

OPINION

연구위원
노성호

DeepSeek(深度求索) AI와 금융산업의 혁신*

최근 주목받고 있는 DeepSeek(深度求索) 인공지능(Artificial Intelligence: AI) 모형의 성공은 생성형 AI 개발 경쟁 구도뿐만 아니라 금융 산업의 연구개발(R&D) 전략에도 중대한 변화를 예고하고 있다. DeepSeek은 대형 언어 모형(Large Language Model: LLM)의 현재까지 가장 발전된 형태인 추론 모형(reasoning model)의 일종으로, 다양한 혁신 기술을 접목하여 높은 학습 효율성을 확보하고, 비교적 적은 비용으로도 뛰어난 추론 성능을 실현할 수 있음을 입증했다. 특히, 모형의 핵심 기술을 외부에 공개함으로써, 기존의 폐쇄적이고 고비용 구조를 유지해 온 선도 기업들과는 달리, 신생 모형들과의 경쟁을 활성화시키며 AI 생태계의 개방성과 역동성을 한층 강화하고 있다. 이러한 흐름은 향후 AI 개발 주기를 더욱 단축시키고, 혁신 기술의 상용화 및 관련 서비스 확산을 촉진하는 계기가 될 것으로 전망된다. 더불어, 금융기관이 기존에 축적한 IT 자산을 전략적으로 활용한다면 외부 기술에 대한 의존도를 줄이고, 디지털 전환에 능동적으로 대응할 수 있으며, 장기적으로는 금융업의 경쟁력 강화와 산업 전반의 부가가치 향상으로 이어질 수 있을 것이다.

지난 1월 발표된 DeepSeek(深度求索) 인공지능(Artificial Intelligence: AI) 모형이 기존의 고성능 AI 모형과 비교하여 저렴한 비용으로 비견할만한 성과를 보여주면서 생성형 AI(Generative AI: GenAI)와 대형 언어모형(Large Language Model: LLM) 개발 경쟁이 가속화되고 있다. 중국 헤지펀드 High-Flyer(幻方)에서 설립한 연구소 DeepSeek에서 개발한 AI 모형은 관련 시장의 선두 주자인 OpenAI에서 내세우는 대표적인 모형보다 낮은 비용으로 비슷하거나 더 나은 성능을 얻고 있다고 주장하여¹⁾ 주목을 받았다. 이와 같은 성과는 미국 기업이 주도하는 AI 개발 경쟁 구도를 크게 변화시킬 수 있는 촉매제가 될 것으로 기대되고 있다.

본 고에서는 DeepSeek AI가 기존 언어모형 대비 우수한 성능을 얻게 된 기술적인 배경과 더불어 해당 모형의 성과가 내포하는 한계를 동시에 살펴보았다. 더불어 DeepSeek의 성공이 향후 AI 개발 경쟁 구도에 미치는 영향과 나아가 금융산업의 연구개발(R&D)을 통한 혁신과 가치 제고 가능성에 대하여 시사하는 바를 논의하고자 한다.

* 본고의 견해와 주장은 필자 개인의 것이며, 자본시장연구원의 공식적인 견해가 아님을 밝힙니다.

1) DeepSeek-AI, 2025, DeepSeek-R1: Incentivizing reasoning capability in LLMs via reinforcement learning.

DeepSeek AI 모형의 기술적 혁신

지난 1월 공개된 DeepSeek은 기반이 되는 LLM 모형인 DeepSeek-V3와 이를 바탕으로 논리적 사고 능력을 강화한 추론 모형(reasoning model) DeepSeek-R1을 공개했다.²⁾ 두 모형 중 성과와 활용도 면에서 우수한 결과를 보인 DeepSeek-R1의 기술적 혁신 요소는 크게 Chain of Thought(CoT: 생각의 사슬), Reinforcement Learning(RL: 강화학습), distillation(지식 증류)의 세 가지로 나누어 볼 수 있다(〈표 1〉 참고).

〈표 1〉 DeepSeek AI의 기술적 혁신 요소

문제점	기술 요소	결과
추론 과정 불분명 결과물의 논리적 연관성	Chain of Thought (CoT)	추론 과정을 명시 답변의 질적인 개선
학습 데이터의 양적, 질적인 부족	Reinforcement Learning (Group Relative Policy Optimization)	비교적 적은 데이터로 학습 효율성 제고
대형화로 인한 학습 시간 및 비용 증가	Distillation	성능을 유지하면서 소형화하여 비용 절감

기존의 LLM이 질문에 대한 답변을 작성하는 복잡하고 확률적인 과정은 외부로 드러나지 않아 사용하는 모형이 어떠한 ‘생각’을 거쳐서 결론에 도달했는지 알기 어려운 문제가 있었다. DeepSeek-R1은 이러한 한계를 극복하기 위하여 논리적 사고 과정을 단계별로 서술하도록 유도하는 CoT 방식을 사용하였다. 예를 들어, 수학 문제를 풀 때 단답형의 출력물을 작성하는 대신에 문제를 푸는 과정을 단계별로 서술하도록 유도하여 답변의 정확도를 높이면서 사용자가 모형의 추론 과정을 볼 수 있도록 하는 접근법이다.³⁾ DeepSeek-R1 모형은 CoT 방식을 적용하여 단순하게 정답을 출력하는 것을 넘어서 모형 스스로 사고 능력을 향상시켜 기존 언어모형 대비 수학, 논리 문제 등에서 높은 성과를 얻을 수 있었다.

CoT 방식은 이미 OpenAI의 o1 모형을 비롯한 추론 모형의 선두 주자를 통해서 효과가 입증되었으나 기존 언어모형과 비교해서도 더욱 많은 학습량이 필요하여 계산에 더 많은 시간과 비용이 소모된다는 점이 한계로 지적되었다. 이를 극복하기 위하여 DeepSeek은 RL의 일종인 Group Relative Policy Optimization(GRPO) 기법을 도입하였는데, 이는 모형이 동일한 과제를 반복적으로 수행하고

2) 생성형 AI 모형 개발을 선도하고 있는 OpenAI의 모형과 비교하면 전자는 GPT-4, 후자는 o1 모형과 경쟁 관계로 볼 수 있다.
3) Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E.H., Le, Q.V., Zhou, D., 2022, Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems* 35, 24824-24837.

높은 빈도로 나온 (정답에 가까운) 답변을 스스로 학습하여 추론 능력을 향상시키는 방법이다. 예를 들어, 수학 문제를 채점할 때 미리 작성한 답안지가 없는 경우 여러 학생들의 답안을 비교하고 이를 종합하여 답안지를 스스로 작성하는 과정에 비유할 수 있다. 이와 같은 과정을 통하여 DeepSeek-R1은 인간의 개입을 최소화하여 비용을 절감하면서 모형의 추론 능력을 높여 시장을 선도하는 모형과 비슷하거나 나은 성과를 얻을 수 있었다.

마지막으로, 대형화된 생성형 AI 모형은 배포 및 활용 단계에서도 많은 연산 자원을 소모하기 때문에 비용이 높게 책정된다는 것이 제약 요인이었다. DeepSeek은 이를 초거대 모형의 학습된 지식을 작은 모형으로 전이하는 distillation 방법을 통하여 극복하였다. 해당 방법론은 모형을 경량화하면서도 성능을 유지하여 기존 추론 모형 대비 적은 연산 자원으로도 작동하도록 한다. 이를 통해서 성능이 입증된 Llama-3, Qwen 등의 경량화 모형을 구축해서 원본 대비 적은 비용으로도 우수한 성능을 보이는 모형을 배포하여 생성형 AI 활용에 따른 비용 부담을 크게 낮출 수 있었다.

DeepSeek AI 성공의 양면성

DeepSeek AI의 성과가 더욱 주목받는 이유는 해당 모형이 오픈소스(open source)로 공개되어 외부 연구진이 모형의 주요 요소를 직접 확인하고 개선할 수 있다는 점이다.⁴⁾ 이와 같은 개방성은 지금까지 대부분의 AI 선도 기업들이 자사의 모형에서 핵심적인 가중치와 학습 방법 등을 철저히 보호하던 방식과 극명한 대조를 이룬다. DeepSeek의 등장 이전까지는 OpenAI의 GPT, Anthropic의 Claude와 같이 미국에서 개발된 폐쇄적인 모형이 성능면에서 가장 앞서있으며 오픈소스 기반 모형은 다소 뒤처져 있다는 것이 일반적인 인식이었다. 그러나 DeepSeek의 사례는 AI 개발 경쟁에서 힘의 균형이 얼마나 빨리 바뀔 수 있는지를 보여주고 있다.

이와 같은 긍정적인 요소에도 불구하고 DeepSeek AI가 선도적인 지위를 오랜 기간동안 유지하기는 어려울 것으로 예상된다. DeepSeek은 Llama, OpenAI의 API와 같이 공개된 기술을 활용하면서 인간의 개입을 최소화하여 비용을 절감하였는데, 이는 해당 모형의 성공 요소에 더하여 인간의 개입을 통해 더욱 개선된 모형을 개발할 여지가 있음을 의미한다. 나아가, 외부 연구진이 공개된 모형을 분석하여 DeepSeek AI의 접근법을 응용할 경우, 더 효율적인 모델을 개발하거나 기존 모델의 성능을 크게 개선할 가능성이 있다.⁵⁾ 이는 역설적으로 AI 개발 경쟁을 심화시켜 DeepSeek AI는 선도적인

4) AI 모형을 공개하는 방식으로는 (1) 학습에 사용된 데이터와 학습 방식을 모두 공개하는 경우와 (2) 학습된 모형의 가중치(weights)를 공개하는 방식으로 나뉜다. 엄밀한 의미에서 전자가 후자보다 더 개방적이지만 메타(Meta)의 Llama가 후자의 방식을 채택한 이후로 두 가지 방식 모두 오픈소스로 흔히 불리고 있다. DeepSeek AI는 학습 자료를 공개하지는 않았다는 점에서 후자에 가까운 경우로 볼 수 있다.

5) 실제로 산업계(Criddle, C., 2025. 3. 2, AI companies race to use 'distillation' to produce cheaper models, Financial Times.)와 학계(Muennighoff et. al., 2025, sl: Simple test-time scaling, *arXiv preprint arXiv:2501.19393*.) 양면에서 모형의 비용 효율을 높이는 연구를 진행하고 있다.

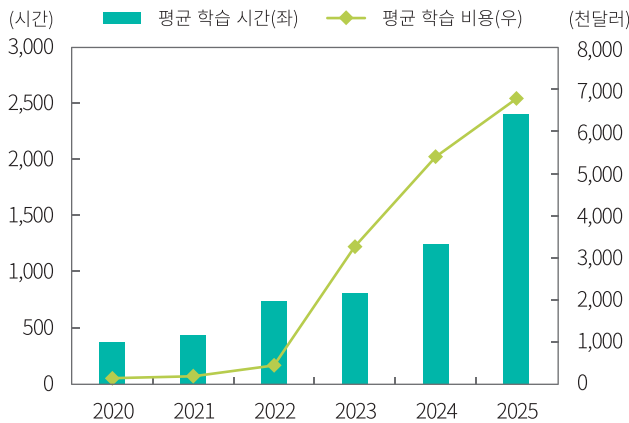
모형으로서 지위를 장기적으로 유지하기 어려울 것으로 예상할 수 있다.⁶⁾

DeepSeek AI가 보여준 성과의 이와 같은 양면성은 향후 생성형 AI 개발 경쟁 구도와 이를 활용하고자 하는 금융업계의 관점에서 중요한 시사점을 전하고 있다.

AI 개발 경쟁 구도의 변화

DeepSeek의 성공은 AI 업계의 경쟁 구도를 변화시킬 것으로 예상된다. 생성형 AI 개발 경쟁은 OpenAI, Google과 같은 선두 기업들이 막대한 투자 자원과 선점 효과로 시장 지배력을 유지할 것이라 예상되었으나,⁷⁾ DeepSeek은 적은 자원으로도 우수한 성능을 구현할 수 있음을 보여주었다. 특히 최신 GPU에 대한 접근이 제한적인 상황에서도 DeepSeek AI 연구진은 기존보다 훨씬 낮은 비용으로 대형 언어모형을 학습시킬 수 있는 방법을 제안하였는데, 이는 지난 수년간 기하급수적으로 증가하고 있는 개발 비용의 추세를(〈그림 1〉 참고) 반전시키는 계기가 될 수 있다. 실제로 DeepSeek-R1의 높은 가격 대비 성능은 이후 공개된 모형들이 경쟁적인 가격을 책정하도록 유도한 것으로 보인다(〈그림 2〉 참고).

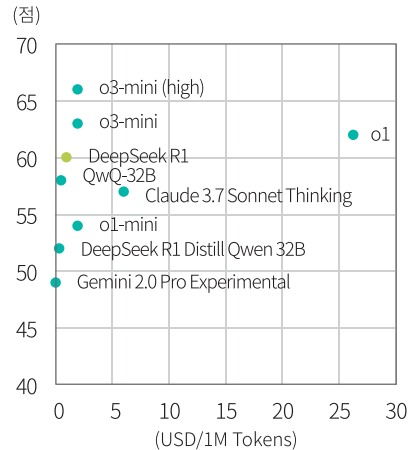
〈그림 1〉 생성형 AI 모형 학습 비용



주 : 해당 연도에 공개된 모든 생성형 AI 모형의 평균 학습 시간 및 비용, 학습 비용은 2023년 달러화 가치를 기준으로 환산하였다.

자료: Epoch AI

〈그림 2〉 모형의 가격과 성능 비교



주 : 2025. 3. 18. 기준 평가점수 상위 10개 모형의 토큰당 가격(가로축) 대비 점수(세로축)를 나타내었다.

자료: Artificial Analysis

6) 실제로, 지난 3월 공개된 LG의 생성형 AI 모형 EXAONE은 성능 면에서 DeepSeek-R1을 비롯한 기존의 추론 모형과 유사하거나 더 나은 성과를 보여서 주목을 받고 있다: LG AI Research, 2025, EXAONE Deep: Reasoning Enhanced Language Models, *arXiv Preprint arXiv:2503.12524*.

7) Korinek, A. & Vipra, J., 2024, Concentrating Intelligence: Scaling and Market Structure in Artificial Intelligence, *NBER Working Paper* no. 33139.

더불어, DeepSeek AI가 오픈소스 접근법을 채택하고서도 우수한 성과를 얻은 사례는 기술의 발전 속도를 더욱 빠르게 하는 계기가 될 수 있다. 오픈소스 모형의 성공은 이를 연구 및 개선하여 폐쇄적인 접근법을 선택한 모형을 성능면에서 빠르게 추격할 가능성을 열었다는 점에서 AI 개발 경쟁에서 후발 주자에 가까운 한국에도 시사하는 바가 크다. 후속 연구를 촉진한다는 측면에서 향후 주도적인 AI 기술은 단순히 우수한 성능을 증명하는 것을 넘어서 협업의 용이성 면에서도 평가받을 것으로 기대한다.

금융산업의 R&D와 혁신 전략

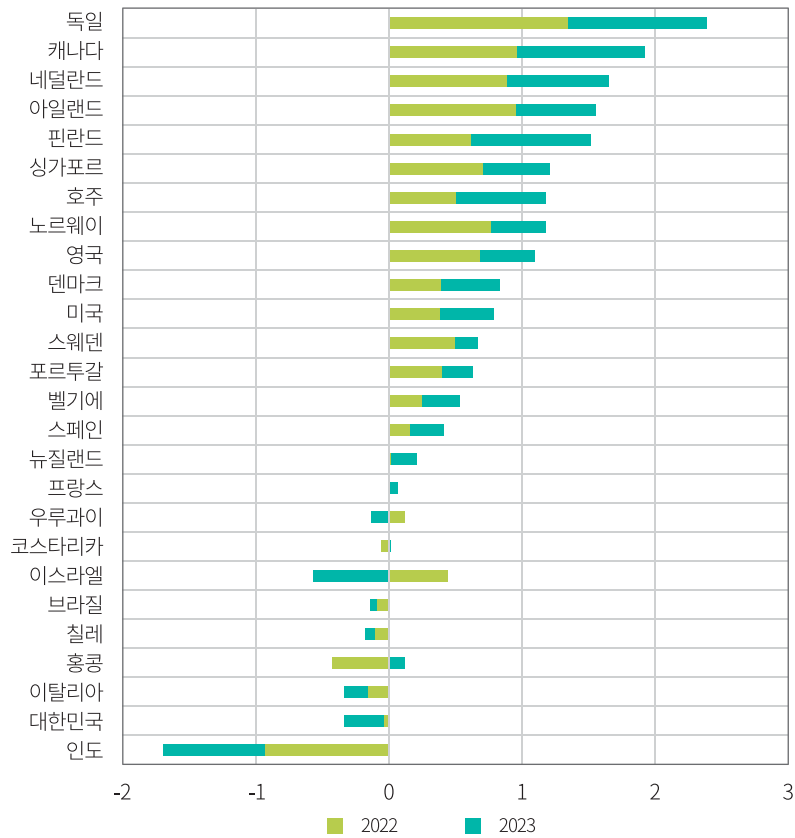
DeepSeek AI의 사례는 금융 기관의 전략적 R&D에 대한 관점의 변화가 필요함을 시사하고 있다. DeepSeek은 모기업인 High-Flyer에서 확보한 GPU 자원을 적극적으로 활용하여⁸⁾ 선진적인 AI 모형을 개발하였다. 특히, 미국의 대중국 수출 제재로 고성능 GPU를 대량으로 확보하기 어려운 상황에서 앞서 설명한 혁신 기술을 활용하여 고성능 모형을 개발할 수 있었던 배경에는 우수한 연구 인력⁹⁾에 더하여 단기적인 수익보다 장기적인 연구 성과를 지향한 점¹⁰⁾을 주목할 수 있다. 이는 금융기관에서 보유하고 있는 IT 인프라와 연구 인력 자원을 효과적으로 활용하여 시장 전반에 영향을 미칠 수 있는 성과를 얻은 사례로 볼 수 있다. 나아가 디지털 혁신의 시대에 금융기관이 외부 IT 솔루션을 활용하는 수동적인 대응을 넘어서 자체적인 기술 역량을 바탕으로 효율성과 안정성이 높은 맞춤형 서비스를 개발하여 시장을 선도하고 기업의 가치를 높일 수 있음을 시사한다.

8) Patel, D., Kourabi, A.J., O’Laughlin, D., Knuhtsen, R., 2025. 1. 31, DeepSeek Debate: Chinese Leadership on Cost, True Training Cost, Closed Model Margin Impacts, SemiAnalysis.

9) Chen, C., How a top Chinese AI model overcame US sanctions, MIT Technology Review.

10) Wu, Z., DeepSeek focuses on research over revenue in contrast to Silicon Valley, Financial Times.

〈그림 3〉 국가별 AI 인재 순유입



주 : LinkedIn 가입자 1만명당 한 해 동안 AI 기술을 보유한 인재의 순유입 비중을 나타내며 양의 값은 인재의 유입, 음의 값은 인재의 유출을 의미한다.
 자료: AI Index(2024), Stanford HAI 연구소

금융기관의 R&D를 통한 장기적인 가치 제고를 위해서는 분명한 목표의 설정과 인프라의 확보가 선행될 필요가 있다. 우선, AI가 가장 큰 영향을 미칠 수 있는 업무 분야를 파악하고, 문제점과 해결 방안을 포함하여 구체적인 목표를 설정해야 한다. 더불어, 전문가를 확보하고 스타트업 및 연구 기관과의 협력을 통해 기술 역량을 강화할 필요가 있다. 국내 금융기관의 IT 인적·물적 자원에 대한 투자는 점진적으로 증가하고 있으나 아직 선진국 대비 부족한 상황이다.¹¹⁾ 특히, 지난 수년간 미국, 캐나다, 유럽에서 AI 전문가 인력의 순유입이 있었던 반면 한국은 순유출 국가로 분류되고 있다(〈그림 3〉 참고).¹²⁾ 이와 같은 추세를 반전시키기 위하여 장기적인 관점에서 연구개발 자원에 대한 전략적인 투자를 지속할 필요가 있다.

11) 노성호, 2023, 생성형 AI에 의한 생산성 혁신과 금융업의 대응 방향, 자본시장연구원 『자본시장포커스』 2023-18.
 12) 해당 자료는 국내에서는 보급률이 상대적으로 낮은 네트워크 플랫폼인 LinkedIn을 통하여 수집되었다는 점에서 인력의 유입 대비 유출이 과도하게 측정되었을 가능성이 있다.