

Strategy Idea



글로벌 투자전략 - 신흥국

Analyst 최설화

02. 6454-4883

xuehua.cui@meritz.co.kr

DeepSeek를 통해 본 중국의 AI 산업

- ✓ 중국의 AI 스타트업 DeepSeek가 아키텍처 개선을 통해 오픈소스 AI 고성능 모델 출시
- ✓ 미국의 기술 견제를 중국기업이 기술 혁신으로 대응한 대표 사례
- ✓ DeepSeek 등장으로 중국 빅테크 AI 투자 확대 → 로컬 컴퓨팅 인프라 관련 기업 수혜

중국 DeepSeek의 급부상

중국 스타트업 DeepSeek가
GPU 효율성으로 시장에서 주목

최근 중국의 AI 스타트업 기업 딥시크(DeepSeek)가 전세계적으로 주목 받고 있다. 2023년에 설립된 동사는 항저우에 본사를 두고 인공지능 일반화(AGI)를 목표로, 상업적 응용보다는 기초 기술 개발에 집중하는 AI Lab이다.

시장에서 주목 받는 이유는 12월 25일에 공개한 오픈소스 DeepSeek V3의 높은 GPU 효율성 때문이다. V3 모델은 14조8,000억개의 토큰으로 사전 훈련되었고, 6,710억 개의 매개변수로 이루어졌다. 이는 4,050억개의 매개변수를 가진 라마 3.1의 약 1.6배에 달하는 크기로 오픈 소스 기준 사상 최대 규모의 LLM이다.

훈련 비용이 미국 기업의 1/10

관건은 훈련 비용이 557만 달러(2,048개의 H800 GPU)로 GPT-4 훈련 비용의 10%에 불과하다는 점이다(표 1). 통상 Open AI의 GPT나 Llama의 최첨단 LLM 모델을 훈련하는데 필요한 수억 달러에 비해 매우 경제적인 것이다.

벤치마크 성능 또한 Meta의 Llama-3.1-405B와 알리바바의 Qwen2.5-72B 등 오픈소스 AI 모델을 능가했을 뿐만 아니라 폐쇄형 선두모델인 GPT-4o 및 Claude 3.5 Sonnet과 비슷했다는 평가이다(그림 2).

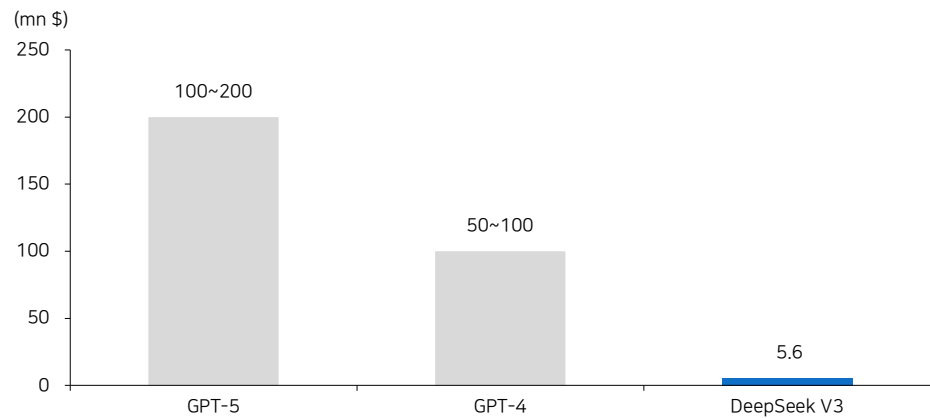
표1 DeepSeek-V3 훈련 비용은 약 557만달러로 추산

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

주: H800 GPU 임대 비용을 시간당 2\$로 가정

자료: DeepSeek

그림1 DeepSeek V3 모델의 훈련비용은 GPT-4의 10% 미만



자료: Chat GPT, 메리츠증권 리서치센터

그림2 DeepSeek V3 모델의 벤치마크 성능, GPT-4o 및 Claude 3.5 Sonnet과 비슷

Benchmark (Metric)	DeepSeek-V3	Qwen2.5 72B-Inst.	Llama3.1 405B-Inst.	Claude-3.5- Sonnet-1022	GPT-4o 0513	
Architecture	MoE	Dense	Dense	-	-	
# Activated Params	37B	72B	405B	-	-	
# Total Params	671B	72B	405B	-	-	
English	MMLU (EM)	88.5	85.3	88.6	88.3	87.2
	MMLU-Redux (EM)	89.1	85.6	86.2	88.9	88
	MMLU-Pro (EM)	75.9	71.6	73.3	78	72.6
	DROP (3-shot F1)	91.6	76.7	88.7	88.3	83.7
	IF-Eval (Prompt Strict)	86.1	84.1	86	86.5	84.3
	GPQA-Diamond (Pass@1)	59.1	49	51.1	65	49.9
	SimpleQA (Correct)	24.9	9.1	17.1	28.4	38.2
	FRAMES (Acc.)	73.3	69.8	70	72.5	80.5
	LongBench v2 (Acc.)	48.7	39.4	36.1	41	48.1
Code	HumanEval-Mul (Pass@1)	82.6	77.3	77.2	81.7	80.5
	LiveCodeBench(Pass@1-COT)	40.5	31.1	28.4	36.3	33.4
	LiveCodeBench (Pass@1)	37.6	28.7	30.1	32.8	34.2
	Codeforces (Percentile)	51.6	24.8	25.3	20.3	23.6
	SWE Verified (Resolved)	42	23.8	24.5	50.8	38.8
	Aider-Edit (Acc.)	79.7	65.4	63.9	84.2	72.9
	Aider-Polyglot (Acc.)	49.6	7.6	5.8	45.3	16
Math	AIME 2024 (Pass@1)	39.2	23.3	23.3	16	9.3
	MATH-500 (EM)	90.2	80	73.8	78.3	74.6
	CNMO 2024 (Pass@1)	43.2	15.9	6.8	13.1	10.8
Chinese	CLUEWSC (EM)	90.9	91.4	84.7	85.4	87.9
	C-Eval (EM)	86.5	86.1	61.5	76.7	76
	C-SimpleQA (Correct)	64.1	48.4	50.4	51.3	59.3

자료: DeepSeek, 메리츠증권 리서치센터

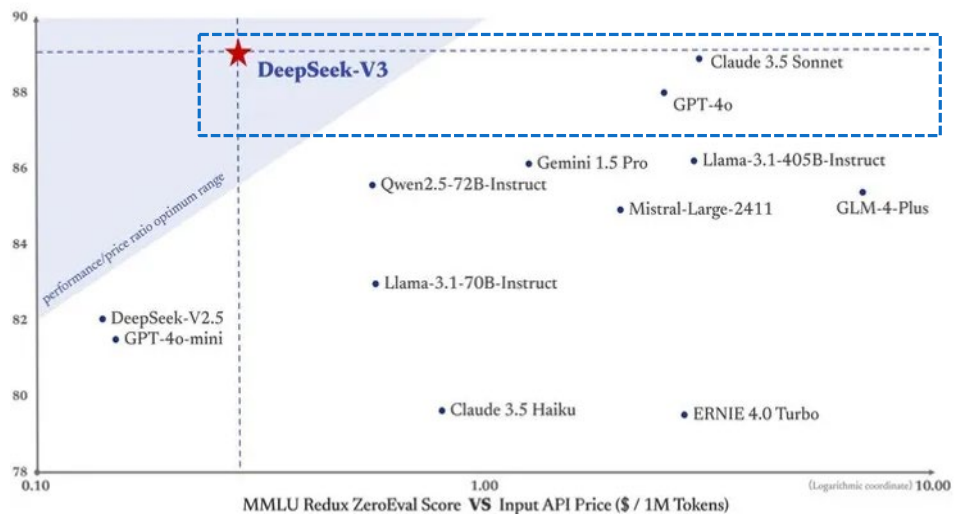
고효율 AI 모델은 아키텍처 개선 덕분

이렇게 고효율의 LLM 모델을 훈련할 수 있었던 원인은 DeepSeek가 독자적인 MLA(Multi-head Latent Attention, 멀티헤드 잠재 어텐션)와 MoE(Mixture of Experts, 전문가 혼합) 아키텍처를 결합 채택하는 기술 혁신을 했기 때문이다.

MLA는 텍스트에서 중요한 세부 사항을 반복적으로 추출함으로써 중요한 정보를 놓칠 가능성을 줄이고 잠재 벡터로 상당히 압축하는 작업이다. 이는 모델에서 결과값의 정확성을 유지하면서 정보를 훨씬 빠르게, 더 적은 메모리로 처리할 수 있는 효율성을 보장해 주었다. 한편 MOE는 작업을 분해해, 전체 모델을 활성화하는 대신 가장 적합한 신경망으로 요청을 보낸다. 즉 6,710억개 매개변수 중 약 340억 개만 활성화 함으로써 성능은 유지하면서도 추론 비용과 메모리 사용량을 크게 절감할 수 있도록 했다.

이러한 아키텍처에 대한 변화를 통해 DeepSeek V3는 결과적으로 <그림 3>에서 보여주는 것처럼 성능 대 가격 비율과 정확도(MMLU Redux Zero-Eval Score) 측면에서 가장 가성비가 우수한 모델임을 입증했다.

그림3 성능 대비 가격 및 정확도에서 DeepSeek V3는 가장 가성비가 우수한 것으로 입증



자료: DeepSeek, 메리츠증권 리서치센터

DeepSeek 등장이 가져온 2가지 변화

최근 중국 AI 스타트업 DeepSeek의 성과는 두 가지 중요한 시사점을 제공한다.

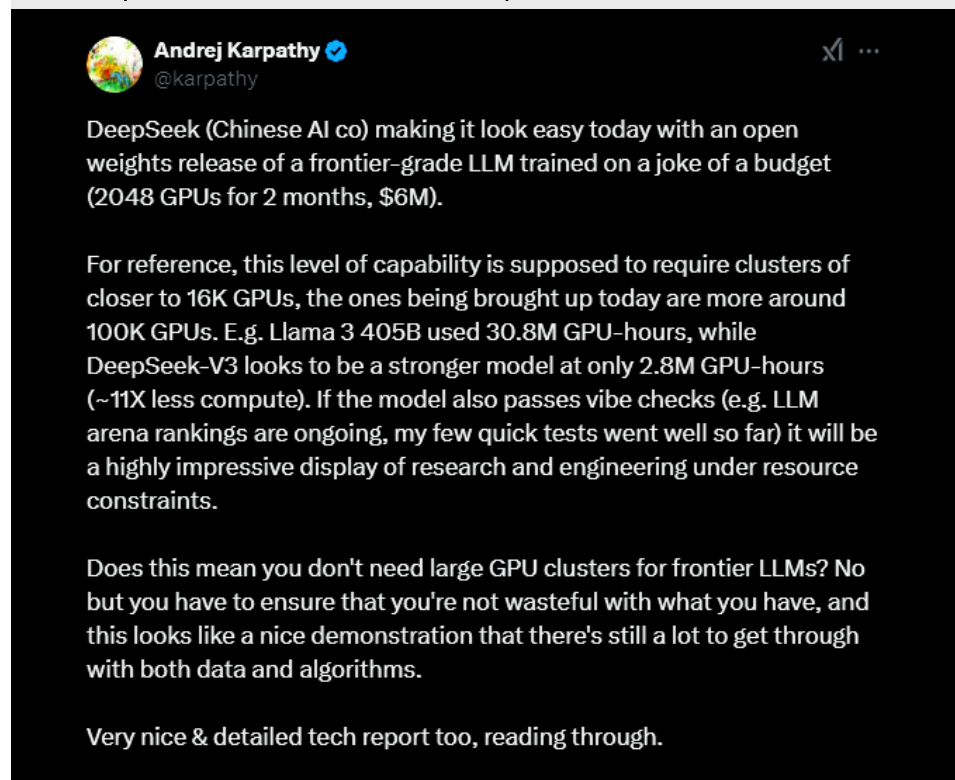
1) 중국 기업은 소프트웨어 개선으로 하드웨어 한계 돌파

첫째, 미국의 기술 제재에도 불구하고 중국의 AI 기업들은 아키텍처 개선을 통해 하드웨어 한계를 극복하며 미국과의 기술 격차를 줄이고 있다.

미국 정부는 중국의 AI 기술 개발을 견제하기 위해 '22년 10월 고성능 AI칩인 H100과 A100 GPU의 수출을 제한했다. 이어 '23년 10월에는 엔비디아가 미국 수출 통제를 우회하기 위해 만든 저사양의 H800과 A800 칩의 수출도 중단했다. 이러한 환경 속에서 DeepSeek는 기술 개선을 통해 고성능을 달성할 수 있는 솔루션을 찾기 시작했다. 그 결과 1년 만에 독자적인 MoE 아키텍처와 어텐션 매커니즘의 변형MLA를 개발하여 높은 성능을 발휘할 수 있는 모델 구현에 성공했다.

Open AI 창립 멤버이자 테슬라의 전 AI 부문 이사인 안드레이 카르파티(Andrej Karpathy)도 X에서 “DeepSeek는 말도 안 되는 비용으로 프론티어급 LLM 출시를 아주 쉬워 보이도록 만들었고, 제약된 자원에서 연구와 엔지니어링을 매우 인상적으로 보여줬다”고 평가했다. 중국 AI 기업들의 소프트웨어 혁신을 통한 격차 축소 노력은 계속될 것이며, 오픈소스 전략은 AI 기술의 접근성과 투명성에 대한 업계 인식을 변화시키며 미국 기업의 글로벌 시장 전략에 영향을 주고 있다.

그림4 DeepSeek V3에 대한 안드레이 카르파티(Open AI 창립 멤버)의 평가



자료: X

2) AI 민주화로 중국 빅테크 기업 투자 확대

둘째, DeepSeek V3 성과는 중국 AI 기업들의 치열한 가격 전쟁 및 잇따른 투자 확대에 연결되면서 로컬 AI 반도체 칩 및 인프라 선두기업에 긍정적으로 작용할 전망이다.

DeepSeek AI 모델의 저렴한 API 가격 책정은 중국 AI 모델 시장에서 가격 경쟁을 촉발했다. 입력 토큰 백만 개당 0.27달러, 출력 토큰 백만 개당 1.10달러라는 파격적인 가격은 경쟁사들의 가격 인하를 유도하고 있다. 이에 알리바바, 바이트댄스, 바이두 등 중국 빅테크도 작년 5월부터 모델 가격을 90% 이상 인하하며 치열한 경쟁에 참여했다(표 2).

표2 2024년 중국 주요 빅테크 기업들의 AI 모델 가격 인하 현황

기업	시점	가격 인하 내용
알리바바	2024년 5월	AI 모델 가격을 최대 97%까지 인하
	2024년 9월	Qwen 2.5 오픈소스 AI 모델 출시
	2024년 12월	Qwen-VK 가격을 최대 85%까지 인하
바이트댄스	2024년 5월	AI 챗봇 모델 '더우바오(Doubao)' 출시. 업계 평균대비 99% 저렴하게 책정
	2024년 12월	더우바오' 영상 이해 모델을 업계 평균대비 85% 낮은 가격에 책정. 1위안(200원)으로 720p 해상도의 이미지 284개 처리 가능
바이두	2024년 5월	어니(Ernie) 기반 스피드와 라이트 모델을 모든 기업 사용자에게 무료로 제공

자료: 현지 언론, 메리츠증권 리서치센터

시장 지배력을 유지하기 위해 중국 빅테크들은 적극적인 기술 개발과 투자 확대에 나섰다. 알리바바는 자체 LLM인 큐원(Qwen) 시리즈를 계속 업그레이드하며 DeepSeek의 혁신에 대응했다. 텐센트는 '23년부터 1/3 투자를 AI 스타트업에 투자하는 방식으로 기업 투자를 강화하고 있다. 이에 '24년 중국 AI 스타트업들이 우후죽순 나타나기 시작했고, 이 중 10억 달러 이상의 기업가치를 평가 받고 있는 기업이 6개에 달하면서 중국의 AI 생태계를 확장해 나아가고 있었다.

표3 2024년 중국 AI 벤처기업 중 10억 달러 이상 평가 받는 6개 비상장기업

기업명	설립일자	편당규모 (억달러)	기업가치 (억달러)	대표 제품	Founder	주요 주주
Zipu	2019.6	8	67	ChatGLM	장펑 (칭화대 컴퓨터공학)	중관춘Tech, 베이징AI산업펀드, Legend Capital 등
Moonshot	2023.4	15	33	Kimi	양즈린 (카네기 멜런대 컴터공학)	Hongshan Capital, Zhen Fund, 알리바바, 텐센트 등
MiniMax	2021.11	-	30	Talkie, Hailuo	양첸제 (SenseTime 前 CEO)	알리바바, 텐센트, IDG Capital 등
Baichuan	2023.4	7	29	Baichuan	왕샤오완 (Sougou 창립자)	베이징AI산업펀드, 상하이AI산업펀드, 텐센트, 알리바바, 샤오미 등
01.AI	2023.5	-	10	Yi-34B	리카이푸 (Google China 전 사장)	알리바바, 동남아재단 등 해외투자자
Stepfun	2023.4	-	-	Step-2	장다신 (MS 전 임원)	상하이국유투자, Qiming Venture, 텐센트 등

자료: Caixin, 메리츠증권 리서치센터

한편 '24년 AI 투자를 공격적으로 늘린 기업은 바이트댄스이다. 한 해에만 AI를 포함한 기술 분야에 총 800억 위안을 투자했다. 이는 바이두, 알리바바, 텐센트의 전체 연간 투자액인 1,000억 위안에 상당한 수준이다. 작년 11월에는 공격적인 마케팅 지출을 통해 AI 챗봇인 더우바오(Doubao)의 중국 내 활성 사용자(MAU) 6천 만명을 확보해 중국에서 가장 인기 있는 AI 챗봇으로 자리매김했다. 이런 선두 지위를 지키기 위해 바이트댄스는 '25년에도 작년의 2배에 달하는 1,600억 위안을 투자할 계획이다(그림 6).

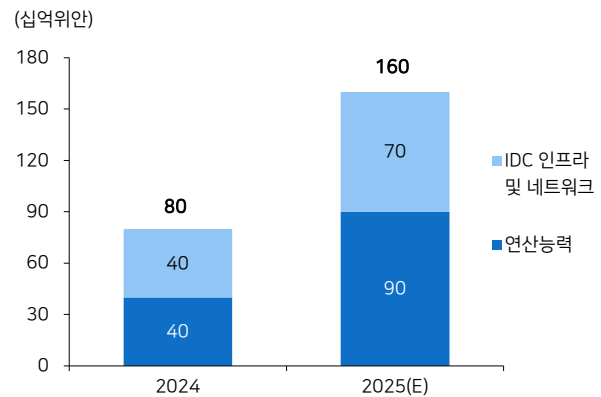
표4 중국 내 주요 AI 모델 월 활성 사용자수 비교

APP	회사	월 활성 사용자수(백만명)
더우바오	바이트댄스	56.5
Kimi	Moonshot	22.8
Wenxiaoyan	바이두	12.0
Xingye	MiniMax	4.5
Maoxiang	바이트댄스	4.4
Zhipu	Zhipu AI	4.4
Tiangong	Kunlun	3.9
iFlytek Spark	iFlytek	2.9
Tongyi	Alibaba	2.8
Midjourney	Dot Matrix Box	1.8

주: 2024년 11월 기준

자료: Caixin, 메리츠증권 리서치센터

그림5 2024~2025년 바이트댄스의 AI 투자액



자료: 현지 언론, 메리츠증권 리서치센터

이는 로컬 컴퓨팅 인프라 관련 기업에 수혜

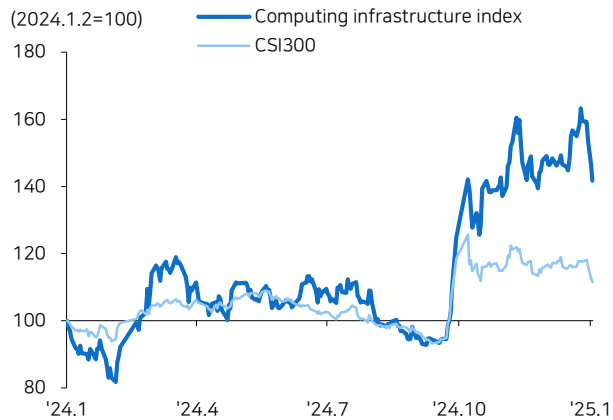
투자의 관점에서 보면 중국의 AI 기업들이 아직 치열한 출혈 경쟁을 하기 때문에 당장 수익화를 하기 어렵다. 또한 빅테크들의 강력한 투자 속에서 어떤 AI 스타트업이 최종 수익화에 성공할 수 있을지도 예단하기 어렵다. 중국 AI 산업에서의 기술 혁신과 발전은 지속될 것이나 아직 특정 기업의 성과에 배팅하기는 어렵다.

다만 중국 AI 산업의 개화와 빅테크 중심의 투자 확대는 로컬 반도체 칩 기업과 데이터센터 인프라 선두기업에는 긍정적인 것이다. 특히 중국의 AI 모델이 기술 혁신을 통해 효율화되면서 저사양의 국산 칩 활용이 더욱 수월해졌다. 트럼프 2기 행정부에서도 미국의 대중국 기술 제재가 더욱 강화될 수 밖에 없으며 국산 대체에 대한 필요성은 더욱 높아지게 된다.

이러한 산업 변화로 작년 10월부터 중국의 컴퓨팅 인프라 지수가 급등하며 시장수익률을 약 30%p 상회했고, 중국의 첫 AI 반도체 상장사인 캄브리콘(cambricon) 주가 또한 3배 이상 급등했다(그림 7).

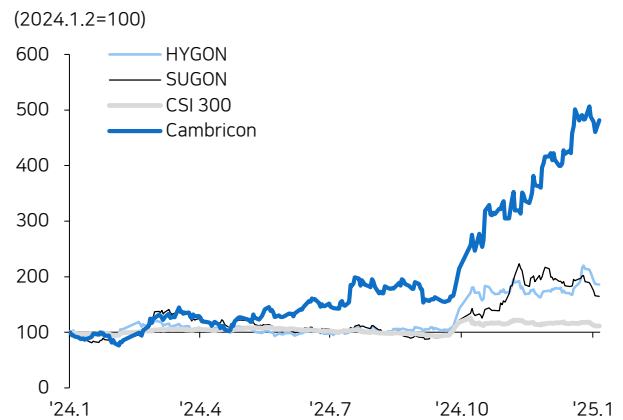
올해에도 바이트댄스 등 빅테크 들의 AI 투자가 빠르게 증가하면서 로컬 인프라 기업들의 수혜가 기대된다. <표 5>에는 중국 컴퓨팅 인프라 관련 대표 기업들이다. 아쉽게도 많은 종목이 아직 외국인 개인투자가 제한된 과창판 또는 차스닥에 상장되어 있어 직접 투자가 어렵다. 대신 이런 종목들이 포함된 과창판50 지수를 대안으로 제시한다.

그림6 '24년 10월부터 컴퓨팅 인프라지수 시장 아웃퍼폼



자료: Wind, 메리츠증권 리서치센터

그림7 중국 컴퓨팅 관련 대표기업 주가 추이



자료: Wind, 메리츠증권 리서치센터

표5 중국 컴퓨팅 인프라 관련 대표 상장기업

구분	기업명	코드	시가총액 (백만원)	2025년 P/E (x)	2025년 EPS 증가율 (% YoY)	기업소개
칩	캠브리콘	688256.SH	259,241	3,035	흑전	중국 최초 AI칩 상장사. 클라우드 서버, 단말기용 AI 핵심칩 개발
	Hygon	688041.SH	315,203	109	43	중국 내 CPU 제조 분야 선두기업, 인텔과 AMD 제품과 경쟁
서버	낭조정보	000977.SZ	67,542	24	25	중국 1위, 글로벌 3위인 서버 제조 기업
	Sugon	603019.SH	94,845	36	25	중국의 슈퍼컴퓨터 전문기업, 슈퍼컴퓨터 시스템 설계 및 제작
응용	iFlytek	002230.SZ	102,803	110	64	중국 최초로 음성 인식 기반 AI 기술 개발 기업
	Cloudwalk	688327.SH	11,614	-28	-	중국 안면 인식 소프트웨어 개발 기업, 독자적 핵심기술 보유
	WPS	688111.SH	121,034	62	28	중국의 대표 오피스 소프트웨어 개발기업, WPS AI로 생산성 향상
	Sensetime	0020.HK	49,220	-17	38	안면인식, 이미지텍스트 식별, 의료 영상 식별 등 AI 기술 제공
IDC	Range Intelligent	300442.SZ	83,079	25	49	대규모 산업단지 데이터센터 운영회사
	AtHub	603881.SH	10,985	49	30	데이터센터 서버 호스팅 서비스 제공 기업

주: 시가총액, P/E는 1월 6일 종가 기준

자료: Wind, Bloomberg, 메리츠증권 리서치센터

Compliance Notice

본 조사분석자료는 제3자에게 사전 제공된 사실이 없습니다. 당사는 자료작성일 현재 본 조사분석자료에 언급된 종목의 지분을 1% 이상 보유하고 있지 않습니다.

본 자료를 작성한 애널리스트는 자료작성일 현재 추천 종목과 재산적 이해관계가 없습니다.

본 자료에 게재된 내용은 본인의 의견을 정확하게 반영하고 있으며, 외부의 부당한 압력이나 간섭 없이 신의 성실하게 작성되었음을 확인합니다.

본 자료는 투자자들의 투자판단에 참고가 되는 정보제공을 목적으로 배포되는 자료입니다. 본 자료에 수록된 내용은 당사 리서치센터의 추정치로서 오차가 발생할 수 있으며 정확성이나 완벽성은 보장하지 않습니다. 본 자료를 이용하시는 분은 본 자료와 관련한 투자의 최종 결정은 자신의 판단으로 하시기 바랍니다. 따라서 어떠한 경우에도 본 자료는 투자 결과와 관련한 법적 책임소재의 증빙자료로 사용될 수 없습니다. 본 조사분석자료는 당사 고객에 한하여 배포되는 자료로 당사의 허락 없이 복사, 대여, 배포 될 수 없습니다.