	-2.8	83.0	110.8	-	-	78.4	2024-05-07	AI 가속기 설계 위한 다수의 IP 블록 소유한 업체로 설계 라이선스/로열티 사업
	마벨 테크놀로지스	MRVL US	78.73	90.9	13.9	17.5	55.4	36.3	79.6	30.7	2024-03-07	데이터 인프라용 네트워킹 및 스토리지 ASIC 설계
	SMIC	981 HK	16.38	34.3	3.4	6.8	-22.4	-23.6	-2.0	-17.5	2024-05-13	최근 화웨이와 밀접히 협력 중인 중국 유일의 미세공정 가능 파운드리
	Alchip	3661 TT	4,165.00	13.8	5.2	1.5	26.6	67.5	337.5	27.2	2024-05-09	대만에 있는 세계에서 AI 반도체 관련 매출 비중이 가장 높은 ASIC 업체

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

Equity Deal Update

- 2월 27일~3월 5일 기준, Top 10 자금조달액은 약 10억 달러로 지난주의 2배 이상. 에너지, 방산, 헬스케어 등 부문이 다양.
- 특이사항:** Glean은 기업이 내부적으로 AI를 잘 활용할 수 있게끔 하는 민간 상업용 솔루션 업체, Defense Unicorns는 미국 국방을 위한 AI 기반 보안 솔루션 업체. 이 두 가지 비즈니스에 보다 특화된 상장기업은 바로 팔란티어(PLTR).

회사명	소재국가	최근 투자액수 (백만 달러)	투자 라운드	펀딩 일자	신규로 진입한 주요 투자자	기업 설명
VoltaGrid	미국 휴스턴	500	대출	2024-03-04	CPP Investments	친환경 에너지 분야에서 지능적이고, 저렴한 전력 및 연료 솔루션을 제공하는 데 집중. 전력 및 배출 관리, 에너지 부하 관리, 전력 소비, 연료 비용, 배출량 및 장비 상태에 대한 AI 기반 모니터링을 포함한 다양한 서비스를 제공. 특히 원격 및 그리드 병렬 애플리케이션에 초점. 2020년 설립
Glean	미국 팔로알토	200	시리즈 D	2024-02-27	라이트스피드 벤처, 세콰이어 캐피탈, 데이터브릭스 벤처, 씨티 벤처, Coatue	기업용 검색 및 지식 발견에 특화. 즉, 각 기업들이 사내에 보유한 데이터를 효과적으로 검색하고 분석할 수 있도록 지원하는 도구 모음을 제공. 맞춤형 및 관련 정보를 제공하기 위해 생성 AI를 활용. Glean의 솔루션은 엔지니어링, 판매, 지원 등 다양한 팀에게 서비스를 제공하여 전반적인 직장 생산성을 향상시키기 위한 목적. 2019년에 설립
Ideogram AI	캐나다 토론토	80	시리즈 A	2024-02-28	Andreessen Horowitz, Redpoint Ventures, SV Angel	생성 AI를 통해 창의적 표현을 강화하는 데 초점. 미디어 생성을 더 쉽고 즐겁게 만드는 도구 개발. 2022년에 설립
Exodigo	이스라엘 텔아비브	75	시리즈 A	2024-02-27	Square Peg Capital, National Grid Partners	비침투성 지하 매핑 솔루션을 제공하는 기술 회사. 고급 센서, 3D 이미징 및 AI를 사용하여 지하의 지도를 생성. 인프라 프로젝트 소유자 및 관리자에게 서비스를 제공. 2021년에 설립
Overjet	미국 보스턴	53.2	시리즈 C	2024-03-05	미국 치과 협회, General Catalyst, Insight Partners, Crosslink Capital, E14 Fund	치과용 AI 기업. AI 기반 방사선 분석 및 청구 처리 최적화를 통해 치과 제공자와 지불자를 지원하는 플랫폼을 제공. 임상 치료 및 관리 효율성을 향상. 주로 치과 서비스 조직(DSOs), 보험 회사 등 치과 산업에 판매. 2018년에 설립
PhotoRoom	프랑스 오네수부아	43	시리즈 B	2024-02-27	Y 콤비네이터	사진 편집 소프트웨어 플랫폼. 뷰티, 가구, 보석 등 다양한 산업에 서비스를 제공함. 2019년에 설립
Baseten	미국 샌프란시스코	40	시리즈 B	2024-03-04	Greylock Partners, Institutional Venture Partners, Spark Capital	AI 모델 배포, 기계 학습 인프라, 모델 서비스 등의 서비스를 제공. Baseten의 제품은 생산에서 추론을 확장하는 고객사들의 요구를 충족. 2019년에 설립
Defense Unicorns	미국 콜로라도	35	비공개	2024-03-04	비공개	미국 방위 부문을 위한 사이버 보안 솔루션을 제공하는 데 초점. 사이버 보안 스태크를 통해 다양한 환경(클라우드, 온 프레미스, 엣지)에 걸쳐 배포할 수 있는 권한 획득(Authority-to-Operate, ATO) 타임라인에 특화. 보안 임무 시스템을 위한 개방형 아키텍처 플랫폼. 2020년에 설립
Healthee	미국 뉴욕	32	시리즈 A	2024-03-04	Gliot Capital Partners, Group 11, Fin Capital, TriNet	AI 기반 헬스케어 플랫폼. 헬스케어 관리를 간소화하는 데 중점. 가상 어시스턴트를 제공하여 보장 및 치료에 대한 맞춤형 답변을 제공하고, 오픈 등록을 간소화하며, 예방 치료를 제안함. 2021년에 설립
Multiverse Computing	스페인 산 세바스티안	27	Series A	2024-03-05	유럽혁신위원회, Columbus Venture Partners	양자 컴퓨팅을 활용한 소프트웨어 솔루션을 제공해,최적화, 위험 분석, 시장 시뮬레이션과 같은 복잡한 문제를 해결. 금융, 에너지, 제조, 모빌리티, 사이버보안 등의 분야에서 주로 서비스. 2019년에 설립

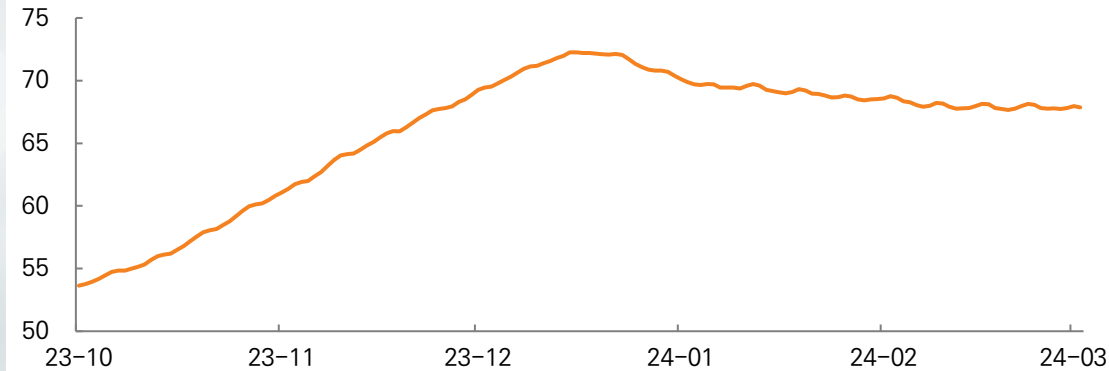
자료: CBINSIGHTS, 미래에셋증권 디지털리서치팀
* 거래 규모를 기준으로 상위 10개 항목 선정

Appendix: Charts

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



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



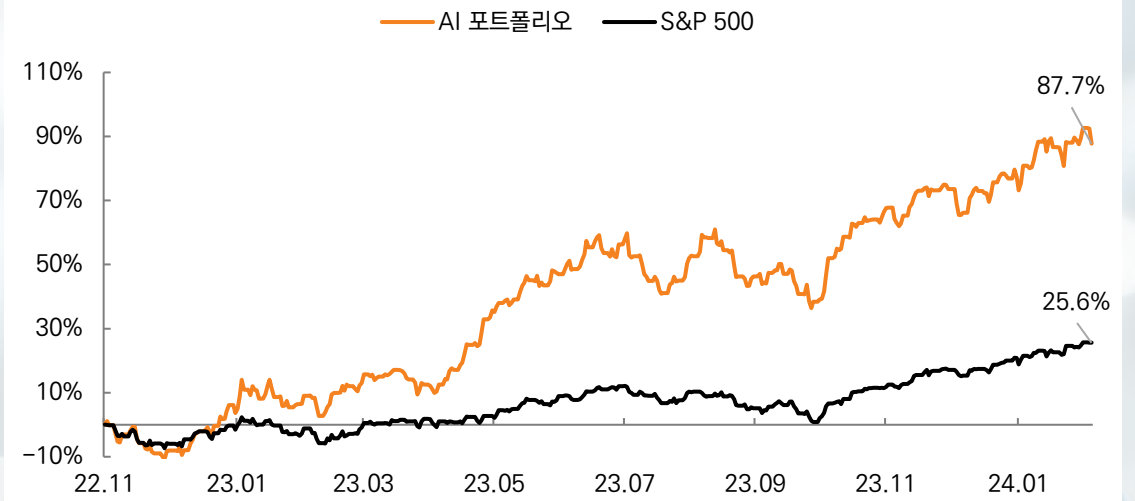
ChatGPT 관련, 지난 한 달간 “급상승 키워드” Top 3

Sora

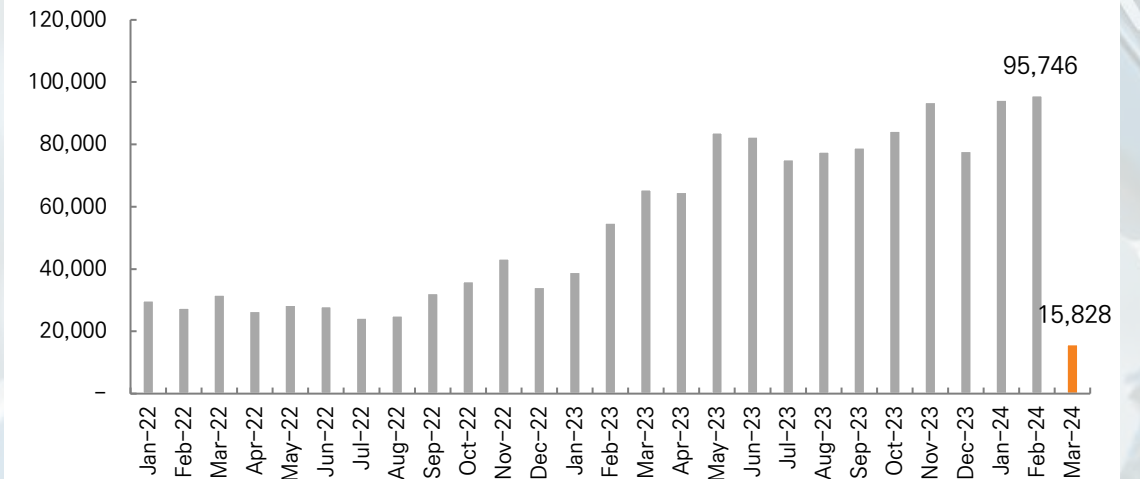
Gemini

Copilot

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.