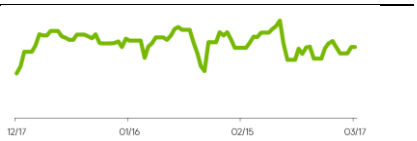
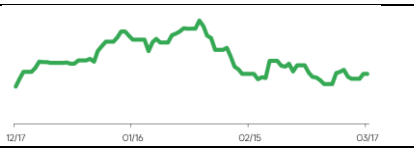
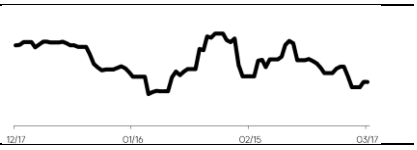
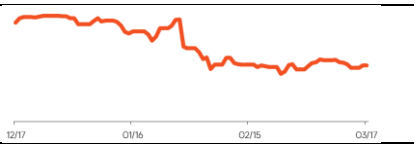
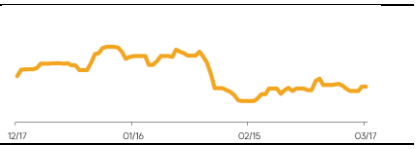
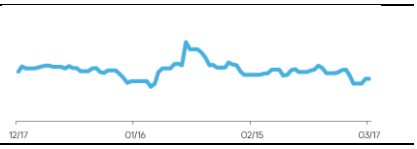
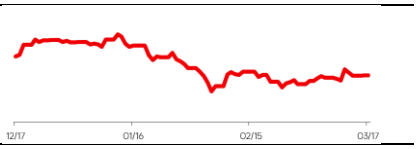
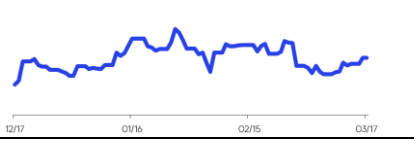
2026-03-18

SK securities

SK증권 해외주식/AI. 박제민
jeminwa@sks.co.kr

	엔비디아	시가총액: 4,452	12MF PER: 22	
		-강한 수요 + Vera Rubin 생산 진행으로 펀더멘탈 견고 -cuDF, cuVF, Agent 에따른 엔터프라이즈 수요 확장 주목		
	알파벳	시가총액: 3,690	12MF PER: 25	
		-AI Full stack 기업, 텍스트 및 이미지 AI 모델 1위 사업자 -AI 로 인한 Search 고도화, Shopping Agent 주목		
	애플	시가총액: 3,712	12MF PER: 29	
		-메모리 가격 상승은 점유율 확보 기회, 엔트리 라인업 확대 - Gemini 결합, 출시 지연으로 AI 출시 기대 국면		
	마이크로소프트	시가총액: 2,970	12MF PER: 23	
		-AI PaaS 1위 사업자, OpenAI 엔진 활용 해자 -Infra 해자보다 고객 접근성, M365 Data 해자에 베팅		
	아마존	시가총액: 2,273	12MF PER: 23	
		-CPU DC 1위 사업자, GPU DC 공격적 증설 진행 중 -OpenAI 통한 Agent 용 DC, Shopping 결합 주목		
	메타 플랫폼스	시가총액: 1,588	12MF PER: 19	
		-AI 도입으로 광고 P,Q 개선. Llama 4 실패 후 투자 확대 -1H26 차세대 AI 모델 발표, AI 글래스 판매량 주목		
	오라클	시가총액: 449	12MF PER: 21	
		-Stargate IaaS 담당 사업자, 수주 절반 이상 OpenAI -OpenAI 사업성, 추가 부채 발행 여부 주목		
	코어위브	시가총액: 46	12MF PER: -	
		-1위 네오클라우드 사업자, 수주 절반 이상 OpenAI -금리 불안정성, 스프레드 상승으로 조달 금리 주목 필요		

주: 시가총액 십억달러

- **지난 주 빅테크 주요 주가 변동은?**

  메타 (-4.9%): 차세대 모델(Avocado) 출시 지연 소식에 하락. 대규모 감원 소식에 잠깐 반등 있었으나 빠르게 되돌림

  오라클 (-5.2%): 실적발표 이후 급등을 지키지 못하며 하락. 아마존 대규모 회사채 발행 성공으로 증권사들 빅테크 회사채 발행 전망 상향. 오라클이 올해 계획 중인 $45~50B 신규 발행 및 리파이낸싱 조달 금리 부담이 부각된 것으로 해석

- **GTC2026, VR 신규 라인업 공개, 변화점은?**

  칩 7 개, Tray 6 개, Rack 5 개, SuperPOD BluePrint: Nvidia 가 권하는 AI factory 참조 설계도 [차트 1] [차트 2]

  SuperPOD blueprint: 총 40 개의 Rack 으로 이루어진 AI Factory (NVL Rack 16 개 + CPU rack 2 개 + BF-4 STX rack 2 개 + SPX rack 10 개 + LPX rack 10 개)

  SuperPod 내 Rack 위치 및 탑재 Tray 종류 [Tray 개수]

  1. NVL72 rack (중앙부 양옆 8 개씩 16 개): 메인 GPU 두뇌 > Vera Rubin Compute Tray [18], NVLink Switch Tray [9]

  2. Vera CPU rack (가운데쪽 SPX rack 양옆 2 개): agent 실행/샌드박스/오케스트레이션 > Vera Compute Tray [32]

  3. BF4 STX rack (양 끝에서 2 번째 2 개): KV cache/context memory 저장소 > STX server Tray [미공개]

  4. Spectrum-6 SPX rack (양끝이랑 가운데 총 10 개): POD 전체 연결망 > Spectrum X CPO Switch [미공개]

  5. LPX rack (양끝에서 3 번째 부터 5 개씩 총 10 개): premium inference 용 초저지연 보조 두뇌 > Groq 3 LPX Compute Tray [32]

  Hopper 에서 Blackwell 이동은 모듈 수준에서 Rack 수준의 인프라로의 이동. 이번 VR 은 AI factory 단위의 라인업 공개

  Agent 로 AI 활용이 복잡해지면서 인프라도 복잡해지는 중. 따라서 단순 GPU FLOPS 로 비교 불가

  엔비디아가 제시한 곡선은 TPS/user 대비 TPS/MW. TPS 는 tokens per second. 즉 토큰을 많이주는 서비스 대비 전성비

  Blackwell 은 Hopper 대비 중간 수준 TPS/user 서비스(reasoning 모델)에서 35 배 수준의 도약

  VR NVL72 rack 은 Blackwell 대비 중간 수준 서비스에서 3 배 수준의 성능 도약

  LPX rack 활용 시 고성능 서비스(높은 TPS/user 기준) 서비스에서 35 배 수준 성능 [차트 3]

  해자가 강한 기존 제품(NVL Rack)에 대한 수요는 확실, 그 외 CPU, BF, LPU rack 에 대한 수요는 검증 필요

  GTC Keynote 에서 CY27 까지 $1T 규모 수요 전망(컨센서스 +20%)에도 부진

  1) 신제품을 활용한 Inference 해자에 대한 의구심 2) 빅테크 Capex 기울기 피크 아웃 등에 대한 우려가 있는 것으로 판단

- **초저지연 칩을 아마존에서도? 엔비디아와의 차이점은?**

  Amazon 은 AWS 에서 Trainium 3 + Cerebras CS-3 조합으로 새로운 inference 서비스를 2026 년 하반기 출시 계획

  Trainium 3 가 prefill, Cerebras 가 answer generation(decode) 을 맡는 구조

  Decode 가 병목이 되고 있는 Agentic workflow 에 대해 Tranium 의 부족 지점을 보충

  2026 년 1 월 OpenAI 는 Cerebras 와 750MW 규모 고속 Compute 파트너십

  Cerebras 를 활용한 GPT-5.3-Codex-Spark 제품이 이미 'real-time coding' 서비스 진행 중 [차트 4]

  초고속 추론 영역에서 엔비디아 의존도를 낮추기 위한 판단

  OpenAI 의 stateful 서버 공급을 위한 움직임으로도 보이며, AWS Bedrock 서비스 차별화에도 활용될 것

  엔비디아의 LPU Rack 과 초저지연 추론(latency-sensitive inference)을 겨냥한다는 점에서 유사

  보도에 따르면 Cerebras-Tranium 은 단순 decode-prefill 만 구분

  LPU Rack 은 disaggregated inference 컨셉. decode 내에서도 attention 은 GPU, FFN, MOE expert 는 LPU 로 구분

  Dynamo SW 활용한 workflow 분류와 배분, ICMS 활용 KV Cache Uploading, Scale up+ out 대역폭 등이 중요

  AWS 도 곧바로 'Disaggregated inference on AWS' 공개. SW 분산 방식으로 HW 해자는 엔비디아 유효 [차트 5]

- **Stargate 부지 확장 중단 뉴스가 오히려 강한 DC 수요를 보여주는 이유는?**

  Oracle, OpenAI 는 텍사스 Abilene 내 Stargate 부지 중 일부 확장안을 중단. 이유는 OpenAI 의 바뀐 요구 조건

  중단된 확장 부지 조건은 600MW 규모, Abilene 지역 1 차 목표인 4.5GW 확장에는 영향이 없는 결정

  확장 계획 중단으로 개발사 Crusoe 가 해당 공단 수요자 물색, Meta 가 임창 방안 검토 및 Nvidia 중재 보도

  추후 Microsoft 도 같은 부지에 대해 검토 중이라는 보도 [차트 6]

  Abilene 8 개 부지 중 2 개는 이미 가동 상태, 신규 확장 부지도 전력 확보가 완료된 상황일 가능성이 높음

  중단 이후 수요자가 바로 나타난다는 점에서 전력 확보 capacity 에 대한 선호가 매우 높음을 보여주는 사례

  OpenAI 는 최근 Amazon 의 투자로 AWS 서버 활용량이 증가(600MW 전후 -> 2GW)하면서 부지 수요 감소 추정

  OpenAI 는 현재 대략 1.9GW 전력을 확보한 것으로 판단, 2026 년 Vera Rubin ~1GW, AWS Tranium 2GW 추가 공급

  2027 년부터는 Broadcom 과 ASIC 이 추가될 예정

Anthropic, 모델사에서 멈추지 않고 컨설팅으로?

Anthropic 은 Blackstone, Hellman & Friedman 등 사모펀드들과 Palantir-like JV 설립을 논의 중

소프트웨어만 파는 게 아니라 컨설팅, 시스템 통합, 업무 자동화 설계, 현장 배치도 판매하는 비즈니스 모델

Blackstone 만 해도 포트폴리오 기업이 250 개 이상. 대형 영업 채널 확보를 통해 JV 가치 극대화

OpenAI 는 내부 FDE(Forward Deployed Engineering)을 앞세워 적극적으로 고객 도입 추진 중

최근 McKinsey·BCG·Accenture 등과 대기업에 AI 도입을 위한 'Frontier Alliance' 구축 [차트 7]

B2B AI 가 소프트웨어 판매에서 확장해 도입 대행 + 맞춤 구축 + 장기 운영으로 확장 중

기존 SaaS 수준의 서비스(GPT Codex, Claude Cowork) 등은 직원을 도와주는 영역이 주요

컨설팅을 가미한 Palantir-like 한 서비스는 워크로드를 재구축하여 효율화 및 자동화

한번 시스템이 구축되고 Fine-tuning 이 이루어지면 재무적, 문화적 전환 비용이 커지는 구조로 판단

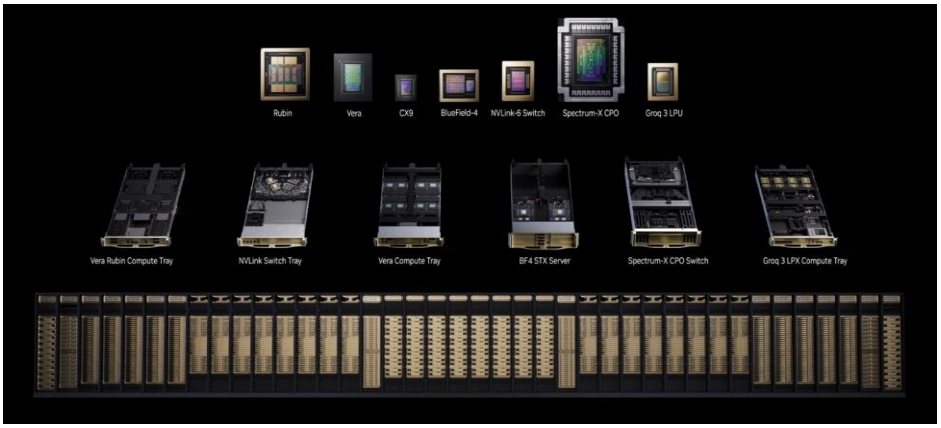
선제적으로 팀을 구축하고 주요 컨설팅사와 협업을시도한 OpenAI 가 아직까지는 해자가 더 클 수 있다고 판단

박제민의 Q&A.I

2026-03-18

SK securities
SK증권 해외주식/AI. 박제민
jeminwa@sks.co.kr

[차트 1] Vera Rubin 신제품 라인업 정리



자료: Nvidia, SK 증권

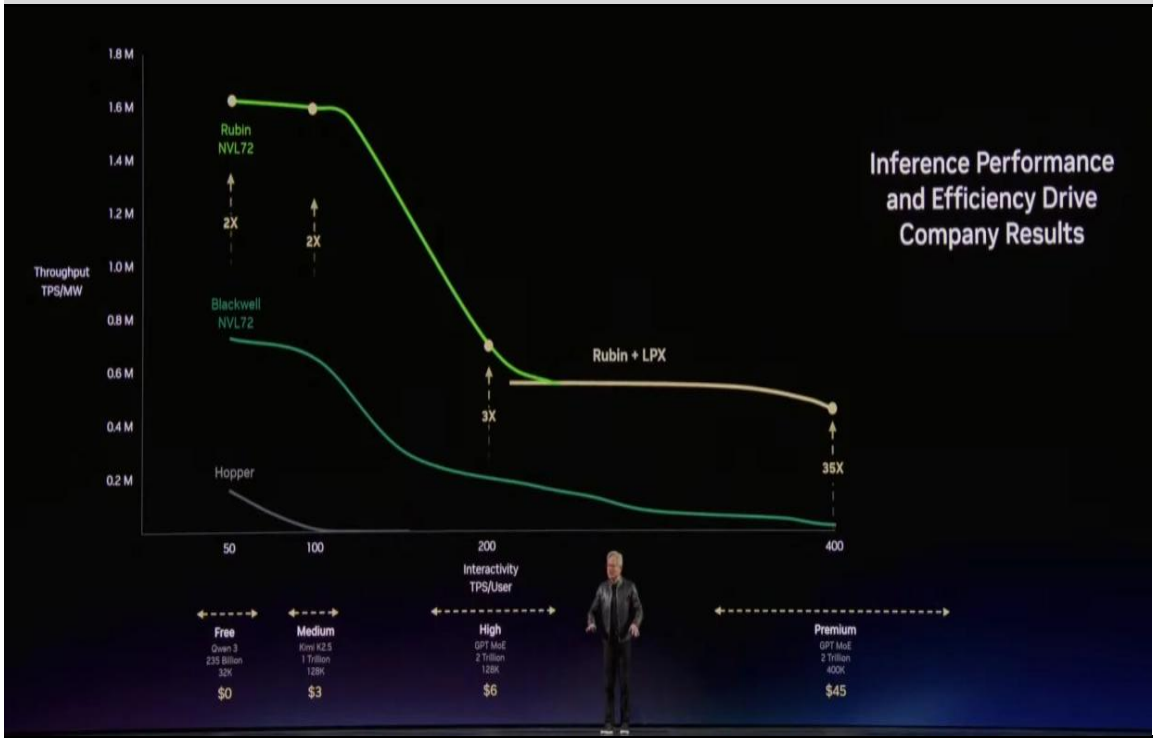
[차트 2] Blackwell 제품 라인업 정리

자료: Nvidia, SK 증권

박제민의 Q&A.I

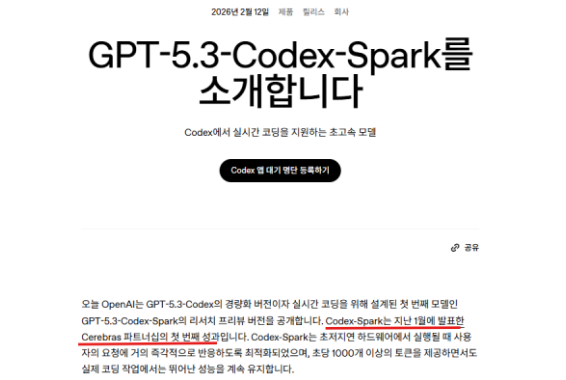
SK securities

[차트 3] Nvidia 제품 세대별 전성비 정리



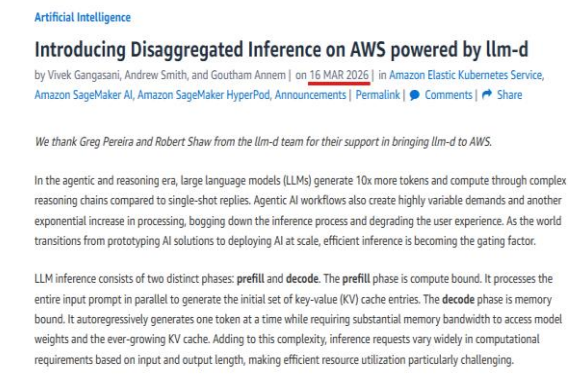
자료: Nvidia, SK 증권

[차트 4] GPT-5.3-Codex-Spark



자료: OpenAI, SK 증권

[차트 5] Disaggregated inference on AWS'

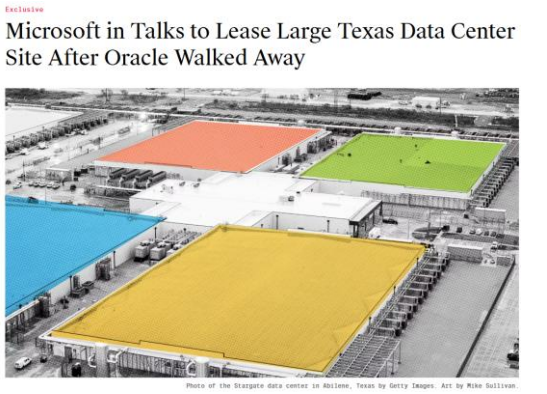


자료: AWS, SK 증권

박제민의 Q&A.I

SK securities

[차트 6] Microsoft 부지 인수 검토

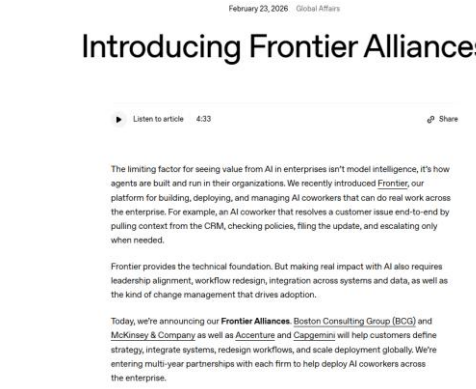

Exclusive

Microsoft in Talks to Lease Large Texas Data Center Site After Oracle Walked Away

Photo of the Stargate data center in Abilene, Texas by Getty Images. Art by Mike Sullivan.

자료: Theinformation, SK 증권

[차트 7] OpenAI Frontier Alliance


February 23, 2026 Global Affairs

Introducing Frontier Alliances

Listen to article 4:33 Share

The limiting factor for seeing value from AI in enterprises isn't model intelligence, it's how agents are built and run in their organizations. We recently introduced Frontier, our platform for building, deploying, and managing AI coworkers that can do real work across the enterprise. For example, an AI coworker that resolves a customer issue end-to-end by pulling context from the CRM, checking policies, filing the update, and escalating only when needed.

Frontier provides the technical foundation. But making real impact with AI also requires leadership alignment, workflow redesign, integration across systems and data, as well as the kind of change management that drives adoption.

Today, we're announcing our **Frontier Alliances**. Boston Consulting Group (BCG) and McKinsey & Company as well as Accenture and Capgemini will help customers define strategy, integrate systems, redesign workflows, and scale deployment globally. We're entering multi-year partnerships with each firm to help deploy AI coworkers across the enterprise.

자료: OpenAI, SK 증권

Compliance Notice

작성자(박제민)는 본 조사분석자료에 게재된 내용들이 본인의 의견을 정확하게 반영하고 있으며, 외부의 부당한 압력이나 간섭없이 신의성실하게 작성되었음을 확인합니다.

본 보고서에 언급된 종목의 경우 당사 조사분석담당자는 본인의 담당종목을 보유하고 있지 않습니다.

본 보고서는 기관투자가 또는 제 3 자에게 사전 제공된 사실이 없습니다.

당사는 자료공표일 현재 해당기업과 관련하여 특별한 이해 관계가 없습니다.

종목별 투자의견은 다음과 같습니다.

투자판단 3 단계(6 개월기준) 15%이상 -> 매수 / -15%~15% -> 중립 / -15%미만 -> 매도