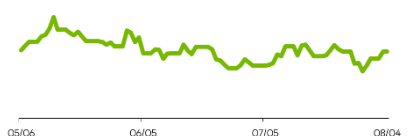
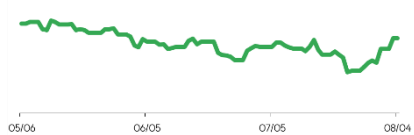
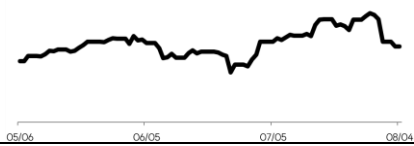
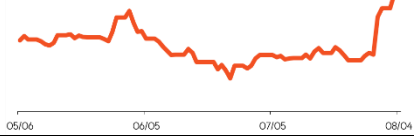
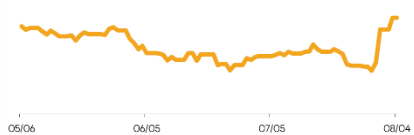
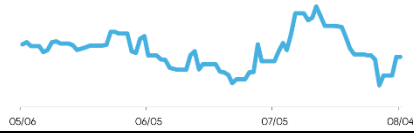
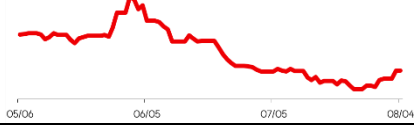




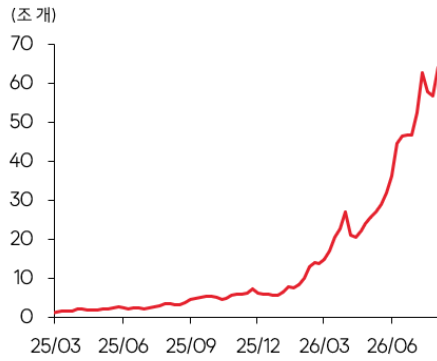
SK증권 해외주식/AI. 박제민
jeminwa@sks.co.kr

	엔비디아	시가총액: 5,001	12MF PER: 23	
		-독보적인 GPU 기술력있으나 ASIC으로 독점력 축소 -Rubin Ultra 기술 전환 및 CPU, LPU 주문 추이 주목		
	알파벳	시가총액: 4,562	12MF PER: 20	
		-TPU 활용 Anthropic 독주로 인프라 해자 인정받는 중 - Coding Agent 역전, 소비자용 Agent 출시 주목		
	애플	시가총액: 4,428	12MF PER: 34	
		-메모리 가격 상승이 브랜드 파워로 극복 가능한지 시험 중 -Siri 고도화로 인한 소비자용 Agent 출시 주목		
	마이크로소프트	시가총액: 3,621	12MF PER: 25	
	소프트	-OpenAI 해자 약화로 인프라 사업부 성장률 축소 -자체 모델 및 커스텀 AI 도입 컨셉 수요 및 기술력 주목		
	아마존	시가총액: 3,064	12MF PER: 28	
		-양대 AI Labs 모두에 공격적 투자를 통한 고객 확보 성공 -인프라 외 Bedrock 등 AlaaS ARR 전망에 주목		
	메타 플랫폼스	시가총액: 1,504	12MF PER: 17	
		-플랫폼 인기 여전, Muse Spark 개발로 유통과 모델 확보 -플랫폼 데이터, 글래스 등 활용한 소비자용 Agent 주목		
	오라클	시가총액: 409	12MF PER: 18	
		-부품 비용, 자본 비용 증가 추세 속 높은 수주 잔고 무색 -인력, 전기 쇼티지에도 매출 전환 성공 여부 주목		
	코어위브	시가총액: 47	12MF PER: -	
		-부품 비용, 자본 비용 증가 추세 속 높은 수주 잔고 무색 -인력, 전기 쇼티지에도 매출 전환 성공 여부 주목		

주: 시가총액 십억달러

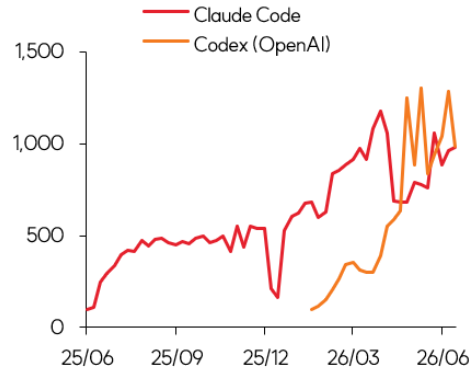
Update Chart

OpenRouter 토큰 사용량



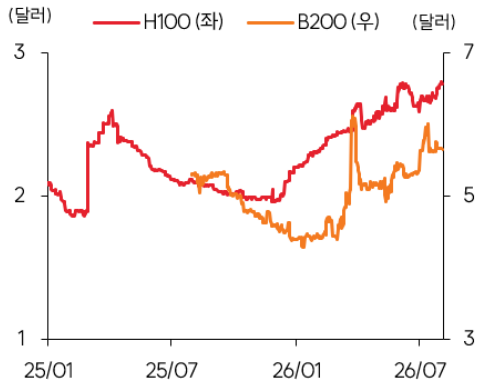
자료: OpenRouter, SK 증권

주요 Coding Agent 활용 강도 (백만회 달성일 = 100)



자료: NPMjs, SK 증권 / 주: 업데이트 및 다운로드 수 기준. Codex 5월 수치 일부 조정

GPU 렌탈 가격 추이 (SiliconData)



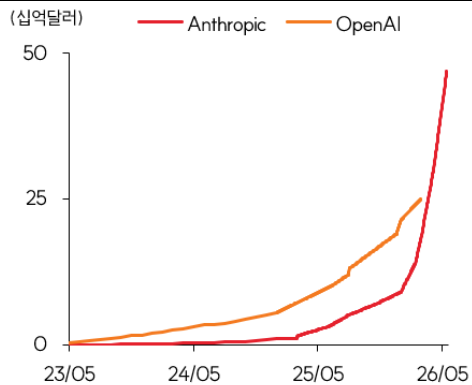
자료: Bloomberg, SK 증권

LLM 지출



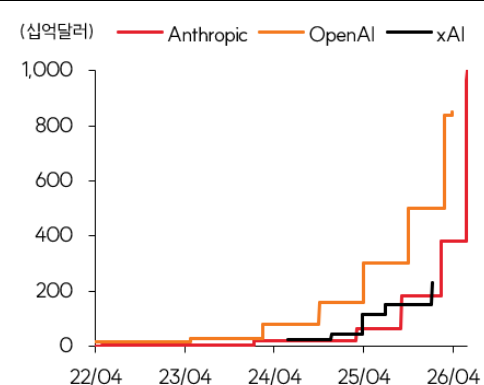
자료: SimilarWeb, SK 증권

주요 AI Labs ARR 추이



자료: EpochAI, SK 증권

주요 AI Labs 기업가치 추이



자료: EpochAI, SK 증권

자료: AMD, SK 증권

현재 확보된 물량 14GW, 고객사는 누구일까?

OpenAI 6GW 멀티제너레이션 계약(주식 워런트 포함)

약 3개월 전 Helios 랙 수령해 GPT-Plus 워크로드 이미 구동 중, MI300 > MI355 > MI450 순차 확대

Meta 최대 6GW 커스텀 GPU 계약(2월 발표), 첫 1GW를 Helios 랙으로 2H26 내 배치

Anthropic 최대 2GW MI450 시리즈, 첫 1GW는 1H27 시작, AMD가 최대 \$5B 지분 투자(마일스톤 연동)

Microsoft는 Azure 프론티어 모델 추론용 대규모 배치, 용량·금액 미공개

Oracle은 MI450 GPU 50,000 장을 3Q26부터 OCI 배치

AMD 언급 기준 Helios 관련 수요 12GW(OpenAI+Meta), Anthropic 포함 시 공개 계약 합계 14GW

AMD 표현으로 "향후 1년간 수 GW 규모 Helios 인프라 배치 예상"

SemiAnalysis는 MI455X GPU capex를 GW당 \$18.7B로 산정

생산 전망이 드러났다면 내년도 매출이 수백억달러 이상 증가 가능한 아이템

VeraRubin 대비 우위, 어디까지 믿을 수 있나?

랙간 단일 기준의 스펙 비교는 아직 부재한 상황

AMD 추정치 기준 FP4 컴퓨트 +15%, HBM 용량 +50%(31TB vs Rubin GPU당 288GB), HBM 대역폭 +6%, 스케일아웃 대역폭 +50%, 톨당 달러 최대 +30%

비교 기준 불일치 존재, AMD는 자사 피크 데이터타입 vs NVIDIA 공표 NVFP4 dense 비교하나 Nvidia는 sparse로 마케팅

확실한 것은 AMD가 CPU, GPU 모두 최초 2nm 양산에 성공하고 있다는 점. MI455, Venice CPU 모두 N2 고정 활용

CPU는 96코어 Venice가 NVIDIA Vera 대비 코어당 성능 20% 우위 주장, NVIDIA 공표 벤치마크 데이터 활용

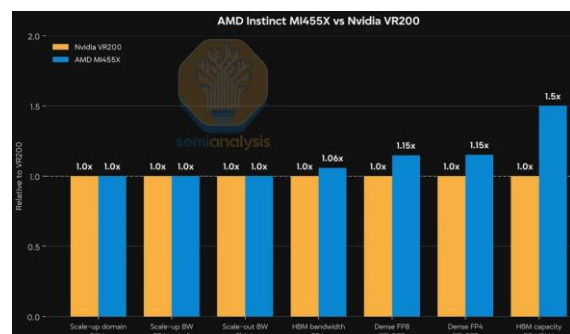
다만 Scale up, out에 필수적인 Switch 칩을 엔비디아는 자체칩을 활용하는 것에 비해 AMD는 Broadcom 외주

시스템 단위 성능은 자체 실리콘 비중이 높고 랙 단위 생산 시작이 오래된 엔비디아가 아직 높을 것으로 판단

AMD Helios vs Nvidia VR200

AMD HELIOS	METRIC	NVIDIA VERA RUBIN NVL72
72 MI455X GPUs	GPUS PER SYSTEM	72 Rubin CPUs
18 Epyc Venice CPUs	CPU	36 Vera CPUs
Up to 2.9 exaflops	FP4 INFERENCE COMPUTE	Up to 3.6 exaflops
Up to 1.4 exaflops	FP8 OR FP6 COMPUTE	Up to 1.26 exaflops
31TB HBM4	GPU MEMORY	20.7TB HBM4
Up to 260TB per second	SCALE-UP BANDWIDTH	260TB per second
ROCm	MAIN SOFTWARE STACK	CUDA
Open rack standards	ARCHITECTURE MODEL	Integrated Nvidia platform

AMD Helios vs Nvidia VR200 (상대 비교)



진짜 변수는 스펙이 아니라 조립

SemiAnalysis는 공급망 체크와 엔지니어링 분석 결과 Helios가 현재 진행형으로 느린 랙 양산 램프를 겪고 있고 백플레인 신뢰성 문제에 부딪히고 있다고 주장

AMD는 올해가 첫 랙 단위 양산, NVIDIA는 이미 랙 3세대를 바라보는 상황, 효율화 격차 존재

스케일업 스위치를 자체 제작하지 않고 Broadcom에 전적으로 의존, 로드맵이 스위치 파트너 실행에 종속

NVIDIA가 Vera Rubin부터 Tray cableless 구조 미적용, AMD는 반영 실패

구조적 열위가 아니라 한 세대 늦게 같은 곡선을 타는 타이밍 리스크로 해석 가능

지연은 매출을 소멸시키는 것이 아니라 지연시키는 것이나, 현 밸류에이션에서는 한 분기 지연도 멀티플에 반영될 가능성

관전 포인트는 랙 출하 대수, ODM(Sanmina·Wiwynn) 물량 코멘트, 그리고 소프트웨어 측 CI 진척도

Helios에서도 보이는 Prefill-Decode 분리 트렌드

AMD는 Helios Rack 출시와 함께 Cerebras와 technical partnership 발표, 지분 투자·금액 조건 미공개

구조: Helios가 prefill 담당, Cerebras WSE가 decode. 단일 disaggregated 추론 워크플로우로 통합

비유하면 Helios가 문제를 읽고 Cerebras가 답을 쓰는 분업

NVIDIA-Groq과 동일한 발상 — GPU가 prefill, NPU가 decode

Intel도 Computex 2026에서 Xeon 오케스트레이션 + NVIDIA Blackwell prefill + SambaNova RDU decode 조합 랙 시연

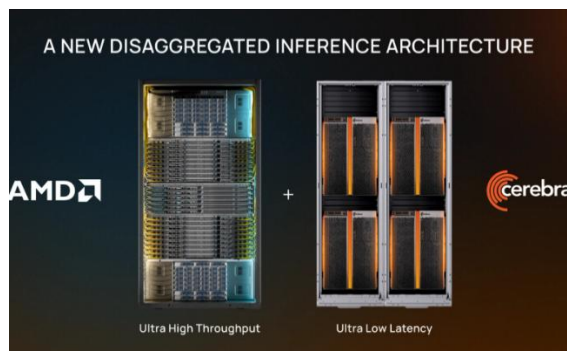
즉, 1~3위 벤더가 모두 같은 아키텍처로 이동 중이며 disaggregated inference가 니치가 아닌 표준으로 굳어지는 국면

Cerebras가 자사 데이터센터에 Helios 도입 > 조인트 솔루션은 2H26 중 Cerebras Cloud 통해 최초 제공

서버 구매자가 자체 AMD 기반 시스템에 WSE를 구성하는 옵션도 언급

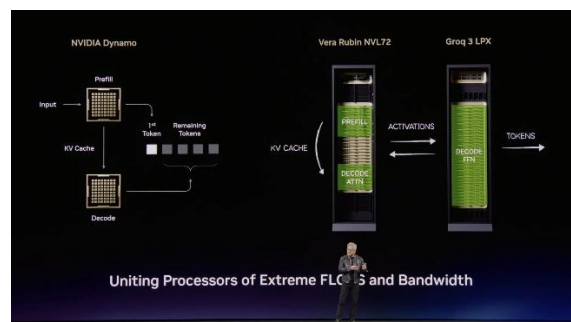
발표 당일 CBRS 주가 4% 이상 상승

AMD Cerebras 연산 구조



자료: 언론 종합, SK 증권

GTC2026에서 발표한 LPU Decoding 구조



자료: Nvidia, SK 증권

다시 부상하는 Nvidia 순환 출자 논란

작년 불거졌던 Nvidia 순환 출자 논란 재점화

(7/26) WSJ, Nvidia가 OpenAI의 오하이오 데이터센터 리스에 약 \$250B 규모 금융 보증 제공을 협상 중이라고 보도
보증 대상은 리스료와 건설채(construction debt), Nvidia 자사 칩은 제외
칩 파이낸싱은 별도 협상 중이며 최대 \$350B 규모로 거론
리스 보증 \$250B + 칩 \$350B 합산 시 총 프로젝트 비용 \$500B 초과, 현재까지 발표된 데이터센터 프로젝트 중 최대
총 10GW 설계, 1단계 약 800MW는 2028년 가동 목표, 64만 가구 전력 소비량 수준
개발·소유는 소프트뱅크 에너지 자회사 SBEnergy, OpenAI는 테넌트
OpenAI는 투자등급(investment-grade) 신용등급 부재, 대주단이 OpenAI 신용만으로는 대출 불가 판단
Nvidia 보증 시 SBEnergy가 테넌트가 아닌 Nvidia 신용으로 차입 가능, 조달 금리·조건 개선
(7/27) Nvidia 주가는 -5%로 마감, Nvidia CDS(신용부도스왑) 스프레드 확대가 월가 화두로 부상

Nvidia의 순환 출자, 유형이 1개가 아니다?

1. Equity(지분 투자): Nvidia가 고객 지분을 취득, 고객은 그 자금으로 Nvidia 시스템 구매
26년 1~4월 AI 에쿼티 투자 \$40B 이상 확약, 직전 회계연도 연간 \$17.5B를 4개월 만에 초과
기존부터 CoreWeave Class A 지분 약 7% 보유
(26년 2월) OpenAI \$30B, Anthropic 최대 \$10B 확약
(26년 3월) Nebius \$2B, Thinking Machines \$2B 라운드 참여
(26년 5월) IREN과 최대 5GW 파트너십

2. Backstop(미판매 캐파 매입 의무): 고객이 캐파를 못 팔면 Nvidia가 매입
CoreWeave 오더폼 초기 가치 \$6.3B, 2032/4/13까지 잔여 미판매 캐파 매입 의무
발동 시 Nvidia는 자기가 판 GPU의 서비스를 되사오는 구조, 실질적으로 매출 취소에 가까움
손실 상한은 계약 규모로 한정되나 발동 시점이 업황 하강기와 일치
TPU와 관련하여 Broadcom이 유사한 구조를 Anthropic과 체결

3. Revenue Share(매출 공유형 벤더 파이낸싱): 선결제 없이 하드웨어 공급, 클라우드 매출을 분배
(26년 7월) Nvidia가 CFO Colette Kress 공동 명의 블로그로 DSX AI 팩토리 파트너십 모델 공식화
1호 파트너 Sharon AI(NASDAQ: SHAZ): 6년 계약, GB300 최대 40,000장, 호주 최대 72MW

2호 파트너 Firmus Technologies: 인도네시아 바탐 360MW 캠퍼스, 최대 170,000 장, 합산 약 210,000 장
 배포된 GPU가 유휴 상태일 경우 Nvidia가 사전 확정 가격으로 미사용 캐파를 임차 또는 매입
 수요 측 참여사로 Baseten, Fireworks AI, Together AI 명시
 항공기 리스나 GE Capital 형 벤더 파이낸싱과 동일한 구조, 논란 이전에 비즈니스 모델 전환으로 해석 가능

4. Guarantee(제3자 채무 보증): 고객의 리스·건설채에 Nvidia 신용을 대어
 오하이오 \$250B가 여기 해당, 칩 파이낸싱 \$350B 논의도 같은 범주
 앞의 세 층위와 달리 손실 상한이 계약 규모 전체이며 자기 자본 대비 레버리지가 걸림
 발동 조건은 테넌트의 리스로 미지급, 즉 OpenAI 신용에 Nvidia 재무제표가 직결
 즉, 젤슨이 올해 초 "OpenAI \$30B가 마지막 투자 에퀴티일 것"이라고 한 것은 수단의 교체

4개의 유형 중 시장이 실제로 걱정해야 할 것은 4번이라고 판단. Nvidia 재무제표를 사실상 OpenAI와 밀과 연결하는 형태
 1번인 Equity는 손실 상한 명확, 2) Backstop은 계약 규모로 한정, 3) Revenue Share는 벤더 파이낸싱 관례에 가까움

Ohio DataCenter 부지



자료: 언론 자료, SK 증권

Nvidia 순환 출자 관련 보도

Nvidia's \$750 Billion in Deals Reignite Circular AI Fears



Jensen Huang departs after speaking at the Nvidia GTC conference in San Jose. Photographer: David Paul Morris/Bloomberg

자료: 언론 자료, SK 증권

Compliance Notice

작성자(박제민)는 본 조사분석자료에 게재된 내용들이 본인의 의견을 정확하게 반영하고 있으며, 외부의 부당한 압력이나 간섭없이 신의성실하게 작성되었음을 확인합니다.
 본 보고서는 기관투자자 또는 제 3자에게 사전 제공된 사실이 없습니다.
 당사는 자료공표일 현재 해당기업과 관련하여 특별한 이해 관계가 없습니다.