GLOBAL
EQUITY

2025. 12. 2

글로벌주식팀
이영진 Analyst, CFA
youngjin91.lee@samsung.com

리서치센터 리포트
바로가기

글로벌 소프트웨어

고래가 돌아왔다. 딥시크의 새로운 모델 DeepSeek-V3.2

- DeepSeek-V3.2 정식 버전과 딥 리즈닝용 고 컴퓨트 모델 V3.2-Speciale 공개
- 오픈소스 모델의 아쉬운 부분을 채우며 프론티어 모델과 유사한 성능 달성
- 벤치마크 외 체감되는 성능 확인이 필요하나, 가성비 측면에서는 여전한 매력도

WHAT'S THE STORY?

DeepSeek-V3.2와 V3.2-Speciale 모델 공개: 딥시크는 9월 말 V3.2-Exp 모델 공개 이후 V3.2 모델 정식 버전과 고 컴퓨트 버전 V3.2-Speciale을 새롭게 업데이트 했다. 두 모델 모두 기존 V3.2-Exp와 동일한 아키텍처를 활용했다. V3.2는 균형 잡힌 추론 성능을 추구하는 모델이다. V3.2-Speciale는 IMO(국제수학올림피아드)와 IOI(국제정보올림피아드) 등에서 금메달 레벨을 달성한 딥 리즈닝(Deep Reasoning) 전용 모델이다. 토큰 사용량이 많기 때문에 API로만 활용할 수 있고, 툴 호출 기능을 지원하지 않는다.

오픈소스 모델이 복잡한 작업에서 능력이 제한되는 이유: 딥시크 측은 추론 모델 등장 이후 오픈소스와 폐쇄형 모델 간의 발전 궤적에서 차이가 발생하며, 성능 격차가 벌어지고 있다고 판단했다. 특히 복잡한 과제 처리 측면에서 폐쇄형 모델이 우월한 능력을 보여주고 있다. 오픈소스 모델이 아쉬운 이유로는 1) 아키텍처 측면에서 vanilla attention 의존 지속, 2) 사후학습 단계에서 컴퓨팅 자원 투자 불충분, 3) AI 에이전트 맥락에서 일반화 능력과 지시를 따르는 능력 부족을 꼽았다.

한계를 넘기 위한 3가지 핵심 기술적 돌파구: 이러한 문제를 해결하기 위한 3가지 기술적 돌파구를 강조했다.

- **DeepSeek Sparse Attention(DSA):** 연산 복잡도를 줄이는 동시에 성능을 유지하는 효율적 어텐션 매커니즘을 새롭게 설계해 적용. 장기 컨텍스트 처리에 최적화
- **확장형 강화학습(RL) 프레임워크:** 견고한 강화학습 프로토콜을 구축 및 사후학습 단계에서 컴퓨트 투입량 대폭 확대. 사전학습 비용의 10% 초과하는 수준을 사후학습에 투입. 이를 통해 상위 모델과 경쟁 가능한 추론 성능 확보
- **대규모 에이전트 작업 합성 파이프라인:** 툴 사용 시나리오에서 추론 능력 통합을 위해 대규모 학습 데이터를 체계적으로 생성하는 합성 파이프라인 개발. 1,800개 이상의 환경과 85,000개의 복잡한 프롬프트 생성 기반의 방대한 합성 데이터가 에이전틱 사후 학습을 가능케하며, 복잡한 상호작용 환경 하에서 일반화 능력과 지시를 따르는 견고성을 향상

(다음 페이지에 계속)

벤치마크가 전부는 아니지만 여전한 가성비: 보여준 벤치마크 상으로는 구글 Gemini 3.0 Pro와 오픈AI GPT-5에 필적하는 성능을 보여주고 있다. 하지만 과적합 가능성을 배제할 수 없기 때문에 실사용 후기를 확인해야한다. 또한 멀티모달이 아닌 언어모델이라는 딥시크 모델의 고질적 한계도 여전히 존재한다.

토큰 효율성 측면의 논란이 있지만, 가성비는 놀라울 정도다. API 가격은 인풋 $0.028(Cache Hit) 및 $0.28(Cache Miss), 아웃풋 $0.42/백만토큰 수준이다. 미국 프론티어 모델 GPT-5.1(인풋 $1.25 및 아웃풋 $10), Gemini 3.0 Pro(인풋 $2 및 아웃풋 $12), Claude Opus 4.5(인풋 $5 및 아웃풋 $25) 대비 현저히 낮다.

인풋 8: 아웃풋 2 비율로 계산해도 GPT-5.1의 1/10 수준이다. 또한 구글과 오픈AI의 IMO 금메달 레벨 모델이 아직 출시되지 않은 상황에서 최초의 IMO 금메달 레벨 모델이 상용화되었다. 더구나 오픈소스로 풀리게 된 것도 상당한 의미가 있다.

스케일링 법칙은 계속된다: 구글 Gemini 3.0 Pro는 사전학습 스케일링이 여전히 유효함을 증명했다. 여기에 딥시크의 V3.2 Speciale는 대규모 컨텍스트를 기반의 강화학습 스케일링이 유효함을 증명했다. 데이터와 컨텍스트를 다듬는 작업은 모델 개선으로 이어지고 있다. 모델 개선 속도가 둔화되면 항상 나오는 것이 "스케일링 법칙은 벽에 부딪혔다"이다. 하지만 AI 산업은 끊임없이 돌파구를 만들어 내고 있다.

한계라고 쓰고 자신감이라고 읽는다: 기술 보고서에서 프론티어 비공개 모델과 비교하면 여전히 한계가 존재한다고 언급했다. 한계와 향후 계획을 짚어보면,

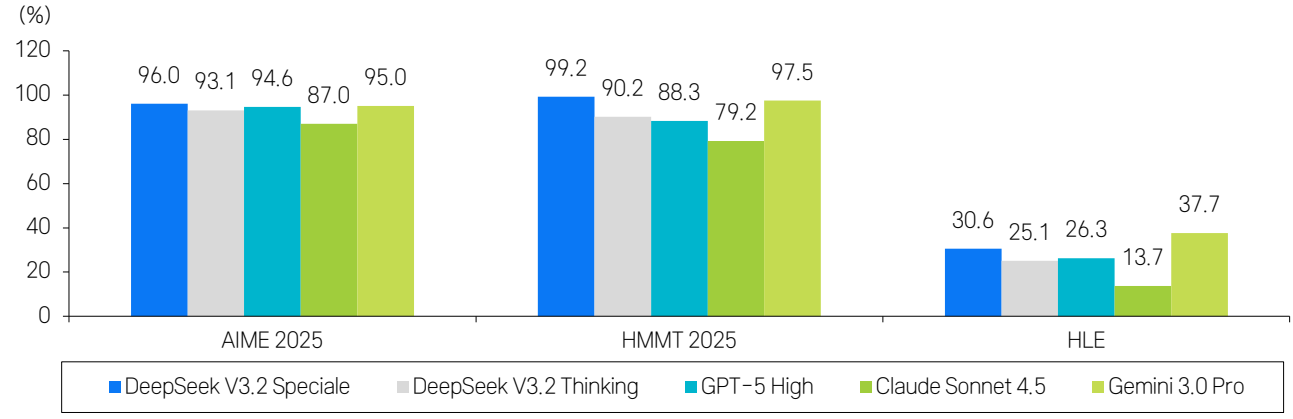
1) 총 학습 FLOPs(연산량)이 적기 때문에 세계 지식 범위(breath of world knowledge)는 여전히 뒤쳐져 있으며, 향후 사전학습 컴퓨트 확장을 통한 지식 격차 해소 계획을 가지고 있다.

2) 모델 아웃풋 퀄리티 상승을 위해 더 많은 토큰이 필요하기 때문에 토큰 효율성이 낮다. 향후 추론 체인 내 지능 밀도(intelligence density)를 최적화해 효율성 향상에 집중할 계획이다.

3) 복잡한 작업 해결 능력은 부족하며, 기초 모델과 사후학습 레시피를 정교화 하려는 동기로 이어진다.

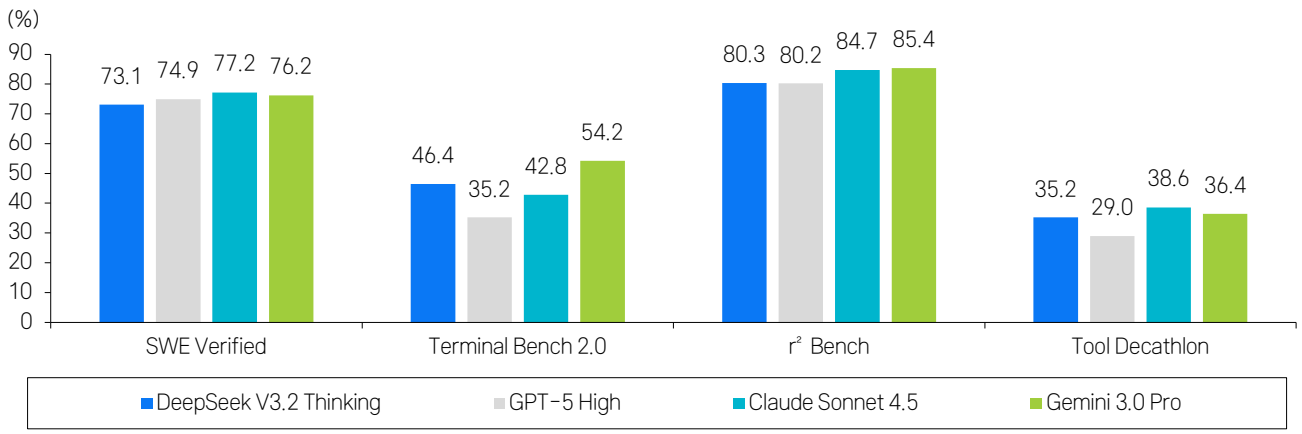
정리해보면, 프론티어 모델의 지식적 우위는 고착화된 본질적 우위가 아니라 컴퓨트(특히 사전학습) 규모 차이에서 기인한 것으로 추론 및 에이전트 능력에서는 유사한 성과를 보이고 있는 만큼, 캐파가 추가되면 효율성을 포함한 성능 측면에서 충분히 따라잡을 수 있다는 자신감으로 해석 가능하다.

추론 능력 평가 벤치마크 기준 딥시크 모델 성능 비교



자료: DeepSeek, 삼성증권

에이전틱 능력 평가 벤치마크 기준 딥시크 모델 성능 비교



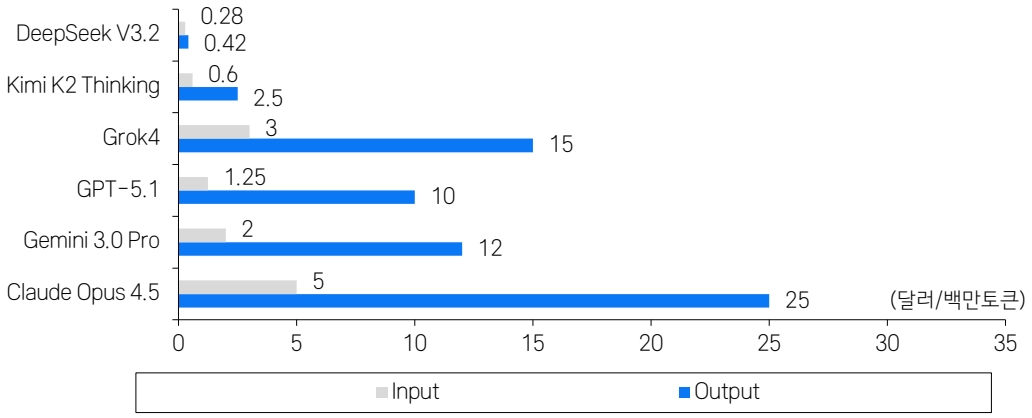
자료: DeepSeek, 삼성증권

딥시크 모델 벤치마크 성능 비교 표

구분	DeepSeek V3.2 Thinking	DeepSeek V3.2 Speciale	GPT-5 High	Gemini 3.0 Pro	Kimi-K2 Thinking
AIME 2025	93.1	96.0	94.6	95.0	94.5
HMMT Feb 2025	92.5	99.2	88.3	97.5	89.4
HMMT Nov 2025	90.2	94.4	89.2	93.3	89.2
IMOAnswerBench	78.3	84.5	76.0	83.3	78.6
LiveCodeBench	83.3	88.7	84.5	90.7	82.6
CodeForces	2,386	2,701	2,537	2,708	-
GPQA Diamond	82.4	85.7	85.7	91.9	84.5
HLE	25.1	30.6	26.3	37.7	23.9

자료: DeepSeek, 삼성증권

딥시크 모델 API 호출 비용 비교(vs 프론티어 모델)



자료: Artificial Analysis, 삼성증권

Compliance notice

- 본 조사분석자료의 애널리스트는 2025년 12월 1일 현재 위 조사분석자료에 언급된 종목의 지분을 보유하고 있지 않습니다.
- 당사는 2025년 12월 1일 현재 위 조사분석자료에 언급된 종목의 지분을 1% 이상 보유하고 있지 않습니다.
- 본 조사분석자료에는 외부의 부당한 압력이나 간섭 없이 애널리스트의 의견이 정확하게 반영되었음을 확인합니다.
- 본 조사분석자료는 당사의 저작물로서 모든 저작권은 당사에게 있습니다.
- 본 조사분석자료는 당사의 동의 없이 어떠한 경우에도 어떠한 형태로든 복제, 배포, 전송, 변형, 대여할 수 없습니다.
- 본 조사분석자료에 수록된 내용은 당사 리서치센터가 신뢰할 만한 자료 및 정보로부터 얻어진 것이나, 당사는 그 정확성이나 완전성을 보장할 수 없습니다. 따라서 어떠한 경우에도 본 자료는 고객의 주식투자의 결과에 대한 법적 책임소재에 대한 증빙자료로 사용될 수 없습니다.
- 본 조사분석자료는 기관투자가 등 제3자에게 사전 제공된 사실이 없습니다.

삼성증권

삼성증권주식회사
서울특별시 서초구 서초대로74길 11(삼성전자빌딩)
Tel: 02 2020 8000 / www.samsungpop.com

삼성증권 Family Center: 1588 2323

고객 불편사항 접수: 080 911 0900

Member of
Dow Jones
Sustainability Indices
Powered by the S&P Global CSA