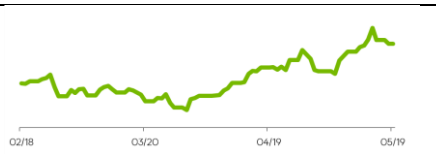
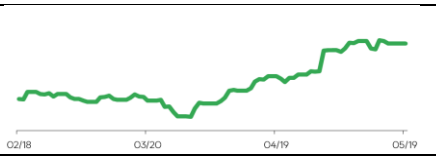
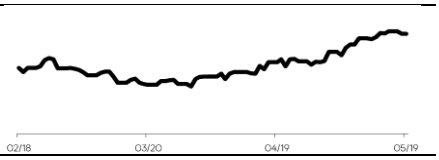
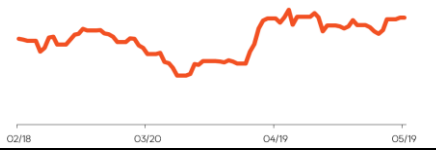
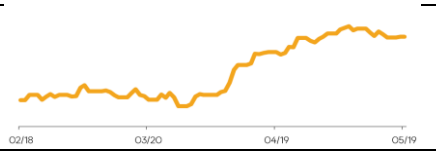
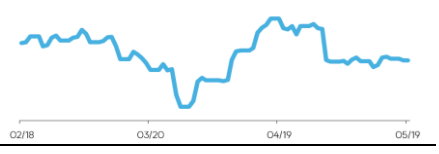
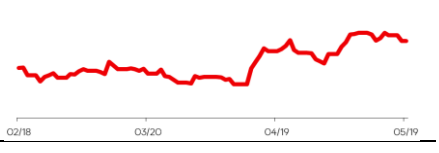
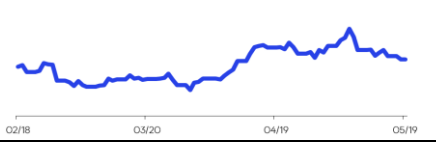
2026-07-01

박제민의 Q&A.I

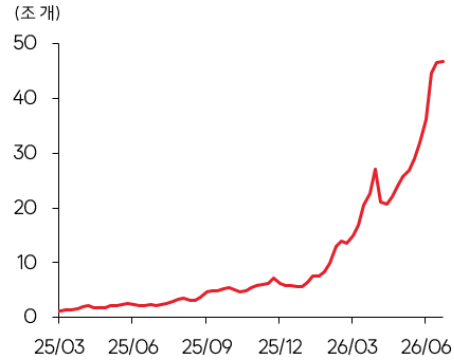
SK securities
SK증권 해외주식/AI. 박제민
jeminwa@sks.co.kr

	엔비디아	시가총액: 5,385	12MF PER: 26	
		-독보적인 GPU 기술력있으나 ASIC으로 독점력 축소 -Rubin Ultra 기술 전환 및 CPU, LPU 주문 추이 주목		
	알파벳	시가총액: 4,787	12MF PER: 28	
		- TPU 활용 Anthropic 독주로 인프라 해자 인정받는 중 - Coding Agent 역전, 소비자용 Agent 출시 주목		
	애플	시가총액: 4,374	12MF PER: 34	
		-메모리 가격 상승이 브랜드 파워로 극복 가능한지 시험 중 -Siri 고도화로 인한 소비자용 Agent 출시 주목		
	마이크로소프트	시가총액: 3,146	12MF PER: 25	
		-OpenAI 해자 약화로 인프라 사업부 성장률 축소 -자체 모델 및 커스텀 AI 도입 컨셉 수요 및 기술력 주목		
	아마존	시가총액: 2,849	12MF PER: 26	
		-양대 AI Labs 모두에 공격적 투자를 통한 고객 확보 성공 -인프라 외 Bedrock 등 AIaaS ARR 전망에 주목		
	메타 플랫폼스	시가총액: 1,552	12MF PER: 17	
		-플랫폼 인기 여전, Muse Spark 개발로 유통과 모델 확보 -플랫폼 데이터, 글래스 등 활용한 소비자용 Agent 주목		
	오라클	시가총액: 557	12MF PER: 25	
		-부품 비용, 자본 비용 증가 추세 속 높은 수주 잔고 무색 -인력, 전기 쇼티지에도 매출 전환 성공 여부 주목		
	코어위브	시가총액: 57	12MF PER: -	
		-부품 비용, 자본 비용 증가 추세 속 높은 수주 잔고 무색 -인력, 전기 쇼티지에도 매출 전환 성공 여부 주목		

주: 시가총액 십억달러

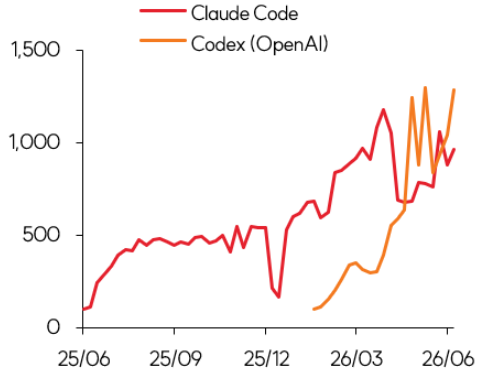
Update Chart

OpenRouter 토큰 사용량



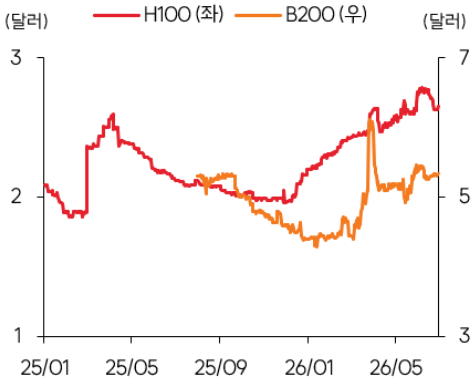
자료: OpenRouter, SK 증권

주요 Coding Agent 활용 강도 (백만회 달성일 = 100)



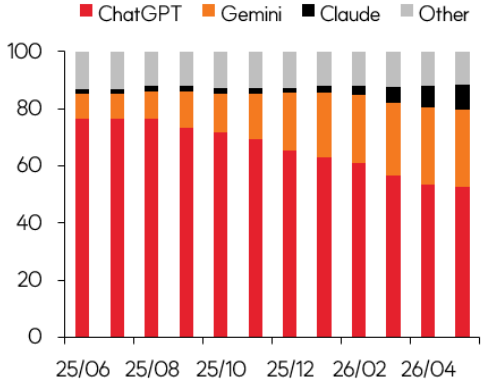
자료: NPMjs, SK 증권 / 주: 업데이트 및 다운로드 수 기준. Codex 5 월 수치 일부 조정

GPU 렌탈 가격 추이 (SiliconData)



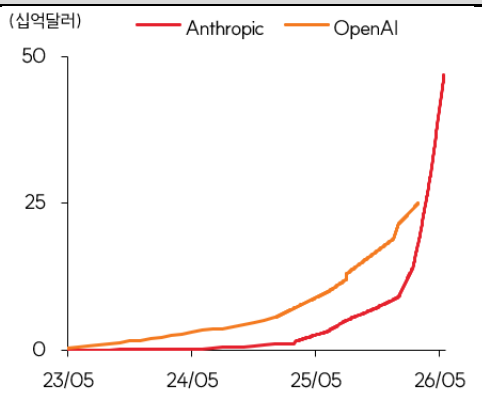
자료: Bloomberg, SK 증권

LLM 별 점유율 추이



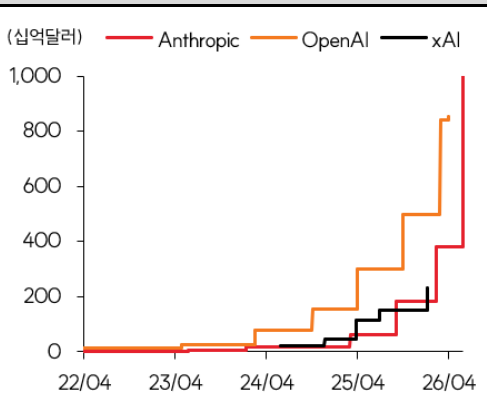
자료: SimilarWeb, SK 증권

주요 AI Labs ARR 추이



자료: EpochAI, SK 증권

주요 AI Labs 기업가치 추이



자료: EpochAI, SK 증권

Coding Agent, 얼마나 왔을까?

Coding Agent 는 Token 사용을 양극화 시키는 중

OpenAI 에서 Agent 활용 동향 보고서 공개 (How agents are transforming work)

1) 여전히 낮은 Coding Agent 사용자 침투율 2) LLM 대비 높은 Coding Agent 의 Token 사용 강도가 보이는 보고서

현재 OpenAI 사용자 중 Codex 사용자는 가파른 성장에도 '개인 고객 중 0.7%', '기관 고객 중 17%'

출력되는 토큰 중 Codex 의 비중은 개인 고객 16%, 기관 고객 63%

0.7%의 개인이 전체의 16%, 기관의 17%가 절반 이상의 토큰을 활용. Agent 사용자가 개인 28 배, 기관 8 배 이상 토큰 소모

2026 년 6 월 기준 상위 1% 사용자는 평균적으로 에이전트 실행 시간(Codex Agent Turn)이 하루 60 시간 초과

한 사용자가 10 개의 Agent 를 6 시간씩 활용하거나 20 개 이상의 Agent 를 3 시간 미만 동안 활용했다는 의미

향후에도 Coding Agent 확산은 '사용자 수 증가'보다 '단일 사용자의 Agent 활용 고도화'에서 나올 것으로 판단

OpenAI 내부 기준 개발자(Engineer) 뿐 아니라 사무직(Finance, HR, Legal) 부문의 활용도 적극적으로 늘어나는 중

사무직의 Codex 활용자 비중은 2026 년 4 월까지 20% 수준에서 두 달만에 90% 수준으로 급등

Codex 활용 효용이 명백히 보이는 수치. 그러나 외부 기관 고객 입장에서는 사무직의 data 를 AI 모델에 입력하기 위한 인프라, 보안, 직원 툴 등이 추가로 필요. 사무직 데이터는 대체로 기업 내부 데이터로 외부 유출에 민감

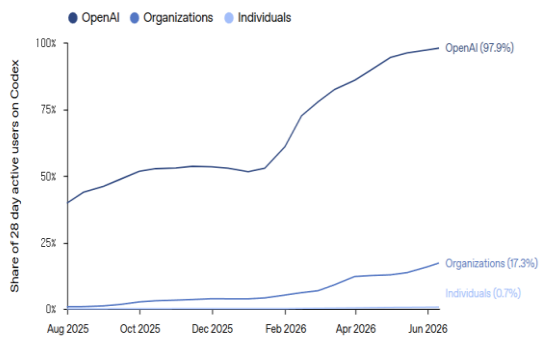
OpenAI 는 외부 업무용 Agent 를 기업 내부 데이터에 병합하기 위해 FDE(Forward Deployed Engineering)팀으로 대응 중

FDE 엔지니어는 고객사에 상주하며 수개월~수년 AI 도입을 위한 문제 해결. 팔란티어가 창안한 역할

OpenAI 는 2026 년 5 월 별도 FDE 법인을 설립하고 AI 컨설팅 기업인 Tomoro 를 인수, 150 명의 컨설턴트 확보

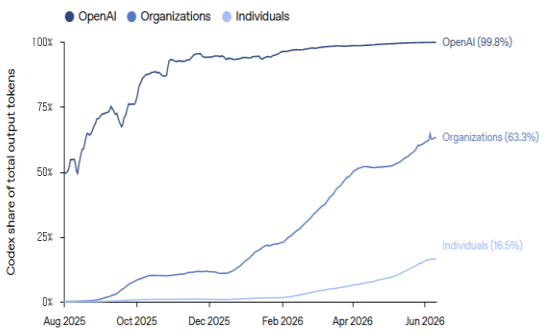
경쟁사인 Anthropic 은 직접 인력을 꾸려 대응하기보다 금융사, 컨설팅사와 JV 로 대응 중

OpenAI 고객 중 Codex 사용 활성 사용자 비중



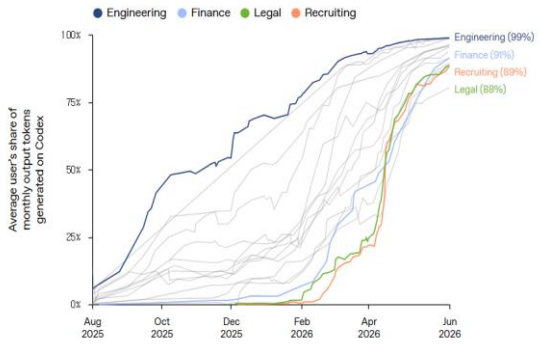
자료: OpenAI, SK 증권

전체 토큰 사용 중 Codex 의 출력 토큰 비중



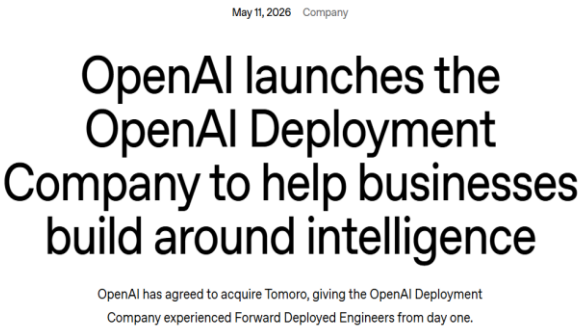
자료: OpenAI, SK 증권

OpenAI 부서별 Codex 활용률



자료: 산업 자료, SK 증권

비개발자용 업무용 Agent 도입을 위해 FDE 법인 창설



자료: OpenAI, SK 증권

Token 사용량이 부담이 되기 시작하자…

Coding Agent 사용자들의 높은 토큰 소모량으로 관련 비용도 커지는 중 (=Anthropic, OpenAI ARR 급증)

Anthropic 은 초기 파트너인 Amazon 에게 더 높은 Claude 사용 비용을 청구하기 위해 재협상을 시도 중이라는 보도

시간 기반(Hours)에서 사용 기반(Tokens) 과금으로 조건 변화 제시. Amazon 은 대응책으로 자체 모델인 Nova 모델을 고려 중

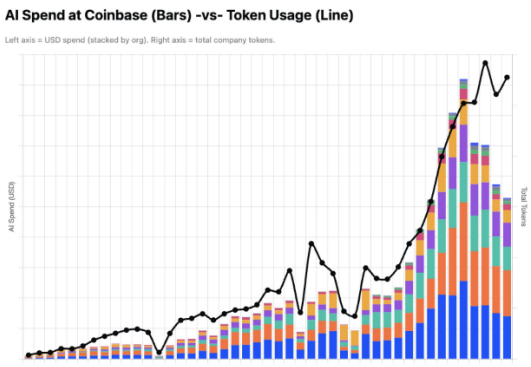
Coinbase CEO 는 급증하는 AI 비용에 맞서기 위한 팁을 X 에 게시. AI 비용 절약 플레이북으로 크게 바이럴

제시된 팁의 골자는 1) 기본 모델을 싼 모델로 활용 (Better Default) 2) 작업별로 맞는 모델 자동 배정 (Better Routing) 3) 반복 요청 효율화 (Better Caching) 4) 쓸데 없는 문맥 축소(Keep Context Lean) 5) ROE 평가를 위해 직원별 사용량 공개

Coinbase 는 5 월 실적 발표에서 700 명 (전사 직원의 14%) 감원을 발표, CEO 가 AI 도구 활용을 직접 원인으로 언급

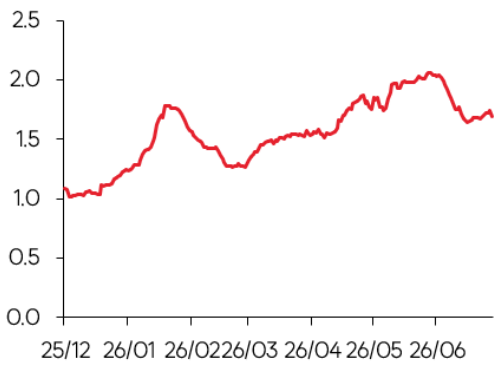
Silicon Data 에서 추정하는 LLM Expenditure Index(LLM 활용 평균 단가)의 하락세도 AI 효율화로 인해 하락 중으로 판단

Coinbase CEO: AI 토큰 비용을 절감하는 법



자료: X(@brian_armstrong), SK 증권

LLM 토큰 평균 가격 단가 추이



자료: Bloomberg, SK 증권

OpenAI 업데이트: IPO 연기와 모델 출시, 자체 칩 준비까지

IPO 연기: 본질은 속도 vs 밸류에이션 방어

OpenAI 가 내년까지 IPO 연기를 고려 중

자문단이 $1T 가치 평가를 위해 기간을 미루거나 가격을 낮출 것을 제안했으나 샘 알트먼이 밸류에이션을 지키는 길 선택

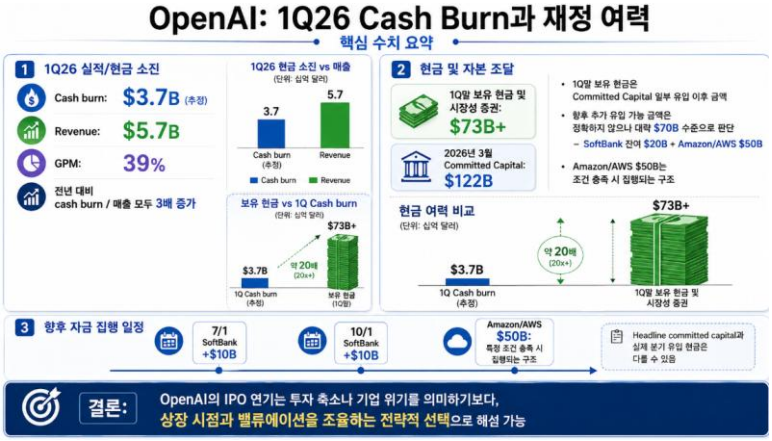
OpenAI 의 1Q26 Cash burn 은 $3.7B(YoY +300%), 동 기간 매출은 $5.7B, GPM 39%

1Q26 기말 보유 현금은 $73B 이상. 올해 3 월 라운드에서 $122B Committed Capital 확정. 이 중 일부만 1 분기 환입 판단

2026 년 7 월, 10 월 에 Softbank 로부터 추가 $20B 집행 예정. Amazon 의 $50B 또한 특정 조건 충족 시 투자 집행

따라서 OpenAI 의 풍부한 재정 여력을 고려할 때 IPO 연기는 투자 축소나 기업 위기를 의미한다고 보기 어려운 상황

OpenAI 재정 여력



자료: TheInformation, SK 증권

ChatGPT 5.6 프리뷰 공개: Mythos 에 버금가는 성능?

GPT 5.5 공시 발표 이후 64 일만에 신규 모델 발표

GPT 5.6 은 3 개 모델로 제공. Sol, Terra, Luna 순으로 하이엔드 모델

중간 모델인 Tera 모델은 GPT 5.5 수준의 성능을 2 배 낮은 가격에 제공한다는 컨셉

트럼프 행정부에서 보안 우려로 GPT-5.6 출시를 단계적으로 하라고 요청, 초기 기간에는 선별된 파트너에게 제공 후 API 확산

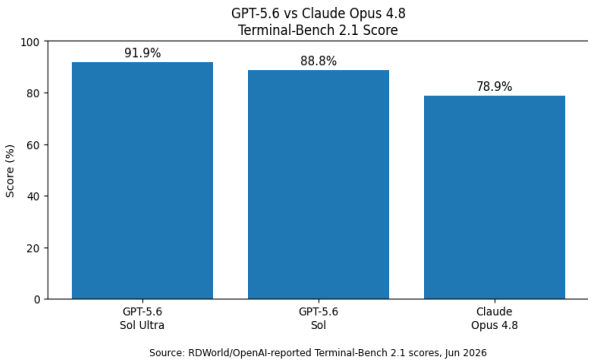
아직 프리뷰 단계로 공개 벤치마크 부족하지만 현재 Claude 대표 모델인 Opus 4.8 대비 Coding Agent 관련 벤치마크 우위

사용자가 반복 입력 구간을 설정할 수 있는 explicit cache breakpoint 기능을 최초 지원. 사용자 비용 절감에 초점을 두는 모습

OpenAI 신규 모델 출시

자료: OpenAI, SK 증권

OpenAI 신규 모델과 Claude 모델 비교



자료: OpenAI, Anthropic, SK 증권

OpenAI 추론 자체칩 할라피뇨: 브로드컴의 시험대

브로드컴 협력 LLM 추론용 AI ASIC 할라피뇨(Jalapeño) 공개

브로드컴 CEO 는 초기 칩아웃 기준 와트당 성능이 Blackwell, Googel TPU 수준이라고 언급

1차 규모 1.3GW ($18B) 전체 10GW 확산 예정. 칩 생산비만 $180B 추정

칩 출하 시 40%를 마이크로소프트가 의무 구매. 해당 계약을 통해 브로드컴이 칩 생산 비용을 위한 금융을 받는 구조 고려

기존 목표는 2026 년 하반기 온라인이었으나 2027 년 이후로 연기

Nvidia, Amazon 을 상대로 칩 공급 교섭력이 높아지는 계기가 될 수 있을지 주목 필요

OpenAI 인프라 관련 계약 정리

파트너 기업	계약 규모	전력 용량
Oracle	2032 년까지 약 $300B	6GW
Microsoft	2032 년까지 $250B	N/A
Broadcom (ASIC)	1차 생산 단계 기준 약 $18B	전체 10GW / 1차 단계 1.3GW
Amazon Web Services	8 년간 $138B	AWS Trainium 용량 2GW 포함
SoftBank / SB Energy	N/A	1.5GW
CoreWeave	5 년간 $22B	N/A
Google Cloud	조건 미공개	N/A
Cerebras	3 년간 $20B 이상	1GW 초과
Advanced Micro Devices	조건 미공개; OpenAI 가 AMD 주식을 수령하는 구조	최대 6GW

자료: TheInformation, SK 증권

DeepSeek Funding 시작: 저가 모델 압박이 가속화되는 중

오픈 소스 모델의 선두 주자 DeepSeek 의 강한 성장 행보가 지속되는 중. 저가 모델의 성능 향상이 가시권

Anthropic 대비 아직 평가가 좋지 못함에도 비슷한 수준의 비용을 과금하는 중인 OpenAI 에게 특히 큰 압박

중국 AI Labs 딥시크가 $50B 이상 밸류에이션으로 $7.4B 조달 (TheInformation)

딥시크는 2026 년 4 월 공개한 DeepSeek V-4 모델이 현재 Gemini 3.1 과 유사한 4~5 위권 수준의 모델 성능을 보여주는 중

5/23 딥시크는 V4-Pro API 가격 75% 인하를 영구화, 100 만 토큰 당 $0.0035~$0.8 (GPT, Anthropic 대비 30 배 저렴)

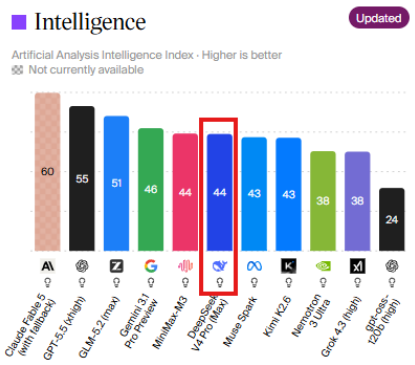
6/25 딥시크가 AI 핵심 R&D 인력을 포함하여 전 부서 인력을 최소 2 배 확대한다는 계획 보도

100 만 토큰 당 가격에서 볼 수 있듯, 저가 모델(오픈소스 모델)의 토큰 당 가격 차이는 30 배 이상

개인 개발자들이 가격 최적화를 주로 활용하는 OpenRouter 순위에는 중국 기업들이 대부분 위치

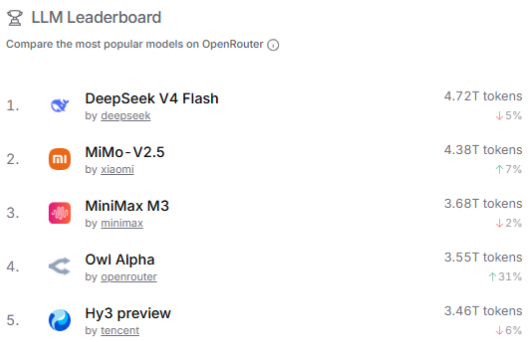
OpenAI 는 Coding Agent 의 압도적 성능 + FDE 팀을 활용한 기업 내부 데이터 활용 퍼포먼스가 필요

Artificial Analysis Intelligence Index



자료: Artificial Analysis, SK 증권

OpenRouter AI 모델 순위



자료: OpenRouter, SK 증권

Compliance Notice

작성자(박제민)는 본 조사분석자료에 게재된 내용들이 본인의 의견을 정확하게 반영하고 있으며, 외부의 부당한 압력이나 간섭없이 신의성실하게 작성되었음을 확인합니다.

본 보고서에 언급된 종목의 경우 당사 조사분석담당자는 본인의 담당종목을 보유하고 있지 않습니다.

본 보고서는 기관투자가 또는 제 3 자에게 사전 제공된 사실이 없습니다.

당사는 자료공표일 현재 해당기업과 관련하여 특별한 이해 관계가 없습니다.