2024. 11.14
그림: Midjourney(엔비디아와 테슬라의 합작 = AI 업계의 표준)

AI Weekly, 특이점이 온다 #39

W-129 to A.G.I.

(AGI's ETA: 2027.05.08

디지털리서치팀 한종목, 김은지
chongmok.han@miraeasset.com
eunji.kim.a@miraeasset.com

MIRAE ASSET
미래에셋증권

Highlight of the Week

- **AI 업계에 다시 겨울이 온다고?**

그렉 브록만의 조기 복귀와 함께 OpenAI의 행보가 가속화. 미국의 AI 주도권 강화를 위해서, 대규모 인프라 투자 계획 등을 담은 Blueprint를 발간. 여기에는 특히 전력 인프라 확충 필요성이 핵심. AI 경제 구역 지정과 원전 건설 허가 간소화 등을 제안. 이것은 xAI와의 경쟁을 의식한 듯. OpenAI는 엔비디아 GB200 서버 10만대 규모의 자체 데이터센터를 내년 완공 목표로 추진 중. 또한, OpenAI는 제품 측면에서 컴퓨터 작업 자동화가 가능한 AI 에이전트 "Operator"를 내년 1월 출시 예정. 한편 Claude 3.5 Opus, Gemini 차세대 모델, GPT-5(Orion)의 성능 정체설이 부상. 이는 기존 스케일링 법칙의 한계점 도달을 시사. 일리야 수츠케버도 사전학습 단계에서의 한계를 인정했으며, 고품질 언어 데이터 고갈이 주요 원인으로 지목됨. 그러나 이는 발전의 정체가 아닌 방향 전환을 의미. OpenAI의 o1 모델이 제시한 "Test-Time Compute" 기반 추론 컴퓨팅이 새로운 패러다임으로 부상 중. 특히 OpenAI는 내부적으로 완전한 o1을 활용해 독점적 추론 데이터를 확보함으로써 경쟁력을 강화할 전망. 타 업체들도 이러한 추론 컴퓨팅으로의 전환이 불가피할 것.

- **일론 머스크는 트럼프 정부의 최고기술책임자**

공화당의 AI 행정명령 폐지 선언과 트럼프의 "AI 해방" 정책은 AI 개발사들의 규제 부담을 크게 경감할 전망. 특히 레드팀 테스트 결과의 정부 제출 의무가 사라지면서 오픈소스 진영, 특히 메타와 엔비디아가 수혜입을 것으로 기대. 다만 바이든 정부가 발표했던 NSM은 대중국 정책의 일환으로 현행 유지될 것으로 보임. 일론 머스크는 비벡 라마스와미와 함께 신설되는 "DOGE" 부처의 공동위원장으로 임명. 연방기관 428개를 100개 정도로 축소하는 것이 핵심 과제가 될 듯. 규제 철폐와 보조금 삭감으로 테슬라, SpaceX 등 자생력 높은 기업이 수혜 예상. 자율주행 라이드쉐어링 시장과 위성 기술 접목으로 새로운 생태계 창출 가능성. 다만 관료조직의 반발이 주요 리스크로 작용할 전망. 테슬라는 중국 최초로 데이터 보안 인증을 획득한 외국기업으로 부상. 데이터 익명화, 현지화 등의 기준을 충족. 이는 중국의 데이터 해외전송 규제 완화 정책과 맞물려 FSD 개발 가속화에 기여할 것으로 전망. 머스크의 미중 양국 내 영향력이 확대되면서 테슬라의 전략적 가치도 상승 중.

- **Paper of the week: The Road Less Scheduled – Meta, Google, Samsung AI Research**

딥러닝 모델의 학습에 있어, 학습률을 결정하는 것은 모델 성능에 매우 중대한 영향을 미침. 그러나 학습 종료 시점(T)을 미리 알아야만 최적의 학습률 스케줄을 설계할 수 있고, 학습 중단 시 재시작이 어려운 등 여러 문제가 존재했음. 이에 스케줄-프리 확률적 경사하강법(Schedule-Free SGD)을 대안으로 제시함. 이는 모델의 현재 파라미터, 해당 시점까지의 평균 파라미터, 그리고 이 두 값의 비율을 고려하여 학습률이 자동으로 조정되도록 한 것. 이는 기존 SGD의 장점은 유지하면서 에러를 더 안정적으로 줄일 수 있음. 특히 매우 큰 학습률을 사용하더라도 안정적으로 학습이 진행되며, 학습 중단 시에도 안정적인 재시작이 가능해 GPU 자원 부족이나 중간 결과 확인 시에도 유연하게 대응할 수 있음.

일	월	화	수	목	금	토
10	11 (위스트론) 실적발표	12 (소프트뱅크) 실적발표	13 (루닛) 실적발표 Post-Industrial Summit Fall 2024(~14)	14 (콴타 컴퓨터) 실적발표	15 (BABA) 실적발표	16
17	18	19	20 (NVDA) 실적발표 Generative AI Summit Toronto	21 (BIDU) 실적발표	22	23
24	25	26 (CRWD) 실적발표	27 AI World Congress(~28)	28	29	30

주: 컨퍼런스 일정은 다음과 같은 기준으로 선별함: 1) 주요 AI 기업이 연사로 참여하거나 스폰서인 경우, 2) 주요 대학 및 기관이 연사로 참여하는 경우 등.
자료: Bloomberg, 외신, 미래에셋증권 디지털리서치팀

AI Issue

AI 업계에 다시 겨울이 온다고?

대선이 끝났고, OpenAI가 다시 고삐를 죄고 있다. 우선, 사장인 그렉 브록만이 돌아왔다. 내년 초에나 복귀할 것으로 보였지만 지난 13일에 회사로 돌아왔다. 상황이 긴박하게 돌아가고 있다는 증거다. 며칠 전 OpenAI는 미국의 AI 전략을 강화하고 중국과의 경쟁에서 우위를 점하기 위한 계획을 발표했다. OpenAI의 청사진을 한 문장으로 압축하면 “대규모/초고속 인프라 투자를 통해 미국이 AI 혁신의 최전선 국가로 남아야 된다"는 것이다. 계획서의 세부 내용에서 에너지 전송 인프라의 계획 및 허가 절차를 간소화하기 위해 "National Transmission Highway Act"을 제안한 것도 흥미롭다. 또한 "AI 경제 구역"을 만들어서 이 구역에서는 원전 건설과 같은 프로젝트 허가를 빠르게 처리할 수 있게 하자고 건의하기도 했다. 이것을 보면 OpenAI 입장에서 현재 무엇이 가장 절실하고 누구를 가장 의식하는지를 알 수 있다. 바로 전력이고, 또 일론 머스크이다. xAI의 대규모 슈퍼컴퓨터 클러스터의 규모와 그 속도는 샘 알트만의 진땀을 빼기 충분했다. 실제로 OpenAI는 xAI에 대항하고자 마이크로소프트에 의존하는 것을 지양하고, 엔비디아 GB200 서버 10만대를 채울 자체적인 대규모 데이터센터를 내년에 완공되도록 소위 "열일"을 하고 있다. 뿐만 아니라, 제품 영역에서도 OpenAI는 존재감을 더 빨리 드러내려고 하고 있다. “Operator”라는 코드명으로 새로운 AI 에이전트를 내년 1월에 출시할 준비를 하고 있기 때문이다. 이 AI는 코드 작성이나 여행 예약과 같이 사용자를 대신하여 컴퓨터로 작업을 수행할 수 있다고 알려져, 앤트로픽의 "Computer Use"와 상당히 유사할 것으로 보인다. 몇 주 전 샘 알트만은 "다음의 거대한 혁신은 에이전트가 될 것"이라고 말했는데, 이 방면에 있어서 당장 내년 1분기부터 선도 업체간 커다란 경쟁이 불 붙을 것은 분명하다.

대선 말고도 최근 AI 업계를 좀 뜨겁게 달구는 주제가 있다. AI 모델의 성능 향상이 일종의 한계에 부딪혔는지에 대해 많은 이야기가 나돌고 있다. 아직 출시가 되기 전인 앤트로픽의 Claude 3.5 Opus와 구글의 Gemini 차세대 모델 성능이 그다지 좋지 못하다는 소문은, ‘스케일링 법칙 회의론자'들의 목소리를 더 크게 하는데 일조했다. 더군다나 OpenAI가 개발중인 GPT-5(Orion으로 이름이 변경됨)조차 성능 향상이 있기는 하나 그 폭이 매우 작다는 소식도 있었다. 이것은 보통 일이 아니다. 스케일링 법칙이 한계가 왔다는 것은 단순히 모델을 크게 만드는 것(더 많은 데이터를 투입하고 더 많은 GPU와 메모리를 사용해서 훈련)이 더 이상 효과적이지 않을 수 있다는 징후를 가리키기 때문이다. 심지어는 지구상 최고의 AI 과학자로 꼽히는 일리야 수츠케버는 최근 인터뷰에서 스케일링이 한계에 도달했다고 인정했다. 이에 대해 많은 사람들은, 고품질의 언어 데이터가 고갈됐기 때문이라는 이유를 대고 있다. 현재 사용중인 데이터셋에 중복된 정보가 많이 포함되어 있는데, 이것이 Gemini의 성능을 저해하는 요인으로 작용한다는 분석도 있다. 또한 데이터 부족을 합성 데이터를 활용해 타개하려는 시도를 하고 있지만, 앞서 언급한 소식들에 따르면 제한적인 효과만 있는 것 같아 보인다. 다시 AI 업계에 겨울이 오는 것인가?

다만, 수츠케버가 말한 것은 "pre-training" 단계를 가리킨다. 그가 정말로 성능 정체를 걱정했다면 초지능을 만들겠다면서 SSI라는 회사를 만들지 않았을 것이다. 또한 OpenAI를 위시해 기업들이 더 많이 더 빨리 데이터센터를 구축하는 것은, 그들이 AI 발전과 그 시장 수요를 낙관한다는 명백한 신호다. 따라서, AI 발전이 멈춘 게 아니라 발전의 방향이 변화하고 있는 것이다. 그 새로운 발전 방향은 OpenAI의 o1이 제시한 "Test-Time Compute", 즉, 추론 컴퓨팅이다. 추론 영역의 스케일링은 단순히 더 많은 데이터를 추가하는 이전의 것과는 완전히 다르다. 사실 o1 개발에서 혁혁한 공을 세운 Noam Brown 또한 "어느 시점에서는 스케일링 패러다임이 무너진다"고 이미 말한 적이 있다. (물론 그는 며칠 전 AI 발전이 둔화될 것이라고 생각지 않는다고 발언) 즉, 사전 훈련에 있어서 일종의 평탄화는 예견된 것이었고, AI 훈련의 패러다임이 바뀌고 있다는 것을 의미한다. 뿐만 아니라 OpenAI는 “완전한 o1”을 내부적으로 사용하면서, AI가 대답으로 내놓는 고품질의 추론 연쇄 과정을 활용해 지속적으로 모델 품질을 갈고 닦을 것이다. 무엇보다 이 데이터는 더 이상 웹에서 공개된 자료가 아니기 때문에 그 배타적 이익은 점점 더 커진다. 즉, 경쟁 업체들도 추론 컴퓨팅으로 전환할 것이고, AI 개발이 정체되고 있다는 내러티브는 곧 약화될 것으로 보인다.

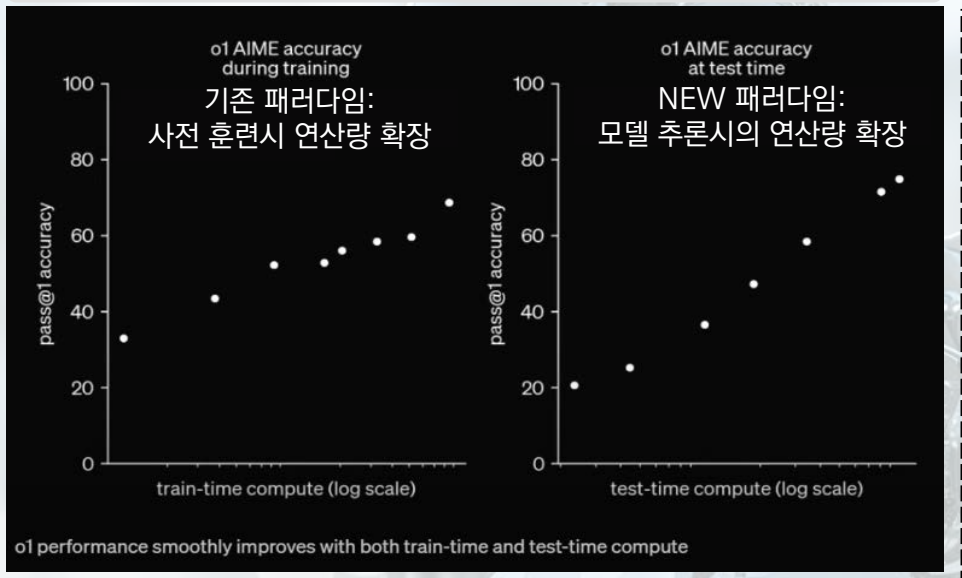
자료: 미래에셋증권 디지털리서치팀

관련 자료: OpenAI

지난 13일, 예정보다는 짧았던 휴가를 끝마친 브록만, '쉴 때가 아니다'

자료: X(@gdb, @polynoamial, @npew), 미래에셋증권 디지털리서치팀

AIME 테스트에서 o1 모델을 두 가지 방법의 스케일링할 때 각 성능 향상 곡선



자료: OpenAI, 미래에셋증권 디지털리서치팀
주: AIME는 가장 우수한 고등학교 수학 학생들을 선발하기 위한 시험으로, 수학 올림피아드 진출자를 가림.

며칠 전 샘 알트만 인터뷰 (Y 콤비네이터)

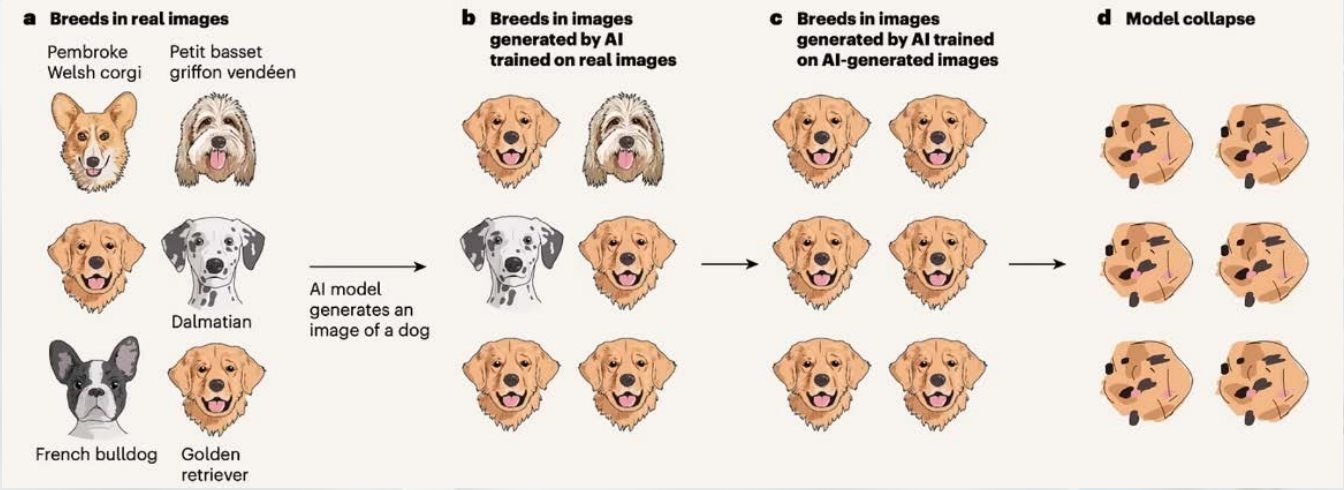
우리는 두 가지 핵심 신념이 있었다: "딥러닝이 실제로 작동한다 + 규모가 커질수록 성능이 더 좋아진다"

모바일, 인터넷, 반도체 혁명, 산업혁명과 같은 큰 변화의 시기에는 신생 기업들이 유리하다. 현재 AI 혁명도 이러한 큰 변화의 시기다. 대기업들은 분기별 계획 주기에 묶여 있다. 심지어 구글은 연간 및 10년 단위의 계획 주기로 움직인다.

우리는 아직 전혀 포화점에 가깝지 않다. AI는 매우 빠르게 훨씬 더 좋아질 것이다. **지금 OpenAI는 창립 이래 처음으로 "무엇을 해야 하는지 알게 된" 시점이다.** AGI 구축까지는 할 일은 많지만, 방향은 명확하다.

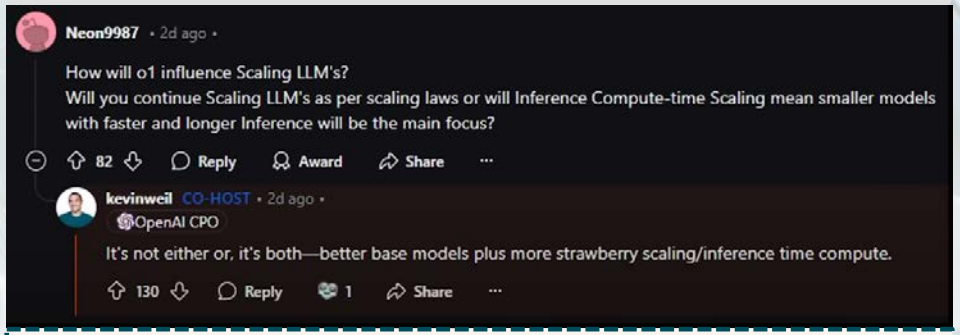
자료: Y Combinator, 미래에셋증권 디지털리서치팀

합성 데이터의 문제점: AI 모델이 이전 세대의 모델이 생성한 데이터로 반복적으로 학습될 때 '모델 붕괴' 현상 발생



자료: Nature 논문 ""AI models collapse when trained on recursively generated data", 미래에셋증권 디지털리서치팀

OpenAI가 새로운 패러다임으로 꺼낸 모델 o1에 관한 CPO의 답변, '난 둘 다'



언어모델의 스케일링 법칙 중 어떤 것(트레이닝 시간 or 추론 시간)에 중점을 둘 것인지에 관한 질문에 대해, OpenAI CPI Kevil Weil의 답변은 아래와 같음.

"더 나은 기본 모델 + 더 많은 추론 시간 컴퓨트 스케일링". 즉, 기본 성능과 추론 능력을 동시에 향상시키겠다는 것이고 **어느 것도 포기하지 않겠다는 말.**

자료: Reddit(@kevinweil), 미래에셋증권 디지털리서치팀

AI Issue

일론 머스크는 트럼프 정부의 최고기술책임자

트럼프의 당선과 공화당의 상하원 장악은 "MAGA" 진영의 발걸음이 앞으로 얼마나 광폭적일지 가늠하기에 충분하다. **공화당은 최근 "우리는 바이든의 AI 행정명령을 폐지할 것이다. 이 행정명령은 AI 혁신을 저해한다. 대신 우리는 표현의 자유와 인간의 번영에 기반한 AI 개발을 지지한다"고 말했다.** 바이든 행정부의 AI 규제 정책에 전면 반대하는 것을 공식화한 것인데, 이어서 **트럼프는 AI 규제 철폐를 통한 "AI 해방" 계획까지 발표했다**. 그들이 해당 AI 행정명령(Executive Order 14110)의 논쟁점으로 꼽은 것은, AI 모델을 개발하는 기업들은 안전 테스트(red-team 과정) 결과를 정부에 의무적으로 제출해야 한다는 것이었다. 그러나 트럼프의 재등장으로 상황은 완전히 바뀌었다. 의무가 사라지면서, AI 개발 기업들의 제약이 대폭 감소돼 AI 기술 발전이 예상보다 훨씬 빠른 속도로 진행될 수 있다. 특히 **오픈소스 진영이 가장 큰 수혜를 입을 수 있고, 그 주인공들은 역시 메타와 엔비디아가 될 가능성이 높다**. 다만, 바이든 정부가 **지난 10월에 발표한 NSM(국가 안보 메모랜덤)은 군사/정보기관의 AI 사용 거버넌스 프레임워크기 때문에 이 행정명령과는 별개로 취급될 것**으로 보인다. NSM이 AGI 개발 경쟁에 관해 비우호 국가(중국 등)와의 전면전을 암시한 것이고, 민주당이든 공화당이든 대중국 정책에 대해서는 큰 이견이 없기 때문이다.

AI를 포함해, 트럼프의 기술 부문 개혁에 대해서 가장 많은 자문을 하는 사람은 지금도, 앞으로도 일론 머스크일 것이다. 그는 트럼프 정부의 새로운 부처인 **"정부 효율화 부처(일명 DOGE)"의 수장으로 임명되었다**. 머스크는 30대 억만장자 기업가 비벡 라마스와미와 공동위원장의 역할을 하면서 내각 외부에서 자문을 제공하게 된다. **"DOGE"의 제 1의 목표는 정부 관료주의 해체와 연방기관 구조조정**이다. 머스크는 지난 10월 당시, '현재 연방기관 숫자가 428개가 되는데, 100개 이하로 줄여야 한다'는 의견을 피력했다. 트럼프는 이번 결정을 "Manhattan Project of our time"으로 명명하면서 일론과 비벡에게 전폭적인 힘을 실어줄 것을 암시했다. **2026년 7월이라는 목표지점을 설정한 만큼, 내년부터는 기대 이상의 규제 철폐와 각종 보조금 지출 삭감이 있을 것이다.** 이 과정에서 가장 수혜를 입을 수 있는 기업은 결국 보조금 없이 **가장 자생력이 높은 기업**이 될 것은 자명한 일이다. 이에 따라 **테슬라와 SpaceX는 시장 재편에 있어 last man standing이 되어 유효시장의 상당부분을 차지할 것으로 사료**된다. 그리고 머스크가 보고 있는 자율주행 기반 라이드쉐어링의 시장 규모는 약 1경원에 해당하는 어마무시한 시장이다. 그리고 테슬라의 AI 기술과 SpaceX의 위성 기술이 접목될 경우 개화될 수 있는 생태계는 가늠하기 힘들 정도로 드넓다. 변화가 큰 만큼 저항이 거셀 것이다. 아메리카 전체에 걸쳐 관료조직의 반발이 예상되는데, 이런 대립과 충돌은 현재 생각하는 것보다 훨씬 극적일 수 있는 큰 리스크 요인이다.

트럼프 정부와 중국 공산당으로부터 동시에 총애를 받는 기업인을 꼽자면 지구상 일론 머스크가 유일할 것이다. MAGA 진영의 핵심 일원이기도 하지만, 일론 머스크는 중국인들에게는 "Faker"와 같다. 본인들이 언젠가 넘어야 할 산인 것이고, 그가 구상한 기술 방향을 정확히 좇고 있다. 앞으로 **중국 정부는 테슬라를 껴안는 제스처를 선보이면서, '자유무역, 선의의 경쟁'과 같은 표어를 강조하며 미국의 보호주의를 압박하지 않을까 생각한다. 그 근거 중 하나는, 중국 정부가 테슬라에게 데이터 보안 인증 획득이라는 선물을 선사했기 때문이다.** 이는 향후 FSD 도입과 시장 확대를 위한 중요한 교두보가 될 것이기 때문에 매우 중대한 이슈다. 테슬라는 중국 관계당국의 엄격한 데이터 규제 기준을 통과한 최초의 외국기업이 되었다. 주요 인증 기준으로는, "차량 외부 얼굴 정보 자동 익명화, 차량 내부 데이터 기본적으로 미수집 설정, 차량 내 데이터에 있어 데이터 처리를 현지화, 개인정보 처리 시 사용자 고지"가 있다. 군사시설 및 정부기관 출입 제한 해제로 중국 국영 기관 고객 확대 가능성이 증가할 것은 물론이다. 더 중요한 것은 **이번 보안 인증 획득을 지난 3월의 중국 사이버공간관리국의 '데이터 해외전송 규제 완화' 정책과 맞물려 생각해야 한다**는 점이다. 이 정책은 "익명화된 데이터"의 경우 해외 전송 규제를 크게 완화해주겠다는 것이었다. 즉, 이제는 **테슬라는 중국에서 얻은 데이터의 익명화 방법에 관해 인증 받았기 때문에, 미국의 Cortex 클러스터로 데이터를 보내 이전과는 다른 규모로 훈련하게 될 것**이다. 결국, **내년 1분기 중국 내 FSD 출시에 관한 타임라인 달성에 성큼 다가섰다고 말할 수 있게 되었다**. 일론 머스크 및 테슬라의 G2 사이에서의 줄타기 곡예는 이제 막 시작되었다.

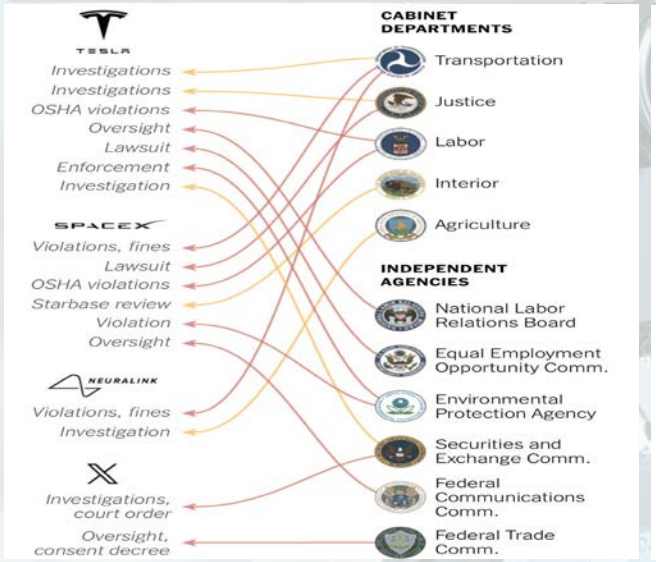
자료: 미래에셋증권 디지털리서치팀

관련 자료: 트럼프 정부와 일론 머스크

일론 머스크가 수장이 된 DOGE(cost-cutting이 목표)의 업무 개시…

자료: X(@elonmusk, @DOGE), 미래에셋증권 디지털리서치팀

일론 머스크의 기업들이 직면 중인 각종 규제 난관들



자료: Jonathan Corum
주: 각각 빨간 선은 규제 관련 충돌, 노란 선은 연방 차원의 조사를 의미

알트코인 약세에도 DOGE 코인의 시가총액은 무려 80조원 돌파

자료: TradingView, 미래에셋증권 디지털리서치팀
주: 도지코인 시가총액 단위는 십억 달러, ETHBTC는 비트코인 대비 이더리움 가격을 의미

자율주행에 관한 데이터 탈민감화 혹은 비식별화(Data Desensitization) 관점에서 중국의 CLA(Closed Loop Analytics) 시스템

데이터 수집 시 매핑 자격을 갖춘 차량에서만 데이터 수집해야 함. 즉, 매핑 데이터 관련 자격 확보가 필요한데 테슬라는 올해 중국 바이두와 관련 협약을 체결했고 모든 차량이 적격 차량.

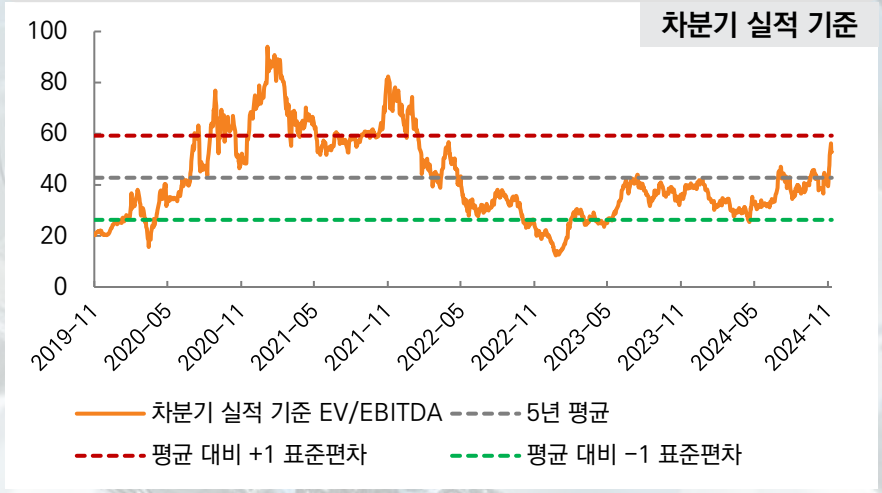
프라이빗 클라우드 기반의 데이터를 안전하게 보관해야 함. 테슬라는 상하이에 클러스터 보유.

데이터 처리에 있어서, 개인정보가 제거된 상태에서의 정확한 라벨링이 중요. 수동 개입은 최소화돼야 개인정보 노출 위험이 감소. 테슬라는 오토라벨링 및 탈민감화 기술을 수년 간 쌓아 옴.

OEM과 공급업체 간의 명확한 데이터 사용 권한도 설정해야 함. 테슬라는 수직적 통합 기업이기 때문에 이러한 권한분배 문제는 없음.

자료: Zhang Mengyu, 미래에셋증권 디지털리서치팀

테슬라의 지난 5년간 EV/EBITDA 추이, '현재 53배, 5년 평균 43배'



자료: Bloomberg, 미래에셋증권 디지털리서치팀

AI Insight

엔비디아의 옴니버스가 AI를 만나면?

지난 11월 7일에 공개된, 젠슨 황 엔비디아 CEO의 인터뷰 중 중요 내용 발췌 (1)

AI로 인해 세상이 바뀌었고 그 시작 단계에 있다. **향후 10년 동안 칩 그 자체가 아니라 (컴퓨팅의) 스케일 측면에서 매년 2~3배의 성능을 발휘할 수 있을 것이다.** 즉, 매년 2~3배씩 비용, 에너지를 절감할 수 있다. **이것이 "하이퍼 무어의 법칙"**이다. 이제 스케일링을 하는 데에는 새 방법이 필요하다. 공정 노드 축소나 3D 적층으로는 부족하다. 특정 SW 작업에 맞춰 최적화된 통합 HW 설계가 필요하다. (우리 팀의 생각: 엔비디아가 각종 산업의 생태계 개발자들과 협력해 가속 컴퓨팅 라이브러리를 만드는데 수십 년간 힘을 써온 이유) 또한, **지금 직면한 컴퓨팅 문제 중 가장 흥미로운 것은 "추론 시간의 변화"다. 복잡한 문제 해결과 추론 능력을 향상시키기 위한 AI 기법으로 Chain of Thought와 Tree Search가 등장했다**. (우리 팀이 지난 9월 발간한 인덱스 보고서 참조) 이 방법에서는, 추론 시간에 집약적인 연산이 지속적으로 소모돼야 하고 또한 몇 초 이내로 AI는 답을 내놔야 한다. 즉, 데이터센터는 매우 높은 처리량의 토큰을 생성해야 하는데다 지연시간도 매우 낮아야 한다. 그런데 높은 처리량과 낮은 지연시간은 상충되는 개념이다. (우리 팀의 생각: GTC에서 FP6/FP4 정밀도 옵션에 대해 강조한 이유. 엔비디아는 OpenAI의 로드맵을 어느 정도 알고 있다는 의미) **이걸 해결하기 위해 NVLink를 만들었다.**

현재 데이터센터의 AI 가속기 인프라는 세분화 되어 있다. OpenAI의 샘 알트만 CEO는 최근에야 (2017년에 출시된) Volta 기반 GPU를 폐기했다고 말했다. 물론 그들은 Ampere(2020년)도 갖고 있고, 다양한 Blackwell 제품들도 받아 보고 있다. 즉, 이 모든 산재된 컴퓨팅 자원을 다 활용해야 한다는 말이다. **엔비디아의 장점은 AI 트레이닝용으로 구축한 인프라가 추후에 인퍼런스에도 사용될 수 있다는 점이다. 즉, AI 훈련에 쓸 수 있다면 추론으로 전용된다는 얘기다**. 게다가, 하드웨어(GPU)는 그대로인 상태에서, **CUDA라는 소프트웨어 계층에서 지속적인 최적화 작업을 수행하면 같은 하드웨어라도 5배 더 나은 성능을 뽑아낼 수 있다**. (우리 팀의 생각: 상위 레벨의 AI 알고리즘이나 애플리케이션은 전혀 수정할 필요도 없이 CUDA 레벨에서의 최적화만으로 5배 성능 향상이 가능했다는 것. 이것이 엔비디아가 HW가 아니라 SW 회사라고 강조하는 이유)

엔비디아는 칩만 배송하는 게 아니라 전체 데이터센터를 구축한다. 이제 새로운 컴퓨팅 단위는 바로 데이터센터다. (우리 팀의 생각: 이론적 최고 성능과 실제 데이터 센터의 성능 간의 차이가 종종 크다는 점을 들어, 각종 NPU 업체들이 엔비디아 GPU보다 성능이 뛰어나다고 주장하는 것은 단지 파워포인트 슬라이드상의 숫자 뿐이라는 얘기) **xAI가 10만 개의 대규모 GPU 클러스터를 매우 빠른 속도로 구축한 사례에 대해서는, 일론 머스크가 1등 공신이지만, 엔비디아의 디지털 트윈 기술 덕분이기도 하다**. (우리 팀의 생각: 다른 경쟁자들은 이를 따라할 수밖에 없음. 실제로 경비행기를 타면서 xAI의 Colossus를 관찰, 모방하려 노력 중. 엔비디아의 옴니버스 또한 데이터센터 건설 경쟁의 수혜자) 먼저 소형 클러스터를 디지털 트윈으로 미리 구성하고 네트워크 구성, 전력, 냉각 등을 포함한 주요 설비와 관련한 모든 것을 시뮬레이션한 덕분에, 실제 클러스터를 구축할 때 걸림돌이 될 만한 요소들을 사전에 예측하고 해결할 수 있었다. **보통 몇 년 걸리는 초대형 컴퓨터 시스템을 단 몇 주 만에 완성할 수 있었고(참고: xAI는 엔비디아 서버를 Colossus에 들여온 지 19일만에 Grok 모델 훈련을 개시), 이러한 규모의 컴퓨터가 이 속도로 구축된 것은 처음이었다**. 우리는 몇 년 전부터, 데이터 센터를 제품으로 팔지는 않지만, 제품으로 취급하는 이니셔티브를 시작했다. (xAI 사례에서 보다시피) 우리는 그럴 능력을 이제 갖추었다. **고객들이 데이터센터 건설을 위한 공간과 전력, 냉각 장치만 있다면, "30일 이내"에 설치를 도울 수 있다. 그래서 엔비디아가 실제로 하고 있는 일이 "AI 파운드리"라는 것이다**. (이것의 유효시장에 관해서는) 1조 달러 규모의 데이터센터 인프라가 현대화되어야 한다.

자료: 미래에셋증권 디지털리서치팀

AI Insight

엔비디아가 했던 모든 것은 사실 AI 때문인 듯 하다

지난 11월 7일에 공개된, 젠슨 황 엔비디아 CEO의 인터뷰 중 중요 내용 발췌 (2)

현재 20~100만 대의 GPU 규모의 클러스터 건설에 대해 OpenAI, Anthropic, xAI, Google, Meta, Microsoft가 시도 중이다. (우리 팀의 생각: 젠슨 황이 생각하는 리딩 AI 업체들) **누가 정상에 먼저 오르지 않으려 하겠나? AGI의 가치는 너무 중요해서 시도하지 않을 수 없다**. (AGI의 유용한 사례 중 하나인) **"AI 기반 칩 설계"의 현재 효율성은 놀랍도록 뛰어나다. 실제로 엔비디아는 AI 없이는 Hopper를 만들 수 없었을 것이다**. AI가 인간보다 훨씬 더 큰 칩 설계 공간(회로 변수 및 조합)을 탐색할 수 있고, 병렬로 동시 탐색할 수 있기 때문이다. (우리 팀의 생각: 화웨이와 구글도 AI가 칩을 설계하는 방법 연구에 관한 노력을 올해부터 본격적으로 나섬) 이런 AI를 포함해, 미래의 업무 환경은 생물학적 직원과 AI 직원이 함께 일하는 하이브리드 형태일 것이다. 특히, SaaS 회사들은 본인들의 툴에 특화된 AI 엔지니어들을 개발할 것이다. **사람들은 SaaS 플랫폼들이 (AI 때문에) 무너질 것이라고 말하지만, 나는 반대로 생각한다.** 그들은 금광 위에 앉아 있는 것과 마찬가지다. 예를 들어, SAP나 Salesforce 같은 기업들은 자신들만의 특화된 AI 에이전트들을 개발할 수 있는 독특한 위치에 있다. 각각의 플랫폼은 자신만의 '언어'와 '문화'를 가지고 있어서, 이를 가장 잘 이해하는 AI 에이전트들을 개발할 수 있기 때문이다. 물론 **엔비디아는 옴니버스를 위한 OpenUSD라는 언어를 갖고 있고, 여기서 가장 뛰어난 AI 에이전트는 바로 엔비디아만이 만들 수 있다.** (우리 팀의 생각: 엔비디아가 AI 에이전트를 구축할 수 있는 플랫폼인 NeMo와 그것을 대량으로 배포할 수 있는 서비스인 NIM에 집중하는 이유가 바로 이것. 게다가 3D 시뮬레이션 및 협업 플랫폼인 옴니버스를 통해 "디지털 트윈" 생태계의 에이전트 시장은 본인들이 최강이 될 것임을 선언한 것과 마찬가지. 엔비디아의 해자는 더 이상 GPU에서만 발휘되지 않을 것. 시장은 1~2년 안으로 이를 이해할 것)

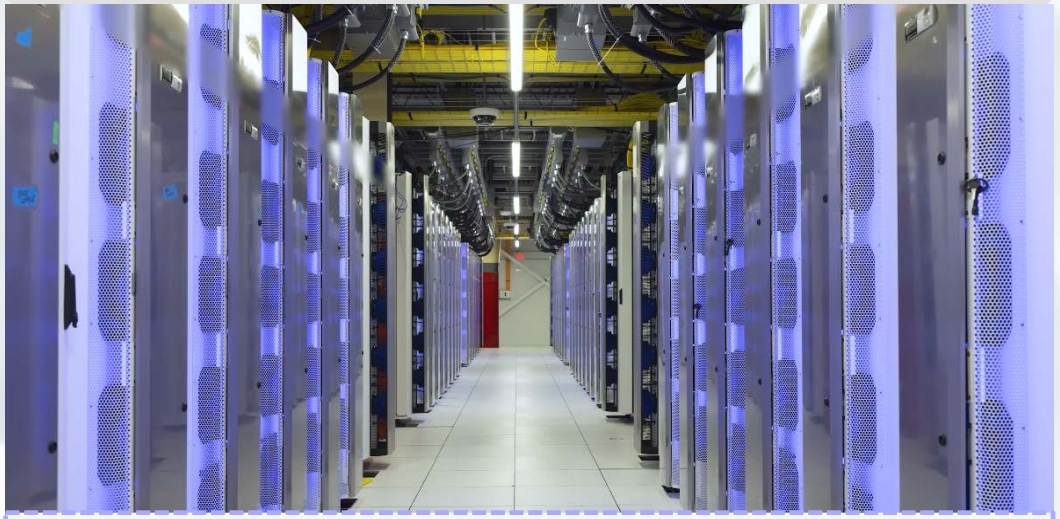
AI는 과학과 공학에 근본적으로 영향을 미치고 있다. 양자 컴퓨팅에서 양자 화학에 이르기까지, 모든 과학 분야를 망라한다. 현재 전 세계 모든 엔지니어와 과학자들의 오늘날 일하는 방식이 미래의 초기 징후라고 생각한다면, 인류는 곧 AI의 거대한 해일이 모든 것을 변화시키는 것을 볼 것이다. 생성 AI가 기반이 되지 않는 과학적 돌파구나 엔지니어링 돌파구는 없을 것이다. 이제 이것을 꽤 확신한다. **아무것도 예전과 같지 않을 것이다**. (우리 팀의 생각: OpenAI가 제시한 AI의 발전 경로의 4번째 단계인 "혁신가"로서의 AI에 대해 젠슨 황은 매우 긍정하는 것. 그의 공격적인 tone 변화가 OpenAI의 추론 강화 언어모델인 o1의 출시 이후 이뤄진 것은 우연이 아닐 것)

AI는 나 자신에게도 선생님이다. ChatGPT나 상황에 따라 Perplexity에 접속해서 공부하고 있다. 심지어 이미 알고 있는 거의 모든 것들, 사실이라고 여기는 것들, 스스로 전문가라고 생각하는 분야의 것들조차도 AI를 통해 다시 한 번 확인한다. 이것이 학습과 지식 검증의 새로운 패러다임이 되었다. (우리 팀의 부탁: 제발 AI 쓰세요) 물론 단순히 개인적인 학습에 국한되지 않는다. **전체 조직이 AI를 통해 변화하고 있다. 예를 들어, 엔비디아는 AI가 칩 설계의 필수적인 "협업 인력"이 됐다. 이것은 단지 시작일 뿐이다. 앞으로 2-3년 안에, 인류는 AI가 주도하는 과학적 돌파구들을 보게 될 것이다**. 양자 컴퓨팅에서부터 신약 개발, 기후 모델링에 이르기까지, 모든 분야가 AI에 의해 변화될 것이다. (우리 팀의 생각: 엔비디아는 양자 시뮬레이션을 위한 가속 라이브러리인 cuQuantum, 신약 개발을 위한 기반 모델인 BioNeMo, 기후 예측 알고리즘인 CorrDiff를 개발) 미래의 과학자나 엔지니어는 **AI 없이 연구하는 것을 상상조차 할 수 없게 될 것이다**. 마치 오늘날 우리가 컴퓨터 없이 연구하는 것을 상상할 수 없는 것과 똑같다. 이것이 바로 **우리가 목격하고 있는 변화의 규모다**. (우리 팀의 생각: 젠슨 황은 1조 달러의 데이터센터 현대화 작업이 끝나면, 그 다음은 수조 달러의 "인텔리전스" 시장이 열린다고 보고 있음)

자료: 미래에셋증권 디지털리서치팀

관련 자료: 엔비디아와 테슬라의 컴퓨터 클러스터링

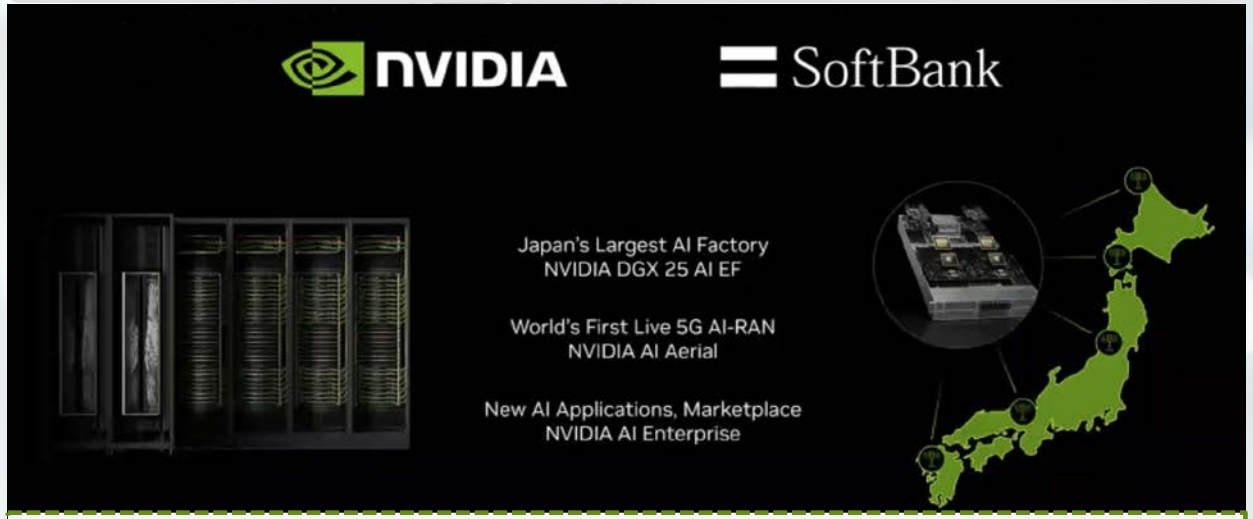
세계 최대 데이터센터인 xAI의 "Colossus"의 내부 모습, '10만 개 이상의 GPU를 갖춤'



엔비디아의 디지털 트윈과 xAI의 실행력으로 122일만에 10만 대 이상 GPU 클러스터가 지어짐.
CSP들이 가진 것의 약 4배의 컴퓨팅이지만, 여전히 건설 중… 전체 건설 목표의 1단계일 뿐!

자료: YouTube(@ServeTheHome), 미래에셋증권 디지털리서치팀

DGX B200 시스템을 도입한 소프트뱅크는 일본 최대의 AI 데이터센터(우선 5G 용도 집중)를 구축할 예정



해당 DGX 시스템은 x86 CPU를 쓰지만, 향후 엔비디아 전용 CPU를 탑재한 GB200으로 전환할 계획
젠슨 황이 강조하는 "소버린 AI"의 필요성에 일본이 화답한 것. 마사요시는 엔비디아가 저평가되었다고 최근 언급

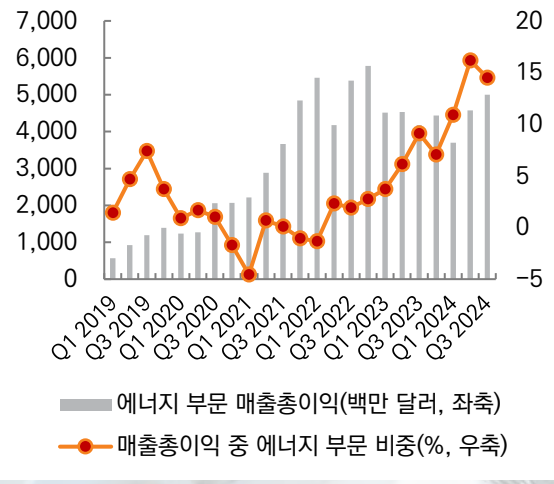
자료: Nvidia, 미래에셋증권 디지털리서치팀

대규모 AI 모델 경쟁에서 필수적인 대규모 컴퓨팅 클러스터는 결국 ESS를 필요로 할 것

전력 변동이 일으키는 문제	– GPU들이 동시에 대규모 연산을 시작할 때 순간적으로 전력 수요 급증. 이런 급격한 전력 수요 변화는 전압 불안정성을 유발 – 불안정한 전압은 GPU의 동작 주파수에 영향. GPU 동작 주파수 변화는 연산 timing에 영향. 병렬 처리되는 연산들 간의 동기화가 흐트러질 수 있음. 즉, 부동소수점 연산의 정밀도 및 gradient 동기화에 영향.
모델 훈련에 미치는 일관성 문제 및 안정성 저하	– Grok 등의 대규모 언어모델에서 필요한 self-attention 계산과 같은 복잡한 정밀 가중치 조정 연산에서 문제가 발생할 수 있음. – 이를 방지하려면 checkpoint 저장 빈도를 높여야 하고, 불안정한 구간에 대한 재학습이 필요해짐. 결국, 전체 훈련 시간은 늘어나게 됨.
테슬라 Megapack을 사용한 해결책	– 전력 공급의 완벽한 안정성 확보. 그리드의 불안정성으로부터 격리 – **훈련 안정성 극대화 및 재현 가능한 결과 보장, 훈련 시간 최적화** 등의 효과를 낼 수 있음.

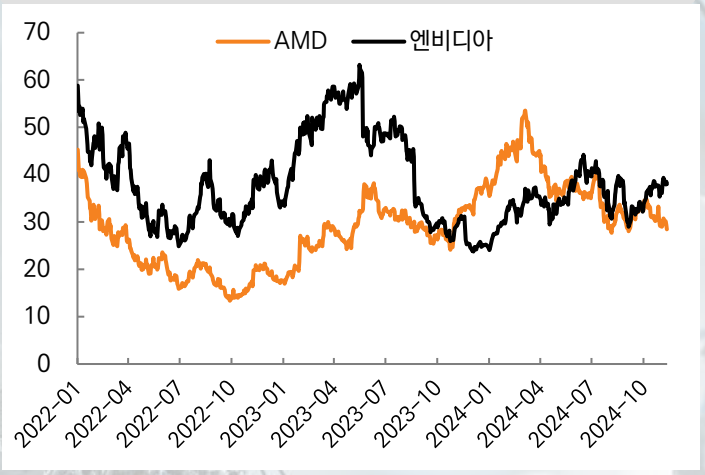
자료: 미래에셋증권 디지털리서치팀

테슬라의 분기별 에너지 부문 매출총이익



자료: Tesla, 미래에셋증권 디지털리서치팀

현재 PER 격차는 엔비디아의 AI 해자에 비하면 적은 듯



자료: Nvidia, 미래에셋증권 디지털리서치팀
주: 선행 12개월 실적 기준 PER 수치

Paper of the Week

The Road Less Scheduled – Meta, Google, Samsung

딥러닝 모델의 학습에 있어, 학습률을 결정하는 것은 모델 성능에 매우 중대한 영향을 미침. 그러나 시간에 따라 학습률을 조절해야 하므로, 학습 종료 시점(T)을 미리 알아야만 최적의 학습률 스케줄을 설계할 수 있음. 그리고 시간에 종속된다는 특성 때문에 **학습 종료 시점이 바뀌면 모델의 최종 성능도 달라질 수 있으며, 학습이 중단되면 재시작 시 어떤 학습률을 사용해야 할지 모른다는 점 등 여러 어려움이 존재**했음.

따라서 학습률 스케줄링을 수식화하기 위한 노력도 있었음. 그러나 이론적으로 최적의 방법으로 알려진 폴리악-루퍼트(Polyak-Ruppert) 평균화 기법이 실제로는 기대만큼의 성능 향상을 보여주지 못해, 모델의 최적 학습률을 결정하는 방법은 현재까지 명확하게 정립되지 않았음.

메타, 구글 및 삼성 AI 연구소는 논문을 통해 **스케줄-프리 확률적 경사하강법**(Schedule-Free Stochastic Gradient Descent, SGD)을 대안으로 제시함. 이는 총 세 가지 값을 모두 고려하는 수식을 통해 자동으로 학습률이 조정되도록 한 것. 특정 시점에서 **모델의 현재 파라미터(z_t), 해당 시점까지의 평균 파라미터(x_t), 그리고 이 두 값의 비율**(y_t, 주로 현재 1: 평균 9)을 세 개의 변수로 두고, **평균값에 비해 현재값이 얼마나 떨어져 있는지를 비교하며 학습률이 자동으로 조정되도록 하는 개념임**. 이는 기존 SGD의 (빠른 학습이라는) 장점은 유지하면서 에러를 더 안정적으로 줄여나갈 수 있음. 현재값은 계속해서 빠르게 변화하지만 평균값이 최종 결과를 보정해주기 때문.

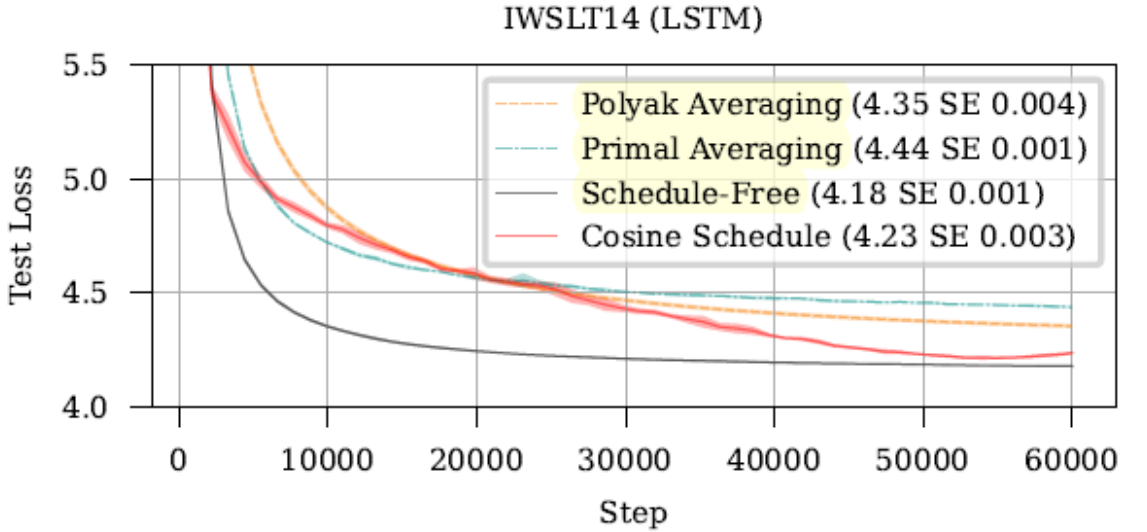
특히 이 과정에서 기존 기법들과 달리 **매우 큰 학습률을 사용하더라도 안정적으로 학습이 진행되며, 중간에 학습을 멈추는 경우에도 안정적으로 재시작할 수 있다는 점** 역시 매우 인상적임. GPU 자원 부족으로 학습이 예기치 못하게 중단되거나, 모델의 중간 결과를 확인하고 추가 학습 여부를 확인하고자 하는 경우에 중단된 시점의 값들만으로 학습을 재개할 수 있어 모델 성능을 확인하고 긴급 상황에 대응하는 데 유연성을 확보할 수 있기 때문.

28개의 벤치마크 문제에 대한 실험 결과, 스케줄 프리 기법은 대부분의 경우 기존 방식인 정교하게 튜닝된 학습률 스케줄보다 더 좋은 성능을 보여주었음. 컴퓨터 비전, 자연어 처리, 음성 인식 등 광범위한 분야에서 그 효과를 입증했으며, MLCommons 2024 대회의 알고리즘 자동 튜닝 부문에서도 우승을 차지하기도 했음.

학습률 설정의 자동화로 실용성을 확보하면서 모델 성능도 개선했지만, 여전히 초기 학습률과 같은 몇몇 하이퍼파라미터는 수동으로 설정해줘야 해 완벽한 자동화라고 보기는 어려움. 다만 사람이 제공하는 학습률보다 **자동으로 변화하는 학습률로 더 좋은 성능을 달성할 수 있었다는 점에서, 다른 하이퍼파라미터들을 자동화함에 따라 추가적인 성능 개선이 가능할 것으로 기대함**.

자료: "The Road Less Scheduled", Meta, Google, Samsung AI Research.

스케줄-프리 기법(검은 선)이 전통적인 평균법들보다 빠르게 수렴하는 모습.
코사인 스케줄링(빨간 선)과 같은 튜닝된 스케줄링 기법보다도 아웃퍼폼하기도



$$y_t = (1-\beta)z_t + \beta x_t$$

$$z_{t+1} = z_t - \gamma \nabla f(y_t, \zeta_t)$$

$$x_{t+1} = (1-c_{t+1})x_t + c_{t+1} z_{t+1}$$

→ 평균값과 현재값에 그래디언트를 적절히 분배

→ 현재 위치에서 γ(학습률)만큼 학습률을 줄이도록 이동

→ 현재까지 지나온 z값들의 평균

자료: "The Road Less Scheduled", Meta, Google, Samsung AI 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)	애플	AAPL US	225.12	4,786.7	1.2	-1.0	1.9	21.1	22.4	17.5	0.40	01-30	온디바이스 AI 관련 엔드유저 데이터보유 및 서비스 배포능력은 타의 추종 불허
	테슬라	TSLA US	330.24	1,491.2	14.5	51.6	58.9	92.1	47.6	32.9	0.34	01-24	자율주행 소프트웨어 FSD v12의 미국과 중국 시장 내 공공 출시 임박
	삼성전자	005930 KS	50,600	302.1	-12.0	-16.8	-34.1	-34.7	-27.2	-34.6	0.16	01-08	온디바이스 AI, HBM, 그리고 파운드리 영역에서 AI 수혜 기대(Galaxy AI)
	퀄컴	QCOM US	160.42	250.7	-7.3	-5.6	-4.8	-12.1	32.2	12.5	0.56	01-31	생성 AI 처리 강화용 온디바이스 AP 및 CPU 설계(X Elite)
	스냅	SNAP US	11.11	26.2	-5.9	-0.2	20.4	-29.5	-1.7	-34.4	0.25	02-06	이미지 생성 모델을 탑재한 증강현실 서비스(AR 필터, AR 렌즈)
비즈니스 효율화	어도비	ADBE US	532.50	329.7	5.5	7.5	-0.5	10.2	-9.8	-10.7	0.43	12-14	세계 최고의 미디어편집 툴로 막대한 생성 AI 서비스 배포능력을 지닌 기업(Firefly)
	세일스포스	CRM US	341.72	459.5	11.3	18.6	34.4	23.5	59.4	30.4	0.34	12-03	세계 최고의 CRM 업체로 생성 AI를 탑재해 사용성 강화(Einstein GPT)
	서비스나우	NOW US	1,045.55	303.0	5.1	11.4	27.7	43.2	63.8	48.0	0.45	01-24	워크플로우 자동화 기업으로 생성 AI를 탑재해 사용성 강화(Now Assist)
	크라우드스트라이크	CRWD US	347.88	120.0	8.7	8.6	40.2	9.0	73.1	36.3	0.39	11-26	기업 고객 대상 생성 AI 챗봇 기반 클라우드 및 엔드포인트 보안(Charlotte AI)
	IBM	IBM US	210.92	274.3	-0.5	-8.9	11.3	28.0	47.6	33.6	0.17	01-29	기업 고객용 AI 어시스턴트(상담사 챗봇 구축, 업무 자동화, 코딩 지원)
	SAP	SAP GY	219.45	400.8	3.1	5.2	14.5	25.8	64.1	59.3	0.31	01-28	기업 고객용 AI 어시스턴트(SAP Business AI: ERP 등 프로세스 지능형 자동화)
	유니티	U US	19.13	10.8	-11.7	-11.5	17.7	-15.7	-32.1	-53.2	0.25	02-26	올해 출시 예정인 차세대 게임엔진에 AI 기능 대거 탑재(Sentis, Muse)
	모더나	MRNA US	42.14	22.8	-18.7	-27.7	-49.2	-66.5	-40.0	-57.6	0.16	02-21	OpenAI와의 협력으로 mRNA 신약 개발 등 모든 업무에 AI 활용(Dose ID GPT)
	루닛	328130 KS	40,400	1.2	3.6	-1.3	11.8	-26.9	-57.1	-50.7	0.08	11-13	AI 기반 암 검출/진단 보조 솔루션(인사이트)
기반 모델	알파벳	GOOGL US	178.88	3,093.6	1.3	9.6	9.1	6.0	35.8	28.4	0.40	01-30	OpenAI와의 AGI 경쟁에 있어 대적할 수 있는 거의 유일한 업체(Gemini)
	메타 플랫폼스	META US	580.00	2,060.4	1.4	-1.7	9.8	24.2	76.7	64.4	0.41	01-31	'거의 완전한' 오픈소스 LLM 개발 및 배포를 선도(Llama)
	알리바바	9988 HK	90.55	313.3	-5.3	-13.9	15.6	12.0	18.4	22.4	0.09	11-15	중국의 하이퍼스케일러이자 중국 최고의 오픈소스 언어모델 성능을 보유(Qwen)
	바이두	9888 HK	83.25	42.2	-7.3	-15.4	1.0	-23.8	-19.8	-28.3	0.13	11-21	중국 내 자연어처리 및 자율주행 모델의 전통적 강자(Ernie Bot, Apollo)
	센스타임	20 HK	1.61	10.2	-7.5	-0.6	50.5	11.0	4.5	38.8	0.08	03-26	중국의 SOTA급 멀티모달 AI 모델을 만들 잠재력이 있는 기업(SenseNova)
	네이버	035420 KS	182,200	29.3	1.4	6.3	15.8	-1.2	-7.9	-18.7	0.09	02-03	자체 개발한 한국형 LLM 개발 및 AI 서비스 출시 예정(HyperCLOVA)
AI Ops	몽고DB	MDB US	300.89	31.3	7.1	2.7	23.8	-16.2	-20.4	-26.4	0.34	12-05	AI 모델을 구축할 때 필수 인프라로 떠오른 '벡터 데이터베이스' 기능 보유(Atlas)
	클라우드플레어	NET US	94.16	45.6	1.8	0.7	15.2	28.0	47.4	13.1	0.39	02-07	트래픽 처리(CDN) 업체로, 실시간 AI 추론 작업에서의 역량 대두(Workers AI)
	팔란티어	PLTR US	60.70	194.4	9.3	39.5	99.7	189.9	208.0	253.5	0.36	02-05	실제 조직 운영시의 의사결정 지원 및 자동화를 위한 AI 개발 플랫폼(AIP)
	포스 패러다임	6682 HK	48.25	4.1	31.1	118.8	2.8	-4.8	-9.8	-3.9	0.01	03-20	실제 조직 운영시의 의사결정 지원 및 자동화를 위한 AI 개발 플랫폼(Sage)

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

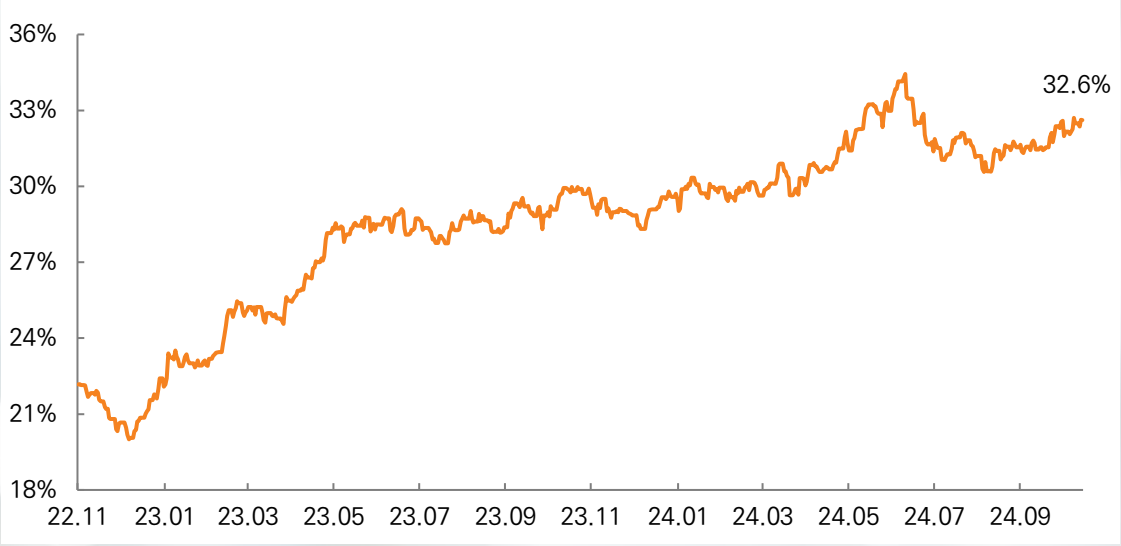
AI Peer Table (2)

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			
클라우드	마이크로소프트	MSFT US	425.20	4,446.9	1.2	2.1	2.9	3.1	16.8	13.7	0.53	01-30	OpenAI의 AI 모델 라이센스에 대한 독점권을 갖고 있는 세계 2위의 CSP
	아마존 닷컴	AMZN US	214.10	3,166.8	3.4	13.4	25.8	14.8	50.2	40.9	0.44	01-31	최고의 CSP로써 작년 9월 Anthropic에 40억 달러 투자하며 AGI 경쟁 참여
	오라클	ORCL US	189.63	739.2	4.6	7.9	41.6	63.8	68.2	82.1	0.40	12-11	기업들의 자체 인프라 수요에 대응한 '온프레미스 서버' DB 관리 시스템의 강자
AI 인프라	엔비디아	NVDA US	146.27	5,047.1	0.5	8.5	26.0	61.8	200.9	195.4	1.00	11-20	AI 훈련은 GPU가 필수, GPU 현존 최강자는 엔비디아
	브로드컴	AVGO US	173.58	1,140.4	-3.3	-4.4	11.5	30.6	86.1	57.2	0.64	12-12	VMware 인수로 클라우드 가상화 역량도 확보한 네트워킹 반도체의 최강자
	AMD	AMD US	139.30	318.0	-4.0	-17.0	-1.3	-7.5	19.3	-5.5	0.64	01-30	HPC CPU에서 업계리딩, GPU에서 엔비디아의 적수로 꼽히는 패스트 팔로워
	인텔	INTC US	24.92	151.2	-0.5	5.8	21.7	-17.8	-33.9	-49.7	0.30	01-24	GPU 중심 서버 설계에는 후발주자. 파운드리 기업으로서의 지정학적 가치 부상
	마이크론	MU US	99.92	155.8	-10.7	-6.5	2.7	-18.6	34.2	17.4	0.55	12-20	HBM에서 SK하이닉스를 추격하는 미국의 메모리 반도체 업체
	SK하이닉스	000660 KS	182,900	133.2	-7.3	-2.5	-1.9	-1.0	41.6	29.9	0.26	01-24	GPU 병목의 키를 쥐고 있는 HBM에 있어 글로벌 선두주자
	시놉시스	SNPS US	557.96	120.6	0.9	3.3	5.5	0.1	6.7	8.4	0.67	12-04	반도체 EDA 업체로, 3D 엔지니어링 시뮬레이터 업체 앤시스와 M&A 논의 중
	암페놀	APH US	72.33	122.7	0.2	10.3	11.9	14.2	68.9	46.8	0.51	01-22	AI 및 HPC 서버에 필요한 고속 I/O, 메모리 및 스토리지 커넥터 제조
	버티브	VRT US	124.53	65.7	3.4	11.3	63.4	31.4	194.9	159.5	0.55	02-21	데이터센터의 열 관리를 위한 액체 냉각 방식에 있어 글로벌 선두업체
	Arm	ARM US	135.09	199.7	-6.6	-10.8	9.1	15.2	161.3	79.8	0.43	02-07	AI 가속기 설계 위한 다수의 IP 블록 소유한 업체로 설계 라이선스/로열티 사업
	마벨 테크놀로지스	MRVL US	90.07	109.7	-0.6	21.7	40.7	30.7	69.1	49.9	0.61	12-03	데이터 인프라용 네트워킹 및 스토리지 ASIC 설계
	SMIC	981 HK	27.15	67.7	-3.9	1.9	61.0	66.6	23.1	36.7	0.07	02-06	최근 화웨이와 밀접히 협력 중인 중국 유일의 미세공정 가능 파운드리
	TSMC	2330 TT	1,035.00	1,162.8	-2.8	-1.0	9.6	26.5	84.2	76.8	0.21	01-20	압도적인 미세공정 점유율을 가진 대만의 반도체 파운드리
	콴타	2382 TT	322.00	53.9	-1.1	13.6	18.8	11.6	59.6	48.2	0.17	11-14	AI 서버 공급망에서 GPU 기반 보드 및 서버 시스템 조립 역할을 하는 대만 업체
	위스트론	3231 TT	121.00	15.2	0.8	11.0	16.9	11.1	26.3	25.7	0.18	01-20	AI 서버 공급망에서 GPU 기반 보드 및 서버 시스템 조립 역할을 하는 대만 업체
	Alchip	3661 TT	2,290.00	7.9	-5.0	19.3	-3.5	-14.7	-29.6	-29.4	0.16	03-03	세계에서 AI 반도체 관련 매출 비중이 가장 높은, ASIC 설계 관련 대만 업체
	소프트뱅크	9984 JP	9,137.00	121.6	-0.9	2.0	14.3	9.7	60.3	45.9	0.19	02-07	동사가 약 25%의 지분을 가진 '비전펀드'는 AI 관련 유니콘 기업에 투자를 집중

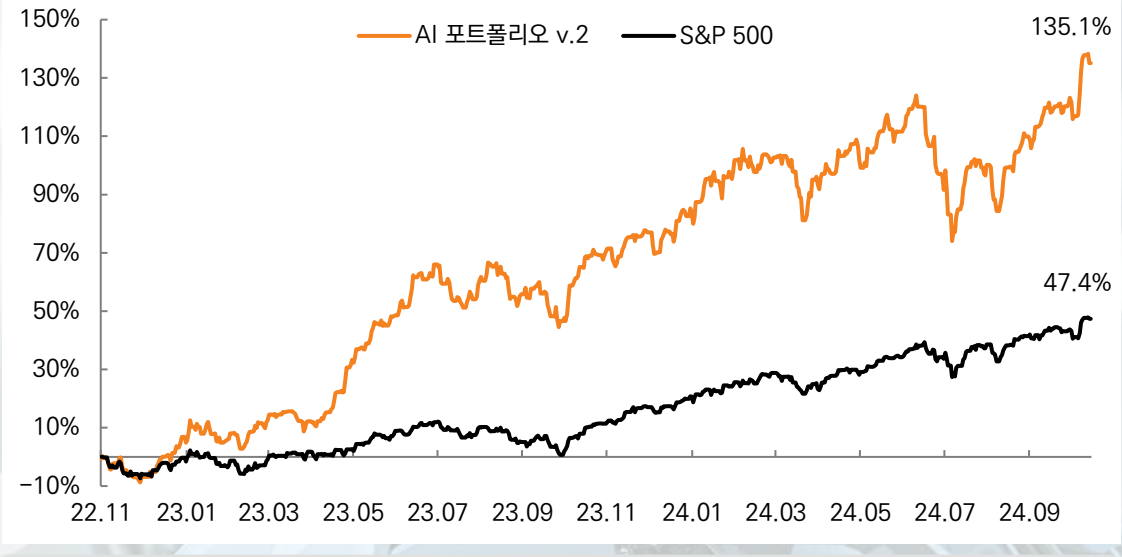
자료: Bloomberg, 미래에셋증권 디지털리서치팀
주: 실적발표 예정일은 수시로 변동될 수 있음

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)

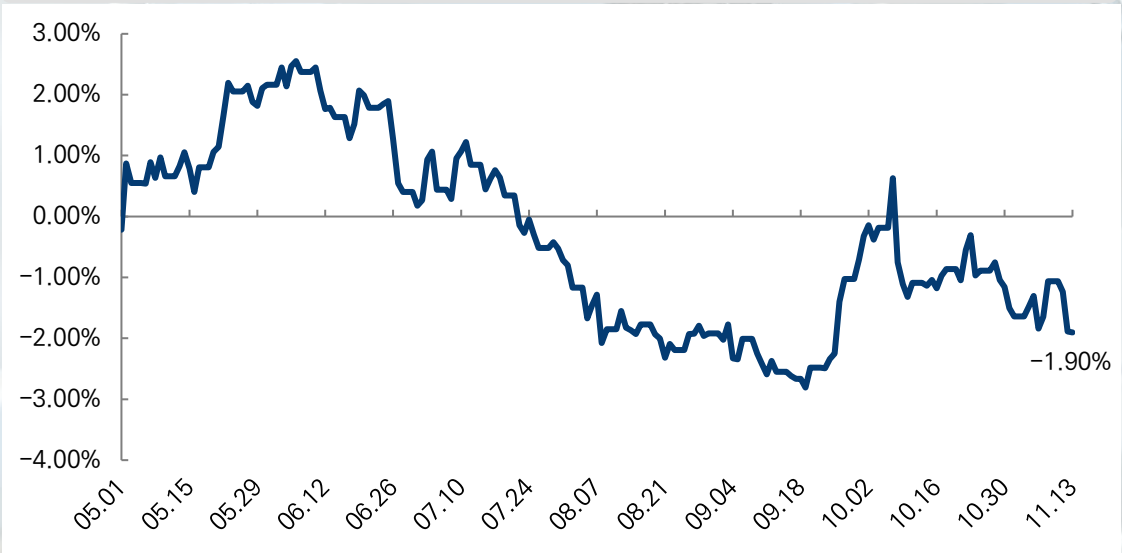


LLM 벤치마크 LiveBench 상위 7개 모델 (동일 모델은 생략)

순위 (전주 변동)	모델명	전체 평균 점수	기관명	라이선스	지식 컷오프
1(-)	o1-preview	66.02	OpenAI	독점 폐쇄소스	2023년 10월
2(-)	Claude-3.5 Sonnet -20241022	60.33	Anthropic	독점 폐쇄소스	2024년 4월
3(-)	o1-mini	59.09	OpenAI	독점 폐쇄소스	2023년 10월
4(-)	GPT-4o	56.03	OpenAI	독점 폐쇄소스	2023년 10월
5(-)	Gemini 1.5 Pro	54.94	구글	독점 폐쇄소스	2023년 11월
6(-)	Llama 3.1-405B	53.24	메타	별도 라이선스 필요 (상업적 사용허가)	2023년 12월
7(-)	Qwen2.5-72B	52.89	알리바바	별도 라이선스 필요 (상업적 사용허가)	2024년 9월

자료: Bloomberg, LiveBench, Huggingface, 미래에셋증권 디지털리서치팀

이전 포트폴리오 버전(v.1) 대비 AI 유니버스 누적 수익률 차이 (%p, 24.04.30 기준)



Compliance Notice

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