

전기전자 (Overweight)

비용효율적 AI의 등장과 IT H/W 영향 (feat. DeepSeek)

Issue Comment

[전기전자] 고의영 2122-9179 ey.ko@imfnsec.com, [IT RA] 손우성 2122-9785 useong@imfnsec.com

Deepseek, 확인이 더 필요한 부분과 IT H/W 섹터에 대한 영향 정리

비용효율적 AI모델의 등장: DeepSeek-V3 (이하 V3)는 6,710억개의 파라미터를 가진 MoE 구조 기반의 LLM이다. MoE 구조이므로 토큰당 활성화 파라미터는 370억개에 불과하다. V3는 강화학습(RL)과 지도학습(SFT)을 활용한 후처리를 거쳐 인간의 선호에 맞게 최적화되었다. 특히 강화학습에서 보상을 통해 성능을 크게 개선할 수 있었다. 강화학습이란 AI가 더 좋은 대답을 할수록 더 높은 보상을 받는 방식으로 학습하는 방법이다. 논문에 따르면, V3는 GPT-4o, Claude 3.5와 비교 가능한 성능을 보여준다. 논란이 된 부분은 V3 훈련을 위해 저사양의 H800 GPU를 활용했으며, 단 \$560만의 비용이 소요되었다는 점이다 (H800 GPU 임대 가격을 시간당 \$2로 가정) [그림1]

V3를 바탕으로 태어난 R1: 한편, DeepSeek-R1 Zero는 V3를 기본 모델로 활용하여 강화학습을 거친 모델이며, 이어서 공개된 DeepSeek-R1 (이하 R1)은 소량의 Cold-Start 데이터를 도입한 모델이다. Cold-Start 데이터란 AI 모델이 기본 사고방식을 할 수 있도록 돕는 초기 학습 데이터를 넣어주는 개념으로, 이를 통해 강화학습을 더 효과적으로 할 수 있다. 결과적으로 R1은 뛰어난 추론 능력을 얻게 되었는데, 이를 Llama3, Qwen2.5 등 여타 오픈소스 모델에 증류하여 매개변수 15억개~700억개의 경량화 AI 모델까지 선보였다. 이 또한 우수한 성능을 기록하였음이 공개됐다 [그림2]

확인이 더 필요한 부분들: 이처럼 비용효율적인 V3, R1의 공개로, 빅테크의 Capex가 높은 상태로 유지될 지 여부에 대한 의구심이 커졌다. 다만 여전히 확인 필요한 부분들이 있다. ① 논문에 따르면 V3 훈련을 위한 \$560만의 비용에는 아키텍처, 알고리즘, 데이터와 관련된 사전 연구 및 실험 비용은 제외되었다. [그림3], ② Financial Times에 따르면 OpenAI는 DeepSeek 훈련에 자사의 모델이 활용되었다는 증거가 있다고 밝혔다. 만약 OpenAI의 모델이 베이스로 활용되었다면, 실제 훈련 비용이 과소 평가되었을 가능성이 있다. ③ 설사 훈련에 필요한 H/W 인프라가 적게 필요하다하더라도, 추론에 필요한 인프라는 여전히 많이 필요할 수 있다. 특히 비용효율적 AI 모델이 AI 어플리케이션의 시대를 연다면 더욱 그럴 수 있다.

IT H/W 섹터에 대한 영향: 앞서 확인이 필요한 부분들이 명확해지기 전까지 단기적으로 수동부품, PCB (FC-BGA, MLB) 등 AI 관련 컴포넌트 주식에 대한 투자 심리에 부정적일 수 있다. 다만, 기회요인도 있다. 특히 스마트폰, 가전 등 IT 디바이스에는 기회일 수 있다. 앞서 살펴보았듯 더 적은 컴퓨팅 자원으로 성능 좋은 AI 모델이 구현했다는 점, 그리고 이를 증류하여 성능이 뛰어난 경량화 AI 모델을 만들어냈다는 점은 On-device AI가 개화함에 있어 중요한 의미를 갖기 때문이다. DeepSeek의 충격이 IT 인프라 관련 주식의 주가를 누른 반면, Apple의 주가가 반등한 이유도 여기에 있는 것으로 판단된다.

당사 커버리지 주식 판단: ① 삼성전자: AI서버(ASIC형 FC-BGA, 서버용 MLCC)에 대한 수요 확대가 주된 투자포인트인 만큼 단기적인 주가 변동성이 있을 수 있다. 다만, 중장기적으로 비용효율적인 AI모델이 On-device AI의 확산, 즉 엣지 디바이스의 추론을 강화할 경우 수혜를 받을 가능성이 크다. 무엇보다도, 여타 AI 관련 주식 대비 밸류에이션 부담이 크지 않다는 점도 장점이다. 따라서 여전히 삼성전기를 가장 좋게 보고 있다. ② LG전자: 동사는 LLM을 활용한 스마트홈 전략을 청사진으로 제시하고 있다. 예컨대, 동사의 AI 허브는 일반 가전을 AI 가전으로 업그레이드하고, 서로 다른 가전을 오케스트레이션 하는 역할까지 수행한다. AI 허브의 두뇌는 GPT와 같은 AI 모델이 담당하는 바, 비용효율적 AI 모델의 등장은 이를 활용하는 동사에게도 긍정적인 수 있다. Apple이 이번 사태의 수혜로 인식되는 것과 유사한 맥락이다. ③ LG이노텍: Apple이 부각 받을 수 있는 환경이므로, LG이노텍과 같은 Apple 공급망에 대한 투자 심리도 단기적으로 개선될 수 있다. 다만, 공급망 내 경쟁심화는 여전히 동사 멀티플을 제한하는 요소로 작용할 전망이다. 펀더멘탈 변화가 더 필요하다.

그림1. DeepSeek-V3를 훈련하는데 소요된 시간과 비용 (H800 기준 총 279만 GPU 시간과 \$560만 소요)

Training Costs	Pre-Training	Context Extension	Post-Training	Total
in H800 GPU Hours	2664K	119K	5K	2788K
in USD	\$5.328M	\$0.238M	\$0.01M	\$5.576M

Table 1 | Training costs of DeepSeek-V3, assuming the rental price of H800 is \$2 per GPU hour.

자료: DeepSeek-AI, iM증권 리서치본부

그림2. DeepSeek-R1으로부터 증류된 AI 모델의 성능

Model	AIME 2024		MATH-500	GPQA Diamond	LiveCode Bench	CodeForces
	pass@1	cons@64	pass@1	pass@1	pass@1	rating
GPT-4o-0513	9.3	13.4	74.6	49.9	32.9	759
Claude-3.5-Sonnet-1022	16.0	26.7	78.3	65.0	38.9	717
OpenAI-o1-mini	63.6	80.0	90.0	60.0	53.8	1820
QwQ-32B-Preview	50.0	60.0	90.6	54.5	41.9	1316
DeepSeek-R1-Distill-Qwen-1.5B	28.9	52.7	83.9	33.8	16.9	954
DeepSeek-R1-Distill-Qwen-7B	55.5	83.3	92.8	49.1	37.6	1189
DeepSeek-R1-Distill-Qwen-14B	69.7	80.0	93.9	59.1	53.1	1481
DeepSeek-R1-Distill-Qwen-32B	72.6	83.3	94.3	62.1	57.2	1691
DeepSeek-R1-Distill-Llama-8B	50.4	80.0	89.1	49.0	39.6	1205
DeepSeek-R1-Distill-Llama-70B	70.0	86.7	94.5	65.2	57.5	1633

Table 5 | Comparison of DeepSeek-R1 distilled models and other comparable models on reasoning-related benchmarks.

자료: DeepSeek-AI, iM증권 리서치본부 주: 1) AIME2024 & MATH-500: 수학 능력 정확도 2) GPQA diamond: 일반 지식 (General-Purpose QA) 질문 정확도 3) LiveCode Bench & CodeForces 코딩 및 프로그래밍 정확도

그림3. DeepSeek-V3 훈련 비용에는 아키텍처, 알고리즘, 데이터와 관련된 사전 연구 및 실험 비용은 제외되어 있음

Lastly, we emphasize again the economical training costs of DeepSeek-V3, summarized in Table 1, achieved through our optimized co-design of algorithms, frameworks, and hardware. During the pre-training stage, training DeepSeek-V3 on each trillion tokens requires only 180K H800 GPU hours, i.e., 3.7 days on our cluster with 2048 H800 GPUs. Consequently, our pre-training stage is completed in less than two months and costs 2664K GPU hours. Combined with 119K GPU hours for the context length extension and 5K GPU hours for post-training, DeepSeek-V3 costs only 2.788M GPU hours for its full training. **Assuming the rental price of the H800 GPU is \$2 per GPU hour, our total training costs amount to only \$5.576M. Note that the aforementioned costs include only the official training of DeepSeek-V3, excluding the costs associated with prior research and ablation experiments on architectures, algorithms, or data.**

자료: DeepSeek-AI, iM증권 리서치본부

Compliance notice

당 보고서 공표일 기준으로 해당 기업과 관련하여,

- 회사는 해당 종목을 1%이상 보유하고 있지 않습니다.
- 금융투자분석사와 그 배우자는 해당 기업의 주식을 보유하고 있지 않습니다.
- 당 보고서는 기관투자가 및 제 3자에게 E-mail등을 통하여 사전에 배포된 사실이 없습니다.
- 회사는 6개월간 해당 기업의 유가증권 발행과 관련 주관사로 참여하지 않았습니다.
- 당 보고서에 게재된 내용들은 본인의 의견을 정확하게 반영하고 있으며, 외부의 부당한 압력이나 간섭 없이 작성되었음을 확인합니다.

본 분석자료는 투자자의 증권투자를 돕기 위한 참고자료이며, 따라서, 본 자료에 의한 투자자의 투자결과에 대해 어떠한 목적의 증빙자료로도 사용될 수 없으며, 어떠한 경우에도 작성자 및 당사의 허가 없이 전재, 복사 또는 대여될 수 없습니다. 무단전재 등으로 인한 분쟁발생시 법적 책임이 있음을 주지하시기 바랍니다.

[투자 의견]

종목추천 투자등급

종목투자자의견은 향후 12개월간 추천일 종가대비 해당종목의 예상 목표수익률을 의미함.

- Buy(매수): 추천일 종가대비 +15% 이상
- Hold(보유): 추천일 종가대비 -15% ~ 15% 내외 등락
- Sell(매도): 추천일 종가대비 -15% 이상

산업추천 투자등급

시가총액기준 산업별 시장비중대비 보유비중의 변화를 추천하는 것임

- Overweight(비중확대)
- Neutral(중립)
- Underweight(비중축소)

[투자등급 비율 2024-12-31 기준]

매수	중립(보유)	매도
92.4%	6.9%	0.7%