2024. 04.24
그림: Midjourney (메타의 Llama-3, '데이터가 왕이다')

AI Weekly, 특이점이 온다 #13

W-148 to A.G.I.

(AGI's ETA: 2027.03.02)

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

- **메타의 Llama-3 출시! '데이터가 왕이고, 훈련은 계속된다'**

메타의 LlaMA-3 모델, 오픈소스 진영에서 GPT-4 수준의 성능을 보이며 AI 헤게모니 변화의 조짐을 보임. 적은 파라미터로 최적 배수 이상의 방대한 데이터를 학습해 기존의 AI 학습 한계에 대한 우려를 불식시켰음. 또한 연말 출시 예정인 405B 모델은 GPT-4를 능가할 것으로 기대. 다만 OpenAI를 따라잡으려면 언어 모델을 넘어 멀티모달 모델의 개발이 필수적일 것으로 보임.

- **TSMC 실적, '꽁꽁 얼어 붙은 핸폰 위로 AI가 걸어다닙니다'**

TSMC, 1분기 실적이 AI 수요 덕분에 호조세 보엿으며, 향후 5년간 AI 프로세서 매출이 연평균 50% 성장해 2028년에는 전체 매출의 20% 이상을 차지할 것으로 전망. AI 가속기 수요에 대응하기 위해 CoWoS 패키징 용량을 대폭 증가시키고 있으며, 에너지 효율 향상을 위해 모든 고객사가 3nm 및 2nm로의 전환을 추진 중. 온 디바이스 AI 확대로 스마트폰 및 PC 교체 주기가 빨라질 것이라는 전망도 제기됨.

- **마크 저커버그 인터뷰: 메타의 화두는? 'AGI (Artificial General Intelligence)'**

메타, LLaMA-3를 발표하며 15조 개의 토큰 학습, 70B/405B 파라미터 등 대규모 모델 개발에 박차 가하는 한편, 오픈소스 공개를 통해 개발 생태계 활성화에도 기여. 검색, 이미지 생성 등 자사 서비스에 활용 중이며, 클라우드 업체 등과 협력해 AI 비즈니스 모델도 모색 중. AGI 개발을 위해서는 멀티모달, 3D, 감정 이해 등이 중요한 요소가 될 것. AI의 잠재적 위험성도 인지하고 있으나, 단기적으로는 유해 콘텐츠 차단 등에 더 집중할 계획. AI 시대의 새로운 기회로 '오픈소스'에 주목. 커뮤니티 기여를 통해 기술 발전을 가속화하고, 대형 기업의 AI 독점을 방지하는 효과도 기대. 다만 AI 모델 자체가 제품이 되는 시점이 오면 비즈니스 모델 재검토가 필요할 것. 궁극적으로 AI는 인류 문명에 획기적 전환점이 될 것이나, 에너지 등 물리적 제약으로 폭발적 발전보다는 점진적 과정을 거칠 것.

- **Paper of the week: "VASA-1", Lifelike Audio-Driven Talking Faces Generated in Real Time(Microsoft Research)**

마이크로소프트, 사람의 이미지와 음성을 바탕으로 음성에 맞게 움직이는 사람의 얼굴을 생성하는 VASA-1 모델 공개. 얼굴의 여러 특징을 구분해 잠재 공간에 표현하는 방식을 채택하고, 디퓨전 트랜스포머 기반 모델을 통해 자연스러운 영상을 생성하는 데 성공. 기존 모델 대비 생성 속도가 빠르고 고개 및 시선 처리가 자연스러우나, 얼굴 외의 부분까지 확장하지는 못했다는 점이 한계.

일	월	화	수	목	금	토
21	22	23 (TSLA) 실적발표	24 (NOW) 실적발표 (META) 실적발표 (IBM) 실적발표 Data Innovation Summit(~25)	25 (SK하이닉스) 실적발표 (MSFT) 실적발표 (INTC) 실적발표 (GOOGL) 실적발표 (SNAP) 실적발표	26	27
28	29	30 (AMZN) 실적발표 (삼성전자) 실적발표 (AMD) 실적발표	1 (QCOM) 실적발표	2 (NET) 실적발표 (AAPL) 실적발표 AISTATS 2024(~4)	3 (EQIX) 실적발표 (네이버) 실적발표	4
5	6 (PLTR) 실적발표 AAMAS 2024(~10)	7 (DDOG) 실적발표 (ANET) 실적발표 ICLR 2024(~11)	8 (ARM) 실적발표 Middle East Banking AI& Analytics Summit 2024	9 (Alchip) 실적발표 (TSMC) 실적발표 (RBLX) 실적발표 (U) 실적발표	10 (루닛) 실적발표	11 CHI 2024(~16)

자료: Bloomberg, 외신, 미래에셋증권 디지털리서치팀

AI Insight

마크 주커버그 "메타의 화두는 'AGI'"

4월 19일, @Dwarkesh Patel의 팟캐스트에 출연한 마크 주커버그의 인터뷰

Q. 메타의 AI 비즈니스 발전 방향?

- 자사 모델을 클라우드 업체들이 서비스로서 제공하게 하고, 매출의 일부를 배분 받는 라이선스 구조이며, **대형 CSP의 사용을 제한한 것은 사용을 막기 위한 것이 아니라, 수익 공유 방식에 대한 논의가 선행되어야 하기 때문.** 따라서 마이크로소프트 Azure나 AWS가 Llama를 재판매할 계획이라면 메타와 수익을 공유해야 할 것. 메타, 타사의 AI 모델이 좋은 성과를 거두는 경우, 이를 활용해 제품 경쟁력을 제고. 또한 **구글과 빙(Bing)의 검색 기능을 통합해 앱 상단에 통합 검색창을 두고 질의 결과를 실시간으로 받아볼 수 있게 할 것**이라 언급. 이는 Llama-2에서는 제한적이었고 수작업이 필요한 기능이었으나, Llama-3에서는 해당 기능을 완전히 통합할 예정.

Q. AI 투자는 어떻게 진행되고 있나?

- AI와 메타버스에 천억 달러에 달하는 막대한 투자를 하는 것은, 사전에 모든 것을 확신할 수는 없으나 성공할 경우 우수한 수익을 거둘 것이라는 명확한 생각 때문. **2022년 CapEx를 크게 늘린 것은** Reels 등에서 팔로우하지 않는 사람의 컨텐츠를 추천할 때 수억 개의 후보군이 존재하게 되므로, **미래의 서비스를 위해 필요한 GPU를 선제적으로 주문했던 것**. FAIR(Facebook AI Research)를 통해 최근 LlaMA 모델의 학습 과정을 공개했는데, 최적 데이터보다 더 많은 데이터를 사용했음. 이는 1) 오픈소스 공개를 위한 퀄리티 확보 2) 다양한 도메인 내 학습 가능성 3)최적점을 넘어도 학습 효과가 지속되는 AI 모델의 잠재력을 시험하기 위한 시도 등으로 생각해볼 수 있음. 이 외에 자선단체 CZI를 통해 과학 분야의 AI 활용에도 힘쓰고 있음.

Q. GPU를 비롯한 AI 인프라 확충 목표는?

- 2022년 당시 대량의 H100 GPU를 구매한 것은 LLM 학습이 아닌 Reels 등의 컨텐츠 훈련을 위한 것이었음. 현재는 **22,000~24,000대의 GPU 클러스터 여러 개를 운영 중**이며, 추천 시스템 및 광고에서의 대형 모델 학습에 사용됨. 메타의 **전체 ML 인프라는 35만 대 수준으로, 이를 연말까지 유지할 것.** 추천 시스템의 추론 효율 향상을 위해 자체 AI 가속칩을 개발 중에 있으며, 장기적으로 GPU를 대체할 가능성이 있다고 생각. 향후 AI 기술의 병목은 GPU 생산, 전력 등 물리적 한계가 될 가능성이 높음. 특히 에너지 문제는 정부 규제와 맞물려 해소가 쉽지 않을 것. 현재 데이터센터는 대부분 50~150MW 수준의 전력을 소모하나, 미래에는 1GW급 데이터센터도 등장할 수 있으며, 이를 위한 에너지 인프라에는 상당한 시간이 소요될 것으로 예상.

Q. AGI에 대해 어떻게 준비하고 있나?

- **10년 전 FAIR를 설립하며 AGI 개발을 위한 기반을 마련**해 왔음. 당시 목적은 자사 서비스 개선이었으나, 최근 대화형 AI, 이미지 생성 AI 등이 주목받으며 AGI의 필요성을 절감. 특히 코딩이 다양한 분야에서의 AI 성능 향상에 기여한다는 사실을 인지하고 본격적인 AGI 개발에 착수. **AGI는 단일 기술이 아닌 다양한 능력의 조합이며, 단계적 발전을 통해 다양한 능력이 더해지는 방식으로 발전할 것**으로 예상. 이를 위해 특히 멀티모달, 3D와 더불어 인간의 뇌의 상당 부분이 감정과 표현에 대한 이해에 할애되고 있어, 이를 별도의 모달리티로 연구할 필요성 존재.

Q. AI의 향후 발전 방향에 대한 생각은?

- **AI가 스스로 합성 데이터를 만들어 학습에 사용하는 것이 미래의 학습 방식이 될 수 있음**. 현 시점에서는 합성 데이터의 생성을 추론에 가깝다고 보지만, 이렇게 생성된 데이터가 학습에 사용된다면 넓은 의미의 학습이라고 볼 수 있을 것. 향후 AI 모델은 단순 질의응답이 아닌 사용자가 지시한 복잡한 임무를 AI가 알아서 처리하는 방식으로 진화할 것. 현재 추세로 10~20년 내에 AI가 대부분의 일에서 인간을 넘어설 것으로 예상하나, 앞서 언급했듯 에너지와 같은 물리적 제약으로 폭발적 발전에는 한계가 있을 것으로 전망. **AI가 소수에 의해 잘못 사용될 경우에 대비하기 위해 AI가 널리 퍼질 수 있도록 오픈소스화하는 것이 중요.** 커뮤니티를 통해 모델 공급사의 기술도 발전할 수 있고, 호환성 확보로 벤더 락인을 피할 수 있다는 장점도 있음.

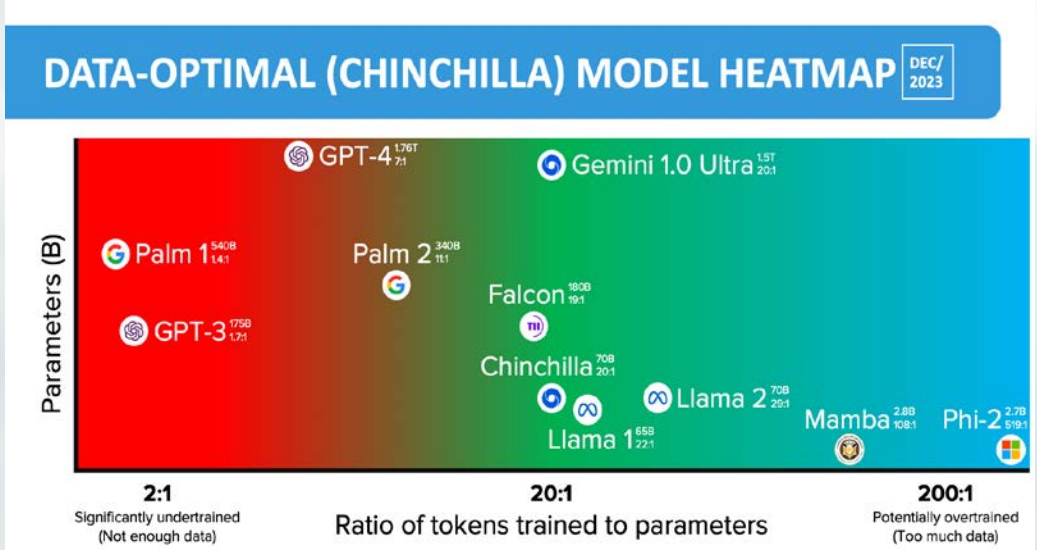
자료: Youtube(@Dwarkesh Patel), 미래에셋증권 디지털리서치팀

AI issues incoming

메타의 Llama-3 출시! '데이터가 왕이고, 훈련은 계속된다'

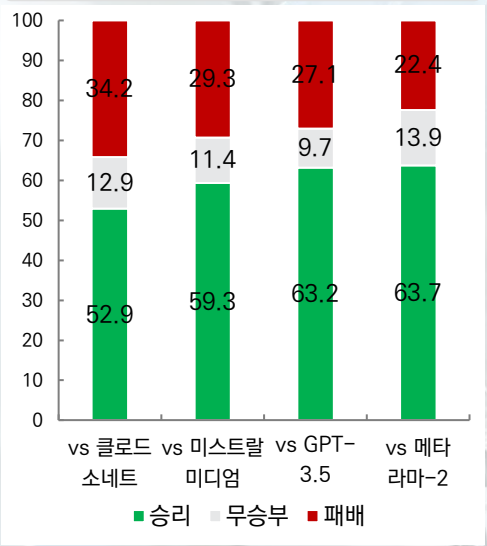
- 드디어 오픈소스 진영에서 엄청난 성능의 언어모델이 출시. 4월 18일에 출시된 메타의 Llama-3는 8B 및 70B 모델 사이즈로 각각 출시. 더 작고 실행 비용이 적은 모델로 품질에 근접. **같은 크기의 다른 모델과 경쟁력이 있다는 말**. LLaMA 3 70B 모델은 Gemini Pro 1.5 및 Claude 3와 성능이 비슷. GPT-4 수준에 도달했다는 것. 중요한 점은, GPT-4는 1.76조 파라미터를 갖고, Llama-3는 700억개. 따라서 벌써, 각 가정에서 애플의 M2 Ultra, M3 Max 등으로 Llama-3-70B를 구동할 수 있다는 후기가 올라오는 중. 다만, 대다수 사람들이 이 정도의 컴퓨팅 성능을 가지고 있지 않고, 여전히 언어모델간 성능이 우선시 될 것이라 판단, 현 헤게모니가 역전될 것으로 보지는 않음. 그러나 **OpenAI가 현재 언어모델에서 갖는 막대한 우위가 오픈소스에 의해 소실되고 있음은 큰 의미**. 한편, Llama-3는 여러가지 병렬화 기법을 결합해 2.4만개의 GPU로 구성된 2대의 맞춤형 클러스터로 훈련. H100 기준 2만개라면 GPT-4 보다 2배 이상의 훈련량임.
- Llama-2가 2조개에 해당하는 토큰 숫자로 훈련된 데 반해, Llama-3는 무려 15조라는 데이터셋으로 8B(80억) 파라미터의 모델로 훈련되었음. "Chinchilla 최적의 스케일링 법칙"에 따르면, 80억 모델 사이즈는 대략 2,000억(25배) 토큰으로 훈련시키는 것이 모델 성능 측면에서 가장 효율적. 그러나 **Llama-3의 파라미터 대비 토큰 비율이 75배. 즉, 적은 파라미터지만 훨씬 많은 데이터로 훈련했음에도 좋은 성능을 낸 것**. 안드레 카파시를 포함한 AI 최고급 전문가들은 이를 매우 긍정적이라고 보고 있음. 왜냐면, **우리가 평소에 사용하는 LLM들이 실질적으로는 매우 과소 훈련(undertrained)된 상태를 입증**하는 것이기 때문. 다시 말해, 더 많은 데이터를 넣으면, 성능향상에 있어 천장에 닿아 수렴할 수 있다는 우려를 불식시키는 것. 이는 OpenAI의 수석 과학자인 일리야 수츠케버가 말한 "AI는 단지 더 배우기를 원할 뿐이다"라는 말과 동일. 이어서 안드레 카파시는, 현재보다 100~1000배 더 많은 데이터로 훈련이 필요할 수도 있다고 주장. 이는 **앞으로 고품질의 데이터가 갖는 입지가 지금보다도 훨씬 더 커질 수 있다는 것**. AGI 경쟁을 위해서 더 많은 데이터를 끌어 모으기 위해 빅테크/국가들의 각축전은 지금보다 훨씬 더 거세질 것.
- Llama-3 중 최고급 모델이자 아직 훈련 중인 **405B는, 완성까지 필요한 연산량이 FLOPS 기준 GPT-4의 약 2배 정도로 계산**. 물론 모델 성능에 영향을 미치는 다른 요소들이 있지만, AI 모델 간에 "공평한", 유일한 방법이라 유의미. 즉, 올 연말에 출시될 메타의 오픈소스 기반 405B 모델은 GPT-4를 쉽사리 제칠 것이라는 결론. 물론, 이에 대해 모든 AI 커뮤니티들은 OpenAI가 GPT-5를 선공개하면서 메타의 빛을 바래게 할 것이라고 예상. 샘 알트만은 지난 15일 인터뷰에서, GPT-5의 성능 향상폭이 GPT-4 때만큼은 될 것이라는 뉘앙스의 발언. 더욱이 우리 팀은 메타가 OpenAI를 따라잡기 위해서는 멀티 모달 모델을 증명해야 된다고 판단. 오픈소스 진영의 'vision' 기능은 GPT-4v보다 훨씬 뒤처져 있는 게 사실. Sora로 대변되는 비디오 모델은 격차가 더 심각. **Embodied AI 및 생성형 미디어를 위해서는 언어 데이터보다는 훨씬 더 많은 이미지/영상 데이터가 필요.** 구글이 유튜브 데이터를 잠그고, 메타가 Reels에 집착하는 것은 이런 이유로 판단.

모델별 파라미터 대비 훈련토큰 비율, 'Chinchilla 스케일링법칙 기준 적정수치는 25:1'



자료: LifeArchitect.ai/chinchilla

라마-3-70B 모델과 peer간 성능비교



자료: Meta, 미래에셋증권 디지털리서치팀
주: Instructed 모델 기준, Human Evaluation 점수

(아직 훈련 중인) "라마-3"와 여러 SOTA 모델들간 성능 비교

벤치마크	Llama-3-405B	클로드 3 오푸스	GPT-4-터보	Gemini 울트라 1.0	Gemini 프로 1.5
MMLU	86.1	86.8	86.5	83.7	81.9
GPQA	48	50.4	49.1	-	-
Human Eval	84.1	84.9	87.6	74.4	71.9
MATH	57.8	60.1	72.2	53.2	58.5

자료: X(@DrJimFan), 미래에셋증권 디지털리서치팀
주: 표의 초록색은 SOTA 대비 비교적 준수한 수준을 의미하고 적색은 열등함을 의미

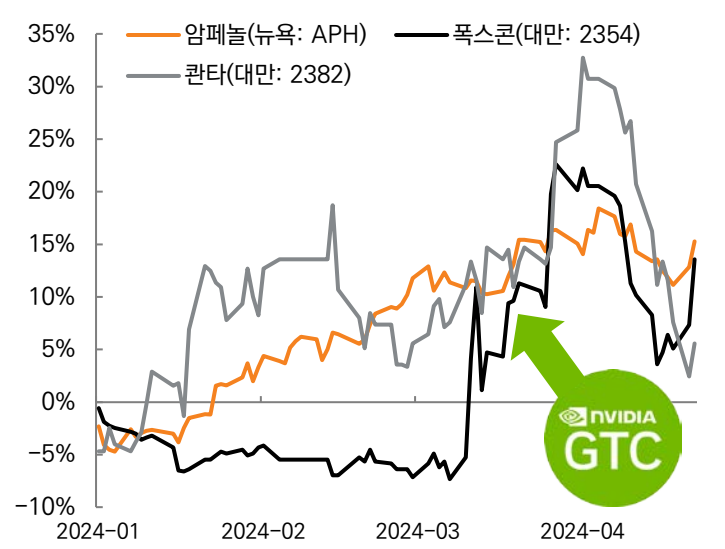
AI issues incoming

TSMC 실적, '꽁꽁 얼어 붙은 핸폰 위로 AI가 걸어다닙니다'

- 엔비디아의 신제품 및 경쟁력에 관해서, 중화권에서의 최근 소식이 매우 흥미로움. 먼저, **엔비디아의 CUDA 생태계가 강한 이유는 스타트업에 대한 영향력 덕분**이라는 것. 엔비디아는 무려 **직원 약 4,500명을 스타트업 지원에 투입**. 여기에는 엔비디아는 AWS 등 클라우드에서 사용할 수 있는 무료 크레딧을 제공하는 방식을 포함. 이러한 AI 스타트업은 CUDA 생태계에 머물게 되고, 따라서 **클라우드 업체들은 엔비디아 서버를 대량 구매해야 하는 상황이라는 점**. 제 아무리 구글이 직접 만든 **TPU가 성능이 좋아도 자체 Gemini 모델 학습 말고는, 외부 스타트업이 VM으로 가져다 쓰기에는 힘든 이유가 바로 이것. CUDA는 사실상 업계 표준**인 셈. 한편, 엔비디아의 새 야심작인 **GB200 참여에 가장 적극적인 ODM 업체는 콴타와 폭스콘**으로 알려짐. 먼저, **폭스콘은 MS와만 거래하면서 슈퍼마이크로를 대체하는 중**으로 파악됨. 최근 SMCI의 주가약세와 맞닿아 있음. 한편, 콴타는 AI 서버 관련 오래된 투자 경험 덕분에, Big 4(MS, GCP, AWS, Meta) 모두 콴타의 GB200 제품에 대해 주문을 하는 상황. **콴타 수석부사장은 올해 서버 출하량이 두 자릿수 성장해 당초 예상보다 상향됐다고 발언**. 또한, 올해 AI 서버가 차지하는 매출비중이 지난해 20%에서 올해 50%까지 상승할 것이라고 전망. 게다가 **GB200을 올 9월에 양산을 개시할 것이라는 말로 중화권에서의 최근 루머의 진위를 확인해 줬음.** 그는 **GB200에 대한 본인들의 생산량은 매분기 늘어날 것**이라 발언. **CSP들은 올 하반기 H100 대신 GB200을 구매할 것**으로 점쳐짐. 테슬라 또한 이번달에 GB200을 대량 주문했다는 정보도 흘려짐. 한편, GB200은 H100과 비교했을 때 설계상의 여러 변화가 있는데, 특히 냉각 솔루션 및 고속 케이블 관련 업체들에 주목할 필요. **GB200에는 기존과 다른 규격의 서버 케이블이 들어가는데, 대표적인 것이 바로 암페놀(티커는 APH)의 케이블.**

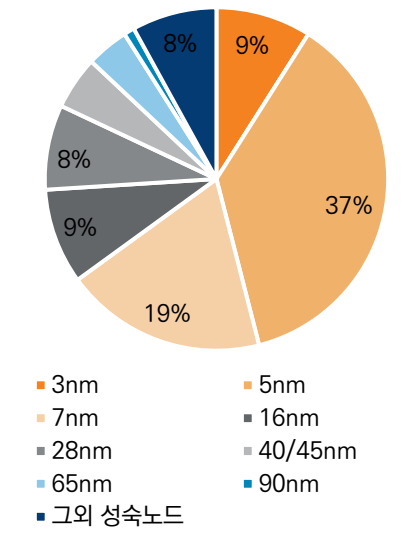
- **TSMC의 1분기 실적은 매출과 순익 모두 호조**. 매출은 **스마트폰의 계절적 영향 불구, 지속적인 HPC(AI 포함)수요에 USD 기준 3.8% QoQ 감소로 선방**. 7nm 이하 미세공정은 웨이퍼 매출의 65%를 차지. **2분기 매출 가이던스는 USD 기준 약 200억을 예상, 이는 QoQ 6%, YoY 27.6%를 의미**. 연간 매출은 USD 기준 20% 초·중반까지 증가할 것으로 예상. TSMC는 AI 관련 데이터센터 수요는 매우 강하다고 강조. 거의 모든 AI 기업들이 에너지 효율적인 컴퓨팅 성능에 대한 끝없는 수요를 해결하기 위해 본인들과 협력 중이라 밝힘. 이에 따라, **여러 AI 프로세서의 매출 기여도**가 올해 2배 이상 증가해 올해 매출의 10% 초반을 차지하고, **향후 5년간 CAGR 50%**로 2028년에는 전체 매출의 20% 이상으로 증가할 것으로 전망. 반면, (CPU 기반) 전통적 서버 수요는 느리고 미지근. 다만, 하이퍼스케일러들의 예산이 전통 서버에서 AI 서버로 이동하는 것은 TSMC에 유리. **TSMC는 2024년 말까지 CoWoS 패키징 용량을 월 4만장으로 작년 말 대비 2.5배 늘릴 예정. 그마저도 여전히 고객의 수요를 충족시키기에는 부족해 2025년 연간 기준 수백만 개를 초과할 가능성까지 제기. 엔비디아 블랙웰에서 사용되는 CoWoS-L 패키징 웨이퍼는 올 하반기 출시 예정.** 현재 모든 AI 가속기는 5나노, 4나노 기술 기반. 다만, 에너지 효율이 중요하기 때문에, 모든 고객이 3나노 및 2나노로 전환하는 추세로 작업 중. 올해 TSMC N3 매출은 작년 매출의 3배 이상이 될 것. 또한, 온디바이스 AI는 가장 중요한 것이 다이 사이즈(결국 전력수요). 이에 따라, **AI 특화 스마트폰 및 PC들의 교체 주기가 앞으로 조금 더 빨라지지 않겠냐는 의견**도 피력.

엔비디아 GB200 NVL 72와 직접 연관된 기업들의 YTD



자료: TradingView, Nvidia, 미래에셋증권 디지털리서치팀

TSMC의 공정별 매출 비중



자료: TSMC, 미래에셋증권 디지털리서치팀

TSMC의 지난 분기 실적 하이라이트, '2분기 매출 가이던스는 약 200달러로 YoY 약 28%'

구분	1Q24	1Q24 기존 가이던스	전분기 대비	전년동기 대비
순 매출 (단위: 십억 미국달러)	18.87	18.0-18.8	-3.8%	+12.9%
순 매출	592.64		-5.3%	+16.5%
매출총이익률	53.1%	52%-54%	+0.1 ppt	-3.2 ppts
영업이익률	42.0%	40%-42%	+0.4 ppt	-3.5 ppts
순이익률	38.0%		-0.2 ppt	-2.7 ppts
주당순이익	8.70		-5.5%	+8.9%
ROE	25.4%		-2.7 ppts	-2.1 ppts
12인치 웨이퍼 기준 판매량 (단위: 천 장)	3,030		+2.5%	-6.1%

자료: TSMC, 미래에셋증권 디지털리서치팀

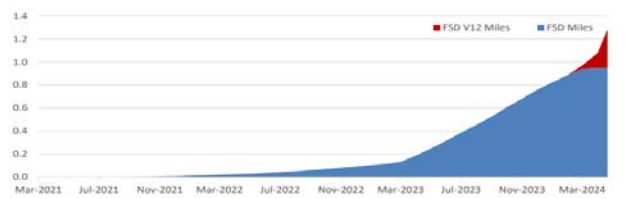
AI issues incoming

테슬라 실적, 'Told you? It's AI'

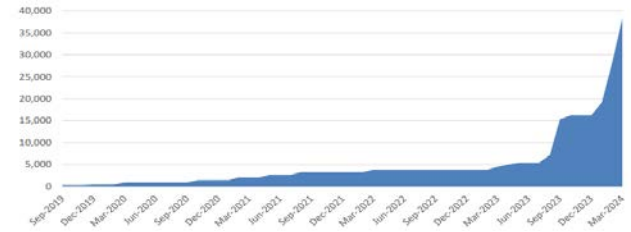
- 실적 발표 전 7일 연속 주가 하락을 경험한 **테슬라는 4월 23일 장 마감 이후 실적을 발표. 매출과 순이익 모두 역성장 했고 순이익의 경우 컨센서스를 하회하기도 했으나 투자자들은 반색**. 이미 4월 초 지난 분기 차량 판매대수가 쇼크 수준으로 나왔기 때문에 모두가 나쁜 소식을 알고 있었던 상황이기 때문. 또한 실적 이후 이뤄졌던 '어닝 콜' 도중 에프터마켓의 주가는 상승 기어를 더 올렸음. **작년 몇 차례에 있었던 다소 "암담/침울한" 분위기의 어닝 콜이 아니라, 대부분 낙관적인 전망/자신감으로 채워진 통화 내용**이었기 때문. 또한 내용 중 곳곳에 나온 여러가지 흥미로운 긍정적인 단서들에 투자자들은 집중하는 모습. 그 단서에 기반해 신제품, 에너지 등 여러 파생된 추측들이 많지만, **가장 확실한 것은 이번 분기부터 테슬라가 AI에 완전히 무게중심을 옮겼다**는 것. 이는 우리 팀이 예상했던 바와 그대로 일치. **우선 AI향 투자금액 자체가 폭증. 테슬라는 지난 분기 28억 달러라는 Capex를 단행했는데, 이중에서 10억 달러를 AI 인프라에 투자**. 이에 따라 잉여현금흐름은 마이너스를 기록하긴 했으나, 여전히 269억 달러 정도의 현금이 있어 우려할 상황은 아님. 앞으로 **AI 슈퍼컴퓨터 "Dojo"의 컴퓨팅 용량을 끌어 올리기 위한 배팅액을 double down 할 것은 불 보듯 뻔함**. 실제로 테슬라가 이번에 공개한 AI 훈련용 컴퓨팅 규모를 보면, **지난 분기 H100 보유량이 폭증**한 것으로 알 수 있음. **작년 말 기준 1.5만개였던 것이 현재 기준 3.5만개를 넘긴 것**. 물론 이 숫자는 H100의 성능에 "맞먹는" 연산 용량이기 때문에 순수하게 H100만을 2만대 구매했다는 것은 아닐 것. 다만, 대부분을 차지하고 자체 실리콘인 D1 칩의 성능이 엔비디아 GPU에는 미치지 못한다는 것이 중론. 또한 일론 머스크의 순수 AI 업체 **xAI도 H100을 구매하고 있다는 것도 유념**. 일론 머스크는 현재 H100을 35만개 가량 보유한 메타, 바로 다음이 테슬라와 xAI가 될 것이라 말한 바 있음. **이번 분기에도 H100 GPU에 대한 범테슬라 진영의 러시는 지속될 것**. 컴퓨팅 용량보다도 중요한, 실제 세상의 데이터에 있어서도 테슬라는 엄청난 해자를 뽐냄. **FSD 베타로 주행한 차량들의 누적 주행 거리가 10억 마일을 달성한지 한 달도 채되지 않아 벌써 약 13억 마일을 달성**한 것. 이는 우리 팀이 짚었듯, "Supervised FSD"가 한달 무료 공개되었기 때문. 놀라운 점은, **구글 웨이모가 창사 이래 여태껏 수집한 전체 데이터의 3~4배에 달하는 데이터량을 테슬라는 지난 한 달 동안 쌓아올린 것.** 이러한 컴퓨팅 용량 및 데이터 확보량의 폭증은 FSD v12의 업데이트 주기를 더욱 빨리 만들게 하는 원동력. **지난 분기에는 FSD와 EAP에서 "자동주차" 기능이 오픈되며, 자율주행 서비스로 벌어들인 이연부채가 매출로 크게 인식**되기도 함. 앞으로 여러 기능들이 해금되면 **이런 회계적 호재들은 계속해서 발생할 가능성 높음**. 일각에서 "실패나 도박"으로 치부한 슈퍼컴퓨터 Dojo에 대해, 실제 사진을 공개함으로써 방어한 점도 흥미로움.

- 이번 실적에서 투자자들이 가장 관심을 줬던 분야는 결국 로보택시. **로보택시는 "저가형 모델 + 차량 공유"의 결합**이라고 이해할 수 있음. 우선, **저가형 모델은 기존 타임라인보다 이른 내년 상반기 출시될 것으로 판단**. 또한 차량 공유에 대해서는, 일론 머스크가 직접 Uber를 언급. 즉, **무인 택시를 내놓기 전에, (투자자들이 우선 원했던 방식으로) 운전자가 있는 상태에서 택시의 기능을 하는 서비스로 먼저 공략하겠다**는 뜻. Uber 운전자가 승객을 태운 상황에서, **FSD가 활용된다면 거리당 요금에 있어서 Uber보다 훨씬 저렴해질 것. Ride-hailing 시장에서 확실한 수요가 발생할 것.**

테슬라의 AI 해자, '실세계 주행데이터, 슈퍼컴퓨터'



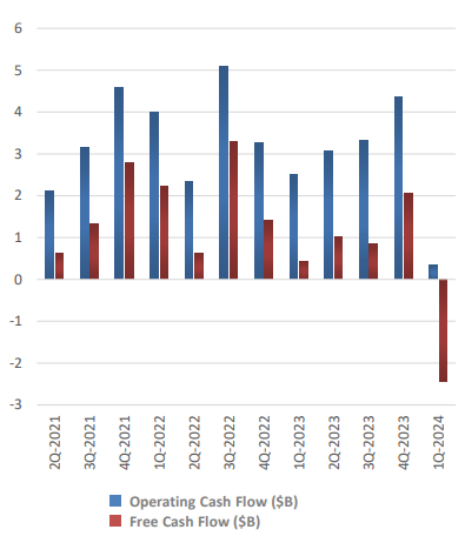
FSD 차량으로 도로를 누빈 누적거리(단위: 십억마일)



AI 훈련용 컴퓨팅 규모(H100로 따질 때 몇 대나 갖고 있나?)

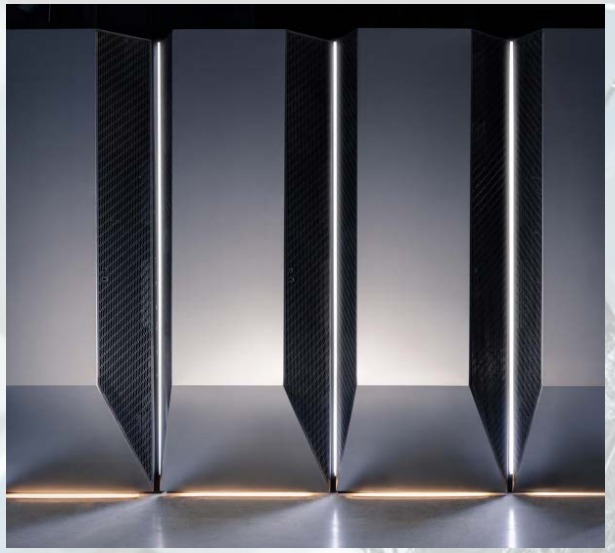
자료: Tesla
주: 위 그래프는 2024년 4월 21일 기준

테슬라의 분기별 현금흐름, 'AI 배팅'



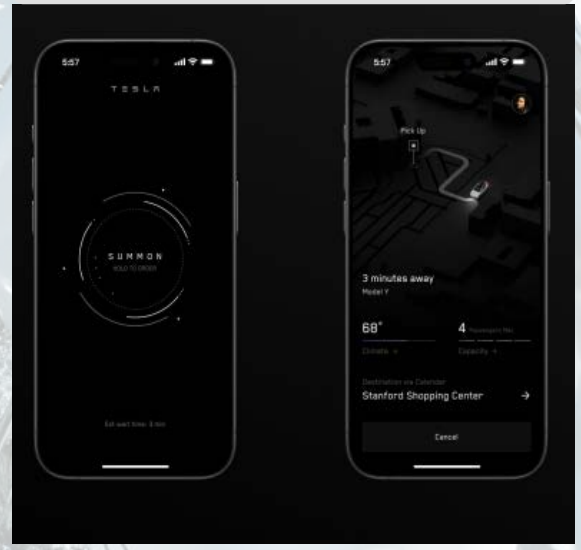
자료: Tesla

테슬라의 AI 슈퍼컴퓨터 "Dojo"의 '실제' 사진



자료: Tesla

이번에 공개한 테슬라 차량공유 앱(예정) 예시



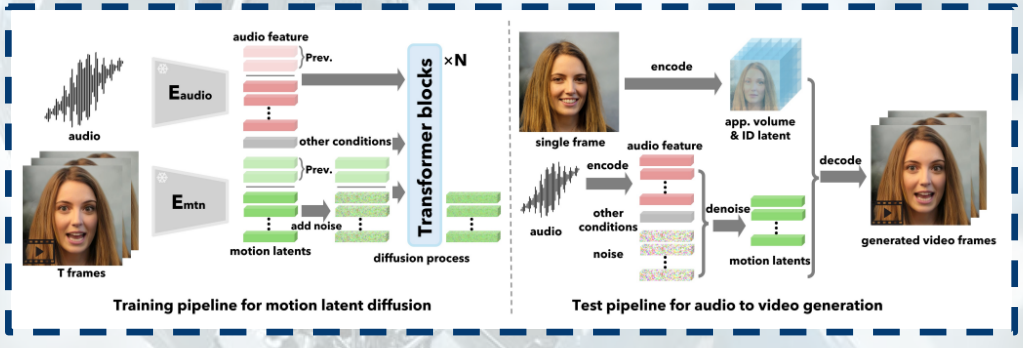
자료: Tesla

Paper of the Week

VASA-1 - Microsoft

✓ Core Contents

- 마이크로소프트, 이미지와 음성을 입력값으로 받아 **자연스러운 입모양과 생동감 있는 표정, 고개 끄덕임까지 실시간으로 생성**하는 모델인 VASA-1 공개.
- 기존 방법들 대비 비디오 품질, 생성 속도 뿐만 아니라 고개 움직임 등까지 구현해 보다 자연스러운 영상 생성이 특징.
- 1)얼굴의 3D 형태와 텍스처 정보, 2)얼굴의 특징적인 외형, 3)머리의 3D 위치 및 방향, 4)입, 표정, 눈 등 얼굴의 동적 변화 를 잠재 변수로 얼굴을 잠재 공간에 표현. 이렇게 되면 **얼굴의 각 속성을 '독립적으로 제어할 수 있음**(disentangled)'.
- 이후 디퓨전 트랜스포머 기반 모델을 사용해 복합적인 표정(Holistic Facial Dynamics Generation, HFDG)과 얼굴의 움직임을 구현. 이 밖에 시선 방향, 얼굴-카메라 거리, 감정 오프셋 등 제어 가능한 조건을 모델에 추가.



✓ Insights

- 이전에 제안된 방식과 달리, **각 부분을 독립적으로 제어할 수 있고, 영상 생성 딜레이를 단축시켰다는 점에서 유용성이 높아졌다**는 데 의의가 있음.
- CAPP(Contrastive Audio and Pose Pretraining) 평가 지표, 음성-머리 움직임 간 동기화 정도를 평가하는 방법으로, 향후 관련 연구에서 유용할 것으로 보임.
- 여전히 <u>머리카락, 옷과 같은 얼굴 외 부분은 움직이지 않아</u> 어색한 부분이 있으며, <u>몸의 움직임은 구현하지 못한다는 한계</u> 존재.
- 향후 원격 협업, AI 튜터, 메타버스 등 다양한 분야에서 새로운 상호작용 옵션을 제공할 수 있을 것. 다만 악용 가능성에 대한 윤리적 고려 필요.

자료: "VASA-1: Lifelike Audio-Driven Talking Faces Generated in Real Time", Microsoft Research

AI Peer Table (1)

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			
최종 소비자 (온디바이스, 앱 등)	애플	AAPL US	166.90	3,541.6	-1.5	-3.1	-14.4	-3.3	1.7	-13.2	0.59	05-02	온디바이스 AI 위한 자체 경량화 언어모델 개발 중(Ajax GPT)
	테슬라	TSLA US	144.68	634.1	-7.9	-15.3	-30.8	-31.8	-12.4	-41.8	0.49	04-23	자율주행 소프트웨어 FSD의 싱글스택 신경망 v12 공공출시 임박
	삼성전자	005930 KS	75,500	450.7	-4.3	-4.3	2.0	10.7	17.5	-3.8	0.13	04-30	'24년부터는 온디바이스 AI에도 중점(Galaxy AI)
	퀄컴	QCOM US	161.35	247.4	-4.3	-5.1	5.7	50.2	40.5	12.1	0.61	05-01	생성 AI 처리 강화용 온디바이스 AP 및 CPU 설계(X Elite)
	스냅	SNAP US	11.39	25.8	9.4	-0.1	-30.3	20.0	13.8	-32.7	0.36	04-25	이미지 생성 모델을 탑재한 증강현실 서비스(AR 필터, AR 렌즈)
	로블록스	RBLX US	36.30	31.9	-0.9	-2.6	-11.0	18.7	-10.8	-20.6	0.43	05-09	생성 AI 탑재해 무료로 'Low code' 지원(캐릭터, 아이템 생성 위한 AI)
AI 비즈니스 솔루션	어도비	ADBE US	472.90	291.1	-0.7	-5.3	-20.8	-12.5	25.2	-20.7	0.58	06-13	생성 AI 기반 미디어 편집 툴(Firefly)
	세일스포스	CRM US	276.68	368.8	-0.3	-10.1	0.1	37.2	39.2	5.3	0.53	05-31	세계 최고의 CRM 업체로 생성 AI를 탑재해 사용성 강화(Einstein GPT)
	IBM	IBM US	182.19	229.5	-0.8	-4.5	5.7	36.3	51.6	12.4	0.31	04-24	기업 고객용 AI 어시스턴트(상담사 챗봇 구축, 업무 자동화, 코딩 지원)
	서비스나우	NOW US	740.95	209.1	-0.5	-4.3	-1.7	36.9	56.5	4.9	0.61	04-24	워크플로우 자동화 기업으로 생성 AI를 탑재해 사용성 강화(Now Assist)
	크라우드스트라이크	CRWD US	299.65	99.6	0.2	-8.5	0.8	67.3	126.6	17.4	0.54	05-31	기업 대상 생성 AI 챗봇 기반 클라우드 및 엔드포인트 보안(Charlotte AI)
	데이터독	DDOG US	126.44	57.7	-0.4	2.8	-2.0	47.6	87.9	4.2	0.52	05-07	생성 AI 기반 클라우드 인프라 및 보안 모니터링(Bits AI)
	다이나트레이스	DT US	46.25	18.8	3.9	-0.2	-20.6	-0.3	8.4	-15.4	0.51	05-17	생성 AI 기반 어플리케이션 성능 관리 및 보안 플랫폼(Davis)
	유니티	U US	23.10	12.3	-3.5	-14.4	-32.6	-12.4	-19.4	-43.5	0.47	05-09	올해 출시 예정인 차세대 게임엔진에 AI 기능 대거 탑재(Sentis, Muse)
	센티넬원	S US	21.18	9.0	1.4	-7.0	-20.2	32.0	14.0	-22.8	0.46	05-31	기업 고객 대상 생성 AI 챗봇 기반 클라우드 및 엔드포인트 보안(Purple AI)
	루닛	328130 KS	52,800	1.5	5.2	-15.0	-27.2	-29.0	100.8	-35.5	0.10	05-10	AI 기반 암 검출/진단 보조 솔루션(인사이트)
기반 모델	알파벳	GOOGL US	158.26	2,717.8	2.5	5.0	7.6	15.9	50.1	13.3	0.56	04-25	OpenAI와의 AGI 경쟁에 있어 대적할 수 있는 거의 유일한 업체(Gemini)
	메타 플랫폼스	META US	496.10	1,729.2	-0.7	-2.6	28.9	58.2	133.3	40.3	0.52	04-24	'거의 완전한' 오픈소스 LLM 개발 및 배포를 선도(Llama)
	바이두	9888 HK	96.35	47.4	3.5	-1.9	-8.3	-7.2	-21.0	-17.0	0.13	05-16	중국 최초/최고 성능의 LLM 보유(Ernie Bot)
	네이버	035420 KS	180,100	29.2	0.3	-4.7	-17.4	-3.3	-4.7	-19.6	0.13	05-03	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)

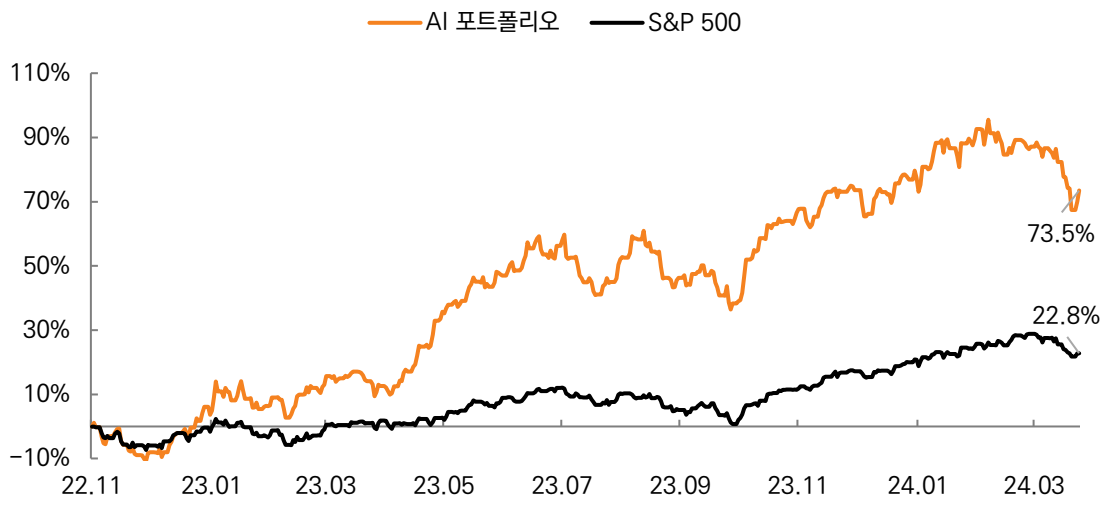
밸류체인	종목명	티커	주가 (현지통화)	시가총액 (조원)	기간별 수익률(%)						엔비디아와의 상관관계 (30일)	실적발표 예정일	AI 밸류체인 관련 한 줄 평
					1W	1M	3M	6M	1Y	YTD			
AIOps	스노우플레이크	SNOW US	151.17	69.4	0.6	-4.9	-26.5	1.7	4.1	-24.0	0.44	05-24	AI 모델 구축, 배포, 관리를 위한 빅데이터 워크로드 플랫폼(Cortex)
	몽고DB	MDB US	366.31	36.7	4.7	3.0	-10.9	7.0	57.3	-10.4	0.55	05-31	AI 모델 구축 시 필수 인프라로 떠오른 '벡터 데이터베이스' 기능 보유(Atlas)
	클라우드플레어	NET US	87.88	40.8	-3.5	-9.0	4.7	49.7	39.9	5.5	0.55	05-02	트래픽 처리(CDN) 업체로, 실시간 AI 추론 작업에서의 역량 대두(Workers AI)
	팔란티어	PLTR US	21.64	65.8	-1.1	-10.5	24.9	33.6	164.5	26.0	0.50	05-06	실제 조직 운영시의 의사결정 지원 및 자동화를 위한 AI 개발 플랫폼(AIP)
	포스 패러다임	6682 HK	49.60	4.1	2.0	-23.8	-1.6	-6.7	-	-1.2	0.14	08-19	실제 조직 운영시의 의사결정 지원 및 자동화를 위한 AI 개발 플랫폼(Sage)
데이터센터	마이크로소프트	MSFT US	407.57	4,161.6	-1.7	-4.9	2.4	24.2	43.8	8.6	0.65	04-25	OpenAI의 AI 모델 라이센스에 대한 독점권을 갖고 있는 세계 2위의 CSP
	아마존 닷컴	AMZN US	179.54	2,562.8	-2.1	0.4	15.1	41.9	67.9	18.2	0.58	04-30	최고의 CSP로써 작년 9월 Anthropic에 40억 달러 투자하며 AGI 경쟁 참여
	오라클	ORCL US	115.09	434.7	-4.6	-9.6	3.2	11.8	22.7	9.9	0.47	06-12	기업들의 자체 인프라 수요에 대응한 '온프레미스 서버' DB 관리 시스템의 강자
	에퀴닉스	EQIX US	761.90	99.4	2.7	-4.9	-4.2	10.1	9.0	-4.9	0.45	05-03	물리적 공간의 데이터센터를 임대/위탁 운영하는 업계 최대 리츠 업체
반도체 인프라	엔비디아	NVDA US	824.23	2,831.6	-5.7	-12.6	37.7	91.8	204.0	66.4	1.00	05-22	AI 훈련은 GPU가 필수, GPU 현존 최강자는 엔비디아
	브로드컴	AVGO US	1,249.19	795.5	-6.0	-7.7	2.3	46.1	101.3	12.4	0.68	06-12	VMware 인수로 클라우드 가상화 역량도 확보한 네트워킹 반도체의 최강자
	TSMC	2330 TT	754.00	824.4	-6.2	-3.9	20.8	40.0	51.7	27.7	0.18	07-22	압도적인 미세공정 점유율을 가진 대만의 반도체 파운드리
	AMD	AMD US	152.27	338.1	-6.8	-15.2	-9.6	52.3	72.2	3.3	0.78	04-30	HPC CPU에서 업계리딩, GPU에서 엔비디아의 적수로 꼽히는 패스트 팔로워
	인텔	INTC US	34.28	200.5	-5.5	-19.5	-29.7	1.9	14.7	-31.6	0.45	04-25	GPU 중심 서버 설계에는 후발주자. 파운드리 기업으로서 지정학적 가치 부상
	마이크론	MU US	112.46	171.1	-7.6	2.1	28.6	68.5	85.0	31.9	0.62	06-28	HBM에서 SK하이닉스를 추격하는 미국의 메모리 반도체 업체
	SK하이닉스	000660 KS	171,000	124.5	-4.3	0.7	20.8	35.1	97.5	20.8	0.22	04-25	GPU 병목의 키를 쥐고 있는 HBM에 있어 글로벌 선두주자
	시놉시스	SNPS US	527.50	110.6	-2.5	-11.2	-3.9	12.8	39.8	2.4	0.71	05-22	반도체 EDA 업체로, 3D 엔지니어링 시뮬레이터 업체 앤시스와 M&A 논의 중
	아리스타 네트웍스	ANET US	251.18	107.9	-4.4	-18.0	-4.1	34.7	61.1	6.7	0.59	05-07	이더넷 중심 고속 네트워킹을 위한 스위치 설계
	Arm	ARM US	96.04	135.5	-21.4	-28.4	25.8	91.3	-	27.8	0.29	05-08	AI 가속기 설계 위한 다수의 IP 블록 소유한 업체로 설계 라이선스/로열티 사업
	마벨 테크놀로지스	MRVL US	63.86	76.0	-6.5	-3.9	-9.5	30.5	64.3	6.1	0.71	05-24	데이터 인프라용 네트워킹 및 스토리지 ASIC 설계
	SMIC	981 HK	14.34	30.4	1.1	-9.2	-7.5	-32.4	-37.7	-27.8	0.03	05-09	최근 화웨이와 밀접히 협력 중인 중국 유일의 미세공정 가능 파운드리
	Alchip	3661 TT	2,860.00	9.5	-1.4	-17.1	-22.3	11.3	137.6	-12.7	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

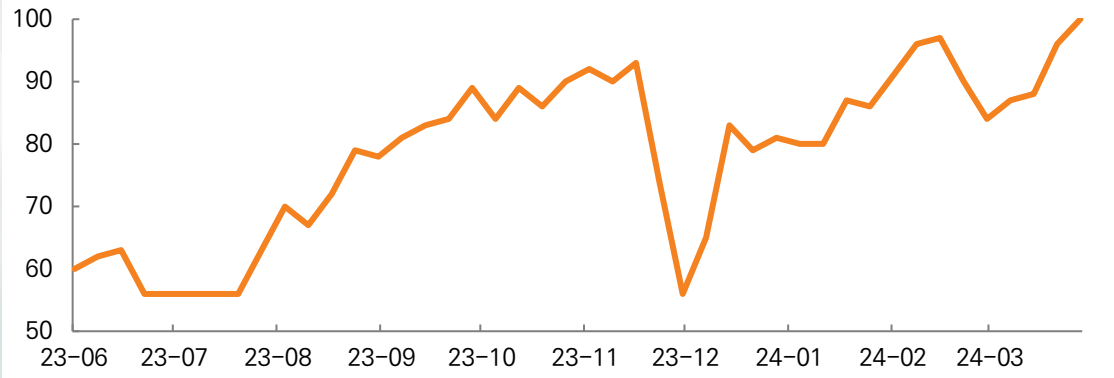
Appendix: Charts

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

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

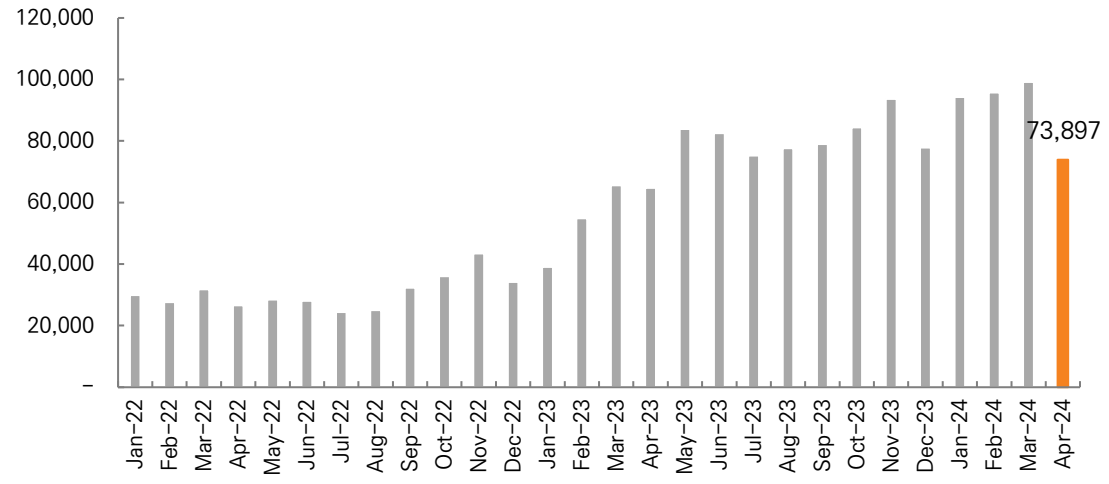


ChatGPT 키워드를 검색한 Google Trends 추이 (지난 90일 평균치 기준)



ChatGPT 관련, 지난 7일간 "급상승 키워드"
Gemini (Google)

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



자료: Bloomberg, CBINSIGHTS, Google Trends, 미래에셋증권 디지털리서치팀

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.