

# 딥시크 변동성 대응전략

## 단기 기회와 중기 변화 요인

### 글로벌 투자전략-선진국

Analyst 황수욱  
02. 6454-4896  
soowook.hwang@meritz.co.kr

RA 박보경  
02. 6454-4861  
pbg811@meritz.co.kr

### 딥시크 모멘트: R1 모델 공개와 충격

1월 20일 중국의 스타트업 딥시크의 AI 모델 R1 출시. 테스트 타임 컴퓨팅, 강화 학습이라는 핵심 기술을 기반으로 미국의 수출 규제에 상대적으로 열악한 AI 인프라 환경에서도 OpenAI의 모델 o1에 필적한 성능을 구사함. 그럼에도 API 비용은 20분의 1 수준. R1이 공개되며 이전 모델이었던 V3의 훈련 비용이 유사한 성능을 구사한 GPT-4 대비 20분의 1이었다는 점도 재부각됨

### 1월 27일 AI 인프라 기업 중심의 주가 하락

1월 27일 미국 주식시장에서는 '막대한 AI 인프라 투자가 필요할까?'라는 의구심에서 관련 종목 주가 급락 발생. 1월 27일 당일 S&P500 수익률 하위 10종목은 모두 AI 인프라(AI 반도체, 전력 인프라) 기업이었음. 반면, AI 인프라 연관성이 낮거나, AI 훈련 비용이 낮아짐으로 AI 확산 가속의 수혜를 볼 수 있는 기업들(AI S/W, 자율주행, 온디바이스)은 오히려 주가가 상승하기도 함

### 1월 27일 이후 회복하는 주가 흐름의 배경과 쟁점

1월 27일 급락했던 AI 인프라 관련주는 대부분 반등에 성공. 이후 주가 반응의 쟁점은 크게 2가지. 1) 중국 모델에 대한 신빙성이 있는지, 2) 중국이 보여준 비용절감 기술의 효율성은 Scaling law(더 많은 AI 하드웨어 투입→더 나은 AI 소프트웨어 성능)을 끝을 의미하는지. V3 모델의 값싼 훈련 비용이 과장되었다는 주장, R1 모델의 미국 모델 distillation(증류) 의혹 제기되며 딥시크에 대한 첫인상의 충격이 약해지기 시작. 비용절감 기술의 효용은 인정이 되나, AI 인프라 투자는 늘어야 한다는 의견이 주류를 이루며 주가 회복 주도

### 딥시크 모멘트의 시사점

딥시크 R1은 OpenAI의 o1과 함께 테스트 타임 컴퓨팅(스케일링)이 거의 처음으로 적용되는 AI 모델. (Test) Time Scaling Law란 1) 추론 단계에서 2) 더 많은 계산 시간을 투입하면 더 나은 결과를 도출할 수 있다는 법칙. Time Scaling Law 하에서 AI 훈련 비용은 줄일 수 있지만, 추론에 필요한 인프라와 비용이 더 필요하게 됨. 딥시크 모델은 Test Time Scaling의 효용성이 무료 버전인 R1을 통해 확산된 계기로, 기술 진보 및 확산의 매개로 이해

### 투자전략: 단기 저가 매수 전략, 중기 추론용 반도체 니즈 가속화 가능성 고려

AI 모델 훈련 최적화를 인정해도 추론용 AI 인프라 확장은 필요. 따라서 딥시크 충격에 따른 AI 인프라 기업 주가 하락은 과도, 단기 저가 매수 전략 유효하다는 판단. 중기적으로는 AI 추론용 인프라에 대한 니즈 앞당길 요인, 관련 종목에 대한 눈높이 상향 가능성 고려

### 딥시크 R1 모델 공개와 충격

#### 1/20 중국 딥시크 R1 공개와 충격

1월 20일 중국의 AI 스타트업 딥시크가 자사의 최신 AI 모델 R1을 공개했다. 일주일 후인 1월 27일 미국 증시는 R1의 성능에 놀랐다. 주요 성능 테스트(벤치마크)에서 가장 우수한 수준의 성능을 보이는 OpenAI의 o1 모델을 뛰어넘거나 거의 비슷한 성능을 보여주었기 때문이다.

#### 더 적은 컴퓨팅 자원을 기술로 극복, 미국과 유사한 수준의 모델 개발 성공

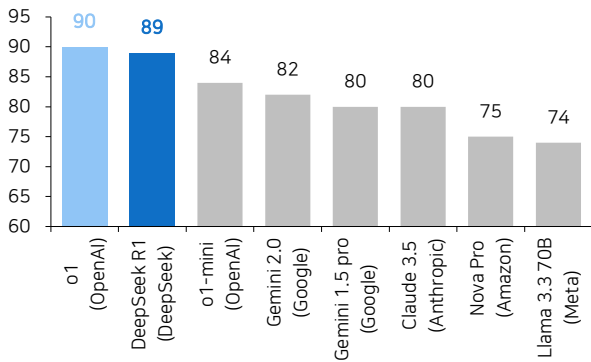
시장이 충격에 빠진 이유는 미국의 AI칩 수출 규제에도 더 적은 비용으로 유사한 성능을 낼 수 있는 효율성 때문이다. OpenAI o1의 20분의 1 수준의 API 비용밖에 소요되지 않고, 미국보다 열악한 AI 하드웨어 환경에서 적은 인풋으로 성과를 낸 점이 놀랍다는 것이다. 리소스의 열위를 극복하게 해준 것은 결국 기술인데, 강화 학습(RL)과 테스트 타임 컴퓨팅(Test Time Computing) 기술로 막대한 자본이 들어간 미국의 AI 모델을 따라잡았다는 데에서 중국의 기술력에 놀란 것이다.

#### 핵심 기술 1) 테스트 타임 컴퓨팅, 2) 강화 학습

여기서 테스트 타임 컴퓨팅이란 문제 해결과정을 분할하고 중간 결과를 점검해 답을 도출하는 방식이고, 강화 학습은 테스트 타임 컴퓨팅 과정에 보상을 주면서 AI가 스스로 정답에 도달하게 할 수 있게 하는 기술이다.

그림1 Artificial Analysis의 AI 모델 Quality Index

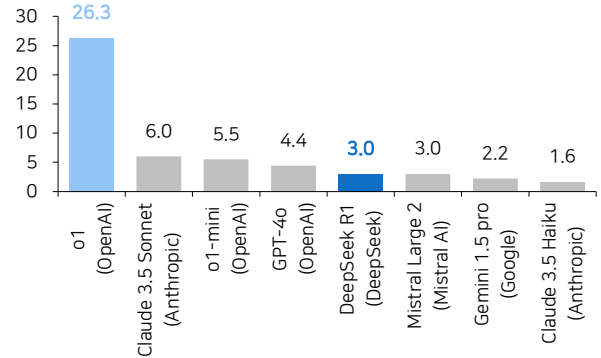
(점수, 높을수록 좋음)



자료: Artificial Analysis, 메리츠증권 리서치센터

그림2 주요 AI 모델별 API 가격

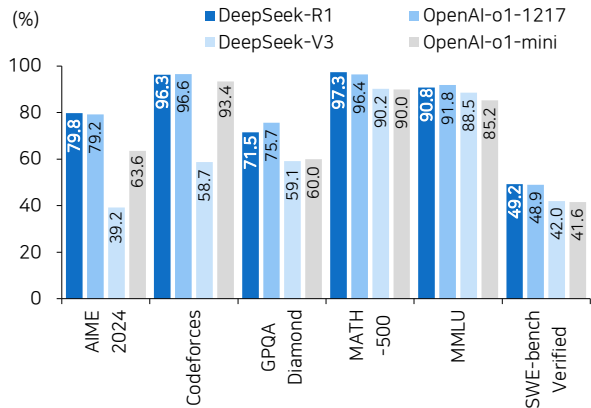
(1M 토큰당 달러 가격, 낮을수록 좋음)



자료: Artificial Analysis, 메리츠증권 리서치센터

그림3 주요 벤치마크에서 모델별 성능 비교

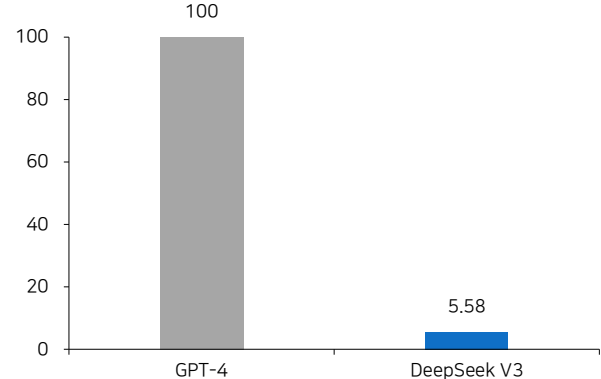
(%)



주: AIME(고등학교 수학 경시), Codeforces(프로그래밍), GPQA(고등 수준 지식 평가), MMLU(언어이해 능력), SWE(소프트웨어 문제해결 능력) 평가  
자료: DeepSeek, 메리츠증권 리서치센터

그림4 다시 화제가 된 DeepSeek V3 모델 훈련 비용

(백만달러)



자료: DeepSeek, Wired, 메리츠증권 리서치센터

### 1월 27일 이후 미국 증시 주가 반응

중국 모델을 본 시장의 처음 생각:  
AI 인프라 투자가 많이 필요할까?

중국의 AI 모델을 보고 미국 시장이 가장 첫번째로 반응한 것은 막대한 'AI 인프라 투자가 필요할까?'였다. AI 인프라 기업을 중심으로 급락하는 종목이 발생했다. S&P500 지수는 전일대비 1.5% 하락하는 데 그쳤다. S&P500 종목 중 1월 27일 당일 351 종목의 주가가 상승했고, 152종목이 하락하며 시장 전체적으로는 상승 종목이 더 많았다. 그러나 필라델피아 반도체 지수가 9.1%나 하락하는 등 AI 반도체 기업 중심으로 주가 낙폭이 컸다.

1/27 AI 인프라 주식을 중심으로  
주가 하락 발생

1월 27일 S&P500 수익률 하위 10 종목으로는 비스트라, 아리스타 네트워크, GE 버노바, 컨스텔레이션 에너지, 판타 서비스, 브로드컴, 엔비디아, 이튼, 오라클, NRG 에너지다. 이들의 공통점은 2023년 이후 AI 인프라 사이클에서 시장을 주도했던 종목들이라는 점이다.

VST, CEG, GEV 등 낙폭 가장 컸던  
종목들은 연초 이후 상승 폭이  
가장 컸던 종목들

당일 가장 낙폭이 컸던 비스트라, 컨스텔레이션 에너지, GE 버노바 등은 연초 이후 주가 상승률이 가장 높았던 종목이다. 트럼프 행정부의 에너지 산업을 중심으로 하는 AI 인프라 산업 부양 의지가 확인되며 주가 상승 추세가 가속화되었다. 이 종목들은 작년말 대비 채 한 달이 되지 않은 기간동안 각각 +55.3%(컨스텔레이션 에너지), +38.6%(비스트라), +27.8%(GE 버노바)나 상승했던 종목들인데, 당일 주가 하락으로 연초 수준까지 하락했다.

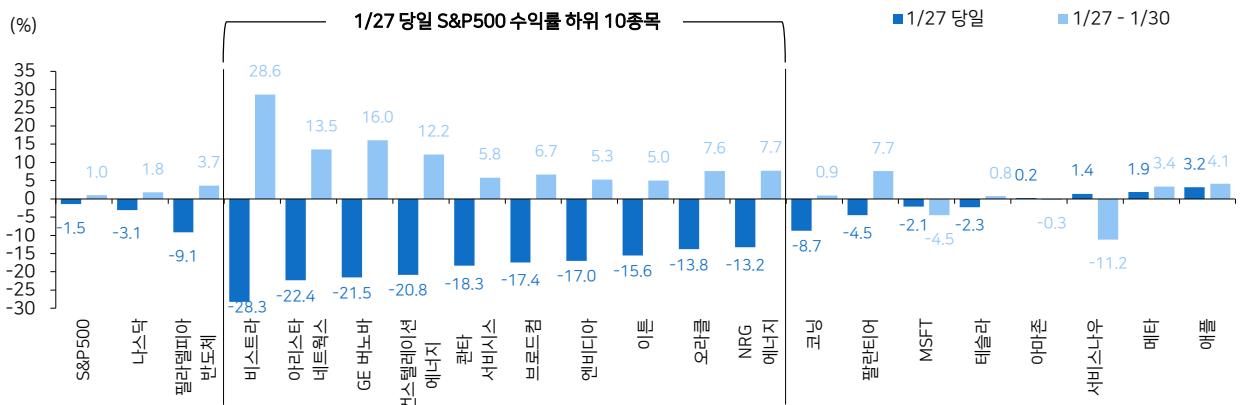
엔비디아, 브로드컴 AI 반도체주는  
연초수준 대비 (-)까지 하락

반면, 같은 기간동안 주가 상승이 크게 두드러지지 않았음에도 엔비디아, 브로드컴과 같은 AI 반도체 기업도 1월 27일 주가 낙폭이 컸다. 브로드컴은 12월 주가 상승이 가팔랐던 것까지 고려하면 '많이 올랐던 종목이 많이 빠졌다'고 해석될 수 있지만, 엔비디아는 상대적으로 최근 주가 모멘텀이 강하지 않았음에도 이날 주가가 17%나 빠지면서 상대적으로 가장 컸던 것으로 풀이된다.

AI usage 관련(자율주행,  
온디바이스 AI, AI S/W) 기업 주가  
상대적으로 양호

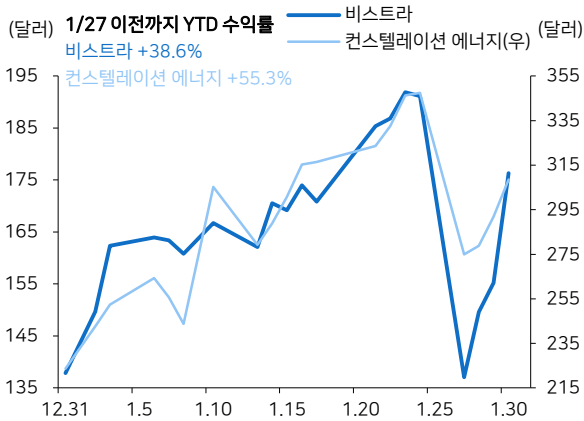
반면 주가에 영향이 크지 않거나 오히려 상승한 종목들도 눈에 띈다. 대표적으로 아마존과 애플, 메타와 같이 AI 인프라보다 AI 서비스에 익스포저가 높은 기업들은 AI 인프라 비용 하락이 수혜가 될 가능성이 있는 기업들은 주가가 올랐고, 팔란티어와 같은 소프트웨어 기업도 1/27에 4.5% 하락했지만 바로 이전 수준을 회복했다.

그림5 연휴 기간 주요 주가 반응 (1/30 종가 기준)



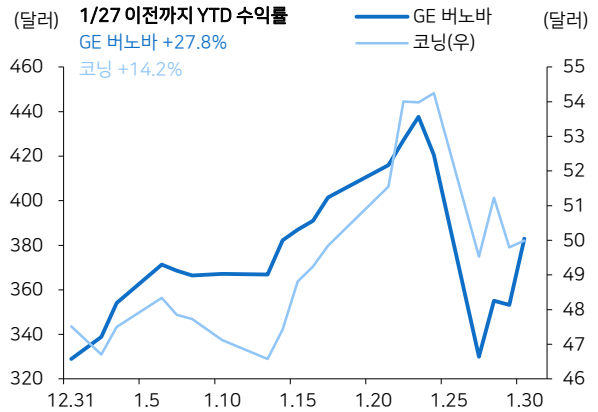
자료: Bloomberg, 메리츠증권 리서치센터

그림6 비스트라, 컨스텔레이션 에너지 주가: 연초 주가 상회



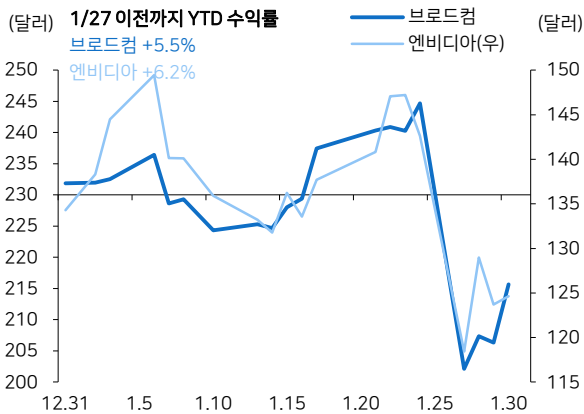
자료: Bloomberg, 메리츠증권 리서치센터

그림7 GE 버노바, 코닝 주가: 연초 주가 상회



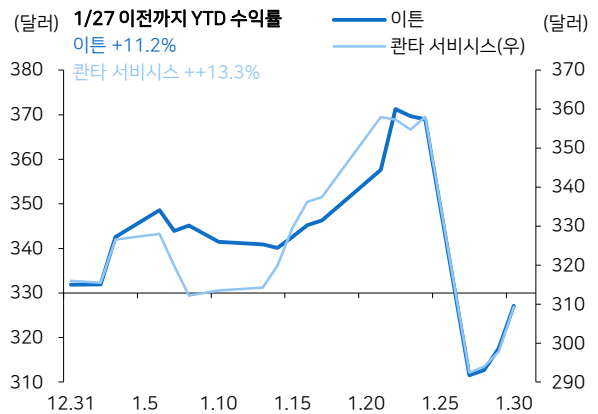
자료: Bloomberg, 메리츠증권 리서치센터

그림8 브로드컴, 엔비디아 주가: 연초 주가 하회



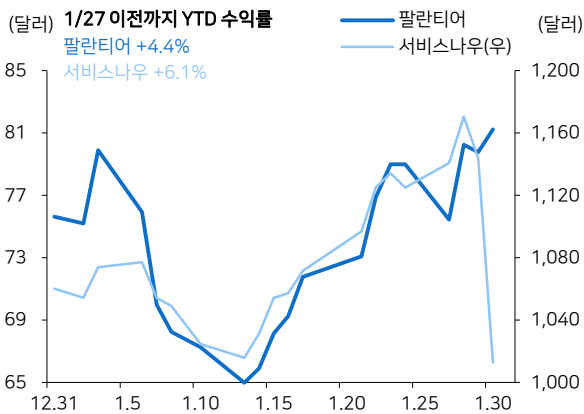
자료: Bloomberg, 메리츠증권 리서치센터

그림9 이튼, 관타 서비스 주가: 연초 주가 하회



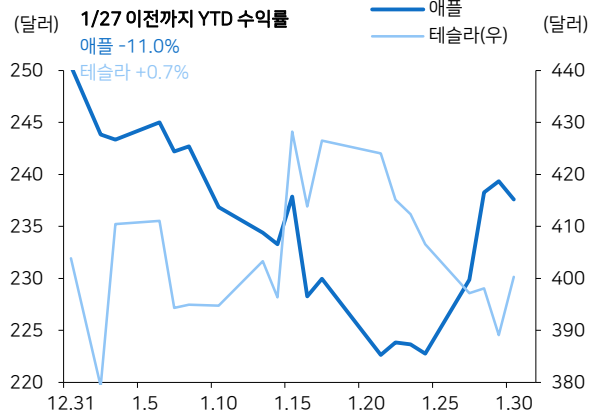
자료: Bloomberg, 메리츠증권 리서치센터

그림10 팔란티어, 서비스나우 주가: 이벤트 영향 없음



자료: Bloomberg, 메리츠증권 리서치센터

그림11 애플, 테슬라 주가: 낙폭 적거나 오히려 반등



자료: Bloomberg, 메리츠증권 리서치센터

### 1/27 이후 주가 회복 흐름의 배경과 제기되는 쟁점

주요 쟁점 2가지:

- 1) 딥시크 모델의 신빙성
- 2) Scaling Law의 종말?

1월 27일 이후 딥시크로 인해 화두가 되는 쟁점은 크게 두 가지다; 1) 딥시크의 모델이 신빙성이 있는지, 2) 중국이 보여준 비용절감 기술의 효율성은 Scaling law(더 많은 AI 하드웨어 투입→더 나은 AI 소프트웨어 성능)는 끝을 의미하는지. 다만 두 쟁점 모두 그렇지 않다는 쪽으로 생각이 모여면서 낙폭이 컸던 AI 인프라 기업들의 주가가 회복 중이다.

딥시크 훈련 비용 과소평가, 미국 모델에 대한 불법적 증류, 보안, 신뢰도 문제 부각 중

딥시크의 신뢰성 문제에 대해서는 먼저 V3 모델의 훈련 비용이 과장되었다는 이야기가 나온다. V3 모델의 훈련 비용 557만 달러는 공식 훈련 만을 포함, 아키텍처, 알고리즘, 데이터에 대한 사전 연구 실험 관련된 비용은 제외되었기 때문에 과소평가 되었다는 것이다. 또한 R1 등 모델을 훈련할 때 Distillation을 했을 가능성을 제기한다. Distillation(증류)이란, 큰 모델의 지식을 작은 모델로 압축하여 학습하는 기법인데, 딥시크가 OpenAI의 모델을 무단으로 활용했다는 의혹이 제기 중이다. 여기에 딥시크의 보안 취약점, 신뢰도, 중국 정부 통제 가능성 등도 언급 중이다.

표1 1/27 충격 이후 DeepSeek에 대한 주요 의견 정리

구분	출처	주요 의견
긍정	Mega Research	- DeepSeek의 모델 비용 대폭 절감 → AI 보급률 상승 및 엣지 디바이스, GPU AI PC 수요 증대로 이어질 것 - 결과적으로 GPU 및 ASIC 칩 수요가 전반적으로 확대될 전망
	Pat Gelsinger	- AI 비용이 낮아지면 시장 수축이 아니라 폭발적으로 확장된다는 기존 컴퓨팅 시장의 교훈을 강조 - DeepSeek의 등장은 엔지니어링 혁신을 보여주는 예시이며, 결국 전체 AI 시장의 확대에 기여할 것
	Nvidia	- DeepSeek의 발전은 엔비디아 칩이 중국 시장에서도 여전히 필수적임을 보여주는 사례 - 추후 단계에서도 대규모 GPU와 고성능 네트워킹이 필요하기 때문에 추후 더 많은 엔비디아 칩이 요구될 것
	BofA	- 엔비디아, 브로드컴, 마벨에 대한 매수 의견 유지. DeepSeek의 비용 효율 이슈가 부각되었지만, 이는 AI 투자를 근본적으로 줄이지는 않을 것. AI 슈퍼사이클은 계속될 것으로 전망
	Jefferies	- DeepSeek의 효율성은 혁명이 아니라 지속적 진화의 일부로 해석 - 소프트웨어 효율성 개선이 AI 도입 속도를 높일 것이며, 과도한 매도세는 기회 요인으로 판단
	Wedbush	- DeepSeek 모멘트는 오히려 매수 기회를 제공. 빅테크가 추진하는 AGI, 자율주행, 로봇틱스는 여전히 엔비디아 등 미국 반도체 기업 중심으로 돌아가므로 장기적인 AI 인프라 투자가 축소되지 않을 것
	Trendforce	- DeepSeek가 보여준 비용 효율성은 AI 인프라의 소프트웨어 중심 최적화 트렌드를 가속화 - 다만 AI 서버 시장은 계속 확대되며, '25년 이후에도 CSP의 대규모 투자 기초가 이어질 가능성이 높음
부정	OpenAI	- DeepSeek가 OpenAI 모델 출력을 '무단으로' 활용했다는 증거가 포착되어 조사 중 - API 사용 규정 위반 가능성, 데이터 제한 우회 시도 의혹 제기
	David Sacks (미 트럼프 행정부 AI 담당관)	- DeepSeek가 OpenAI 모델로부터 지식 증류를 통해 기술을 확보했을 가능성이 높으며 지적 - OpenAI는 이에 매우 불쾌할 것이라는 언급으로, IP 침해, TOS 위반 등 우려 제기
	WSJ, CNBC Reuters	- DeepSeek의 주장(훈련비용 약 560만 달러, 최고 수준 성능 달성)에 대해 불투명하거나 과장되었을 가능성 제기 - 제재가 있는 환경에서 어떤 칩을 사용했는지 명확하지 않으며, 실제 비교 지표 부족
	Citigroup	- DeepSeek 모멘트가 'OpenAI를 560만 달러에 만들었다' 라는 식의 해석은 과장 - 실제 대규모 첨단 모델 구현 시 여전히 고사양 GPU와 대규모 클러스터가 필요하며, 미국 기업이 계속 우위
	국태증권리서치본부	- DeepSeek가 AI 보급화에 기여하는 것은 사실이지만, 엔비디아에 미치는 영향은 제한적 - 주로 저비용 모델로 특정 사례에 대응 가능해도, 고성능 부문은 여전히 GPU, ASIC 수요가 크다는 견해
	Oracle	- DeepSeek의 저비용 공개가 OpenAI 구독형 모델에 부정적 영향을 미칠 수 있으나, 고성능 맞춤형 고객층에서는 여전히 부족하다는 평가. 완전한 대체보다는 부분적 시장 잠식 정도로 그칠 것이라는 견해
포춘지	- 딥시크 신뢰도 문제 부각. 정보 검증에서 83%의 높은 실패율 기록, 보안상 취약점 다수 발견 - '23년 10월 이후 데이터 학습 불가, 거짓 정보를 학습하도록 쉽게 조작 가능, 중국 정부의 통제정책이 원인	

자료: 언론 보도 종합, 메리츠증권 리서치센터 정리

**딥시크 충격에도 AI 인프라 투자 기대가 유지되는 이유**

모델의 증류(distillation)은 고성능 모델의 지속 필요성 함의

특히 딥시크의 OpenAI 모델 distillation 가능성은 오히려 고성능의 LLM 모델의 지속적인 개발이 필요함을 의미하기 때문에 고성능 AI 모델 개발에 대한 니즈, Scaling law에 따라 AI 인프라 투자의 지속적인 필요성을 함의한다.

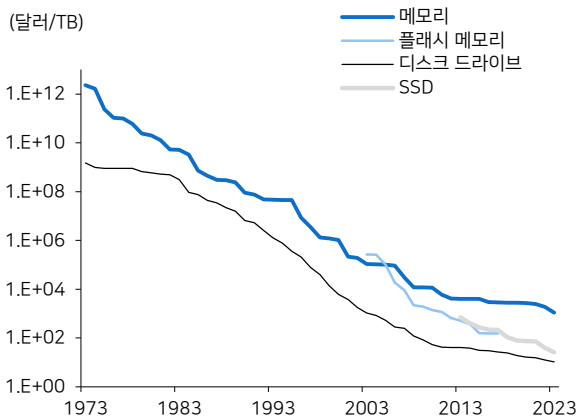
비용절감은 기술확산의 전형적 패턴, AI 인프라 투자 확대 저해 요인이 아니라는 평가 주류

한편, 2)의 관점 중 일부인 비용절감 기술력은 인정되는 분위기이다. 다만 기술 효율 개선에 따른 비용 절감은 기술 확산의 전형적인 패턴이지, AI 인프라 투자 확대를 저해하는 요인은 아니라는 방향으로 시장의 생각이 모이고 있다. 딥시크 충격 이후 엔비디아를 비롯한 주요 IB들은 딥시크 모멘트가 오히려 AI 인프라 기업들에 대한 매수 기회를 제공하고 있다고 평가했다. 딥시크의 기술이 AI 보급에 기여하는 것은 사실이지만, 여전히 고사양 GPU가 필요하며, 기술 진보에 따른 비용 하락은 시장 수축이 아니라 폭발적인 확장이 있었다는 기존 컴퓨팅 시장의 교훈을 강조했다.

Test Time Scaling 이해 필요

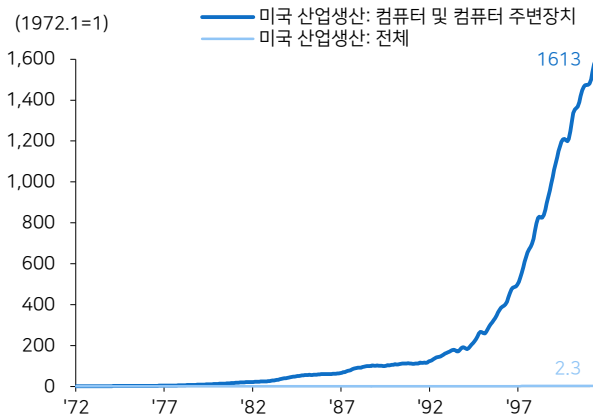
표면적으로 딥시크의 기술 진보는 AI Scaling law, AI 학습시 컴퓨팅 자원(모델 크기, 데이터 규모, 컴퓨팅 파워)이 증가함에 따라 모델 성능이 향상된다는 개념의 종말과 AI 인프라 투자가 불필요한 것처럼 보인다. 그러나 딥시크에 활용된 제논황, 샘 올트먼이 최근 자주 언급하는 Test time scaling에 대해 이해하면 추론용 AI 컴퓨팅 자원에 대한 투자 중심의 AI Capex 확대의 필요성을 이해할 수 있다.

**그림12 PC 사이클: exponential 비용 절감은 기술진보 과정**



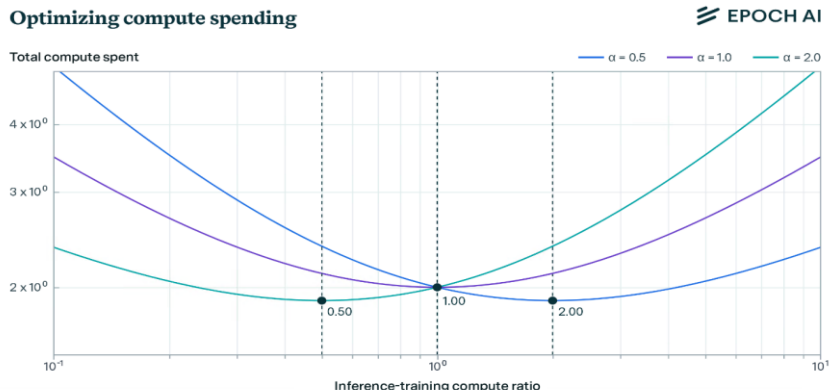
자료: Our World in Data, 메리츠증권 리서치센터

**그림13 비용 절감은 기술 확산을 가속시키는 요인일 뿐**



자료: Federal Reserve, 메리츠증권 리서치센터

**그림14 Epoch AI: AI 최적 컴퓨팅 비율은 훈련과 추론에 5:5로 할당하는 것**



주: AI 훈련과 추론에 할당하는 컴퓨팅 비율이 5대5가 최적임을 나타내는 그림  
 자료: Epoch AI, "Optimally Allocating Compute Between Inference and Training" ('24.3.29)

Scaling Law:  
 학습 단계의 더많은 컴퓨팅 자원  
 → 더 나은 모델  
 Time Scaling Law:  
 추론 단계의 더많은 시간  
 → 더 나은 결과

전통적인 Scaling Law 에서는 1) 학습 단계에서 2) 더 많은 AI 학습 자원(모델 파라미터 수, 학습 데이터, 연산 자원(컴퓨팅 프로세서))이 확보될수록 더 강력한 AI 모델을 만들 수 있다고 강조한다. 이와 대비되는 Time Scaling Law 는 1) 추론 단계에서, 2) 더 많은 계산 시간을 투입하면 더 나은 결과를 도출할 수 있다는 법칙이다. Time Scaling Law 하에서는 AI 훈련 비용은 줄일 수 있지만, 추론에 필요한 인프라와 비용이 더 필요하게 된다.

딥시크 R1 은 OpenAI의 o1 처럼  
 Test Time Scaling 적용 모델

딥시크 R1은 OpenAI의 o1과 함께 Test Time Scaling이 거의 처음으로 적용되는 AI 모델이다. o1의 유료 사용자라면 경험해보았을 것인데, 모델이 생각하는 과정을 거쳐 적게는 수 초, 길게는 수십 초간 '생각'을 하다가 답을 내놓는다. 이 생각하는 시간은 Test Time Scaling이 적용되어 나타나는 현상으로 추측한다. 그리고 GPT-4를 사용하다가 o1를 사용하면, 획기적인 성능 향상을 체험할 수 있을 것이다. 딥시크 R1도 마찬가지다. 아마 o1의 유료 사용이라는 진입장벽이 Test Time Scaling의 성과를 무료로 체험하게 해준 딥시크 R1의 역량을 더 놀라게 했을 것이다.

두 모델의 공통점:  
 느린 결과산출속도, 높은 레이턴시

R1의 함의:  
 Test Time Scaling 성과 체험을  
 확산시킨 첫 무료 모델

이 두 모델은 공통적으로 주요 AI 모델 중 가장 낮은 결과 산출 속도, 가장 높은 레이턴시를 보인다. 탁월한 추론 결과를 보여주지만, '생각'하는 시간(Test Time)이 추가적인 자원으로 필요한 것이다. 따라서 딥시크 R1의 함의는 Test Time Scaling 기술 성과의 체험을 확산시킨 첫 무료 모델이라는 데 있다고 평가한다.

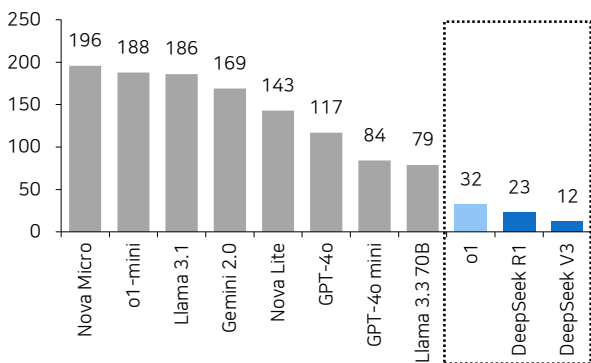
표2 Scaling Law와 Time Scaling Law 비교

구분	전통적인 Scaling Law	Time Scaling Law(Test Time Scaling)
초점	- 훈련(Training) 과정 - 모델 크기, 학습 데이터, 연산 자원의 양	- 추론(Inference) 과정 - 예측/생성시 연산 자원 - 시간 투자량
요점	- 모델을 더 크게+더 많은 데이터+더 많은 GPU → 더 강력한 모델	- 추론시 더 많은 계산(시간) 투입 → 더 나은 결과(정확/정교)
산업 영향	- AI 거대 모델 경쟁 가속 - AI 훈련 비용 대폭 증가 - 엔비디아 GPU 등 수요 폭증	- 훈련 비용 뿐만 아니라 추론 비용 또한 확대 - 클라우드/GPU/ASIC 추론 인프라에 대한 투자 지속
기술 예시	- GPT-4 등 모델 훈련 비용 수억 달러 - LLaMA 등 주요 빅테크의 초거대 모델	- Beam Search 확장, 재추론, Prompt Chaining으로 품질 향상 - 일부 모델(o1, R1)에서 Test time Scaling 개념 시도
진화 방향	- 파라미터 수를 지속적으로 늘림 - 거대 데이터셋 확보 - 컴퓨팅 파워의 제한이 시장을 결정	- 추론 효율성과 유연성 강화 - 주어진 상황에서 필요한 만큼 '추론 스케일'을 확장해 성능 강화

자료: 메리츠증권 리서치센터 정리

그림15 주요 AI 모델 결과 산출 속도

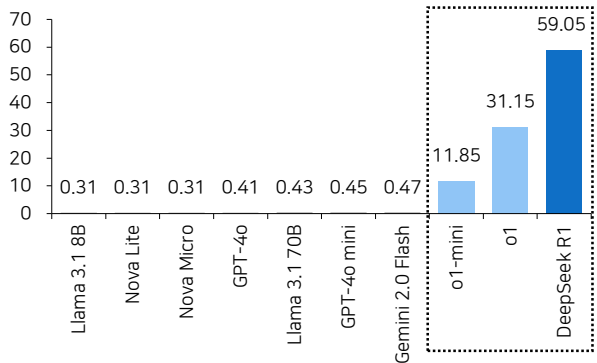
(output token per second, 높을 수록 빠름)



자료: Artificial Analysis, 메리츠증권 리서치센터

그림16 주요 AI 모델 레이턴시(지연 속도)

(seconds to first token received, 낮을수록 빠름)



자료: Artificial Analysis, 메리츠증권 리서치센터

딥시크 모멘트: AI 추론 컴퓨팅 니즈 가속화시키는 요인으로 해석

신기술의 체험 확산은 해당 기술에 대한 더 많은 니즈를 유발할 것이다. 따라서 이번 이슈의 AI 하드웨어 측면에서 시사점은, AI 소프트웨어의 확산으로 AI 컴퓨팅의 중심이 학습(Training)에서 추론(Inference)으로 넘어가는 현상을 더 가속화시키는 요인으로 해석해야 할 것이다. 이 경우 AI 하드웨어 중에는 추론에서 상대적으로 강점을 보이는 ASIC에 대한 니즈 가속 시점이 앞당겨질 수 있다.

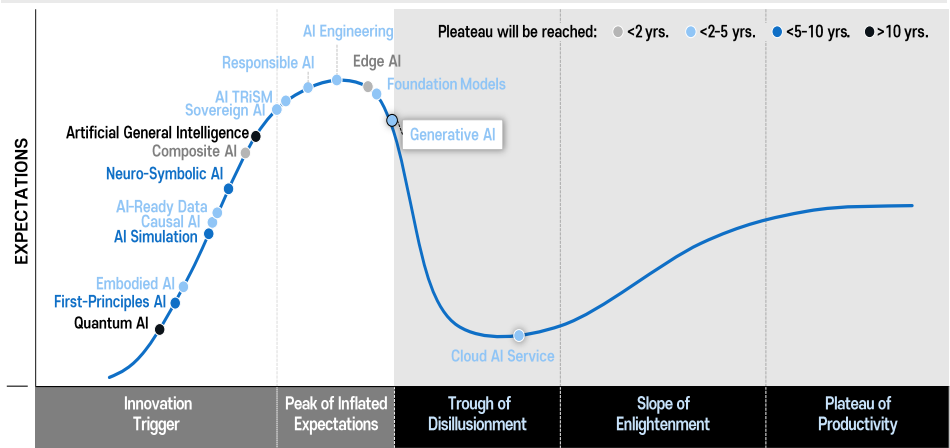
Gen AI 단계에서 나타난 현상, AI 학습에 대한 수요가 약해진 것으로 해석하기 어려움

그러면 AI 학습에 대한 수요의 중요성은 약해지는 것이라고 해석해야 할까? 좀 더 큰 그림에서는 이번 딥시크 이슈가 Generative AI에서 나타난 현상임을 인지할 필요가 있다. Generative AI는 가트너의 hype cycle 상 이미 피크를 지나고 있는 기술이다. 새로운 기대보다는 실질적인 성과를 보여주어야 하는 기술의 영역이 된 셈이다. AI 학습에 대한 니즈가 더 큰 분야는 Generative AI 다음 기술들일 것이다.

Physical AI에서는 고성능 AI 학습 컴퓨팅 니즈 여전히 있을 것

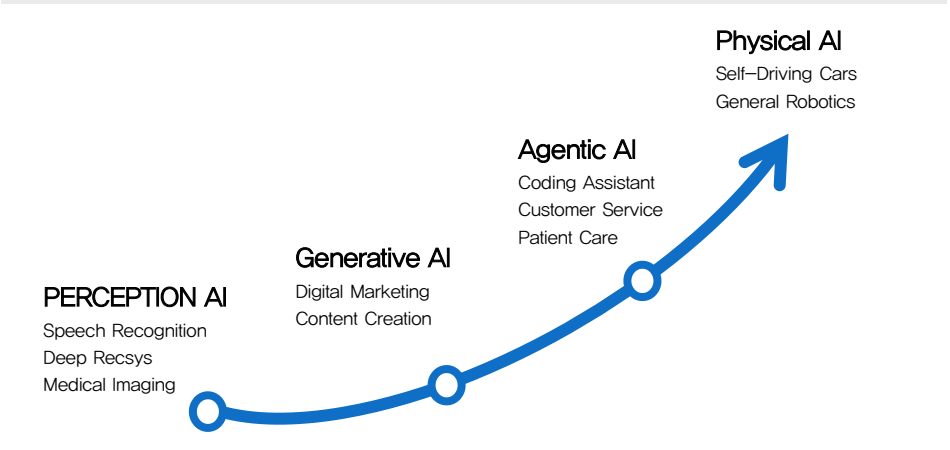
엔비디아 젠슨 황이 올해 1월 CES 기조연설에서 이미 강조했듯, AI 기술 발전 단계에서 Generative AI 이후 중착점은 Physical AI다. AI 모델은 더 다양한 데이터에 대한 효과적인 학습을 기반으로, 빠른 반응으로 인간을 물리적으로 보조해야 하는 자율주행, 로봇의 영역에서 경쟁 중이다. 고성능 AI 학습칩에 대한 니즈는 아직 중착점이 아니라고 봐야 할 것이다.

그림17 AI hype cycle 상 peak-out 중인 Generative AI



자료: Gartner(2024.11.11), 메리츠증권 리서치센터

그림18 글로벌 AI 산업은 Generative AI 다음 페이지 경쟁으로 넘어가는 중



자료: CES 젠슨황 기조연설, 메리츠증권 리서치센터



딥시크 충격 이후 주요 기술 기업 실적발표

딥시크 이슈에도 AI CAPEX에 대한 의견을 유지하는 빅테크

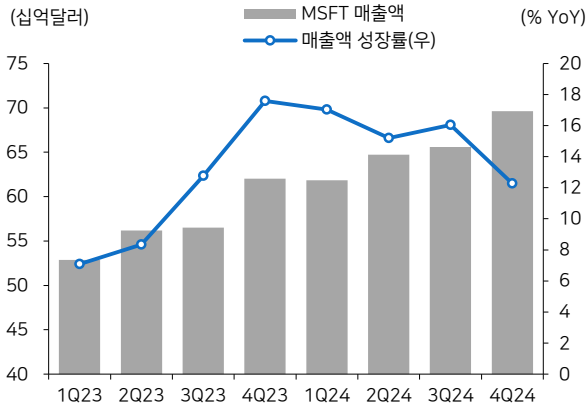
AI 인프라 기업들의 Capex에 대한 생각은 실적 컨콜에서도 드러난다. 이번주 실적 발표한 마이크로소프트, 메타는 딥시크 관련한 질문에 모델 비용 절감에 따른 B2C 확산 기대, 여전히 대규모 Capex 투자가 전략적 우위를 제공할 것이라는 관점을 유지했다. 서비스나우와 같은 소프트웨어 기업은 딥시크 모멘트를 실적 컨콜에서 AI 어플리케이션 개발 투자를 늘릴 수 있는 기회로 해석했다.

표3 DeepSeek R1 발표 이후 AI 관련 기업 실적 컨콜 주요 내용 정리

구분	기업명	매출	EPS	가이드스	주요 코멘트
AI 인프라	MSFT	696억 (하회)	3.23 (하회)	매출: 698.1억 - PBP: 299.3억 - IC: 299.3억 - MPC: 130억 - Azure: 31~32% YoY	<p><b>Intelligence Cloud(IC):</b> 255억 달러 (+19% YoY). Azure +31%</p> <p>AI Business는 130억 달러(+175% YoY)의 Run Rate 달성. - 데이터센터 내 모든 레이어별로 성장이 지속되는 중. 인프라 차원에서는 <b>고객사의 Migration 지속</b> 중. 활용처 차원에서 Copilot은 구독 기업들이 <b>활용 인원 수를 10배 수준으로 확대</b>했으며, 개별 직원의 <b>활용도도 전분기 대비 60% 증가</b></p> <p>가이드스 하회 배경: Non-AI Azure 부문의 프로그램 도입이 원활하지 않았기 때문. On premise, Enterprise &amp; partner Services 등도 마찬가지로</p> <p>Capex 관련: 1) 학습과 추론 모두 생산성 개선 진행 중. 2) 한편 AI 수요가 기하급수적으로 늘어날 텐데, MSFT가 집중하는 부분은 전세계적으로 서비스를 공급하는 것. 이를 위해서는 <b>매출 성장과 capability 성장의 균형을 맞춰 가야 할 것</b></p> <p><u>DeepSeek 관련: 딥시크로 인해 비용이 절감되면 어플리케이션 개발이 활성화 될 것이고 어플리케이션 소비도 덩달아 증가할 것. 모델이 저렴해지면서 AI PC 등을 통해 보급된다면, 서버와 PC 사업을 모두 영위하는 MSFT에는 좋은 뉴스일 것</u></p>
	META	484억 (상회)	8.02 (상회)	매출: 395억~418억 CAPEX: 600-650억	<p><b>Family of Apps:</b> 473억 달러 (+21% YoY) - 전체 광고는 +6% YoY 증가했으며, 광고당 단가는 +14% 증가.</p> <p>가이드스: - Family of Apps: 1) 일간 활성화 수준이 페이스북, 인스타그램, 왓츠앱 모두에서 성장을 보이는 중, 2) <b>AI 기반 맞춤형 광고를 통해 수익화 관련 효율성을 개선</b>시키고 있음. Thread와 같이 아직 수익화 시도가 이루어지지 않은 부문에서도 <b>맞춤형 광고 효과 누릴 수 있을 것</b></p> <p>Capex: Llama의 성능 개선 및 생성형 AI 제품의 <b>활용처 확대</b>를 우선순위에 둘 것</p> <p><u>DeepSeek 관련: 저렴한 모델이 등장했다고 해서 컴퓨팅 리소스가 덜 필요하다는 것은 아님. 오히려 추론 과정에서 더 많은 컴퓨팅 자원을 활용함으로써 더 나은 서비스 품질을 제공할 수 있음. 즉 대규모 인프라를 구축하는 역할이 높은 품질의 서비스를 제공하기 때문에 여전히 Capex 대규모 투자가 전략적 우위를 제공할 것.</u></p>
자율주행	TSLA	257억 (하회)	0.73 (하회)	-	<p>마진 축소 관련: 평균 판매 가격 하락과 더불어 3분기 출시된 FSD 차량의 영향이 4분기에 인식되며 마진이 축소됨. 다만 차량 당 비용 가격을 35,000달러 선으로 절감. 그러나 신제품 출시 및 공장 업데이트로 인해 몇 주 동안 생산이 이루어지지 않을 수 있으며, 그 결과 마진이 영향을 받을 수밖에 없음</p>
AI SW	NOW	30억 (하회)	3.67 (상회)	구독 매출: 30억	<p><b>구독 매출:</b> 28억 +21% YoY, CRPO: +22% YoY. - 2개의 신규 대형 고객사 확보(ACV 1억달러 이상), 이는 AI 플랫폼의 스케일링이 시작되었음을 의미 - Now Assistant (Agentic AI)의 경우 +150% QoQ의 <b>딜 개수 확대</b>를 보임</p> <p>현재로서는 신규 에이전트 개발보다는 <b>채택률을 높이는 것이 우선순위</b></p> <p><u>DeepSeek 관련: 모델들이 상품화되면 고객들의 입장에서 인프라, LLM 등에 대한 투자 부담이 줄어들고, 어플리케이션 개발에 대한 투자가 늘어날 수 있는 기회</u></p>
	IBM	176억 (상회)	3.92 (상회)	(‘25년 가이드스) 매출: +5% YoY	<p><b>Software:</b> 790억 달러 (+11% YoY) - Red Hat (+17% YoY)가 가장 크게 성장에 기여. 6분기 연속 두자릿수 성장률을 기록 중, 이는 <b>비용 최적화를 위한 클라우드 솔루션에 대한 수요가 강함</b>을 방증 - 신규 제품 출시도 소프트웨어 성장에 유의미한 영향을 끼침 - 소프트웨어에 힘입어 <b>AI 포트폴리오 규모가 올해 20억 달러 확대</b>됨</p>
AI 인프라	GLW	39억 (상회)	0.57 (상회)	(1분기 가이드스) 매출: 39억  (‘25년 가이드스) 매출: 145억 Capex: 13억	<p><b>Optical Communications:</b> 14억 달러 (+51% YoY) - <b>사이클 반등</b>에 따른 매출 확대. 이를 위해 6월에 신규 <b>생성형 AI 제품</b>을 공개. - 매 분기 신규 생성형 AI 제품에 대한 수요 증가했으며, 그 결과 <b>옵티컬 사업부 내 Enterprise 매출 비중이 4분기에 +93% YoY 확대</b> - 도시 간 AI 데이터센터 연결: Lumen Technologies와 협업하여 생성형 AI 광섬유를 처음으로 외부 공장에 설치</p>

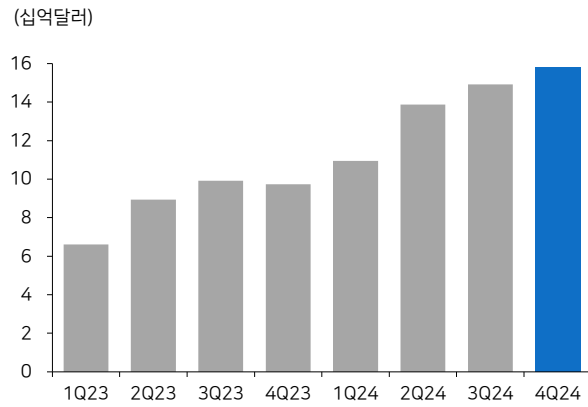
주: 괄호는 컨센서스 상하회 여부  
자료: Bloomberg, 메리츠증권 리서치센터

그림19 마이크로소프트 AI 매출액 및 성장률



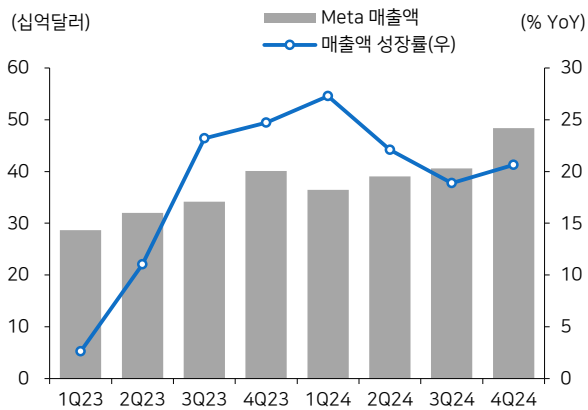
자료: Bloomberg, 메리츠증권 리서치센터

그림20 마이크로소프트 CAPEX 금액



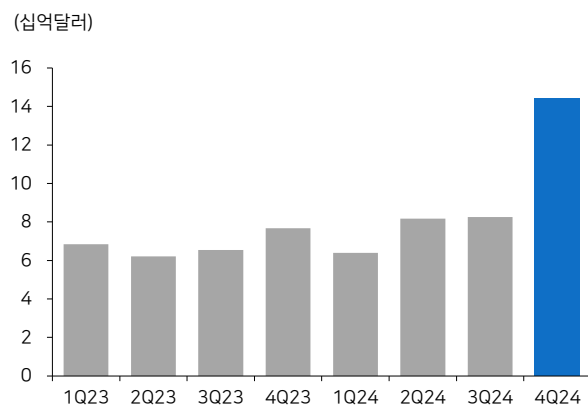
자료: Bloomberg, 메리츠증권 리서치센터

그림21 메타 매출액 및 성장률



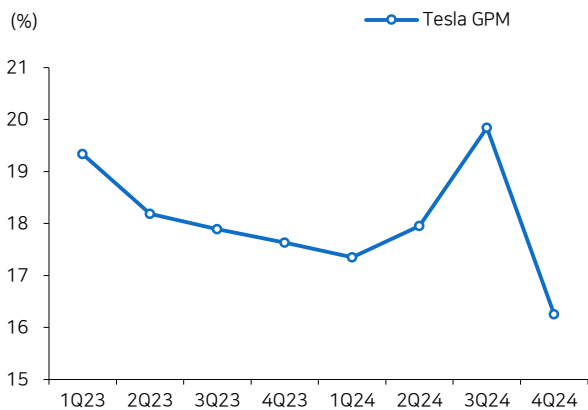
자료: Bloomberg, 메리츠증권 리서치센터

그림22 메타 CAPEX 금액



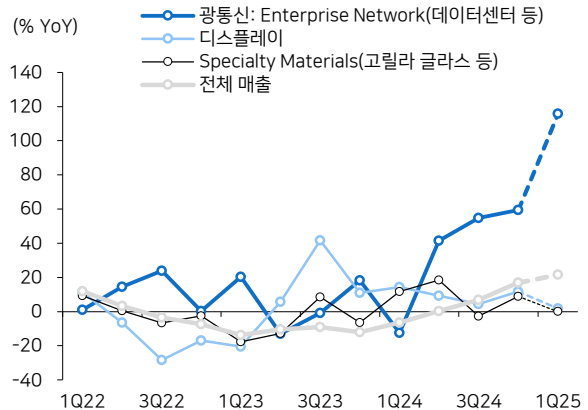
자료: Bloomberg, 메리츠증권 리서치센터

그림23 테슬라 GP 마진: 일시적 요인에 의한 부진



자료: Bloomberg, 메리츠증권 리서치센터

그림24 코닝 사업부별 매출액 성장률



주: 점선은 다음 분기 컨센서스  
자료: Bloomberg, 메리츠증권 리서치센터

**대응: 단기 저가 매수 전략, 중기 추론용 반도체 리레이팅 기대**

1/27 주가 반응 과도, 낙폭 과대  
종목 위주의 저가매수 전략 유효

이상의 분석을 정리하면, 딥시크 R1으로 AI 학습의 비용 절감 가능성은 확인이 되었지만, Test Time Scaling이 적용되었음을 고려할 때 AI 추론용 컴퓨팅 자원의 니즈의 가속화를 고려해야 한다. 따라서, AI 컴퓨팅에 필요한 인프라 투자 확대는 여전히 필요하며, AI 학습에 필요한 컴퓨팅도 딥시크 모멘트가 Generative AI 단계에서 나타났음을 고려할 때, Physical AI 등 더 고차원의 AI 모델 학습에 필요한 컴퓨팅 수요는 여전히 유효하다고 본다.

Test Time Scaling 경험 확산,  
추론용 반도체 ASIC 리레이팅 기대

즉, 1월 27일 AI 인프라 기업을 중심으로 나타난 주가 하락은 과도한 반응이며 낙폭이 컸던 종목들을 중심의 저가 매수 전략이 유효하다는 판단이다. 단기적으로는 실적 발표를 앞둔 AI 인프라 기업 중심으로 관심 가질 필요가 있어 실적 발표 일정을 함께 체크하는 것이 필요하다. 특히 Test Time Scaling에 직접 수혜가 예상되는 AI 추론용 반도체, ASIC 관련한 종목들에 대한 리레이팅이 가속화될 수 있음을 고려해야 한다.

표4 2월 주요 실적발표 일정

MON	TUE	WED	THU	FRI
3 Palantir (PLTR)	4 Alphabet (GOOGL) AMD (AMD) Paypal (PYPL)	5	6 Amazon (AMZN)	7
10	11	12 Cisco (CSCO) Vertiv Holdings (VRT)	13 Applied Materials (AMT)	14
17	18 Arista Networks (ANET)	19	20	21
24 ONE OK (OKE) BWX Technologies (BWXT)	25 Constellation Energy (CEG)	26 Nvidia (NVDA) Salesforce (CRM) Vistra Group (VST) NRG Energy (NRG)	27 Quanta Services(PWR)	28

자료: Nasdaq, 메리츠증권 리서치센터

### Compliance Notice

본 조사분석자료는 제3자에게 사전 제공된 사실이 없습니다. 당사는 자료작성일 현재 본 조사분석자료에 언급된 종목의 지분을 1% 이상 보유하고 있지 않습니다. 본 자료를 작성한 애널리스트는 자료작성일 현재 추천 종목과 재산적 이해관계가 없습니다. 본 자료에 게재된 내용은 본인의 의견을 정확하게 반영하고 있으며, 외부의 부당한 압력이나 간섭 없이 신의 성실하게 작성되었음을 확인합니다.

본 자료는 투자자들의 투자판단에 참고가 되는 정보제공을 목적으로 배포되는 자료입니다. 본 자료에 수록된 내용은 당사 리서치센터의 추정치로서 오차가 발생할 수 있으며 정확성이나 완벽성은 보장하지 않습니다. 본 자료를 이용하시는 분은 본 자료와 관련한 투자의 최종 결정은 자신의 판단으로 하시기 바랍니다. 따라서 어떠한 경우에도 본 자료는 투자 결과와 관련한 법적 책임소재의 증빙자료로 사용될 수 없습니다. 본 조사분석자료는 당사 고객에 한하여 배포되는 자료로 당사의 허락 없이 복사, 대여, 배포 될 수 없습니다.