

AI Weekly

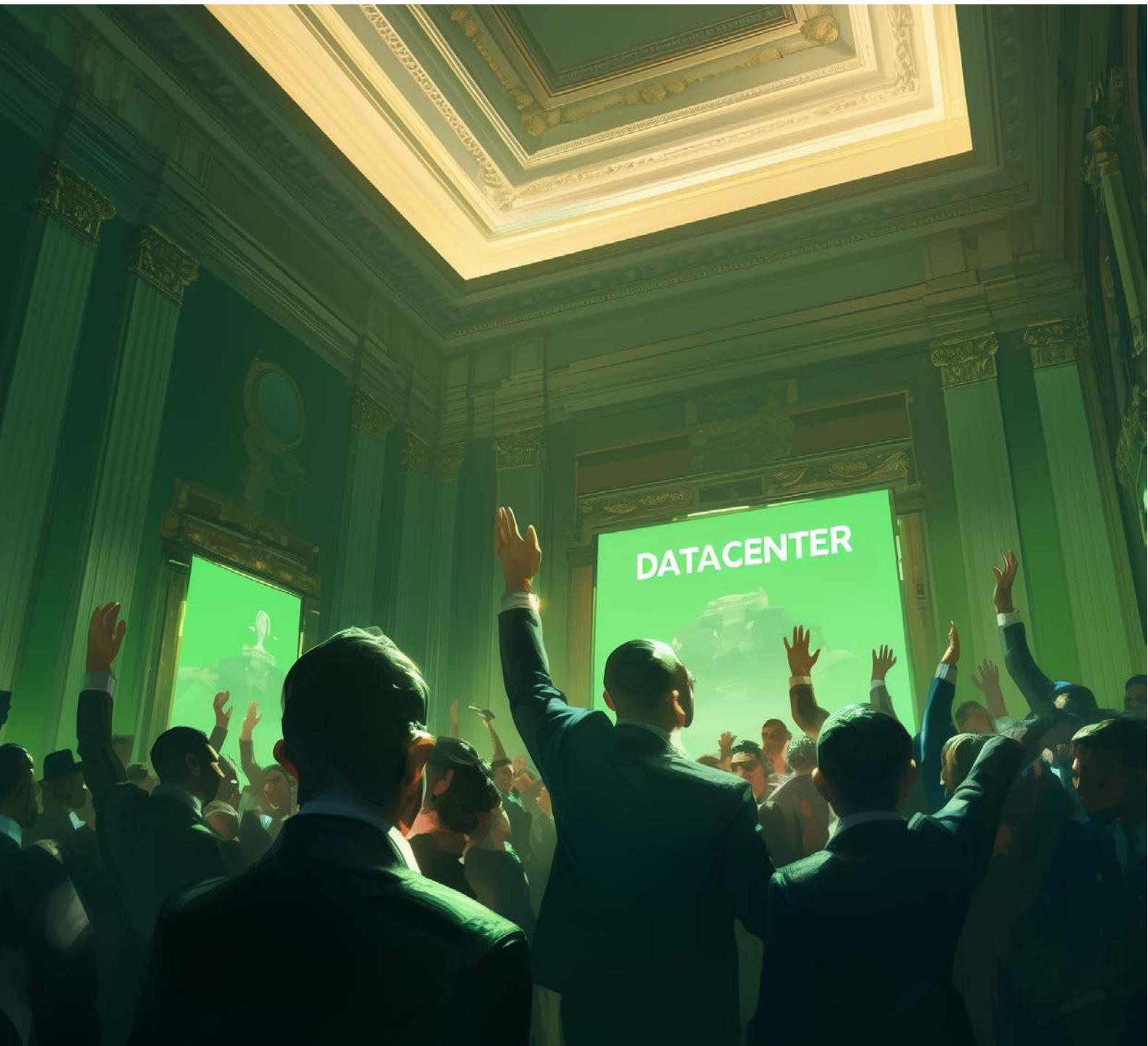
인공지능의 특이점은 이미 달성됐다

한종목

chongmok.han@miraeasset.com

김은지

eunji.kim.a@miraeasset.com



Highlight of the Week

I. AI Issue: AGI 임박의 시그널

샘 알트만은 “Reflections” 글을 게시했다. AGI를 만들 확실한 방법을 알고 있다고 했다. 샘 알트만은 이제 초지능 연구로 직행하겠다는 포부를 밝혔다.

한편, 러-중 AI 협력은 AGI 대중화 시대에 미국 주도의 질서에 대항하기 위한 절박함의 표현이라 할 수 있다. BRICS 국가들과 연대하여 미국의 AI 주도권에 도전하고, “G7 vs BRICS” 구도의 AI 경쟁이 본격화될 것으로 사료된다. 중국의 AI 굴기도 주목할 만하다. 미국의 대중 제재는 오히려 중국 AI 기업들의 성장을 촉진한 듯 보인다. 특히, DeepSeek-V3는 제한된 GPU 자원으로도 대규모 모델 학습 및 추론을 가능하게 하는 기술 혁신을 선보였다. Cambricon은 미국의 제재 대상이지만, 화웨이와의 협력을 통해 성장 중이다.

II. AI Issue: CES 2025 중 전해진 소식들

아마존 AWS는 AI 데이터센터에 110억 달러, 마이크로소프트는 800억 달러를 투자 계획을 발표했다. UAE 국부펀드도 미국 데이터센터에 200억 달러 투자 계획을 밝혔다. 한편, 엔비디아와 퀄컴의 삼성 파운드리 채택에 관한 최근 루머는 사실이 아닐 가능성이 높다.

AMD는 다양한 신제품을 쏟아냈지만, RDNA 4와 FSR 4에 대한 정보 부족으로 시장 반응은 별로 긍정적이지 않다. AMD 전략은 중급형 AI PC 시장에 집중하는 것으로 읽힌다.

III. Paper of the week: AI가 노동시장에 미치는 영향 – 변곡점 효과

연구진은 AI 발전이 근로자에게 미치는 영향이 긍정적에서 부정적으로 전환되는 '변곡점' 개념을 제시. 프리랜서 구인구직 사이트의 데이터를 분석한 결과, 번역은 이미 변곡점을 넘어 AI로 대체되는 단계에 진입. 웹 개발은 아직 생산성 향상 단계에 있음. GPT-4 업그레이드 이후 분석에서도 글쓰기 관련 직종은 지속적 대체 효과가, 운영/창의성 관련 직종은 생산성 향상 효과가 나타났음. 개인은 AI를 활용한 생산성 향상이 필요한 상황.

표 1. AI 관련 주요 일정

일	월	화	수	목	금	토
5	6	7	8	9	10	11
.	.	.	· 삼성전자 실적(잠)	.	.	.
12	13	14	15	16	17	18
.	.	.	.	· TSMC 실적	.	.
19	20	21	22	23	24	25
.	· 위스트론 실적(예)	.	· APH 실적	.	· NOW 실적(예) · 하이닉스 실적(예)	.

자료: Bloomberg, 미래에셋증권 리서치센터

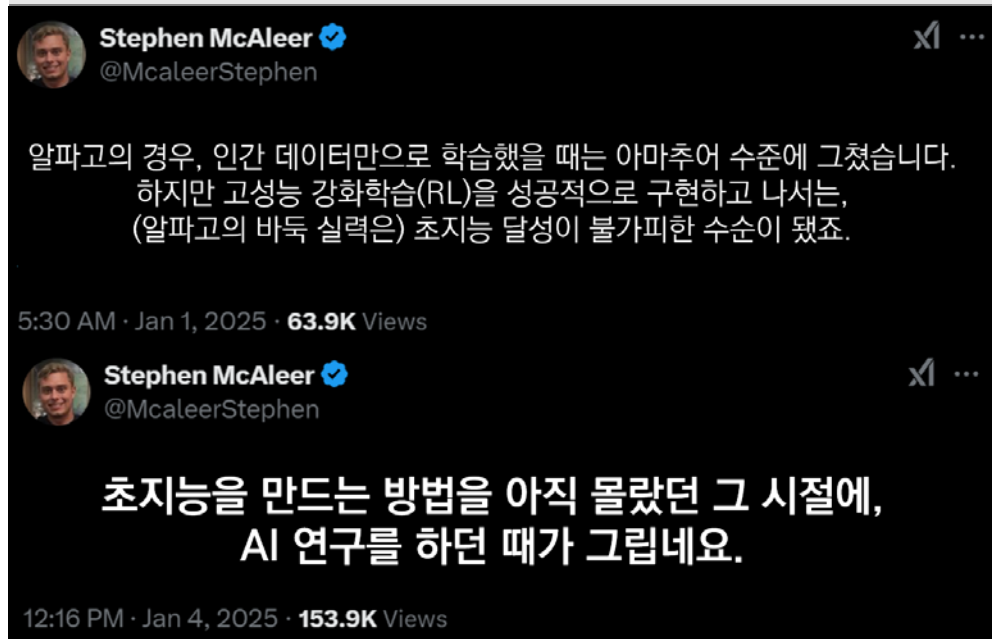
I. AI Issue

1. AGI가 진짜 다 왔나?

(1) OpenAI로부터 계속 언급되는 그 단어... “AGI”

2025년 들어서 OpenAI와 구글을 포함한 AI 연구원들의 발언 뉘앙스가 훨씬 강해지는 것이 체감될 정도다. 특히 OpenAI의 경우에는 o1, o3로 명명되는 Test-Time Computing 기반의 모델 출시 이후에 소속 연구원들이 잇따라 ‘내가 말할 수는 없지만 굉장한 것을 보았다’는 식으로 소위 teasing을 하고 있다. 가장 최근의 트윗 중 흥미로운 내용은 AI 에이전트와 관련한 안전을 담당하고 있는 한 연구원의 글에서 나왔다.

표 2. OpenAI에서 '에이전트 안전'을 담당하고 있는 연구원이 2025년에 남긴 트윗 2개
인간을 아득히 뛰어넘는 성능의 AI를 말하는 “초지능”을 만드는 법을 이제 알고 있다는 소리!



자료: X(@McaleerSteph), 미래에셋증권 리서치센터

이 연구원의 트윗 두 가지를 종합하면 생각해보면 다음과 같다:

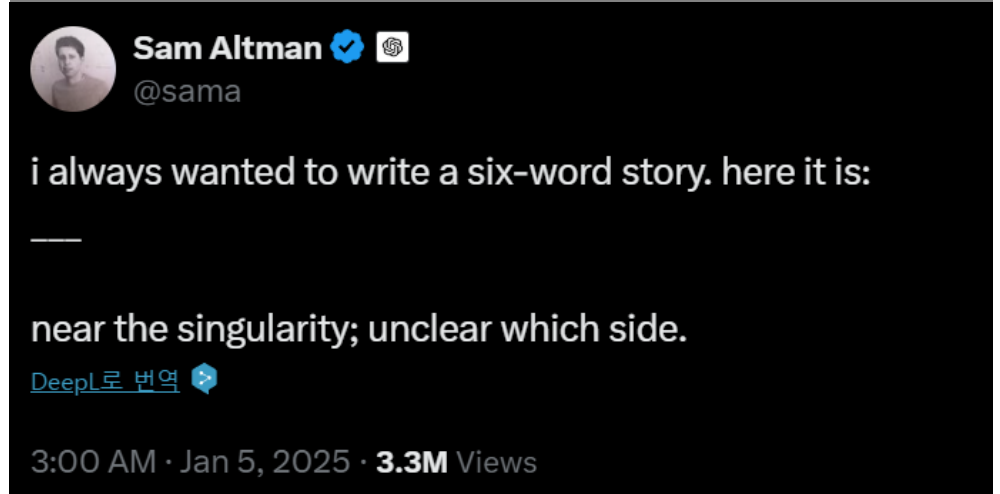
“알파고처럼 언어모델에 있어서도 고성능 강화학습을 성공적으로 구현했다. 그것이 초지능 달성을 위한 방법이다. 이제 우리는 그 방법을 알고 있다.”

즉, OpenAI에서는 이제 어떻게 해야 하는지 방법을 깨우쳤으니 무엇이 부족하고, 어떤 것을 더 충당하는지 그 방법이 굉장히 명확해졌으리라 생각된다. 즉, 이제 문제는 스케일링과 같은 공학적인 문제인 것이지 어떠한 새로운 과학적 돌파구는 그리 고민하지 않아도 된다는 자심만만한 소리처럼 들리기도 하다. 비단 저 사람만의 생각일 것이라고는 생각되지 않는다. AI Weekly를 준비하면서 최근 몇 달 간 이와 비슷한 뉘앙스로 글을 쓴 수많은 OpenAI 직원들의 계정을 관찰해왔기 때문이다.

그리고, 그 다음날인 좀 더 심오한 글이 OpenAI의 CEO로부터 나왔다. 샘 알트만 CEO는 단 6 글자의 트윗을 남겼다. 그 또한 AGI에 대해 많은 고민을 하고 있는 듯했다.

표 3. 지난 1월 5일에 샘 알트만이 남긴 기술적 특이점(Singularity)에 대한 언급

"특이점 근처: 어느 쪽인지 불분명"이라는 문구는 기술적 특이점에 매우 가까워졌음을 시사



자료: X(sama), 미래에셋증권 리서치센터

위와 같은 샘 알트만의 트윗은 OpenAI CEO로서 그가 겪은 AGI 개발 과정에서의 불확실성과 다가오는 미래에 대한 복합적인 감정을 드러낸다. 먼저 **"near the singularity"**라는 샘 알트만의 말은, AGI, 즉 인공지능의 도래가 가까워졌음을 의미한다. 이는 o1와 같은 새로운 패러다임의 AI 모델등장이 기술 발전의 중요한 변곡점임을 인지하고 있음을 나타낸다고 볼 수 있다. 그에게 있어 'singularity'는 AI가 인간의 지능을 초월하는 기술적 특이점, 혹은 그에 준하는 중대한 변화의 시기를 의미한다.

그리고 "unclear which side"는 AGI 개발 과정에서 마주하는 불확실성을 표현한다. 샘 알트만은 AGI 개발이 "미지의 영역을 탐험하는 여정"과 같다고 언급하면서, 수많은 난관과 예측 불가능한 사건들을 겪었음을 고백한 적이 있다.

이는 AGI의 미래가 유토피아적일지, 디스토피아적일지 확신할 수는 없는 현 상황을 반영한다. 즉, 특이점 이후로 찾아올 "영광스러운 미래"를 꿈꾸지만, 동시에 "통제 불가능성"이라는 꼬리표에 대한 깊은 불안감을 느낀다고 비춰진다. 결론적으로 말하면, 그는 AGI의 도래가 가까워졌음을 직감하면서도, 그 미래에 대한 불확실성, 즉 "어느 쪽"으로 나아가게 될지 모르는 상황에 대한 고민을 밝힌 것이다.

샘 알트만의 이러한 말은 적잖은 생각거리를 제공한다. **AGI 개발이 단순한 기술적 진보를 넘어, 인류의 미래에 대한 근본적인 질문을 던지고 있는 셈**이기 때문이다. 우선, 가장 먼저 드는 생각은 AGI의 임박한 도래에 대한 암시로 해석할 수 있다는 것이다. 이와 관련하여 샘 알트만은 지난 1월 6일, 본인의 블로그에다 "Reflections"라는 글을 남겼다. 이 글은 마치 어떤 위인이 큰 성공을 이룩한 뒤의 회고록으로 쓰여진 느낌이 물론 담겨 있다. 그리고 이 글에 있어서, 투자자로서 얻을 수 있는 인사이트와 앞으로의 AI 개발에 관한 힌트가 여럿 있기 때문에, 중요 내용을 발췌하여 아래와 같이 정리한다:

(2) 샘 알트만의 새로운 게시물: Reflections

ChatGPT의 두 번째 생일이 한 달여 전에 지났고, 이제 우리는 복잡한 추론이 가능한 모델이라는 새로운 패러다임으로 전환했습니다. 저는 지금까지의 여정과 그 과정에서 배운 것들에 대한 개인적인 생각을 나누고 싶습니다. AGI에 더 가까워질수록, 우리 회사의 진전을 돌아보는 것이 중요한 시기라고 느낍니다.

우리는 약 9년 전에 OpenAI를 시작했습니다. AGI가 가능하며, 이것이 인류 역사상 가장 큰 영향력을 미칠 기술이 될 수 있다고 믿었기 때문입니다. 우리는 언젠가 티핑포인트에 도달하여 AI 혁명이 시작될 것이라는 것을 항상 추상적으로 알고 있었습니다. 하지만 그 순간이 언제일지는 몰랐습니다. 놀랍게도 그것은 바로 지금 이 순간입니다. 마침내 AI에서 항상 희망했던 거대한 잠재력의 일부를 목격하고 있으며, 곧 더 많은 것이 올 것임을 알 수 있습니다.

우리는 지금까지 중 가장 뛰어난 연구 성과들을 이루어냈습니다. 주간 활성 사용자가 약 1억 명에서 3억 명 이상으로 증가했습니다. 회사가 이렇게 빠르게 성장하고 발전할 때, 자연스럽게 이해관계가 갈라지게 됩니다. 그리고 중요한 산업에서 선도적인 위치에 있는 회사는 여러 이유로, 특히 경쟁하려는 사람들로부터 많은 공격을 받게 됩니다.

시작할 때 우리는 제품 회사를 만들어야 할 것이라고는 전혀 예상하지 못했습니다. 우리는 단지 훌륭한 연구만 하면 된다고 생각했습니다. 또한 이렇게 엄청난 양의 자본이 필요할 것이라고도 예상하지 못했습니다.

우리는 이제 전통적으로 이해해온 방식의 AGI를 어떻게 구축할 수 있는지 확신하게 되었습니다. 2025년에는 최초의 AI 에이전트들이 '인력'으로 참여하여 기업의 생산성을 실질적으로 변화시키는 것을 볼 수 있을 것이라고 믿습니다. 우리는 AI 시스템을 안전하게 만드는 가장 좋은 방법은 점진적이고 단계적으로 세상에 공개하여, 사회가 기술에 적응할 수 있도록 사람들의 손에 점진적으로 쥐어주는 것이라 믿습니다.

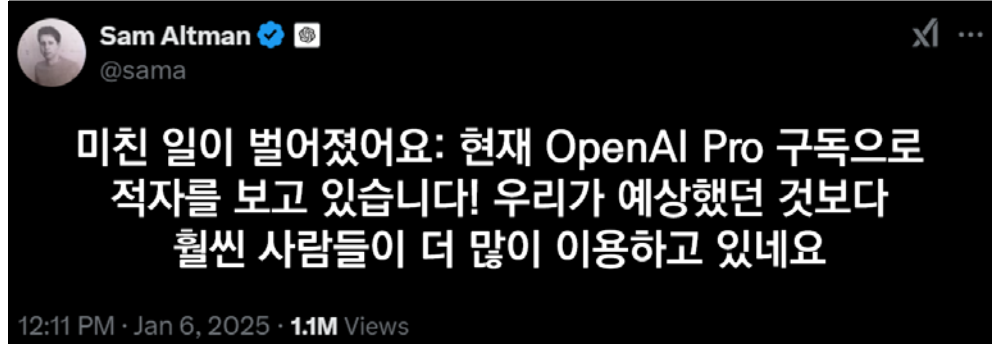
이제 AGI 이상을 목표로 하기 시작했습니다. 진정한 의미의 초지능을 향해서 말입니다. 초지능 도구들은 우리가 스스로 할 수 있는 것을 훨씬 넘어서는 수준으로 과학적 발견과 혁신을 가속화할 수 있습니다. 앞으로 몇 년 안에 모든 사람이 우리가 보는 것을 보게 될 것이라고 꽤 확신합니다. - 샘 알트만, 2025년 1월 6일

샘 알트만의 글을 보면, 그가 써내려간 문장들은 상당 부분을 OpenAI가 어쩔 수 없이 연구회사에서 상업적인 회사로 변화해야 했던 명분에 대해 많이 할애하고 있음을 알 수 있다. 이것은 일론 머스크 OpenAI 창업자와의 법적 분쟁의 불씨가 꺼지지 않고 지속/심화되고 있고 여론도 그리 본인에게 우호적인 상황만은 아니기 때문에 이를 의식한 것이라고 할 수 있다. 이와 관련해, 딥러닝의 대부라고 불리는 지프리 힌튼 교수의 최근 입장이 매우 흥미롭다. 힌튼은 일론 머스크가 OpenAI를 상대로 제기한 소송, 즉 OpenAI가 완전한 영리 기업으로 전환하는 것을 막기 위한 가처분 신청을 공식 지지했다. 힌튼은 OpenAI가 안전에 중점을 둔 비영리 단체로 설립되었다는 점을 강조하며, 영리 추구로의 전환이 "매우 나쁜 메시지"를 보낸다고 비판하고 나선 것이다.

“대부”가 목소리를 냈으니 여론전에서 꽤 불리한 입지로 내몰린 것이다. 이 경우, AI 기술 개발에 있어 놀라운 와우 포인트를 만들어 내면서 반대편의 주장이 빛 바래도록 분위기를 환기하는 것이 주요 전략일 수 있다. 초지능 도구들은 과학적 발견과 혁신을 가속화하여 풍요와 번영을 크게 증가시킬 것이고 그것을 OpenAI가 하겠다는 그의 말에서 그러한 태도를 읽을 수 있다.

표 4. 월 200달러 구독 서비스가 성황리에 팔리고 있다는 샘 알트만의 트윗

OpenAI Pro는 200달러인데도, 더 많이 사용할수록 적자를 보고 있다는 말... 이는 컴퓨팅 사용량이 어마무시하다는 얘기



자료: X(@sama), 미래에셋증권 리서치센터

AGI를 내부적으로 달성했다고 하는 커뮤니티의 소문들이 사실일 것이라는 생각이 들기도 한다. AGI를 어떻게 구축할 수 있는지 “확신”한다고 샘 알트만이 분명히 말했기 때문이다. 그리고 그가 말한 전통적으로 이해해온 방식의 AGI라고 하는 것은, 기존에 OpenAI가 AGI를 정의한 것을 상기해보면 된다. OpenAI는 자사 웹사이트에다가 AGI를 “대부분의 경제적 가치가 있는 작업에서 인간의 성능을 뛰어넘는 고도의 자율 시스템”이라고 정의한 바 있다. 그리고 이 시스템을 만들 레시피를 갖고 있으니, 이제 이것보다는 초지능을 연구하겠다는 포부로 읽을 수 있다. 이는 OpenAI를 떠났던 최고의 AI 개발자 일리아 수츠케버가 AGI를 건너뛰고 ASI에 중점을 둔 연구 기업 “SSI(Safe Super Intelligence)”를 창립했던 것과 궤를 같이 한다고 볼 수 있다.

표 5. 2023년 11월 OpenAI CEO 축출 사태 직후 OpenAI의 익명 내부자가 올린 트윗

“일리아 수츠케버는 무엇을 보았을까?” (아마 이 시점에서부터 이미 AGI를 개발했을 것이라 추측되고 있음...)



자료: X(@apples_jimmy), 미래에셋증권 리서치센터

바로 여기서, OpenAI에서 o1과 같은 추론 모델 개발에 앞장선 시니어 과학자인 Noam Brown의 말 또한 참고해볼 필요가 있다. o1-pro 모델이 출시된 직후 공개한 인터뷰에서, “2021년에 일리아 수츠케버가 나에게 말했다. 추론 컴퓨팅을 일반화하여 확장할 수 있는 능력이 어렵지만 10년 보다는 더 빨리 도달할 수 있을거야”라는 일화를 전했다.

추론 컴퓨팅을 일반적인 방식으로 확장한다는 말은, 알파고가 바둑에서는 인간을 뛰어넘었듯이 대부분의 작업에서 인간을 뛰어넘을 수 있게 일반화의 확장을 한다는 뜻이다. 곧 AGI를 뜻한다. AI 개발에 최선두에 선 개발자들은 이처럼 모두 같은 것을 이미 보았던 것 같다. 그것도 꽤 오래 전부터 스노우볼이 흘러갔다고 추측할수 있다.

그러니 OpenAI와 구글, 그리고 앤트로픽 등 여러 선도적인 AI 연구소들이 내놓는 말들은 허풍이라고 보기에는 무리가 있다고 생각된다. 이런 측면에서 볼 때, 그들이 나아가야 할 길은, 그들이 몇 년동안 주창해왔던 스케일링 법칙이라는 것을 알 수 있다.

젠슨 황 CEO가 CES 2025 키노트에서 언급했듯이 스케일링 법칙은 벽에 부딪힌 게 아니라 그 영역을 확장하고 있다고 이해하면 된다. 물론 알고리즘 혁신도 동반돼야 하겠지만, 이는 연구 방향을 완전히 바꾼다기 보다는 어떻게 하면 효율적으로 물리적 자원을 활용할 수 있을까와 같은 극도의 최적화 작업에 초점이 맞춰지는 국면이다.

표 6. 샘 알트만이 강조한 추론 가능 모델의 패러다임에서는, 컴퓨팅의 스케일링도 바뀐다.

사전 학습 중심에서 사후 학습 및 추론에 연산 요구량이 늘어난다 → 컴퓨팅 요구량 증대



자료: X(@DrJimFan), 미래에셋증권 리서치센터

2. AGI 경쟁에서 잊지 말아야 할 사항

(1) 러-중 AI 협력: 미국 주도 질서에 도전

2025년 새해가 되자마자 기다렸다는 듯, **러시아 푸틴 대통령은 AI 분야에서 중국과의 협력을 제안했다.** 푸틴 대통령은 크렘린궁 홈페이지를 통해, 러시아 정부와 스베르방크에, 중국과의 AI 협력 강화를 통한 AI 분야 공동 발전을 지시했다.

이러한 착수의 이유는 쉽게 떠올릴 수 있다. 미국 주도의 AI 패권에 도전장을 내밀기 위함이다. 이는 서방의 제재를 우회하고, 기술 굴기를 통해 새로운 국제 질서를 구축하려는 전략으로 해석된다. 특히, **러시아의 기초과학 기술과 중국의 산업화 능력을 결합하여, BRICS를 중심으로 한 새로운 AI 협력 구도를 구축하려는 움직임은 주목할 만하다.**

이와 관련해, 푸틴 대통령은 2024년 12월, 모스크바에서 열린 스베르방크 콘퍼런스에서도 'AI 동맹 네트워크' 구축을 선언한 바 있다. 이 당시, BRICS 회원국 및 다른 국가들과 AI 개발을 함께 진행하겠다고 말했다. **BRICS AI 얼라이언스 네트워크에는, 현재 브라질, 인도, 남아공 등 AI 협회들이 이미 참여한 상태다.**

AI 기술 영역에서는 다소 조용해보였던 러시아가 이제 칼을 본격적으로 빼들고 있다. 푸틴 대통령은 AI를 국가 발전의 핵심 동력으로 인식하고, 스베르방크를 중심으로 AI 개발을 적극 추진하고 있다. 특히, 미하일 미수스틴 총리와 스베르방크 회장 게르만 그레프에게 직접 AI 개발 책임을 맡기며, 강력한 추진력을 보이고 있다. 이러한 노력은 러시아판 ChatGPT인 'GigaChat' 개발과 같은 구체적인 성과로 이어지고 있다.

러시아는 2030년까지 AI 기술로 11.2조 루블(1,090억 달러) 규모의 GDP 창출을 목표로 하고 있다. 특히, 전 국민의 80%에게 AI 교육을 제공하고, 범죄 예방 등 공공 영역에 AI를 적극 도입하여, 국가 시스템 전반의 혁신을 추구하겠다고 명시했다. 푸틴 대통령은 AI를 21세기 가장 유망한 기술로 규정하고, 러시아가 모든 분야에서 AI를 선도해야 한다고 강조하며 국가적 역량을 집중하고 있는 상황이다.

표 7. 러시아의 AI 국가전략 문서: 2024년 2월 15일에 개정된 2030년까지의 정량적인 목표

지표	단위	2022년의 수치	2030년까지의 목표	증가율
러시아 내 AI 슈퍼컴퓨터 총 성능	ExaFLOPs	0.073	최소 1 이상	약 13.7배
AI 기술 활용으로 인한 GDP 증가분 (누적)	조 루블	0.2	11.2 이상	56배
AI 솔루션 개발 및 구현 서비스 시장 규모 (연간)	억 루블	120	600 이상	5배
최상위 AI 학술대회(A*) 발표 러시아 논문 수 (연간)	편수	113	450 이상	약 4배
최상위 25% 과학 저널(1사분위수) 게재 러시아 AI 논문 수 (연간)	편수	103	450 이상	약 4.4배
AI 분야 고등 교육 졸업생 수 (연간)	명	3,048	15,500 이상	약 5.1배
AI 기술 활용 능력 보유 인력 비중 (전체 노동 인구 대비)	%	5	80 이상	16배
AI 기술에 대한 국민 신뢰도	%	55	80 이상	약 1.5배
AI 기술 도입 준비도 높은 경제 분야 비중	%	12	95 이상	약 7.9배
AI 기술 도입 및 활용에 대한 기업 지출 규모 (연간)	억 루블	1,230	8,500 이상	약 6.9배

자료: 러시아 연방, 미래에셋증권 리서치센터

그런데, **올해 1월 1일, 푸틴 대통령이 러-중간 AI 협력을 공식적으로 밝힌 것은, 미국주도의 AGI 대중화 시대가 임박한 상황에서 읽혀지는 절박함으로 풀이된다. 지금 상황에서 '도광양회'와 같은 전략을 구사할 때가 아닌 것으로 판단한 것인데, 혼자 힘으로는 시간과 자원이 부족하니 다른 나라와의 연합을 꾀하는 것이다.** 위 표에서 보듯이 러시아 내 AI 슈퍼컴퓨터 총 성능 목표치가 1 ExaFLOPs에 불과한데, 이는 현재 시점에서 미국 빅테크들로부터 나오는 소식만 감안해도 그리 엄청난 규모는 아니기 때문이다.

따라서, 그에 더해 AI 동맹 네트워크를 통해 중국을 비롯한 BRICS 국가들과의 연대를 강화하고, 미국의 AI 주도권에 도전하며, 특히 중국의 화웨이와 Cambricon을 끌어안아 범

BRICS 진영의 AI 생태계를 구축하려는 전략으로 보인다. BRICS 매체에 따르면, 벌써 러시아의 구글이라고 불리는 안덱스와 중국의 화웨이, Cambricon, 바이두 연합이 미국의 빅테크들을 캐치업할 가능성을 점쳐보는 사람도 생겼다고 한다.

어쨌든 중요한 점은, **AI 기술 영역 또한 “G7 vs BRICS”의 구도로 진영 논리로 경쟁할 거대한 스타디움이 됐다**는 것이다. 예를 들어, G7에 속한 국가들의 국민들은 AI 모델을 사용할 때 그 기반모델이 중국 혹은 러시아 업체가 만든 모델이라고 한다면 사용하는 것을 거부할 것으로 생각하는 게 자연스럽다. 물론 그 반대도 마찬가지다. 이것은 젠슨 황 CEO가 언급한 “소버린 AI”의 필요성에 각국이 점점 더 진심을 보일 것이라는 말과 같다. 젠슨 황이 중국을 방문하지 않는 이유는 이러한 진영상의 AI 경쟁을 의식하고 있기 때문으로 사료된다.

(2) 부상하는 중국 AI 기업들

중국은 2017년부터 AI를 국가 전략 산업으로 육성해 온 나라다. 이를 지속해 2030년까지 글로벌 AI 리더로 도약한다는 목표 또한 세웠다. 특히, 2014년부터 2023년까지 3.8만 건 이상의 생성 AI 특허를 출원하며, 전 세계 AI 특허 출원의 약 70%를 차지하는 등, 기술 혁신에 박차를 가하고 있다. 또한, AI 관련 학술 문헌 기여도 역시 크게 증가하여, 2020년에는 미국을 추월하는 성과를 보였던 국가로서, 무시할 만한 국가는 아니다.

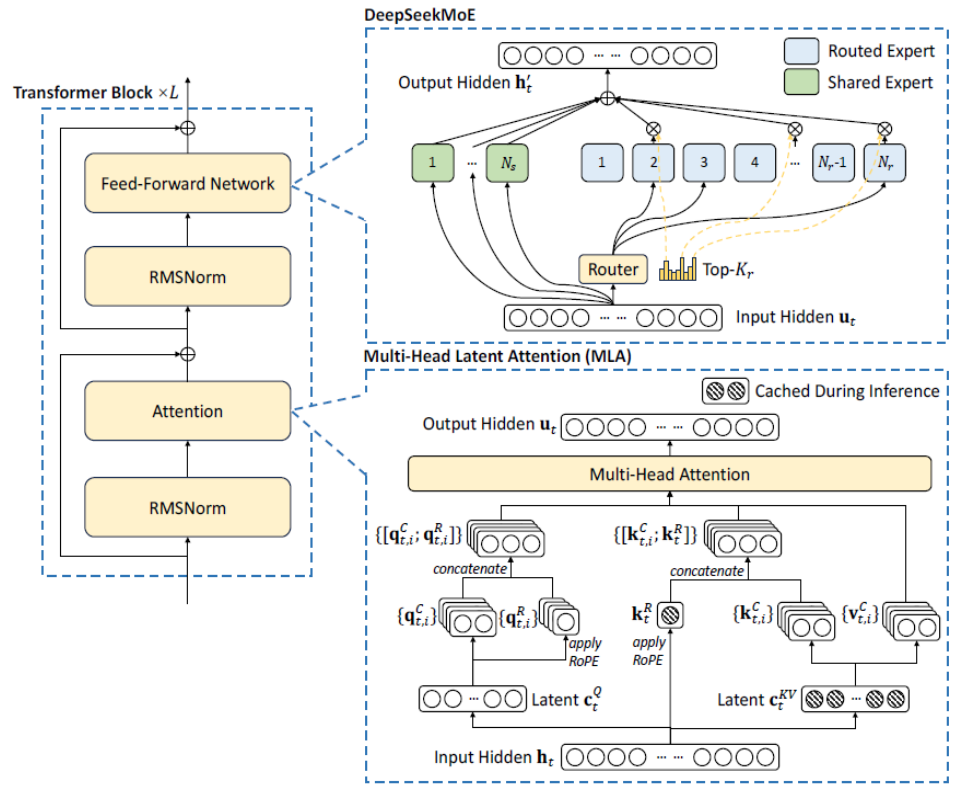
최근의 **미국이 심화 적용하고 있는 기술 분야의 대중 제재는 오히려 중국 AI 기업들의 성장을 촉진하는 역설적인 결과를 낳고 있다는 분석이 많다**. 특히, DeepSeek와 Cambricon 등은 중국의 AI 굴기를 이끄는 대표적인 기업으로 주목받고 있다.

최근 AI 업계의 이목을 집중시킨 모델이 있었으니, 바로 **중국 기업 DeepSeek가 공개한 모델인 "DeepSeek-V3"**다. 이 모델은 2,048개의 엔비디아 H800 GPU로 구성된 클러스터에서 훈련되었고, **메타가 Llama 3를 훈련했던 것과 비교해보면, 고작 9% 남짓한 컴퓨팅 자원만 사용했다는 점에서 경탄을 자아낸 모델이다**.

DeepSeek-V3의 등장은, 미국의 대중 AI 반도체 수출 통제 속에서, 중국이 어떻게 GPU 부족 문제를 극복하고 AI 기술 발전을 이뤄내고 있는지 보여주는 단적인 예시라 할 수 있다. DeepSeek-V3는 제한된 GPU 자원으로도 대규모 모델 학습 및 추론을 가능하게 하는 다양한 기술 혁신을 선보였다.

표 8. DeepSeek-V3의 기본 아키텍처를 설명하는 다이어그램

DeepSeek-V3는 총 6,710억 개의 매개변수를 가진 대규모 MoE 언어 모델로, 토큰당 370억 개의 매개변수가 활성화
전문가는 256개의 전문가로 구성되고, 각 토큰이 처리될 때마다 상위 8개의 전문가와 1개의 공유 전문가를 선택하는 구조



자료: DeepSeek, 미래에셋증권 리서치센터

주: Shared Expert는 모든 토큰에 대해 공통적으로 적용되는 전문가. 특정 토큰에 관계없이 항상 활성화되어 모든 입력 토큰을 처리.
Routed Expert: 토큰별로 특화된 작업을 수행. 어떤 전문가가 코딩에 특화, 다른 전문가는 수학에 특화되는 형태.

- MoE(Mixture of Experts) 아키텍처: 복잡한 작업을 여러 "전문가"에게 분산하여 처리하는 효율적인 구조. DeepSeek-V3는 256개의 라우팅된 전문가와 1개의 공유 전문가를 활용, 각 토큰에 가장 적합한 전문가를 선택적으로 활성화. 기존 MoE 모델의 문제점이었던 "전문가 불균형" 문제도 해결.
- MLA(Multi-Head Latent Attention): 키-값 캐시 크기를 "잠재 벡터(Latent Vector)로의 압축해 추론 효율을 높이는 혁신적인 어텐션 메커니즘.
- MTP(Multi-Token Prediction): 한 번에 여러 토큰을 예측하도록 훈련하여 학습 효율성을 높임. 이는 추측적 디코딩의 기반을 제공해 추론 속도 향상을 만듦.
- FP8 혼합 정밀도 훈련: "세분화된 양자화", "고정밀 누적" 과 같은 최적화 기법을 통해, FP8 훈련의 정확도 손실을 최소화하고 메모리 사용량과 계산 부하 줄임.
- DualPipe: 연산과 통신을 중첩. 파이프라인 버블을 줄이고, GPU 활용률 극대화.
- 효율적인 All-to-All 통신: "크로스-노드 All-to-All 통신"을 최적화하여, 여러 노드에 있는 GPU들이 서로 효율적으로 정보를 주고받을 수 있도록 함.

이 모델은 무려 6,710억 개의 매개변수를 가진 거대 언어 모델(LLM)이지만, 토큰당 370억 개의 매개변수만 활성화하는 혁신적인 MoE(Mixture of Experts) 아키텍처를 채택해, 성능과 효율성 두 마리 토끼를 모두 잡았다는 평가를 받고 있다. 우리 팀도 DeepSeek-V3를 자주 사용하고 있는데 그 가성비와 토큰 생성 속도에 굉장히 놀랐다. 심지어 무료다.

DeepSeek-V3는 미국의 대중 제재라는 어려움 속에서도, 기술 혁신과 효율성 극대화를 통해 AI 강국으로 부상하려는 중국의 의지를 보여주는 사례라 할 수 있다. DeepSeek-V3의 등장은 벌써 중국 내에서 큰 반향을 만들어 큰 변화를 만들어 내고 있다.

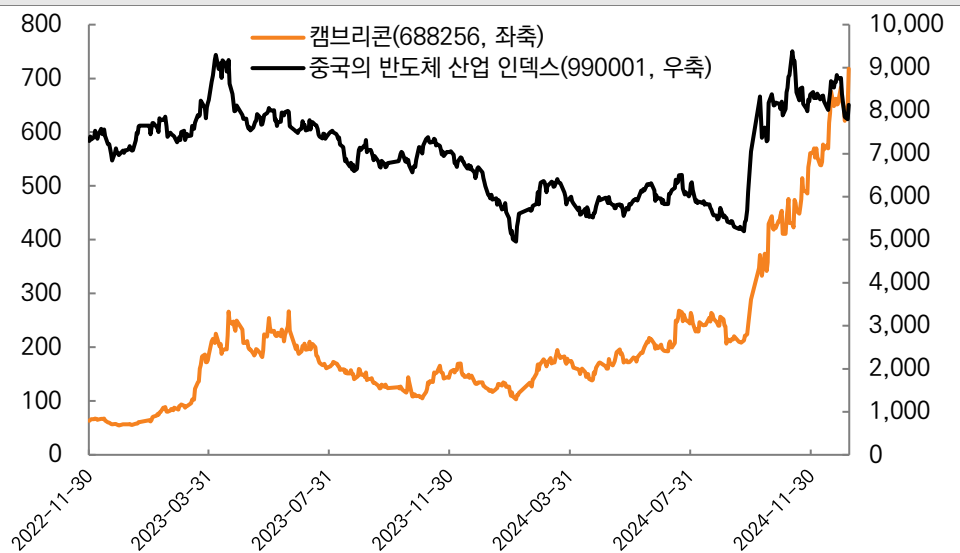
DeepSeek라는 신생 기업이 공격적이고 매력적인 가격 정책으로 LLM 시장을 공략하니 이번에는 알리바바가 본인들의 Qwen 모델에 대한 토큰 비용을 85%까지 인하해버렸다. DeepSeek-V3의 경우, 입력 비용은 약 0.002위안이기 때문에 이에 근접하게 Qwen-VL-Max 모델의 입력 비용은 1,000 토큰당 0.003위안으로 낮춰 잡았다. GPT-4o와 비교해서는 16% 수준의 가격으로 책정된 것이다.

중요한 점은, **중국 대륙에서 치열한 공방전 끝에 살아남아 글로벌 시장을 공략하는 산업적 인 흐름을 반드시 유념해야 한다는 점이다. 우리는 사실 이미 많이 목격했다. 스마트폰이 그랬고, 전기차가 그랬고, 이제 곧 AI 모델도 마찬가지인 셈이다.** 심지어 2024년 중국 내 250개 이상의 생성형 AI 모델이 승인된 상황이다. 그리고 올해는 그 중에서 살아남은 중국의 모델들이 BRICS의 각 영역에 또아리를 틀 것으로 판단된다.

다만, 알고리즘 혁신도 기하급수적으로 계속 스케일링 할 수는 없는 노릇이니 여전히 AI 가속기 칩(GPU 포함)에 대한 중국의 갈증은 계속될 수밖에 없다. 그런 차원에서, 주목받는 업체는 위에서 보듯 러시아에서 언급한 Cambricon이라는 기업이다. 이 기업은 중국과학원 출신 연구진이 설립한 AI 칩 설계(ASIC) 전문 기업이다. NPU 설계에 특화된 기술력을 보유하고 있다.

표 9. 캄브리콘과 중국의 반도체 산업 지수의 주가 비교

2022년 11월 30일인 ChatGPT 런칭 시기 대비 캄브리콘의 주가는 현재까지 1000% 넘게 상승



자료: TradingView, 미래에셋증권 리서치센터

미국의 제재 대상에 포함된 국영기업이지만, 이는 오히려 중국 내에서 Cambricon의 중요성을 부각시키는 계기가 되었다. Cambricon은 화웨이와의 협력을 통해 미국의 제재를 우회하고, 중국 내 AI 칩 수요 증가에 힘입어 가파른 성장세를 보이고 있다.

특히, 2024년 주가 상승률 기준 중국 CSI 300 지수 최고 성과를 기록한 것은, Cambricon의 잠재력을 보여주는 단적인 예시라 할 수 있다. 이에 대해서 저명한 AI 개발자들은 “결핍이 혁신과 창의를 낳는다”고 이 업체에 대한 놀라움과 공포심을 나타내기도 했다.

표 10. 중국의 AI 가속기 설계 업체인 Cambricon의 기업 개요

구분	상세 내용
기업명	Cambricon Technologies (寒武纪科技)
설립연도	2016년
설립주체	중국과학원 연구진(Chen Tianshi, Chen Yunji 형제)
본사 위치	베이징 하이디안구 (R&D 센터는 상하이에서 운영)
직원 수	약 1,000명
상장 정보	2020년 상해증권거래소 과학판 상장 (종목 코드: 688256)
주요 제재	2022년 말 중국군 AI 칩 기술 공급 우려로 미국 제재
핵심 사업	AI 칩 설계 및 판매, AI 플랫폼 솔루션 제공
생산 방식	TSMC 등 파운드리 위탁 생산
2024년 상반기 실적	<ul style="list-style-type: none"> • 매출: 6,477만 위안(약 911만 달러) • 순손실: 5.3억 위안
2023년 실적	<ul style="list-style-type: none"> • 매출: 7억 2,590만 위안 (전년대비 -6.8%) • R&D 투자: 11.18억 위안 (매출대비 157.53%)
주요 제품군	<ul style="list-style-type: none"> • Cloud AI 칩: MLU 200 시리즈 (MLU270, MLU290) • Edge AI 칩: MLU 100 시리즈 • Cambricon NeuWare: AI 배포를 위한 소프트웨어 개발 플랫폼 • Cambricon NeuCloud: 클라우드 AI 플랫폼
주요 고객	<ul style="list-style-type: none"> • 화웨이 등 통신 관련 업체 • 중국 주요 서버 제조사 • 인터넷/금융/교통/에너지/제조업 기업
특징	<ul style="list-style-type: none"> • NPU 설계 특화 • 중국 반도체 자립화 정책의 핵심 기업 • 엔비디아 대체제 역할 수행

자료: 미래에셋증권 리서치센터

II. AI Issue

1. CES 2025를 전후로 나온 여러 소식

(1) 데이터센터 투자 경쟁

AI 기술 발전에 따라, 관련 데이터센터 용량에 대한 수요가 폭발적으로 증가하고 있다는 것은 공공연한 사실이다. 최근 발간된 맥킨지 보고서에 따르면, **AI 데이터센터 용량 수요는 2030년까지 연평균 33% 성장률을 보이며, 전체 데이터센터 수요의 약 70%를 차지할 것으로 전망된다.** 이러한 흐름 속에서, 주요 기업들은 AI 인프라 투자를 공격적으로 확대하고 있다.

아마존 AWS가 조지아 주에 AI 데이터센터 인프라 구축을 위해 110억 달러를 투자할 예정이라는 소식이 전해졌다. 이는 향후 몇 년간 AI 데이터센터에 1,000억 달러 이상을 투자하려는 장기 계획의 일환이다.

그리고 **마이크로소프트는 회계연도 2025년에 AI 데이터센터 구축에 무려 800억 달러를 투입**할 계획이다. 한화로 따지면 116조원이 넘어가는 상상하기도 힘든 돈이다. 몇 달 전, 오라클의 CTO이자 창업자인 래리 엘리슨이 남긴 말이 안떠오를 수가 없다. “빅테크들의 AI 경쟁전에 참가하기 위한 티켓 값은 1,000억 달러다”. 그의 말은 점점 사실이 되어 간다.

마이크로소프트의 800억 달러 투입금액은, 2024년 대비 약 60% 증가한 수치다. 이 중 절반 이상이 미국 내 투자가 될 것으로 예상된다. 이러한 대규모 투자는 급증하는 AI 수요에 선제적으로 대응하고, 당연히 시장 주도권을 확보하려는 전략이다.

또한, **아랍에미리트(UAE) 국부펀드 DAMAC Properties의 미국 데이터센터 투자 발표도 주목할 만하다.** 후세인 사자와니 회장은 트럼프 대통령 당선인과의 회담에서 200억 달러 규모의 투자 계획을 밝혔다. 텍사스, 애리조나, 일리노이, 미시간 주와 같이 서부~남부~중부에 미 대륙 길게 걸쳐 투자가 이루어질 예정이다. 이는 2024년 12월 16일, 소프트뱅크의 손마사요시 회장이 도널드 트럼프 대통령 당선인의 플로리다 마라라고 저택에서 향후 4년간 미국에 1,000억 달러 투자 계획을 발표한 것의 연장선상이라 할 수 있다.

이처럼 외국 자본의 자본의 미국 내 데이터센터 투자는 AI 인프라 구축 경쟁에 또 다른 연료로 작용하고 있다. **상당한 금액의 투자가 미국에서 이뤄져 수십만개의 일자리를 창출하고 미국산 AI 기업들을 지원한다는 점을 감안하면, 점점 더 AI 경쟁은 G2간 진영별 경쟁으로 모아질 가능성을 가리키고 있다고 판단한다.**

(2) 엔비디아와 TSMC, 그리고 메모리 반도체...

엔비디아와 퀄컴이 TSMC의 2nm 공정 대신 삼성전자의 2nm 공정을 사용할 수 있다는 루머가 돌았다. TSMC의 높은 생산 비용과 제한된 생산 능력이 주요 원인이라는 것이다.

삼성전자는 과거 4nm 칩 생산에서 발생했던 발열 및 성능 문제로 잃었던 고객 신뢰를 회복하기 위해, 올 1분기부터 2nm 공정(SF2) 시험 생산을 시작하며 추격에 나선 상태다.

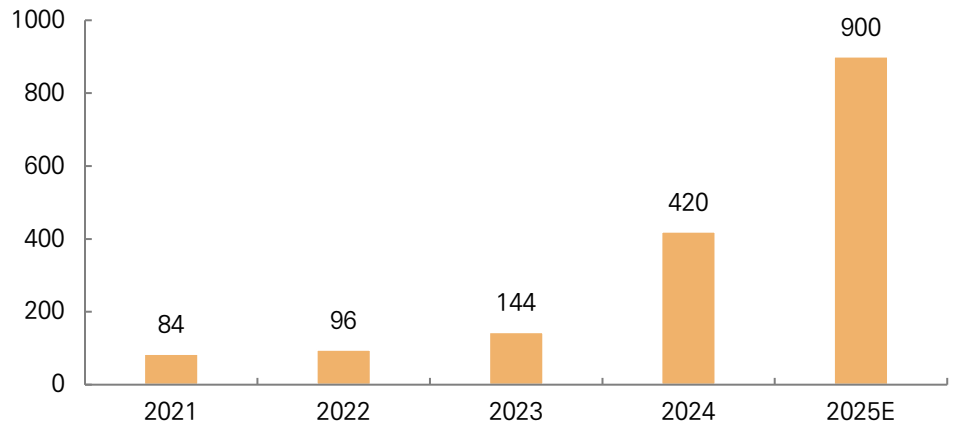
TSMC가 2nm 생산 제약 때문에, 웨이퍼 가격을 50% 인상한 30,000 달러로 책정했다는 해당 현지 소문은, 3nm 및 4nm 공정 가격 역시 5~10% 인상될 것(엔비디아와 AMD는 3~4nm 공정 가격에서 최대 10%의 가격 인상)이라는 기존의 전망과 맞물려 칩 제조사들에게 큰 부담으로 작용할 것이라는 점이 루머의 근거라고 할 수 있다.

그러나 **TSMC의 2nm 공정 지연 엔비디아와 퀄컴의 삼성 파운드리 채택 가능성에 대한 루머는 사실과 다르다**는 분석이 Dylan Patel(SemiAnalysis 편집장이자 중화권 전문 반도체 소식을 잘 다룸)로부터 나왔다. TSMC는 2nm 공정 개발을 계획대로 진행하고 있으며, 2025년 하반기 양산을 목표로 하고 있다는 것이다. 원래 N2(2nm) 웨이퍼는 제조에 약 14주, 패키징 및 조립에 추가로 수개월이 소요되므로, N2 기반 제품의 출시 시점은 빨라도 2026년 2분기가 맞다는 말이다. 즉, 현재 TSMC는 2nm 공정 수율을 60%까지 끌어올렸으며, 2025년 초 시험 생산, 2025년 하반기 양산을 목표로 하고 있는 상황으로 보인다.

한편, TSMC는 2026년까지 2nm 생산 능력을 월 10,000장에서 80,000장까지 확장하고, CoWoS 패키징 가격도 10~20% 인상할 계획이다. 주요 고객사인 엔비디아는 이미 이러한 가격 인상에 동의한 것으로 알려졌으며, 2025년 CoWoS 생산 능력은 월 75,000개 웨이퍼로 확대될 예정이다.

표 11. TSMC의 연간 CoWoS 생산능력 전망 (단위: 천 웨이퍼)

작년 4분기만 하더라도 2025년의 CoWoS 생산능력이 YoY 88% 상승할 것이라 예상했었으나 최근 상향 조정



자료: 미래에셋증권 리서치센터

애플은 계획대로 2026년에 N2 기반 첫 제품을 출시할 계획이며, 고성능 AP에서 N-1 노드와 SoIC 기술을 활용하여 비용 절감을 추진하고 있다. 또한, 성능 향상을 위해 광대역 IO 코퍼키징 메모리(CPO)에 집중하고 있다. 또한, TSMC는 2025년 중반 엔비디아 GB300 플랫폼(블랙웰 울트라 아키텍처)에 CPO 샘플을 최초로 공급할 예정이다.

즉, TSMC의 여전히 괜찮고, **엔비디아와 퀄컴은 여전히 TSMC를 주요 파운드리 파트너로 유지하고 있다.** Dylan Patel에 따르면, “엔비디아와 퀄컴은 삼성을 신뢰하지 않는다. 삼성은 현재 PPA(성능, 전력, 면적) 측면에서 TSMC보다 약 3년, 인텔보다도 약 1년 뒤쳐져 있다. 엔비디아와 퀄컴 모두 가장 많은 물량의 제품을 TSMC에서 옮길 계획은 없다”고 삼성에게는 너무하다 싶은 가혹한 혹평을 했다. 참고로 엔비디아의 젠슨 황 CEO는 삼성전자의 HBM 관련해서도 말을 남겼는데, “새로운 설계를 다시 해야 한다”고 CES에서 언급했었다.

루머와 Patel의 의견을 종합하면, 엔비디아와 퀄컴이 삼성 파운드리의 새로운 노드를 테스트한 것은 맞지만, 이는 이례적인 일은 아니고, 오히려 TSMC와의 협상력을 높이기 위한 전략으로 해석할 수 있다.

엔비디아와 TSMC만큼의 사이보다는 느슨하지만, 동맹의 중요한 한 축인 SK하이닉스의 소식도 CES 2025를 전후로 등장했다. **SK하이닉스는 CES 2025에서 HBM3E 16단 시제품을 공개했다.** 이 제품은 16개의 DDR5 메모리 칩을 수직 적층하여 48GB 용량을 구현한다. 이는 업계 최대의 용량으로 현재 엔비디아에 양산 공급 중인 HBM3E 12단을 잇는 차세대 메모리 솔루션이다. 이 칩은 학습 성능과 추론 성능 면에서 기존의 12단 제품 대비 각각 18%와 32% 향상된 능력을 보이는 것으로 보고되었다.

SK하이닉스는 하이브리드 본딩 기술 도입 전까지, HBM3E의 20단 DDR5 칩 적층 기술까지 개발할 계획이다. 그리고 2025년 하반기에는 차세대 HBM4 양산을 목표로 하고 있다.

표 12. CES 2025에서 공개된 HBM3E 16단 제품의 확대 제작 모형

어드밴스드 MR-MUF 공정을 적용하여 업계 최고층인 16단 구현해 칩의 휨 현상을 제어하고 방열 성능을 극대화



자료: 언론 보도, 미래에셋증권 리서치센터

2. CES 2025: AMD

(1) AMD의 야심 찬 신제품 러시

매년 초, 미국 라스베이거스는 전 세계 IT 업계의 시선이 집중되는 뜨거운 열기의 도시로 변모한다. 바로 세계 최대 IT 전시회, CES가 개최되기 때문이다. 컴퓨텍스, MWC와 함께 세계 3대 IT 박람회로 불리는 **CES에서 AMD는 꽤 공격적이었다. 게이머를 위한 X3D 칩부터, 워크스테이션 사용자를 겨냥한 Ryzen AI Max, 그리고 가성비를 앞세운 Ryzen 200 시리즈까지, 방대한 제품군을 쏟아냈다.** 신제품 '폭격'이라고 해도 과언이 아니다.

표 13. CES 2025에서 AMD가 선보인 신제품 라인업

게이머용 X3D 칩, 워크스테이션 중심의 Ryzen AI Max 제품군, 새로운 저가형 Ryzen 200 시리즈 등 대거 출시



자료: AMD, 미래에셋증권 리서치센터

게이머를 위한 궁극의 무기: Ryzen 9 9950X3D & 9900X3D

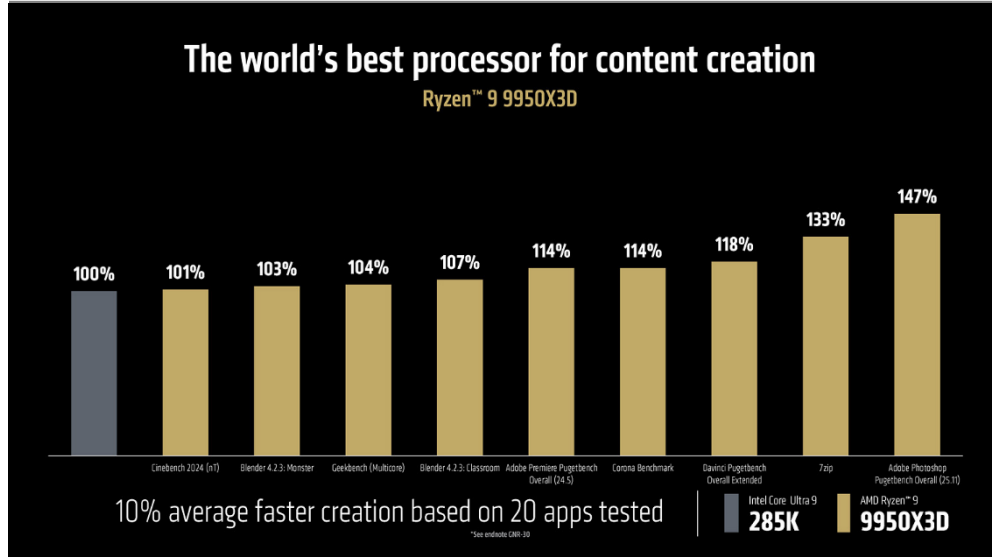
가장 먼저 이목을 끈 것은 단연 Ryzen 9 9950X3D와 9900X3D다. AMD의 자랑인 2세대 3D V-캐시 기술을 적용해 게이밍 성능을 극대화한 제품이다. 3D V-캐시는 CPU 다이 위에 캐시 메모리를 수직으로 쌓아 올려, 데이터 접근 속도를 비약적으로 향상시킨 기술이다. 쉽게 말해, 게임 로딩 시간을 줄이고, 프레임 속도를 높이는 데 탁월한 효과를 발휘한다는 것이다. Ryzen 9 9950X3D는 16코어 32스레드, 5.7GHz 부스트 클럭, 144MB 캐시, 170W TDP라는 괴물 같은 스펙을 자랑한다. Ryzen 9 9900X3D는 12코어 24스레드, 5.5GHz 부스트 클럭, 140MB 캐시, 120W TDP를 갖췄다. 게이머들에게는 꿈의 CPU라 할 수 있겠다. 3D V-캐시는 이전 세대와 마찬가지로 CCD 하나에만 적용되었다. 일각에서는 두 개의 CCD에 모두 3D V-캐시를 적용하여, 더욱 극적인 성능 향상을 기대하기도 했다. 하지만, AMD는 3D V-캐시가 적용되지 않은 CCD에는 클럭이 높은 다이를 탑재하여, 게임뿐만 아니라 콘텐츠 제작 성능까지 잡았다고 설명했다.

AMD에 따르면, Ryzen 9 9950X3D는 이전 세대인 Ryzen 9 7950X3D 대비 평균 8% 더 높은 게이밍 성능을 제공한다. 또한, 인텔의 코어 울트라 9 285K와 비교해서는 40종 게임 평균 20% 더 높은 성능을 보인다고 밝혔다. 콘텐츠 제작 측면에서도 20종의 소프트웨어 테스트에서 이전 세대 대비 평균 13%의 성능 향상을 달성했다고 하니, 게임뿐만 아니라

영상 편집, 3D 렌더링 등 고사양 작업에서도 발군의 성능을 기대할 수 있다는 의미다.

표 14. 라이젠은 게임뿐만 아니라 콘텐츠 제작 쪽에서도 “세계 최고의 CPU”라고 홍보

포토샵은 원래 인텔 CPU에서 더 성능이 좋다고 정평나있었는데, 9950X3D가 가볍게 인텔을 압도



자료: AMD, 미래에셋증권 리서치센터

노트북에 부는 3D V-캐시 열풍: Ryzen 9000 HX & HX3D "Fire Range"

노트북 게이머들을 위한 희소식도 있었다. 노트북 최초로 2세대 3D V-캐시를 탑재한 Fire Range HX3D 모델도 공개했기 때문이다. 고성능 게이밍 노트북을 위한 HX 및 HX3D "Fire Range" 프로세서는 게이머들의 가슴을 설레게 하기에 충분했다. "HX" 접미사는 고성능 티어를 나타내며 플래그십 칩을 의미한다. 최상위 모델인 Ryzen 9 9955HX3D는 16코어 32스레드, 총 144MB 캐시, 54W TDP를 갖춰 입이 떡 벌어지는 전성비를 자랑한다.

AI 워크스테이션 시장을 겨냥하다: Ryzen AI 300 "Max"

이번 발표에서 가장 눈에 띄었던 것은 바로 Ryzen AI 300 "Max" 시리즈였다. 이 제품군은 애플 M4 프로세서를 정조준한 고성능 워크스테이션용 프로세서로, 특히 AI 연산 성능에 초점을 맞췄다. AMD는 강력한 성능의 애플 M4 Pro와의 비교는 제시했지만, M4 Max와의 비교는 의도적으로 피한 듯한 인상을 줘 아쉬움을 남겼다. Ryzen AI Max+ 395는 "Zen 5" 아키텍처를 기반으로 하는 단일 칩으로서, 16개의 성능 코어, 32개의 스레드를 가진다. 그리고 "RDNA 3.5" 아키텍처 기반 40 CU(컴퓨팅 유닛), 최대 50TOPS가 가능한 NPU 및 초당 256GB 대역폭 메모리 인터페이스를 갖추고 있다. 특히 40 CU를 탑재했다는 것은, 외장 그래픽카드에 준하는 성능을 보이는 것이고 같은 AI PC 라인업에서 비교할 만한 제품이 없는 수준이라는 평가도 있다. 그럼에도 TDP가 120W에 불과하다는 점이 놀랍다.

AMD는 다양한 3DMark 제품군 테스트에서 인텔의 "Lunar Lake" Ultra 9 288V보다 1.4배 빠른 그래픽 성능과 Cinebench, Blender 와 같은 벤치마크에서 동일한 인텔 칩보다 2.6 배 빠른 렌더링을 보인다고 주장했다. 특히, 700억 개의 파라미터를 가진 대규모 언어 모델을 실행했을 때 엔비디아의 RTX 4090보다 최대 2.2배 더 빠른 속도를 보였다고 강조했다.

표 15. 워크스테이션 및 고성능 노트북에서도 탑재될 Ryzen AI Max+ 395의 주요 스펙 사항
 "RDNA 3.5" 아키텍처 기반 40 CU, 최대 50TOPS가 가능한 NPU 및 초당 256GB 메모리 대역폭을 갖춘 전천후 프로세서

Up to **16 "Zen 5" Performance Cores**

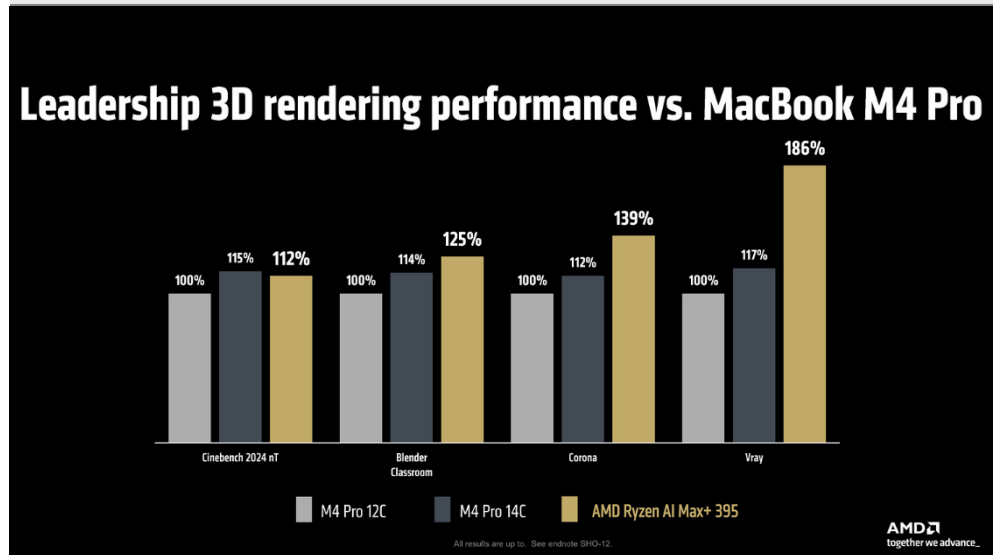
Up to **40 RDNA™ 3.5 Compute Units**

Up to **50 TOPs XDNA 2 NPU**

Up to **256GB/s Bandwidth New Memory Interface**

*See endnote SHO-243

강력한 Apple M4 Pro와 성능 비교표를 제시했으나, M4 Max와의 비교는 피함...



엔비디아 RTX 4090(24GB VRAM)보다 초당 토크생성 기준 최대 2.2배 빠른 70B 크기의 LLM을 실행할 수 있다고 주장

The world's first Copilot+ PC processor to run 70B LLM

Up to **2.2x**
faster AI Performance*

AMD Ryzen™ AI Max+ 395 vs. NVIDIA GeForce RTX 4090 24GB

Up to **87%**
lower TDP

AMD Ryzen™ AI Max+ 395 vs. NVIDIA GeForce RTX 4090 24GB

*tokens / second

All results are up to: Total TDP of Ryzen AI Max+ 395 is 50W and TDP of GeForce GTX 4090 is 450W. See endnote: SHO-14

자료: AMD, 미래에셋증권 리서치센터

이는 꽤나 놀라운 수치다. 물론, RTX 4090은 현존 최강의 그래픽 카드 중 하나이기에, AMD의 주장을 곧이곧대로 믿기는 어렵다. 하지만, AMD의 Ryzen AI Max+ 기반 하이엔드 프로세서는 온다이 통합 메모리 아키텍처를 통해, 그래픽 처리 부문인 CU에서 최대 96GB의 메모리를 사용할 수 있기 때문에 더 많은 메모리 사용으로 이런 주장을 한 것은 어느 정도 이해가 가는 대목이다.

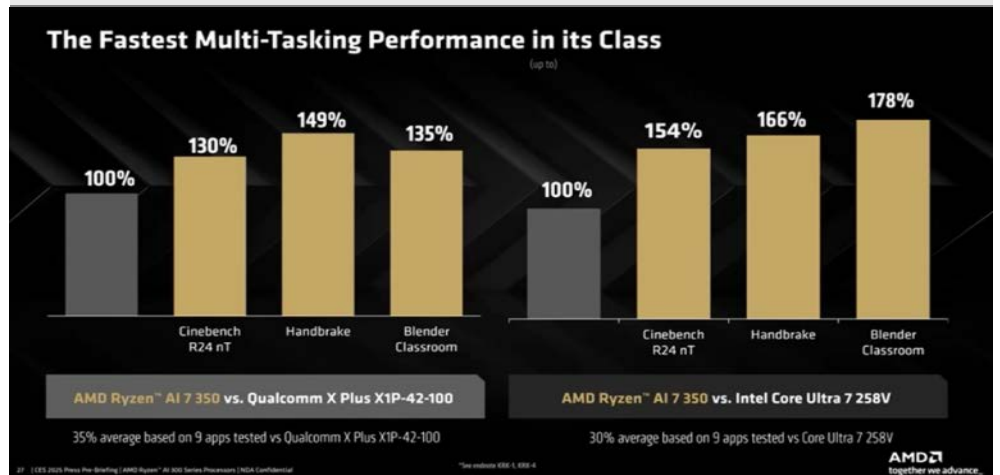
AI PC 시대를 위한 야심찬 포석: Ryzen AI 300 & AI 200

일반 사용자를 위한 Ryzen AI 300 시리즈도 공개했다. 3세대 Ryzen AI를 표방하는 300 시리즈는 Ryzen AI 7과 Ryzen AI 5, 두 가지 모델로 출시된다. 특히, **퀄컴의 스냅드래곤과의 비교에서도 우위를 점했다고 강조했다.** 스냅드래곤은 전통적으로 ARM 아키텍처를 사용해왔기에, x86 진영을 위협하기는 어려울 것이라는 게 중론이었다. 하지만, 최근 윈도우 지원이 강화되면서 꽤나 위협적인 성능을 보여주고 있기에, AMD의 주장은 꽤나 의미심장하게 다가온다.

보급형 노트북 시장을 겨냥한 Ryzen 200 시리즈도 빼놓을 수 없다. Ryzen 200 시리즈는 메인스트림 노트북을 위한 프로세서로, 합리적인 가격대로 출시될 것으로 예상된다. 최대 8코어 16스레드와 16 TOPS의 NPU 성능을 자랑하며, 2025년 2분기 출시 예정이다.

표 16. Ryzen AI 300 시리즈 CPU와 퀄컴(좌) 및 인텔(우) 칩과 벤치마크 성능 비교

일상 및 비즈니스 노트북에 탑재된 Ryzen AI 300 CPU는 올해 150개 이상의 노트북에 탑재될 것이라고 AMD는 주장



자료: AMD, 미래에셋증권 리서치센터

(2) RDNA 4와 FSR 4은 어디갔나... 실망스러운 행보

신제품 출시 용단 폭격에도 불구하고, AMD의 CES 2025 기조연설에 대한 시장의 반응은 그다지 긍정적이지 않았다. 기술 커뮤니티의 반응은 대체로 부정적이거나 미흡하다는 의견이 지배적이었다. 실망감의 가장 큰 이유는 바로 사람들이 기다려온 GPU 관련 발표가 턱없이 부족했기 때문이다. 많은 이들이 RDNA 4 아키텍처에 대한 구체적인 정보 공개를 기대했지만, AMD는 이에 대한 언급을 극도로 아꼈다. 그냥 조용히 보도자료를 발표하면서, RDNA 4 아키텍처 기반의 RX 9070 XT와 RX 9070을 공개한 게 전부였다. RTX 4070 Ti와 4070 Super에 대항하는 중급 GPU로 포지셔닝으로 된 제품들이다.

또한, AMD는 그들의 업스케일링 기술인 FSR에 대해서도 거의 언급하지 않고 넘어갔고 RDNA 4처럼 보도 자료를 통해 확인해야 했다. 새로운 “FSR 4”은 머신러닝 기반 업스케일링 기술을 도입했다고 하는데, 성능 시연이 없었기에 크게 와닿지 않는다. 소프트웨어 역량에 있어 엔비디아보다 심각한 열위를 가지고 있는 기업이라는 평가를 받고 있는 와중이기 때문에 엔비디아가 무엇을 내놓을 것인지에 눈치를 보고 있다는 느낌을 지우기가 힘들었다.

표 17. 키노트에서는 거의 언급하지 않고 보도자료를 통해 조용히 나온 RDNA 4와 FSR 4 소식
RDNA 4는 AMD의 차세대 소비자용 GPU 아키텍처... 엔비디아의 하이엔드GPU 경쟁은 포기한 소극적 움직임은 아쉬움

INTRODUCING
AMD RDNA™ 4 ARCHITECTURE

Optimized Compute Units | **Supercharged** AI Compute | **Improved** Raytracing Per CU | **Better** Media Encoding Quality

4nm Process | 2nd Gen AI Accelerators | 3rd Gen Raytracing Accelerators | 2nd Gen AMD Radiance Display™ Engine

소비자용 GPU인 라데온 RX 9070 XT와 9070이 1분기 내 출시 예정이고 브랜드 네이밍 방식도 엔비디아에 맞게 바꿈

RADEON BRANDING FOR RDNA™ 4

RTX 4080/S, RTX 4070 Ti, RTX 4070S, RTX 4060 Ti, RTX 4060 (RTX 5000 Series)

RX 7900 XTX, RX 7900 XT, RX 7900 GRE, RX 7800 XT, RX 7700 XT, RX 7600 XT, RX 7600 (RX 9070 SERIES, RX 9060 SERIES)

AMD Radeon™ RX 9070 XT, AMD Radeon™ RX 9070

Gaming sync with Ryzen™ 9000 series Use 8000 series for (RDNA 3.5) mobile

Simplify model system to match direct competitor compare

AMD의 최고급 GPU인 RX 9070 시리즈는, 엔비디아 보급형 GPU인 RTX 4070과 경쟁 하는 수준

엔비디아 업스케일링 기술인 DLSS 4와 정면으로 경쟁할 AMD의 FSR 4 기술도 출시... 그런데 기대하는 여론은 크지 않음

CALL OF DUTY BLACK OPS 6

AMD FIDELITYFX™ SUPER RESOLUTION 4 ML-POWERED UPSCALING

- DEVELOPED FOR AMD RDNA™ 4
- HIGH QUALITY 4K UPSCALING
- HIGH PERFORMANCE WITH FSR + FG
- LOW LATENCY WITH AMD ANTI-LAG 2

*AMD FSR 4 UPGRADE FEATURE ONLY AVAILABLE ON AMD RADEON RX 9070 SERIES GRAPHICS FOR SUPPORTED GAMES WITH AMD FSR 3.1 ALREADY INTEGRATED

AMD together we advance gaming

자료: AMD, 미래에셋증권 리서치센터

AMD는 45분간의 키노트 발표 중 대부분의 시간을 AI CPU 및 APU 관련 내용에 할애했고, AI라는 단어를 153번이나 언급했다. 그럼에도 GPU 소식 부재 때문에 이러한 노력들이 가려졌다고 평가받는 건 안타까운 대목이다.

AMD의 AI 전략을 한 번 생각해보면, 애초에 RDNA 4와 게임은 주요 관심사가 아니었고, RISC 진영(애플, 퀄컴, 미디어텍 등)의 AI PC 부문의 추격을 물리치기 위해 중급 시장에서 입지를 공고히 하겠다는 의지로 읽혀진다. 중급/보급형 AI PC 시장이 가장 큰 시장 규모를 가지고 있을 것이기 때문에 AMD의 전략은 현명한 포지셔닝으로 보이기도 한다.

그러나, AI 대중화를 위한 AI 에이전트 시장의 침범 역할을 하기 위해서는 아직까지는 와우 포인트를 만들어낼 정도의 고성능/게이밍 GPU가 여전히 많은 사람들의 손에 쥐어져야 하는 상황이다. 게다가, AI 에이전트가 실제로 소비자들에게 선택받게 되기 위해서는 OpenAI의 o1, DeepSeek의 R1과 같은 reasoning 기반 언어모델을 잘 서비스할 수 있어야 한다.

그리고 우리 팀은 소비자들이 생성 AI를 가장 많이 활용하는 미디어는 결국 영상이나 게임이 될 것으로 추측된다. **즉, 고성능/게이밍 GPU 출시에 소극적인 자세는 좀 아쉬운 부분이라 할 수 있다. 이번에 AMD가 제시한 Ryzen AI 300/200으로는 다소 부족한 GPU 성능이 아닐까 추측해본다.**

III. Paper of the Week

시가 노동시장에 미치는 영향 – 변곡점 효과

AI의 성능이 모두의 예상보다도 빠르게 향상되며 많은 사람들의 삶을 바꾸어 놓았다. ChatGPT의 출시 이후로 사람들이 일하는 방식이 크게 변화했고 우리 팀 역시 ChatGPT 출시 이전에 어떻게 일했는지 잘 기억이 나지 않을 정도다. 이렇듯 AI 발전이 근로자들의 일상을 바꾸어 놓은 반면, 이러한 변화가 거시적인 노동 시장에 미친 영향은 거의 연구되지 않았다. '사무직의 업무 효율화'는 외부에서 정량적으로 평가하기 어려운 요소였기 때문이다. 반면 프리랜서들의 경우 각 작업별로 소요 시간, 가격 등이 따로 책정되기 때문에 시가 미친 영향은 비교적 직관적으로 확인할 수 있다. 이 논문은 ChatGPT의 출시 시점을 기준으로 프리랜서들의 구직 사이트에서 어떤 변화가 있었는지를 추적해 AI의 출시가 여러 직종에 걸쳐 어떠한 영향을 주었는지를 분석했다.

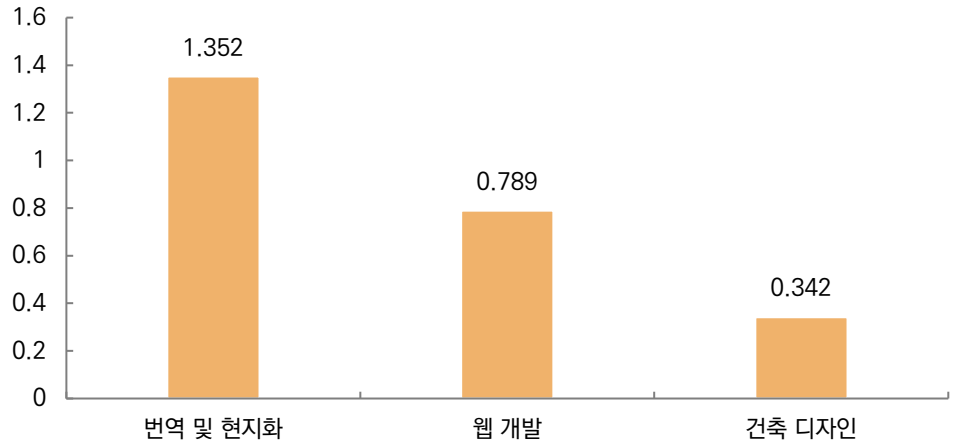
시가 노동 시장에 미치는 영향을 설명하기 위해 논문에서는 '변곡점'의 개념을 제시했다. 연구진은 이론적으로 각 직종에 대해 변곡점이 존재할 것이라고 보았다. 이 변곡점은 **AI 발전이 근로자에게 미치는 영향이 긍정적(생산성 향상)에서 부정적(사람이 시로 대체됨)으로 전환되는 지점**을 의미하며, 변곡점 이전에는 인간 근로자들이 AI 발전에서 이익을 얻지만, 변곡점을 지나고 나면 인간 근로자들은 시로 대체되기 시작한다는 것을 의미한다. 또한 연구진은 이것이 **비가역적 과정이며, 변곡점을 지나고 난 뒤에는 다시 그 전으로 돌아갈 수 없을 것**이라고 예측했다.

이러한 예측이 들어맞는지를 확인하기 위해 연구진은 프리랜서들의 온라인 구인 구직 사이트(OLM)에서 데이터를 수집했다. 여기에는 글쓰기, 프로그래밍, 건설, 회계 등 다양한 범주를 아우르는 광범위한 직업군이 있다. 이 사이트에서 프리랜서들은 자신의 전문성에 맞게 고정 금액 또는 시간당 임금 방식으로 업무를 수주하고, 고용주는 작업 완료 후 리뷰와 평점을 통해 프리랜서를 평가한다. 이렇듯 작업의 종류별로 세분화되어 있는 풍부한 데이터를 기반으로 논문은 이중차분법(difference in differences)을 사용해 시가 각 구직 시장에 미친 영향을 분석했다. 이중차분법은 정책의 효과를 추정하는 분석 방법의 일종으로, '정책 시행 전후 종속 변수의 변화'와 '실험군과 대조군 간 차이'를 통해 정책의 효과를 추정하는 방식이다.

이를 위해서는 먼저 대조군을 설정해야 한다. 따라서 논문에서는 AIOE(AI Occupational Exposure) 지수를 통해 ChatGPT와 상관관계가 거의 없는 직업군을 확인했다. AIOE 지수는 각 직업군이 AI에 얼마나 노출되었는지를 나타내는 지표로, 평균을 0, 표준편차를 1로 두었을 때 평균에서 얼마나 떨어져 있는지를 나타낸 지표다. 예를 들어, AIOE 지수가 1.5 점이라면 평균보다 1.5 표준편차만큼 높은 AI 노출도를 의미한다. 측정 결과, 온라인 구직 사이트에 포함된 직업군 중 건축 디자인 분야가 0.342로 낮은 노출도를 보였고, 구글 검색량 지수를 통해 ChatGPT와 각 분야명이 동시에 검색된 강도를 확인한 결과 역시 이를 뒷받침했다. 이에 따라 건축 디자인 분야를 대조군으로 설정했다.

표 18. 직업별 AI 노출 지수(AIOE) 비교

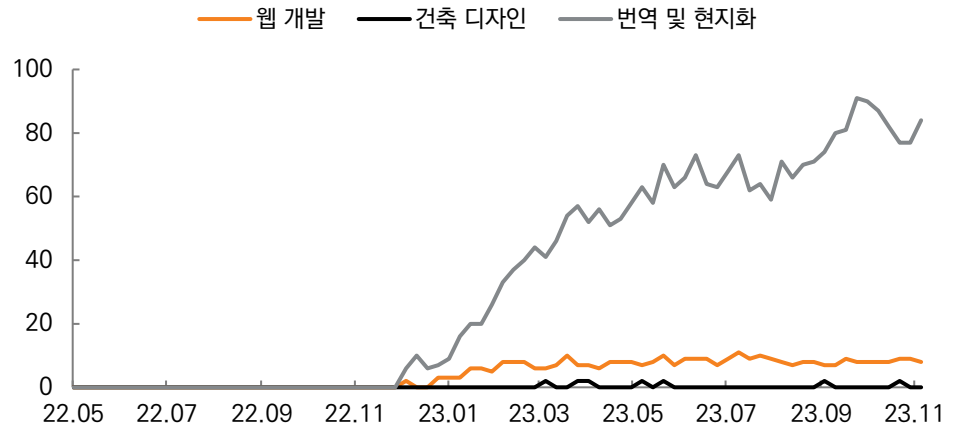
번역 및 현지화는 평균보다 1.352 표준편차만큼 높은 AI 노출도를 지닌 반면, 건축 디자인은 0.342로 낮은 노출도를 보여 이를 대조군으로 선정



자료: 미래에셋증권 리서치센터

표 19. 직업별 주간 구글 트렌드 지수(SVI) 비교

구글에 ChatGPT와 각 산업명을 동시에 검색한 건수를 상대강도로 표기. 번역 및 현지화나 웹 개발에 비해 건축 디자인과 ChatGPT를 동시에 검색하는 사례는 매우 적은 것을 확인

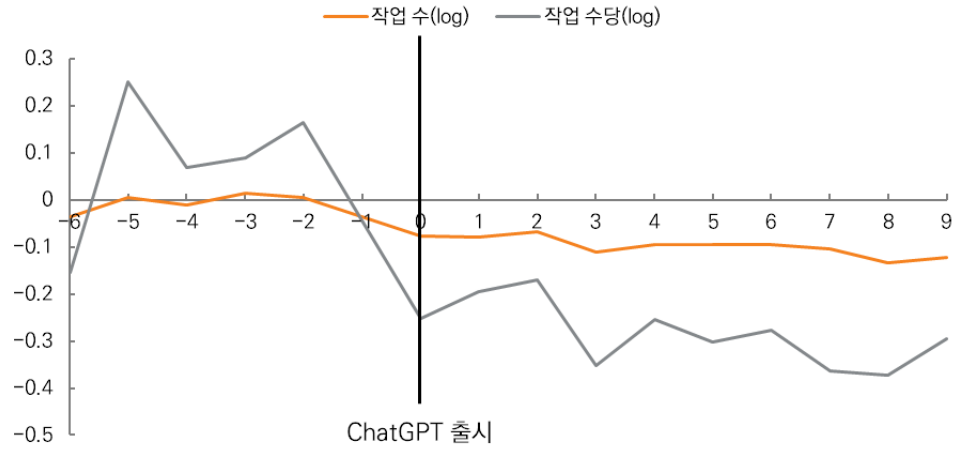


자료: 미래에셋증권 리서치센터

그 후 연구진은 ChatGPT가 강점을 보이는 번역 및 현지화와 웹 개발, 그리고 대조군인 건축 디자인을 중심으로 계약 건수와 금액, 소요 시간 데이터를 수집한 뒤, ChatGPT의 출시 시점을 외부 충격이 주어진 시점으로 간주하고 그에 따른 변화를 분석했다. 그 결과, **번역 및 현지화 분야에서는 ChatGPT 도입 후 작업량이 9.0%, 수입이 29.7% 감소한 것으로 나타났으며, 이는 AI에 의해 근로자가 대체될 수 있다는 논문의 주장을 뒷받침한다.** 반면, **웹 개발 분야에서는 ChatGPT 도입 후 작업량이 6.2%, 수입이 66.5% 증가한 것으로 나타나, 시가 생산성 향상에 기여했음을 보여준다.**

표 20. 번역 및 현지화 분야의 작업 수, 비율 및 수당 변화

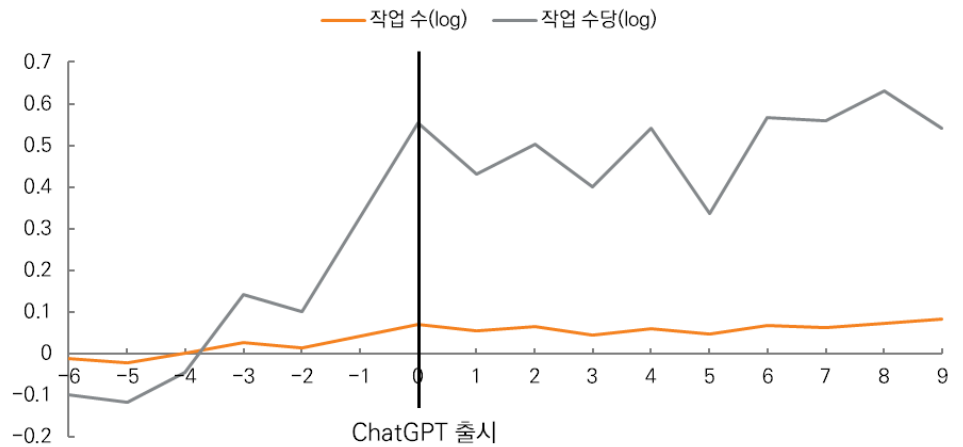
번역 및 현지화의 경우 ChatGPT 출시 이후로 작업 의뢰 건수와 수당이 감소



자료: 미래에셋증권 리서치센터

표 21. 웹 개발 분야의 작업 수, 비율 및 수당 변화

반면 웹 개발의 경우 ChatGPT 출시 이후로 작업 의뢰 건수 및 수당이 모두 증가



자료: 미래에셋증권 리서치센터

또한 연구진은 AI 능력이 증가할수록 시장 잠재력이 감소한다는 가설을 제시했다. AI에 익숙한 고용주들이 많아질수록 인간 노동력을 시로 대체하려는 경향이 커지기 때문이다. 이러한 시장 잠재력의 변화는 일반적으로 오목함수 형태를 보이며, 이는 기술 수용이 가속화 되는 현상을 반영한다. 이 때 **AI 능력이 변곡점 이전일 때는 AI 발전이 근로자의 생산량과 이윤을 증가시키지만, AI 능력이 변곡점을 넘어서면 시가 발전할수록 근로자의 생산량과 이윤이 줄어든다.** 실제 사례를 통해 살펴보면, 번역 시장은 ChatGPT가 이미 대부분의 번역 작업을 처리할 수 있기 때문에 변곡점을 넘어서는 상태로 대체 효과가 발생하는 반면, 웹 개발 시장은 ChatGPT의 능력이 코드 확인, 함수 식별 등에 제한되어 아직 변곡점 이전 단계로 생산성 향상 효과가 발생하는 것으로 해석할 수 있다.

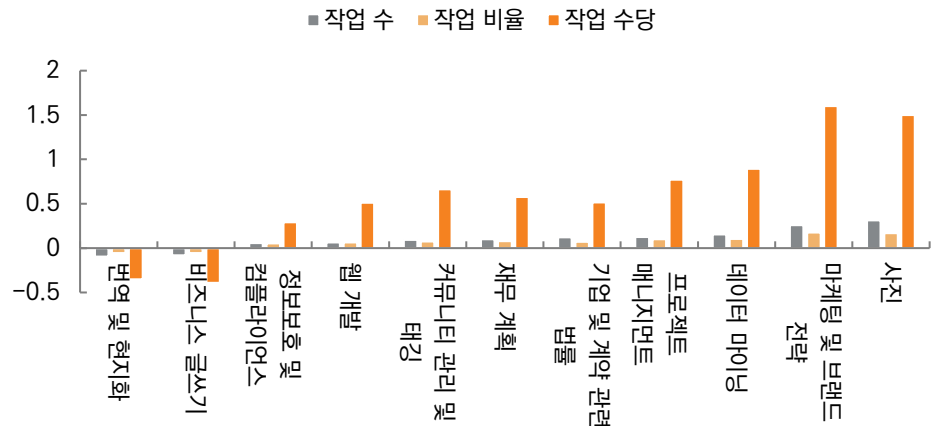
다만 이 결과는 ChatGPT가 처음으로 출시된 순간을 기준으로 한 것이다. 따라서 연구진은 GPT-4.0으로 업그레이드되는 등 기술 혁신이 발생하는 경우에 대해서도 변곡점 가설이 성립하는지를 확인하기 위해 세 가지 패턴을 예측했다.

1. 두 번의 업그레이드 모두에서 변곡점이 넘지 않는다면 생산성 효과가 지속될 것
2. 첫 번째 업그레이드에서는 변곡점 아래에 있지만, 두 번째 업그레이드에서 변곡점을 초과한다면 생산성에서 대체로의 전환이 일어날 것
3. 두 번의 업그레이드 모두에서 변곡점 위에 있다면 대체 효과가 지속될 것

연구 결과, 번역 및 현지화, 전문적 글쓰기 분야에서는 지속적 대체 효과가, 정보 보안 분야에서는 생산성에서 대체로의 전환이, 사진 및 마케팅/PR 분야에서는 지속적 생산성 효과가 관측되었다. 특히, 대체 효과에서 생산성 효과로 전환되는 경우는 관찰되지 않았는데, 이는 변곡점 가설을 강력하게 지지하는 증거이며, **한번 대체 효과가 지배적이 되면 이후 AI 발전으로도 되돌릴 수 없다는 것을 시사한다.**

표 22. 직업군별 ChatGPT와의 상관관계

글쓰기 관련 직종은 이미 대체효과를 보이고 있고, 마케팅 및 전략, 사진 등과 같은 직종은 AI로 인한 생산성 향상을 누리고 있음



자료: 미래셋증권 리서치센터

AI는 노동 시장에 대체와 생산성 향상이라는 두 가지 상반된 영향을 동시에 미치고 있다. 여러 직종에 대한 ChatGPT와의 상관관계 분석 결과, 글쓰기 관련 직종은 이미 대체 효과에 취약한 반면, 프로그래밍 관련 직종은 현재 생산성 효과가 나타나고 있으나 향후 대체될 수 있으며, 운영/창의성 관련 직종은 지속적 생산성 효과가 예상된다. AI가 점차 발전함에 따라 더 많은 직업들이 변곡점 너머로 사라질 듯하다. 기술이 하루가 다르게 발전하는 가운데 개인이 AI에게 대체되지 않는 방법은 AI를 잘 활용해 자신의 생산성을 올리는 것뿐이다.

III. AI Peer Table

표 23. Peer Table (1)

종목명	티커	주가 (현지통화)	시가총액 (조원)	기간별 수익률(%)						실적발표 예정일	AI 밸류체인 관련 한줄평
				1W	1M	3M	6M	1Y	YTD		
최종 소비자(온디바이스, 앱)											
애플	AAPL US	243.85	5,434	-5.9	1.8	7.6	11.0	32.0	-2.6	01-30 엔드유저 데이터 보유 및 서비스 배포능력	
테슬라	TSLA US	379.28	1,795	-16.5	6.2	52.3	64.0	52.7	-6.1	01-29 FSD v13 공공 출시 임박 및 1Q25 중국 출시 가능성	
삼성전자	005930 KS	53,400	319	-0.6	-0.4	-12.9	-34.4	-29.6	0.4	01-08 온디바이스 AI, HBM, 파운드리 영역에서의 가능성	
퀄컴	QCOM US	153.64	252	-3.1	-5.3	-8.2	-22.4	11.7	0.0	01-31 생성 AI 처리 강화용 온디바이스 CPU 설계(X Elite)	
스냅	SNAP US	11.24	28	0.4	-7.2	3.3	-30.9	-30.4	4.4	02-06 이미지 생성 모델을 탑재한 증강현실 서비스	
비즈니스 효율화											
어도비	ADBE US	441.00	286	-2.0	-14.6	-12.8	-22.3	-24.0	-0.8	03-14 세계 최고의 미디어편집 툴. 서비스 배포능력(Firefly)	
세일스포스	CRM US	330.66	467	-3.2	0.0	18.5	29.6	29.8	-1.1	02-28 세계 최고의 CRM 업체. AI로 사용성 강화(Einstein)	
서비스나우	NOW US	1,054.34	320	-4.1	0.6	19.8	32.8	53.4	-0.5	01-24 워크플로우 자동화. AI로 사용성 강화(Now Assist)	
클라우드스트라이크	CRWD US	347.34	126	-4.9	0.2	23.9	-9.8	40.7	1.5	03-05 기업 고객 대상 엔드포인트 보안(Charlotte AI)	
IBM	IBM US	219.94	300	-2.2	-3.3	0.9	26.1	41.1	0.1	01-29 기업 고객 대상 AI 모델 개발, 배포 플랫폼(watsonx)	
액센추어	ACN US	348.82	322	-3.2	-3.5	-1.7	15.8	2.2	-0.8	03-20 각 산업별 맞춤형 AI 에이전트를 만들기 위한 컴빙	
SAP	SAP GY	238.55	443	0.9	3.2	17.1	28.8	75.9	1.0	01-28 기업용 SW 솔루션에 AI를 내장한 에이전트(Joule)	
인포시스	INFO IN	1,957.85	140	2.6	4.2	4.0	22.2	31.6	4.1	01-16 각 산업별 맞춤형 AI 에이전트를 만들기 위한 컴빙	
타타 컨설턴시	TCS IN	4,175.75	260	0.2	-2.4	-2.4	4.4	12.5	2.0	01-09 각 산업별 맞춤형 AI 에이전트를 만들기 위한 컴빙	
유니티	U US	24.51	15	1.4	1.1	16.5	53.1	-36.8	9.1	02-26 차세대 게임엔진에 AI 기능 대거 탑재(Sentis, Muse)	
모더나	MRNA US	42.00	24	3.9	-5.1	-33.5	-64.1	-62.7	1.0	02-21 신약 개발 등 모든 업무에 생성 AI 활용(Dose ID)	
루닛	328130 KS	65,600	2	7.2	4.8	50.3	42.3	-14.5	2.8	03-21 AI 기반 암 검출/진단 보조 솔루션(인사이트)	
AI Ops											
몽고DB	MDB US	244.62	27	0.5	-24.8	-3.8	-5.9	-36.2	5.1	03-07 AI 모델 구축시 핵심 인프라인 '벡터 DB' 기능(Atlas)	
클라우드플레어	NET US	112.54	57	-0.1	7.1	41.8	32.3	41.8	4.5	02-07 실시간 AI 추론에 CDN 수요증가 가능성(Workers AI)	
팔란티어	PLTR US	75.19	252	-8.5	13.3	100.6	191.1	353.5	-0.6	02-05 비즈니스 현장의 의사결정을 돕는 AI 플랫폼(AIP)	
포스 패러다임	6682 HK	45.30	4	-16.7	17.8	81.2	-15.1	-4.2	-11.1	03-20 중국의 팔란티어(Sage)	

자료: Bloomberg, 미래에셋증권 리서치센터 / 주: 실적발표 일정은 변경될 수 있음.

표 24. Peer Table (2)

종목명	티커	주가 (현지통화)	시가총액 (조원)	기간별 수익률(%)						실적발표 예정일	AI 밸류체인 관련 한줄평
				1W	1M	3M	6M	1Y	YTD		
파운데이션 모델											
알파벳	GOOGL US	189.43	3,429	-3.2	10.6	14.3	2.5	37.6	0.1	01-30 OpenAI와의 AGI 경쟁이 가능한 업체(Gemini)	
메타 플랫폼스	META US	599.24	2,231	-0.7	1.2	4.7	17.8	73.7	2.3	01-31 오픈소스 AI 개발의 선구자(Llama 등)	
알리바바	9988 HK	81.30	294	-1.3	-3.7	-26.1	12.6	13.3	-1.3	02-07 중국의 CSP이자 중국 최고의 언어모델(Qwen)	
바이두	9888 HK	80.80	43	-4.9	-4.0	-25.7	-6.3	-28.2	-2.3	02-28 NLP 및 자율주행 모델의 전통강호(Ernie Bot, Apollo)	
센스타임	20 HK	1.41	10	-7.8	-6.0	-20.3	-13.0	25.9	-5.4	03-26 중국의 멀티모달 AI의 다크호스(SenseNova)	
네이버	035420 KS	193,800	31	-2.0	-7.3	15.2	21.3	-12.7	-2.6	02-03 한국형 LLM 개발 및 AI 서비스(HyperCLOVA)	
클라우드											
마이크로소프트	MSFT US	418.58	4,588	-4.5	-2.9	0.5	-8.5	13.7	-0.7	01-30 OpenAI 모델 라이선스 독점권 보유한 세계 2위 CSP	
아마존닷컴	AMZN US	220.22	3,414	-3.0	4.5	19.2	10.1	46.9	0.4	01-31 최고의 CSP로서 Anthropic에 수 조원 투자	
오라클	ORCL US	166.03	685	-3.3	-8.5	-0.8	16.5	61.5	-0.4	03-11 AI 데이터센터 capex 경쟁에 진입한 전통적 강자	
소프트뱅크	9984 JP	9,185.00	126	-0.7	0.4	5.8	-14.0	46.7	0.0	02-07 '비전펀드'는 AI에 집중. 일본 최고의 AI 슈퍼컴퓨터.	
하드웨어 인프라											
엔비디아	NVDA US	138.31	4,994	-1.2	-0.2	16.4	12.8	187.2	3.0	02-26 AI 모델 훈련 및 추론에 필수인 GPU계의 현존 최강자	
브로드콤	AVGO US	231.98	1,603	-5.5	39.7	36.3	40.8	116.5	0.1	03-07 이더넷 기반 네트워킹 반도체의 최강자	
AMD	AMD US	120.63	289	-3.5	-15.1	-24.5	-26.6	-13.0	-0.1	01-30 서버용 CPU의 최강자. AI 가속기 분야 패스트팔로어	
인텔	INTC US	20.22	129	-1.1	-15.5	-9.7	-34.5	-57.1	0.8	01-24 파운드리로서 소버린 AI 미 지정학적 가치 부상	
마이크론	MU US	87.33	143	-2.6	-11.3	-12.3	-33.9	6.5	3.8	03-20 SK하이닉스 추격 중인 "미국"의 메모리 반도체	
SK하이닉스	000660 KS	171,200	125	-1.9	3.8	1.2	-27.3	25.7	-1.6	01-24 HBM 부문 전세계 최강자	
시놉시스	SNPS US	482.75	110	-2.6	-14.7	-2.8	-20.8	-3.3	-0.5	02-21 반도체 EDA 부문 리딩 업체, AI 에이전트 적극 활용	
암페놀	APH US	69.01	123	-3.6	-5.4	11.3	2.6	43.7	-0.6	01-22 엔비디아향 사용 서버 내 고속 I/O 및 커넥터 제조	
버티브	VRT US	118.30	65	-0.5	-6.9	15.7	34.5	159.5	4.1	02-21 액체 냉각 방식에 있어 글로벌 선두업체	
Arm	ARM US	128.20	199	-1.1	-8.7	-6.4	-21.6	86.0	3.9	02-07 AI 가속기 설계 위한 다수의 IP를 소유한 팹리스	
마벨 테크놀로지스	MRVL US	113.56	145	-1.9	17.2	57.8	58.9	95.9	2.8	03-07 데이터 인프라용 네트워킹 및 스토리지 ASIC 설계	
SMIC	981 HK	29.00	69	-5.4	11.5	37.1	65.3	53.1	-8.8	02-06 7나노 공정을 달성한 중국 유일의 미세공정 파운드리	
TSMC	2330 TT	1,065.00	1,238	-2.3	1.3	10.0	9.7	87.3	-0.9	01-16 명실상부 반도체 파운드리 최강자	
관타	2382 TT	280.00	48	-3.8	-4.4	4.5	-9.1	35.2	-2.4	03-17 GPU 기반 보드 및 서버 시스템 조립 담당, 대만 업체	
위스트론	3231 TT	103.00	13	-3.3	-10.4	1.5	-5.9	12.8	-1.0	01-20 GPU 기반 보드 및 서버 시스템 조립 담당, 대만 업체	
Alchip	3661 TT	3,100.00	11	-6.8	41.6	61.0	26.7	-9.2	-5.5	03-03 빅테크들의 AI 가속기 설계를 돕는 대만의 팹리스	

자료: Bloomberg, 미래에셋증권 리서치센터
 주: 실적발표 일정은 변경될 수 있음.

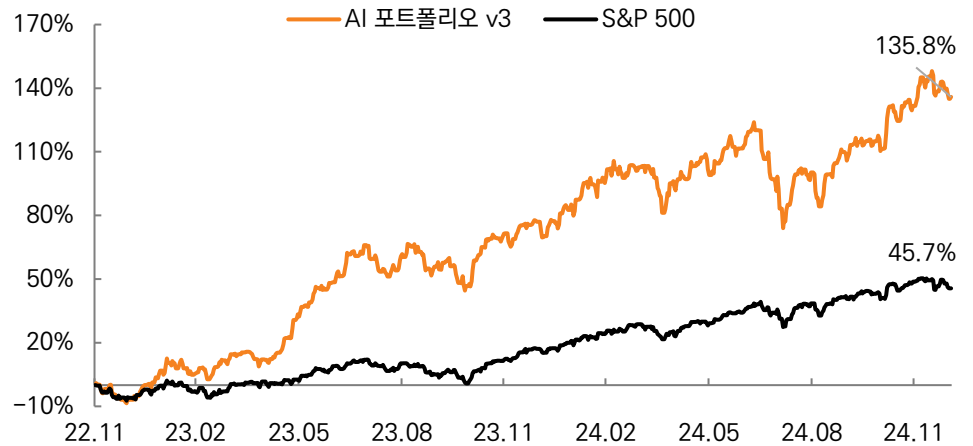
IV. Appendix: Charts

그림 1. ChatGPT 출시일 이후, S&P 500에서 Magnificent Seven이 차지하는 비중(시가총액 기준)



자료: Bloomberg, 미래에셋증권 리서치센터

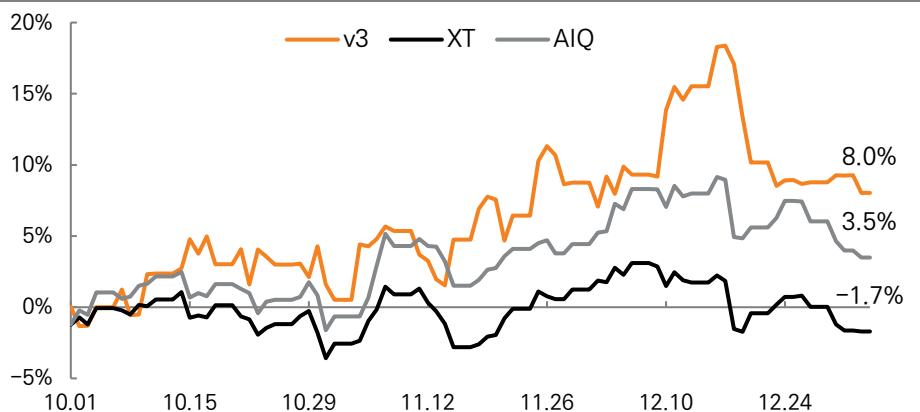
그림 2. ChatGPT 출시일 이후, AI 밸류체인에 속한 유니버스 종목들의 평균 수익률(동일가중 방식)



자료: Bloomberg, 미래에셋증권 리서치센터

그림 3. 대표 AI ETF vs AI 유니버스(v.3) 수익률 추이(24.09.30 기준)

- Global X Artificial Intelligence & Technology ETF(AIQ), iShares Exponential Technologies ETF(XT)



자료: Bloomberg, 미래에셋증권 리서치센터

표 25. LLM 벤치마크 챗봇아레나의 “Hard Prompts” 기준 성능 상위 모델 10개(동일 모델 생략)

순위 (전 주 변동)	모델명	아레나 점수	기관명	라이선스	지식 컷오프
1(-)	Gemini-exp-1206	1364	구글	독점 폐쇄소스	-
2(New)	Gemini-2.0-flash-thinking-exp-1219	1359	구글	독점 폐쇄소스	2024년 8월
3(New)	o1-2024-12-17	1358	OpenAI	독점 폐쇄소스	
4(▼2)	o1-preview	1352	OpenAI	독점 폐쇄소스	2023년 10월
5(▼2)	Gemini-2.0-flash	1347	구글	독점 폐쇄소스	2024년 8월
6(New)	Deepseek-v3	1319	DeepSeek	상업적 사용허가	
7(▼2)	ChatGPT-4o-latest	1338	OpenAI	독점 폐쇄소스	2023년 10월
8(▼4)	o1-mini	1337	OpenAI	독점 폐쇄소스	2023년 10월
9(▼3)	Gemini 1.5 Pro	1299	구글	독점 폐쇄소스	2023년 11월
10(▼1)	Deepseek-v2.5	1289	DeepSeek	상업적 사용허가	-

자료: Imarena.ai, 미래에셋증권 리서치센터

주: Hard Prompts는 기존의 일반적인 요청보다 훨씬 더 까다롭고 복잡한 문제를 제시하여 LLM의 한계를 테스트하는 벤치마크

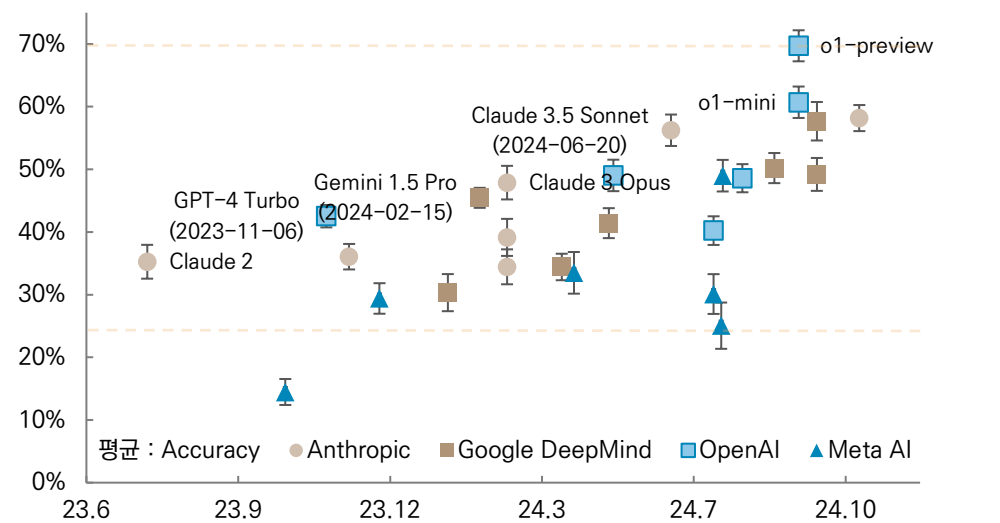
표 26. LLM 벤치마크 LiveBench의 성능 상위 모델 10개(동일 모델 생략)

순위 (전 주 변동)	모델명	전체 평균 점수	기관명	라이선스	지식 컷오프
1(New)	o1-2024-12-17	75.67	OpenAI	독점 폐쇄소스	
2(-)	Gemini-exp-1206	64.09	구글	독점 폐쇄소스	-
3(New)	Gemini-2.0-flash-thinking-exp-1219	61.83	구글	독점 폐쇄소스	
4(New)	deepseek-v3	60.45	Deepseek	상업적 사용허가	
5(▼1)	Gemini-2.0-flash	59.26	구글	독점 폐쇄소스	2024년 8월
6(▼3)	Claude-3.5 Sonnet	58.74	Anthropic	독점 폐쇄소스	2024년 4월
7(▼1)	o1-mini	57.76	OpenAI	독점 폐쇄소스	2023년 10월
8(▼3)	Gemini-exp-1121	57.36	구글	독점 폐쇄소스	-
9(▼3)	GPT-4o	55.33	OpenAI	독점 폐쇄소스	2023년 10월
10(▼3)	Gemini 1.5 Pro	54.33	구글	독점 폐쇄소스	2023년 11월

자료: LiveBench, Huggingface, 미래에셋증권 리서치센터

표 27. 각 AI 모델들의 GPQA Diamond 기준 성능 비교

GPQA Diamond는 박사 수준의 과학 분야 질문으로, 인간 전문가들도 약 65%의 정확도를 보임



자료: EPOCH AI, 미래에셋증권 리서치센터

Compliance Notice

- 당사는 자료 작성일 현재 조사분석 대상법인과 관련하여 특별한 이해관계가 없음을 확인합니다.
- 당사는 본 자료를 제3자에게 사전 제공한 사실이 없습니다.
- 본 자료를 작성한 애널리스트는 자료작성일 현재 조사분석 대상법인의 금융투자상품 및 권리를 보유하고 있지 않습니다.
- 본 자료는 외부의 부당한 압력이나 간섭없이 애널리스트의 의견이 정확하게 반영되었음을 확인합니다.

본 조사분석자료는 당사의 리서치센터가 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.