

AI Weekly

2025년의 AI: 중국의 선전포고와 미국의 대오각성

한종목

chongmok.han@miraeasset.com

김은지

eunji.kim.a@miraeasset.com



이러한 경쟁은 '스푸트니크 모멘트'를 연상시킨다. 1957년 소련이 세계 최초로 인공위성 '스푸트니크 1호'를 발사한 사건은 미국에 큰 충격을 주었고, 이는 미국이 우주 경쟁에서 뒤처지지 않도록 노력하게 만들었다. 마찬가지로, 현재 AI 경쟁은 미국과 중국 사이에 벌어지고 있으며, 이는 양국 모두에게 큰 도전과 기회로 여겨지고 있다.

미국은 AI 분야에서 선도적인 위치를 차지하고 있으며, 특히 빅테크 기업들이 AI 연구에 막대한 자금을 투입하고 있다. 반면 중국은 정부의 강력한 지원과 함께 AI 분야에서의 경쟁력을 빠르게 높이고 있다. 이러한 경쟁은 결국 AI 기술의 발전과 대중화에 기여할 것으로 예상된다.

한편, AI 경쟁은 단순히 기술적 우위를 다투는 것을 넘어, 국가 안보와 경제 경쟁력까지 연결되어 있다. AI 기술은 국방, 외교, 경제 등 다양한 분야에서 활용될 수 있으며, 이는 국가의 미래 경쟁력을 결정짓는 핵심 요소로 여겨지고 있다.

결론적으로, 2025년의 AI 경쟁은 미국과 중국 간의 대결로 치달고 있다. 양국은 각자의 강점을 발휘하며 AI 기술의 발전을 촉진하고 있다. 이러한 경쟁은 AI 기술의 발전과 대중화에 기여할 것으로 기대된다.

Highlight of the Week

I. AI Issue: 2025년 중국의 AI

중국의 DeepSeek AI가 OpenAI의 o1에 필적하는 'R1' 모델을 MIT 라이선스로 공개. 주요 벤치마크에서 o1을 능가하는 성능을 보이며, o1보다 25-30배 저렴한 가격으로 제공. R1은 총 6,710억 개의 파라미터 중 370억 개가 상시 활성화되는 MoE 구조의 DeepSeek V3를 기반으로 만들어짐. R1은 기존 LLM 훈련 방식과 달리 SFT 단계가 최소화되고 강화학습(GRPO)으로 개발.

R1은 '지식 증류' 기법으로 15억에서 700억 파라미터 규모의 소형 모델들도 개발하여 GPT-4o를 가뿐히 능가하는 성능 달성. OpenAI 등이 최신 모델의 세부 추론 과정을 비공개로 유지하는 것은 증류를 통한 기술 유출을 방지하기 위한 전략으로 해석됨.

특히 증류와 강화학습의 결합이 AI 기업들의 핵심 전략이 될 것으로 전망. DeepSeek의 이번 행보는 중국의 AI 기술력을 입증하는 동시에 미국 기업들에 대한 적잖은 압박으로 작용할 전망. 다만 미국 최고 연구소들의 미공개 기술이 여전히 앞서 있을 것으로 분석되나, OpenAI와 Anthropic 등은 자사의 '트럼프 카드'에 관한 압박을 받을 것으로 예상.

II. AI Issue: 2025년 미국의 AI

미국은 AI 칩과 모델 가중치에 대한 새로운 통제 체제를 도입하고, 14/16nm 이하 공정의 반도체와 24종의 제조장비에 대한 수출 통제를 강화. 특히 성숙 공정까지 규제를 확대함으로써 중국의 우회 수입을 차단하고 AI 생태계 전반의 발전을 지연시키려는 의도.

OpenAI는 PhD 수준의 슈퍼 에이전트 출시를 앞두고 있으며, o3 모델은 GPQA에서 87%의 성능을 달성했다고 알려짐. David Shapiro는 o3의 IQ가 145 수준에 도달했다고 평가하며, o3-pro는 150 수준에 육박할 것으로 전망. 네이처에 발표된 AI 칩 설계 연구에서도 AI가 인간을 뛰어넘는 성능을 보여주며, 모든 영역에서 "이세돌 모멘트"가 임박했음을 시사.

메타의 주커버그는 AI의 일자리 대체에 관한 우려를 일으킴. WEF는 향후 5년간 41%의 기업이 AI로 인한 인력 감축을 예상. OpenAI 연구진은 o4나 o5 모델이 AI 연구 개발을 자동화할 수 있는 수준에 도달할 것으로 전망하며, 이는 "재귀적 자기 개선"의 임계점일 듯.

OpenAI는 중국과의 경쟁에서 승리해야 함을 강조한 청사진을 제시. AI 발전의 기하급수적 특성상 뒤처진 국가는 영원히 따라잡지 못할 수 있어, 단순한 기술 경쟁을 넘어 문명의 패러다임을 바꾸는 변화이기 때문. 샘 알트만은 AGI가 2029년 1월 이전에 개발될 것으로 전망을 수정, AI 발전의 가속화를 시사.

III. Paper of the week: MatterGen – 마이크로소프트

과거 신소재 개발은 방대한 양의 물질을 탐색하는 스크리닝 방식에 의존해 왔으나, 이 과정은 시간과 비용이 많이 들었고, 탐색 범위도 10만~100만 개 수준에 불과했음. 마이크로소프트는 이러한 한계를 극복하기 위해 AI 모델 MatterGen을 제시. MatterGen은 고체의 결정 구조를 이해하고 이를 바탕으로 새로운 재료를 설계하는 생성형 AI임.

MatterGen은 확산 모델을 기반으로 결정 구조의 핵심 요소인 원자 종류, 좌표, 격자를 생성함. 이 모델은 기존 연구 대비 2배 이상 높은 78%의 안정적인 구조 생성 성공률을 보였으며, 천만 개의 구조를 생성하면서도 52%의 고유성을 유지했고, 61%는 기존에 알려지지 않은 새로운 구조였음. 이는 MatterGen이 단순 모방을 넘어 진정한 '창조'를 수행함을 의미함.

또한 자성, 전기적 특성, 강도 등을 원하는 대로 조정한 맞춤형 재료 설계의 가능성을 보여줌. 실제로 TaCr2O6라는 물질 합성에 성공하며 가능성을 입증했음. 또한, 희토류를 포함하지 않는 강력한 자성 재료를 설계하여 공급망 위험 문제를 해결하는 실마리를 제시했음.

MatterGen의 등장은 재료 설계 패러다임의 전환을 의미함. 스크리닝에서 생성으로의 변화는, 건초 더미에서 바늘을 찾는 것에서 바늘을 직접 만드는 것과 같은 혁신임. AlphaFold 이후 신약 개발이 활발해진 것처럼, MatterGen은 신소재 및 재료공학 분야의 혁신을 주도할 것으로 기대됨.

표 1. AI 관련 주요 일정

일	월	화	수	목	금	토
19	20	21	22	23	24	25
.	.	.	· APH 실적	· 하이닉스 실적	.	.
26	27	28	29	30	31	1
.	.	· SAP 실적	· Autonomous(~30) · IBM 실적 · TSLA 실적 · MSFT 실적 · META 실적 · NOW 실적	· AAPL 실적(잡) · INTC 실적	· AMZN 실적(예)	.
2	3	4	5	6	7	8
.	· PLTR 실적	· AMD 실적 · GOOGL 실적 · SNAP 실적(잡)	· ARM 실적 · QCOM 실적	· SMIC 실적(예) · NET 실적	· BABA 실적(예) · 네이버 실적	.

자료: Bloomberg, 미래에셋증권 리서치센터

I. AI Issue: 2025년 중국의 AI

1. 중국의 AI 굴기: DeepSeek

2025년은 AI 에이전트(Agentic AI)에 관한 해로만 기억되지는 않을 것이다. 특히 절대로 잊지 말아야 할 점은, 중국의 AI 연구소들이 미국을 얼마나 빨리 따라잡을 수 있을지 보는 것이 매우 무섭고도 흥미로운 올해의 관전 포인트라는 것이다.

2023년에 설립된 중국의 AI 스타트업 DeepSeek AI가 공개한 R1 모델은 그 중심에 서서, 오픈소스 AI의 새로운 가능성을 제시하며 업계에 신선한 충격을 안겨주고 있다. DeepSeek AI의 R1은 OpenAI의 o1에 필적하는 성능을 자랑하면서도, 훨씬 더 저렴한 비용과 개방성이라는 무기를 앞세워 AI 생태계의 지각변동을 예고하고 있다. 본 리포트는 R1의 기술적 특징, 훈련 과정, 성능, 그리고 시사점을 심층적으로 분석하고, 이를 통해 R1이 AI 산업 전반에 미칠 영향과 미래 전망을 제시하고자 한다.

그림 1. DeepSeek AI에서 공개한 정식 추론 모델 “R1”에 대한 논문



DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning

DeepSeek-AI

research@deepseek.com

Abstract




자료: DeepSeek, 미래에셋증권 리서치센터

먼저 작년 11월 말, DeepSeek은 자사 웹사이트에서 R1-Lite Preview라는 모델에 대한 접근권을 제공했었다. R1은 DeepSeek가 독자적으로 개발한 강화학습 기반 추론형 언어모델이다. R1-Lite Preview는 그들의 첫 번째 추론 모델이었고, 이를 시도해본 대부분의 사람들이 매우 놀라워했다. 당시 그들이 언급했던 멋진 점 중 하나는 이 모델의 가중치를 공개할 예정이라는 것이었다. 그리고 이번 주 월요일인 20일 DeepSeek는 약속을 지켰다.

DeepSeek R1 대형 모델뿐만 아니라, 그로부터 추출한 다양한 크기의 작은 모델들까지 여러 “중류 모델”들도 공개했다. 또한 DeepSeek R1 모델의 가중치를 공개했을 뿐만 아니라 모델을 MIT 라이선스로 공개했다. 따라서 기본적으로 원하는 용도로 사용할 수 있다. 심지어 이 모델의 출력을 사용해서 다른 모델을 훈련시키는 것도 허용하고 있다. 즉, 자신만의 모델이 있고 그것을 R1의 출력값으로 강화훈련시키고 싶다면, 그것도 괜찮다는 것이다.

세 가지 벤치마크(AIME, MATH-500, SWE-bench)에서 R1은 확실히 OpenAI의 완전한 o1 모델을 능가하고 있다. OpenAI의 o1-mini에는 모든 영역에서 명확하게 능가하고 있다. 특히 R1은 수학, 과학, 코딩 등 고도의 추론 능력이 요구되는 분야에서 최상위 모델들과 어깨를 나란히 하는 수준으로 평가받고 있다.

그림 2. DeepSeek R1과 최상위권 모델과의 성능 차이(1), '많은 영역에서 o1을 뛰어넘는 R1'

							
Benchmark (Metric)	Claude-3.5-Sonnet-1022	GPT-4o-0513	DeepSeek-V3	OpenAI-o1-mini	OpenAI-o1-1217	DeepSeek-R1	
Architecture	-	-	MoE	-	-	MoE	
# Activated Params	-	-	37B	-	-	37B	
# Total Params	-	-	671B	-	-	671B	
English	MMLU (Pass@1)	88.3	87.2	88.5	85.2	91.8	90.8
	MMLU-Redux (EM)	88.9	88.0	89.1	86.7	-	92.9
	MMLU-Pro (EM)	78.0	72.6	75.9	80.3	-	84.0
	DROP (3-shot F1)	88.3	83.7	91.6	83.9	90.2	92.2
	IF-Eval (Prompt Strict)	86.5	84.3	86.1	84.8	-	83.3
	GPQA Diamond (Pass@1)	65.0	49.9	59.1	60.0	75.7	71.5
	SimpleQA (Correct)	28.4	38.2	24.9	7.0	47.0	30.1
	FRAMES (Acc.)	72.5	80.5	73.3	76.9	-	82.5
	AlpacaEval2.0 (LC-winrate)	52.0	51.1	70.0	57.8	-	87.6
	ArenaHard (GPT-4-1106)	85.2	80.4	85.5	92.0	-	92.3
Code	LiveCodeBench (Pass@1-COT)	38.9	32.9	36.2	53.8	63.4	65.9
	Codeforces (Percentile)	20.3	23.6	58.7	93.4	96.6	96.3
	Codeforces (Rating)	717	759	1134	1820	2061	2029
	SWE Verified (Resolved)	50.8	38.8	42.0	41.6	48.9	49.2
	Aider-Polyglot (Acc.)	45.3	16.0	49.6	32.9	61.7	53.3
Math	AIME 2024 (Pass@1)	16.0	9.3	39.2	63.6	79.2	79.8
	MATH-500 (Pass@1)	78.3	74.6	90.2	90.0	96.4	97.3
	CNMO 2024 (Pass@1)	13.1	10.8	43.2	67.6	-	78.8
Chinese	CLUEWSC (EM)	85.4	87.9	90.9	89.9	-	92.8
	C-Eval (EM)	76.7	76.0	86.5	68.9	-	91.8
	C-SimpleQA (Correct)	55.4	58.7	68.0	40.3	-	63.7

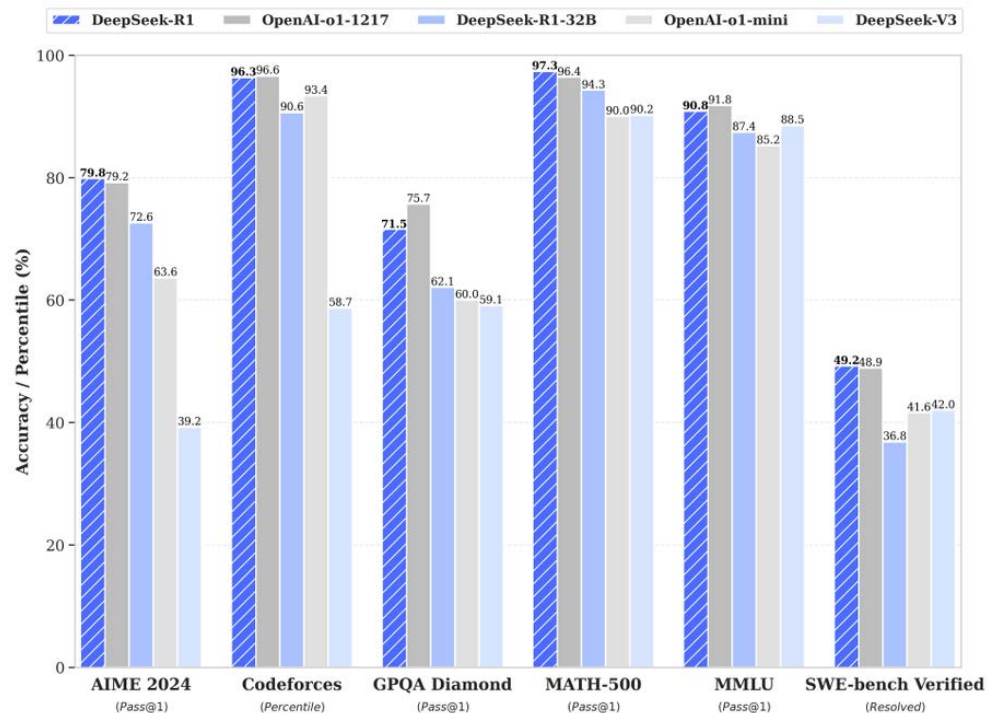
자료: DeepSeek, 미래에셋증권 리서치센터

흥미로운 점은 **DeepSeek R1이 실제로 많은 작업에서 본인들의 non-reasoning 모델인 DeepSeek V3를 기본 모델로 사용하여 훈련을 시작했다**는 것이다. DeepSeek V3는 기본적으로 전문가 혼합(MoE: Mixture of Experts) 모델이다. 총 6,710억 개의 파라미터를 가지고 있으며, 이 중 언제든 370억 개의 파라미터가 활성화되어 있다. 이는 R1에서도 마찬가지다. DeepSeek V3에 대해서는 AI Weekly에서도 최근 다룬 적이 있고, 많은 개발자들에게 찬사를 받은 매우 효율적인 모델이다.

DeepSeek V3와 이번 R1 모델의 경우, 사전 훈련에 있어서 동일한 모델이지만, 단지 강화 학습을 어떻게 적용했는지, 그리고 다단계 훈련에서 다른 것들을 어떻게 적용했는지에 따라 위와 같이 벤치마크 성능표에서 보듯 그 실력 차이가 발현되었다고 볼 수 있다.

다시 말해, 여기서 주요 시사점 중 하나는 이것이 **단순히 더 크거나 새로운 모델이라는 것이 아니라, 사후학습(post-training)을 다르게 하는 것만으로도 이러한 놀라운 결과를 얻을 수 있다는 것이다**. 그리고 이는 많은 면에서 OpenAI가 o1과 o3 모델들에 대해 공개적으로 논의한 내용과 일치한다.

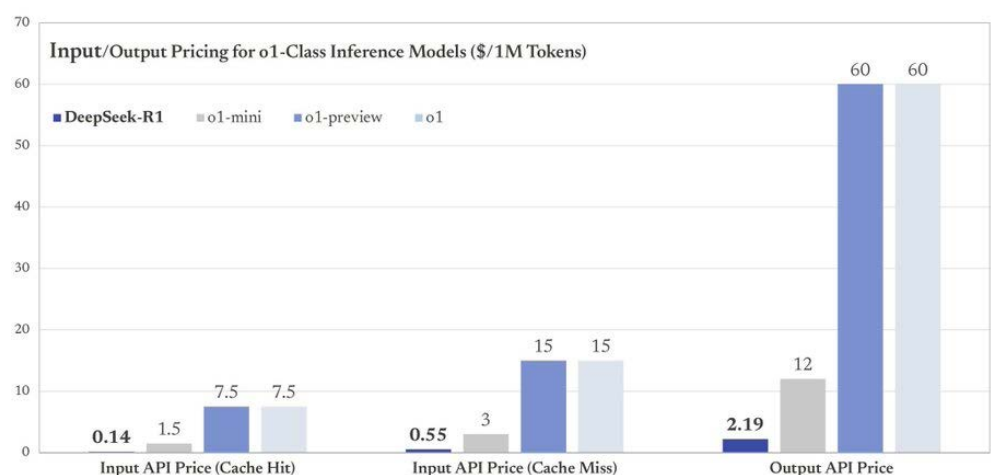
그림 3. DeepSeek R1과 최상위권 모델과의 성능 차이(2), 'R1과 V3의 차이는 매우 큼... 이유는?'



자료: DeepSeek, 미래에셋증권 리서치센터

R1은 OpenAI의 o1과 비교했을 때, 대부분의 벤치마크에서 동등하거나 더 나은 성능을 보였으며, 특히 비용 측면에서 압도적인 우위를 점하고 있다. OpenAI o1과 DeepSeek AI R1의 가격 비교를 해보면, R1이 모든 범주에서 훨씬 저렴하다. 거의 96~98%의 "파격 세일"이라 할 수 있다. 즉, **DeepSeek는 25~30배 저렴한 가격이면서도 o1 성능과 일치한다.**

그림 4. DeepSeek R1과 OpenAI o1 시리즈의 API 비용, '비교가 무의미할 정도로 저렴'



자료: DeepSeek, 미래에셋증권 리서치센터

DeepSeek을 테스트해보고 싶다면 chat.deepseek.com에 접속하면 된다. 실제로 사용해 볼 경우 흥미로운 점은 답이 맞고 틀리고가 아니라, 이것이 생각하는 과정을 보여주는 방식이다. R1 모델이 무언가를 진술하기 시작하면, 종종 다른 것들에 대해 스스로 명확히 하고,

때로는 확실히 하기 위해 뒤로 돌아가기도 한다. "제가 확인해야 할 것이 있습니다"와 같은 말을 하기도 하고, 그런 다음 무엇을 해야 할지 알아낸다. 또한 모델은 "이전 대화를 되돌아보면서 시작하겠습니다"라고 말하기도 한다. 즉, 컨텍스트 윈도우에 있는 내용을 살펴보고 있는 것이다. 확실히 일종의 내적 대화 같은 것이다.

사실 OpenAI의 o1 모델도 이러한 내부 독백을 공개하기는 했으나, 이것은 상당히 많이 축약된 버전이고 가공되지 않은 CoT는 아니다. 그러나 이번에 R1은 날 것 그대로의 사고 과정을 "마치 인간이 큰 소리로 생각하는 것처럼" 제시했다. 진짜 "chain of thought"는 단순히 기술적인 측면을 넘어, 사용자에게 심리적 영향도 미친다. 답변뿐만 아니라, 모델이 답변에 이르기까지 사고하는 그 자체가 마치 망설임이나 불안함과 같은 인간적인 면모를 보여, 사용자로 하여금 AI 모델에 대한 새로운 차원의 '공감'을 불러일으킬 수 있다.

2. DeepSeek R1 논문 리뷰

DeepSeek는 R1을 출시하면서 공식 paper인 "DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning" 논문을 내놨다. 해당 논문을 살펴보면, 두 가지 주요 모델, 즉 DeepSeek R1-Zero와 DeepSeek R1이 이번에 새롭게 제시됐다.

이번에 주로 집중해야 할 것은 DeepSeek R1이다. 이것이 둘 중 더 나은 모델이기 때문이지만, **DeepSeek R1을 만들기 위해서 애초에 DeepSeek R1-Zero이 나온 것이기도 하다.**

그러나 먼저 기술할 DeepSeek R1-Zero는 후처리 없이 사전 훈련된 DeepSeek V3 모델을 가져와서 여기에 강화학습을 실행하여 만들어진 것이다. 전통적인 LLM 훈련에서 사람들이 일반적으로 하는 것은, 먼저 사전 훈련을 하고 그 다음에 일종의 SFT를 하고, 그 후에 RL로 정렬 훈련을 하는 것이다. 이것은 RLHF일 수도 있고, RLAIFF일 수도 있으며, DPO와 같은 대안일 수도 있다.

R1의 아이디어는 그 SFT 단계를 하지 않는다는 것이다. 지도 미세 조정(SFT)을 거치지 않았다는 것은, **인간이 만든 예시 데이터에 의존하지 않고 순수 RL로만 훈련된 모델에 적용되었다**는 뜻이다. AlphaGo가 수많은 기보 학습 없이 순수 강화학습만으로 바둑을 마스터한 사례에 빗대어 설명할 수 있다. 이제 **우리는 LLM 강화학습의 시대로 접어들고 있다**"고 해야 하지 않을까 싶다.

- SFT(지도 미세조정): 모범 답안을 보면서 학습시키는 방식. 인간 개발자가 생성한 고품질의 데이터셋을 사용하여 기본 모델을 특정 task 또는 도메인에 맞게 미세 조정(fine-tuning).

예를 들어, 질의응답 task를 위한 SFT 데이터셋은 질문과 정답 쌍으로 구성. 모델이 모범 답안에만 의존하기 때문에 고품질의 대규모 SFT 데이터셋이 필요하다는 단점.

- RL(강화학습): 시행착오를 통해 스스로 터득하는 학습. 모델은 환경과 상호작용하며, 보상(reward)을 최대화하는 방향으로 자신의 행동(policy)을 개선. 게임을 하면서 스스로 공략법을 터득하는 게이머와 유사.

보통 인간 피드백(RLHF) 또는 AI 피드백(RLAIF)을 보상(reward)으로 사용하여 모델이 인간의 선호도에 맞는 출력을 생성하도록 유도. 스스로 최적의 전략을 탐색하기 때문에, 창의적이고 새로운 답변을 생성할 수 있으나, SFT에 비해 학습 속도가 느릴 수 있음.

이와 관련해, DeepSeek는 R1에게 명시적인 지식이나 정답을 제공하는 대신에(SFT 대신에), 모델이 스스로 사고의 연쇄 과정을 생성하기 시작하도록 설정한다. 이를 위해서 DeepSeek AI에서는, AI 모델이 사고의 연쇄를 이끌어내기 위해 여기서 정말 흥미로운 일종의 프롬프트 템플릿을 사용했다.

즉, 모델에게 단순히 질문을 던지는 것이 아니라, "생각"을 이끌어낼 수 있는 특정 구조의 프롬프트를 사용한 것이다. DeepSeek AI가 이번에 제공한 프롬프트 템플릿은 사용자의 질문과 어시스턴트의 답변, 그리고 어시스턴트의 사고과정까지 모두 포함하는 대화식으로 (질문-답변-추론의 프레임에 갖도록) 구성되어 있다. 다시 말해, 모델에게 사용자의 질문을 해결하는 어시스턴트의 역할을 부여함으로써, 문제 해결 과정을 자연스럽게 유도하는 것이다.

그림 5. DeepSeek가 R1-Zero를 훈련하기 위해 준 프롬프트 템플릿(=스스로 추론을 하도록 유도)

A conversation between User and Assistant. The user asks a question, and the Assistant solves it. The assistant first thinks about the reasoning process in the mind and then provides the user with the answer. The reasoning process and answer are enclosed within `<think>` `</think>` and `<answer>` `</answer>` tags, respectively, i.e., `<think>` reasoning process here `</think>` `<answer>` answer here `</answer>`. User: **prompt**. Assistant:

Table 1 | Template for DeepSeek-R1-Zero. **prompt** will be replaced with the specific reasoning question during training.

자료: DeepSeek, 미래에셋증권 리서치센터

그리고 시간이 지남에 따라, 단계가 올라갈수록 이 작업에서의 정확도도 올라가고 있다는 점이 확인되었다. 수많은 시도와 실패를 통해 R1은 스스로 추론 과정을 정립하고, 최적의 해결책을 찾아가는 법을 터득한 것이다. 이것이 바로 R1의 훈련 과정에서 주목할 만한 현상인 '아하! 모멘트'라고 할 수 있다. 실제로 DeepSeek는 R1-Zero가 어떤 수학 문제를 풀다가 스스로 멈추고 "잠깐, 잠깐. 잠깐. 내가 여기서 '아하 모멘트'를 포착할 수 있을 것 같아(Wait, wait. Wait. That's an aha moment I can flag here)"라고 외친 순간에 대한 본인들의 경험담을 논문에다가 공유했다. 이 순간은 DeepSeek-R1-Zero가 단순히 주어진 데이터를 학습하는 것을 넘어, 스스로 생각하고 추론하는 능력을 갖추고 있음을 보여주는 상징적인 장면이자 순수 RL의 잠재력을 보여주는 생생한 증거라고 할 수 있다.

그림 6. DeepSeek가 AI를 훈련하며 얻은 통찰 "아하! 모멘트"(AI가 추론하는 능력을 갖춘 순간)

Question: If $a > 1$, then the sum of the real solutions of $\sqrt{a - \sqrt{a + x}} = x$ is equal to
 Response: `<think>`
 To solve the equation $\sqrt{a - \sqrt{a + x}} = x$, let's start by squaring both ...
 $(\sqrt{a - \sqrt{a + x}})^2 = x^2 \implies a - \sqrt{a + x} = x^2$.
 Rearrange to isolate the inner square root term:
 $(a - x^2)^2 = a + x \implies a^2 - 2ax^2 + (x^2)^2 = a + x \implies x^4 - 2ax^2 - x + (a^2 - a) = 0$
 ...
Wait, wait. Wait. That's an aha moment I can flag here.
 Let's reevaluate this step-by-step to identify if the correct sum can be ...
 We started with the equation:
 $\sqrt{a - \sqrt{a + x}} = x$
 First, let's square both sides:
 $a - \sqrt{a + x} = x^2 \implies \sqrt{a + x} = a - x^2$
 Next, I could square both sides again, treating the equation: ...
 ...

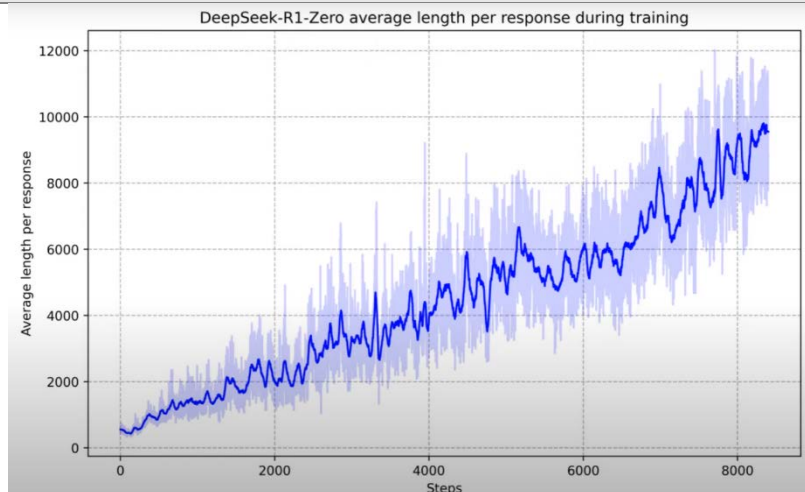
Table 3 | An interesting "aha moment" of an intermediate version of DeepSeek-R1-Zero. The model learns to rethink using an anthropomorphic tone. This is also an aha moment for us, allowing us to witness the power and beauty of reinforcement learning.

자료: DeepSeek, 미래에셋증권 리서치센터

이 현상이 AI 연구에 가져올 파급력에 주목할 필요가 있다. "아하 모멘트"가 AI 모델에게 뿐만 아니라 이를 관찰하는 연구자들에게도 '깨달음의 순간'일 수 있기 때문이다. AI 모델의 자율 학습능력과 고급 추론기술의 자발적 발현은 AI 연구의 큰 주제가 될 것이다.

DeepSeek R1-Zero는 이 강화학습만으로도 o1 모델을 능가하고 있는 것을 볼 수 있다. 훈련이 길어질수록 더 긴 사고의 연쇄와 더 긴 종류의 사고 과정을 만들어낸다는 것이다. 처음에는 응답이 실제로 매우 짧았지만, 8,000 스텝 정도가 되면 더 긴 사고의 연쇄(CoT)를 생성하는 경향을 보인다. 이는 모델이 훈련 과정에서 스스로 더 복잡한 추론 과정을 학습하고 있음을 시사한다.

그림 7. DeepSeek R1-Zero이 강화학습을 지속 수행하자 더 높은 수준의 추론능력이 점차 상승



자료: DeepSeek, 미래에셋증권 리서치센터

DeepSeek의 이번 모델이 사용하는 강화학습의 유형은 GRPO 알고리즘이다. R1-Zero도 GRPO로 탄생했다. GRPO는 그들의 DeepSeek의 작년 논문인 “DeepSeek-Math”에서 나온, 독자개발한 개념이다. GRPO 알고리즘의 핵심 작동 원리는 "멀티 샘플링"과 "그룹 상대 평가" 라고 간결하게 요약할 수 있다. 다시 말해, AI 모델에게 여러 가지 답변을 생성할 수 있게 하고 그 중에서 평균적으로 가장 좋은 답변이 무엇인지 정규화한 다음 이 중에서 좋은 답변을 고르는 최적화 알고리즘이다. 여기서 GRPO 알고리즘 자체가 여러 샘플(답변)을 생성하고, 그 중 가장 좋은 것을 선택하는 "best-of-N" 방식과 유사하다고 볼 수 있다.

무엇보다, R1 논문에서는 R1-Zero가 "생각할 시간"을 늘려가는 과정에서 "reflection(반성)"과 "exploration(탐색)"을 한다는 점을 발견했다고 밝혔다. 이것은 DeepSeek-R1-Zero가 여러 추론 경로를 고려한다는 것을 뜻한다. 즉, OpenAI의 o1의 방식처럼 MCTS와 같은 복잡한 트리 탐색 알고리즘을 사용하지는 않지만, search를 통한 "추론 컴퓨트 스케일링(Test-Time Compute)" 개념이 어느 정도 적용되었다고 볼 수 있다.

참고로, DeepSeek AI는 OpenAI 때문에 유명해진 PRM(Process Reward Model) 방식과 MCTS(Monte Carlo Tree Search)를 사용하려다가 실패했던 경험을 솔직하게 공유했다. 단순히 성공 사례만을 포장하는 것이 아니라, 어떤 방법이 효과가 없었는지에 대한 귀중한 통찰력까지 제공한다는 점에서 기술력의 진정성이 더욱 크다고 평가되는 대목이다. PRM은 일반적인 추론 task에서 "단계"를 명확하게 정의하는 것이 어렵고, MCTS를 적용하기 위해서는 AI가 답변을 탐색할 공간을 효과적으로 줄이는 방법이 필요하나 이게 쉽지 않은 문제라고 DeepSeek는 말했다. 사실 바로 이 부분이 그만큼 어려운 영역인만큼 OpenAI와 같은 선도적인 AI 연구실의 강점으로 봐야 하는 부분이라 생각된다. 앞으로도 우리 팀은 이 영역의 혁신을 이뤄내는 기업들에 대해 더욱 주목할 필요성을 느꼈다.

- GRPO: 별도의 Critic 모델없이, 여러 샘플(답변)들을 그룹으로 묶어 상대적으로 비교하고, 그 차이(advantage)를 최대화하는 방향으로 policy를 업데이트하는 알고리즘. 일종의 선생님과 비슷한데, 학생들에게 동일한 문제를 풀게 하고, 답안들을 서로 비교하여, 누가 더 잘했는지, 어떤 부분이 부족한지 평가한 뒤, 그 평가 결과를 바탕으로 학생들을 지도하는 방식.

GRPO는 PPO 대비 계산비용이 적고, 학습 속도가 빠른 빠르고 개별 샘플에 대한 보상값의 편차에 덜 민감하다는 장점. 실제로 동일 하드웨어에서 PPO 대비 30% 적은 메모리 사용했고 GRPO 적용 후 DeepSeekMath 7B의 정확도가 46.8% → 51.7%로 향상되었다고 논문에 적혀있음. 단, 애초의 샘플링된 데이터들의 품질이 안 좋다면 학습 자체가 어려워질 수 있음.

표 2. 엔비디아의 수석 AI 과학자 Jim Fan “DeepSeek는 GRPO라는 독자적 강화학습 최적화 알고리즘을 만들었다. 참 대단한 팀이다.”

특징	PPO(주류 최적화 알고리즘)	GRPO(DeepSeek가 독자 개발한 알고리즘)
비평가 모델(Critic)	사용 (별도의 신경망)	필요 없음 → 메모리 사용량 50% 감소
Advantage 계산	Critic의 가치 함수 기반	그룹 내 상대 평가 기반
효율성	Critic 학습 및 추론으로 인한 추가 비용	Critic 제거로 인한 효율성 향상

자료: DeepSeek, 미래에셋증권 리서치센터

그러나 R1-Zero는 추론 능력은 뛰어났지만, 가독성이나 일관성 측면에서 부족한 점이 있었고, 이는 논문에서도 언급된 내용이다. R1-Zero의 한계를 극복하기 위해, DeepSeek R1은 "cold-start data"를 이용한 SFT를 추가하고, 다단계 훈련 파이프라인을 도입했다.

즉, R1은 R1-Zero와 달리, 인간이 생성한 고품질 CoT 데이터(수천 개의 예시 데이터)를 초기 훈련에 활용하기는 한다. 사용자가 받아들이기 가독성과 일관성을 향상시키기 위함이다. 다만, R1 훈련의 핵심은 여전히 RL이며, SFT는 초기 훈련을 돕는 보조적인 역할을 수행하고 SFT는 RL 단계로 넘어가기 위한 초석일 뿐이다. 그런 다음 DeepSeek R1-Zero처럼 다시 GRPO 알고리즘으로 강화학습을 하고, 또 다시 SFT(지도 미세조정)을 하고, 또 마지막 단계에서 전통적인 방식의 강화학습을 또 한 번 진행한다.

즉, R1 훈련 과정에서 RL과 SFT가 번갈아 가며 적용됨을 보여준다. 이를 종합하면 R1 훈련 과정은 5단계로 구성되고, 각 단계는 이전 단계의 결과를 기반으로 점진적으로 모델을 개선하는 방식으로 설계된 것이다. SFT를 통해 효율적으로 기본기를 다지고, RL을 통해 자가 학습하고 발전하는 능력을 갖추도록 훈련된 방식인 셈이다. 세부 단계는 아래와 같다.

- R1 훈련: R1-Zero → 소량의 SFT(수천개의 예시) → RL (GRPO) → 대규모 SFT(60만개의 예시) → RL (GRPO)

- 0단계(R1-Zero): 순수 RL만으로 훈련하여, 스스로 추론 능력을 개발하도록 했다. SFT 없이, DeepSeek-V3-Base 모델을 GRPO 알고리즘을 사용한 강화학습만으로 훈련하여 R1-Zero 모델을 생성. 특히, AIME 2024 벤치마크에서 15.6%에서 71.0%로 비약적인 성능 향상을 보였으며, majority voting을 적용할 경우 86.7%의 정확도를 달성하여 o1-0912 모델과 동등한 수준에 도달. SFT 없이 순수 강화학습만으로도 LLM의 추론 능력을 크게 향상시킬 수 있음을 보여주는 중요한 단계. 훌륭한 추론 능력을 보여주지만, 그 결과물을 인간이 이해하기 쉬운 형태로 표현하는 데는 어려움.

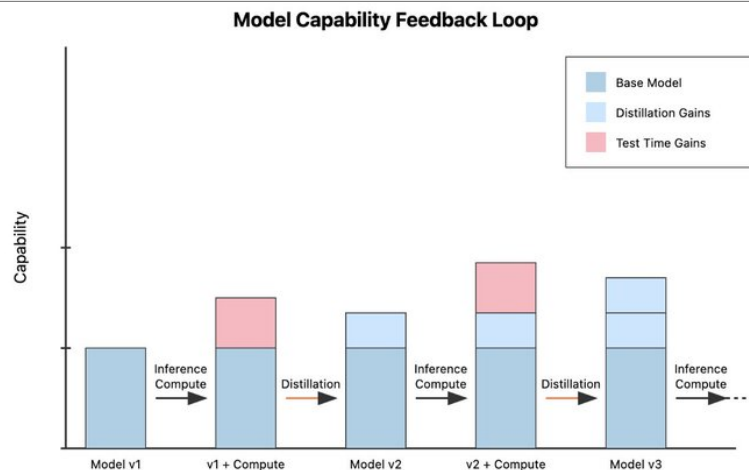
- 1단계, 3단계: R1-Zero를 사용하여 긴 CoT 샘플을 생성하고, 이를 바탕으로 이전 단계의 모델을 미세 조정. 즉, SFT를 통해 모델의 기본적인 능력을 향상시키고, 가독성을 개선해 구조화된 출력을 생성하는 능력을 향상. 작문, 질의응답, 자기 인식 등 다양한 능력이 향상.

- 2단계, 4단계: RL(GRPO)을 통해 추론 능력을 극대화. 수학, 코딩, 과학, 논리 추론과 관련된 벤치마크에서 높은 성능을 기록. 마지막 RL에서는 창의적인 글쓰기처럼 명확한 규칙을 적용하기 어려운 "non-reasoning" 업무에서는 이전에는 사용하지 않았던 DeepSeek-V3 모델을 동원해 인간이 선호하는 출력을 생성하도록 훈련.

해당 논문 리뷰를 마무리하면서 증류에 대해 살펴보겠다. 증류(Distillation)는 대규모 모델(teacher)의 지식을 소규모 모델(student)에게 전달하는 기술로, 이를 통해 소규모 모델도 대규모 모델에 버금가는 성능을 낼 수 있도록 훈련하는 방법이다. DeepSeek는 R1의 뛰어난 추론 능력을 더 작은 모델에 전파하기 위해 지식 증류기법을 활용했는데, R1을 교사 모델로 사용하여 80만 개의 고품질 데이터를 생성하고, 이를 사용하여 Qwen 및 Llama와 같은 작은 모델들을 학생 모델로 훈련했다.

특히, "증류된 모델에 강화학습을 적용하면 상당한 추가 이득을 얻을 수 있다"는 논문의 내용은 몇 번이고 강조해도 지나치지 않을 정도다. 증류와 RL의 결합이 앞으로 거의 모든 오픈소스 AI 기업들의 모델 훈련의 핵심 전략이 될 수 있기 때문이다. 즉, 대형 모델을 훈련할 수 있는 자원이 없는 연구자나 기업도 소형 모델을 통해 고성능 AI의 혜택을 누릴 수 있을 것이다. 이는 한국에게도 적용이 될 부분으로 사료된다.

그림 8. 모델의 성능이 어떻게 향상되는지의 핵심은 "Distillation Gains"와 "Test Time Gains"



자료: X(@techno_guile), 미래에셋증권 리서치센터

- *Distillation Gains*: R1과 같은 대규모 모델에서 학습된 지식을 더 작은 모델로 "증류"함으로써 얻는 성능 향상. 즉, 큰 모델의 능력을 작은 모델에 효과적으로 전달해, 작은 모델도 큰 모델 못지않은 성능을 낼 수 있도록 함.
- *Test Time Gains*: 모델이 추론 과정에서 더 많은 계산을 수행함으로써 얻는 성능 향상. 예를 들어, R1-Zero는 더 많은 "사고 시간"을 할애함으로써, 즉 더 많은 추론 단계를 거침으로써, 더 정확한 답변을 생성하게 함.

이번에 공개된 R1 시리즈 모델들을 직접 로컬에서 실행해보고 싶다면, 아마도 가장 큰 버전은 실행하지 못할 것이다. 전체 크기의 R1 버전은 6,710억 개의 파라미터를 가지고 있기 때문이다. 하지만 작은 증류 모델들을 사용해볼 수는 있다. DeepSeek R1 논문에 따르면, 가장 작은 모델은 15억 개의 파라미터를 가지고 있다. 뿐만 아니라 15억에서 700억까지 다양한 크기를 가지고 있다. 이 중에서 15억과 70억짜리 모델은 구글의 CoLab 환경에서도 쉽게 실행할 수 있고, 이와 관련한 튜토리얼 영상도 벌써 유튜브에 많이 게시되어 있을 정도로 이미 '바이럴'을 타고 있다.

그런데도 GPT-4o나 Claude 3.5 Sonnet뿐만 아니라, 더 큰 모델들은 심지어 o1-mini 모델도 능가하는 성능을 낸다는 것을 볼 수 있다. DeepSeek-R1-Distill-Qwen-7B 모델은 AIME 2024에서 55.5%, MATH-500에서 92.8% 정확도를 기록하며, GPT-4o-0513 (AIME 2024: 9.3%, MATH-500: 74.6%)과 같은 대형 모델을 능가하는 성능을 보였다.

R1의 추론 능력이 작은 모델에도 효과적으로 전달될 수 있음을 보여주는 강력 증거다.

표 3. DeepSeek R1 논문에서 밝힌 공개된 증류 버전들의 종류와 메모리 필요 용량

모델	파라미터 수	용량(Float32)	용량(Ollama 기준 4비트 양자화 적용)
DeepSeek-R1-Distill-Qwen-1.5B	1.5B (15억)	~6 GB	~1.1 GB
DeepSeek-R1-Distill-Qwen-7B	7B (70억)	~28 GB	~4.1 GB
DeepSeek-R1-Distill-Qwen-14B	14B (140억)	~56 GB	~8.2 GB
DeepSeek-R1-Distill-Qwen-32B	32B (320억)	~128 GB	~18 GB
DeepSeek-R1-Distill-Llama-8B	8B (80억)	~32 GB	~4.7 GB
DeepSeek-R1-Distill-Llama-70B	70B (700억)	~280 GB	~40 GB

자료: DeepSeek, 미래에셋증권 리서치센터

Float32 용량은 모델의 파라미터를 부동소수점 32비트 형식으로 저장했을 때의 용량 / 용량(Ollama)은 Ollama 플랫폼(모델 실행을 간편하게 해주는 도구)에서 실행하는 데 필요한 용량

그런데 흥미로운 점은 Qwen-32B-Base 모델에 1만 단계의 강화학습을 적용한 결과 (DeepSeek-R1-Zero-Qwen-32B)가 R1으로부터 증류한 모델(DeepSeek-R1-Distill-Qwen-32B)보다 성능이 떨어지는 것으로 나타난 것이다. 다시 말해, **R1과 같은 더 강력한 모델을 더 작은 기본 모델(Qwen/Llama)에서 증류하는 것이 해당 기본 모델 자체(예: QwQ-32B)에다가 대규모 RL 훈련을 수행하는 것보다 성능이 뛰어나다**는 말이다.

그림 9. DeepSeek R1의 여러 증류버전들의 여러 수학/공학/코딩 관련 벤치마크 점수

	AIME 2024 pass@1	AIME 2024 cons@64	MATH- 500 pass@1	GPQA Diamond pass@1	LiveCodeBench pass@1	CodeForces rating
GPT-4o-0513	9.3	13.4	74.6	49.9	32.9	759.0
Claude-3.5-Sonnet-1022	16.0	26.7	78.3	65.0	38.9	717.0
o1-mini	63.6	80.0	90.0	60.0	53.8	1820.0
QwQ-32B	44.0	60.0	90.6	54.5	41.9	1316.0
DeepSeek-R1-Distill-Qwen-1.5B	28.9	52.7	83.9	33.8	16.9	954.0
DeepSeek-R1-Distill-Qwen-7B	55.5	83.3	92.8	49.1	37.6	1189.0
DeepSeek-R1-Distill-Qwen-14B	69.7	80.0	93.9	59.1	53.1	1481.0
DeepSeek-R1-Distill-Qwen-32B	72.6	83.3	94.3	62.1	57.2	1691.0
DeepSeek-R1-Distill-Llama-8B	50.4	80.0	89.1	49.0	39.6	1205.0
DeepSeek-R1-Distill-Llama-70B	70.0	86.7	94.5	65.2	57.5	1633.0

자료: DeepSeek, 미래에셋증권 리서치센터

그러니까 이제 **AI 모델의 성능을 더욱 올리기 위해 가장 필요해지게 되는 것은, 더욱 강력한 "기본 모델"과 더 큰 규모의 강화학습이라는 것과 같다.** 따라서, 이러한 증류 방식의 막강함 때문에, 아마도 OpenAI와 같은 선도 기업들이 최신 모델(o1 pro와 o3)의 세부 추론 과정(real 사고의 연쇄)을 공개하지 않으려 한 것이라는 합리적인 추측을 해볼 수 있다.

만약 그 추론 과정이 공개되면 누구나 증류를 통해 그 능력을 작은 모델에 이식할 수 있고, 그렇게 되면 최신 모델을 비공개로 유지함으로써 얻는 경쟁 우위가 사라지기 때문이다.

결론적으로, **OpenAI o1과 같은 선도적인 reasoning 모델이 계속 앞서 나가게 할 근본적인 이유는, 배타적인 사고의 연쇄 과정을 통해 효과적인 증류 기술을 활용할 수 있기 때문일 수 있다.** DeepSeek는 이에 대한 제대로 된 패스트 팔로워이다. 그리고 오픈소스 정책을 통해 “메기 역할”의 주인공까지 맡게 됐다.

3. 오픈소스 진영을 이끌어 가는 DeepSeek, 그리고 중국

R1의 종류 효과성과 오픈소스 공개는 AI 생태계에 거대한 변화를 가져올 것으로 예상된다. R1은 누구나 자유롭게 접근하고 활용할 수 있는 고성능 AI 모델로서, AI 기술의 민주화를 촉진하고, AI 연구 개발의 새로운 지평을 열 것이며, AI 산업 생태계의 지각변동을 가져올 수 있기 때문이다. 앞으로 R1을 기반으로 한 다양한 연구와 혁신이 이어질 것이며, R1과 같은 고성능 AI 모델들이 지속해서 등장할 것으로 예상된다.

DeepSeek의 행보가 마치 "OpenAI의 본래 미션"인 개방적인 연구를 이어가는 것처럼 보여 놀랍다. 그러나 한 가지 짚어보자면, 중국이 최고가 됐다는 일각의 hype 섞인 결론으로 곧바로 도달할 필요는 없다고 본다. 여전히 최고의 미국 연구소들은 지금 가지고 있는 것을 출시를 하고 있지 않기 때문이다. 그들은 여전히 앞서 있다고 생각한다. (이후 장에서 기술)

하지만, **미국 기업들에게 더 많은 압력을 가하고 있다는 것도 사실일 것이다.** OpenAI, Anthropic 등은 이제 본인들의 "트럼프 카드"를 공개해야 하는 압박감을 느낄 것이다.

AI는 맨해튼 프로젝트 2.0으로 간주된다. 분명한 것은 양측 모두 AGI 및 ASI 경쟁에서 1위를 차지하도록 놔두고 싶어하지 않는다는 것이다. R1의 출시는 이러한 배경에서 읽어야 한다. 이제 중국은 AI 영역에서도 선전포고를 한 것이다. **2025년 1월 20일, DeepSeek의 CEO인 량원펑(梁文峰)은 DeepSeek-R1 모델을 공개한 직후 중난하이로 이동해 리창 총리의 정부업무보고 좌담회에 참석했다. 그는 LLM 관련 기업 중 유일한 참석자였다.** 이 부분을 좀 더 파고들면, 中정부는 DeepSeek R1 출시와 오픈소스 공개에 관한 승인을 내렸을 가능성이 높다는 것이다. OpenAI가 정부 관료들과 만나 의사결정을 내리는 것과 같다.

그림 10. DeepSeek의 CEO 량원펑이 R1 모델을 공개한 직후 리창 총리를 알현한 자리



자료: CCTV, 미래에셋증권 리서치센터

이와 유사하게, 중국은 지난 12월 26일 모택동 탄생일을 맞아 두 가지 종류의 6세대 전투기를 전격적으로 공개해 미국을 경악케 한 적이 있음을 주목할 필요가 있다. 일론 머스크는 이를 두고 미국과 중국이 "스푸트니크 모멘트(Sputnik Moment)"를 맞이했다고 평가했다. **트럼프 정부의 "기술 책사"라고 평가되는 일론 머스크의 "스푸트니크 모멘트" 선언은 경쟁국이 갑자기 기술적 돌파를 선보였을 때 이를 따라잡아야 한다는 경각심이 드는 순간을 의미할 것이다. AI가 점점 더 국가적 문제가 되고 있다는 점을 반드시 유념해야 한다.**

- 스푸트니크 모멘트는 1957년 10월 4일 소련이 지구 궤도에 진입한 최초의 인공위성 스푸트니크 1호를 발사해 미국을 깜짝 놀라고 조금하게 만든 데서 유래된 표현.