

2025.2 vol.3

ISSUE REPORT



중국의 AI 혁명: 생성형 AI와 DeepSeek가 뒤흔드는 미래

한중과학기술협력센터 정리(鄭麗) 연구원

CONTENTS

1. 중국 생성형 AI의 정책, 투자와 역량
2. 중국의 생성형 AI와 DeepSeek의 혁신
3. DeepSeek가 바꾼 게임의 규칙 : 글로벌 AI경쟁의 新 풍경
4. 요약과 시사



Issue/Report는 한중과학기술협력센터가 중국 과학기술계의 주요 이슈를 요약·정리하여 발행하는 자료입니다.





한중과학기술협력센터

ISSUE REPORT



KOREA-CHINA
SCIENCE & TECHNOLOGY
COOPERATION CENTER

2025.2 vol.3

중국의 AI 혁명: 생성형 AI와 DeepSeek가 뒤흔드는 미래

CONTENTES

1. 중국 생성형 AI의 정책, 투자와 역량 2
2. 중국의 생성형 AI와 DeepSeek의 혁신 4
3. DeepSeek가 바꾼 게임의 규칙 : 글로벌 AI경쟁의 新 풍경 .. 15
4. 요약과 시사 18

1

중국 생성형 AI의 정책, 투자와 역량

▶ DeepSeek R1의 등장은 미국은 물론 글로벌 AI업계에게 일종의 충격을 주고 있는데, 우선 후발국의 '고기능, 저비용' AI 전략의 가능성을 보여주었으며, 오픈소스 전략과 파격적인 채용 조건으로 자국 내 육성된 젊은 인재를 확보하며 중국 AI 굴기를 전세계에 알리는 신호탄으로 평가

- 중국 AI 굴기의 기저에는 '17년 '차세대 인공지능 발전 계획'에서부터 '22년 'AI 응용 혁신 촉진'을 통한 고도 경제 발전, '24년 'AI+' 행동 계획과 같은 일관된 정책 지원이 존재

▶ (정책 측면) 중국은 AI 발전을 국가 전략으로 설정하고, 생성형 AI 산업의 성장과 확산을 위해 대규모 모델 개발을 지원하고 AI 기술의 상용화를 촉진하는 데 재정적·행정적 지원을 적극 마련하고 있음

- 특히 '차세대 인공지능 발전 계획(2017)에서 발표된 비전에는 일찌기 2030년까지 세계적 AI 강국으로 도약하기 위한 '3단계 로드맵(三步走戰略)'을 제시한 바가 있음
 - '20년까지 AI 기술과 산업을 세계 선진 수준으로 육성하고, '25년까지 기초 이론 혁신과 핵심 기술 선도를 추진하며, '30년까지 AI 전 분야에서 세계 최고 수준을 달성해 글로벌 AI 혁신 중심국으로 도약하는 것을 목표로 함
 - 또한 베이징, 상하이, 광둥 등 주요 지역에서도 AI 대규모 모델과 관련 지원 정책을 연이어 발표하며, 지역별 AI 산업 육성을 위한 전략적 접근을 확대

〈중국 정부의 AI '3단계 발전 로드맵(2017)'과 연도별 주요 정책〉

1단계(2020년) : 기반 마련	2단계(25년) : 돌파형 혁신	3단계(30년) : 선도형 발전
- AI 기초 이론 안정화 - 핵심 인력 양성에서 획기적 돌파구 마련. - 일부 응용 분야에서 세계 선도,	- AI 기초 이론에서 획기적 돌파구 마련. - 일부 응용 분야에서 세계 선도, - 산업 고도화 및 경제 전환의 동력	- 글로벌 AI혁신 중심지로 도약 - 이론·기술·응용에서 최고 수준 도달, - 지능형 경제 및 사회 구축

구분	일시	발표 기관/지역	AI 지원정책
국가 차원	2017.07	국무원	'차세대 인공지능 발전 계획'
	2021.09	과기부	'차세대 인공지능 윤리 규범'
	2022.07	과기부 외	'AI 응용 혁신 촉진을 통한 고도 경제 발전 가속화 지침'
	2023.07	공업신식화부 외	'생성형 인공지능 서비스 관리 조치'
	2024.09	네트워크보안기술위원회	'인공지능 보안 거버넌스 프레임워크'
지방 차원	2023.05	베이징시 정부	'베이징시 범용 인공지능 혁신 발전 조치'
	2024.07	베이징시 발전개혁위원회	'베이징시 '인공지능+' 촉진 행동계획('24-'25)'
	2023.06	상하이시 경제정보화위원회	'상하이시 인공지능 대규모 모델 혁신 발전 촉진 조치('23-'25)'
	2023.11	광둥성 정부	'광둥성 인공지능 산업 혁신에 대한 실시 의견'
	2023.06	안후이성 정부	'안후이성 범용 인공지능 혁신 발전 3개년 행동 계획('23-'25)'
	2023.07	충칭시 정부	'충칭시 AI 산업 고도화 활용 사례 기반 발전 계획('23-'25)'

* 출처: 24.11, 중국인터넷정보센터, '생성형 AI 응용 발전 보고서'



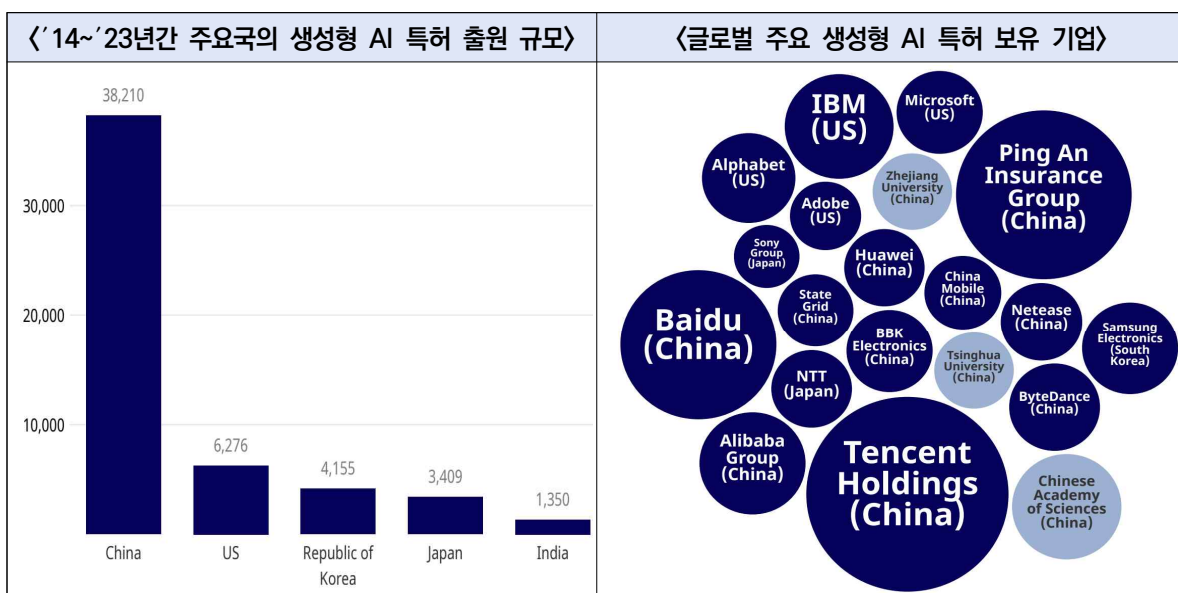
▶ (투자 측면) '24년 1~3분기 동안 중국에서 발생한 AI 관련 투자 및 금융 조달 건수는 총 504건이며, 투자 금액은 약 812억 위안(약 16조 1,165억 원)에 달함

- 분야별로는 AI 응용과 전통 산업 응용이 각각 27.4%(138건)로 가장 큰 비중을 차지했으며, 로봇 관련 분야가 23.6%(119건), AI 하드웨어와 기술은 15.7%(79건)를 기록
- 지역별로는 베이징이 총 436억 위안(53.7%)으로 가장 많은 투자가 이루어졌으며, 상하이(13.8%), 광둥성, 선전, 저장성 등이 50~80억 위안의 투자 규모를 기록

▶ (역량 측면) 세계지적재산권기구(WIPO)에 따르면, 2014~2023년 전 세계 생성형 AI 특허 출원 건수는 5만 4,000건이며, 이 중 3만 8,210건(약 70%)이 중국에서 출원되며, 6,276건을 기록한 미국 대비 약 6배 규모로 세계 1위를 차지

* 한국은 3위(4,155건), 일본(3,409건)과 인도(1,350건)가 각각 4위와 5위를 기록

- (특허) 세계 10대 생성형 AI 특허 출원 기관 중 6곳이 중국 기업으로, 텐센트(2,074건)가 글로벌 1위를 기록했으며, 평안보험(1,564건), 바이두(1,234건), 중국과학원(607건), 알리바바(571건), 바이트댄스(418건)이 포함되었음



▶ (산업 측면) 24년 상반기 기준, AI 기업 수는 전년 대비 35.65% 증가(4,500개 이상)했고, 190개 이상의 생성형 AI 모델이 등록·운영되어 약 6,000억 위안에 이르는 산업 규모로 전망됨

- 국가인터넷정보판공실이 발표한 '생성형 AI 서비스 등록 정보'에 따르면 '24년 11월까지 등록된 생성형 AI 제품은 총 309개로, 베이징(31.1%), 상하이(27.2%), 광둥성(11.7%)이 높은 점유율을 기록

〈중국 생성형 AI 제품 등록 수 상위 10개 지역〉

순위	지역	수량	순위	지역	수량
1	베이징시	96	6	쓰촨성	9
2	상하이시	84	7	귀저우성	5
3	광둥성	36	8	후난성	4
4	저장성	25	9	산둥성	4
5	장쑤성	18	10	천진시	4

* 출처: 중국인터넷정보센터, ‘생성형 AI 응용 발전 보고서’, p4

2

중국의 생성형 AI와 Deepseek의 혁신

2.1 춘추전국 시대를 맞이한 중국의 생성형 AI 모델

▶ 중국의 생성형 AI 산업은 빠르게 성장하고 있으며, AI 모델을 상용화함에 따라 범용 모델을 비롯하여 특정 산업과 디바이스 환경에 특화된 다양한 AI 모델을 개발·확산

- 중국에서 대표적인 범용 AI 대모델로는 커다쉬페이(科大訊飛, iFLYTEK)의 ‘싱훠(訊飛星火)’, 바이두(百度)의 ‘원신이옌(文心一言)’, 알리바바(阿里巴巴)의 ‘통이첸원(通義千問)’ 등이 있음

〈중국 대표적 AI 대규모 모델 개요〉

모델	출시 시간	개발 회사/AI 플랫폼	핵심 기능
 讯飞星火 iFLYTEK SPARK	2024.1	커다쉬페이 ‘星火(싱화)V3.5’	- 자연어 이해 및 지식 추론을 기반으로 텍스트 생성, 코드 작성, 논리 추론 등 7대 핵심 기능을 제공 - AI 기반 문서·PPT 자동 생성, 의료·정책·일반 상식 질의응답, 이미지 분석 등 주요 분야 활용됨
 文心一言	2023.10	바이두 ‘원신이옌4.0’	- 자연어 이해, 텍스트 생성, 논리 추론, 기억 능력 등 4대 핵심 기능 제공 - 문학·광고 카피 작성, 검색 질의응답, 다중 모달 콘텐츠 생성 및 수리 논리 연산 지원
 Qwen	2023.10	알리바바 ‘통이첸원2.0’	중·영문 이해, 수학적 추론, 코드 해석 등 자연어 처리 기반 모델 제공하고 기업용 보고서 생성, 문서 요약 및 교육·학습 보조 활용
 容联云 CLCOPEN	2023	룽롄윈(容联云) ‘츠투(赤兔)’	기업 맞춤형 AI 기반 고객 응대 및 마케팅 자동화 지원하며, 상담 응답 분석, 데이터 기반 비즈니스 인사이트 제공, 스마트 FAQ 및 음성 분석 기능 포함
 midu	2023	미두(蜜度) ‘원슈(文修)’	- 전문 교정 및 문서 감수를 위한 AI 기반 언어 모델 제공 - 뉴스·출판·교육기관의 자동 교정, 문서 품질 개선, AI 기반 감수 서비스 지원
 用友企业服务大模型 YonGPT	2023	용우(用友) ‘YonGPT’	- 기업용 AI 서비스 모델로, 업무 프로세스 자동화 및 데이터 분석 지원 - 스마트 문서 생성, 기업 경영 분석, AI 기반 인재 채용 및 업무 프로세스 최적화



모델	출시 시간	개발 회사/AI 플랫폼	핵심 기능
“写易” 智能创作引擎	2023	인민망(人民网) '셔이(写易)'	- 정부·언론·공공기관을 위한 AI 기반 문서 작성 및 교정 지원 - 정책 문서 작성, 뉴스 기사 보조, 맞춤형 문서 자동 생성 기능 포함
BlueLM 蓝心大模型	2023.11	비보(vivo) '란신(蓝心)'	- 스마트폰 등 단말기에서 실행 가능한 AI 모델 제공 - 개인화된 AI 비서, 자연어 처리 기반 검색·질의응답, 스마트 문서 작성 및 콘텐츠 생성 지원
KIMI	2023.10	Moonshot AI Kimi	- 생성형 AI 기반 스마트 어시스턴트로, 텍스트 생성, 코드 작성, 학술 논문 해석, 법률 분석 및 장문 처리 지원 - 학술 연구 보조, 법률 상담, API 개발 문서 분석, 창작 및 업무 생산성 도구로 활용

* 출처: 24.4, 인민망, '2024 중국 AI 대규모 모델 산업 발전 보고서', p15~27

2.2 Deepseek의 설립과 혁신

▶ DeepSeek는 중국 쿼트 헤지펀드 회사 High-Flyer의 범용인공지능(AGI) 연구 조직으로 출범했으며, '23년 7월 독립 법인으로 분사되었음

- '15년 중국 저장성 항저우에서 설립된 High-Flyer는 중국 4대 쿼트 펀드 중 하나로 AI 기술을 활용하여 투자 알고리즘을 자동화하는 등 기술 기반 투자 모델을 구축해 왔으며, 이러한 연구 개발 역량을 확대하고 기술적 우위를 유지하기 위해 DeepSeek를 설립
- DeepSeek는 High-Flyer의 지원하에 외부 투자 없이 독립적인 운영을 유지. 또한, DeepSeek는 High-Flyer가 보유한 대규모 컴퓨팅 인프라(GPU 클러스터 등)를 활용할 수 있으며, 최대 50,000개의 NVIDIA Hopper GPU를 운영할 수 있는 능력을 갖추고 있음
- DeepSeek는 '23년 설립 이후 단기간 내에 중국 AI 업계에서 가장 주목받는 기업 중 하나로 자리 잡았으며, 혁신기초 연구 및 오픈소스 전략을 중심으로 발전하고 있음

〈참고: DeepSeek의 발전 연혁〉

- '23년 7월: DeepSeek 창립. 초기 연구진은 High-Flyer의 AI 전문가들과 젊은 연구원들로 구성됨
- '23년 11월 2일: 최초의 오픈소스 코드 모델 'DeepSeek Coder' 출시
- '23년 11월 29일: 670억 개의 매개변수를 갖춘 대형 언어 모델(LLM) 'DeepSeek LLM' 공개
- '24년 5월 7일: 두 번째 오픈소스 혼합 전문가(MoE, Mixture of Experts) 모델 'DeepSeek-V2' 발표
- '24년 12월 26일: 6,710억 개 매개변수를 보유한 'DeepSeek-V3' 출시
- '25년 1월 20일: 최신 추론 모델 'DeepSeek-R1' 출시
- '25년 2월 4일: DeepSeek가 140개국 애플 앱스토어 다운로드 순위 1위를 기록

▶ DeepSeek의 R&D 인력은 총 139명으로, 대부분 칭화대, 베이징대 등 중국 최고 대학 출신의 20-30대 젊은 인재들로 구성되어 있으며, 기존 기업들과는 달리 상업적 모델보다 기초 연구 및 기술 혁신에 집중하는 방식을 추구하고 있음

● DeepSeek 창립자인 량원펑(梁文鋒)은 '15년에 설립된 퀀트 헤지펀드 회사인 High-Flyer(幻方量化)를 창립하였으며, AI 기술을 투자 거래에 적용하는 데 주력해 왔음

〈High-Flyer의 기술 혁신과 성장 과정〉

- 2008년, 머신러닝을 활용한 퀀트 헤지펀드 투자 연구 시작, 데이터 축적 본격화
- 2015년, High-Flyer 설립, 퀀트 헤지펀드 사업 공식 출범
- 2016년, 딥러닝 기반 첫 거래 모델 도입, AI 주도 투자 전략 개발
- 2017년, 전면적인 AI 자동화 거래 시스템 구축
- 2019년, 운용 자산 규모 1,000억 위안 달성, 중국 4대 퀀트 헤지펀드 기업으로 자리매김
- 2023년, 범용 인공지능(AGI) 연구를 위한 'DeepSeek' 설립, 기술 확장 추진

〈참고: DeepSeek 창립장 량원펑 개요〉



- ▷ 량원펑(梁文鋒)은 1985년 중국 광둥성에서 출생하였으며, 17세에 저장 대학교에 입학 후 '10년 정보통신공학 석사 학위를 취득
- ▷ '13년, 양자 투자기업 '항저우 야크비(雅克比)투자관리유한회사' 공동 창립
- ▷ '15년 '항저우 High-Flyer 테크놀로지'를 설립하여 AI 기반 양자 투자 사업을 본격적으로 확대
- ▷ '23년 DeepSeek 설립, AGI 분야 확장
- ▷ '25년 1월 DeepSeek R1 공개, 글로벌 AI 시장 진출

* 출처: 25.2.12, 텐센트망, <https://news.qq.com/rain/a/20250211A07MZF00>

〈항저우시 본부 (<https://www.high-flyer.cn/fund/>)〉





〈DeepSeek의 핵심 개발자 (일부)〉

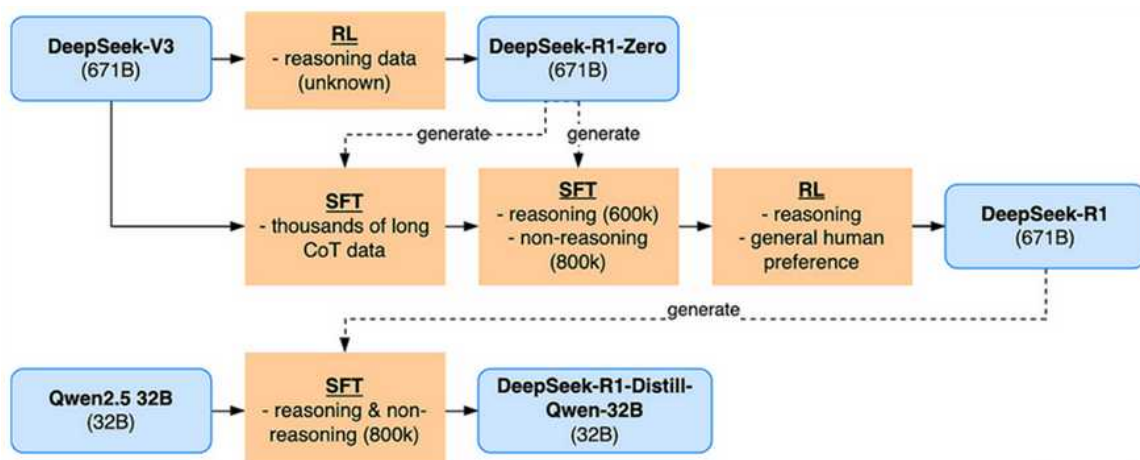
구분	이름	주요 경력
1	가오화조(高華佐)	- 베이징대학교 물리학과 졸업 - MLA 아키텍처 핵심 기여자 - AI 기술 및 물리학적 접근법 융합하여 DeepSeek 기술 발전 주도
2	뤄푸리(羅福莉, 1995년 생)	- 베이징사범대 컴퓨터공학 학사, 베이징대학교 계산언어학 석사 2019년 ACL 국제 학회에서 8편 논문 발표 (2편은 제1저자) - 알리바바 다모원에서 VECO 모델 개발 및 AliceMind 오픈소스 주도 2022년 DeepSeek 합류, DeepSeek-V2 MoE 대형 모델 개발 참여
3	펑왕딩(曾旺丁)	- 베이징우전대 졸업 - DeepSeek에서 GRPO 알고리즘 개발 주도 - 강화학습 및 DeepSeek-Math 프로젝트 핵심 연구 수행
4	샤오즈홍(邵智宏)	- 칭화대학교 CoAI 연구팀 박사 - DeepSeek-Math, DeepSeek-Coder-v2 등 프로젝트 참여 - 마이크로소프트 연구소 근무 경험 보유
5	주치하오(朱琪豪)	- 베이징대학교 박사 - 코드 학습 및 자동 수정 연구, CCF-A급 논문 16편 발표 - DeepSeek-Coder-V1 개발 주도, 오픈소스 코드 모델 연구
6	다이다마이(代達勤)	- 베이징대학교 컴퓨터과학 박사(2024년 졸업) - DeepSeek-V2 MLA 및 DeepSeek-Math GRPO 프로젝트 핵심 참여
7	왕빙쉬안(王炳宣)	- 베이징대학교 졸업 - DeepSeek LLM v1부터 주요 연구 및 개발 참여

* 출처: (25.2.8. 高校大数据与人工智能推进联盟) 起底DeepSeek背后的8位年轻人

2.3 DeepSeek의 주요 모델, 비용 우위 및 핵심 기술

▶ **[주요 모델]** 대규모 언어모델(LLM), 시각-언어 모델(VL), 코드 생성 AI 및 수학 특화 AI 등에 기반하여 DeepSeek-V3에서 시작해 DeepSeek-R1 모델이 완성되는 과정으로 혁신했고, 특히 Qwen2.5 32B 모델부터 Distillation을 거쳐 DeepSeek-R1-distill-Qwen-32B 모델 생성

〈DeepSeek 핵심 모델 출시의 마일스톤〉



- DeepSeek 시리즈는 2024년부터 다양한 특화 모델(자연어, 코드 생성, 수학, 시각-언어 등)을 출시하며 각 분야에 최적화된 성능을 보이는데, GPT-3.5, GPT-4, Llama2 등 최신 AI와 경쟁하며, 대규모 데이터셋과 첨단 기술(RL, MoE)을 활용하여 효율성 극대화
- 최신 모델들은 코드 자동 완성, 수학적 추론, 이미지 생성 및 다중 모달 분석 등 다양한 실무 응용 분야에서 우수한 성능을 발휘

〈DeepSeek의 주요 모델 및 특징〉

모델명	출시 시간	주요 특징 및 활용 분야
DeepSeek LLM	'24.1.5	- DeepSeek의 최초 대규모 언어 모델(LLM), 670억 개의 파라미터 포함 2조 개의 토큰 데이터셋을 활용하여 기초부터 학습 - 중국어 및 영어 이해 능력 우수 - 성능: Llama2 70B Base 및 GPT-3.5 성능 초과
DeepSeek-Coder	'24.1.25	- 코드 생성 및 자동 완성 기능에 특화된 모델 - 다양한 프로그래밍 언어 지원 - 주요 활용 분야: 프로젝트 단위의 코드 자동 완성 및 개발 지원
DeepSeekMath	'24.2.5	- DeepSeek-Coder-v1.5 7B 기반, 수학 문제 해결을 위한 특화 모델 5,000억 개의 토큰으로 학습, GPT-4 수준의 수학적 추론 가능 - 주요 활용 분야: 수학적 추론 및 계산 지원
DeepSeek-VL	'24.3.11	- 오픈소스 기반 시각-언어 모델(Vision-Language, VL) - 고효율 시각 데이터 처리 및 분석 가능 - 주요 활용 분야: 다중 모달 AI 태스크, 이미지 인식 및 설명 생성
DeepSeek-V2	'24.5.7	- 혼합 전문가 모델(Mixture of Experts, MoE) 아키텍처 기반 - 훈련 및 추론 비용 최적화, 생성 효율성 향상 - 성능 개선: 생성 처리량(Throughput) 대폭 향상
DeepSeek-Coder-V2	'24.6.17	- 코드 생성 및 수학적 추론 기능 강화 - 지원하는 프로그래밍 언어의 범위 확장 - 주요 활용 분야: 복잡한 프로그래밍 작업 및 AI 기반 코딩 지원
DeepSeek-VL2	'24.12.13	- 향상된 다중 모달 이해 및 시각-언어 통합 분석 기능 - 주요 활용 분야: 크로스 모달 분석 및 생성 태스크
DeepSeek-V3	'24.12.26	6710억 개의 파라미터를 보유한 MoE 모델 - 훈련 시 37B 파라미터만 활성화하여 실행 효율성 극대화 14.8조 개의 토큰을 활용한 사전 학습 - 성능: 지식 기반 태스크 및 콘텐츠 생성 능력 강화, 생성 속도 및 처리 효율성 대폭 개선
DeepSeek-R1	'25.1.20	- 강화 학습(RL) 기반 최적화 모델 - 최대 128,000 토큰 컨텍스트 길이 지원 - 프로그래밍 및 복잡한 입력 처리에 특화 - 성능: 추론 및 코드 생성 태스크에서 OpenAI GPT-4 및 Claude 3.5 대비 우수한 성능 발휘
Janus-Pro, JanusFlow	'25.1.27	- Janus-Pro-7B: 고성능 이미지 생성 AI 모델 - JanusFlow-1.3B: 다중 모달 AI 분석을 위한 모델 - 주요 활용 분야: 이미지 생성 및 다중 모달 분석, DALL-E 3 및 Stable Diffusion 대비 성능 우위

* 출처: DeepSeek의发展历史, https://blog.csdn.net/Bl_ack/article/details/145461360



〈참고: DeepSeekR1 모델 개요〉

▶ **(개요)** 중국 AI 기업 딥시크(DeepSeek)는 '25년 1월 20일 대규모 추론형 AI 모델 'R1' 및 'R1-Zero'를 공개했으며, 해당 모델을 허깅페이스(Hugging Face)를 통해 오픈소스로 배포했음. R1 시리즈는 오픈AI의 'o1'과 동등한 수준의 성능을 제공하면서도 비용을 대폭 절감한 점이 특징

▶ **(특징)** R1 시리즈는 딥시크 V3 모델을 기반으로 미세조정된 전문가 혼합(MoE) 모델로, 6,170억 개의 매개변수 중 약 340억 개만 활성화되도록 설계되었으며, 강화학습(RL) 기반 추론 모델, 추론 성능 및 경제성 확보, 단계적 최적화를 통한 성능 개선 등 3가지 특징이 갖추고 있음

① 비용 절감형 강화학습(RL) 기반 추론 모델

- ☞ R1-Zero는 지도학습 미세조정(SFT) 없이 시행착오를 통한 학습(RL)으로 사고 연쇄(CoT) 기반의 추론 기능 구현
- ☞ 강화학습에 GRPO(Group Relative Policy Optimization) 프레임워크를 적용하여 추가 평가 모델 없이 자체 최적화 가능

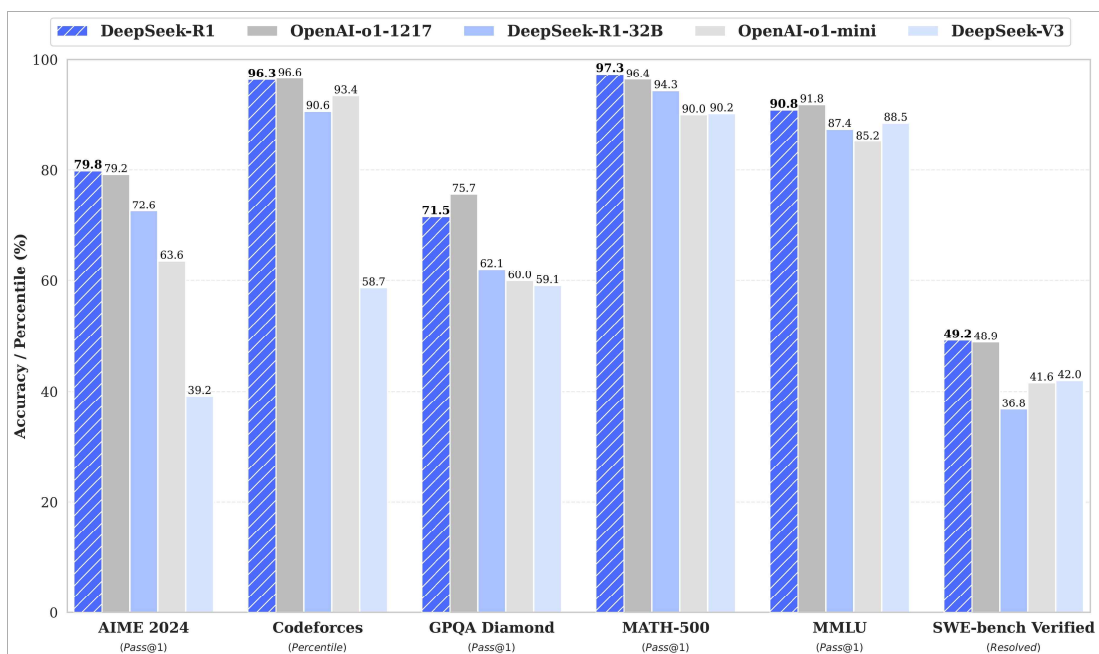
② 추론 성능 및 경제성 확보

- ☞ 벤치마크 평가에서 오픈AI o1과 동등하거나 우수한 성능 기록, 특히 수학 관련 AIME 2024(79.8%) 및 MATH-500(97.3%)에서 o1보다 높은 정확도 확보
- ☞ 비용 절감 효과가 크며, o1 대비 추론 비용 90~95% 절감
 - 오픈AI o1: 입력 토큰 100만 개당 15달러 / 출력 토큰 100만 개당 60달러
 - 딥시크 R1: 입력 토큰 100만 개당 0.55달러 / 출력 토큰 100만 개당 2.19달러

③ 단계적 최적화를 통한 성능 개선

- ☞ R1-Zero는 초기 모델에서 응답 반복, 가독성 저하, 언어 혼합 등의 문제 발생, 이를 개선하기 위해 강화학습 이전 단계에서 기초 지식을 제공하는 사전학습 데이터 추가 및 다단계 학습 적용

〈DeepSeek 모델과 기존 주류 AI의 성능 비교〉



* 출처: DeepSeek홈페이지, <https://www.deepseek.com/>

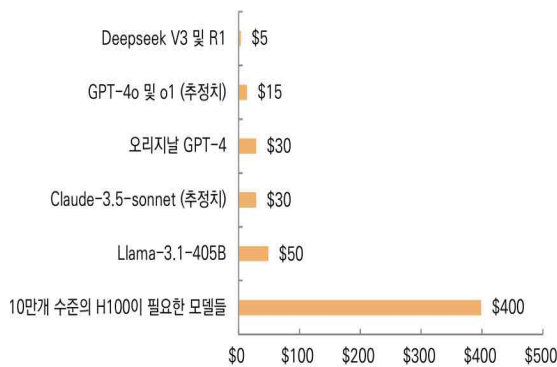
▶ [비용우위] 혁신적 학습 비용절감에도 비판적 논란은 존재

- DeepSeek에 대한 평가 중에 ‘단지 2050개 정도의 GPU만을 사용해 558만 달러만을 들여, 모델 훈련을 끝냈다’는 부분에 대해서는 DeepSeek의 공식 논문에 따르면 ‘마지막’ 훈련 비용만 포함된 것이라는 견해가 주류

* 참고로 OpenAI의 GPT-4의 훈련 비용은 6천만 달러, 앤트로픽의 Claude 3.5 Sonnet는 3천만 달러 수준

- 딥시크의 혁신과 비용절감에도 불구하고, 반도체 관련 저널 SemiAnalysis 등 일각에서 "DeepSeek가 약 10,000개의 H800과 약 10,000개의 H100에 액세스할 수 있다고 분석하며, H100, H800, H20, 그리고 A100까지 총 6만개의 GPU를 보유하고 있으며, 이를 TCO(총소유비용)로 환산하면 4년 동안의 총소유비용(TCO)은 25억 7,300만 달러(약 3.4조원)로 추정

〈H100 훈련시간에 따른 언어모델 훈련비용과 DeepSeek의 컴퓨팅 TCO 추정〉

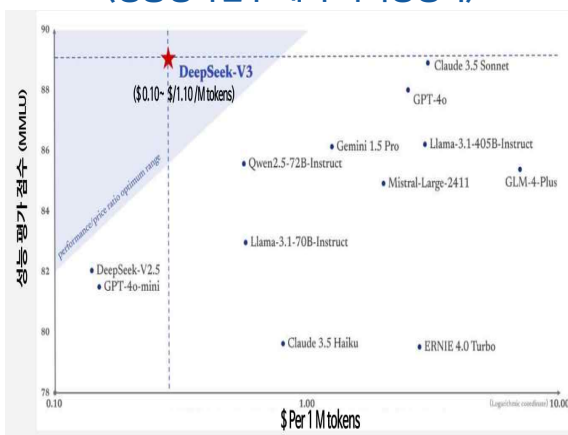


항목(unit)	A100	H20	H800	H100	전체
강제성적연산(년)	4	4	4	4	
GPU 숫자(개)	10,000	30,000	10,000	10,000	60,000
엔비디아 GPU 평균매개(달러)	\$13,500	\$12,500	\$20,000	\$23,000	
GPU당 서버 설비 투자금액 (달러)	\$23,716	\$24,228	\$31,728	\$34,728	
총 서버 설비 투자금액(백만달러)	\$237	\$727	\$317	\$347	\$1,629
운영 비용(백만달러)	\$157	\$387	\$170	\$230	\$944
4년에 걸친 총 소유비용(백만달러)	\$395	\$1,114	\$487	\$577	\$2,573

* 출처 : SemiAnalysis, 미래에셋증권 리서치센터

- 다만, 딥시크는 기계가 데이터를 생성하고, 스스로 학습하는 강화학습(RL) 모델을 활용하면서 학습 (training)과추론(inference)비용을 획기적으로 절감 → 다른 AI 모델 대비 저렴한 이용료 책정

〈성능평가점수 대비 가격경쟁력〉



〈DeepSeek V3 API 가격(Pay-as-you-go)〉

DeepSeek-V3 API Pay-as-you-go pricing

Input :

cache hit
\$0.07 \$0.014 / 1M tokens

cache miss
\$0.27 \$0.14 / 1M tokens

Output :

\$1.10 \$0.28 / 1M tokens



The promotional period will end after February 8, 2025, at 16:00 UTC

* 출처 : <https://api-docs.deepseek.com/news/news1226>



- ▶ **[핵심 기술]** DeepSeek는 미국의 대중국 반도체 및 AI 기술 규제 강화 속에서 지식 증류와 전문가 혼합(MoE) 기반 트랜스포머 모델, 강화 학습(RL) 등 혁신적인 기법을 적용하여 연산 효율성을 극대화하고 논리적 추론 능력을 강화한 AI 모델을 구현

〈참고: 미국의 대중국 반도체·AI 기술 통제 및 DeepSeek의 대응 전략〉

- ▶ 미국 정부는 기술 봉쇄를 통해 중국의 전략적 기술 경쟁력을 약화시키려는 목표로 '22년부터 중국의 반도체 및 AI 기술 발전을 견제하기 위해 강력한 수출 통제 정책을 시행
- 엔비디아(NVIDIA)의 A100, H100 등 고성능 GPU의 대중국 수출을 금지하고, 여러 중국 AI 기업을 미국 상무부의 '실체 리스트(Entity List)'에 추가하는 등 규제를 지속 강화
- '25년 초에는 '임시 최종 규칙(Interim Final Rule)'을 통해 대만 TSMC 등의 파운드리가 중국 기업을 위한 반도체 생산을 제한하도록 조치하고, 고대역폭 메모리(HBM) 등의 첨단 반도체 기술에 대한 통제도 한층 강화
- ▶ 이러한 제한적인 환경 속에서도 DeepSeek는 개방형 기술(Open Source)과 최적화된 모델 설계를 통해 고성능 AI 모델을 개발하는 데 성공했음
- DeepSeek의 접근 방식은 기존의 '고성능 연산 자원(High-Performance Computing, HPC)에 의존하는 AI 학습 방식'에서 벗어나, 상대적으로 낮은 성능의 중급 GPU를 활용하여 고성능 AI 모델을 구현하는 방향으로 발전하고 있음
- 특히 DeepSeek의 개방형 기술 전략(Open Source)으로 기존의 물리적 하드웨어의 이동을 차단하는 방식으로 이루어졌으나, 소프트웨어 기반의 알고리즘 혁신과 개방형 생태계를 통해 지식과 기술을 자유롭게 공유하는 방식을 선택

① **확장형 트랜스포머 (Multi-head Latent Attention)**

- DeepSeek가 MLA(Multi-head Latent Attention)를 활용한 주요 목적은 메모리 최적화, 필요한 순간에만 압축된 데이터를 원래 상태로 복원하여 사용하며 미국의 고성능 반도체 수출 제한에 대응하기 위한 전략적 선택
- DeepSeek는 전체 입력을 다시 참고하는 대신 입력 문장에서 유의미한 내용만 선별적으로 추출하여 계산함으로써 기존 Multi-Head Attention과 유사한 성능을 유지하면서 GPU 메모리 사용량과 연산 비용 등 리소스 사용 최소화와 하드웨어 제약을 극복해서 저성능 반도체로도 높은 성능을 달성할 수 있는 구조 설계

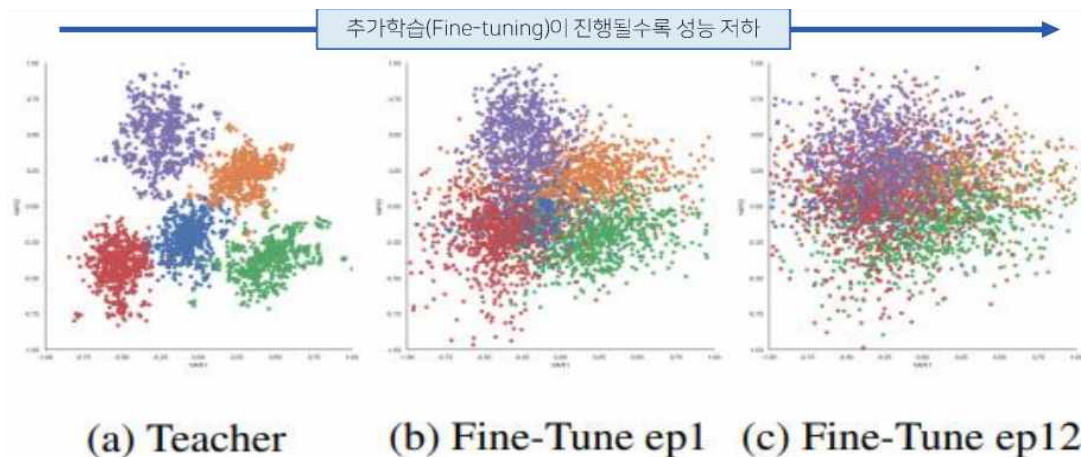
② **지식 증류 및 모델 경량화 (Knowledge Distillation & Model Compression)**

- DeepSeek는 지식 증류와 모델 경량화를 통해 AI 모델의 성능을 유지하면서도 연산 비용 및 하드웨어 자원을 효과적으로 절감하는 데 성공하였으며, 이러한 기술을 바탕으로 고성능·저비용 AI 모델의 대중화를 가속화하고 있음

① 지식 증류 (Knowledge Distillation) : DeepSeek는 대규모 모델의 지식을 소형 모델로 효과적으로 전이하는 지식 증류 기법을 활용

* 기존 AI 모델은 마치 회의 내용을 전부 다 받아적는 비서와 같이 회의 참석자들의 모든 말, 소소한 대화, 심지어 쓸모없는 잡담까지 모두 기록하는데, 이렇게 하면 시간이 오래 걸리고, 리소스가 많이 소요되는 문제가 발생했지만, MLA는 마치 핵심만 요약해서 기록하는 똑똑한 비서와 같이 대화의 중요한 포인트만 정리하고 압축하고 불필요한 세부 정보는 무시하여 빠르고 정확한 판단을 할 수 있게 지원하는 기술

- 파라미터 개수 및 연산량이 큰 대형 모델의 핵심 기능을 경량 모델로 이전하여, 더 적은 연산 자원으로도 복잡한 작업 수행이 가능하도록 설계됨
- 예를 들어, DeepSeek 모델 중 일부는 16억 개의 파라미터만으로도 대규모 모델 수준의 고급 추론 작업을 수행할 수 있음
- 증류(Distillation)을 통해 만들어지는 경량화 모델은 추론 속도 개선에 장점이 있지만, 실시간으로 변하는 물리환경과 AGI 수준의 환경을 추가로 학습할 때 용량이 작아 내용 습득이 어려울 수도 있단 점도 존재



② 모델 경량화 (Model Compression) : DeepSeek는 모델 압축 기법을 적용하여 계산 부담을 줄이고 메모리 사용량을 최소화함으로써, 제한된 하드웨어 환경에서도 원활하게 모델을 실행할 수 있도록 지원. 특히, 모바일 기기 및 엣지 컴퓨팅 환경에서도 우수한 성능을 유지할 수 있도록 최적화됨

③ 전문가 혼합 (Mixture of Experts, MoE)

● MoE 아키텍처를 기반으로 DeepSeek는 제한된 자원 내에서도 대규모 모델을 효율적으로 운영할 수 있도록 설계되었으며, 최고 수준의 성능과 유연성을 동시에 확보

- ① 작업 분할 및 전문가 협업: MoE 아키텍처는 복잡한 문제를 여러 개의 하위 작업으로 분할하여, 각 전문가 네트워크(Expert Network)가 특정 분야의 작업을 처리할 수 있도록 설계되었음
- 예를 들어, 언어 모델에서 문법 이해를 담당하는 전문가 네트워크와 창의적 문장 생성 및 논리 추론을 담당하는 전문가 네트워크가 협업하여 최적의 결과를 도출함



* 기존 모델(일반 트랜스포머): 한 명의 셰프가 모든 요리를 다 만들어야 합니다. 피자, 스테이크, 디저트까지 전부 맡기 때문에 시간이 오래 걸리고 효율이 낮음

* MoE 모델: 피자는 피자 전문가, 스테이크는 스테이크 전문가, 디저트는 디저트 전문가가 각각 맡아서 조리합니다. 이렇게 필요할 때만 해당 전문가를 호출하기 때문에 더 빠르고, 더 나은 품질의 결과를 제공

- ② 파라미터 효율화: 기존 딥러닝 모델이 모든 입력에 대해 전체 파라미터를 활성화하는 것과 달리, MoE 기반 모델은 특정 작업 수행 시 필요한 전문가 네트워크만 활성화하여 연산을 최적화함
- 예를 들어, DeepSeek-V3 모델은 총 6,710억 개의 파라미터를 보유하고 있으나, 특정 입력 처리 시 370억 개의 파라미터만 활성화되며 효율적인 연산 수행
- ③ 동적 라우팅 및 부하 균형: DeepSeek의 MoE 모델은 Expert Choice(전문가 선택) 라우팅 알고리즘을 활용하여, 작업별로 적절한 전문가 네트워크를 동적으로 할당하여 최적의 성능을 유지하고, 특정 전문가 네트워크에 과부하가 발생하는 문제를 방지하여 연산 효율성을 극대화함

④ 강화 학습 및 보상 시스템 (RL: Reinforcement Learning)

- DeepSeek는 기존의 모범 답안을 보고 배우는 지도 학습(SFT: Supervised Fine-Tuning) 대신 스스로 여러 가지 시도를 통해 정답을 찾는 강화 학습(RL: Reinforcement Learning) 중심의 미세 조정을 적용하여 모델의 논리적 사고 능력을 향상
- ① 강화 학습 적용: DeepSeek는 학습된 강화 메커니즘 및 환경 피드백을 활용하여 모델의 의사 결정 능력을 지속적으로 최적화하고 있음. 특히, 추론 및 복잡한 문제 해결에서 강화 학습 기법을 적용하여 모델의 사고 능력을 향상
- ② 규칙 기반 보상 메커니즘: 일반적인 신경망 보상 모델과 차별화된 규칙 기반 보상 시스템을 개발하여, 모델이 명확한 규칙에 따라 학습할 수 있도록 유도하며, 훈련 효율을 높이고 논리적 추론이 필요한 작업에서의 성능을 강화함

〈 DeepSeek V3의 공학적 혁신 주요 포인트 (출시 : 24.12.26) 〉

혁신 영역	핵심 기술	주요 이점
아키텍처 설계	- MLA (압축된 KV 캐시)	- 메모리 사용량 80~90% 감소
	- DeepSeekMoE (동적 편향 조정)	- 추론 속도 향상
분산 훈련 최적화		- MoE 활용 효율성 최적화
	- 동적 워크 스케줄링 커스터마이징	- 파이프라인 병렬화 효율 향상
메모리 최적화		- IB/NVLink 대역폭 최대화
	- RMSNorm 재계산	- 계산-통신 오버헤드 최소화
	- MLA 업-프로젝션	- 활성화 값 저장시 메모리 절약
정밀도 최적화	- MTP 모듈 공유	- 중복 메모리 할당 방지
	- FP8 훈련	- 추측적 디코딩 가능
	- 동적 스케일링	- 메모리 대역폭 요구 감소
지식 전달		- 계산 효율성 향상
	- DeepSeek-R1 기반 지식증류	- 정밀도 손실 최소화
		- R1의 추론 능력을 계승

* 출처 : Deepseek 홈페이지 참조

2.4 글로벌 주요 AI모델과 비교

▶ DeepSeek는 미국 기반 주요 대규모 언어 모델(LLM)인 ChatGPT, Gemini, Claude와 비교하여 효율적인 아키텍처, 비용 효율성 등을 통해 AI 시장에서 강력한 경쟁자로 부상

- 가격 및 접근성 : DeepSeekR1은 MIT 라이선스 기반의 오픈소스로 무료제공되며, OpenAI GPT-4 및 Google Gemini와 같은 폐쇄형 모델 대비 기업 맞춤형 활용이 가능하도록 혁신
- 컨텍스트 길이 : 128K 토큰 지원으로 GPT-4(32K) 대비 4배 긴 문맥을 처리 가능, Google Gemini Flash Thinking(1M)보다 짧지만 실용적인 성능을 보장하며 혁신

〈글로벌 주요 대규모 언어 모델의 비교〉

항목	ChatGPT(미국)	Gemini(미국)	Claude(미국)	DeepSeek(중국)
				
제작사	OpenAI	Google DeepMind	Anthropic	DeepSeek
특장점	- GPTs: 맞춤형 도구 제작, 공유 및 활용 - canvas: 글쓰기, 코딩 특화 모드	- Gmail, Google Drive 등 탐색 답변 - Double-check response: 웹 검색을 통해 답변 재검증	- 문장력 우수 - Artifact: 코드 실행 미리 보기	- 추론 효율화: 671B→37B - 종류 모델 제공: 1.5B~70B - R1 성능 GPT o1 유사
무료 사용 모델	GPT-4o mini, DALL·E3	Gemini 1.5 Flash Gemini 2.0 Flash Experimental Imagen3	Claude 3.5 Sonnet	- DeepSeek LLM - DeepSeek Coder - DeepSeek Coder V2 - DeepSeek Math - DeepSeek VL.V2.V3 - DeepSeek R1 - Janus (multimodal) - Janus (이미지 생성)
유료 사용 모델	Plus: 월 20\$ Pro: 월 200\$	Premium: 월 29,000₩	Pro: 월 20\$	
웹 검색	가능 (Bing)	가능 (Google)	불가능	가능
파일 출력	가능 (내부 파이썬 실행 지원)	가능 (Google Drive 연동)	가능 (csv, md 파일 가능, MS 오피스 파일 불가)	불가능
대화 내용 저장	저장 (옵트아웃 가능)	저장 가능	저장 (Trust & Safety 데이터만)	저장 (모델 개선 가능성 있음)
IP 및 위치 데이터	저장	저장	저장	저장
데이터 저장 위치	미국 서버	미국 및 글로벌 서버	미국 서버	중국 내 서버
법적 관할권	미국 법 적용	미국 법 적용	미국 법 적용	중국 법 적용

* 출처: <https://docsbot.ai/models/compare/deepseek-r1/gpt-4>



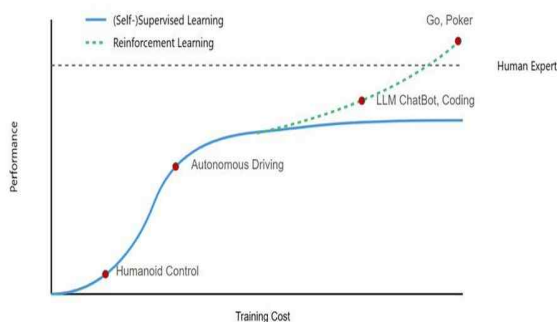
3 DeepSeek가 바꾼 게임의 규칙 : 글로벌 AI경쟁의 新 풍경

▶ DeepSeek의 저비용·고성능 AI 모델 출시는 AI 산업 전반에 큰 파장을 일으키며, 특히 비용 효율성의 향상, 오픈소스 혁신의 중요성 부각, 데이터 보안 등 새로운 도전 이슈가 촉발

▶ **[고효율, 저비용의 트렌드]** 두 가지 학습 방법(지도학습, 강화학습)에 따른 AI 모델의 성능 변화와 저사양, 고효율 AI모델로 글로벌 AI 기술패권의 새로운 패러다임이 형성

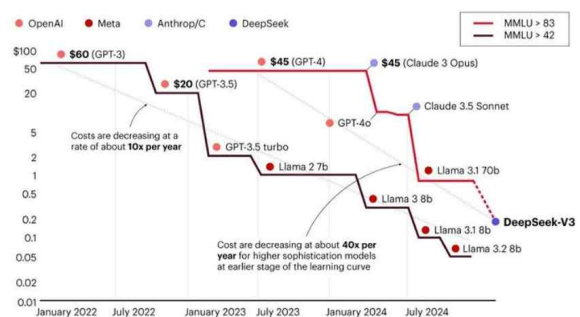
- 미국 내에서는 기존 대규모 자본 투입 없이도 고성능 AI 개발이 가능하다는 점이 확인되면서, 과잉 투자 논란과 함께 프로젝트의 타당성을 재검토해야 한다는 의견이 제기되고 있고, 특히, 고가 GPU 의존도를 줄이고 저전력·고효율 칩에 대한 수요가 증가할 것으로 전망하며, 이에 따라 AI 인프라 투자보다는 알고리즘 최적화 전략이 더 중요한 요소로 부각
- 한편, 딥시크의 부상으로 미국 AI 정책의 변화를 촉진하고 있으며, 오픈AI 등 기존 폐쇄적인 개발 전략에 대해 보다 더 개방형 협력 모델로의 전환이 이루어져야 한다고 강조하고, 미국 정부는 대중 반도체 수출 규제를 강화하면서도 AI 인프라에 대한 투자를 병행하고 있으며, 소프트뱅크를 중심으로 최대 250억 달러 규모의 AI 자금 조달 계획이 추진되는 등 전략적인 대응이 본격화

〈AI 지도학습과 강화학습의 비용 비교〉



* 출처 : Zhengyao Jiang, LLM training cost

〈AI 추론비용 곡선〉



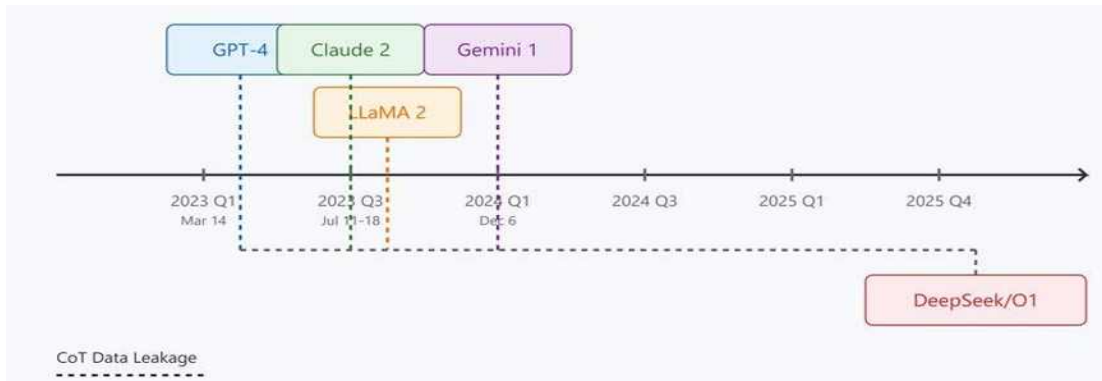
* 출처 : Bain & company

▶ **[미-중간 AI경쟁 가속화]** 딥시크(deepseek)'가 1월 20일 기존의 10분의 1 비용으로 개발한 R1이 오픈AI의 최신 모델(o1)보다 더 뛰어난 성과를 기록하며 중국의 추격과 미국의 견제가 심화

- 트럼프 정부가 1월 22일 스타게이트 프로젝트를 발표한 직후, 딥시크의 R1이 발표(1/26)되었는데, 챗GPT와 MS가 딥시크의 데이터 오용에 대한 조사를 1월 31일 착수했고, 미국, 이탈리아 등이 딥시크 사용금지를 연이어 발표하는 등 딥시크 쇼크에 각국의 대응이 신속하게 진행 중
- 특히, 중국 정부의 적극적인 AI 투자와 관련 기업들의 신기술 발표가 이어지면서, 미국은 반도체 수출 통제 강화, AI 기술 보호, 데이터 활용 규제 등 다각적인 조치를 검토 중이며, 연방의회 경제안보검토위

원회(USCC)에서도 일찍이 ‘맨해튼 프로젝트’와 같은 프로그램을 수립하고 자금을 지원해야 한다”라며, 주요 AI, 클라우드, 데이터센터 기업이 미국 AGI 리더십의 목표에 부합하는 속도와 규모로 정책을 추진할 수 있도록 관련 자금을 지원해야 한다”라고 의회에 제안(24.11)

〈미국 AI 모델의 혁신과 데이터를 기반으로 등장한 DeepSeek〉



출처 : Zhengyao Jiang, LLM training cost, 미래에셋증권 리서치센터

- 기업측면에서는, 오픈AI의 샘 알트만 CEO도 1월 20일 DeepSeek R1가 공개된 직후, o3-mini 모델을 무료로 공개(2월 1일)했는데, 특히 데이터보안을 이유로 미국, 이탈리아, 일본 등 각국 기업과 정부가 딥시크 AI이용을 금지하고 있는 점을 고려하여, 이번에 공개하는 o3-mini는 ‘GPT-4o’를 크게 앞지르는 보안성을 보이며, 중국으로 데이터 전송, 광범위한 개인정보 수집 등 보안에 대한 우려 목소리가 나오는 딥시크와 차별화¹⁾

▶ **[AI 경량화로 산업 융합 가속화]** AI 개발 비용 감소로 신흥 시장 기업들의 AI 도입이 가속화되고, AI와 융합하는 에너지, 자동차, 반도체, 금융 등 신규 수요 시장이 등장하면서 글로벌 AI 및 기술 산업의 판도가 변화될 전망

〈DeepSeek의 등장과 AI+산업 융합의 영향〉

구분	주요 산업	주요 영향
1	기술 및 AI 산업	- Bain & Company의 분석에 따르면, DeepSeek는 혼합 전문가(MoE) 아키텍처, 최적화된 증류 기술, 강화 학습 등을 통해 훈련 및 추론 비용을 크게 절감했으며, 이는 AI 모델 개발의 비용 효율성을 획기적으로 개선한 것으로 평가됨²⁾ - Domino Data Lab의 AI 전략 책임자인 Kjell Carlsson 박사는 ‘DeepSeek의 성공은 AI의 가치가 인프라의 규모나 모델의 독점성이 아닌 효과적인 활용에 있다는 강력한 교훈을 주고 있다’고 지적했으며, 이는 AI 산업에서 오픈소스 혁신의 중요성과 기존 기업들의 취약성을 보여주는 사례임³⁾
2	에너지 산업	DeepSeek의 고효율 AI 모델은 기존 대비 연산 처리에 필요한 전력 소모를 크게 절감할 수 있어, 데이터센터 및 클라우드 서비스의 에너지 소비 최적화에 기여할 것으로 예상됨.

1) ‘中 딥시크’ 보안 듯…오픈AI, 추론 모델 ‘o3 미니’ 무료 배포 [출처:중앙일보] <https://www.joongang.co.kr/article/25311037>



구분	주요 산업	주요 영향
		이에 따라 전력 수요 증가에 대한 기존 전망이 조정될 가능성이 있으며, 주요 에너지 기업들은 AI 기반 전력 수요 예측 및 자원 최적화 방안을 재검토하고 있음 ⁴⁾
3	반도체·하드웨어	Erste Asset Management의 분석에 따르면, 반도체 산업의 추가 발전 가능성을 면밀히 모니터링해야 한다고 지적했으며, 특히 AI 애플리케이션에 필요한 컴퓨팅 파워가 줄어들 것으로 예상되면서, 이에 따른 영향을 주시해야 한다고 강조⁵⁾
4	소프트웨어·클라우드	Erste Asset Management의 기술 주식 전문가 Ruttenstorfer는 'AI 애플리케이션을 서비스에 성공적으로 통합하고 있는 소프트웨어 기업들이 수혜자가 될 것'이라고 전망했으며, Salesforce, SAP, ServiceNow 등의 기업들이 AI 기반 제품을 더 낮은 가격에 제공할 수 있게 되어 고객 수용도를 높이고 마진을 개선할 수 있을 것으로 예상됨
5	자동차 및 통신 산업 ⁶⁾	- 중국의 자동차 제조사인 Great Wall Motor는 DeepSeek AI를 차량 내 인공지능 시스템에 도입하여 지능형 운행 보조 시스템 및 자율주행 기술을 고도화하는 방안을 추진 중. 특히, 차량 내 음성 인식, 내비게이션, 실시간 데이터 분석 등의 분야에서 AI 모델이 적용될 전망 - 중국 내 주요 이동통신사(China Mobile, China Unicom, China Telecom)는 DeepSeek AI를 활용한 네트워크 최적화 및 고객 서비스 개선 방안을 검토 중이며, AI 기반 데이터 분석을 통해 통신 품질을 개선하고, 지능형 고객 응대 서비스(Chatbot) 및 맞춤형 콘텐츠 추천 기능을 강화할 계획
6	사이버보안 산업	The National Cybersecurity Alliance의 Cliff Steinhauer는 DeepSeek의 기술적 진보가 데이터 거버넌스와 프라이버시 프레임워크에 중요한 고려사항을 제기한다고 지적했으며, 특히 중국 AI 기업들이 운영되는 규제 환경의 특수성을 고려할 때, 국제 사용자들의 데이터 보호에 대한 우려가 증가할 것으로 예상됨⁷⁾
7	금융 산업	Deutsche Wealth의 분석에 따르면, DeepSeek의 혁신으로 인해 대규모 AI 투자의 필요성에 대한 의문이 제기되면서 기술 및 관련 섹터에 급격한 시장 반응이 있었음. 이는 AI 관련 주식의 밸류에이션에 대한 재평가로 이어질 수 있음⁸⁾
8	신흥 시장	East Capital의 분석에 따르면, DeepSeek의 등장은 신흥 시장에도 영향을 미칠 것으로 예상되며, 특히 AI 개발 비용의 감소로 인해 신흥 시장 기업들의 AI 도입이 가속화될 수 있음. 이는 글로벌 AI 경쟁 구도에 변화를 가져올 수 있는 요인임⁹⁾

2) <https://www.bain.com/insights/deepseek-a-game-changer-in-ai-efficiency/>

3) <https://www.techradar.com/pro/security/this-is-a-wake-up-call-the-deepseek-disruption-10-experts-weigh-in>

4) https://www.wsj.com/tech/ai/power-energy-stocks-ai-deepseek-39a1e3be?utm_source=chatgpt.com

5) <https://blog.en.erste-am.com/what-effects-could-deepseek-have-on-the-technology-sector/>

6) https://www.reuters.com/technology/amid-deepseek-frenzy-chinese-companies-detail-use-ai-2025-02-09/?utm_source=chatgpt.com

7) <https://solutionsreview.com/the-emergence-of-deepseek-instant-insights-from-industry-experts/>

8) <https://www.deutschewealth.com/en/insights/investing-insights/investing-themes/perspectives-memo-market-perspectives-on-deepseek.html>

9) <https://www.eastcapital.com/insights/the-impact-of-deepseek-on-emerging-markets>

4

요약과 시사

▶ 미국의 반도체·AI 기술 통제에 대응 하는 중국 AI의 우회추격 (彎道超車)

- 미국은 GPU·AI 반도체 수출 규제를 강화하며 중국의 AI 산업 발전을 견제했으며, 이에 따라 중국은 하드웨어 의존도를 줄이고 혼합 전문가(MoE) 아키텍처 및 지식 증류 기술을 활용해 상대적으로 낮은 성능의 GPU에서도 고성능 모델을 구현하며, 미국의 기술 봉쇄를 우회하는 AI 개발 방식을 확립함

▶ DeepSeek는 중국 정부의 일관된 정책과 High-Flyer의 과감한 투자가 결합된 결과

- 중국 정부는 AI를 국가 전략 기술로 지정하고, 5개년 계획 및 연구개발(R&D) 투자 확대, 자연과학기금(NSFC) 지원 등을 통해 AI 산업을 적극 육성해왔으며, DeepSeek의 모회사인 High-Flyer는 AI 기반 퀀트 트레이딩 기업으로, 자체 GPU 클러스터 및 대규모 데이터셋 보유 등 DeepSeek의 기술적 성장을 위한 강력한 인프라를 제공하며, R1의 혁신을 창출

▶ 저비용, 고사양 AI의 가능성을 시사하며, 글로벌 AI경쟁의 신평경 촉발

- 또한, DeepSeek에 대한 글로벌 평가에서는 기술적 성과를 인정하면서도 데이터 보안, 검열, 지정학적 경쟁 등의 우려가 지속적으로 제기되는 상황으로 실제로 MS와 OpenAI는 딥시크가 OpenAI의 API를 통해 대규모 데이터를 무단 추출한 정황을 조사 중이며, 미국, 호주, 이탈리아, 일본, 한국 등은 정부 시스템에서 딥시크 사용 금지와 앱스토어 접근 차단을 발표

▶ DeepSeek가 AI 추격국에 희망을 주기도 하지만, 제본스 역설¹⁰⁾로 연결될 가능성도 존재

- DeepSeek는 후발국에 혁신의 유인을 제공하지만, 챗봇, 자율주행, 로보틱스와 같이 더 거대한 영역에서 융합할 경우, 이에 필요한 컴퓨팅량은 무한대로 증가하며, 모델의 훈련을 넘어 추론(reference)하는 데는 훨씬 더 많은 GPU가 소요되어 인프라 의존도를 높이기도 할 것으로 전망

* 현재 화웨이의 Ascend 시리즈로는 중국 국내 수요 충족도 어려우며 향후 엔비디아, 반도체 등 중국의 외부 AI인프라 의존도가 더욱 높아져서, 역으로 미국의 대중국 견제에 새로운 초크 포인트가 될 가능성도 존재

요소	미국	기타 국가(EU, 한국, 일본 등)	중국
AI 연구	Market Maker (OpenAI, 구글, Anthropic 등)	각국 내 적은 규모의 Share Taker (소버린 AI 관점)	Fast Follower (DeepSeek, Alibaba, Moonshot 등)
반도체	엔비디아 AMD 및 인텔 반도체 장비업체	EU (IMEC 등 연구, 노광장비 ASML) 한국 (메모리, 파운드리) 대만 (TSMC R&D) 및 일본 (TSMC 공장)	자립화 몰두 (화웨이, Cambricon, SMIC)
데이터센터 인프라	세계 최대로 확장 중 송전/배전 그리드가 문제	각국별로 확장 중이나 규모는 제한적	세계 최대 규모로 구축 가능한 잠재력
정부 지원	규제 혁파, 리쇼어링 (Stargate 등) * 트럼프는 몇 가지 행정명령에 벌써 서명	상대적으로 소극적 일본이 그 중 가장 적극적	세계에서 가장 적극적이고 규제도 유연 * 2024년부터 반도체 산업에 약 1,600억 달러 투자 발표

10) Jevons paradox, 기술진보로 효율이 높아지면 자원의 사용을 줄이는 것이 아니고 오히려 늘어난다는 역설



참고 문헌

1. 중국인터넷정보센터, 「생성형 AI 응용 발전 보고」, 2024.11.
2. 세계지식재산권기구(WIPO), 「중국, 생성형 AI 특허 출원 세계 1위」
<https://baijiahao.baidu.com/s?id=1803725572532698372&wfr=spider&for=pc>
3. 인민망, 「2024 중국 AI 대규모 모델 산업 발전 보고」, 2024.4
<http://lib.ia.ac.cn/news/newsdetail/68799>
4. The New York Times, "Do DeepSeek's A.I. Advances Mean US Tech Controls Have Failed?", 2025.01.28.
<https://www.nytimes.com/2025/01/28/business/economy/deepseek-china-us-chip-controls.html>
5. 텐센트망, 2025.02.12., DeepSeek创始人梁文锋的量化前史
<https://news.qq.com/rain/a/20250211A07MZF00>
6. 中国AI新贵DeepSeek, 「DeepSeek 뒤에 있는 8명의 젊은 인물」, 2025.02.08.
<https://zhuanlan.zhihu.com/p/21610485987>
7. CNN, "DeepSeek is giving the world a window into Chinese censorship and information control", 2025.01.29.
<https://edition.cnn.com/2025/01/29/china/deepseek-ai-china-censorship-moderation-intl-hnk/index.html>
8. Center for Strategic and International Studies (CSIS), "Choking off China's Access to the Future of AI"
<https://www.csis.org/analysis/choking-chinas-access-future-ai>
9. 중앙일보, 「中 딥시크 보란 듯…오픈AI, 추론 모델 'O3 미니' 무료 배포」
<https://www.joongang.co.kr/article/25311037>
10. 자유아시아방송(RFA), 「DeepSeek 부상, 더 강력한 규제 초래할까?」, 2025.01.30.
<http://www.rfa.org/cantonese/news/us-sanction-china-deepseek-ai-nvidia-chip-security-01302025154819.html?encoding=simplified>
11. CNN, "China hits out at latest US effort to block Beijing's access to chips", 2024.12.02.
<https://edition.cnn.com/2024/12/02/tech/china-us-chips-new-restrictions-intl-hnk/index.html>

12. CSDN 블로그, 「DeepSeek的发展历史」

https://blog.csdn.net/Bl_ack/article/details/145461360

13. DeepSeek 공식 홈페이지: <https://www.deepseek.com/>

14. SemiAnalysis & 미래에셋증권 리서치센터, 「AI 반도체 시장 전망」

15. Zhengyao Jiang & 미래에셋증권 리서치센터, 「LLM 훈련 비용 분석」

16. Jevons paradox, 「기술진보로 효율이 높아지면 자원의 사용을 줄이는 것이 아니고 오히려 늘어난다는 역설」.

17. CNBC, "In AI chip trade war with China, there's one big mistake US can't make", 2025.02.11.

<https://www.cnbc.com/2025/02/11/deepseek-ai-chip-export-ban-trade-war-us-wont-win.html>

18. 法制日报, 「DeepSeek 확산, AI 산업 혁신 발전에 필요한 규제 균형」

19. Telecom Review Asia, DeepSeek-R1 Shakes Up the AI Industry

<https://telecomreviewasia.com/news/featured-articles/4835-deepseek-r1-shakes-up-the-ai-industry>

20. Tecex, "The Impact Of US Export Regulations On Nvidia AI Chips To China"

<https://tecex.com/ai-chips-trade-wars-sanctions/>

21. CSET, 「중국 AI 연구 개발 재정 지출: 예비 연구 결과」.

22. Cimphony, "U.S. AI Chip Export Restrictions: Impact on Nvidia, AMD"

<https://www.cimphony.ai/insights/us-ai-chip-export-restrictions-impact-on-nvidia-amd>

23. DeepSeek: A Game Changer in AI Efficiency?

<https://www.bain.com/insights/deepseek-a-game-changer-in-ai-efficiency/>



필자 소개

▶ 정 리 (鄭麗)

현) 한중과학기술협력센터 연구원

miouly@naver.com

2025.2 vol.3

한중과학기술협력센터

ISSUE REPORT

중국의 AI 혁명: 생성형 AI와 DeepSeek가 뒤흔드는 미래

| 발행일 | 2025. 2.

| 발행인 | 한중과학기술협력센터

주소 : 北京市朝阳区酒仙桥路甲12号电子城科技大厦1308

TEL : 86)10-6410-7876 / 6410-7886

WEB : <http://www.kostec.re.kr>