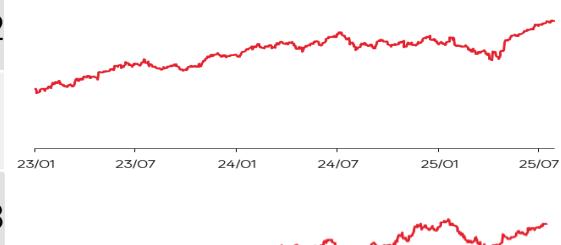
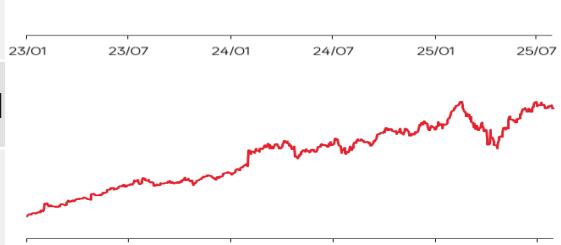


AI WEEKLY

SK증권 해외주식 박제민



종목 (티커)	시가총액(B, \$)	수익률 (%), 1W/1Y/커버리지 이후)			주가 추이 (23.01.01~)
엔비디아 (NVDA)	4,282	1.0	69.2	58.5	
		<ul style="list-style-type: none"> - Azure로 AI 시대 하이퍼스케일러 포지션 구축, B2B 비즈니스 틀인 Office의 AI 침투율 제한은 염려 사항 - AI Action Plan, 중국 수출 허용 등 호재로 최고가 지속 간접. 대부분 호재 반영 주가로 판단 			
마이크로소프트(MFST)	3,809	0.3	21.2	32.2	
		<ul style="list-style-type: none"> - 추론 시대 도입에 따라 GPU 사용량 증가 지속, Test time scaling으로 훈련 수요도 여전 - 실적발표 직전 최고 주가 간접 중, Edge Copilot Mode 공개로 브라우저 경쟁 참여 뉴스 			
아마존(AMZN)	2,452	-0.5	27.1	26.8	
		<ul style="list-style-type: none"> - 물류센터 추가 건설을 통한 배송 최적화 지속 진행 중, AWS 역시 비용 최적화 관점이 우선시 - 실적발표 앞두고 신중한 투자자들로 약보합, ServiceNow이 서버 GCP로 다변화 뉴스도 악재로 작용 			
알파벳 (GOOGL)	2,371	1.9	15.1	23.1	
		<ul style="list-style-type: none"> - 플랫폼 AI 제품 증가에 따른 Subscription 사업부 확장 주목. Search 부문 경쟁 우려는 지속 - 호실적 발표 이후 준수한 주가 흐름, AI Mode 미국에 이어 영국 롤아웃 			
메타 (META)	1,760	-2.1	51.1	24.1	
		<ul style="list-style-type: none"> - GPU 전환, LLM 도입으로 피드, 광고 엔진 강화 진행 중. 광고 생성 엔진 가동 시작 - EU 정치 뉴스 중단, 슈퍼인텔리전스 관련 뉴스 지속. 단기적 이익 상승보단 리스크 부각되며 하락 			

AI WEEKLY

SK증권 해외주식 박제민



백악관의 AI 경쟁에 적극 개입, 청사진 공개

백악관, AI Action Plan 공개



Winning the Race AMERICA'S AI ACTION PLAN

자료 : WhiteHouse, SK증권

미국 백악관, AI 전략 계획(Action Plan) 공개. 향후 행정부 정책의 청사진

계획은 7/23 공개. 백악관은 “향후 수 주~수 개월 내 90건 이상의 연방 정책 행동”을 추진하는 로드맵 성격임을 명시.
3개의 핵심 축 존재

1. 혁신 가속화: 불필요한 연방 규제, 지침 재검토 및 폐지. 주별 규제 강할 경우 AI 예산 지급 제한
2. AI 인프라 구축: DC 인프라 인허가 간소화, 반도체 제조 규제 간소화, 전력망 안정화 및 고도화(지열, 핵분열, 핵융합 등 우선)
3. 국제 외교: 동맹국에 풀스택 패키지 수출 촉진. 국제 기구 내 중국 영향력은 견제. 수출 통제를 제조 서브 시스템까지 확대

AI Infra: 1) 인프라 규제 완화 2) 윤리, 데이터 등 AI 학습 허들 완화 3) AI 도입 샌드박스로 추론 수요 긍정적 4) 중국 견제 지속으로 인프라 계층 독주
하이퍼스케일러: 1) 인프라 규제 완화 (DC 공급 병목 해소) 2) AI Infra 수요 증가(IaaS 가치 증가) 3) 오픈 소스 장려는 부정적 포인트 작용 가능

AI 서비스: 1) AI 도입 샌드박스 도입으로 수요 증가 2) AI 평가, 해석 관련 생태계로 신규 BM 창출 가능 3) 오픈소스 확산으로 신규 AI 서비스 출시 용이
발표 이후 5거래일간 NVDA +3.5%, AMD +11% Broadcom +3% 상승. 젠스황, 'AI와 에너지 중요성 가장 먼저 인식한 대통령이 트럼프' 발언

혁신 가속화, 인프라 구축, 국제 외교 3 축으로 구성

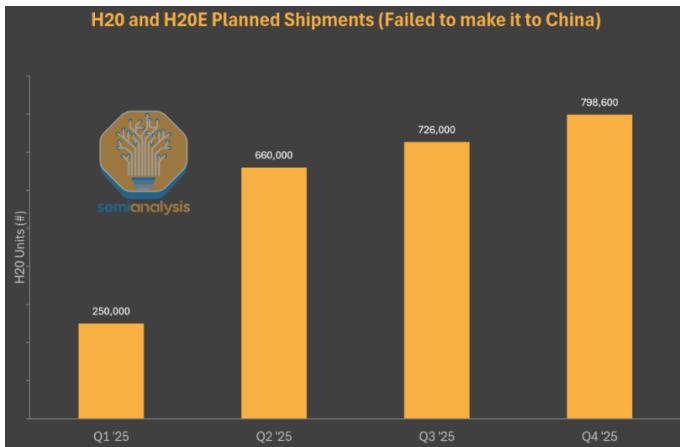
America's AI Action Plan has three pillars: innovation, infrastructure, and international diplomacy and security. The United States needs to innovate faster and more comprehensively than our competitors in the development and distribution of new AI technology across every field, and dismantle unnecessary regulatory barriers that hinder the private sector in doing so. As Vice President Vance remarked at the Paris AI Action Summit in February, restricting AI development with onerous regulation "would not only unfairly benefit incumbents... it would mean paralyzing one of the most promising technologies we have seen in generations."² That is why President Trump rescinded the Biden Administration's dangerous actions on day one.

자료 : Whitehouse, SK증권



엔비디아 중국 수출 완전 허용, H2O 재생산 돌입

기존 H2O 판매량 전망



자료 : Semianalysis, SK증권

엔비디아 H2O 라이선스 승인 획득, 재생산 주문까지

- (7/15) 엔비디아 블로그 통해 H2O 판매 라이선스 승인 자신감 표출, 주가 4% 상승 (H2O 재고 60~70만개, 2024년 총 100만 판매)
- (7/19) 엔비디아가 재고 외의 H2O 생산 계획 없음. 이미 철수한 TSMC 생산 라인의 재전환에 최대 9개월 소요 (Trendforce)
- (7/29) 기존 재고 판매 계획 철회, TSMC에 H2O 칩 30만개 신규 주문. 기존 보유분에 더해 총 100만개 이상 재고 확보 예정 (Reuters)
- (7/29) 백악관 경제보좌관, 인터뷰에서 트럼프 대통령 엔비디아칩 중국 수출 허용 발언. 기존 강했던 수요 + 라이선스 리스크 해소로 추가 주문 가능
- 엔비디아는 1) H2O 판매 수익도 필요하지만 2) Scaling Law 지속을 위해 중국의 AI 학자들의 AI 접근 보장 필요 3) CUDA 독점적 지위 지속 필요

NVIDIA 및 Huawei 주요 GPU 시스템 스펙 비교

기업명	NVIDIA GB200 NVL72	Huawei CloudMatrix 384
메모리 용량	13.8TB(HBM3E)	49.2TB(HBM)
메모리 대역폭	576 TB/s	1,229 TB/s
시스템 비용	\$3M	\$8.2M
전력 소비	145 kW	559 kW

자료 : Reuters, SK증권

중국 내 H100 수리 산업 부흥, Huawei Cloudmix 384 시스템 실물 공개

중국에서 H100/A100 등 금수 대상 Nvidia 칩의 수리 수요 급증, 공식적인 공급, 3년 보증 AS 불가로 밀수품 기반 수요 급증

H100 출시 이후 2년이 지나고 있는 상황, GPU 혹사로 일반적인 수명 (2~5년)보다 짧은 시기부터 수리 수요 급증

7/26 WAIC에서 CloudMatrix 384 실물 공개. 현재 수대 납품, 하반기 상용화 예정

NVL 72 대비 칩 성능 낮으나 시스템 단위 성능 충분(BF16 연산 성능, 메모리 용량 3.6배, 대역폭 2.1배), 전성비 뒤처지나 중국 전력 충분하여 수요 존재



신규 Scaling Law: Reinforcement Learning이란?

RL을 이용한 강화학습 원리, LLM의 필터로 작용



자료 : Medium AI, SK증권

Scaling Law 다변화로 요구 연산량 증가 및 연산 종류 다변화



자료 : OpenAI, SK증권

Scaling Law 다변화, Grok 4로 다시 조망받는 Reinforcement Learning

전통적인 LLM 훈련 방식의 성능 개선이 둔화되면서 RL(Reinforcement Learning)이 대안으로 부상

RL은 목표 달성을 보상을 주고, 잘못된 행동엔 패널티를 주는 방식, 주로 인간 피드백이나 문제풀이 예제를 통해 학습 진행

코딩, 수학 등 정답이 명확한 분야에서 잘 작동하나, 자동화가 어렵다는 단점

Salesforce, Amazon, DoorDash 등은 최근 앱을 복제한 디지털 환경(Gym)에서 RL 모델 훈련. Agent 주문, 리드 등록 등 실제 앱 사용 훈련 Gym을 통해 명시적으로 훈련하지 않은 앱도 자연스럽게 적응 가능, 그러나 구축에 막대한 컴퓨팅 양 필요

Elon Musk의 xAI 등은 RL에 동등한 컴퓨팅 리소스 투입 주장, OpenAI 연구원은 향후 RL에 훨씬 많은 컴퓨팅 양 사용 예고

8월 공개 예정인 GPT-5 또한 RL을 활용하여 특히 코딩 영역에서 성능이 크게 개선되었다는 보도 존재

RL은 DC 단위 효율적 운영 중요, Dynamo SW 등 인프라 소프트웨어 역할 중요. Rack scale에서 DC scale로 인프라 해자가 이동할 것으로 전망