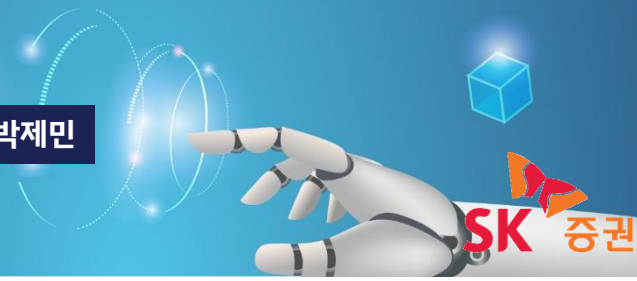

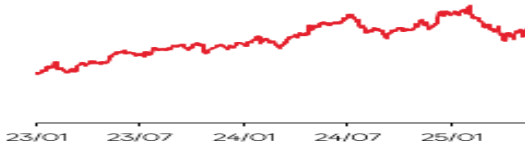

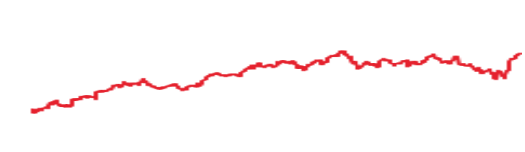
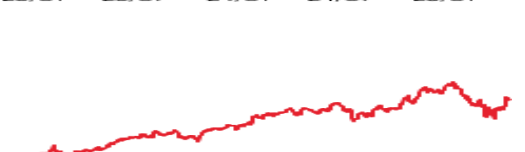
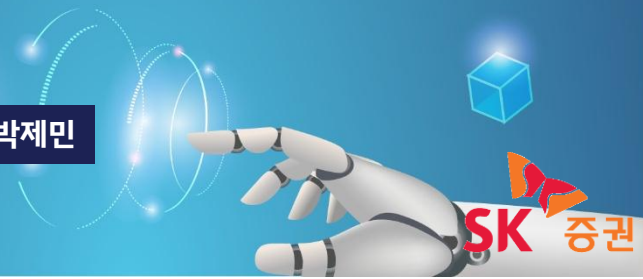


AI WEEKLY

SK증권 해외주식 박제민

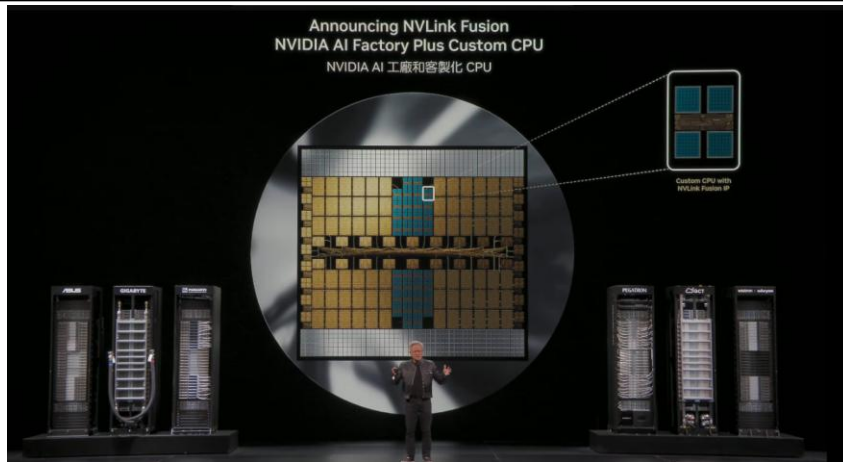


종목 (티커)	시가총액(B, \$)	수익률 (% , 1W/1Y/커버리지 이후)			주가 추이 (23.01.01~)
마이크로소프트 (MSFT)	3,410	1.3	7.8	18.3	
- Azure로 AI 시대 하이퍼스케일러 포지션 구축, Office의 저조한 AI 침투율은 염려 사항 - 지난 주 실적 발표 이후 꾸준한 상승 추세 지속 중, Build 호평이 주가 모멘텀을 만들진 못하는 모습					
엔비디아 (NVDA)	3,306	0.1	43.0	22.4	
- 추론 시대 도입에 따라 GPU 사용량 증가 지속, Test time scaling으로 훈련 수요도 여전 - 지난 주 중동 투자 지속으로 매출 안정성 확보되며 강세, 신용등급 강등 등 매크로에 가장 탄력적으로 반응					
아마존 (AMZN)	2,188	-1.9	12.3	13.2	
- 물류센터 추가 건설을 통한 배송 최적화 지속 진행 중, AWS 역시 비용 최적화 관점이 우선시 - 미중 관세 완화 이후 강한 되돌림, 아직 주가가 AWS보다 스토어 부문 위주로 움직이는 경향					
알파벳 (GOOGL)	2,028	0.7	-5.0	4.6	
- 플랫폼 AI 제품 증가에 따른 Subscription 사업부 확장 주목. Search 부문 경쟁 우려는 지속 - 지난 주 주가는 이번 주 구글 I/O를 앞두고 쉬어가는 모습, 반독점 관련 잡음 지속					
메타 (META)	1,610	-1.9	36.5	17.0	
- GPU 전환, LLM 도입으로 피드, 광고 엔진 강화 진행 중. 광고 생성 엔진 가동으로 비즈니스 모델 고도화 - 지난 주 주가는 쉬어가는 모습, 미중 관세 완화의 강한 되돌림 이후 숨고르는 기간					



Computex: 이제 ASIC도 품는 엔비디아

NVLINK Fusion: ASIC 우려 일부 완화 가능



자료 : NVIDIA, SK증권

DGX Spark: 연구자용 미니 컴퓨터



자료 : NVIDIA, SK증권

젠슨황 기조연설 주요 포인트: 1. NVLINK Fusion / 2. RTX Pro / 3. DGX Spark / 4. Physical AI

1. NVLINK Fusion: ASIC을 엔비디아 컴퓨팅, 네트워킹 생태계 내에 편입할 수 있도록 돕는 기술

1) AI를 위한 경쟁 보다 '확산'에 방점 2) 동일 환경 내 성능 격차 자신감 3) 성공적 ASIC에도 네트워크 판매를 위한 포섭 4) 신규 생태계 견제 효과

2. RTX Pro: GPU 중심 AI 공장(GB200/GB300)과 전통 x86 서버의 중간층을 채우는 하이브리드 제품. Agentic AI를 위한 포석

3. DGX Spark: AI 미니컴퓨터. 바로 B2C 확산을 도모하기엔 오버스펙, 일부 대학교 및 연구소 위주로 판매될 것으로 추정

4. Physical AI: CES, GTC에서 신규 플랫폼 (Omniverse, Groot)를 선보인 것에 비해 이번에는 성능 개선만 공개. SW 기를 마련된 것으로 판단

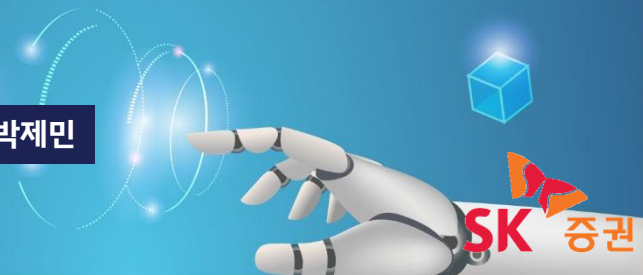
Physical AI의 요구 토큰 양은 기존 LLM 대비 매우 높아 AI Capex 증설 논리 강화. 향후 추론 비용 감소에 따른 적용 가능성 추적 필요

퀄컴은 데이터센터 CPU 복귀 발표

Grace CPU처럼, Arm 기반 고성능 CPU + GPU 통합 아키텍처에 대응하는 전략, 메타 및 사우디 AI 기업 'Humain'과 협업도 진행 중
과거 철수했던 데이터센터 CPU 시장에 애플 출신 엔지니어 영입 후 재도전하는 모습.

AI WEEKLY

SK증권 해외주식 박제민



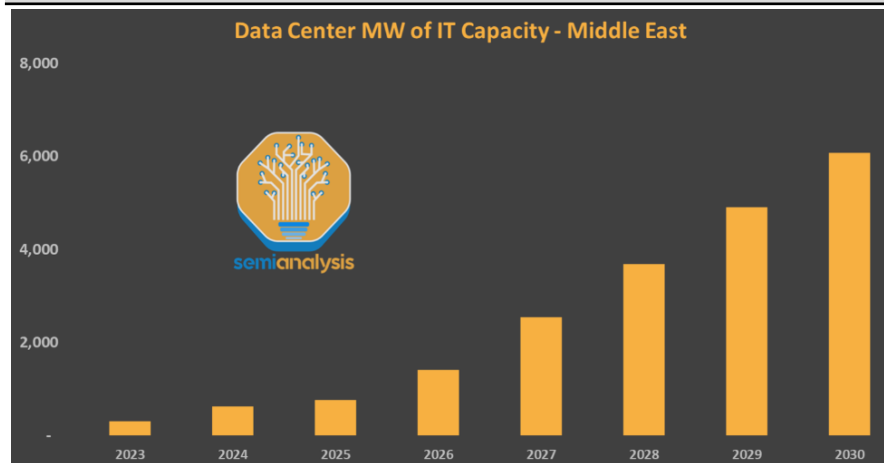
중동 고객사 확보로 AI Capex 우려 타파!

아부다비의 셰이크 자이드와 협상 이후 악수하는 트럼프



자료 : Reuters, SK증권

중동 데이터센터 투자 계획



자료 : Semianalysis, SK증권

AI 기술 + 오일 머니 전환을 원하는 중동, 저가 전력 + 투자금을 원하는 미국의 인프라 행보 지속

사우디 \$600B, UAE \$200B 이상 미국에 투자 계획 발표. 주요 투자 분야는 데이터센터, 에너지 인프라 투자

엔비디아는 1) 2027년까지 매년 최첨단 칩 50만개 판매 가능 (1년 생산량의 10% 수준) 2) 중동 지역 투자 유치로 자국 데이터센터 산업 안정성 증가 3) 중국 시장 잃을 위기이나 기존 중국 시장으로 인식되던 중동 시장을 확보함으로써 Upside 확보

특히 관세로 인한 CSP들의 AI Capex 우려가 지속되던 중 외부 자금 유입으로 이를 많은 부분 해결 가능해졌다는 점이 긍정적

엔비디아 매출 중 xAI, Stargate, 중동 지역 매출이 유의미하게 올라오면서 빅테크 4사 50% 수준의 매출 비중이 더 건전해질 것으로 판단

미국은 5/13 화웨이 Ascend AI 칩 사용이 미국 수출 통제 지침을 어긴다고 발표, 이는 중동 지역을 CUDA 생태계로 완전히 잡아두기 위한 움직임

AI WEEKLY

SK증권 해외주식 박제민

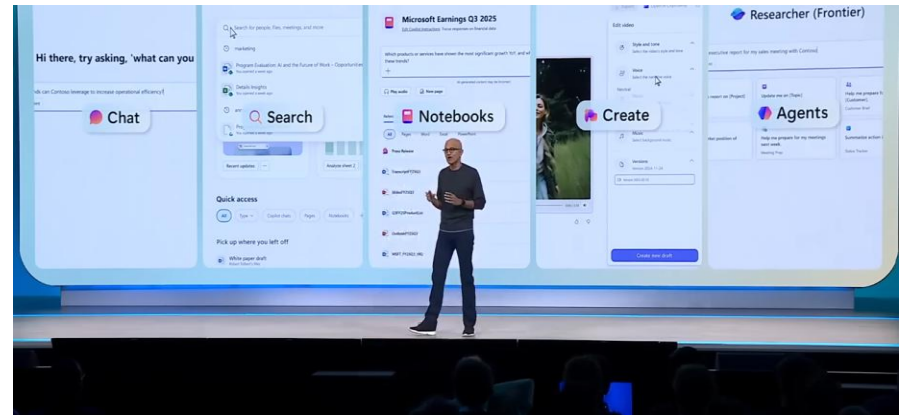
Microsoft Build: 전 층위에서 선택지를 늘리는 Azure

Microsoft Azure에 채택 가능한 LLM 선택지 1900개 이상



자료 : Microsoft, SK증권

Microsoft 365 Copilot의 App Redesign 발표, 통합된 UI 최초 출시



자료 : Microsoft, SK증권

Azure Foundry에서 활용할 도구 증가, 전반적 플랫폼 상태 개선(보안, 호환성, 추가기능)

Azure의 전략인 '최대한 다양한 서비스 제공으로 AI Tool의 플랫폼이 되자'는 기조 재확인

Foundry 관련하여서 Grok 추가 등 추가 이벤트 존재, MCP 및 멀티 모달 Router로 기능 확장 강조

Fabric, Cosmos DB, RAG 개선을 통해 현재 Enterprise AI 도입의 병목인 Data core 해결하려는 모습

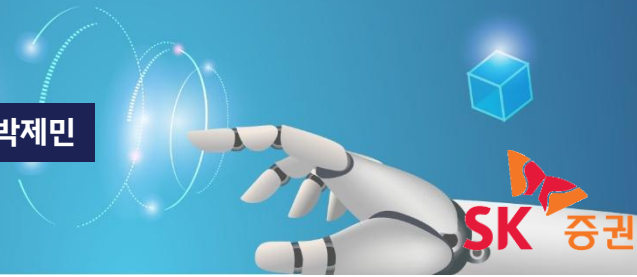
Microsoft 365 Copilot Redesign, 반전 포인트가 될 수 있을까

통합 앱으로 구성된 Office Copilot 공식 출시. Chat, Search, Notebooks, Create를 하나의 창으로 통합한 UI 발표

기존 프리뷰만 있던 제품들 공식 출시, Notebook 기능과 Copilot Tuning 기능 추가. 특히 Tuning 기능의 경우 Office에 RAG 기능이 추가되는 격

Microsoft Office Copilot 침투율이 제한적인 현재(<2%) 해당 Redesign이 변곡점이 되줄 수 있을지 지켜봐야

Build 2025를 통해 볼 수 있는 Azure Foundry의 락인은 1) Agent ID 2) Fine-tuning, RAG 등을 통한 데이터 종속 3) 보안

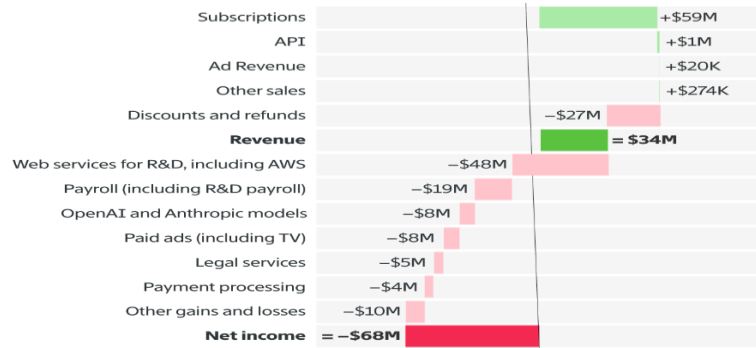


LLM 사업자가 돈 버는 법, 진화 중인 Test-time Thinking

Perplexity 2024년 손익 계산서 추정

Perplexity's Plotline

How the AI search engine made—and lost—money in 2024

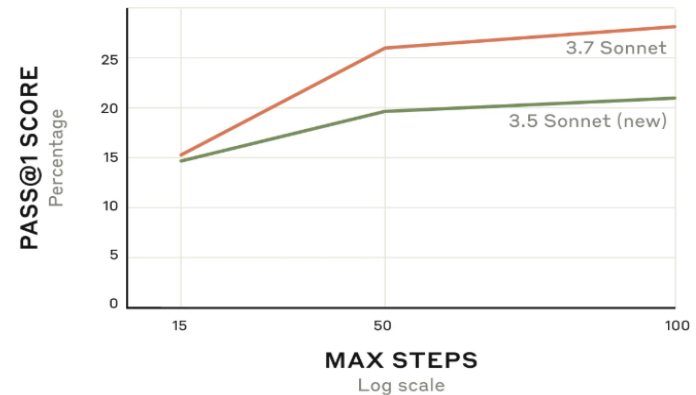


Note: Selected categories
Chart: Shane Burke • Source: The Information reporting

자료 : Theinformation, SK증권

독자적 'Extended Thinking' 기술로 Test time thinking 적용

OSWorld



자료 : Anthropic, SK증권

Perplexity, LLM 서비스만으로는 돈이 안되는 상황

작년 \$34M 매출을 올렸지만, \$65M의 적자를 기록. 무료 이용자들에게 제공하는 추론 비용이 매출을 훨씬 상회

AWS 비용(50%) + 모델/API 사용료(25%) 합치면 매출의 75% (Theinformation 추정)

추론 비용 공격적으로 낮아질 것으로 보이나, 그만큼 요구 서비스 정도도 늘어날 것으로 예상

희망적인 부분은 2025년 기준 연환산 매출 \$100M 돌파, 광고·커머스 진출, 60% 수준의 매출총이익률 확보

순수 영리 추구 AI 서비스 사업자가 어떤 행보를 보이는지 주목 필요 (OpenAI, xAI는 모두 이사회, 경영진이 공익 목적을 강조하며 활동 중)

Anthropic은 여전히 'test-time compute 기반의 강화된 추론 모델' 전략에 베팅 중

'추론 ↔ 도구 사용'을 오가는 멀티스텝 사고 능력을 강화 (self-correcting reasoning loop)하는 기술

기존 Reasoning 모델보다 오류를 인식하고 수정한다는 점에서 발전 포인트 존재

Reasoning 모델은 KV Cache 사용량 증가, 중간 메모리 요구량 증가. 해당 기술 위주의 확산 시 AI 서비스에 있어 메모리 사용량 증가 가능