

2025. 07. 16

금융으로
만나는 새로운 세상

IBKS In-Depth Report

닥컴이 예언한 NEXT AI

인터넷/게임 **이승훈**

02) 6915-5680

dozed@ibks.com

RA **박은명**

02) 6915-5527

eunp9622@ibks.com



IBK기업은행 금융그룹

IBK투자증권

본 조사분석자료는 당사 리서치본부에서 신뢰할 만한 자료 및 정보를 바탕으로 작성한 것이나 당사는 그 정확성이나 완전성을 보장할 수 없으며, 과거의 자료를 기초로 한 투자참고 자료로서 향후 주가 움직임은 과거의 패턴과 다를 수 있습니다. 고객께서는 자신의 판단과 책임하에 종목 선택이나 투자시기에 대해 최종 결정하시기 바라며, 본 자료는 어떠한 경우에도 고객의 증권 투자 결과에 대한 법적 책임소재의 증빙자료로 사용될 수 없습니다.

CONTENTS

AI, 진짜는 시작도 안 했다.....	3
Key Chart.....	5
1. 반복되는 테크 사이클.....	15
2. 닷컴 이후 기술 확산 구조.....	18
3. AI의 현재 위치.....	27
4. 2025년 AI 트렌드	31
5. NEXT AI: 2차 설치기와 전개기 전망.....	46
6. AI산업의 캐즘 가능성이 크지만 붕괴는 없다.....	53
7. 2025년 대한민국의 AI정책.....	62
8. 21세기 AI 기술 진화.....	65
9. 기업분석.....	82
NAVER (035420).....	82
카카오 (035720).....	91
알파벳 (GOOGL US).....	97
메타 (META US).....	102
마이크로소프트 (MSFT US).....	107
아마존 (AMZN US).....	111
에스티로더 (EL US).....	116

AI, 진짜는 시작도 안 했다

AI의 현주소 진단과
차세대 AI 전망

이 보고서의 아이디어는 2022년 11월 OpenAI가 ChatGPT를 출시한 이후, 글로벌 빅테크 기업들이 고유의 초거대언어모델(LLM)을 어떤 전략으로 발전시켜왔는지를 체계적으로 정리하는 데서 출발했다. 다양한 측면에서 논의를 전개한 끝에, 현재 AI 산업이 기술 전환의 어느 단계에 위치해 있는지 진단하고, 다음 세대 인공지능(NEXT AI)의 방향을 전망하는 데 초점을 맞췄다.

- 1) 닷컴부터 AI 이전 분석
- 2) 현재 AI 산업 위치 진단
- 3) LLM 진화 과정
- 4) AI 트렌드 · 향후 전망

▲닷컴 이후 기술 전환의 경로 분석, ▲현재 AI 산업의 위치 진단, ▲최근 AI 트렌드 도출을 통해 NEXT AI의 방향성 제시, ▲주요 기업들의 LLM 진화 계보를 정리한다. 아울러 AI 산업의 거품 붕괴 가능성과 한국의 AI 정책 현황, 국내외 주요 AI 기업들의 전략도 함께 담았다.

IBKS 기술 확산 4단계:
① 1차 설치기 ② 캐즘
③ 2차 설치기 ④ 전개기

우선 기술 전환의 맥락을 설명하기 위해 칼로타 페레스(Carlota Perez) 교수의 ‘기술 혁명과 금융 자본’ 이론을 응용(설치기, 전개기의 2단계)하였다. 1995년 닷컴 혁명의 시작부터 혁신, 버블 붕괴, 모바일·클라우드 확산에 이르는 흐름을 분석한 결과, 이를 발전시켜 IBK투자증권 고유의 4단계(1차 설치기, 캐즘, 2차 설치기, 전개기) 기술 전환 프레임워크를 제안한다.

AI 현재 위치,
캐즘 단계 진입

현재 AI 산업은 ‘캐즘(Chasm)’ 단계에 진입한 것으로 판단된다. 데이터센터 규모와 ChatGPT 사용자 수는 급격히 확대되었지만, 아직 AI 서비스는 기존 인터넷·모바일 환경에서 창출된 부가가치를 넘어서지 못하고 있다. 이는 마치 인터넷 보급 이전의 PC 활용, 아이폰 출시 이전의 피쳐폰 사용 시기와 유사한 상태라 할 수 있다.

2025년 AI 변곡점
딥시크 등장

2025년 1월 딥시크의 R1 모델 공개는 대규모 AI 투자의 정당성에 의문을 제기하며 투자 심리를 위축시켰고, 최근 메타의 MSL(Meta Superintelligence Labs) 설립은 인건비 상승과 오픈소스 진영의 불확실성을 증폭시키고 있다. 미국과 중국 간 AI 주도권 경쟁도 가속화되고 있으며, 중국의 양적 성장은 이미 미국을 위협할 수준에 도달했다.

Next AI:
에이전틱(Agentic) AI와
피지컬(Physical) AI

NEXT AI의 단초는 최근 주목받고 있는 ‘에이전틱 AI(Agentic AI)’와 ‘피지컬 AI(Physical AI)’에서 포착된다. Agentic AI는 AI가 단순한 챗봇 수준을 넘어 인간의 의도를 최소한으로 파악하고, 다양한 소프트웨어와 하드웨어 도구를 활용해 자율적으로 목표를 달성하는 시스템으로의 진화를 뜻한다. 이러한 AI가 로봇 등 하드웨어와 결합해 실질적인 물리적 작업을 수행하는 것이 Physical AI이다.

AI 밸리,
닷컴 버블과는 다르다

주식 시장에서 AI 산업의 붕괴는 과거 닷컴버블과 같은 방식으로 재현되기는 어려울 것으로 판단된다. 현재 빅테크 기업들은 AI 외에도 확실한 수익 모델을 보유하고 있으며, 상장된 AI 스타트업 수도 제한적이다. 무엇보다 AI 기술은 이미 상용화가 진행 중이며 성숙 단계에 접어들었다. 물론 금리·경제 환경에 따라 AI 관련 기업의 밸류에이션 하락과 스타트업 도산이 이어질 수 있으나, 이 과정을 거친 이후에는 과거와는 차원이 다른 형태의 서비스 기반에서 폭발적인 성장이 가능할 것으로 기대된다.

(다음 페이지로)

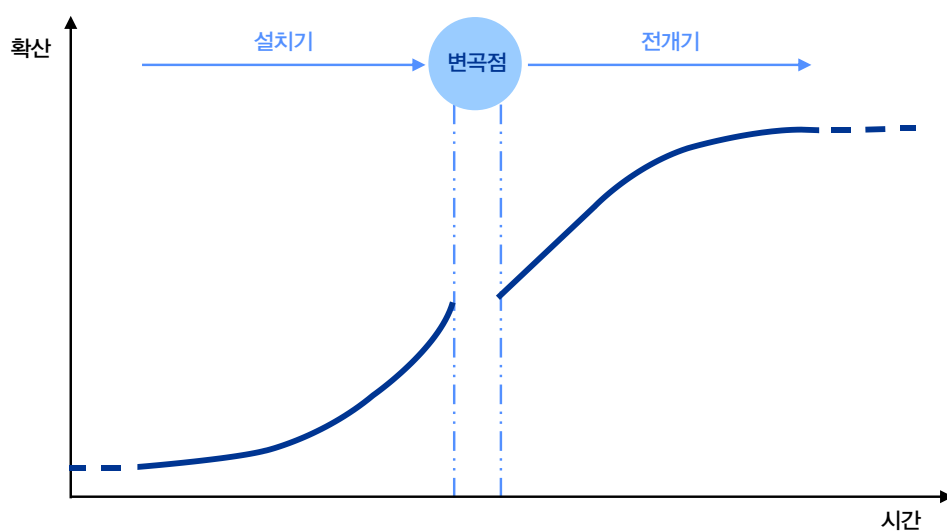
AI 정부 정책 본격화	<u>한국의 AI 산업은 과거 2년간 정책적 공백이 있었으나, 민간기업이 축적해온 자본과 기술 위에 이제부터 국가 정책이 본격적으로 결합되며 새로운 성장의 르네상스가 열릴 것으로 보인다.</u>
2022년 이후 AI 초기: LLM 고도화, 멀티모달 진화, 폐쇄 전략	2022년 이후 빅테크 기업들은 LLM의 매개변수를 극단적으로 확장하는 한편, 텍스트 중심의 AI에서 이미지·오디오·비디오를 아우르는 <u>멀티모달 AI로 진화</u> 하였다. 막대한 자본 투입 속에 메타를 제외한 구글, OpenAI, NAVER 등은 <u>주력 모델을 비공개로 운영</u> 하며, 자사 <u>주요 서비스에 자체 LLM을 통합</u> 하는 전략을 택하고 있다. 그러나 여전히 <u>대규모 데이터센터 및 AI 연구개발을 위한 투자</u> 가 필수적인 상황이다.
국내 기업 수혜 기대: NAVER, 카카오	국내에서는 자체 LLM과 플랫폼을 모두 보유한 NAVER, 국내 최대 사용자 기반과 생활 밀착형 서비스를 가진 카카오(향후 OpenAI 협업 기대)를 최우선 추천주(Top Picks)로 제시한다.
AI 혁신 주도 해외 기업: 알파벳(구글), 메타, 아마존	해외 기업 중에서는 지속적으로 AI 혁신의 전환점을 주도해온 알파벳, AGI를 넘어 ‘슈퍼 인텔리전스’ 시대를 준비 중인 메타, LLM 클로드(Claude)의 개발사 앤트로픽(Anthropic)의 최대 주주이자 Physical AI 구현 범위가 가장 넓은 아마존을 관심 종목으로 제시한다.

Key Chart

1. 반복되는 테크 사이클

카를로타 페레즈의 기술·경제 패러다임

그림 1. 페레즈는 기술이 '설치기-전개기'의 사이클로 확산된다고 주장

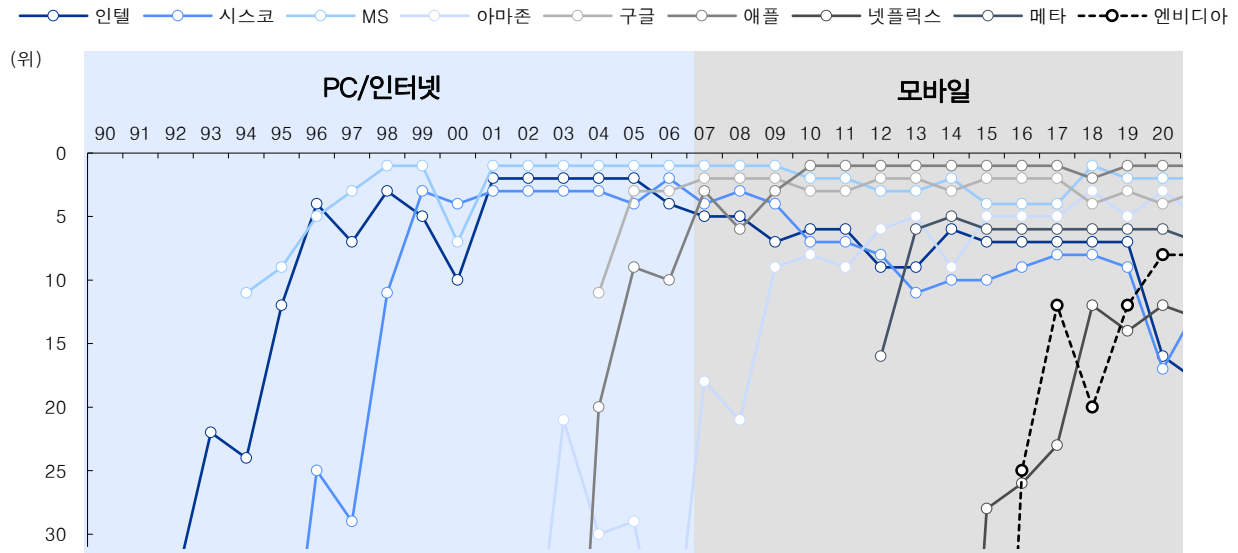


자료: 『Finance and technical change A long-term view』, IBK투자증권

2. 닷컴 이후 기술 확산 구조

닷컴/모바일 시대에도 반복되는 구조

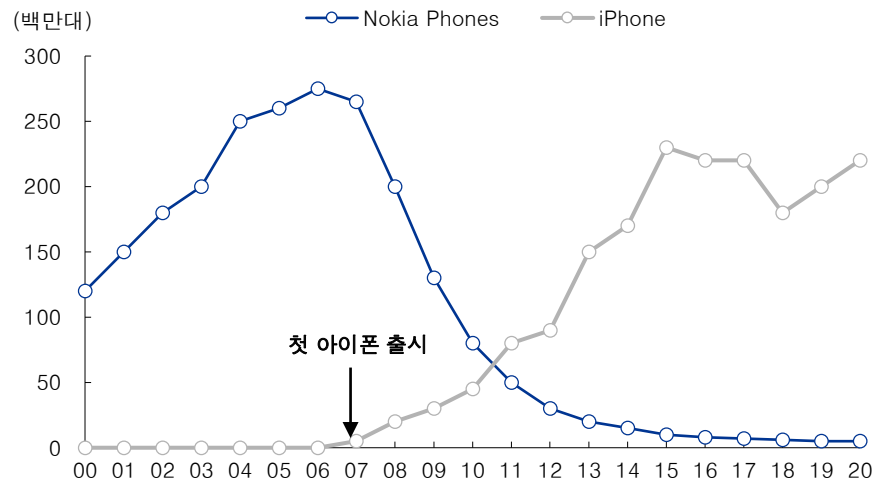
그림 2. 닷컴과 모바일 시대에도 사이클에 따라 대장주가 변동



자료: Bloomberg, IBK투자증권

캐즘(Chasm): 수요 정체 구간

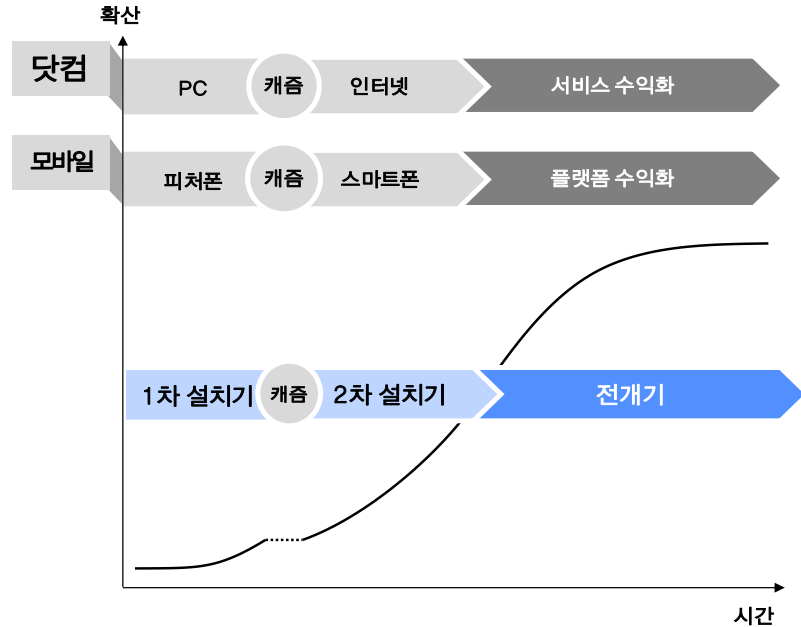
그림 3. 모바일 초기에는 노키아 폰 판매량이 높았지만, 추후 아이폰 판매량이 추월함



자료: IBK투자증권

IBKS가 제시하는 테크 사이클: 1차 설치기-캐즘-2차 설치기-전개기

그림 4. 4단계는 각각 ‘기존 기술의 설치-정체-패러다임 전환-서비스 대중화’를 의미



자료: IBK투자증권

3. AI의 현재 위치

‘캐즘’에 돌입한 국면

그림 5. 현재의 AI 산업이 캐즘에 돌입했다고 보는 이유

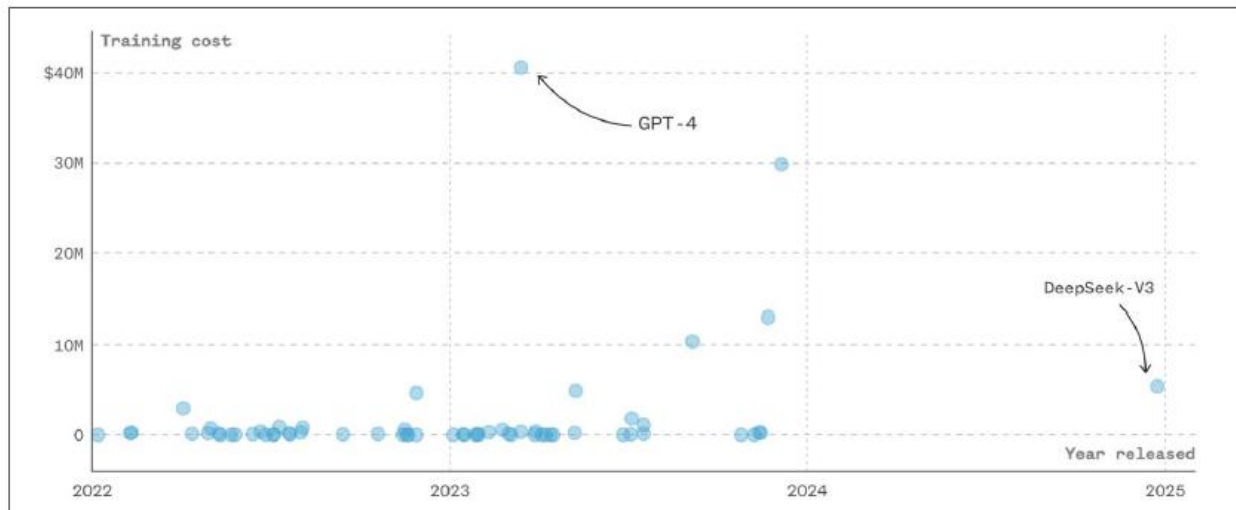
- 1 자본 투자가 인프라 설치와 기술 개발에 집중
- 2 기존 서비스의 연장선에 머물러 있음
- 3 유료 수익 모델 전환의 어려움

자료: IBK투자증권

4. 2025년 AI 트렌드

1) AI산업의 변곡점

그림 6. 딥시크발 낮은 훈련 비용 충격



자료: BOND, IBK투자증권

표 1. 메타 슈퍼인텔리전스 인력 흡수 현황

구분	주요 내용
조직 신설	Meta Superintelligence Labs(MSL) 출범, 기존 AI 조직 통합
핵심 영입 인재	Alexandr Wang(Scale AI 창업자), Nat Friedman(전 GitHub CEO), Daniel Gross, Ruoming Pang(애플 AI 총책임자) 등
경쟁사 영입 인원	OpenAI, Anthropic, Google 등에서 10명 이상 핵심 연구자 영입 스위스 OpenAI 연구소 3인 합류
보상 패키지	최대 1억 달러 계약금, 4년간 3억 달러 스톡옵션, 무제한 컴퓨팅 리소스 등
투자 규모	2025년 한 해에만 AI 인재 및 인프라에 143억 달러(약 19조 원) 투자
업계 반응	경쟁사들도 인재 유출 방지책 마련, AI 인재 확보 경쟁 심화

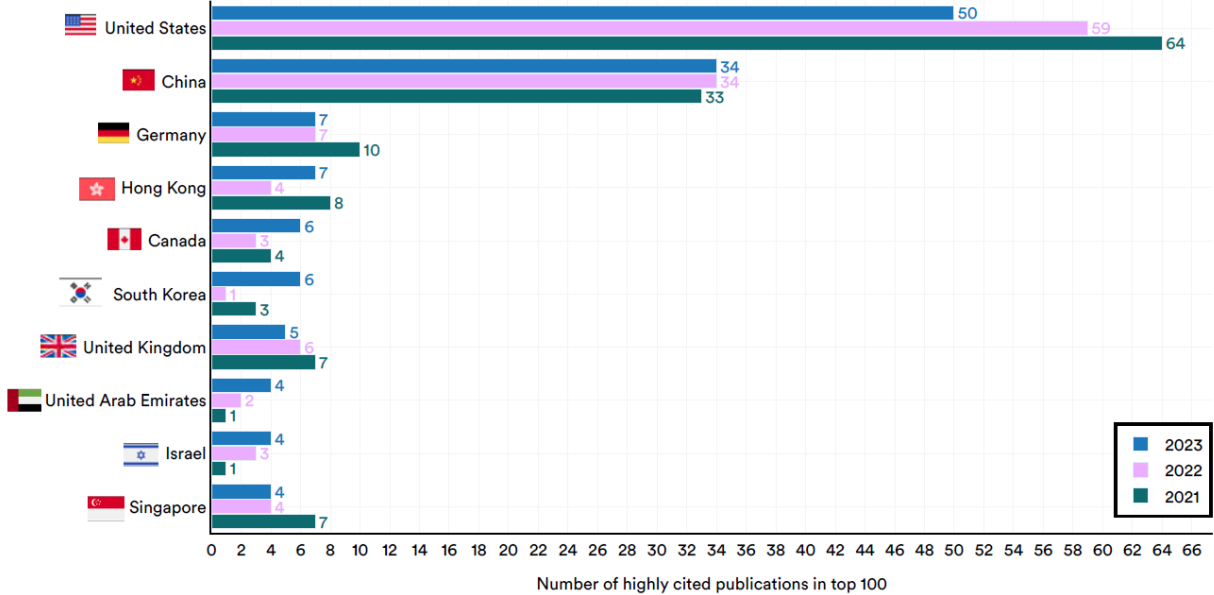
자료: 언론 보도 종합, IBK투자증권

2) AI Index 2025 보고서에 담긴 인사이트

그림 7. 상위 인용 AI 논문 100편의 글로벌 비교

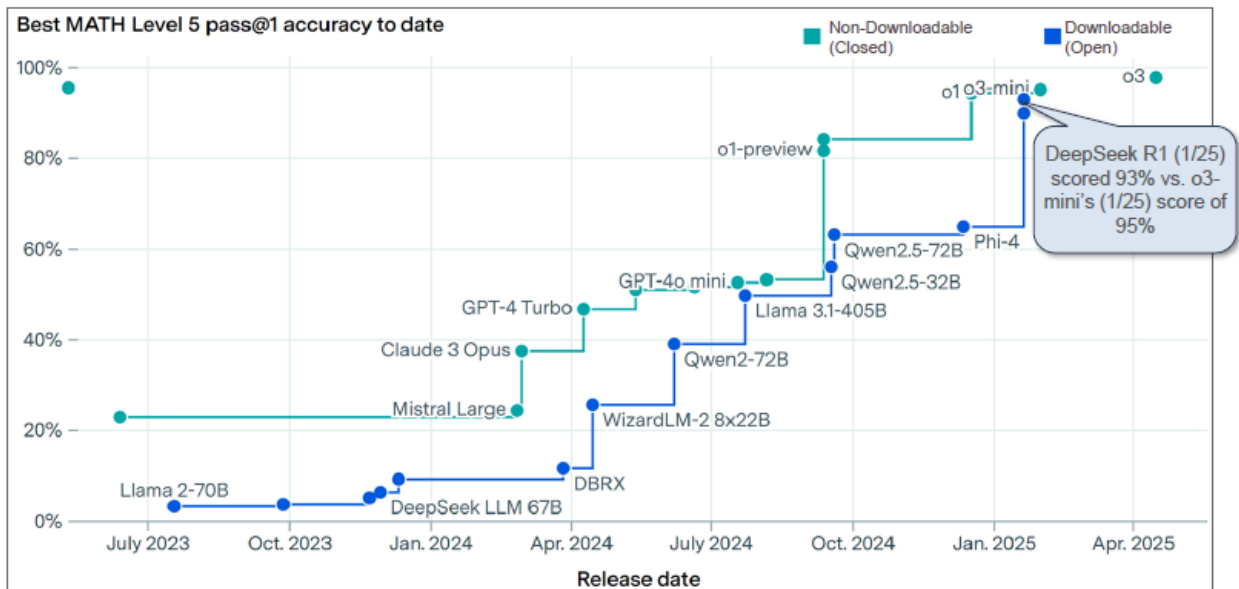
Number of highly cited publications in top 100 by select geographic areas, 2021-23

Source: AI Index, 2025 | Chart: 2025 AI Index report



자료: Stanford HAI 『AI Index Report 2025』, IBK투자증권

그림 8. 오픈소스모델 성능 향상에 기여

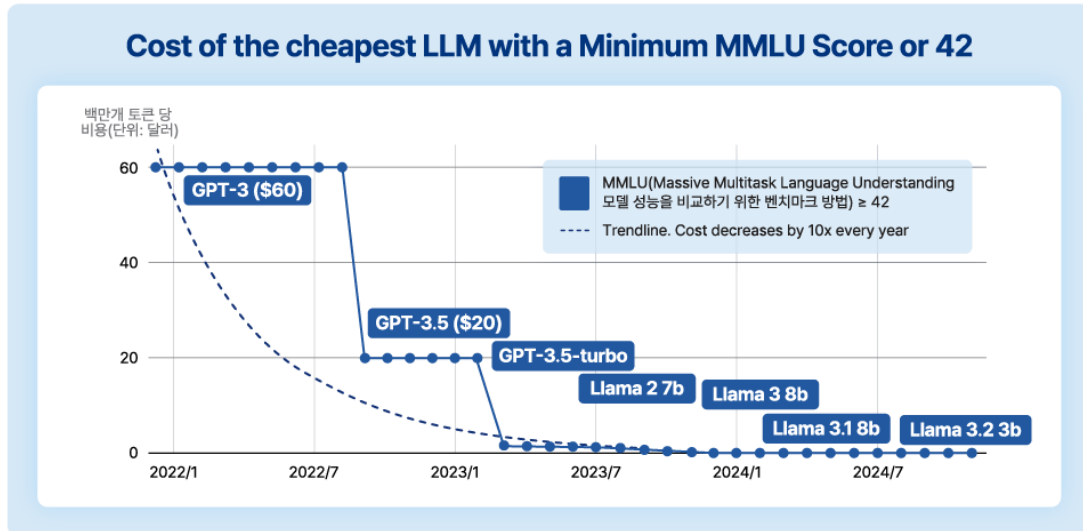


자료: BOND, IBK투자증권

5. NEXT AI: 2차 설치기와 전개기 전망

1) 최근 AI 산업에서 ‘추론’의 중요성이 급부상

그림 9. LLM추론 비용 하락



자료: IBK투자증권

2) Agentic AI

표 2. 샘알트만이 제시한 AI의 5단계

단계	설명
Chatbot	단순 대화형 AI. 질문에 답하거나 정보를 제공하는 수준의 AI. 인간과 자연스럽게 대화할 수 있지만, 복잡한 추론이나 자율적 행동은 불가능
Reasoner	논리적 추론 및 단계별 문제 해결이 가능한 AI. 복잡한 수학, 과학, 프로그래밍 등에서 단계적 논리 추론을 수행하며, 자신의 추론 과정을 설명할 수 있음
Agent	실제 세계에서 자율적으로 업무를 수행하는 AI. 인간의 개입 없이 특정 작업을 완수할 수 있으며, 소프트웨어 엔지니어링 등 실제 업무 현장에서 가상 동료로 활동할 수 있음
Innovator	새로운 과학적 지식이나 혁신을 스스로 창출하는 AI. 기존 지식을 조합하거나 새로운 아이디어를 만들어내는 창의적 능력을 갖춘
Organization	인간 조직의 복잡한 기능을 복제하는 다중 에이전트 시스템. 여러 AI가 협력하여 하나의 조직처럼 움직이며, 인간 조직의 역할과 구조를 대체할 수 있음

자료: Google, IBK투자증권

표 3. AI 에이전트와 에이전틱 AI의 비교

특징/분류	AI 에이전트	에이전틱 AI
자율성 수준	특정 규칙에 따라 반응하며, 제한된 자율성을 가짐	최소한의 인간 개입으로 복잡한 목표를 자율적으로 추구.
목표 복잡성	특정 업무 수행 및 사용자 요청에 반응	복잡하고 다단계의 목표를 설정하고 달성.
핵심 능력	환경 인지, 특정 작업 수행, 즉각적인 반응	도구 사용, 자기 반성, 계획 수립, 다중 에이전트 협업.
학습 및 적응	주로 정해진 데이터나 규칙 기반 학습	동적으로 환경에 적응하고, 경험을 통해 학습 및 개선.
의사결정	사전에 정의된 규칙이나 학습된 패턴에 따라 결정	고도화된 추론과 계획을 통해 스스로 의사결정.
예시	콜센터 상담 도우미, AI 비서, 개인화 챗봇	자율 금융 자문 시스템, 산업별 맞춤형 에이전트

자료: Google, IBK투자증권

표 4. LLM과 LAM비교

항목	LLMs	LAMs
핵심 기능	패턴 기반 텍스트 생성	API 및 인터페이스를 통해 작업 실행
기본 기술	고유한 모델 아키텍처 (트랜스포머)	독자적인 아키텍처는 아직 없음
학습 방식	방대한 텍스트 데이터셋을 통해 완전 학습	별도 학습 없이 LLM에 도구만 추가
모델 상태	널리 배포되었으며 지속적으로 학습 중	대부분 개념적 수준, 현재 LLM 위에 구축됨
자율성	반응형: 명시적 프롬프트 필요	준자율형: 단단계 작업을 수행 가능
상호작용 방식	채팅 또는 텍스트 기반 인터페이스	앱 기반 또는 명령어 기반 상호작용
대표 사례	GPT-4, Claude, Llama, Gemini	Rabbit R1(LLM 기반), AutoGPT 일부, Devin(제한적 자율성)
기술적 현실	기술적으로 확립됨	LLM + 커스텀 인터페이스 + 앱 연결 방식

자료: Google, 언론 보도 종합, IBK투자증권

6. AI산업의 캐즘 가능성이 크지만 붕괴는 없다

현재 AI 산업의 차별점

표 5. 닷컴 버블(1999~2001)과 현재 AI 붐(2023~2025)의 주요 특징

비교 항목	1999~2001 닷컴 버블	2023~2025 AI 붐
투자 열풍 양상	.com 불은 IPO 500여 개, 첫날 주가 3~4배 ↑ 빈발; VC 자금 40% 인터넷으로	OpenAI 등 유니콘 등장; VC 자금 2024년 \$560억 (2배 ↑); AI 언급 기업 주가 급등 (테마 편승)
수익 구조/모델	상당수 기업 매출 · 이익 부재, 사용자는 있으나 수익화 실패	빅테크는 AI 즉각 수익 적으나 본업 이익 견조; 스타트업도 API 판매, 구독 모델 등 시도 (아직 초기)
거품 원인	저금리 자금 과잉, 투기/과신; “선점하면 이익 따라온다” 논리; 광고 등 과지출	금리인하 등 유동성이 다시 증가. 혁신 과대평가 가능성; “AI 패권 잡아야” 경쟁 심리; 인재 · GPU 쟁탈 비용 부담
거품 붕괴 촉발	금리 인상 + 실적 미스에 투심 급변 (내재가치 의심 폭발); 자금 경색으로 줄도산	일각 버블 경고 증대; 수익 부진하거나 경제 둔화 시 주가 조정 가능성

자료: Google, 언론 보도 종합, IBK투자증권

그림 10. 닷컴과 AI시기의 NASDAQ100, S&P500 밸류에이션 비교

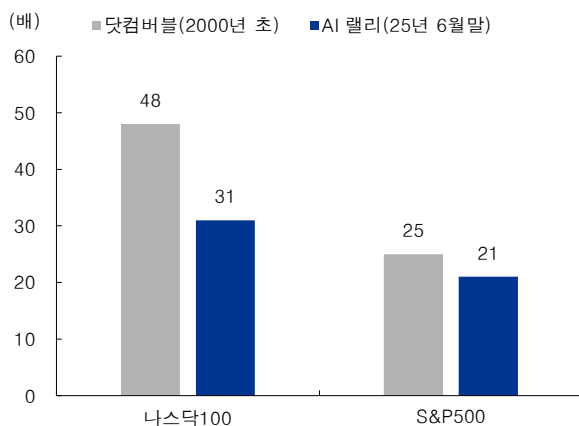
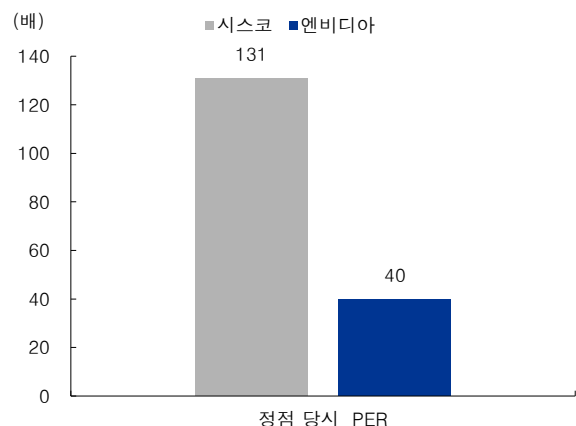
자료: Bloomberg, IBK투자증권
주: 12개월 선행 PER

그림 11. 정점 당시 시스코와 엔비디아의 밸류에이션 비교

자료: Google, IBK투자증권
주: 12개월 선행 PER

7. 2025년 대한민국의 AI정책

새 정부, AI 강국 도약 준비 중

표 6. 새 정부는 AI 도약을 위해 100조 원 투자 계획을 밝히며 민간과 협력 선언

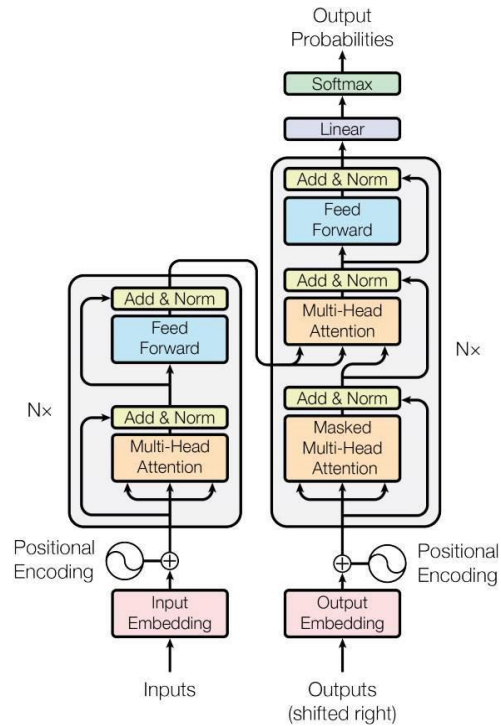
정책	공약	실제 진행 현황(25/7/14 기준)	목표 시기
소버린 AI 개발	한국어자체 데이터 기반 국가 주도 AI 모델 개발	과기정통부 주도 R&D 로드맵 수립 단계 예산 2조원 내외 책정 논의 중	2027
AI 예산 증액, 민간 투자 유도	AI 예산 증액 민간 투자 100조 원	국책은행 등 협력형 AI 펀드 조성 계획 발표 (10조원 규모)	2026
데이터센터 건설	'AI 고속도로' 구축	SK-AWS가 울산에 100MW급 AI 슈퍼 데이터센터 공동 건설 착수 (7조 원 규모)	2027년
GPU 확보	최신 GPU 최소 5만 개 이상 확보	국가 AI 컴퓨팅센터에 고성능 GPU 1만 대 확보 추진 중 추가 조달 계획 진행 중	2025년 말까지 1만 대 확보 목표
'모두의 AI' 프로젝트	범용 AI 확산, 민간 협력형 프로젝트	디지털 인재 양성 및 중소기업 대상 AI 솔루션 확산 전략 반영	-
AI 대표기업 육성	국가대표 AI 기업 발굴 및 투자	관련 기업 발굴 및 지원 위한 민관 합동 프로그램 논의 중	-
정부 주도 데이터 클러스터	정부 직접 데이터 클러스터 조성	AI 컴퓨팅 센터 구축에 2조 원 투입 예정 고성능 연산 인프라 확보 계획 포함	2026년 초
대통령 직속 AI 정책 조직 신설	대통령실 산하 'AI정책책임수석' 신설	2025년 6월 신설 완료	2025.6 완료
인공지능 기본법 조기 제정	AI 법제 정비, 위험기술 규제, 데이터 활용 기준 마련	'AI 기본법' 국회 본회의 통과	시행시기 조율 중
중기부 전면 개편	중기부 디지털 역량 강화 및 AI 산업 지원 구조 개편	사디지털 전문가 장관 후보 지명 혁신 펀드·지원책 마련 시사	2025년 하반기

자료: 언론 보도 종합, IBK투자증권

8. 21세기 AI 기술 진화

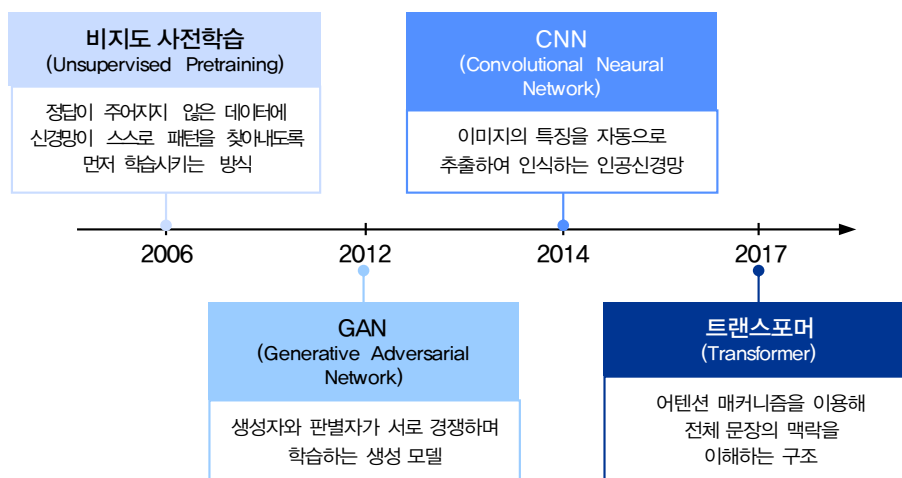
1) 생성형 AI 기술 진화

그림 12. 2017년 구글의 Transformer 원리



자료: 언론 보도 종합, IBK투자증권

그림 13. 생성형 AI의 초석이 된 AI기술



자료: IBK투자증권

표 7. 2018년 이후 구글, OpenAI, Meta 및 NAVER의 생성형 AI 관련 연대기

연도	회사	모델/서비스	특징	사양/기술
2018	Google	BERT	양방향 인코더 기반, 문맥 이해 강화	3.4억 파라미터, 인코더 구조
2018	OpenAI	GPT-1	최초의 GPT, 비지도 사전학습	1.17억, BookCorpus
2019	OpenAI	GPT-2	본격 언어 생성, 파라미터 10배 ↑	15억, WebText
2019	Meta	RoBERTa	BERT 개선, 성능 최적화	BERT 동일 구조, 마스킹 방식 변경
2020	OpenAI	GPT-3	Few-shot/Zero-shot 강화, 상용 API 시작	1,750억, 다중 도메인 데이터
2020	Meta	BlenderBot 1.0	대화형 모델, 장기기억+외부문서 통합	약 9.4억, 오픈도메인 대화
2021	Google	LaMDA	대화 특화 언어모델	디코더 구조, 맥락 이해 대화 강화
2021	NAVER	HyperCLOVA	한국어 GPT-3 계열, 상업용	2,040억 파라미터, 한국어 중심
2021	OpenAI	InstructGPT	RLHF 도입, 사용자 지시 이해	GPT-3 기반, 피드백 최적화
2022	Google	PaLM	초거대 모델, 디코더 전용 구조	5400억, Pathways 구조
2022	Meta	OPT	GPT-3 대응 공개 모델	66억~175억, 오픈소스
2022	OpenAI	GPT-3.5 / ChatGPT	GPT-3 기반 대화 최적화, 대중화	RLHF 기반 튜닝, ChatGPT 서비스
2022	DeepMind	Chinchilla	compute 최적 모델, 학습량 강조	70억 파라미터, 1.4조 토큰
2023.2	Meta	LLaMA 1	경량화 + 고성능, 연구 목적 공개	7B~65B, 오픈소스 (비상업용)
2023.3	Google	Bard (LaMDA 기반)	ChatGPT 대응 챗봇, 첫 공개	LaMDA Lite 기반, 검색 연동
2023.3	OpenAI	GPT-4	멀티모달 입력 (텍스트+이미지)	약 1T+, 128K 토큰 컨텍스트
2023.5	NAVER	HyperCLOVA X	한국어 reasoning 강화, 서비스형 전환	다국어 6조 토큰 학습, CLOVA X/Cue: 출시
2023.7	Meta + MS	LLaMA 2	상업 이용 가능 오픈소스, Azure 연동	7B/13B/70B, RLHF 기반
2023.12	Google	Gemini 1.0	본격 멀티모달, PaLM → Gemini 전환	최대 100만 토큰, 텍스트 · 이미지 · 코드 · 비디오
2024.4	Meta	LLaMA 3 (8B/70B)	GPT-4급 성능, 오픈소스 상업화	frontier-level 성능, 공개
2024.5	OpenAI	GPT-4o	음성+비전 실시간, 저지연 멀티모달	오디오 직접 처리, 1M+ 컨텍스트
2024.6	NAVER	CLOVA X Tools	플러그인 · 검색툴 · 토크링 기능 탑재	외부 시스템 통합 실행형 AI
2025.1	Google	Gemini 2.5 Pro	성능 개선 및 SGE 검색 연계 강화	추론 · 시뮬레이션 고도화, Duet · Gmail 탑재
2025.4	Meta	LLaMA 3.1 (405B)	최초 frontier급 오픈소스, AGI급 성능	4050억 파라미터, 완전 공개 예정

자료: IBK투자증권

표 8. 빅테크 기업의 LLM 모델 및 서비스 개선 과정의 공통점

공통 전략 항목	설명
1. 초거대 언어모델 경쟁	GPT 시리즈(OpenAI), Gemini(Google), LLaMA(Meta), HyperCLOVA(NAVER) 등 수십억~수천억 파라미터 규모의 LLM 개발을 통해 기술 경쟁 주도
2. 멀티모달 AI 확장	텍스트 중심에서 이미지 · 오디오 · 비디오 등 복합 입력을 처리하는 멀티모달 AI로 진화 중 (GPT-4o, Gemini 1/2, LLaMA 3.1 등)
3. 자사 서비스 통합	AI 모델을 검색(Google, NAVER), 광고(Google, Meta), 클라우드(OpenAI, Google), 커머스(NAVER) 등 핵심 서비스에 내재화
4. AI 인프라 투자 확대	GPU, 슈퍼컴퓨터, 데이터센터 등 고성능 인프라 구축에 수십~수백억 달러 투자 (OpenAI Stargate, Google TPU, NAVER GAK 세종 등)
5. API 및 플랫폼 전략	ChatGPT API, Gemini API, Azure + LLaMA API, NAVER Cloud API 등 클라우드 기반 개발자 생태계 확대 전략 추진

자료: IBK투자증권

1. 반복되는 테크 사이클

카를로타 페레즈의 기술·경제 패러다임

설치기와 전개기로 보는
역사적 기술 혁신 사이클

카를로타 페레즈(Carlota Perez) 교수는 2024년에 『Finance and Technical Change: A Long-term View』라는 논문을 발표하며 기술 혁신과 금융자본의 상호작용을 역사적 관점에서 분석했다.

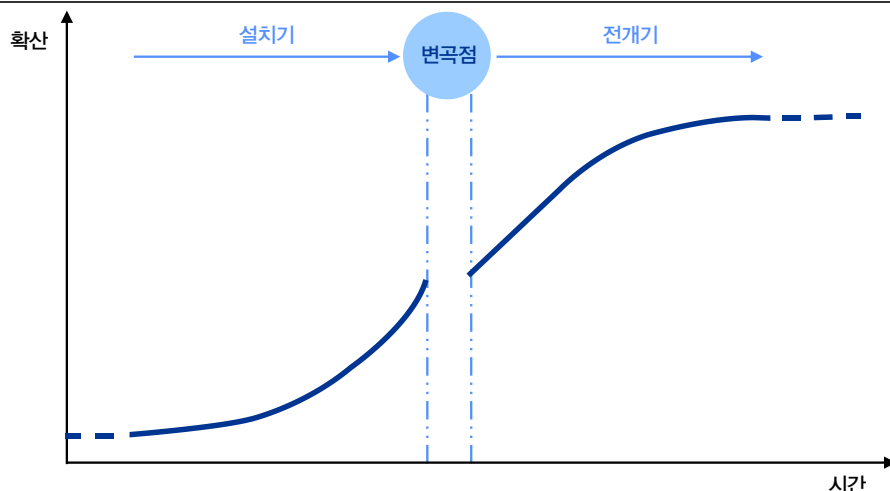
해당 논문의 핵심 골자는, 새로운 기술 패러다임(Technological Paradigm, 특정 시대를 지배하는 기술 시스템)이 등장하고 사회 전반에 확산되는 과정이 일정한 패턴을 따른다는 점이다. 이 패턴은 크게 두 시기로 구분된다. 첫 번째는 기술이 등장하고 급속히 확산되는 설치기(Installment Period)이며, 두 번째는 기술이 어느정도 설치되고 사회 및 경제 전반에 안착하는 전개기(Deployment Period)이다.

설치기는 초기 기술이 본격적으로 확산되기 시작하는 시기로, 대규모 인프라 구축과 함께 막대한 금융자본이 유입되는 특징을 가진다. 이 시기의 자본은 주로 금융자본(은행 등 투자자를 통해 유입된 자본)으로, 수익모델이 완전히 정립되기 이전임에도 기술에 대한 '기대'만으로 투자자들이 몰려든다. 이는 과잉투자로 연결되며, 일정 시점 이후에 그 거품은 꺼지게 된다.

버블이 꺼진 이후에는 금융자본이 빠져나가고, 대신 생산자본(기업이 생산을 위해 자체적으로 투입하는 자본)이 주도하는 전개기가 시작된다. 이 시점부터는 기술 기업이 실질적인 수익 모델을 만들어내고, 그 기업이 생존하며 산업을 재편한다. 기술은 사회 전반에 정착하며, 혁신적인 기술과 금융 사이에서 시너지를 창출한다.

페레즈는 이러한 구조를 통해 기존 슈페터 이론에서 상대적으로 경시되었던 금융가(Banker)의 역할을 재조명했다. 금융자본은 단순히 생산 자금을 지원하는 역할 뿐만 아니라, 기술 혁신 초기에 기술이 발전할 수 있도록 하는 결정적인 역할을 한다고 보았다. 특히 버블조차도 단순한 시장 실패가 아니라, 장기적으로 기술 생태계 형성을 위한 필연적 조정이라고 해석했다.

그림 14. 페레즈는 기술이 '설치기-전개기'의 사이클로 확산된다고 주장



자료: 『Finance and technical change A long-term view』, IBK투자증권

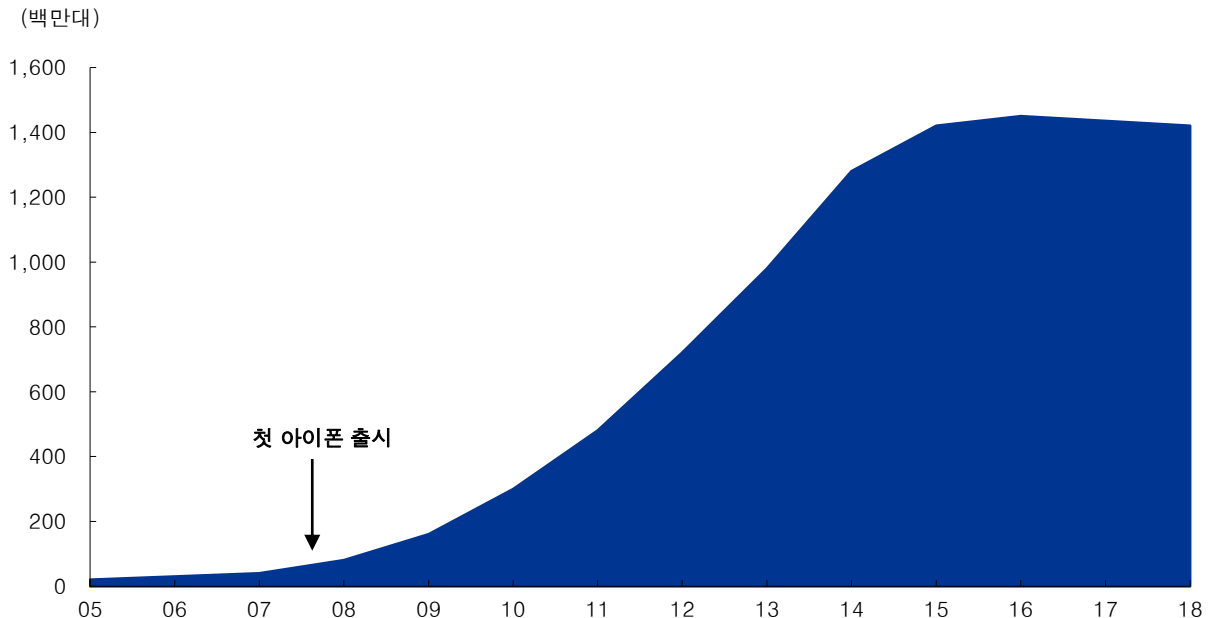
1) 설치기, 기술의 폭발적 확산 단계

설치기, 대규모 인프라 투자와
금융자본 집중의 시기

설치기의 특징은 두 가지로 설명된다. 첫째, 인프라 중심의 대규모 투자가 발생한다는 점이다. 이 시기의 자본은 주로 금융자본이며, 새로운 기술이 상용화되기 전임에도 불구하고 미래 수익에 대한 ‘기대’만으로 막대한 자금이 유입된다. 예를 들어, 19세기 중엽 철도 인프라, 20세기 초 전기·자동차 기반 시설, 1990년대 인터넷 네트워크 구축 등이 이에 해당한다. 둘째, 불균형의 시기라는 점이다. 기술은 급격히 발전하고 있지만, 사회는 아직 따라가지 못한다. 새로운 기술이 기존 기술과 충돌하면서, 산업별로 신기술을 받아들이는 속도에 불균형이 발생한다. 이는 과도한 낙관론과 회의론을 동시에 유발하며, 과잉투자와 버블로 이어진다.

페레즈는 설치기가 중요한 이유가 다음과 같다고 설명한다. 첫째, 기술적 토대가 형성되는 시기이다. 신기술을 중심으로 새로운 가치사슬이 형성되며, 이후 전개기에서 실제 수익이 창출되는 기반이 된다. 둘째, 금융자본이 집중되는 시기이다. 초기 자금은 주로 금융투자를 통해 조달되는데, 이는 생산자본이 신기술에 대한 확신을 가지기 전까지는 리스크를 감수할 수 없기 때문이다. 이 과정에서 투기가 발생할 수 있으나, 동시에 기술 확산 속도를 높이는 데 기여한다. 셋째, 버블이 붕괴되며 금융자본이 빠져나가는 시기이기도 하다. 설치기 후반에는 시장의 조정이 시작하며 다음 단계인 전개기로의 전환이 준비된다.

그림 15. 설치기, 스마트폰 출하량도 유사한 흐름을 보임



자료: JPR, IBK투자증권, 주: 스마트폰 출하량

2) 전개기, 기술의 안정적 정착 단계

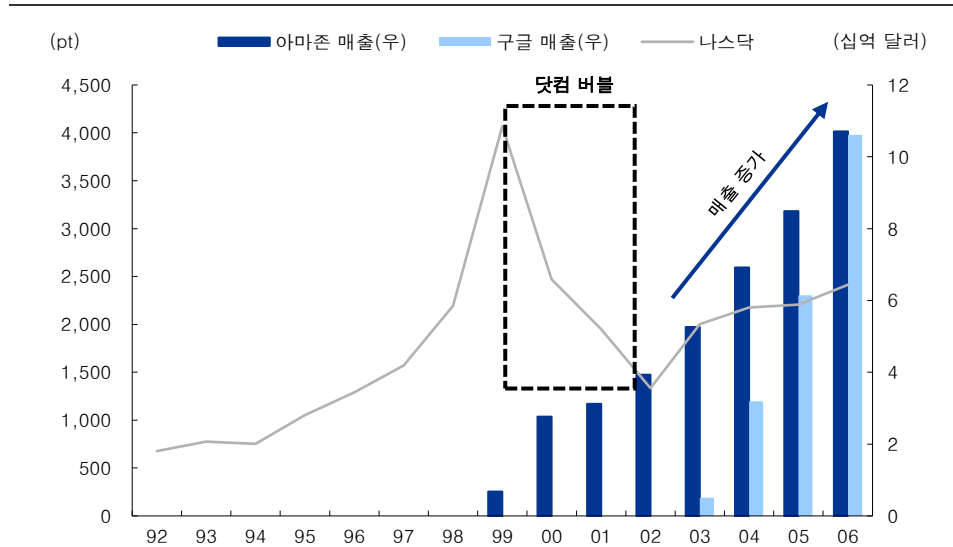
설치기, 수익 모델 정착과
생산자본 자생의 시기

전개기(Deployment Period)는 기술 확산 구조의 두 번째 단계로, 설치기에 확산된 신기술이 본격적으로 사회에 정착되는 시기라고 정의한다. 이 단계에서는 기술에 대한 막연한 ‘기대’가 실제 ‘수익’으로 전환되며, 새로운 생태계가 형성된다.

전개기의 특징은 세 가지로 설명된다. 첫째, 생산자본 중심 투자가 이루어진다. 기술적 불확실성이 줄어들고 수익모델이 가시화되면서, 금융자본의 비중은 줄고 기업의 투자가 본격화된다. 이는 과잉투자와 투기적 성격이 강했던 설치기와는 대비되는 안정적 성장 국면이다. 둘째, 사회적 조정이 이루어진다. 정부, 산업, 노동시장, 교육 시스템 등이 신기술에 맞춰 변화하며, 새로운 기술 패러다임에 적응한다. 셋째, 생산성이 향상된다. 기술이 도입되는 설치기에서는 GDP 성장이 본격화되지 않지만, 전개기에 들어서면 기술이 산업 전반에 자리잡아 생산성을 극적으로 개선한다.

페레즈는 전개기가 중요한 이유가 다음과 같다고 설명한다. 첫째, 기술이 상업화되는 시기이다. 초기에는 불확실했던 수익모델이 가시화되며, 기술은 성숙기에 접어든다. 소비자들은 신기술을 기반으로 한 제품 및 서비스를 일상적으로 이용하게 된다. 둘째, 사회 전체가 기술을 수용하는 시기이다. 노동시장·교육·법제도·인프라 등이 신기술에 맞게 정비되면서, 기술의 혜택이 특정 기업이나 산업을 넘어 전 사회적으로 분배된다. 셋째, 불균형 해소와 안정화가 이루어지는 시기이다. 설치기 후반에 버블이 붕괴된 이후, 시장은 재정비되며 지속가능한 성장 궤도에 진입한다. 정부의 정책적 개입도 효과를 발휘하며, 자본시장(투자자)과 실물경제(기업)의 연계가 강화된다.

그림 16. 닷컴버블 이후 아마존, 구글의 매출이 비약적으로 증가



자료: Bloomberg, IBK투자증권

2. 닷컴 이후 기술 확산 구조

1980~1990년대 초반: 개인용 컴퓨터(PC)의 등장

1980년대, PC 확산과
디지털 생태계 기반 형성

1980년대는 정보기술(IT) 산업의 태동기이자, 현대 디지털 경제의 기초가 마련된 시기로 평가된다. 특히 1981년 IBM의 개인용 컴퓨터(IBM PC) 출시를 기점으로, 기업 전산실에 국한되었던 컴퓨터 사용이 사무실과 가정으로 확대되며 정보 접근이 용이해졌다. 이후 마이크로소프트(Microsoft)가 운영체제(OS) 시장에서 지배력을 확보하면서, 본격적인 소프트웨어 생태계가 형성되기 시작했다.

당시의 컴퓨터는 텍스트 기반 인터페이스(DOS 환경)에서 GUI(그래픽 사용자 인터페이스)로의 전환이 이뤄지던 과도기였으며, 사용자는 워드프로세서, 스프레드시트, 데이터베이스 등 기본적인 정보처리 소프트웨어를 주로 사용했다. 1985년 출시된 Windows 1.0은 아직 미완성된 형태였으나, 이후 Windows 3.0(1990) 출시를 통해 보다 직관적인 컴퓨팅 환경이 제공되며 시장의 폭발적인 반응을 이끌어냈다.

하드웨어 관점에서는 인텔이 8086, 80286, 80386 등 프로세서 라인업을 통해 고성능 연산 시장에서 점유율을 확대했다. 하드디스크(HDD), 플로피디스크 등 저장장치 기술 또한 진화를 거듭했다. 컴퓨터 가격 또한 점차 하락 안정화되며, 가정에서도 구매가 가능한 수준까지 진입하게 된다. 당시 네트워크 인프라가 매우 한정적이었기 때문에, 대부분의 컴퓨터는 오프라인에서만 주로 활용되었다.

그림 17. 1981년 출시된 IBM의 PC



자료: Google, IBK투자증권

그림 18. 1982년 출시된 인텔의 80286(PC용 마이크로 프로세서)



자료: Google, IBK투자증권

1990년대 중반~2000년 초반: 인터넷의 대중화

1990년대, 인터넷 상용화와
네트워크 인프라 투자 급증

1990년대 중반은 인터넷이 상용화되고 정보 접근 방식이 획기적으로 변화된 시기이다. 오프라인 환경에 머물러 있던 정보 접근 방식은 ‘온라인’으로 전환됐다. 이는 인터넷 포털, 검색, 전자상거래, 커뮤니티 플랫폼이 출현하는 배경이 되었다.

가장 결정적인 변곡점은 1993년 ‘모자이크(Mosaic)’ 브라우저의 등장이었다. 이후 넷스케이프(Netscape Navigator), 마이크로소프트의 인터넷 익스플로러(IE) 등이 연달아 출시되며 웹을 기반으로 하는 정보 탐색이 가능해졌다. 웹 브라우저는 일반 대중이 HTML 기반 정보를 시각적으로 불러올 수 있게 했다. 인터넷이 정부기관이나 연구소의 전유물에서 대중적 인프라로 확장되기 시작했다.

미국에서는 통신 인프라도 확장됐다. 미 국방부의 ARPANET 실험에서 파생된 NSFNET이 민간으로 확산되며, ISP(인터넷 서비스 제공자) 산업이 빠르게 성장했다. 일반적인 다이얼업 접속(전화 회선을 이용하여 인터넷에 접속하는 방식)에서 케이블·ADSL 등 고속 인터넷 접속이 시작됐다. 한국은 1999년 하나로통신이 ADSL 상용화했다. 초고속인터넷 가입자가 폭발적으로 증가하며, OECD 국가 중 인터넷 보급률 1위를 기록하게 된다.

인터넷 플랫폼 서비스도 생겨났다. 야후(Yahoo)는 웹디렉토리 기반의 포털을 제시했고, 아마존(Amazon)은 온라인 서점에서 시작해 범용 전자상거래 플랫폼으로 확장했다. 이베이(eBay)는 C2C(Customer-to-Customer) 경매 플랫폼을 정립했고, AOL은 메신저 및 뉴스 서비스를 제공했다. 이 기업들은 대부분 트래픽을 선점하고 난 후에 광고 수익화 또는 중개 수수료화의 방식으로 수익 모델을 실험했고, 이는 이후 플랫폼 비즈니스 모델의 토대 자리 잡는다.

투자 관점에서는 과잉 투자가 일어났다. 1995~2000년 사이 나스닥(NASDAQ)에는 매년 수백 개의 닷컴 기업들이 상장되었고, 수익성보다는 잠재적 효과에 대한 ‘기대’만으로 기업가치가 과도하게 부풀려졌다. 대표적으로 Pets.com, Webvan, eToys 등의 사례는 수익모델이 부실한 상태에서도 고평가를 받았으며, 이는 2000년 닷컴 버블 붕괴의 직접적 원인으로 작용했다.

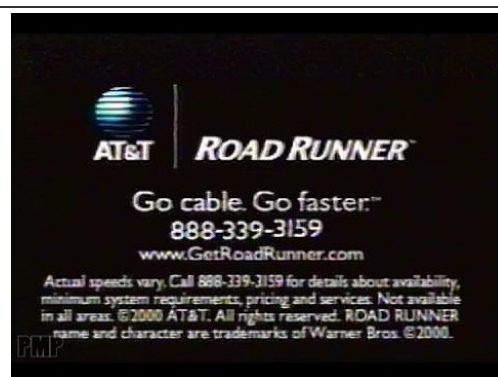
버블 붕괴 이후 시장은 실질적 경쟁력을 갖춘 기업을 중심으로 재편되었다. 아마존은 물류 인프라에 선제적으로 투자하며 생존에 성공했고, 구글은 검색 알고리즘(페이지랭크)과 키워드 기반 광고(AdWords)를 통해 수익구조를 확보했다.

그림 19. 1995년 출시된 Microsoft의 IE(인터넷 익스플로러)



자료: Google, IBK투자증권

그림 20. 2000년 출시된 AT&T의 Road Runner(초고속 인터넷)



자료: Google, IBK투자증권

2000년대 중반: 모바일(피쳐폰)의 도입

2000년대, 3G 피쳐폰 확산과
통신사 중심 생태계 형성

2000년대 중반은 글로벌 ICT 시장이 고정형(PC·유선전화) 중심에서 이동형(3G 등 모바일) 기반으로 전환되는 과도기였으며, 하드웨어·통신·콘텐츠 생태계 전반에서 구조적 변화가 동시다발적으로 전개되었다. 특히 이 시기는 스마트폰으로 전환되기 직전까지, 피쳐폰(Feature Phone)을 중심으로 한 음성·문자 기반의 마지막 모바일 단말기 시대로 정의된다.

기술적으로는 2.5G(GPRS, cdma2000 1x)에서 3G(WCDMA, EV-DO)로의 진화가 이루어졌다. 이를 기반으로 텍스트 중심의 통신(문자, 이메일, 웹 브라우징)에서 이미지·멀티미디어 기반 콘텐츠 소비가 확대되기 시작했다. 그러나 당시 네트워크 속도는 여전히 제한적이었고, 고속 데이터 전송을 통한 스트리밍·앱 실행 등은 본격화되지 못한 상태였다.

단말기 제조사 간 경쟁은 하드웨어 중심의 성능 경쟁으로 전개됐다. 노키아, 모토로라, 삼성전자, LG전자 등 글로벌 제조사들은 폴더폰, 슬라이드폰, 초소형폰 등 다양한 폼팩터(제품의 외형)를 개발 및 출시했다. 카메라 화소·디스플레이 크기 개선 등을 통해 시장점유율을 확대해 나갔다. 특히 노키아는 심비안(Symbian) 기반 OS를 활용한 피쳐폰 시장을 선도했다. 한국 기업으로는 삼성전자가 ‘애니콜’ 브랜드를 통해 국내 시장을 사실상 독점하다시피 했다.

통신사 수익모델도 본격화되었다. 이동통신 3사는 단말기 유통, 요금제 설계, 콘텐츠 플랫폼 운영을 모두 수직계열화하며 폐쇄형 모바일 생태계를 형성했다. 대표적으로 국내에서는 SK텔레콤의 ‘NATE’, KTF의 ‘매직엔’, LG텔레콤의 ‘ez-i’가 생겼으며, 이를 통해 벨소리·배경화면·모바일 게임·화보 등 유료 콘텐츠가 유통됐다. 통신사는 이 구조를 통해 CP(Content Provider)로부터 수익을 분배 받는 구조를 설계했다. 유선 인터넷 환경에서의 ‘광고 기반 플랫폼’과는 다른 형태의 폐쇄적 과금 구조를 유지한 것이다.

그림 21. 2000년 출시된 노키아의 9110i 커뮤니케이터



자료: Google, IBK투자증권

그림 22. 2004년 출시된 노키아의 9300



자료: Google, IBK투자증권

2007년~2010년 후반: 스마트폰으로의 전환

2000년대 중반,
스마트폰 등장과
앱 생태계 구축

2007년 애플의 첫 아이폰 출시를 기점으로, 글로벌 ICT 산업은 단말기 경쟁에서 플랫폼 경쟁 체제로 본격 전환되었다. 기존 피쳐폰이 하드웨어 스펙 중심으로 진화했던 것과는 달리, 스마트폰은 UX·콘텐츠·앱개발 측면에서 패러다임을 근본적으로 바꾸는 사건이었다. 앱 생태계 참여 장벽을 낮춤으로써 전 세계 개발자들이 직접 서비스 유통에 참여할 수 있는 구조를 마련했다. 2008년 앱스토어(App Store) 출시 이후 수백만 개의 앱이 등록되었고, B2C(Business-to-Business)·C2C(Consumer-to-consumer) 중심의 앱 기반 플랫폼 비즈니스 모델이 본격화되었다.

구글은 2008년 안드로이드(Android) 운영체제를 공개하며, 제조사 및 통신사와의 협업을 통해 iOS에 대응하는 오픈소스 기반 생태계를 조성했다. 삼성전자, LG전자, HTC, 화웨이 등 글로벌 제조사들이 안드로이드 OS를 채택하며, 향후 글로벌 스마트폰 시장에 iOS와 Android의 양자 구도가 고착되었다.

스마트폰 보급률이 가파르게 상승하며 사용자의 정보 탐색, 커뮤니케이션, 콘텐츠 이용 방식이 모바일로 전환되었다. 포털 이용은 점차 감소하고, 모바일 앱을 통한 콘텐츠 소비(유튜브, 인스타그램, 카카오톡 등)가 일상화되었다. 푸시 알림, 위치 기반 서비스(GPS 등), 카메라 결합 등 스마트폰이 가져온 신기능은 새로운 UX 설계를 가능하게 했다.

비즈니스 모델도 개발되었다. 대표적으로는, 플랫폼형(우버, 에어비앤비, 쿠팡 등), 광고형(페이스북, 인스타그램, 틱톡 등), 구독형(넷플릭스, 스포티파이 등), B2B형(슬랙, 줌, 모바일 ERP 등)이 있다.

한편, 통신사의 영향력은 급격히 약화되었다. 기존 피쳐폰 시기에는 통신사가 요금제, 단말기 유통, 콘텐츠 플랫폼을 수직계열화했으나, 스마트폰 시기에는 플랫폼 사업자(App 개발사, OS 공급자)와 제조사가 수익의 대부분을 차지하는 구조로 재편되었다. 통신사는 단순 인프라 공급자로 전락하였다.

그림 23. 2007년 출시된 애플의 iPhone 1세대



자료: Google, IBK투자증권

그림 24. 2008년 출시된 애플의 앱스토어



자료: Google, IBK투자증권

2010년대: 클라우드 컴퓨팅의 확산(산업의 기반)

2010년대, 클라우드 확산과
SaaS 모델 등장

2010년대는 정보를 저장하고 처리하는 방법이 온프레미스(On-Premise: 자체적으로 IT 인프라를 보유)에서 클라우드(Cloud: 외부 업체로부터 IT 인프라를 빌려서 사용)로 전환되었던 시기다. ICT 산업의 비용 구조, 운영 방식, 기술 진입장벽이 변화하기 시작했고, 이 시기에는 데이터의 수집·처리·분석 역량이 기업의 경쟁력으로 부상했다.

클라우드 인프라가 급속히 확산되었다. 2006년 Amazon Web Services(AWS)의 EC2·S3 서비스가 출시된 이후, 2010년대 중반부터는 마이크로소프트(MS Azure), 구글(GCP) 등이 본격적으로 가세했다. 퍼블릭 클라우드(인터넷을 통해 컴퓨팅 리소스를 공유해서 사용) 기술은 초기에 스타트업에서 주로 사용했으나, 이후 대기업에서 공공기관까지 클라우드 이전(Cloud Migration: 기업이나 조직의 데이터, IT 인프라 등을 온프레미스 환경에서 클라우드 환경으로 옮기는 과정)을 단행하며 전 산업으로 확대되었다.

SaaS(Software-as-a-Service, 서비스형 소프트웨어) 모델도 등장했다. Dropbox, Google Workspace, Zoom, Salesforce 등 클라우드 기반 협업툴·CRM(고객관리)·비즈니스 솔루션은 생산성 혁신을 견인했다. 특히 코로나 팬데믹 이후 원격근무와 비대면 업무가 시행되면서 SaaS를 도입하는 기업이 많아졌다.

클라우드 기반 구조는 빅데이터, AI(인공지능), ML(머신러닝)의 발전에도 기여했다. Hadoop(여러 개의 컴퓨터를 하나로 묶어 대용량 데이터를 처리하는 기술), Spark(하둡과 같은 대용량 분산처리 기술이나 더 빠른 속도를 제공) 등 대용량 분산처리 기술과 AWS Sagemaker, Google Vertex AI 등 AI 개발 플랫폼이 등장했다.

그림 25. 2006년 출시된 아마존의 AWS



자료: Google, IBK투자증권

그림 26. 2020년 출시된 구글의 Workspace



자료: Google, IBK투자증권

사이클마다 달라지는 대장주

기술 진화 따라
시장 주도주도 순환됨

기술 사이클이 변화할 때마다 주식시장의 주도 종목도 교체되었다. 실제로 나스닥 시가총액 상위 기업 변화를 살펴보면, 기술 발전에 따라 대장주 역시 순차적으로 변화한 흐름이 포착된다.

개인용 컴퓨터(PC) 보급 확산과 함께 인텔, 마이크로소프트 등 하드웨어 및 운영체제 기업이 시총 상위를 차지했다.

인터넷 인프라 투자 사이클이 도래하며 시스코를 비롯한 네트워크 장비 업체들이 시장 주도주로 부각되었다. 이후 검색·이커머스·콘텐츠 플랫폼 등 Web 2.0이 급부상하며 구글과 아마존 시총이 급등했다.

피쳐폰과 통신 인프라가 확산되는 시기에는 노키아가, 이후 스마트폰이 확산되는 시기에는 애플이 모바일 대장주로 부각되었다. 스마트폰 생태계가 정착되며 유튜브(구글), 페이스북(메타), 넷플릭스, 아마존 등 소프트웨어 기업이 시장 지배력을 확대했다.

클라우드 컴퓨팅 시장이 확대되면서 MS, 아마존 등이 시총 상위를 차지했다.

최근에는 생성형AI 확산을 기점으로 엔비디아(NVIDIA)가 단기간 내 급등하는 현상이 나타났다.

즉, 기술 사이클이 바뀔 때마다 시장의 주도주는 이동했다. 투자자 관점에서 기술 변화에 선제적으로 대응해야 투자 타이밍을 포착할 수 있을 것이며, 장기적으로도 포트폴리오 리밸런싱의 방향성을 결정짓는 주요 변수로 작용할 것이다.

그림 27. 기술 발전의 변곡점에 따라 나스닥 내 시가총액 상위 기업이 변화함



자료: Bloomberg, IBK투자증권, 주: 10위권 내 종목만 표시

캐즘(Chasm): 수요 정체 구간

새로운 기술 패러다임 등장 전
정체기, 캐즘 구간

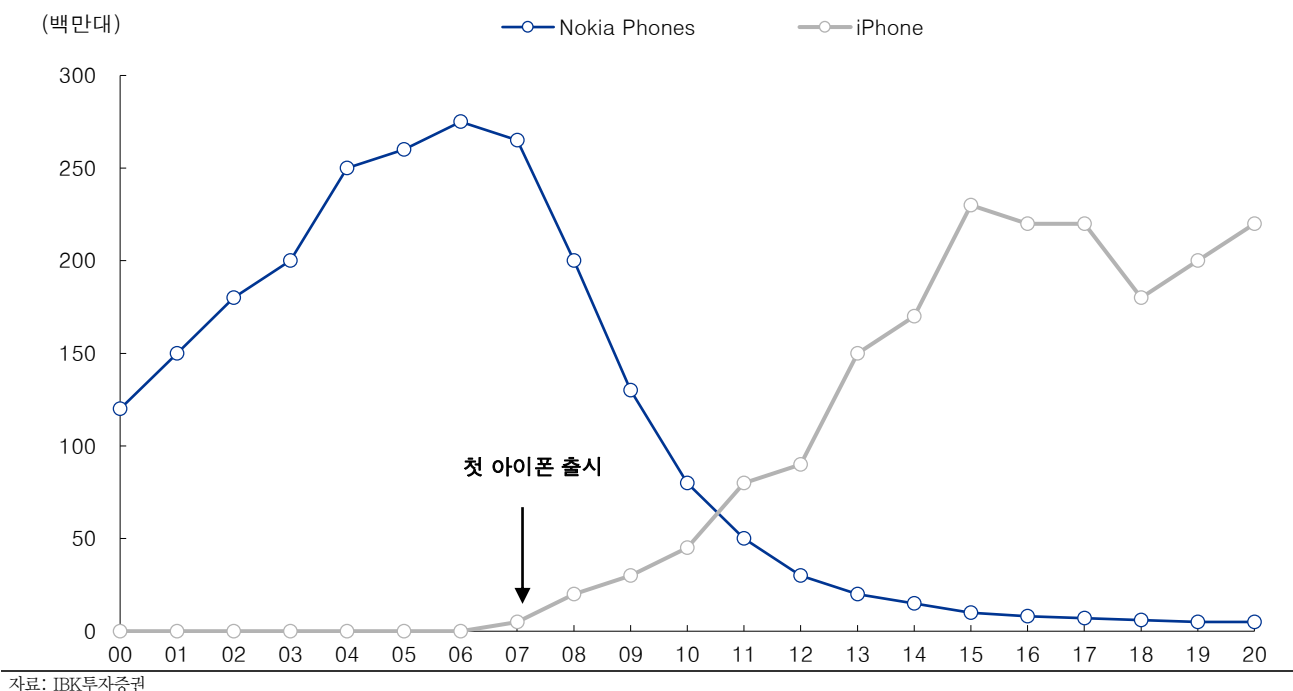
기술의 확산은 직선적이지 않다. 초기 보급이 빠르게 이뤄졌더라도, 사용자 경험이 제한적이거나 실질적 효용이 미비할 경우 수요는 정체된다. 이처럼 기술은 설치되었지만 활용되지 않는 비선형적 구간이 바로 **캐즘**이다.

캐즘은 모바일 사이클에서 가장 뚜렷하게 관찰됐다. 피쳐폰의 대중화 이후 스마트폰으로 넘어가기 직전 특히 2006~2008년 사이, 노키아폰 판매가 정점을 찍은 이후 급격히 하락하고, 아이폰 판매가 본격적으로 증가하기 전까지의 공백기는 대표적인 캐즘 구간으로 해석된다.

2000년대 초반부터 노키아·모토로라 등 피쳐폰 제조사들이 시장을 주도하며 단말기 보급률은 빠르게 증가했다. 2006년까지 미국 내 노키아폰 판매량은 연간 260만 대 이상으로 정점을 찍었다. 하지만 당시 피쳐폰은 문자, 벨소리, 간단한 게임 정도의 기능에 그쳤고, 인터넷 접속은 제한적이며 요금 부담도 컸다. 소비자는 휴대폰을 ‘가지고는 있지만’ 더 이상 체감할 만한 새로운 기능이 없었다.

이러한 상황은 사용자 효용 정체로 이어졌고, 그 결과 판매량은 2007년 이후 급격히 감소한다. 반면, 2007년 첫 아이폰이 출시되고 새로운 패러다임을 열었고 2010년에는 노키아의 시장 점유율을 대체하는 수준 이상으로 성장했다. 즉, 기술 전환의 가능성이 보였지만, 대중들이 반응하기 전까지의 몇 년 간은 캐즘 구간이었다.

그림 28. 모바일 초기에는 노키아 폰 판매량이 높았지만, 추후 아이폰 판매량이 추월함



IBKS가 제시하는 테크 사이클: 1차 설치기-캐즘-2차 설치기-전개기

IBKS 기술 확산 4단계 제시:
설치기-캐즘-2차 설치기
-전개기 구조

IBK투자증권은 페레즈의 ‘설치기-전개기’ 구조를 한층 세분화하여 ‘1차 설치기-캐즘-2차 설치기-전개기’라는 4단계 사이클을 제시한다.

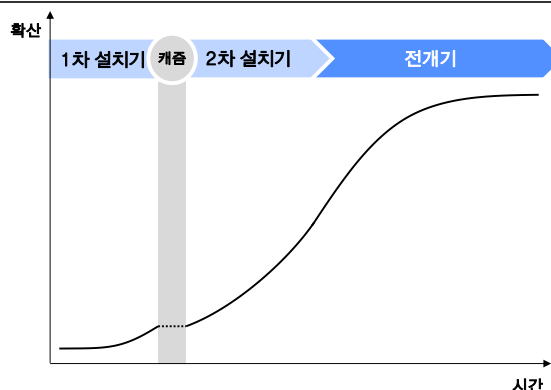
1차 설치기는 기존 기술의 연장선에서 이루어지는 초기 확산 단계다. 이 시기에는 이미 존재하던 기술의 개선이나 조합을 통해 기술이 빠르게 설치된다. 하지만 이 시점에서 기술은 기존 방식의 효율을 높이는 수준에 그치며, 소비자에게 체감 가능한 새로운 가치를 제공하는 데에는 한계가 있다. 기존 시장의 연장선에서 실적이 형성되기 때문에 시장은 투자 대비 수익에 대한 의문을 품기 시작한다.

이런 초기 확산을 지나면 **캐즘**이 발생한다. 캐즘은 기술이 얼리어답터를 넘어 대중으로 전환되는 데 실패하면서 수요가 정체되는 구간이다. 소비자 입장에서는 기술이 존재하지만 이를 사용해야 할 ‘이유’가 부족하고, 기업 입장에서는 수익 모델이 명확하지 않아 투자가 위축된다. 기술은 설치되어 있지만 활용되지 않는 이 구간은, 자본이 회의적으로 변하고 시장의 기대감이 일시적으로 무너지는 전환점이 된다.

이후 기술의 본질적 전환이 이루어지면서 **2차 설치기**가 시작된다. 이 구간에는 기존 기술의 단순 확장이 아닌, 완전히 새로운 기술 패러다임이 등장한다. 기술의 성격 자체가 변화하면서, 사용자 경험과 시장 구조에 근본적인 변화가 발생하고, 자본 역시 다시 한 번 빠르게 유입되기 시작한다. 이때 나타나는 기술은 기존 기술과의 연속성보다 단절성에 가깝기 때문에, 선점에 성공한 기업은 산업 지형을 빠르게 장악할 수 있다. 디바이스, 플랫폼, 서비스 전반에 걸친 생태계 구축이 본격화되며, 기존 주도 기업이 밀려나고 새로운 리더가 등장하는 교체 현상이 나타난다.

마지막으로 **전개기**는 기술이 광범위하게 확산되는 단계다. 이 시기에는 사용자가 기술을 일상적으로 받아들이고, 서비스가 폭발적으로 등장하며 새로운 시장에서 실질적인 수익화가 본격화된다. 초기에는 금융자본이 주도했던 시장이, 점차 생산자본 중심으로 옮겨가며 자립적인 생태계가 형성된다. 기술은 더 이상 낯설거나 실험적인 것이 아니라 일상에 필수적인 요소가 되며 산업과 생활 전반을 재편하는 수준으로 확산된다.

그림 29. IBK투자증권은 ‘1차 설치기-캐즘-2차 설치기-전개기’ 사이클을 제시



자료: IBK투자증권

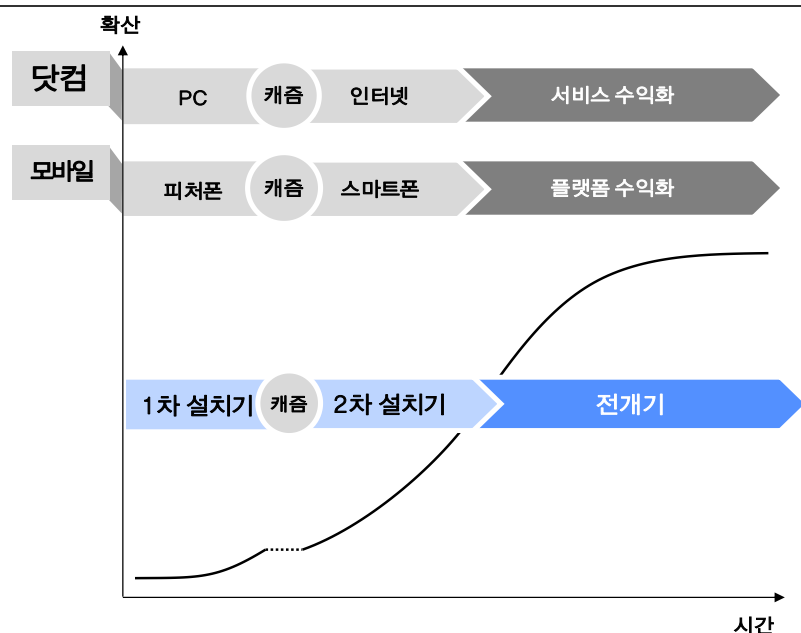
IBKS 사이클로 다시 본 닷컴·모바일 시대

닷컴 · 모바일 시대,
4단계 확산 경로로 재구성

이 프레임으로 분석해보면, 닷컴 사이클의 **1차 설치기**는 가정용 PC가 빠르게 보급되던 시기로 볼 수 있다. 이 시기에는 컴퓨터가 사회 전반에 되었다. 이후 **인터넷이 대중적으로 확산되기 전까지 수요가 정체되는 캐즘** 구간이 이어졌다. 기술은 깔렸지만 사람들이 실제로 활용할만한 ‘이유’가 부족했던 시기다. 이후 **2차 설치기**는 **인터넷 인프라가 본격적으로 확산된 시기였다**. 시스코(Cisco)를 중심으로 한 네트워크 장비 기업이 주목받았다. 그러나 여전히 인터넷은 속도나 연결 안정성 측면에서 한계가 있었고, 사용자 효용 역시 제한적이었다. 결정적인 전환점은 초고속 인터넷의 등장이다. 이 시점에서 언제 어디서든 빠르고 안정적인 웹 접근이 가능해지면서 인터넷은 필수 기술로 자리 잡았다. 이로 인해 **구글·아마존 같은 플랫폼 기업이 수익 모델을 갖추어 서비스 생태계를 구축하는 전개기로 진입했다**.

모바일 사이클 역시 동일하게 4단계로 해석된다. **1차 설치기**는 피쳐폰이 보편화되던 시기로, 통신 기기의 보급률은 높았지만 제한된 기능 탓에 기술의 활용도는 낮았다. **캐즘**은 바로 그 기능적 제약으로 인해 수요가 정체된 구간이었다. 그러다 **아이폰을 필두로 한 스마트폰의 등장은 기술적 패러다임을 전환시킨 2차 설치기에 해당하며**, 모바일 생태계의 가능성을 열어주었다. 하지만 진정한 **전개기**는 무제한 데이터 요금제가 등장한 이후였다. 이는 사용자의 데이터 소비 장벽을 제거하며 유튜브·페이스북 등 모바일 중심 플랫폼의 수익화와 사용자 확산을 견인했다. 결국 모바일도 닷컴과 마찬가지로 ‘기존 기술의 설치-정체-패러다임 전환-서비스 대중화’라는 일관된 경로를 따라 확산되었음을 확인할 수 있다.

그림 30. 4단계는 각각 ‘기존 기술의 설치-정체-패러다임 전환-서비스 대중화’를 의미



자료: IBK투자증권

3. AI의 현재 위치

1) 2022년 11월 이후: 생성형 AI의 급성장

ChatGPT 이후
생성형 AI 대중화 시작

2022년 11월, OpenAI는 GPT-3.5 기반 ChatGPT를 전 세계에 공개하며 생성형 AI 시대의 문을 열었다. 텍스트 생성 능력뿐 아니라 대화형 인터페이스가 일반 사용자에게 친숙하게 다가오면서, 출시 두 달 만에 1억 명 이상이 이용하는 이례적 성장을 기록했다. 이 시기를 기점으로 생성형 AI는 연구실을 벗어나 대중화되기 시작했다.

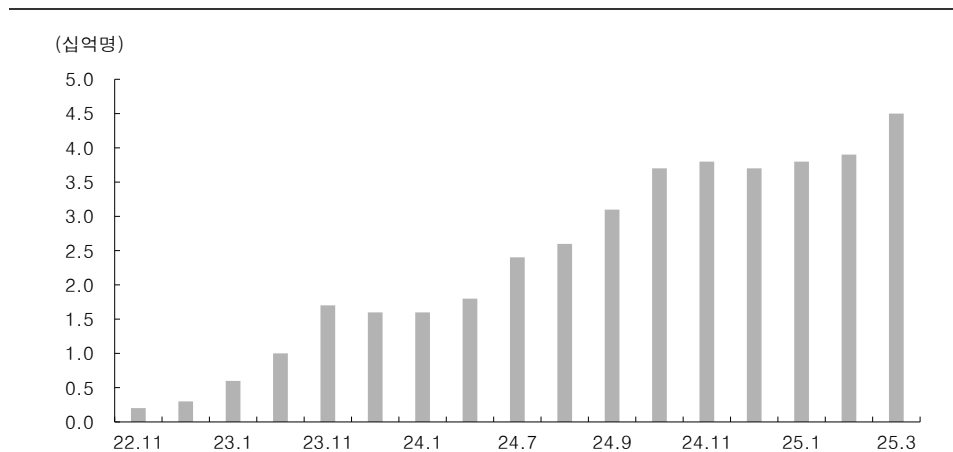
2023년 3월에는 GPT-4가 등장하며 정확성, 추론 능력, 코드 작성 등에서 비약적인 성능 향상을 보였다. 특히 멀티모달 입력(텍스트+이미지) 기능이 도입되며, AI의 활용 범위가 확장되었다. 같은 해 Google은 Bard(현재 Gemini), Anthropic은 Claude, Meta는 LLaMA 시리즈를 공개하며 경쟁이 본격화되었다. 이 시기 오픈소스 모델이 급증하고, Hugging Face를 중심으로 학습·배포 생태계가 활성화되었다.

표 9. ChatGPT 출시 당시 특징

구분	주요 내용
출시일	2022년 11월 30일
기본 모델	GPT-3.5 (InstructGPT 계열, RLHF 적용)
핵심 기술	RLHF (Human Feedback을 통한 강화학습)
주요 특징	- 대화에 최적화된 자연스러운 응답 - 대화 맥락 유지 가능 - 복잡한 프롬프트 없이도 높은 응답 품질 제공
지식 한계	2021년 중반까지 학습된 데이터 기준
사용 편의성	자연어로 쉽게 질문 가능, 기술적 배경 불필요
안전성	민감한 주제에 대한 응답 제한 및 필터링 적용
서비스 모델	초기 무료 베타 형태로 제공 (가입만 하면 사용 가능)
반응	출시 5일 만에 100만 사용자 돌파, 폭발적 성장

자료: 언론 보도 종합, IBK투자증권

그림 31. ChatGPT 사용자 추이



자료: Demandsage, IBK투자증권

2) 현재의 AI, 아직 1차 설치기에서 캐즘으로 진입하는 것일수도

생성형 AI, 인프라 중심
1차 설치기 국면

현재 생성형 AI 산업은 '1차 설치기'에 머물러 있다고 볼 수 있다. 그 근거는 다음 세 가지 측면에서 명확하게 드러난다.

첫째, 자본 투자가 인프라와 기술 개발에 집중되어 있다. AI 열풍 이후 투자 자금은 대부분 GPU, 데이터센터, AI 서버 등 하드웨어 인프라를 고도화하는데 사용 중이다. 또한 AI 인재 확보 경쟁이 치열해지면서 수백, 수천억원의 보상이 제시되고 있다. Microsoft, Amazon, Google, Meta 등 빅테크 기업들 역시 2023~2024년 사이 AI CapEx를 50% 이상 늘렸지만, 이 투자 대부분은 인프라 확충에 사용되었다. 이러한 인프라 투자로 인해 엔비디아 시가총액이 급등했으며 과거 닷컴 초기 인텔 시가총액이 급증했던 것과 유사하다.

그림 32. 닷컴시대 1차 설치기에는 인텔의 시가총액이 급증

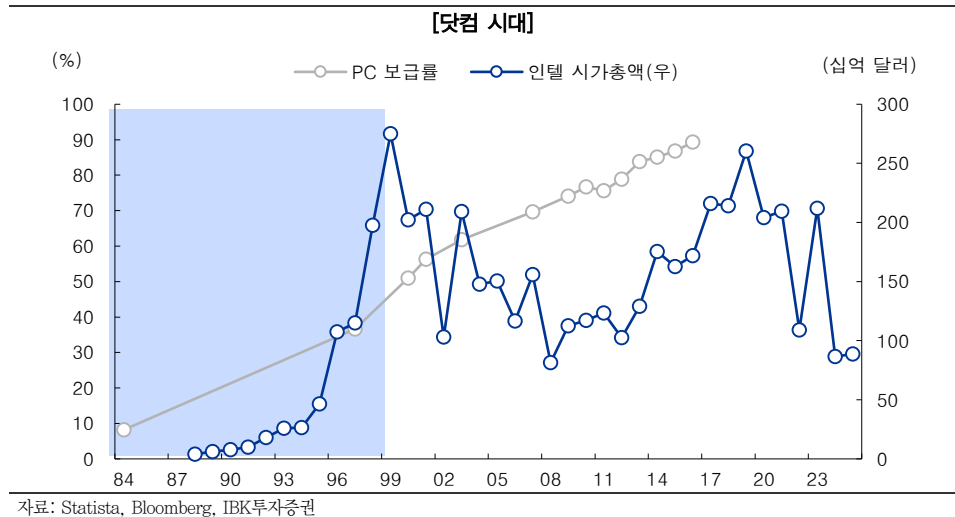
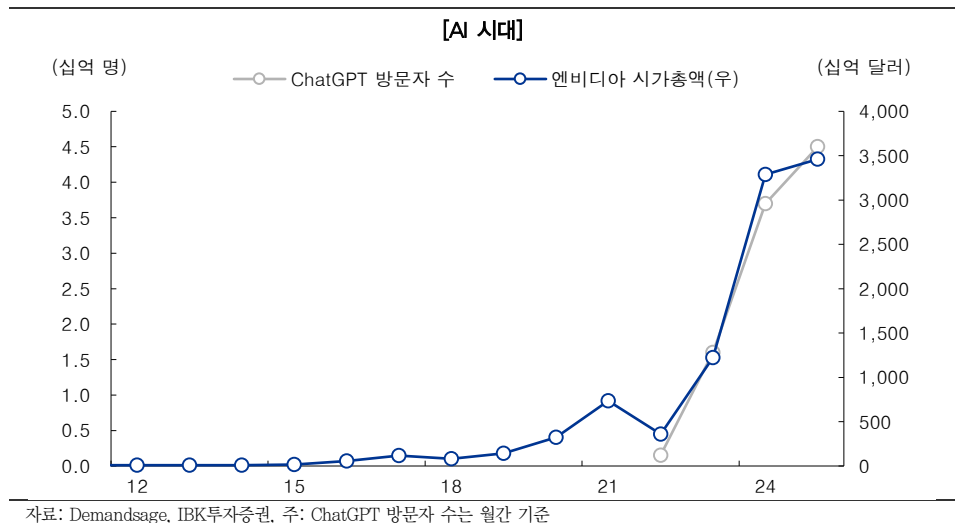


그림 33. 현재의 AI는 1차 설치기 단계와 유사



AI 활용, 기존 서비스 확장에
그치는 상황

둘째, AI의 사용이 기존 인터넷, 모바일 서비스의 연장선에 머물러있다. ChatGPT 등 챗봇 형태의 AI 서비스 이용자가 급증했다. 지브리 스타일의 이미지 생성이 유행이 되면서 2025년 4월 주간 평균 이용자수가 1.5억명을 상회했으나 한달 내 이용자가 빠르게 이탈된 것으로 나타났다. ChatGPT, Perplexity 등의 서비스는 기존 검색 서비스가 고도화된 측면이 강하다. 최근 산업 자동화, 업무 효율화에 AI가 적용되고 있으나 일반인들이 체감할 수 있는 AI서비스는 아직 부재한 상태이다.

생성형 AI,
뚜렷한 수익 모델 부재

셋째, 수익 모델이 부재하다. 대부분의 LLM 서비스는 여전히 무료 혹은 저가 요금제 기반이며, 광고·구독 등 전통적 수익 모델이 아직 제대로 작동하지 않고 있다. 기업 도입도 일부 Copilot류의 툴에 국한되며, 확장성이나 실효성 측면에서 아직 회의적인 시선이 존재한다. 이는 과거 유튜브·페이스북이 트래픽만으로 가치를 인정받던 시기와 유사한 양상이다.

그림 34. ChatGPT의 지브리 스타일 이미지 생성



자료: 언론 보도 종합, IBK투자증권

그림 35. 빅테크는 AI 인프라 확대를 위해 CapEx를 평균 +50% 이상 늘렸으나, ROI는 전년보다 감소

		CapEx	→	Free Cash Flow Margin	vs.	Revenue
Microsoft	C2023	\$35B		30%		\$228B
	C2024	\$56B		27%		\$262B
	Y/Y Change	+58%		-10%		+15%
Amazon	C2023	\$53B		6%		\$575B
	C2024	\$83B		5%		\$638B
	Y/Y Change	+57%		-8%		+11%
Alphabet (Google)	C2023	\$32B		23%		\$307B
	C2024	\$52B		21%		\$350B
	Y/Y Change	+63%		-8%		+14%
Meta Platforms (Facebook)	C2023	\$27B		33%		\$135B
	C2024	\$37B		33%		\$165B
	Y/Y Change	+38%		<1%		+22%

자료: BOND, IBK투자증권

3) '캐즘'에 돌입한 국면

현재 AI 산업 캐즘 국면 진입:

- 1) 낮은 활용도
- 2) 유료화 전환 한계 봉착
- 3) 사용자 피로도 증가

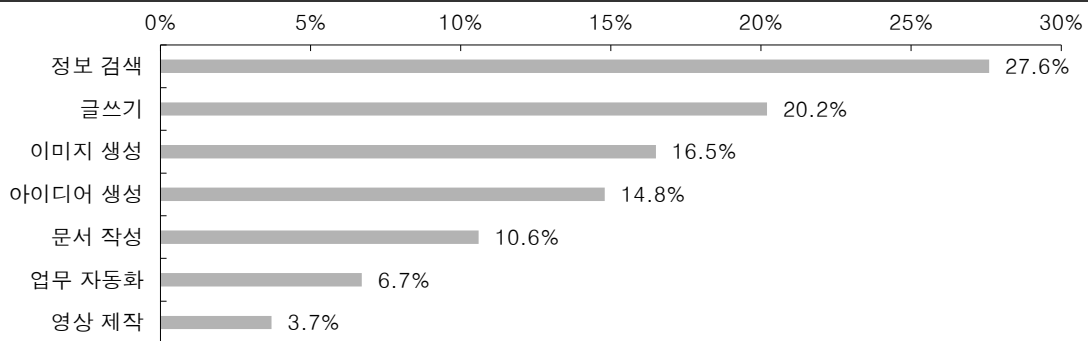
2024년 하반기부터는 기술 확산의 속도에 제동이 걸리고 있다. 이는 전형적인 캐즘 현상으로 해석할 수 있다. 지금 AI 산업은 다음과 같은 이유로 캐즘 구간에 접어들고 있다고 판단된다.

첫째, 초기 기대에 비해 실제 활용도가 낮다. 많은 기업이 생성형 AI를 PoC(기술이 실제로 가능한지 시험해보는 단계) 수준에서 도입해 봤지만, 실제 업무 효율성과 ROI가 기대에 못 미치는 경우가 많다. 기술은 신기하지만, 실질적인 도입 '이유'가 없는 것이다.

둘째, 유료 수익 모델 전환의 어려움이 명확해지고 있다. 대부분의 사용자는 무료로 AI를 체험했다. AI 기업이 유료 전환이나 장기적 구독 유도를 시도하고 있지만, 그 전환율은 낮다. 기업 고객 역시 LLM API의 반복 호출 비용 부담이 커서 본격적인 도입을 망설이고 있다.

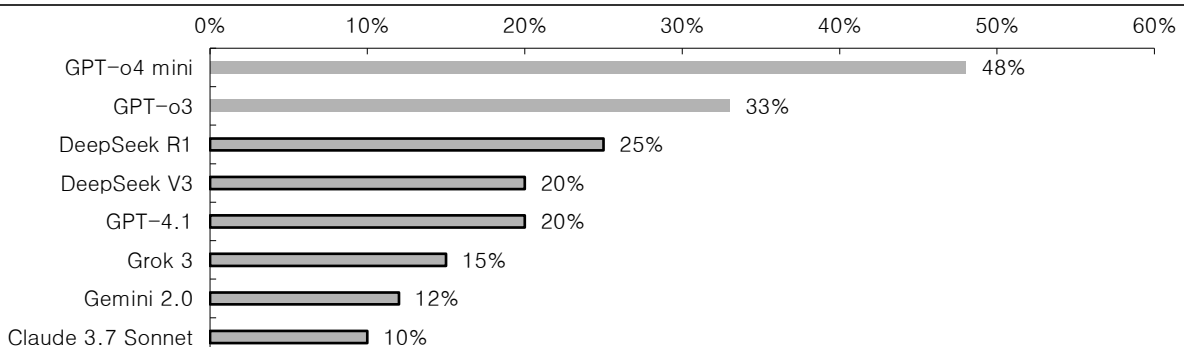
셋째, 사용자 피로도가 증가하며 캐즘을 심화시키고 있다. 초창기에는 호기심으로 몰렸던 사용자도, 반복된 Hallucination(AI 모델이 그럴듯하지만 사실이 아닌 내용을 생성하는 현상) 이슈나 개인정보 우려, 저작권 분쟁 등으로 실질적인 사용은 제한되고 있다.

그림 36. 생성형 AI를 가장 많이 활용하는 분야(AI 사용은 기존 기능에 국한되는 상태)



자료: 합사우트 글로벌, IBK투자증권

그림 37. AI 모델 별 Personal QA 할루시네이션 확률은 높은 편



자료: 언론 보도 종합, IBK투자증권, 주1: PersonQA는 사람과 관련된 사실 기반 질문을 중점적으로 평가하는 벤치마크, 주2: 검은 테두리는 추정치

4. 2025년 AI 트렌드

1) AI 산업의 변곡점

DeepSeek Impact: 투자자들의 인식의 변화 초래

딥시크, 저비용 고효율로
AI 판도를 흔들다

딥시크(DeepSeek)는 인공지능(AI) 산업에 저비용·고효율 모델을 제시하며 혁명적인 변화를 가져온 중국 AI 스타트업이다. 2025년 1월 20일 DeepSeek R1 모델을 공개했다. 딥시크는 기존 AI 모델들이 막대한 연산 자원과 고성능 하드웨어, 특히 고가의 GPU에 크게 의존했던 것과 달리, 소프트웨어 최적화와 독창적인 학습 및 추론 기법을 통해 하드웨어 의존도를 크게 줄이고 성능을 향상시키면서 AI 산업에 큰 충격을 가했다. 딥시크의 등장은 AI 산업 전반에 걸쳐 여러 혁명적인 변화를 촉발했다.

첫째, AI 모델 성능이 막대한 컴퓨팅 자원과 고비용 투자에 비례한다는 '규모의 법칙'에 균열을 냈다. 약 550만 달러의 저비용으로 메타의 Llama 4를 능가하고 OpenAI의 o1 모델과 유사한 성능을 달성했다. 이는 중소기업이나 스타트업도 고성능 AI를 개발할 수 있는 현실적인 가능성을 열어주었다. 결국 투자자들 역시 그동안 AI 업계의 과도한 투자에 의구심을 가지게 되었고 이후 AI 스타트업의 투자 유치가 어려워진다.

둘째, 오픈소스 모델을 공개함으로써 AI 기술의 민주화를 가속화했다. 이는 전 세계 개발자들이 딥시크의 기술을 자유롭게 활용하고 개선할 수 있는 생태계를 조성하여 AI 기술의 발전을 촉진하고, AI 시장의 확대 가능성을 제시했다.

셋째, AI 개발의 핵심 과제가 방대한 데이터 축적에서 고품질 데이터 선별, 정리, 정제로 전환되는 계기가 되었다. 딥시크는 정교하게 설계된 데이터 선택 파이프라인을 통해 최적화된 데이터 세트를 구축하며 효율성과 비용 절감을 동시에 달성했다. 이는 AI가 단순히 더 큰 모델을 만드는 것을 넘어 데이터를 관리하고 처리하는 능력에 달려 있음을 보여주었다.

그림 38. 딥시크 정보

딥시크(DeepSeek)	
창업자	 량원평 1985년생 중국 광둥성 잔장시 출생 저장대학 전자공학 학사 및 정보통신공학 석사
개발 과정	2015년/ 대학 친구들과 헤지펀드 한팡펀트 설립 2023년 7월/ 중국 항저우에 딥시크 설립 2023년 11월/ 첫 번째 오픈소스 AI 모델 딥시크 코더 출시 2023년 11월/ 딥시크 LLM 출시 2024년 5월/ 딥시크 V2 출시 2024년 12월/ LLM 딥시크 V3 출시 2025년 1월/ 추론 모델 딥시크 R1 출시
특징	미국 수출 규제를 피하기 위한 엔비디아의 저성능 저가형 H800칩 사용 오픈AI 등 대형 IT 기업에 비해 상대적으로 저렴한 개발 비용 180여명 규모의 적은 인력으로 개발

자료: 머니투데이, IBK투자증권

표 10. OpenAI o1과 DeepSeek R1 비교

구분	Open AI o1 (미국)	DeepSeek R1 (중국)
설립	2015년	2023년
AI 개발 비용	1억 달러 (추정)	557만 6,000달러 (이전 모델 V3 기준)
연구 개발 인력 (추정)	1,200명	139명
모델 공개	폐쇄형	오픈소스(MIT)
API 비용 (1M 토큰 출력 기준)	\$ 60	\$ 2.19
투입된 반도체	· 엔비디아 고성능 칩 · H100 1만 개 이상 (추정)	· 엔비디아 저사양 칩 · H800 2,000여 개
모델 아키텍처	MoE 구조 (추정)	MoE 및 MLA와 같은 기법 활용
보안성	· 보안 정책 : 미국 및 영국의 AI 안전 가이드라인에 따라 체계적인 보안 기준을 따름 · 데이터 관리 : 다양한 규제 기관과 협력하며, 사용자 데이터 보호와 악의적 사용 방지를 위한 모니터링 체계를 갖춘	· 검열 및 제한 : 민감한 정치적 주제에 대해 엄격한 검열 및 응답 제한이 적용됨 · 데이터 관리 : 프라이버시 우려 (*개인정보 이슈 참고)
장점	· 안정성과 신뢰성 · 풍부한 부가 기능 : 음성 모드, 이미지 생성 등 다양한 기능 · 커뮤니티가 활성화되어 있음 · 정기적인 업데이트로 최신 정보 반영	· 합리적인 가격과 접근성 : 초당 60개 이상의 토큰 생성 · 모델을 공개하여 커스터마이징 가능 · 투명한 추론 과정 : CoT과정을 노출하여 모델의 문제 해결 과정을 이해하고 개선할 수 있는 기회 제공
약점	· 비용이 상대적으로 비쌈 · 모델이 폐쇄되어 있음	· 커뮤니티가 작음 · 다국어 잘 수행하지 못함 · (강력한 추론이라고 하지만) 추론의 경우 o1-preview보다 잘하지만, o1보단 성능이 떨어짐 · 추가 검증이 필요

자료: 한컴테크, IBK투자증권

학습 방식의 기여

GRPO 도입, 비용 95%↓
속도 2배↑ 달성

딥시크는 AI 학습 방식에 있어 여러 혁신을 도입하며 AI 산업에 기여했다 기존 AI 모델들이 주로 활용해온 인간 피드백 강화학습(RLHF)을 배제하고, 순수 강화학습(RL) 기반의 자율적 진화 방식을 도입했다. 특히, AI 모델이 스스로 정답을 찾아가도록 유도하는 '그룹 상대 정책 최적화'(GRPO) 학습법을 사용했다. GRPO는 여러 개의 응답 후보군을 그룹으로 묶고, 응답 간 상대 순위 정보만을 활용하여 정책(Policy)을 최적화하는 방식이다. 기존의 RLHF는 개별 응답에 대해 절대적인 보상을 부여하고 그에 따라 정책을 업데이트하기 때문에 평가 항목과 점수 저장, 비교 등의 작업이 필요한 반면 GRPO는 응답 그룹 내의 상대 순위만 활용하며, 절대 보상이나 외부 평가자의 점수가 없이 그룹 내 가장 우수한 응답일수록 더 높은 확률을 할당하도록 정책을 조정한다. 이러한 방식은 학습 과정에서 메모리 사용량을 75% 절감하고, 속도를 두 배 향상시키며, 학습 비용을 95% 이상 절감하는 성과를 달성했다.

전문가 혼합(MoE)으로
최소 연산에도 성능 유지

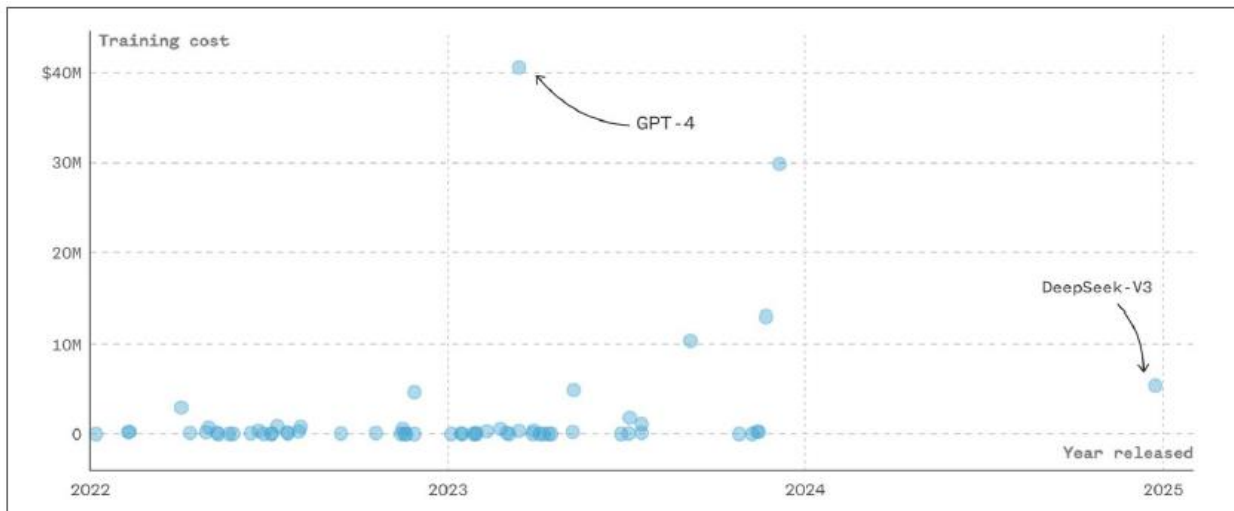
또한 딥시크는 '전문가 혼합(MoE, Mixture of Experts) 기법을 학습에 적용하여 모델의 크기를 키우지 않고도 성능을 향상시키며 계산 비용을 절감하는 효율적인 모델 개발을 실현했다. 이 방식에서 전체 모델은 수십억~수백억 개의 파라미터를 가진 여러 전문가(신경망 내부의 일부 서브모델)로 구성된다. 입력이 들어오면, 게이트(Gating Network)가 그 입력에 가장 적합한 소수의 전문가를 선택하고, 이 선택된 전문가들만 활성화되어 연산 수행하여 계산 효율을 극대화한다.

표 11. OpenAI(RLHF)와 DeepSeek R1(GRPO)비교

비교 항목	Open AI (RLHF)	DeepSeek R1(GRPO)
학습 방식	지도학습 + 강화학습 (PPO)	강화학습 (GRPO)
필요한 모델 수	4개 (AI, 평가, 리워드, 레퍼런스)	3개 (AI, 리워드, 레퍼런스)
GPU 사용량	매우 높음 (4개 모델 운영)	상대적으로 낮음 (평가 모델 제거)
학습 비용	높음	낮음
학습 안정성	하이퍼파라미터 조정이 어려워 Mode Collapse 가능성이 있음	상대적으로 안정적
데이터 의존도	보상 모델을 훈련하기 위해 직접적인 인간 피드백이 필요	그룹별 응답을 상대적으로 평가하여 정책을 최적화

자료: 한컴테크, IBK투자증권

그림 39. 딥시크발 낮은 훈련 비용 충격



자료: BOND, IBK투자증권

추론 방식의 기여

CoT·MoE로 비용↓ 성능↑ 달성

딥시크는 '전문가 혼합'(MoE) 기술을 추론 과정에 활용하여, 전체 모델을 구동하는 대신 특정 작업에 필요한 AI 부분만 활성화하여 컴퓨팅 자원을 절약했다. 이를 통해 모델의 크기를 키우지 않고도 성능을 향상시킬 수 있었다.

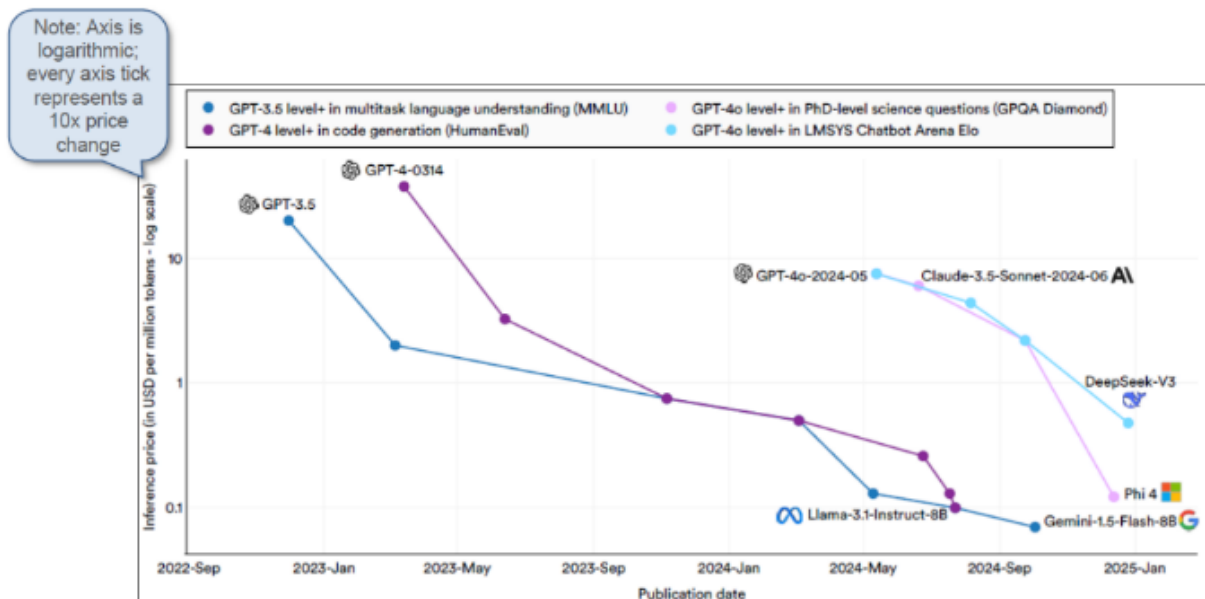
한편 딥시크 R1 모델은 여러 번의 사고 과정을 반복하며 스스로 검토하는 '체인오브소트(Chain of Thought)'와 같은 다단계 추론 기법을 적용하여 복잡한 문제 해결 능력을 획기적으로 높였다. 이러한 접근 방식은 기업들이 자체 데이터를 활용하여 고성능 AI 모델을 경제적으로 구축할 수 있는 로드맵을 제시했다. 특히, 딥시크의 R1 모델은 기존 오픈AI의 o1 모델과 유사하거나 일부 지표에서 능가하는 성능을 보이면서도, 획기적으로 낮은 추론 가격을 제시하여 대규모 언어 모델(LLM)의 광범위한 보급 가능성을 시사했다.

표 12. 딥시크의 기술 간의 관계

개념	역할	핵심 목적
GRPO (Group Relative Policy Optimization)	학습 방법	여러 답변 중 상대적으로 더 좋은 응답을 기준으로 모델을 학습함
CoT (Chain of Thought)	추론 방식	정답을 내기까지의 사고 과정을 언어로 드러내며 문제를 해결함
MoE (Mixture of Experts)	모델 구조	여러 전문가(서브모델) 중 상황에 맞는 일부만 선택적으로 활성화하여 연산 효율과 성능을 동시에 확보함

자료: Google, 언론 보도 종합, IBK투자증권

그림 40. AI 추론 비용 급감



자료: BOND, IBK투자증권

메타의 MSL(Meta Superintelligence Labs) Impact

AI 생태계, 메타발 초지능
인력 경쟁 시동

: AI인재 쟁탈전으로 AI스타트업 비용 구조 악화 & 오픈소스 진행의 위기

AI산업은 무한 생존 경쟁 구간에 진입

메타의 무제한 인재 유치 경쟁은 AI산업의 인력 생태계를 와해시키면서 수익성이 낮은 AI기업들의 파산이 급증할 것으로 전망된다. AI 서비스의 비즈니스 모델이 구축되지 않은 상황에서 AI의 사용자가 증가하는 것은 기업의 비용 구조를 악화시킬 수밖에 없다. 특히 메타발 인력 인플레이션은 비상장 기업 중심으로 부담을 가중시킬 것이다.

슈퍼인텔리전스 랩 신설과 조직 개편

2025년 6월, 메타는 슈퍼인텔리전스 개발을 위한 새로운 조직인 Meta Superintelligence Labs(MSL)를 출범했다. 기존의 AI 연구조직(FAIR, Llama 개발팀 등)을 통합해, 슈퍼인텔리전스(AGI, 초지능) 연구에 집중하는 형태로 조직을 재편했다. 이 과정에서 마크 저커버그가 직접 인재 영입과 조직 개편을 진두지휘하며, AI 분야에서 주도권을 잡으려는 의지를 보여줬다. 최근 LLM 성과가 기대에 미치지 못하면서, 저커버그는 슈퍼인텔리전스 개발 가속화를 위해 업계 최고의 인재를 대거 영입하는 전략을 선택했다. 2025년 한 해에만 AI 인재와 인프라에 143억 달러 이상을 투자하고 있으며, 인재 유치 경쟁이 AI 업계 전체로 확산될 것으로 예상된다. 이렇게 과감한 투자가 진행된 후에 Meta의 AI를 오픈소스로 계속 공개할지는 의문이다. 그동안 메타에 의존이 심했던 오픈소스 진행의 AI개발자 및 스타트업의 불확실성이 커질 것이다.

대규모 인재 영입과 파격적 보상

경쟁사인 OpenAI, Google DeepMind, Anthropic 등에서 대규모로 인재를 영입하고 있다. 최근에는 Scale AI의 창업자 알렉산더 왕을 최고 AI 책임자(Chief AI Officer)로 영입했고, 전 GitHub CEO인 Nat Friedman, 전 Safe Superintelligence CEO인 Daniel Gross 등 업계 유명 인사들도 합류했다. 이들에게 최대 1억 달러에 달하는 계약금과 스톡옵션, 무제한 컴퓨팅 리소스 등 업계 최고 수준의 조건을 제시하며, 인재 확보에 공격적으로 나서고 있다.

표 13. 메타 슈퍼인텔리전스 인력 흡수 현황

구분	주요 내용
조직 신설	Meta Superintelligence Labs(MSL) 출범, 기존 AI 조직 통합
핵심 영입 인재	Alexandr Wang(Scale AI 창업자), Nat Friedman(전 GitHub CEO), Daniel Gross, Ruoming Pang(애플 AI 총책임자) 등
경쟁사 영입 인원	OpenAI, Anthropic, Google 등에서 10명 이상 핵심 연구자 영입 스위스 OpenAI 연구소 3인 합류
보상 패키지	최대 1억 달러 계약금, 4년간 3억 달러 스톡옵션, 무제한 컴퓨팅 리소스 등
투자 규모	2025년 한 해에만 AI 인재 및 인프라에 143억 달러(약 19조 원) 투자
업계 반응	경쟁사들도 인재 유출 방지책 마련, AI 인재 확보 경쟁 심화

자료: 언론 보도 종합, IBK투자증권

2) AI Index 2025 보고서에 담긴 인사이트

美·中 AI 경쟁 격화

비용은 급락

AI Index 2025는 스탠퍼드대 인간중심 인공지능연구소(HAI)가 매년 발간하는 글로벌 AI 현황 분석 보고서로 최근에 8번째 보고서를 발간했다. 이 보고서는 AI의 기술 발전, 경제적 영향, 정책과 규제, 사회적 인식 등 전반적인 분야를 포괄한다.

연구개발 경쟁: 미·중의 양강 구도 속 비용의 혁신적 하락

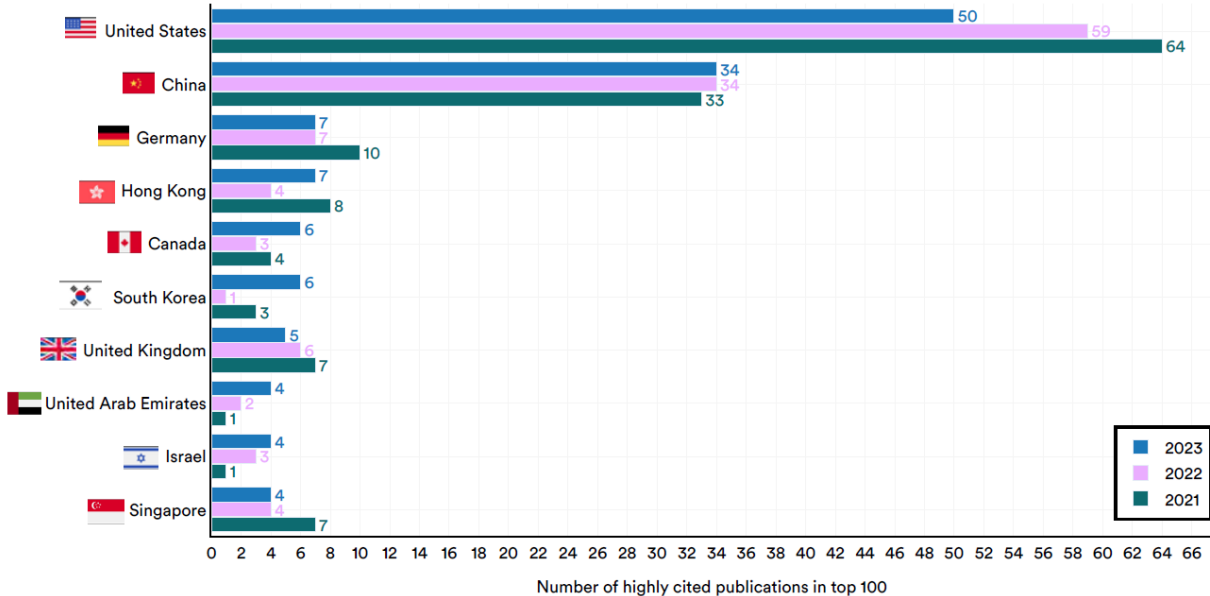
AI 연구개발 경쟁은 미국과 중국이 주도하고 있다. 논문 수에서는 중국이 23.2%로 세계 1위인 반면 고인용 논문 상위 100편 중 절반 이상은 미국이 차지하고 있어 질적 영향력 면에서는 미국이 우위다. 특히 관련해서 중국이 2023년 전체 AI등록 특허의 69.7%를 차지하며 총량에서 압도적인 선두를 기록 중이다. 한국은 인구 대비 AI특허수(10만명 당 17.3개)에서는 1위를 기록했다. 미국은 Epoch AI 선정 '주목할 만한 AI 모델' 중 40개를 차지하며 중국(15개), 유럽 전체(3개)를 크게 앞서고 있다.

최근 LLM모델 관련해서 주목할 점은 학습·추론 비용이 급격히 하락했다는 점이다. 예를 들어 100만 토큰당 쿼리 비용은 2022년 20달러에서 2024년 0.07달러로 280배 하락했다. 이는 AI의 대중화를 촉진할 수 있는 핵심 요소이다.

그림 41. 상위 인용 AI 논문 100편의 글로벌 비교

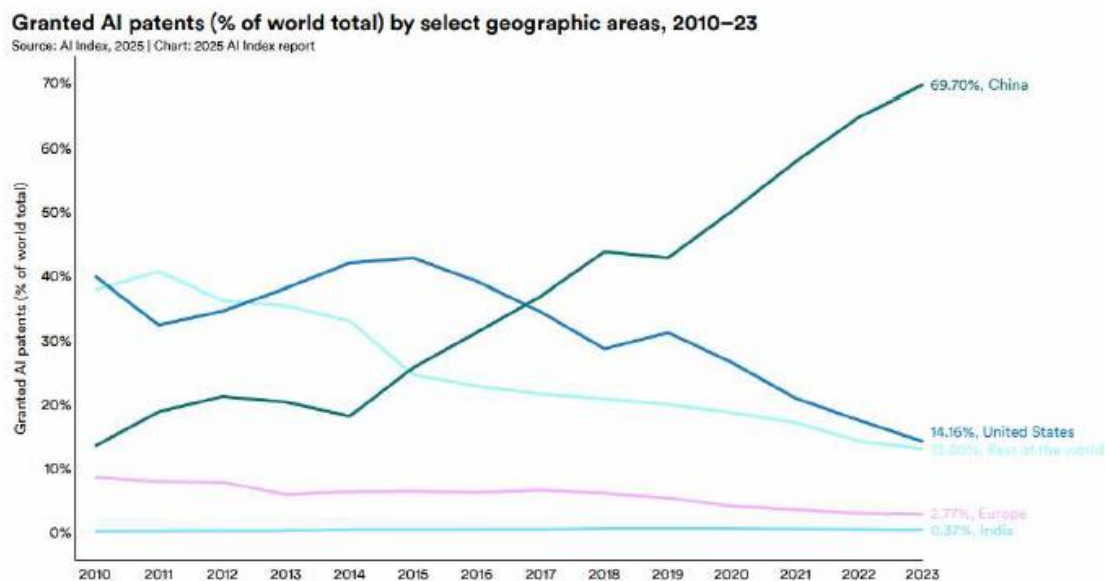
Number of highly cited publications in top 100 by select geographic areas, 2021-23

Source: AI Index, 2025 | Chart: 2025 AI Index report



자료: Stanford HAI 『AI Index Report 2025』, IBK투자증권

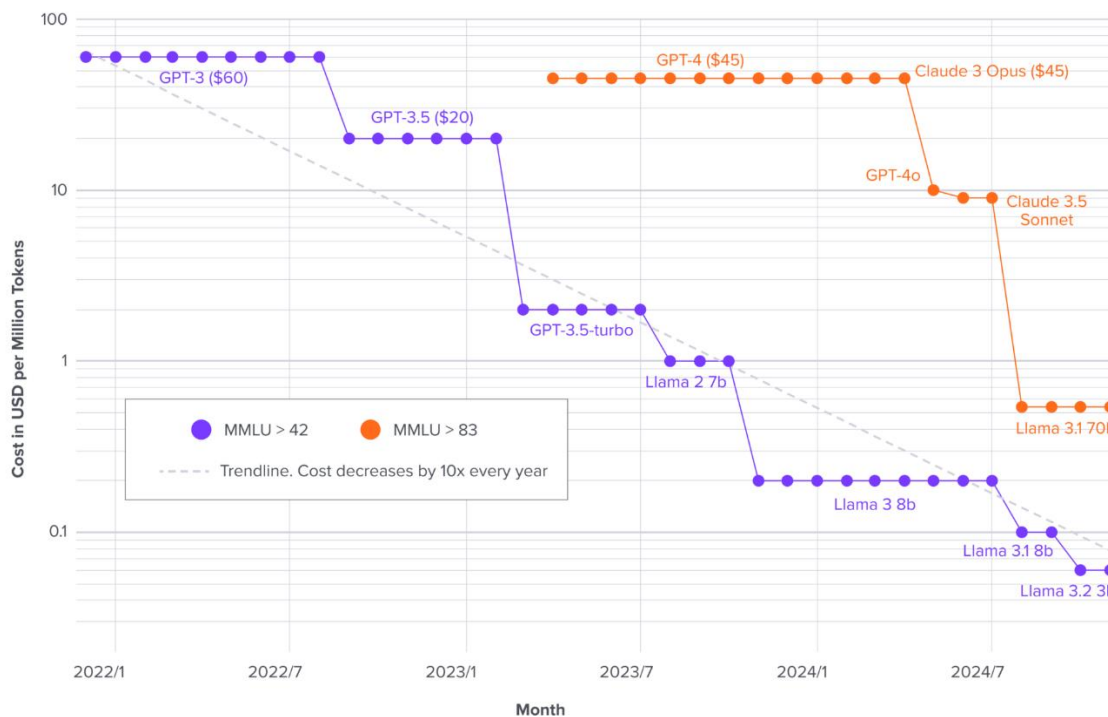
그림 42. 전체 AI 등록 특허 추이



자료: Stanford HAI 『AI Index Report 2025』, IBK투자증권

그림 43. LLM모델의 학습 및 추론 비용 급감

Cost of the Cheapest LLM with a Minimum MMLU Score (Log Scale)



자료: a16z, IBK투자증권

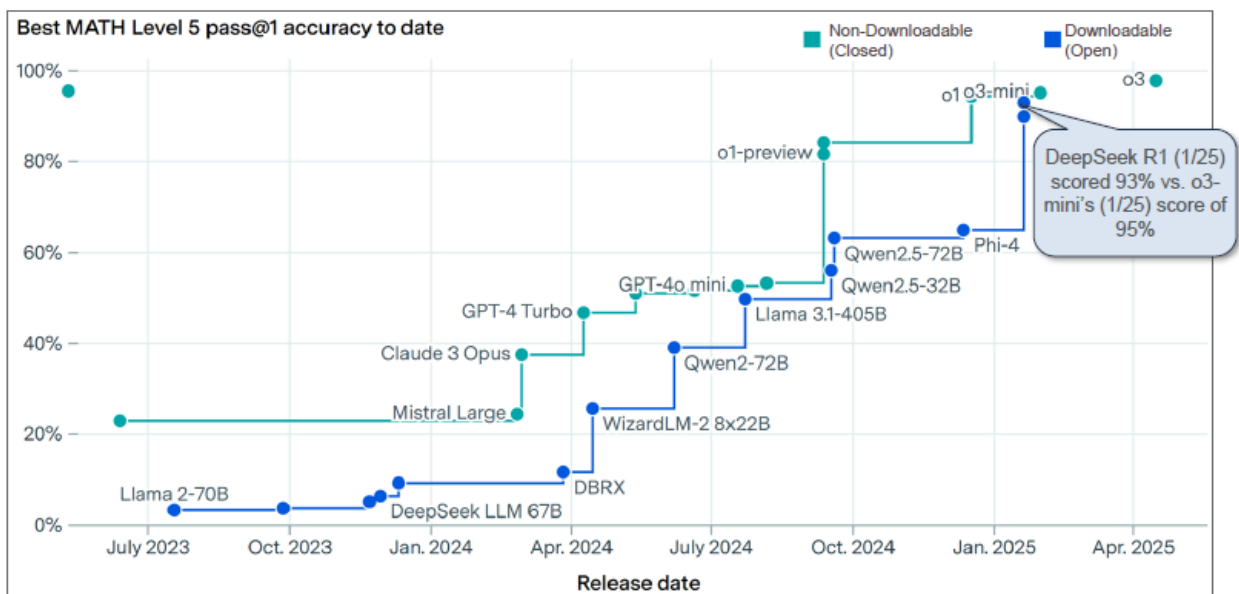
기술 성능: 모델 간 격차 감소, 상향 평준화 추세

AI 성능 급등에도 불구하고
여전히 고차원 추론에 한계

AI 성능은 급속도로 발전 중이며, 고난이도 벤치마크 점수도 1년 만에 최대 67%p 향상되었다. 개방형 모델과 폐쇄형 모델 간 성능 차이도 15.9%에서 1.7%로 축소되었고, 미국과 중국 간 성능 격차도 다수 벤치마크에서 한 자릿수로 좁혀지고 있다. 경량화 모델도 발전하여, 2024년 기준 Phi-3-mini(3.8B)가 GPT-3 수준의 성능을 달성하기도 했다.

한편 AI 에이전트가 부각되면서 관련된 성능 향상이 빠르게 진행 중이다. AI 에이전트는 고차원적인 추론, 예를 들어 연쇄적 사고나 논리·계획을 포함하는 문제에서는 여전히 제한이 존재하며, 짧은 시간의 작업에서는 인간보다 우수하지만 장기 작업에서는 성능이 저하되는 것으로 나타났다. AI는 결정적이고 반복적이며, 목표가 명확한 작업에서는 매우 강력하다. 그러나 장기적 전략 수립, 목적의 수정, 창의적 탐색이 필요한 상황에서는 인간의 직관·경험·전략적 사고가 뛰어난 것을 반증한다.

그림 44. 오픈소스모델 성능 향상에 기여



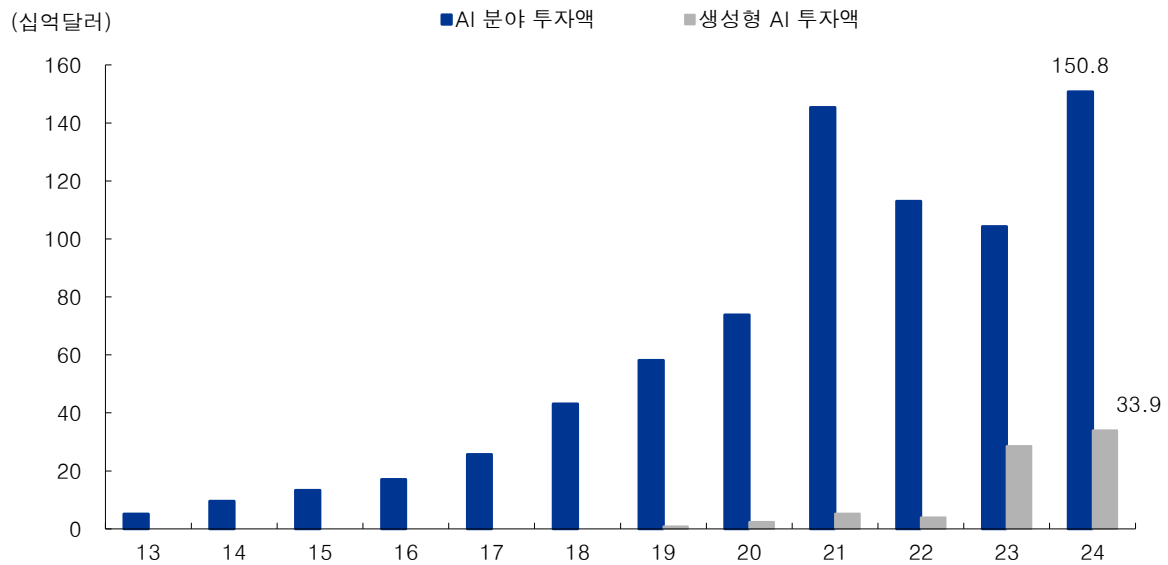
자료: BOND, IBK투자증권

경제: 민간투자 증가, 한국은 상대적 정체

2024년 AI 투자 1,508억 달러,
미국 72% 차지

2024년 글로벌 민간 AI 투자액은 전년 대비 44.5% 증가한 1,508억 달러에 달했다. 이 중 미국이 1,091억 달러로 전체의 72%를 차지하며 압도적인 우위를 보였고, 생성형 AI 투자도 339억 달러로 전체의 20%를 점유했다. 한국은 전체 투자액 13.3억 달러로 11위를 기록하며 전년보다 순위가 하락(‘23년 9위, 13.9억 달러)했다. 다만 신규로 투자받은 AI 기업 수는 52개(7위)로 꾸준히 증가 중이다.

그림 45. AI 및 생성형 AI에 대한 글로벌 민간 투자액



자료: Stanford HAI 『AI Index Report 2025』, IBK투자증권

4) 빅테크 3사의 최신 모델 및 서비스

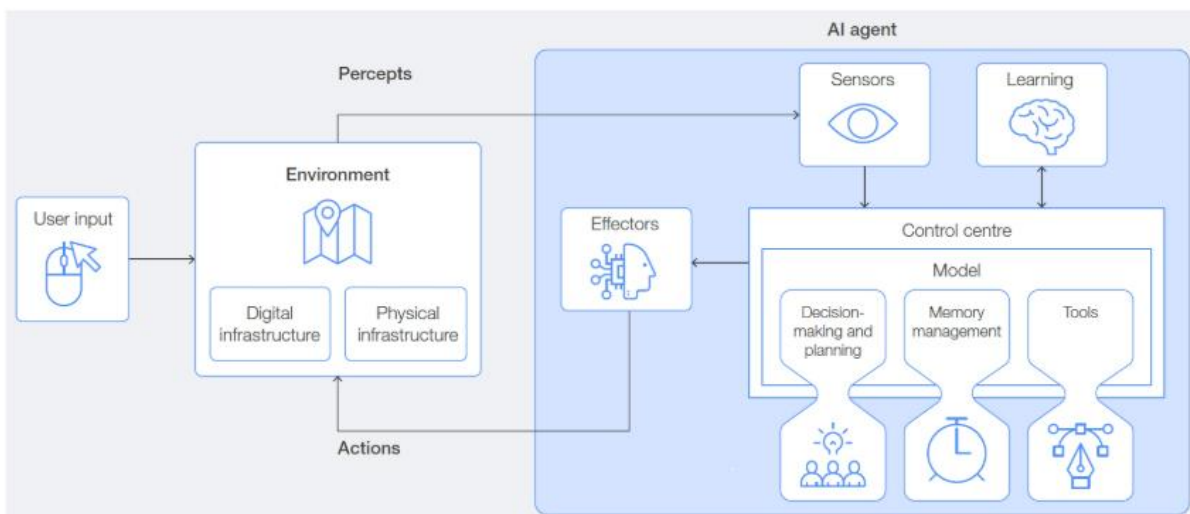
3사의 최신 모델과 서비스의 공통점

AI 에이전트 및 슈퍼 어시스턴트화

AI, 정보 제공자에서
실행형 비서(Agent)로 진화

세 기업은 공통적으로 인공지능을 단순한 정보 제공자에서 능동적인 행위자(Agent)로 진화시키는 데 초점을 맞추고 있다. Google은 Project Astra를 통해 웹 탐색, 예약, 코딩 등 실제 작업을 수행하는 AI 에이전트를 개발하고 있으며, 이를 검색 서비스에 통합하고 있다. OpenAI는 ChatGPT를 일정 관리, 이메일 작성, 데이터 처리 등 다기능 슈퍼 어시스턴트로 전환하고 있으며, 이는 단순 대화형 모델에서 일상적 의사결정을 돕는 실행형 AI로의 진화이다. Meta 역시 LLaMA 4 기반 AI를 WhatsApp, Instagram 등과 통합하여 개인 맞춤형 상호작용을 강화하고 있으며, 스마트 안경을 통해 실시간 AI 비서화를 추진한다. 이러한 변화는 향후 AI가 사용자의 일상 속 비서로 자리잡고, 디지털 에이전트 경쟁이 새로운 플랫폼 패권의 핵심이 될 것임을 시사한다.

그림 46. Key Components of Advanced AI Agents



자료: World Economic Forum, IBK투자증권

모델 세분화 및 제품 포트폴리오 다양화

빅테크, AI 모델 다변화로
원가 최적화 대응

빅테크 3사는 공통적으로 ‘하나의 최상위 모델’ 전략에서 탈피하여 성능, 속도, 비용, 용도별 모델 포트폴리오를 운영하고 있다. 이는 사용자의 니즈에 맞춘 모듈형 공급 구조로 진화하고 있다는 신호이며, 학습과 추론 비용이 급증하면서 공급 원가를 낮추기 위한 선택지를 제공하는 것이다. Google은 Gemini 2.5를 Pro, Flash, Flash-Lite로 세분화하여 고성능·속도·비용 간 균형을 맞췄다. OpenAI는 o3, o3-pro, o4-mini, GPT-4.1 등 다양한 스펙트럼의 모델을 병행하며, 고급 사용자는 고성능 모델을, 일반 사용자는 경량 모델을 선택할 수 있도록 설계했다. Meta는 LLaMA 4를 Scout, Maverick 등 전문가 수와 파라미터 기반으로 구분하여 다양한 수요에 대응한다.

표 14. 2025년 3사의 최근 LLM모델의 구분

구분	Google (Gemini 2.5)	OpenAI (O3 시리즈)	Meta (Llama 4 시리즈)
고성능	Gemini 2.5 Pro - 1M 토큰 컨텍스트 - 텍스트, 이미지, 오디오, 비디오 멀티모달 - 고도 추론, 에이전트 작업, 대규모 데이터 처리	O3-Pro - 최상위 추론/분석 성능 - 복잡한 수학, 코딩, 과학, 멀티모달 - 가장 깊은 reasoning, 신뢰도 최상	Llama 4 Maverick - 17B 활성/400B 총 파라미터 - 128 전문가(MoE), 1M 토큰 - 텍스트+이미지 네이티브 멀티모달
일반	Gemini 2.5 (기본/Flash) - 1M 토큰, 멀티모달 - 고성능 대비 비용·속도 최적화 - 다양한 서비스·API에 통합	O3 - 고성능 reasoning, 멀티모달 - 200,000+ 토큰 컨텍스트 - 실시간 검색·코딩·디버깅 등 도구 통합	Llama 4 Scout - 17B 활성/109B 총 파라미터 - 16 전문가, 10M 토큰 - 효율성과 성능 균형
경량	Gemini 2.5 Flash-Lite - 최저 지연, 비용 효율 - 모바일·임베디드 등 경량화	O4-Mini / O3-Mini - 속도·비용 최적화 - 24% 빠르고 63% 저렴	Llama 4 Scout (경량) - 17B 활성 파라미터 - 단일 GPU, 대량 사용자 지원

자료: Google, 언론 보도 종합, IBK투자증권

멀티모달 AI로의 확장

멀티모달 AI, 인간 수준의
직관을 위한 핵심 전략

모든 기업이 공통적으로 멀티모달(Native Multimodal) AI를 전략의 핵심으로 삼고 있다. 멀티모달 AI는 단일 언어 중심에서 인간 수준의 직관과 문맥 이해로 이동하고 있으며, 이는 향후 AI의 범용성과 상호작용 깊이를 결정짓는 핵심 요소가 될 것이다. 현실 세계 자체가 멀티모달이기 때문이다. 인간은 텍스트만이 아니라 시각, 청각, 움직임 등을 통합적으로 받아들이고 반응한다. 인간처럼 이해하고 대응하는 에이전트형 AI(Agentic AI)나 휴먼 인터페이스 AI를 만들기 위해선 네이티브 멀티모달이 필수이다.

그림 47. Gemini Live 실험



자료: Google, IBK투자증권

AI 플랫폼 생태계 및 하드웨어 통합

AI, 소프트웨어를 넘어
디바이스 통합 시도

세 기업 모두 AI를 단일 모델이나 소프트웨어 기능에 머무르지 않고, 하드웨어와 플랫폼 전반에 통합하려는 전략을 강화하고 있다. 향후 AI가 클라우드에 머무르지 않고, 모바일/웨어러블/IoT 디바이스로 확장되며 ‘AI-OS’ 생태계 전쟁으로 변질 가능성을 시사한다. Google은 Android XR Glasses, Google Beam, Gemini 통합 검색을 통해 AI가 내장된 플랫폼 생태계를 구축 중이다. OpenAI는 하드웨어 스타트업 ‘io’ 인수를 통해 독립적인 스마트 디바이스 생태계를 구축하려 하며, Apple이나 Google에 종속되지 않는 자체 플랫폼 전략을 추진하고 있다. Meta는 Ray-Ban, Oakley, Hypernova 스마트 글래스를 통해 AI 기반 착용형 디바이스 시장을 선도하고 있으며, 하드웨어와 LLaMA 모델을 결합해 일상 속 AI 구현에 집중하고 있다.

그림 48. 구글이 2025년 I/O 컨퍼런스에서 발표한 AR스마트안경 프로젝트 아우라



Project Aura | Android XR

자료: Google, IBK투자증권

그림 49. OpenAI가 애플 전 CDO인 조너선 아이브가 설립한 스타트업 ‘io’를 약 9조원에 인수



자료: TechCrunch, IBK투자증권

Google의 최신 LLM 모델 및 AI 전략 (2025년 발표)

Google, 사고형 모델과
XR 기반 AI 플랫폼 강화

Google은 2025년 Google I/O에서 Gemini 2.5 모델군을 발표하며 AI 역량을 핵심 경쟁력으로 내세웠다. Gemini 2.5는 '사고형 모델(Thinking Models)'이라는 개념을 도입하여, 응답 생성 전에 내부적인 추론 과정을 거치는 방식으로 성능과 정확도를 크게 향상시켰다. 이는 기존 LLM이 프롬프트에 즉시 반응하던 방식과 달리, 문제 해결을 위한 중간 사고 단계를 수행한다는 점에서 큰 전환점이다.

Gemini 2.5 모델군(Pro, Flash, Flash-Lite)은 공통적으로 사고형 아키텍처를 기반으로 하지만, 성능·비용·속도라는 세 가지 축에서 각각 차별화된 모델을 제공한다. Pro는 최고 성능, Flash는 속도와 비용 효율성, Flash-Lite는 최저 지연 시간에 초점을 맞춘다. 이는 단일 최상위 모델이 아닌, 다양한 목적에 맞춘 세분화된 모델 라인업으로 AI 시장이 진화하고 있다는 신호이다.

Gemini 2.5는 텍스트, 오디오, 이미지, 비디오 등 다양한 입력을 이해할 수 있는 네이티브 멀티모달 기능을 탑재하였다. 엔드 투 엔드 학습 구조를 기반으로 여러 모달리티를 통합적으로 처리할 수 있도록 설계되었다.

또한 Google은 AI 에이전트 기술(Project Mariner, Jules, Astra)을 통해 단순 질의응답을 넘어서 웹 상의 복잡한 작업 수행과 실시간 환경 인지 능력을 강화하고 있다. 이들은 Google Search에도 통합되어 검색 경험의 근본적 전환을 유도하고 있다.

하드웨어 분야에서도 Android XR Glasses를 통해 실시간 번역과 내비게이션 기능을 지원하며, Google Beam은 2D 비디오를 3D로 변환하는 AI 기반 통신 플랫폼으로 진화하고 있다.

그림 50. 구글 2025 I/O



자료: Google, IBK투자증권

OpenAI의 최신 LLM 모델 및 AI 전략 (2025년 발표)

OpenAI, 고성능 LLM과
스마트 디바이스로 생태계 확장

OpenAI는 2025년에도 다양한 LLM 모델을 출시하며 기술 리더십을 강화하고 있다. 주요 모델은 다음과 같다.

OpenAI o3: 2025년 4월 16일 출시된 o3는 OpenAI의 가장 강력한 '추론 중심 모델'이다. 코딩, 수학, 과학, 시각 인식 등 다방면에서 최고 수준의 성능을 보이며, 강화 학습 기반의 '사고 사슬(chain of thought)'을 내장하여 복잡한 문제 해결 능력을 제공한다.

- OpenAI o4-mini: 같은 날 출시된 o4-mini는 경량 모델로, 빠르고 비용 효율적인 추론에 특화되어 있다. 수학·코딩 등에서 크기 대비 우수한 성능을 보인다.
- OpenAI o3-pro: 2025년 6월 10일 ChatGPT Pro 및 API를 통해 공개된 o3의 상위 버전으로, 더 깊고 신뢰할 수 있는 응답을 제공하도록 설계되었다.
- GPT-4.1: 2025년 5월 14일 출시되었으며, GPT-4o 대비 더 정밀한 지시 수행과 웹 개발에 강점을 보인다. 최대 100만 토큰의 컨텍스트를 지원하여 대규모 문서 분석에 적합하다.

OpenAI는 고성능 모델에는 프리미엄 가격을 적용하는 동시에, o3-mini와 같은 경량 모델을 저가로 제공하는 전략을 취하고 있다. 이는 기술력뿐 아니라 시장 세분화, 가격 전략 등 비즈니스 전략을 LLM 개발의 핵심 요소로 통합한 사례이다.

OpenAI는 2025년 상반기, ChatGPT를 사용자 맞춤형 슈퍼 어시스턴트로 진화시키는 전략을 공개하였다. 일정 관리, 이메일 작성, 데이터 분석 등 실제 업무 지원이 가능한 다기능 플랫폼으로 변화하고 있다. 2025년 5월, 애플 전 CDO 조너선 아이브가 설립한 하드웨어 스타트업 'io'를 약 65억 달러에 인수함으로써, AI 기술을 스마트폰 등 물리 디바이스에 통합하고, 플랫폼 독립적인 생태계를 구축하려는 구체적 시도를 본격화하고 있다.

Meta의 최신 LLM 모델 및 AI 전략 (2025년 발표)

Meta, 초지능 조직 출범과
Llama 4로 AI 대중화 가속

Meta는 2025년 6월, 초지능 개발을 목표로 하는 Meta Superintelligence Labs(MSL)를 공식 출범하였다. 이는 FAIR, Llama 팀, 차세대 연구팀을 통합한 조직으로, AGI 개발에 집중하고 있다. 업계 인재들을 대거 영입하며, 공격적인 연구 투자와 인재 확보 전략을 병행하고 있다.

2025년 4월 5일, Meta는 Llama 4 모델군(Scout, Maverick)을 발표하였다. 이들은 Meta의 첫 오픈 웨이트 기반 네이티브 멀티모달 모델이며, MoE(Mixture-of-Experts) 구조와 1000만 토큰 이상의 컨텍스트 길이를 제공한다.

Llama 4 기반의 Meta AI는 Facebook, Instagram, WhatsApp에 통합되어, 사용자 맞춤형 상호작용 기능을 강화하고 있다. 또한 Ray-Ban, Oakley, Hypernova 스마트 안경 등을 통해 AI 기능을 일상 속 디바이스로 확장하고 있다. 특히 Hypernova는 HUD를 통해 텍스트 출력, 알림, AI 응답 등을 실시간으로 제공한다.

Meta는 LlamaCon 2025에서 오픈소스가 AI의 미래임을 강조하며, 독점 모델이 아닌 민주화된 접근을 지향하고 있다. Llama 4는 더욱 빠르고, 다국어 처리 능력이 강화되었으며, 광범위한 맥락 이해 능력을 제공함으로써 접근성의 장벽을 낮추는 데 기여하고 있다.

Meta는 ‘엔터테인먼트와 소셜 연결의 융합’을 AI 전략의 핵심으로 삼는다. 대규모 인프라 투자와 함께, 몰입형 콘텐츠 경험을 제공하고, 실시간 협업 및 커뮤니케이션 도구를 개발함으로써 디지털 공간에서의 소셜 상호작용을 혁신하려는 전략을 강화하고 있다.

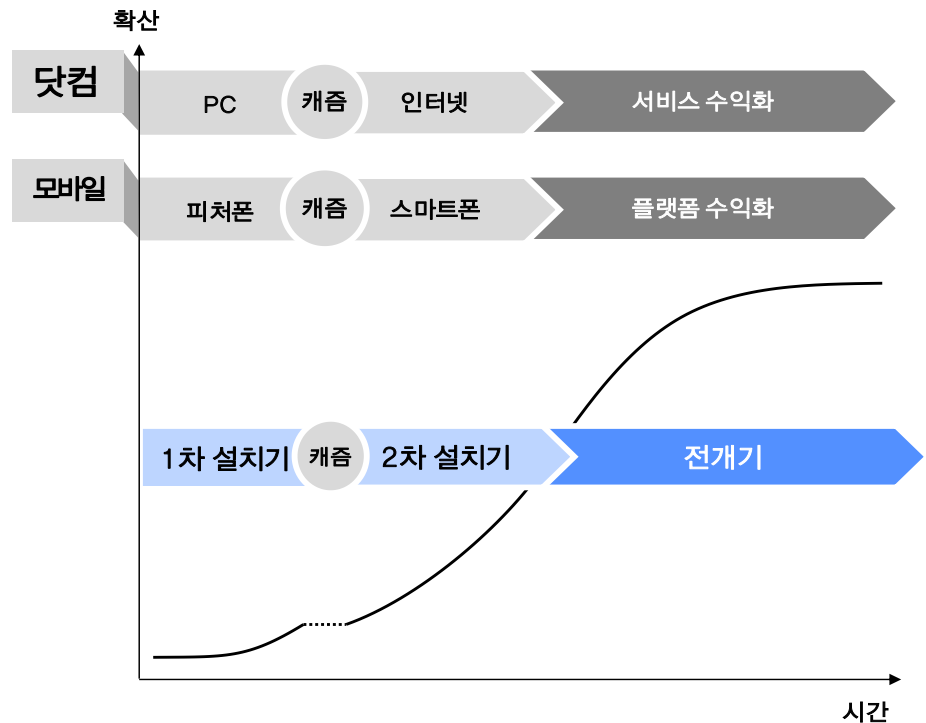
5. NEXT AI: 2차 설치기와 전개기 전망

AI 산업의 2차 설치기와 전개기의 주역 찾기

AI 산업, 캐즘 통과 중
2차 설치기 수혜주 주목

AI 산업은 현재 1차 설치기를 지나 캐즘의 단계에 진입한 것으로 판단된다. 1차 설치기의 주역은 NVIDIA를 대장으로 전력기기 등 데이터센터 인프라 혁신과 관련된 기업들이었다. 과거 모바일의 2차 설치기 주역은 애플의 아이폰이었다. 그 당시 이 혁신적인 제품을 예상하는 것조차 어려웠고 출시 이후에도 존재 가능성에 부정적 시각이 강했다. AI산업의 2차 설치기와 전개기의 주역을 예상하기는 어렵지만 최근 주목받는 트렌드, 기술 등을 통해 그 범주를 예측하고자 한다.

그림 51. 4단계는 각각 '기존 기술의 설치-정체-패러다임 전환-서비스 대중화'를 의미



자료: IBK투자증권

1) 최근 AI 산업에서 '추론'의 중요성이 급부상

AI 경쟁의 중심축,
'추론 최적화'로 이동

최근 인공지능(AI) 산업의 핵심 무게추가 '학습(Training)'에서 '추론(Inference)'으로 빠르게 이동하고 있다. AI 모델은 한 번 학습된 뒤, 수십억 번의 추론 과정을 통해 사용자와 상호작용하며 실제 가치를 창출한다. AI 모델의 학습은 막대한 계산 자원과 시간, 비용이 소요되는 고정비 성격의 작업이다. 반면, 추론은 학습된 모델을 다양한 실제 상황에서 사용하는 과정으로, 사용자마다 매 요청마다 반복 수행되기 때문에 연산 비용이 누적되고 예측 불가능한 수준으로 증가한다. GPT-4나 Gemini 같은 대형모델이 서비스되는 구조에서는 학습은 수개월에서 수년에 한 번 이뤄지는 반면, 추론은 하루에도 수십억 번씩 발생하는 '현장형 연산'이다. 따라서 산업적으로 보면, 학습은 일회성 설비 투자(CAPEX)에 가깝고, 추론은 지속적인 운영비용(OPEX)에 해당한다.

음성, 이미지, 텍스트 등 다양한 모달을 하나의 모델이 통합적으로 처리하는 '네이티브 멀티모달' 모델이 등장하고, 대화형 에이전트나 자율적 행동 AI(Agentic AI)의 등장이 가속화되면서, 추론 과정에서 요구되는 처리 능력과 연산 복잡도가 비약적으로 증가하고 있다. 이러한 상황에서는 단순히 대규모 데이터를 학습하는 것보다, 실시간으로 다양한 상황에 유연하게 대응하는 추론 시스템을 얼마나 잘 설계했는지가 AI 경쟁력을 결정짓는 요소가 되고 있다. 이에 따라 글로벌 빅테크 기업들도 모델을 여러 버전으로 출시하면서 추론 최적화를 통해 비용을 낮추는 전략을 강화하고 있다.

AI 추론, 클라우드에서
엣지로 확산

향후 AI 추론 인프라는 클라우드 기반에서 점차 엣지 디바이스(모바일, PC, 임베디드 칩)로 확산될 것으로 보인다. 이는 AI 모델의 분산 활용과 실시간 응답성 향상이라는 수요에 대응한 것으로, 삼성, 애플, 퀄컴 등의 반도체 기업들도 자체 NPU(신경망 처리 유닛)를 통해 추론 전용 하드웨어 개발에 속도를 내고 있다. 동시에 기업 고객들은 추론 단가를 낮추기 위해 자체 경량화 모델이나 오픈소스 모델을 도입해 API 의존도를 줄이려는 움직임을 보이고 있다.

표 15. AI모델에서 학습과 추론의 관계

항목	학습(Training)	추론(Inference)
비용 구조	초기 고정비 (수천억~조원 GPU 투자)	지속적 변동비 (API 비용, GPU 시간당 요금)
빈도수	수개월~수년에 한 번	매일 수십억 건
기술 방향	대규모 데이터 학습, 모델 구조 혁신	속도 최적화, 경량화, 캐시, 양자화
대표 사례	GPT-4, Gemini 1.5 Pro 학습	ChatGPT, Copilot, SGE, Claude 사용
핵심 기술	MoE, RLHF, CoT, Retrieval	Flash Attention, LoRA, Quantization, Edge AI

자료: Google, 언론 보도 종합, IBK투자증권

2] Agentic AI

OpenAI의 샘알트만이 제시한 AI의 5단계

AI, Reasoner 단계를 넘어
Agent로 진화

샘알트만이 제시한 AI 5단계의 주요 특징은, AI가 단순히 대화만 할 수 있는 수준(챗봇)에서 출발해, 점차 논리적 추론과 복잡한 문제 해결(Reasoner), 실제 업무를 자율적으로 처리하는 능력(Agent), 창의적으로 새로운 지식과 아이디어를 만들어내는 단계(Innovator), 그리고 여러 AI가 협력해 인간 조직처럼 움직이는 대규모 시스템(Organization)으로 진화한다는 점이다. 각 단계는 AI의 자율성과 창의성, 그리고 인간 사회에 미치는 영향이 점점 커진다는 특징을 가진다. 최근 ChatGPT형 AI는 Reasoner 단계를 진입해서 Agent를 지향한다.

표 16. 샘알트만이 제시한 AI의 5단계

단계	설명
Chatbot	단순 대화형 AI. 질문에 답하거나 정보를 제공하는 수준의 AI. 인간과 자연스럽게 대화할 수 있지만, 복잡한 추론이나 자율적 행동은 불가능
Reasoner	논리적 추론 및 단계별 문제 해결이 가능한 AI. 복잡한 수학, 과학, 프로그래밍 등에서 단계적 논리 추론을 수행하며, 자신의 추론 과정을 설명할 수 있음
Agent	실제 세계에서 자율적으로 업무를 수행하는 AI. 인간의 개입 없이 특정 작업을 완수할 수 있으며, 소프트웨어 엔지니어링 등 실제 업무 현장에서 가상 동료로 활동할 수 있음
Innovator	새로운 과학적 지식이나 혁신을 스스로 창출하는 AI. 기존 지식을 조합하거나 새로운 아이디어를 만들어내는 창의적 능력을 갖춘
Organization	인간 조직의 복잡한 기능을 복제하는 다중 에이전트 시스템. 여러 AI가 협력하여 하나의 조직처럼 움직이며, 인간 조직의 역할과 구조를 대체할 수 있음

자료: Google, IBK투자증권

① AI 에이전트

AI 에이전트,
반응형 개인 비서

AI 에이전트는 센서를 통해 환경을 인지하고, 이러한 인지와 자체 규칙에 따라 행동할 수 있는 로봇 또는 컴퓨터 프로그램을 일반적으로 지칭한다. 이는 환경으로부터 감각 입력을 받아들이고 환경을 변화시키는 행동을 수행한다는 의미에서 '상황 인지' 특성을 가진다. AI 에이전트는 주로 사용자와의 상호작용이나 특정 작업 수행에 초점을 맞추어 특정 업무에 대한 실시간 지원을 제공하는 반응형 AI 시스템으로 볼 수 있다.

② 에이전틱 AI의 정의 및 주요 특징

에이전틱 AI,
자율형 협력 시스템

에이전틱 AI는 인공지능 분야의 새로운 방향으로, 최소한의 인간 개입으로 복잡한 목표를 추구하도록 설계된 자율 시스템을 의미한다. 이러한 시스템은 진화하는 환경에서 동적으로 작동할 수 있도록 적응력, 고급 의사결정 능력 및 자율성을 보여준다. 에이전틱 AI는 대규모 언어 모델(LLM)을 중심 뇌로 사용하여 환경을 인지하고, 계획하며, 기억하고, 행동하는 방식을 취한다.

표 17. 구글 Cloud에서 제시한 에이전틱 AI의 사례

분야	에이전틱 AI 활용 예시
Customer service	고객 문의 처리, 문제 해결, 맞춤형 지원 제공 등을 통해 인간 상담사가 더 복잡한 문제에 집중할 수 있도록 지원
Supply chain management	판매, 재고, 운송 등 다양한 데이터를 분석해 공급망 최적화, 수요 예측, 물류 자동화를 수행
Healthcare	의료 기록, 연구 논문, 임상시험 데이터를 분석해 진단, 치료 계획 수립, 신약 개발 지원
Financial services	시장 데이터, 고객 행동, 재무제표를 분석해 사기 탐지, 리스크 평가, 투자 전략 자동화
Software development	코드 생성, 디버깅, 테스트 자동화를 통해 개발 속도 향상 및 코드 품질 개선

자료: Google, IBK투자증권

표 18. AI 에이전트와 에이전틱 AI의 비교

특징/분류	AI 에이전트	에이전틱 AI
자율성 수준	특정 규칙에 따라 반응하며, 제한된 자율성을 가짐	최소한의 인간 개입으로 복잡한 목표를 자율적으로 추구.
목표 복잡성	특정 업무 수행 및 사용자 요청에 반응	복잡하고 다단계의 목표를 설정하고 달성.
핵심 능력	환경 인지, 특정 작업 수행, 즉각적인 반응	도구 사용, 자기 반성, 계획 수립, 다중 에이전트 협업.
학습 및 적응	주로 정해진 데이터나 규칙 기반 학습	동적으로 환경에 적응하고, 경험을 통해 학습 및 개선.
의사결정	사전에 정의된 규칙이나 학습된 패턴에 따라 결정	고도화된 추론과 계획을 통해 스스로 의사결정.
예시	콜센터 상담 도우미, AI 비서, 개인화 챗봇	자율 금융 자문 시스템, 산업별 맞춤형 에이전트

자료: Google, IBK투자증권

③ 에이전틱 AI의 핵심 프로토콜로서 MCP와 A2A 기술 부각

MCP는 외부 도구 연동,
A2A는 에이전트 간 직접 통신

클로드의 MCP(Model Context Protocol)와 구글의 A2A(Agent-to-Agent) 프로토콜은 에이전틱 AI 시스템의 상호 운용성과 협업을 강화하기 위한 핵심적인 개방형 표준이다. 이 두 프로토콜은 에이전틱 AI 생태계의 기반을 형성하여, 다양한 AI 에이전트 간의 안전하고 확장 가능한 통신을 가능하게 하며 외부 도구 및 리소스에 대한 접근을 용이하게 한다. MCP가 도구 중심인 반면 A2A는 에이전트 중심으로 차이가 존재한다. 비유하면 MCP는 컴퓨터(LLM)가 주변기기(외장하드, 모니터) 등과 연결해주는 기술인 반면 A2A는 여러 컴퓨터(에이전트)가 서로 직접 연결하여 데이터를 교환하게 하는 네트워크 통신과 유사하다.

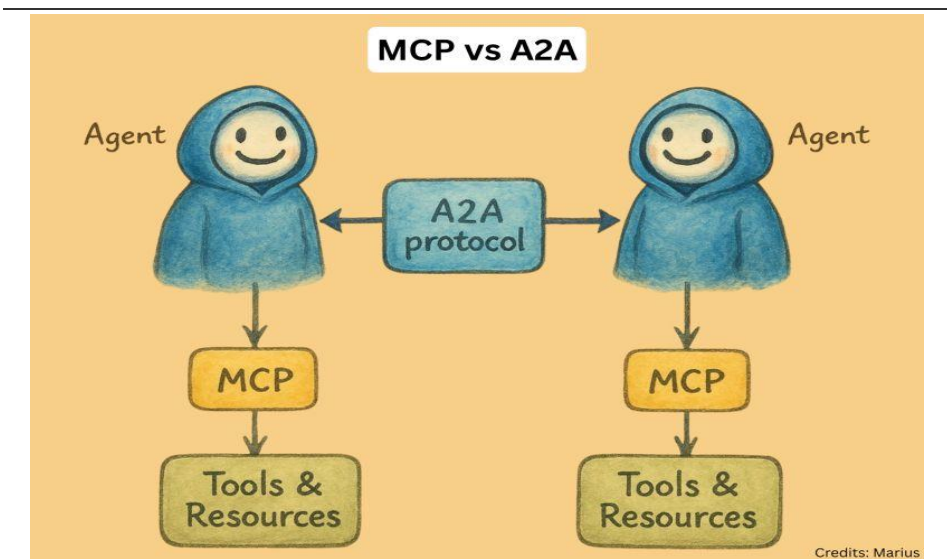
클로드의 MCP(Model Context Protocol)

클로드의 MCP는 앤스로픽(Anthropic)이 도입한 표준화된 인터페이스로, AI 모델이 외부 도구 및 리소스와 원활하게 상호작용할 수 있도록 설계되었다. 이 프로토콜은 서로 다양한 시스템 간에 갇혀있는 정보를 연결하고 상호 운용성을 촉진하는 데 중점을 둔다. 새로운 보안 문제를 야기할 수 있으므로 엄격한 분석과 완화를 위한 보안 메커니즘을 제공한다.

구글의 A2A(Agent-to-Agent) 프로토콜

A2A 프로토콜은 2025년 4월에 구글이 오픈 소스 프로젝트로 출시한 개방형 표준이다. 이 프로토콜은 이기종 AI 에이전트 간의 안전하고 확장 가능한 통신 및 협업을 가능하게 한다.

그림 52. MCP와 A2A 비교



자료: IBK투자증권

Credits: Marius

④ 에이전틱 AI와 Physical AI의 관계(LLM에서 LAM으로 진화)

에이전틱 AI,
Physical AI 시대의 선행 조건

Physical AI는 인공지능(AI)이 물리적 세계와 직접 상호작용하고, 센서와 액추에이터를 통해 환경을 인지하고 행동하는 능력을 갖추는 것을 의미한다. 전통적인 AI가 주로 신호 처리 및 데이터 기반으로 제한되어 온 것과 달리, Physical AI는 물리 법칙을 이해하고 이를 AI 모델에 통합하여 현실 세계의 문제 해결 능력을 향상시키는 것을 목표로 한다.

에이전틱 AI와 Physical AI는 인공지능의 자율적인 상호작용 능력을 현실 세계에 구현하는 데 있어 상호보완적인 관계를 가진다. 에이전틱 AI는 주로 소프트웨어 내에서 자율적 의사결정과 행동 계획 기능을 수행하는 반면, Physical AI는 이러한 자율적인 인공지능 시스템이 로봇, 자율주행차, 드론 등 물리적 세계에서 직접 동작할 수 있도록 하는 기술 분야를 의미한다. 대규모 언어 모델(LLM)이 물리적 세계 정보를 처리하고 추론하는 능력을 확장하기 위한 'Penetrative AI' 개념(LLM이 단순히 글만 이해하는 것이 아니라, 현실 세계의 데이터까지 해석하고, 그 정보를 바탕으로 실제 상황을 추론하는 능력을 갖추게 하는 것을 의미)은 IoT 센서 데이터를 해석하고 물리적 영역의 작업에 대해 추론하는 LLM의 능력을 탐구한다. 이를 통해서 사이버-물리 시스템에 인간 지식을 통합하는 새로운 방법을 가능하게 한다.

최근 1차원의 LLM에서 3차원 적용을 위한 LAM(Large Action Model, 대형 행동 모델)이 주목받고 있다. LLM(대형 언어 모델), LAM(대형 행동 모델), 그리고 physical AI(물리적 인공지능)는 AI의 진화 과정에서 서로 긴밀하게 연결되어 있다. LLM이 '생각'과 '이해', LAM이 '디지털 행동', physical AI가 '실제 세계의 행동'을 담당하며, 세 기술은 언어 이해에서 디지털 실행, 나아가 물리적 환경 제어까지 AI의 역할을 점차 확장해가는 흐름 속에 있다.

표 19. LLM과 LAM비교

항목	LLMs	LAMs
핵심 기능	패턴 기반 텍스트 생성	API 및 인터페이스를 통해 작업 실행
기반 기술	고유한 모델 아키텍처 (트랜스포머)	독자적인 아키텍처는 아직 없음
학습 방식	방대한 텍스트 데이터셋을 통해 완전 학습	별도 학습 없이 LLM에 도구만 추가
모델 상태	널리 배포되었으며 지속적으로 학습 중	대부분 개념적 수준, 현재 LLM 위에 구축됨
자율성	반응형: 명시적 프롬프트 필요	준자율형: 단단계 작업을 수행 가능
상호작용 방식	채팅 또는 텍스트 기반 인터페이스	앱 기반 또는 명령어 기반 상호작용
대표 사례	GPT-4, Claude, Llama, Gemini	Rabbit R1(LLM 기반), AutoGPT 일부, Devin(제한적 자율성)
기술적 현실	기술적으로 확립됨	LLM + 커스텀 인터페이스 + 앱 연결 방식

자료: Google, 언론 보도 종합, IBK투자증권

표 20. Physical AI 관련 기업

회사명	설립 연도	본사	주요 서비스 및 특징	추가 정보
Boston Dynamics	1992	미국 매사추세츠주 월섬	Spot(로봇개), Atlas(휴머노이드) 등 첨단 로봇에 AI 적용. 산업 자동화, 물류, 보안 등 다양한 분야 진출. MIT에서 스펀오프.	2021년 현대자동차그룹이 지분 80% 인수. DHL과 협력하여 Stretch 로봇을 전 세계적으로 배치 예정.
Anduril Industries	2017	미국 캘리포니아주 코스타메사	미국의 첨단 방위 기술 기업으로, 자율 시스템과 인공지능(AI) 기반 국방 솔루션 개발에 특화. 2017년 오클러스 VR의 창업자 팔머 럭키(Palmer Luckey) 등이 창업	미국 국방부 및 동맹국에 기술 판매. 2024년 매출 10억 달러, 직원 수 3,500명. Founders Fund, Andreessen Horowitz 등으로부터 투자 유치.
Waabi	2021	캐나다 토론토	AI 기반 자율 트럭 운송 솔루션. 실제 도로 환경에서 자율주행 기술 상용화. 독점 소프트웨어로 상업용 배송 경로에서 자율 주행 자동화.	Uber, Khosla Ventures, Volvo Group 등으로부터 총 2억 8,400만 달러 투자 유치. AI 선구자 Raquel Urtasun이 설립.
Figure	2022	미국 캘리포니아주 산호세	범용 업무 수행이 가능한 휴머노이드 로봇 개발. 노동력 부족 문제 해결과 생산성 향상에 중점. Figure 01, Figure 02 프로토타입 공개. BMW와 협력하여 자동차 제조 시설에 로봇 배치 예정. OpenAI와 협력했으나 현재는 종료.	2024년 2월 Jeff Bezos, Microsoft, Nvidia, Intel, Amazon, OpenAI 등으로부터 6억 7,500만 달러 투자 유치 (기업 가치 26억 달러).
RLWRLD	2022	대한민국 서울	로봇용 AI 파운데이션 모델 개발. 실제 환경에서 적용 가능한 로봇 지능에 집중. 로봇이 인간처럼 보고, 생각하고, 손을 사용할 수 있는 능력 부여.	2025년 4월 시드 라운드에서 1,480만 달러 투자 유치. FuturePlay, Global Brain, Hashed 등 투자 참여.
Physical Intelligence (Pi)	2024	미국 캘리포니아주 샌프란시스코	로봇에 실제 물리적 지능을 부여하는 π (파이제로) 모델 개발. 로봇이 식품 포장, 세탁물 정리, 달걀 포장 등 다양한 작업 수행. 일반 목적 AI를 로봇 및 기타 물리적 장치에 통합하는 데 중점.	Thrive Capital, Lux Capital 등으로부터 총 4억 7,000만 달러 투자 유치 (기업 가치 24억 달러). OpenAI의 지원을 받는 엔지니어 그룹으로 구성.

자료: IBK투자증권

그림 53. Anduril Industries에서 개발한 Barracuda(바라쿠다) 자율 무인 잠수정(AUV)



자료: Anduril Industries, IBK투자증권

그림 55. Figure사의 가사도우미로봇 Helix



자료: Figure, IBK투자증권

그림 54. Waabi의 자율주행 트럭



자료: Waabi, IBK투자증권

그림 56. Physical Intelligence (Pi)사의 Pi0 동영상 캡처



자료: Physical Intelligence, IBK투자증권

6. AI 산업의 캐즘 가능성이 크지만 붕괴는 없다

1) 빅테크 기업들의 AI 투자 확대와 부족한 수익

AI 인프라 CAPEX 급증,
성과는 제각각

글로벌 주요 빅테크 기업들의 CAPEX(자본지출) 지표를 보면, 2024년 알파벳(구글), 아마존, MS, 메타 등 4사 합산이 2,460억 달러(약 320조 원, 한국 2024년 정부 예산의 약 50%)에 달했으며, 2025년에는 3,200억 달러(약 440조 원)를 넘어설 전망이다. 이는 2023년 약 1,510억 달러 대비 2배 이상 증가한 수치이다.

아마존은 2024년 770억 달러에서 2025년 1,000억 달러 이상의 CAPEX를 계획하며, 대부분을 AWS 클라우드와 AI 인프라 확충에 투입할 예정이다. 메타 또한 2024년 390억 달러에서 2025년 600억~650억 달러로 투자액을 늘려 생성형 AI와 본업(광고/메타버스) 모두를 지원할 방침이다. 구글 모회사 알파벳은 2025년 약 750억 달러의 CAPEX를 계획하며 AI 중심의 데이터센터 확충에 나섰다. MS는 2025년 800억 달러를 AI 데이터센터에 투입할 것이라고 밝혔는데, 이는 미국 내 대규모 AI 인프라 구축 경쟁을 주도하는 투자이다.

투자 대비 당장의 수익성은 기업마다 차이가 있다. 엔비디아의 경우 AI 붐의 직접적인 수혜로 매출과 이익이 폭발적으로 증가하여 투자자들이 비교적 낙관적인 시각을 보이고 있다. 반면 MS, 구글, 아마존 등은 전문학적인 투자에 비해 즉각적인 수익 창출이 제한적이어서 투자자들이 우려를 나타내기도 했다.

MS의 Azure OpenAI와 Copilot 등의 매출이 2024년에 폭발적으로 증가하여 전년 대비 +175% 성장했고 연간 130억 달러 규모에 이르렀다. 아마존 또한 AWS 내 AI 서비스 수요 증가로 2024년 클라우드 매출이 +19% 성장하고 영업이익이 크게 늘었다. 반면 구글의 Bard, Gemini 등은 아직 수익화 단계가 초기이거나 간접효과(검색 광고 유지 등)에 그치고 있다. 메타는 생성형 AI를 직접 판매하지는 않지만, AI 추천 시스템을 고도화하여 페이스북/인스타그램의 광고 효율을 높이고 신규 AI 기능(예: 이미지 생성)을 도입하여 사용자 참여를 끌어올림으로써 간접 수익에 기여하고 있다.

표 21. 글로벌 빅테크 기업의 AI투자 및 매출

기업 (분야)	2024년 CAPEX	2025년 CAPEX 계획	AI 관련 매출 및 성장률 (최근)
아마존 (AWS 클라우드)	\$770억 (추정)	\$1,000억+ (대다수 AWS/AI 인프라)	AWS 2024년 매출 \$1,076억(+19% YoY) 2024년 영업이익 \$398억 (전년 \$246억)
마이크로소프트 (Azure)	미공개	\$800억 (AI 인프라)	Azure OpenAI 고객 18,000개 (+7,000 ↑, 2023년 하반기) AI 사업 연간 매출 \$130억(run-rate, +175% YoY)
알파벳 (Google)	애널리스트 추정 \$580억	\$750억 (데이터센터 증설)	Cloud 부문 2023년 흑자 전환 (분기 영업이익 \$1.9억) AI 제품 (Bard, Gemini 등) 직접 매출 미미
메타 (Facebook)	\$390억	\$600~650억 (AI + 메타버스)	AI 직접 매출 적으나, AI 추천 · 타게팅으로 광고 효율 ↑ ; 2024년 광고 매출 · 이용자 지표 반등
엔비디아 (NVIDIA)	\$13억 (FY2023, 제조 자산 투자)	(TSMC 위탁 생산으로 설비투자 낮음)	2024년 Q3 매출 \$181억 (+206% YoY) 분기 순이익 \$100억(사상 최대)

자료: IBK투자증권

2) 생성형 AI 서비스 상용화 현황과 이용자 추이

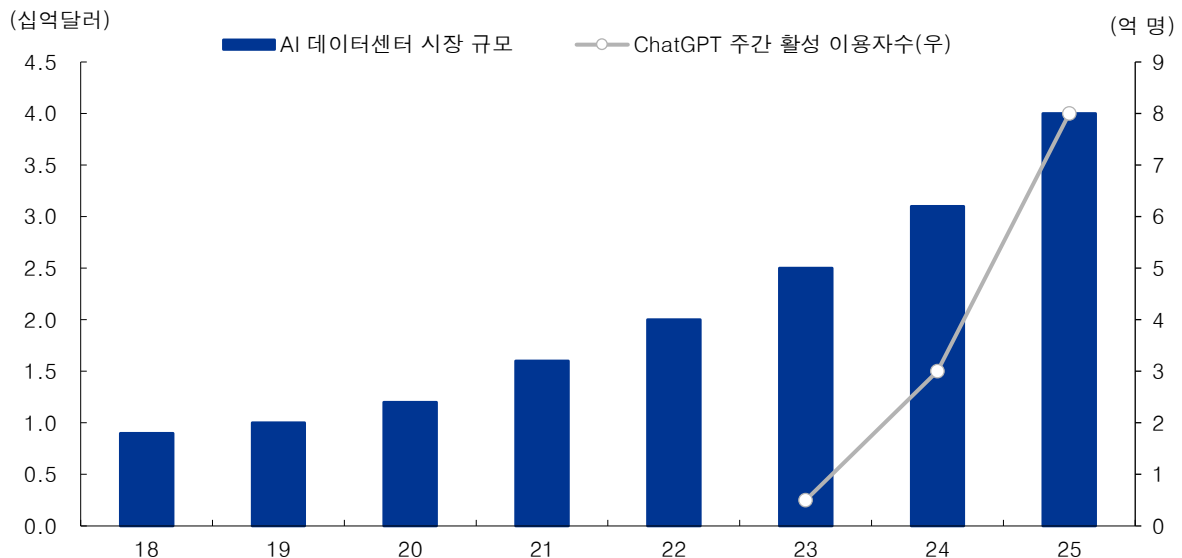
ChatGPT의 폭발적 대중화에도
아직 수익화는 과제

ChatGPT를 필두로 한 생성형 AI 서비스들은 전례 없는 속도로 대중화되었다. 2022년 말 출시된 ChatGPT는 출시 두 달 만에 월간 1억 명 이상의 사용자를 끌어모아 역사상 가장 빠른 소비자 앱 성장 사례로 기록되었다. 2023년 들어서도 사용량이 꾸준히 증가하여 2025년 중반에는 주간 활성 이용자 수가 8억 명에 육박했다는 추정치도 있다. MS에 따르면 “Azure OpenAI 서비스는 출시 1년 만에 18,000개 이상의 기업 고객을 확보하여 6개월 만에 고객 수가 11,000→18,000으로 급증했다”고 한다. 이는 OpenAI의 GPT 모델을 기업용 API나 MS 클라우드를 통해 도입한 조직 수를 뜻하며, 포춘 500 기업의 절반 이상이 이미 Azure OpenAI를 도입했거나 실험 중인 것으로 알려져 있다.

그러나 활발한 사용량 대비 수익화율은 아직 낮은 편이다. 예컨대 ChatGPT의 경우 유료 구독(ChatGPT Plus) 전환율이 5% 내외에 불과하다는 분석이 있어, 대다수 사용자가 무료 버전을 쓰고 있는 실정이다. 대신 OpenAI는 API 판매를 통해 기업으로부터 매출을 올리고 있는데, 2023년 약 2억 달러 매출에서 2024년 10억 달러 달성을 목표로 공격적인 영업을 전개하고 있다. MS는 2023년 말부터 기업용 Microsoft 365 Copilot(Office에 GPT 기능 통합)을 사용자당 월 30달러에 판매하기 시작했는데, 포춘 500 상당수가 이를 시험 도입하여 잠재 매출원이 가시화되고 있다. 구글도 Workspace에 생성 AI 기능(문서 요약, 자동 완성 등)을 추가하여 유료 플랜을 제시했다. 이용자 규모의 성장세는 뚜렷하지만, 수익 모델 확보와 사용자 유료 전환이 향후 핵심 과제로 남아 있다.

LLM기반의 주요 AI기업의 GPT 형태 서비스는 기존 인터넷, 모바일 환경의 서비스의 연장선에 머물러있다. 따라서 AI 데이터센터의 확장과 GPT서비스의 보급은 앞서 이야기했던 1차 설치기에 해당하는 것으로 판단한다.

그림 57. AI 데이터센터 확장과 GPT 보급은 1차 설치기



자료: Horizon Grand View Research, Demandsage, IBK투자증권, 주: 2025년 AI 데이터센터 시장 규모는 전망치

3) AI 모델 개발·운영의 비용 급증

AI 모델 운영비 과중,
비용절감 필요성 대두

아직 뚜렷한 수익모델이 부재한 상황에서 AI의 사용이 증가할수록 AI테크 기업들의 부담이 가중될 것이다. 거대 AI 모델의 학습(training)과 추론(inference)에는 막대한 비용이 든다. OpenAI의 사례를 보면 2023년 초 분석에 따르면 ChatGPT (GPT-3.5 기반)의 추론 비용만 하루 70만 달러(약 9억 원)에 달하며, GPT-4 모델을 사용할 경우 비용은 그보다 훨씬 높아질 것으로 추정되었다. 이는 거대 언어 모델이 매 요청마다 수천억 개의 연산을 수행해야 하며, 이를 위해 고성능 GPU 서버가 대규모로 동원되기 때문이다. OpenAI의 2022년 재무에서 5.4억 달러의 손실이 발생한 것도 GPT-4 개발 및 ChatGPT 출시 전후로 컴퓨팅 자원 비용이 폭증한 영향이 컸다.

AI 모델 추론 비용이 특히 큰데, 전문가들은 “대규모 모델 배포 시에는 추론(서빙) 비용이 훈련 비용을 주 단위로 상회한다”고 지적한다. 즉 한 번 거대한 모델을 만들어 놓으면, 이를 수백만 사용자가 지속적으로 질의응답에 쓰는 과정에서 전기료, 클라우드 사용료 등 운영비가 훈련비보다 더 많이 든다는 것이다. 실제로 구글 모기업 Alphabet의 회장 John Hennessy는 “자사 Bard 챗봇에 한 번 질의 답변하는 데 드는 비용이 일반 검색 질의의 10배”라고 밝혔다.

현재 무료 혹은 낮은 요금제로 사용자 기반을 확장하는 전략이 많은데, 장기적으로 과금 구조를 만들지 못하면 사업 지속이 어렵기 때문이다. 이에 따라 AI관련 매출을 늘리는 동시에 추론 과정 비용을 줄이기 위한 노력이 진행 중이다. OpenAI는 매출 확대를 위해 API 사용료와 ChatGPT 유료판을 도입했고, MS와 구글 등도 유료 업셀링 전략을 병행하고 있다. 또한 빅테크 기업들은 비용 절감을 위해 전용 AI 하드웨어 개발에 나섰다. 예컨대 MS는 “Athena”라는 자체 AI 반도체 프로젝트를 2019년부터 비밀리에 진행해왔으며, 2024년~2025년경 데이터센터에 적용해 엔비디아 GPU 의존을 줄이고 비용을 낮추려는 것으로 알려져 있다. 구글은 자사 TPU(Tensor Processing Unit)를 4세대까지 발전시켜 내부 AI 서비스에 활용 중이고, 아마존도 Inferentia, Trainium 등 맞춤형 칩을 AWS에 도입했다. 모델 경량화와 최적화 알고리즘 연구도 병행되어, 추론 비용을 낮추는 다양한 기법(모델 압축, 저정밀 연산 등)이 개발 중이다.

비용과 관련해 또 하나의 지속 가능성 이슈는 에너지 및 환경 부담이다. 거대 모델 훈련에는 수백~수천 메가와트시(MWh)의 전력이 소모되고 탄소 배출도 상당하다는 연구들이 있다. 다만 빅테크 기업들은 재생에너지 사용을 늘리고, 효율적 데이터센터 설계로 전력당 연산량을 높이는 등 친환경 측면도 신경 쓰고 있다. 결론적으로, 현시점 생성형 AI 기업들은 고비용 구조를 감수하며 시장을 개척하고 있으며, 향후 수익 모델 정착과 비용 효율화 기술 확보 여부가 이 산업의 장기 지속 가능성을 좌우할 것이다. 투자자들도 “AI 인프라 지출이 언제 매출 및 이익 성장으로 이어질지 분명히 제시되어야 한다”며 시간을 두고 지켜보는 분위기다.

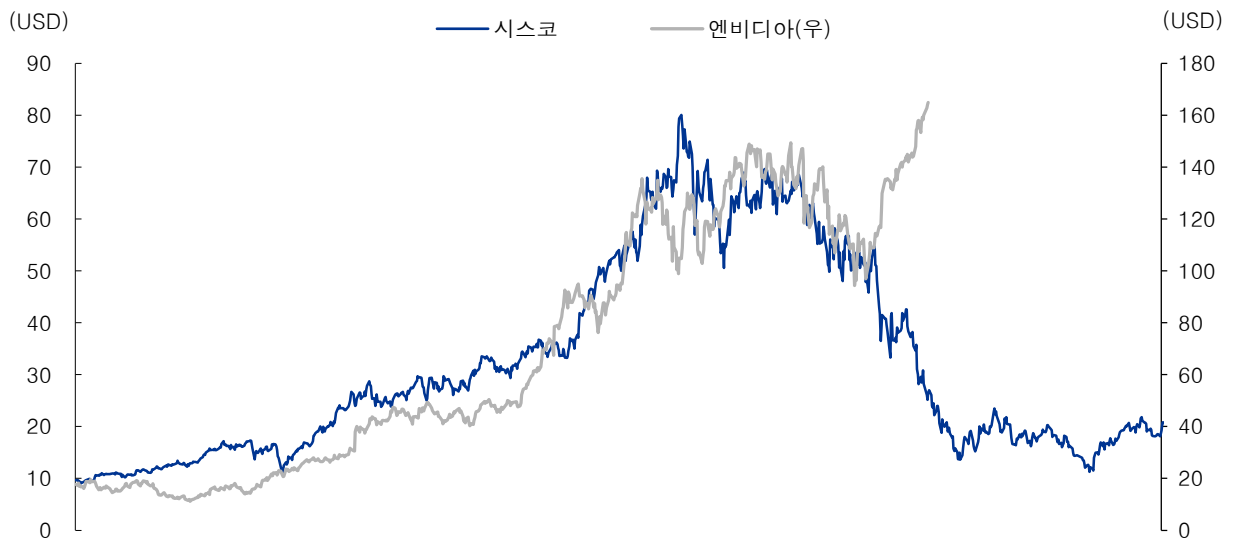
4) 2023~2025년 AI 기업 주가 동향: 과열 vs 실적, 버블 논쟁

AI 테마 랠리,
실적 대비 과열 정도 높지 않아

2023년에 들어 미국 증시는 “AI 열풍”에 힘입어 대형 기술주 중심의 랠리가 전개되었다. 2022년 10월 저점 대비 2024년 중반까지 S&P 500 지수는 50% 이상 상승했고, 기술주 중심의 NASDAQ 지수는 급등하며 사상 최고치를 갱신 중이다. 특히 AI 반도체 시장을 독점한 NVIDIA의 주가가 폭발적인 상승을 보이면서 4조 달러를 돌파하여 글로벌 시가총액 1위 기업이 되었다.

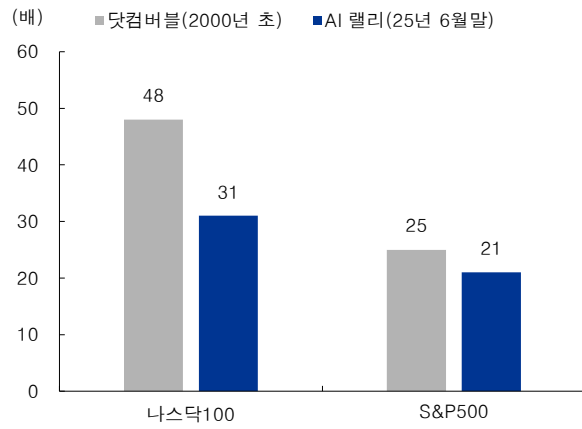
주요 AI 관련주의 높아진 밸류에이션은 버블 논쟁을 촉발한다. NASDAQ 100 지수의 12개월 선행 PER(주가수익비율)이 2025년 6월 말 약 31배로 상승하여, 2000년 버블 정점기(약 48배)에는 못 미치나 역사 평균 대비 높아졌다. S&P 500 전체의 PER도 21배로 장기 평균보다 높지만, 2000년 전후 25배보다는 아직 낮다. 특히 NVIDIA의 경우 2023년 말 PER이 40배 수준인데, 이는 Cisco가 2000년 3월 정점 당시 131배였던 것과 비교되며 “실적 대비 과열 정도는 당시보다 덜하다”는 분석도 있다. 실제로 이익 전망 측면에서 보면, 2023년과 2024년 현재 AI기업들의 EPS(주당순이익) 예상치가 시장 평균보다 빠르게 증가하고 있어, 1999년~2000년 당시 이익 없는 회사들까지 천문학적 주가를 받았던 상황과는 상이하다. 한편 과거 닷컴버블 당시 닷컴기업들이 우구죽순 상장했지만 AI관련 기업의 상장은 상대적으로 제한적이고 엄격한 기준 아래에서 진행된다.

그림 58. Cisco와 NVIDIA의 주가 추이 비교



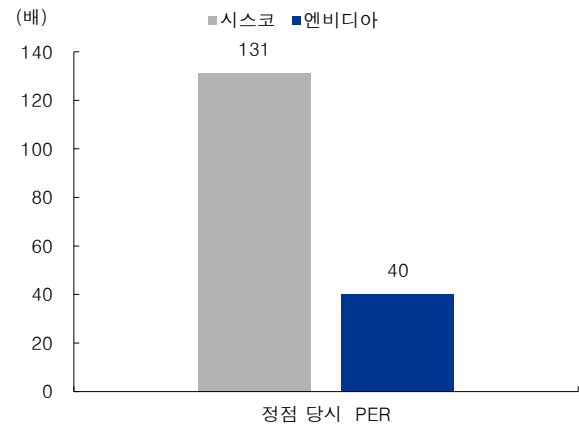
자료: Bloomberg, IBK투자증권

그림 59. 닷컴과 AI시기의 NASDAQ100, S&P500 밸류에이션 비교



자료: Bloomberg, IBK투자증권
주: 12개월 선행 PER

그림 60. 정점 당시 시스코와 엔비디아의 밸류에이션 비교



자료: Google, IBK투자증권
주: 12개월 선행 PER

5) 생성형 AI 스타트업 투자 규모와 VC들의 Exit 전략

생성형 AI 열풍,
버블이 꺼진다면 비상장부터

ChatGPT 등장 이후 전 세계 벤처 투자 자금이 생성형 AI 스타트업으로 쏠리는 현상이 두드러졌다. 2023년 한 해 생성형 AI 관련 스타트업 투자액은 전 세계적으로 약 291억 달러에 달했으며, 2024년에는 그 규모가 560억 달러로 2배 가까이 증가해 사상 최고치를 기록했다. 과거 닷컴 버블 시기에는 대규모 IPO(기업공개)를 통해 VC들이 단기간에 지분을 현금화하는 사례가 많았지만, 현재 생성형 AI 스타트업들은 아직 대부분 비상장 상태로 자금 조달을 받고 있다.

2022년~2023년 거시경제 여건상 IPO 창구가 좁았고, 2023년 후반부터 Arm, Klaviyo 등 일부 기업공개가 재개되며 2025년 이후 AI 유니콘들의 상장 러시 가능성도 거론된다. 그러나 현재 인프라, 인력 확보 과정에서 천문학적인 비용이 사용되기 때문에 비상장 AI기업들이 지속하기 어려운 환경이 심화될 것이다. 상장된 대형 빅테크 기업들은 AI이외 충분한 매출원이 존재한다. 그러나 비상장 기업들은 비즈니스를 구축하기 어렵기 때문에 적자폭이 커질 가능성이 높다. 따라서 AI산업의 버블 붕괴는 비상장 내 기업들을 중심으로 확산될 가능성이 높다.

표 22. 2024년 포브스(Forbes) 선정 유망 AI 스타트업

순위	스타트업	산업	누적 투자 금액
1	Sierra	고객 응대(CS)	\$110M
2	Harvey	법률	\$106M
3	Perplexity	리서치	\$102M
4	ElevenLabs	음성 합성(TTS)	\$101M
5	DeepL	번역	\$100M

자료: Visual Capitalist, IBK투자증권

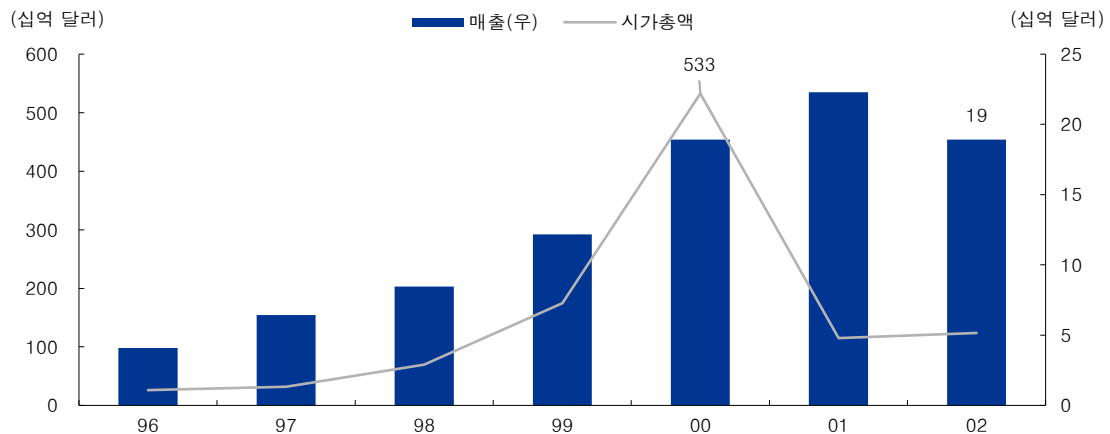
주: 2024년까지 누적 투자 금액

6) 닷컴 버블 시기 주요 기술기업의 주가 하락 원인

닷컴 버블, 시스코(Cisco)
매출 증가에도 주가 90% 급락

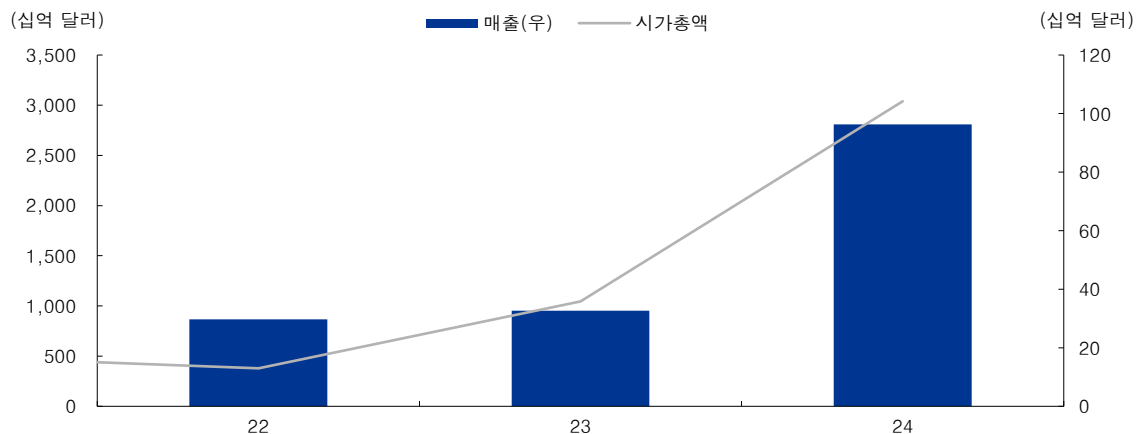
닷컴 버블(Dot-Com Bubble)은 1995년부터 2000년 초까지 인터넷 기반 신생 기업들에 투자 열풍이 불면서 주가가 비정상적으로 치솟았다가, 2000년 3월 정점을 찍고 2001년~2002년 급락한 사건이다. 이 시기 Cisco, Intel, Oracle, Dell과 같은 IT 핵심 기업들도 주가가 폭등했는데, 버블 붕괴 시에는 이들조차 예외 없이 주가가 80% 이상 폭락했다. Cisco Systems의 예를 보면 1995년~2000년 매출이 20억 달러 → 190억 달러로 850% 성장하는 호실적을 냈음에도, 주가는 그보다 훨씬 과열되어 5년간 3,800% 상승 후 시가총액 5,000억 달러로 세계 1위를 기록했다. 그러나 2000년 3월 이후 시장이 급격히 냉각되자 Cisco 주가는 2년 만에 79달러 → 9.50달러로 88% 폭락했다. 흥미로운 점은, Cisco의 실적이 붕괴한 것이 아니었다는 것이다. 버블 붕괴 후인 2001년 매출은 220억 달러로 오히려 증가했고 2002년에도 190억 달러로 2000년 수준을 유지했지만, 투자자들이 더 이상 미래에 높은 멀티플을 지불하지 않으면서 주가가 급락한 것이다. 결국 원인은 기대치의 붕괴였다. 이들 기업은 탄탄한 수익 구조를 가지고 있었지만, 주가에 반영된 미래 예상이 지나치게 낙관적이어서 성장률 둔화나 주문 감소 등 작은 실망에도 주가 폭락이 불가피했다.

그림 61. 시스코의 시가총액 상승과 연간 실적 상승 추이



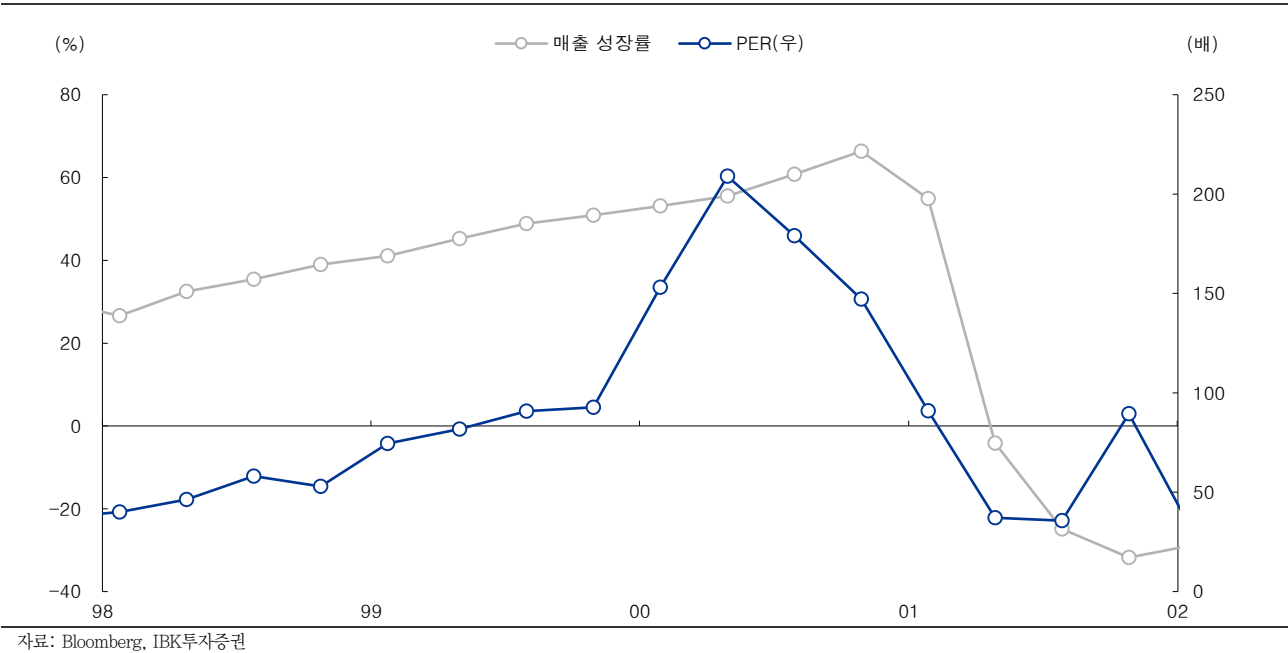
자료: Bloomberg, IBK투자증권

그림 62. 엔비디아의 시가총액 상승과 연간 실적 상승 추이



자료: Bloomberg, IBK투자증권

그림 63. 시스코의 매출 성장률과 밸류에이션 추이



기 현재 AI 산업의 차별점

AI 붐, 닷컴 버블과 달리
탄탄한 실적 기반 구조

현재 AI 산업은 2000년대 초의 닷컴 거품과 몇 가지 중요한 측면에서 다르다.

첫 번째, 실제 매출과 이익을 내는 핵심 플레이어 존재: 닷컴 버블 당시에는 시가총액 상위 기업 중에 이익을 내는 곳이 드물었고, 매출 대비 과대평가된 경우가 많았다. 그러나 현재 AI 붐을 이끄는 NVIDIA, MS, 구글, 아마존 등은 이미 크고 안정적인 수익원을 가진 기업들이다. 즉 시장의 주도주들이 맹목적인 꿈의 주식이 아닌, 실적 뒷받침이 있는 기업들이라는 점은 안정 요소다.

두 번째, 기술의 성숙도: 2000년 전후 인터넷은 아직 미성숙한 기술(느린 접속, 제한된 콘텐츠)이었다. 반면 AI는 딥러닝 기술이 10여 년 연구를 거쳐 2020년대 중반 상당한 성능 임계점을 넘은 상태다. 또한 클라우드 인프라, GPU 등 뒷받침 기술 생태계도 성숙하여 서비스 제공 비용이 시간이 갈수록 빠르게 감소하는 추세다.

세 번째, 상장사 구성의 차이: 닷컴 버블은 무수히 많은 신생 기업들이 IPO에 나서면서 공개 시장에서 거품이 형성되었다. 그러나 AI 붐에서는 대부분의 신생 기업들이 아직 비상장이며, 주식 시장에서는 빅테크 위주로 테마가 형성되었다. 상장 진입장벽이 높아진 덕에 투자자 보호가 일부 되고 있는 셈이다. 물론 이에 따라 거품이 끼는 범위가 제한되고, 버블 붕괴 시 파급도 비교적 국한될 수 있다.

8 결론 및 전망

단기 조정 후,
에이전틱 및 피지컬 AI 중심
신시장 도래 전망

단기(1~2년): AI 산업 내 조정 국면 가능성이 높다. 경기 여건과 금리, 투자 심리에 따라 일부 거품 종목 급락 및 스타트업 중심으로 구조조정이 일어날 수 있다. 인프라 투자가 증가하면서 실적이 급성장한 종목은 실적 성장이 시장의 기대치를 하회하는 순간에 접근하고 있는 것으로 판단된다. 상장되어 있는 주요 AI 기업은 밸류에이션이 하락할 수 있으나 닷컴 버블과 같은 폭락 가능성은 적을 것이다.

중장기: 생존한 AI 선도 기업들이 시장을 주도하고, 거품이 꺼진 자리에 건강한 성장 생태계가 정착할 전망이다. 닷컴 시기 인터넷의 확산, 모바일 시기 스마트폰의 보급처럼 기존 시장과 단절된 신규 시장을 구성할 수 있는 Agentic, Physical AI가 폭발적으로 확산될 수 있다. 이 시기는 기존 빅테크 이외의 신규 기업이 시장의 주도권을 잡을 것이다. 현재 예상할 수 없는 새롭고 거대한 AI 산업이 나타날 것으로 전망된다.

표 23. 닷컴 버블(1999~2001)과 현재 AI 붐(2023~2025)의 주요 특징

비교 항목	1999~2001 닷컴 버블	2023~2025 AI 붐
투자 열풍 양상	.com 붐은 IPO 500여 개, 첫날 주가 3~4배 ↑ 빈발; VC 자금 40% 인터넷으로	OpenAI 등 유니콘 등장; VC 자금 2024년 \$560억 (2배 ↑); AI 언급 기업 주가 급등 (테마 편승)
수익 구조/모델	상당수 기업 매출 · 이익 부재, 사용자는 있으나 수익화 실패	빅테크는 AI 즉각 수익 적으나 본업 이익 견조; 스타트업도 API 판매, 구독 모델 등 시도 (아직 초기)
거품 원인	저금리 자금 과잉, 투기/과신; "선점하면 이익 따라온다" 논리; 광고 등 과지출	금리인하 등 유동성이 다시 증가. 혁신 과대평가 가능성; "AI 패권 잡아야" 경쟁 심리; 인제 · GPU 팽탈 비용 부담
거품 붕괴 촉발	금리 인상 + 실적 미스에 투심 급변 (내재가치 의심 폭발); 자금 경색으로 출도산	일각 버블 경고 증대; 수익 부진하거나 경제 둔화 시 주가 조정 가능성

자료: Google, 언론 보도 종합, IBK투자증권

7. 2025년 대한민국의 AI정책

새 정부, AI 강국 도약 준비 중

정책-민간 연결로
소버린 AI 도약

한국은 이미 강력한 정책 드라이브와 민간의 뛰어난 기술 역량을 확보했다.

현재 한국은 ‘소버린 AI’ 개발을 계기로, 국가 차원의 AI 전략을 전면 재편하고 있다. 실제로 새 정부는 한국어와 국내 데이터를 기반으로 한 자체 언어모델 개발을 추진 중이며, AI 반도체와 고성능 GPU 인프라 확보, 산업별 AI 적용 확산 등을 병행하고 있다.

정책-인프라-시장 간 유기적 연계를 통한 국가 전략의 진화를 의미한다.

표 24. 새 정부는 AI 도약을 위해 100조 원 투자 계획을 밝히며 민간과 협력 선언

정책	공약	실제 진행 현황(25/7/14 기준)	목표 시기
소버린 AI 개발	한국어자체 데이터 기반 국가 주도 AI 모델 개발	과기정통부 주도 R&D 로드맵 수립 단계 예산 2조원 내외 책정 논의 중	2027년
AI 예산 증액, 민간 투자 유도	AI 예산 증액 민간 투자 100조 원	국책은행 등 협력형 AI 펀드 조성 계획 발표 (10조원 규모)	2026년
데이터센터 건설	'AI 고속도로' 구축	SKAWS가 울산에 100MW급 AI 슈퍼 데이터센터 공동 건설 착수 (7조 원 규모)	2027년
GPU 확보	최신 GPU 최소 5만 개 이상 확보	국가 AI 컴퓨팅센터에 고성능 GPU 1만 대 확보 추진 중 추가 조달 계획 진행 중	2025년 말까지 1만 대 확보 목표
'모두의 AI' 프로젝트	범용 AI 확산, 민간 협력형 프로젝트	디지털 인재 양성 및 중소기업 대상 AI 솔루션 확산 전략 반영	-
AI 대표기업 육성	국가대표 AI 기업 발굴 및 투자	관련 기업 발굴 및 지원 위한 민관 합동 프로그램 논의 중	-
정부 주도 데이터 클러스터	정부 직접 데이터 클러스터 조성	AI 컴퓨팅 센터 구축에 2조 원 투입 예정 고성능 연산 인프라 확보 계획 포함	2026년 초
대통령 직속 AI 정책 조직 신설	대통령실 산하 'AI정책책임수석' 신설	2025년 6월 신설 완료	2025.6 완료
인공지능 기본법 조기 제정	AI 법제 정비, 위험기술 규제, 데이터 활용 기준 마련	'AI 기본법' 국회 본회의 통과	시행시기 조율 중
중기부 전면 개편	중기부 디지털 역량 강화 및 AI 산업 지원 구조 개편	사디지털 전문가 장관 후보 지명 혁신 펀드·지원책 마련 시사	2025년 하반기

자료: 언론 보도 종합, IBK투자증권

1) AI 뉴딜 프로젝트

AI 뉴딜,
데이터·인재 생태계 강화

정부는 AI 생태계의 핵심 자원인 양질의 데이터 확보와 혁신 인재 양성에 적극적으로 나서고 있으며, 세계적 수준의 AI 개발을 위한 민간의 노력을 집중 지원한다. 더 나아가 범용 AI 모델을 모든 국민이 일상에서 이용할 수 있도록 보급을 확산하고, 산업별로 특화된 AI 개발도 지원하여 국민이 생활 곳곳에서 AI 혜택을 누릴 수 있도록 할 것이다. 이재명 정부는 AI에 100조 원 규모의 투자를 목표로 하고 있으며, 이는 공공·산업·국민 생활 전반의 AI 확산과 지역 균형, 포용을 위한 'AI 뉴딜'을 포함하는 '대한민국 AI 국가전략 3.0'의 핵심이다.

한국판 뉴딜 프로젝트의 3대 과제 중 하나인 '디지털 인프라 구축'은 AI 학습과 개발의 기본이 되는 '데이터 전 주기 인프라 강화'와 '데이터 수집 활용 확대'를 포함한다. AI 시대의 핵심 기반 시설인 데이터센터 구축과 관련 네트워크 강화에 집중하고 있다. 또한, AI 기술이 특정 산업에만 머무르지 않고 '전 산업으로 AI 융합 확산'되도록 지원하는 정책 방향도 분명히 드러난다. AI 서비스가 원활하게 작동하고 대용량 데이터를 빠르게 전송하기 위해 필수적인 '5세대(5G) 인프라 조기 구축'과 이를 바탕으로 하는 '5G+ 융복합 사업 촉진' 역시 AI 생태계를 지원하려는 중요한 과제로 제시하고 있다.

그림 64. 한국형 뉴딜 정책

3대 프로젝트	10대 중점 과제
디지털 인프라 구축	•데이터 전 주기 인프라 강화 •데이터 수집 활용 확대 •인공지능(AI) 데이터 인프라 확충 •전 산업으로 AI 융합 확산 •5세대(5G) 인프라 조기 구축 •5G+ 융복합 사업 촉진
비대면 산업 육성	•비대면 서비스 확산 기반 조성 •클라우드 및 사이버 안전망 강화
산업간접자본(SOC) 디지털화	•노후 국가기반시설 디지털화 •디지털 물류 서비스 체계 구축

자료: 중앙일보, IBK투자증권

2) K-AI: 독자 AI 파운데이션 모델 프로젝트

K-AI, LLM 너머
LAM까지 기술 확장

K-AI는 대한민국이 인공지능 분야에서 기술 주권을 확보하고 글로벌 AI 강국으로 도약하기 위한 핵심 전략이다.

1. 개발 전략: 실력 있는 정예팀과 경쟁 기반의 DARPA형 추진

실력 있는 AI 정예팀을 중심으로 민간 협력을 통해 추진하는 것을 핵심으로 한다. 특히 'DARPA형' 추진 방식을 채택하여, 정부가 명확한 목표를 제시하면 참여팀이 자율적으로 개발 전략을 구현하는 방식이다. 또한, 6개월 단위로 경쟁형 평가를 통해 참여팀을 선별하고, '무빙 타겟(Moving Target)' 방식으로 목표를 유연하게 조정하며 최적의 결과물을 도출하는 고도화된 방식을 적용한다.

2. 기술 확장: LLM에서 LAM으로, 오픈소스 지향 및 글로벌 진출

초기에는 언어 기반의 LLM(Large Language Model)에 집중하지만, 점차 멀티모달(다중 양식)을 지원하는 LMM(Large Multimodal Model), 그리고 나아가 행동(액션)까지 처리하는 LAM(Large Action Model)으로 고도화하거나 병행 개발을 지향한다. 이는 AI 모델이 단순히 정보를 이해하는 것을 넘어, 실제 환경에서 복잡한 작업을 수행할 수 있도록 발전시키겠다는 비전이다.

3. 활용 확산: 국민의 AI 접근성 증진 및 사회적 편익 확보

'대국민 사용 평가제도(User Based Review)'를 운영하여 사용자 피드백을 적극 반영하고, '대국민 콘테스트(Contest)' 등을 통해 국민들에게 정례적으로 프로젝트를 홍보할 예정이다. 궁극적으로는 AI 발전의 사회적 편익을 확보하여 국민 AI 접근성을 높이는 데 기여하고자 한다. 더불어 오픈소스 지향을 통해 민간의 다양한 AI 서비스 출시를 유도하며, AI 대전환 기반의 경제 성장을 목표로 공공·경제·사회 AX(AI 전환)를 촉진할 계획이다.

그림 65. 과기정보통신부의 '독자 AI 파운데이션 모델 프로젝트 (K-AI)'

2 추진 방향



자료: 과기정보통신부, IBK투자증권

8. 21세기 AI 기술 진화

1) 생성형 AI 기술 진화

딥러닝의 부활과
LLM 중심 AI 혁신 가속화

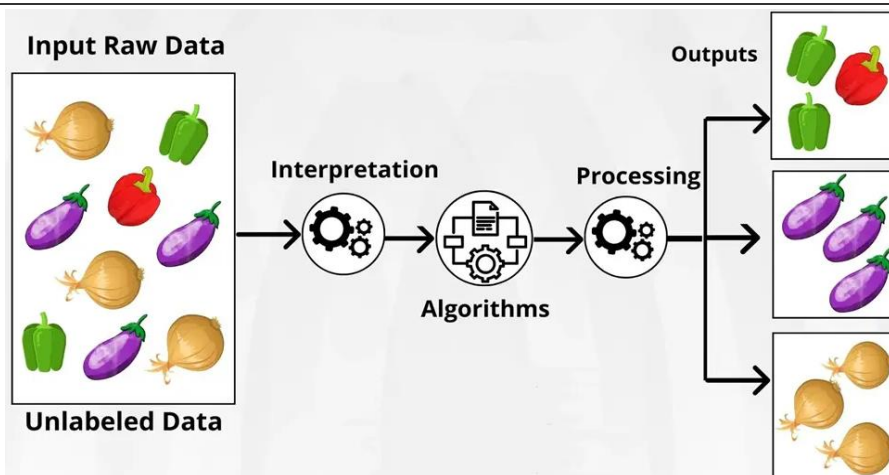
21세기 들어 인공지능(AI)은 인간의 언어, 시각, 행동을 이해하고 모방하는 기술로 발전해왔다. 특히 2017년 Transformer 구조가 등장한 이후, 대형언어모델(LLM: Large Language Model) 기반의 AI는 새로운 혁신의 중심에 서게 되었다. GPT, BERT, PaLM, LLaMA, HyperCLOVA 등 모델들은 언어 생성, 검색, 코딩, 번역, 대화 등 광범위한 영역을 빠르게 혁신하고 있으며, 이 흐름을 주도하는 주체는 미국의 OpenAI, Google, Meta 등이며 한국에서는 NAVER가 글로벌 경쟁력을 확보했다.

2006년 생성형AI 발전의 초석을 마련

2006년 이전까지 인공지능 연구 분야에서 '신경망(Neural Network)'이라는 개념은 한동안 큰 주목을 받지 못하고 있었다. 하지만 '딥러닝의 대부'라고 불리는 캐나다 토론토 대학교의 제프리 힌튼 교수는 2006년에 중요한 논문인 "A Fast Learning Algorithm for Deep Belief Nets"를 발표하며, '딥러닝(Deep learning)'이라는 개념을 다시금 조명했다. 힌튼 교수는 신경망을 효율적으로 학습시키는 새로운 방법(비지도 사전학습 방식)을 제시했다. '비지도 사전학습 방식'이란, 정답이 따로 주어지지 않은 방대한 양의 데이터를 신경망이 스스로 분석하고 패턴을 찾아내도록 먼저 학습시키는 것을 의미한다.

2012년 알렉스넷(AlexNet)의 이미지넷(ImageNet) 대회 우승은 컴퓨터가 이미지를 훨씬 더 정확하게 인식하고 분류할 수 있게 해주는 'CNN(Convolutional Neural Network)' 기술이 크게 발전하는 전환점이 되었다. CNN은 이미지에서 눈, 코, 입, 귀 같은 특징들을 자동으로 찾아내고 학습하는 데 특화된 인공지능 신경망이다. 기존에는 사람이 일일이 특징을 정의하고 컴퓨터에 알려줘야 했지만, CNN은 이 과정을 스스로 학습하기 때문에 방대한 이미지 속에서 더 정교하고 복잡한 패턴을 발견하여 인식 성능을 극적으로 끌어올릴 수 있었다.

그림 66. 비지도 사전학습 방식



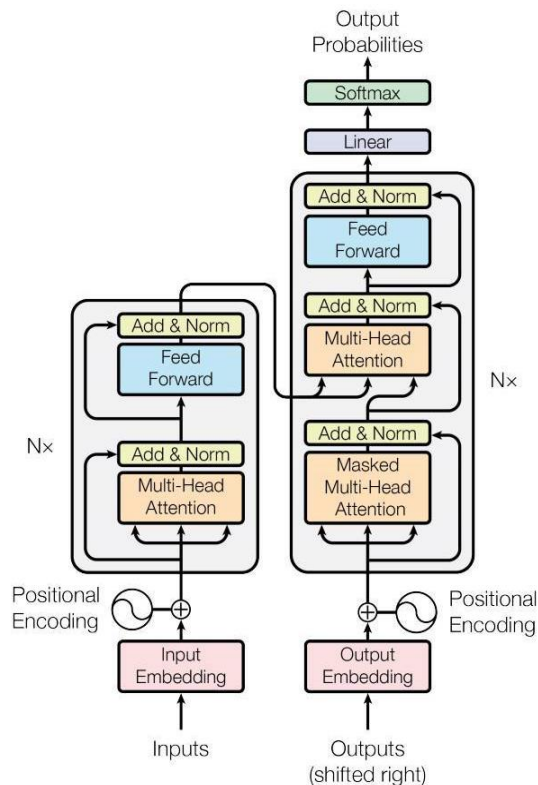
자료: 언론 보도 종합, IBK투자증권

GAN·Transformer,
생성형 AI 기술의 출발점

생성형 AI의 초석은 2014년 GAN(Generative Adversarial Network)의 제안으로 다져졌으며, 이미지 생성 기술의 발전에 크게 기여했다. GAN은 '생성자(Generator)'와 '판별자(Discriminator)'라는 두 개의 인공지능이 서로 경쟁하며 학습하는 독특한 구조를 가지고 있다. 생성자는 가짜 이미지를 만들고(마치 위조범처럼), 판별자는 이 이미지가 진짜인지 가짜인지 구별하는 역할(마치 경찰처럼)을 한다. 생성자는 판별자조차 진짜와 구분하기 어려운 매우 사실적인 가짜 이미지를 만들어낼 수 있게 된다. 이를 통해 존재하지 않는 사람의 얼굴, 새로운 스타일의 그림 등 다양한 고품질 이미지를 생성하는 기술이 폭발적으로 발전하게 되었다.

LLM 시대를 연 기술적 기초는 2017년 구글에서 발표된 'Transformer' 논문으로 마련되었다. 이 논문의 제목은 "Attention Is All You Need"였다. 트랜스포머가 등장하기 전에는 주로 '순환 신경망(RNN)'이라는 기술이 언어 처리에 사용되었는데, 문장을 순서대로 처리하기 때문에 긴 문장에서는 앞부분의 정보를 잊어버리거나 병렬 처리가 어려워 학습 속도가 느렸다. 하지만 트랜스포머는 '어텐션(Attention)'이라는 혁신적인 메커니즘을 도입하여 이 문제를 해결했다. 어텐션은 문장 속의 어떤 단어가 다른 단어와 얼마나 관련이 깊은지를 스스로 파악하는 능력이다. 예를 들어, "사과를 먹고 기분이 좋다"라는 문장에서 '사과'와 '먹고'의 관련성, 그리고 '사과'와 '기분'의 관련성을 동시에 파악할 수 있게 된 것이다. 덕분에 트랜스포머는 문장 전체의 맥락을 한 번에 이해하고, 병렬 처리가 가능해져 학습 속도와 효율성을 크게 높였다. 이러한 트랜스포머 구조는 이후 GPT, BERT, PaLM, LLaMA 등 현재 대형 언어 모델(LLM)들의 핵심 기반이 되었다. 이듬해인 2018년 OpenAI GPT-1과 Google BERT가 발표되면서 본격적인 자연어 처리 모델 경쟁이 시작되었다. 2020년에는 GPT-3가 대규모 퓨샷(few-shot) 학습 시연을 통해 AI의 상용화 가능성을 크게 부각시켰다.

그림 67. 2017년 구글의 Transformer 원리



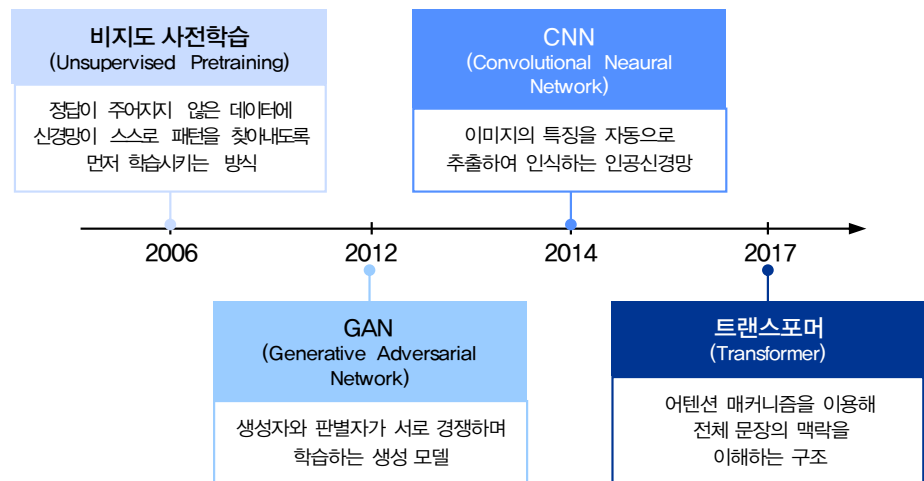
자료: 언론 보도 종합, IBK투자증권

표 25. 2000년 이후 인공지능 관련 주요 사건

연도	주요 사건	세부 내용
2006	Geoffrey Hinton, 딥러닝 재조명	논문: A Fast Learning Algorithm for Deep Belief Nets → 신경망 기반 학습 부활
2012	AlexNet, ImageNet 대회 우승	CNN 기반 이미지 인식 성능 혁신적 향상 → 딥러닝 붐 확산
2014	GAN 제안	Generative Adversarial Network → 생성 AI 기술의 초석
2017	Transformer 논문 발표	Attention Is All You Need → LLM 시대의 기술 기반 확립
2018	GPT-1, BERT 발표	OpenAI: GPT-1 / Google: BERT → NLP 분야 경쟁 본격화
2020	GPT-3 공개	1750억 파라미터 기반, few-shot 학습 성능으로 상용화 가능성 부각
2022	ChatGPT 출시	사용자 1억 명 돌파 → 생성형 AI의 대중화 전환점
2023	Gemini, LLaMA 2 · 3 발표	Google: Gemini / Meta: LLaMA 2, 3 → 멀티모달, 오픈소스 기반 경쟁 심화

자료: IBK투자증권

그림 68. 생성형 AI의 초석이 된 AI기술



자료: IBK투자증권

2) 글로벌 빅테크 및 NAVER의 생성형AI 진화 및 전략

표 26. 2018년 이후 구글, OpenAI, Meta 및 NAVER의 생성형 AI 관련 연대기

연도	회사	모델/서비스	특징	사양/기술
2018	Google	BERT	양방향 인코더 기반, 문맥 이해 강화	3.4억 파라미터, 인코더 구조
2018	OpenAI	GPT-1	최초의 GPT, 비지도 사전학습	1.17억, BookCorpus
2019	OpenAI	GPT-2	본격 언어 생성, 파라미터 10배 ↑	15억, WebText
2019	Meta	RoBERTa	BERT 개선, 성능 최적화	BERT 동일 구조, 마스크 방식 변경
2020	OpenAI	GPT-3	Few-shot/Zero-shot 강화, 상용 API 시작	1,750억, 다중 도메인 데이터
2020	Meta	BlenderBot 1.0	대화형 모델, 장기기억+외부문서 통합	약 9.4억, 오픈도메인 대화
2021	Google	LaMDA	대화 특화 언어모델	디코더 구조, 맥락 이해 대화 강화
2021	NAVER	HyperCLOVA	한국어 GPT-3 계열, 상업용	2,040억 파라미터, 한국어 중심
2021	OpenAI	InstructGPT	RLHF 도입, 사용자 지시 이해	GPT-3 기반, 피드백 최적화
2022	Google	PaLM	초거대 모델, 디코더 전용 구조	5400억, Pathways 구조
2022	Meta	OPT	GPT-3 대응 공개 모델	66억~175억, 오픈소스
2022	OpenAI	GPT-3.5 / ChatGPT	GPT-3 기반 대화 최적화, 대중화	RLHF 기반 튜닝, ChatGPT 서비스
2022	DeepMind	Chinchilla	compute 최적 모델, 학습량 강조	70억 파라미터, 1.4조 토큰
2023.2	Meta	LLaMA 1	경량화 + 고성능, 연구 목적 공개	7B~65B, 오픈소스 (비상업용)
2023.3	Google	Bard (LaMDA 기반)	ChatGPT 대응 챗봇, 첫 공개	LaMDA Lite 기반, 검색 연동
2023.3	OpenAI	GPT-4	멀티모달 입력 (텍스트+이미지)	약 1T+, 128K 토큰 컨텍스트
2023.5	NAVER	HyperCLOVA X	한국어 reasoning 강화, 서비스형 전환	다국어 6조 토큰 학습, CLOVA X/Cue: 출시
2023.7	Meta + MS	LLaMA 2	상업 이용 가능 오픈소스, Azure 연동	7B/13B/70B, RLHF 기반
2023.12	Google	Gemini 1.0	본격 멀티모달, PaLM → Gemini 전환	최대 100만 토큰, 텍스트 · 이미지 · 코드 · 비디오
2024.4	Meta	LLaMA 3 (8B/70B)	GPT-4급 성능, 오픈소스 상업화	frontier-level 성능, 공개
2024.5	OpenAI	GPT-4o	음성+비전 실시간, 저지연 멀티모달	오디오 직접 처리, 1M+ 컨텍스트
2024.6	NAVER	CLOVA X Tools	플러그인 · 검색툴 · 토크링 기능 탑재	외부 시스템 통합 실행형 AI
2025.1	Google	Gemini 2.5 Pro	성능 개선 및 SGE 검색 연계 강화	추론 · 시뮬레이션 고도화, Duet · Gmail 탑재
2025.4	Meta	LLaMA 3.1 (405B)	최초 frontier급 오픈소스, AGI급 성능	4050억 파라미터, 완전 공개 예정

자료: IBK투자증권

공통점

빅테크 공통점:
LLM 개발, 멀티모달,
서비스 연계, 인프라 투자,
API 및 플랫폼

초거대 언어 모델(LLM) 개발 및 경쟁: 네 회사 모두 수십억에서 수천억 파라미터에 이르는 자체 LLM을 개발하고 있으며, 이를 통해 AI 기술 경쟁에서 우위를 점하려 했다. GPT 시리즈, Gemini, LLaMA, HyperCLOVA 등이 대표적인 예시이다.

멀티모달 AI로의 확장: 초기 텍스트 기반 모델에서 발전하여, 이미지, 오디오, 비디오 등 다양한 형태의 데이터를 이해하고 생성하는 멀티모달 AI로의 확장을 공통적으로 추구한다.

자사 서비스와의 연계 및 통합: 개발된 AI 모델을 단순히 독립적인 서비스로 제공하는 것을 넘어, 각 회사의 핵심 서비스(검색, 광고, 클라우드, 쇼핑 등)에 깊이 통합하여 기존 서비스의 기능과 사용자 경험을 혁신하는 데 활용한다.

AI 인프라 투자 확대: 대규모 모델 학습과 서비스 운영을 위한 고성능 컴퓨팅 인프라(GPU, 슈퍼컴퓨터) 구축 및 확장에 막대한 투자를 단행하고 있다.

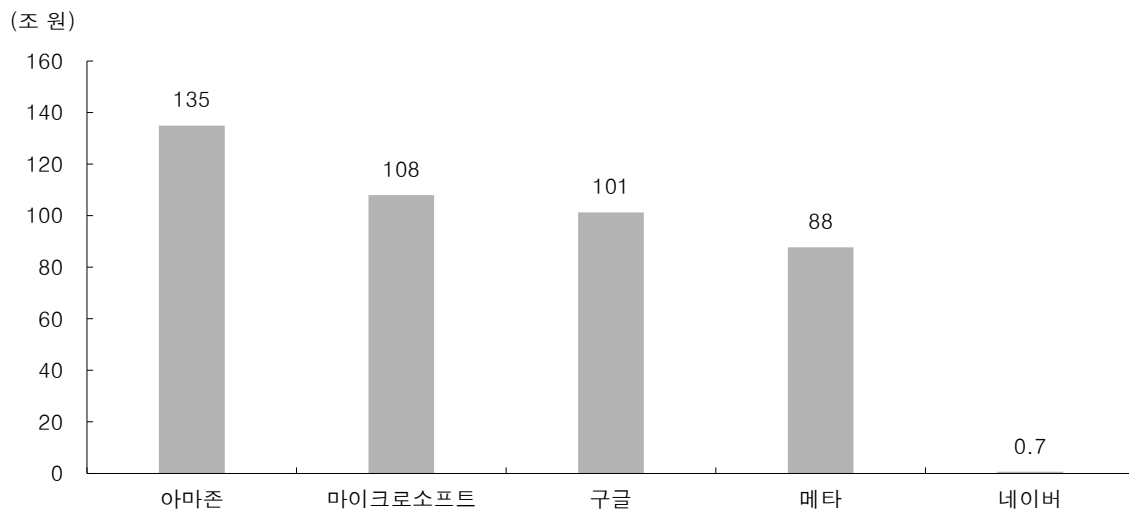
API 및 플랫폼 전략: 개발자들이나 기업이 자신들의 AI 모델을 활용할 수 있도록 API를 제공하거나 클라우드 기반의 AI 플랫폼을 구축하여 생태계를 확장하려 한다.

표 27. 빅테크 기업의 LLM 모델 및 서비스 개선 과정의 공통점

공통 전략 항목	설명
1. 초거대 언어모델 경쟁	GPT 시리즈(OpenAI), Gemini(Google), LLaMA(Meta), HyperCLOVA(NAVER) 등 수십억~수천억 파라미터 규모의 LLM 개발을 통해 기술 경쟁 주도
2. 멀티모달 AI 확장	텍스트 중심에서 이미지 · 오디오 · 비디오 등 복합 입력을 처리하는 멀티모달 AI로 진화 중 (GPT-4o, Gemini 1/2, LLaMA 3.1 등)
3. 자사 서비스 통합	AI 모델을 검색(Google, NAVER), 광고(Google, Meta), 클라우드(OpenAI, Google), 커머스(NAVER) 등 핵심 서비스에 내재화
4. AI 인프라 투자 확대	GPU, 슈퍼컴퓨터, 데이터센터 등 고성능 인프라 구축에 수십~수백억 달러 투자 (OpenAI Stargate, Google TPU, NAVER GAK 세종 등)
5. API 및 플랫폼 전략	ChatGPT API, Gemini API, Azure + LLaMA API, NAVER Cloud API 등 클라우드 기반 개발자 생태계 확대 전략 추진

자료: IBK투자증권

그림 69. 기업별 AI투자 비용 추정



자료: IBK투자증권, 주: 원/달러 환율 1,350원 기준으로 환산

차이점

오픈소스 정책:

빅테크 차이점:
AI 전략, 오픈소스, 수익화 방식

Meta: 가장 적극적인 오픈소스 전략을 펼친다. LLaMA 시리즈를 통해 모델 가중치와 코드를 공개하여 학계 및 개발자 커뮤니티의 참여와 혁신을 유도하며, 오픈소스 생태계의 주도권을 확보하려 한다.

Google & OpenAI: 기본적으로 폐쇄형 모델 전략을 취한다. 자체 개발한 최상위 모델(Gemini, GPT-4)은 직접 서비스나 API를 통해서만 접근을 허용하며, 기술의 통제와 상업적 활용에 중점을 둔다.

NAVER: HyperCLOVA X는 내부 서비스 통합 및 B2B 플랫폼 중심의 폐쇄형 모델을 지향하지만, 한국어 특화 모델의 강점을 살려 국내 생태계 협력을 모색한다.

사업 모델 및 수익화:

OpenAI: ChatGPT Plus 구독 모델과 API 판매를 통해 소비자 직접 서비스 및 개발자 중심의 수익 모델을 빠르게 구축했다.

Google: 검색, 광고, 클라우드 등 기존 핵심 사업에 AI를 심층 통합하여 경쟁력을 강화하고, 이를 통해 새로운 수익원을 창출한다. Bard/Gemini와 같은 챗봇은 자사 서비스 연동을 통한 사용자 경험 개선에 중점을 둔다. 구글 클라우드와의 시너지 매출 상승 중이다.

Meta: LLaMA와 같은 모델을 오픈소스로 공개하여 생태계 확장에 중점을 두며 현재 AI기술은 신규 매출원을 창출하는 것보다는 업무효율화를 통한 비용 통제에 기여한다. 장기적으로 자사 플랫폼 내의 AI 서비스(스마트 글래스, 광고 자동화 등)를 고도화하고 새로운 시장을 창출하려는 전략을 구사한다.

NAVER: 한국어/한국 문화에 최적화된 모델을 기반으로 자사 서비스(검색, 쇼핑, 번역 등) 고도화와 함께 B2B 중심의 AI 플랫폼(NAVER Cloud) 사업을 추진하며 국내 시장 및 아시아 시장을 공략한다.

표 28. 차이점

기업	오픈소스 정책	전략 요약	사업 모델 및 수익화	전략 요약
Meta	가장 적극적	LLaMA 1~3 모델 가중치 및 코드 공개. 연구 · 상업용 모두 허용. 오픈소스 생태계 주도	플랫폼 강화 업무 효율화	오픈소스를 통한 생태계 확대 → Meta AI 기반 서비스(광고, 스마트글래스 등) 내재화 및 장기적 수익 창출
Google	폐쇄형 중심	Gemini 등 최상위 모델은 폐쇄형. 일부 연구 모델만 제한 공개. API 중심 접근 허용	기존 사업 통합형 클라우드 시너지	검색 · 광고 · 클라우드에 LLM 심층 통합 → 사용자 경험 개선 및 기존 서비스 매출 확대
OpenAI	완전 폐쇄형	GPT-4/4o는 비공개. 사용은 ChatGPT · API로 제한. 기술 통제 및 상업 보호 중시	직접 수익 모델	ChatGPT Plus 구독, GPT/DALL · E/Whisper API 판매로 빠른 매출 확보, 소비자+기업 대상 이중 수익 구조
NAVER	제한적 폐쇄형	HyperCLOVA X는 비공개. 내부 서비스 · B2B API 중심. 국내 파트너십 확대 모색	B2B 중심 수익 모델	검색 · 쇼핑 · 번역 등 자사 서비스 고도화 + NAVER Cloud 기반 B2B API 판매로 국내 및 아시아 공략

자료: IBK투자증권

3) LLM 기술 진화 계보: 기업별 타임라인, 모델, 전략

Google

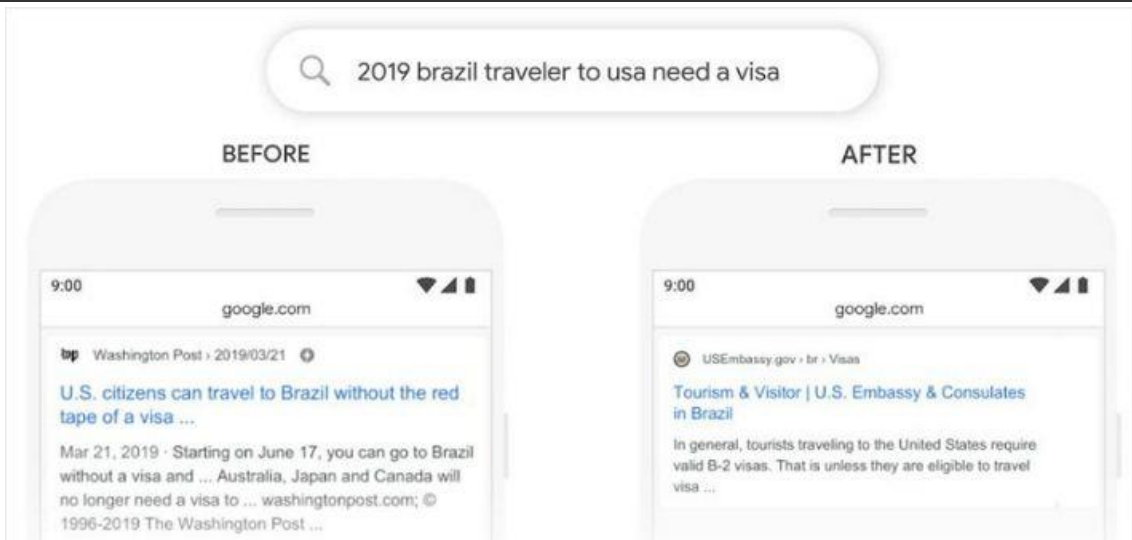
구글, BERT 기반
LLM 연구 본격화

BERT (2018): 구글은 양방향 인코더 구조의 트랜스포머 모델인 BERT(Bidirectional Encoder Representations from Transformers)를 공개하며 자연어 이해 분야에서 당시 최고 수준의 성능을 달성했다. BERT는 문맥을 양방향으로 파악하여 단어의 의미를 더욱 정확하게 이해한다. 이전의 언어 모델들은 보통 문장을 왼쪽에서 오른쪽으로, 즉 '단방향'으로만 읽으며 다음 단어를 예측하는 방식이었다. 문장의 앞뒤 전체를 함께 고려하여 이해하는 '양방향' 방식 덕분에, 단어의 중의성을 해결하고 문맥을 훨씬 더 정확하게 파악할 수 있게 되었고, 이는 구글 검색 엔진에 적용되어 검색 품질을 크게 향상시키는 데 기여했다. Transformer 계열 연구 확장

(2019~2020): 2019년부터 2020년까지는 T5, Meena 등 텍스트 생성 및 대화형 LLM 연구를 꾸준히 확장하며 다양한 언어 처리 과제에 대한 접근법을 모색했다.

그림 70. BERT 도입 이후 구글 검색의 전후 변화

(전: 키워드의 유사성만 판단 후 뉴스 노출 / 후: 검색 쿼리의 의도를 파악 후 미국 대사관 노출)



자료: IBK투자증권

LaMDA, 자연스러운
대화 흐름에 초점

LaMDA (2021): 구글은 대화형 AI 기술에 집중하여 2021년 LaMDA(Language Model for Dialogue Applications)를 공개하며 사람처럼 유창하게 대화를 이어가는 능력을 선보였다. BERT는 언어의 이해, GPT 언어 생성에 중점을 두었으나 LaMDA는 합친 대화를 중점으로 개선한 언어모델이다. LaMDA는 마치 사람과 대화하는 것처럼 자연스럽게 맥락에 맞는 답변을 생성하는 데 특화된 언어 모델이다. 일반적인 언어 모델이 단순히 정보를 전달하는 데 중점을 둔다면, LaMDA는 대화의 '흐름'을 이해하고 질문의 의도를 파악하여 실제 대화에서 나타나는 개방적이고 자유로운 질문에도 자연스럽게 반응하도록 설계되었다. 예를 들어, "달에 대해 이야기해 줘"라고 물으면 달의 특성뿐 아니라 "달에 가보고 싶니?"와 같은 후속 질문을 유도하거나, "달의 표면은 어때?"라는 추가 질문에 대해 이전에 언급했던 '달'의 맥락을 유지하며 답변하는 식이다.

PaLM (2022): 2022년에는 구글의 초거대 언어 모델인 PaLM(Pathways Language Model)을 발표했다. PaLM은 5,400억 개라는 엄청난 수의 '파라미터'를 지닌 모델이다. PaLM은 특히 '디코더 전용(Decoder-Only) 트랜스포머' 구조를 가지고 있는데, 이는 마치 문장을 처음부터 끝까지 자연스럽게 '생성'하는 데 특화된 인공지능 모델이라고 이해할 수 있다. 사용자에게 질문이나 지시가 주어지면, PaLM은 그 뒤에 이어질 가장 적절하고 논리적인 답변이나 내용을 창의적으로 만들어내는 데 뛰어난 성능을 보인다. '디코더 전용'이라는 것은 과거의 트랜스포머 모델이 문장을 이해하고(인코더) 생성하는(디코더) 두 부분을 모두 사용했던 것과 달리, 생성 부분에만 집중하여 더 효율적이고 강력한 텍스트 생성 능력을 갖추게 되었다는 의미이다.

표 29. BERT, GPT, LaMDA의 비교

항목	BERT	GPT	LaMDA
개발사	Google	OpenAI	Google
구조	Transformer Encoder	Transformer Decoder	Transformer Decoder
방향성	양방향(Bidirectional)	단방향(Left-to-right)	단방향
주요 목적	이해 (NLU)	생성 (NLG)	대화 생성 (Conversational AI)

자료: Google, 언론 보도 종합, IBK투자증권

표 30. OpenAI의 LLM모델과 PaLM모델의 파라미터 차이 비교

모델명	파라미터 수	발표 시기	주요 특징
GPT-1	1.17억	2018	초기 시도, 규모 작음
GPT-2	15억	2019	본격적인 언어 생성 가능
GPT-3	175억	2020	ChatGPT 이전 모델, few-shot 성능 우수
PaLM	5,400억	2022	초거대, Google의 대형 자연어 처리 모델

자료: Google, 언론 보도 종합, IBK투자증권

Chinchilla 법칙,
학습량이 모델 성능을 좌우

Chinchilla(친칠라) 법칙 (2022): 2022년 DeepMind가 발표한 모델 Chinchilla에서 유래한 개념으로, 대형 언어 모델의 성능을 높이기 위해서는 파라미터 수보다 학습 데이터의 양이 더 중요하다는 주장을 담고 있다. 파라미터는 AI 모델이 학습을 통해 데이터에서 얻는 가중치(weight)와 편향(bias) 등 내부 변수이다. 기존에는 GPT-3처럼 수천억 개의 파라미터를 가진 모델이 성능을 좌우한다고 여겨졌지만, Chinchilla는 약 70억 개의 파라미터에 1.4조 개의 토큰(한 사람이 하루에 10만 토큰(약 책 한 권 분량)을 읽는다고 해도, 1.4조 토큰을 다 읽으려면 140만 일이 걸리고 이는 약 3,800년이 넘는 시간)을 학습시켜 GPT-3보다 더 나은 성능을 보였다. 이 결과는 동일한 계산량을 사용할 경우, 모델 크기보다는 학습량(토큰 수)이 성능에 더 큰 영향을 준다는 것을 보여준다. 생성형 AI에서 파라미터 수가 많아질수록 모델이 더 복잡해지고 성능이 향상될 수 있지만, 이에 따라 학습과 추론에 필요한 연산 자원, 전력, 하드웨어, 그리고 클라우드 API 사용료 등 다양한 비용이 기하급수적으로 증가한다.

Bard, PaLM 2 기반 챗봇
이후 Gemini로 통합

Bard와 PaLM 2 (2023): 2023년에는 ChatGPT의 등장에 대응하여 자체 챗봇 서비스인 Bard를 출시하고, 이어서 더욱 강력한 PaLM 2 모델로 업그레이드했다. Bard는 Google이 2023년 3월 정식 출시한 대화형 AI 챗봇 서비스로, OpenAI의 ChatGPT에 대응하기 위한 Google의 전략적 제품이다. Bard가 첫 공개(2023.2) 데모에서 허블 망원경에 대한 잘못된 정보를 답변해 Google 주가가 하락한 사건 발생이 발생했다. Bard는 Google의 대규모 언어모델(LLM)인 PaLM → PaLM 2 → Gemini를 기반으로 진화한다. Bard는 ChatGPT에 상응하는 제품이름이고 PaLM은 초기 모델 이름이다. 2023년 이후 제품 이름과 모델 이름이 모두 Gemini로 통합된다.

표 31. 구글이 발표한 Chinchilla(친칠라) 법칙

항목	기존 GPT-3	Chinchilla
파라미터 수	1750억	약 70억
학습 토큰 수	3000억 tokens	1.4조 tokens
성능(언어 이해/생성 등)	기준 수준	GPT-3보다 우수
메시지	파라미터 수만 키우는 건 비효율	학습량(데이터 양)이 중요

자료: Google, 언론 보도 종합, IBK투자증권

표 32. 구글의 Bard와 OpenAI의 ChatGPT-3.5 비교

구분	구글 Bard (초기, LaMDA 기반)	OpenAI ChatGPT (초기, GPT-3.5 기반)
출시 시점	2023년 03월 21일	2022년 11월 30일
기본 모델	LaMDA의 경량화된 버전	GPT-3.5
정보 업데이트	Google 검색 엔진에 연결되어 비교적 최신 정보 접근 가능	2021년 9월까지의 데이터로 학습, 최신 정보에 제한적
답변 속도	비교적 빠른 답변 속도	단어 단위 생성으로 인해 상대적으로 느리게 느껴질 수 있음
출처 제공	답변에 사용된 정보의 출처를 제공하는 기능 (초기부터 강조)	초기에는 답변 출처를 명확히 제공하지 않음
정확성 및 일관성	초기에는 정보 오류 및 일관성 부족 문제 발생 (출시 초 주가 하락 이슈)	비교적 높은 정확성과 일관성을 보였으나, 때때로 "환각" 현상 발생
당시 평가	출시 초기에는 챗GPT 대비 정보 오류와 완성도에서 부족하다는 평가가 있었으나, 구글의 방대한 데이터와 기술력 기반으로 빠른 개선 기대	혁신적인 성능으로 큰 주목을 받으며 AI 챗봇의 대중화를 이끌었으나, 정보 최신성 부족이 단점으로 지적됨

자료: Google, 언론 보도 종합, IBK투자증권

구글, Gemini로 멀티모달
LLM 경쟁 본격화

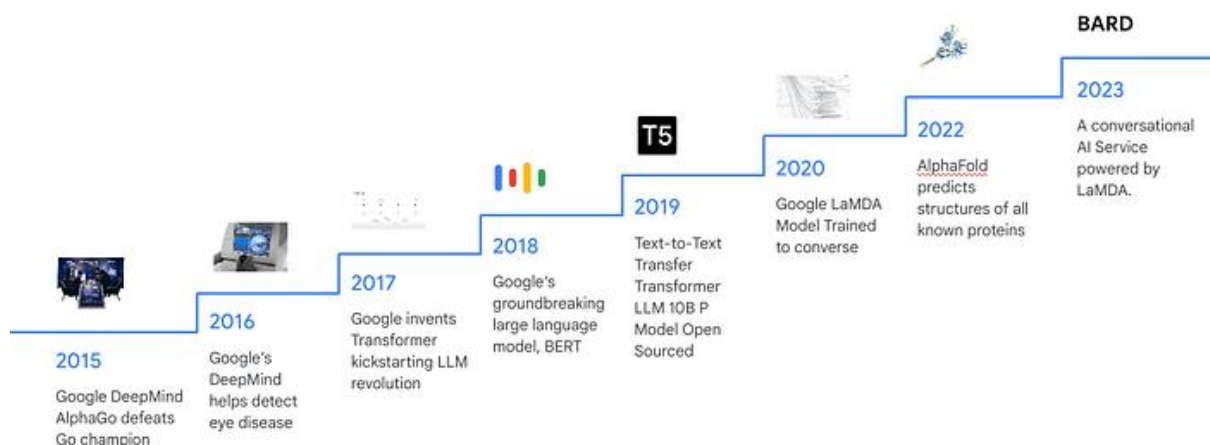
Gemini (2023): 2023년 말, 구글은 차세대 초거대 언어모델인 Gemini를 공식 발표하며 LLM 경쟁에서 본격적으로 존재감을 드러냈다. Gemini는 Ultra, Pro, Nano 세 가지 크기 버전으로 제공되며, 텍스트는 물론 이미지, 오디오, 비디오, 코드 등 멀티모달 입력을 자연스럽게 처리할 수 있도록 설계되었다. 또한 구글은 Gemini를 Gmail, YouTube, Google Search, Android, Pixel, Chromebook 등 자사 서비스 전반에 깊이 통합하여, 단순히 챗봇을 넘는 범용 AI 도우미로서의 활용을 확대하고 있다.

향후 전망:

AI 퍼스트 전략,
Gemini로 전방위 확장

구글은 'AI 퍼스트(AI First)' 기조 아래 모든 서비스에 거대 언어 모델(LLM)을 심층 통합하여 혁신을 주도하고 있으며, 궁극적인 목표는 사용자의 맥락을 깊이 이해하고 다양한 기기에서 사용자를 대신해 행동을 수행할 수 있는 "범용 AI 어시스턴트"를 개발하는 것이다. 이 비전의 중심에는 Gemini 모델의 획기적인 발전이 있으며, 이는 실제 세계의 다양한 측면을 시뮬레이션하여 계획을 세우고 새로운 경험을 상상할 수 있는 '월드 모델'로 진화하는 것을 목표로 한다. 구글은 검색 경험을 혁신하는 SGE(Search Generative Experience)와 협업 도구에 AI 기능을 추가하는 Duet AI 등을 통해 제품 전반의 기능을 고도화하고 있으며, 2025년에 약 750억 달러(약 108조 원)의 막대한 자본을 AI 서비스 강화 및 인프라 확장에 지출할 예정이다. Gemini 2.5 Pro, Flash, Flash-Lite 등 다양한 Gemini 모델을 출시하여 텍스트, 이미지, 비디오, 오디오를 동시에 처리하며 추론 능력을 향상시키고 있으며, 개발자를 위한 Gemini CLI, 이미지 생성을 위한 Imagen 4 등도 제공한다. 또한, AI 기반 광고 도입 및 서비스 구독 모델을 통해 AI 기술에 대한 투자를 회수하고 새로운 수익원을 창출하려는 움직임도 보인다.

그림 71. 구글의 AI 혁신 과정



자료: Google, 언론 보도 종합, IBK투자증권

OpenAI

OpenAI, GPT로
대규모 언어모델 시대 개막

GPT-1 (2018): OpenAI는 대규모 언어 모델 분야의 선구자로서 GPT 시리즈를 통해 인공지능의 발전 방향을 제시해왔다. 2018년, 약 1억 1천만 개의 파라미터를 가진 대규모 언어 모델인 GPT-1을 발표하며 비지도 사전학습(Unsupervised Pre-training)을 통해 언어 이해 성능을 크게 높였다.

GPT-2 (2019): 이듬해인 2019년에는 GPT-1을 10배 확장한 약 15억 개의 파라미터를 가진 GPT-2를 공개하며 언어 생성 능력이 비약적으로 향상되었음을 보여주었다.

GPT-3, 소수 예시만으로도
고성능 구현

GPT-3 (2020): 2020년에는 1,750억 개 파라미터로 혁신적으로 규모를 확대한 GPT-3를 발표하며 제로샷(Zero-shot) 및 퓨샷(Few-shot) 학습만으로도 뛰어난 성능을 보이는 것을 증명했다. 이는 GPT-3가 새로운 작업을 수행할 때 별도의 학습 데이터를 거의 필요로 하지 않음을 의미한다. 또한 "프롬프트 엔지니어링"을 통해 모델의 동작을 제어하고 원하는 결과를 얻을 수 있음을 보여주었다. 이후 발간된 논문 'Emergent Abilities of Large Language Models(Wei et al., 2022)'에서 LLM의 스케일이 일정 임계치를 넘으면, 갑작스럽게 새로운 능력(Emergent abilities)이 나타나는 현상을 정량적으로 설명했다. 스케일이 단순한 양적 증가가 아닌 질적 변화도 초래함을 보여줌.

- 제로샷(Zero-shot) 학습은 AI 모델이 이전에 전혀 보지 못한 유형의 작업이나 질문에 단 한 번의 예시 없이, 혹은 아무런 추가 학습 없이도 스스로 해결할 수 있는 능력을 말한다. 이는 모델이 매우 광범위한 일반 지식과 언어 이해 능력을 갖추고 있음을 보여준다.
- 퓨샷(Few-shot) 학습은 단 몇 개의 예시만 주어진더라도 해당 작업을 즉시 학습하고 수행할 수 있는 능력을 의미한다. 몇 가지 예시를 주고 새로운 유형의 변환을 요청하면, 모델이 그 몇 가지 예시를 통해 패턴을 파악하여 새로운 문장도 변환할 수 있다.

표 33. GPT-1,2,3 비교

모델	출시 연도	파라미터 수	학습 데이터	주요 특징	성능 및 한계
GPT-1	2018	1.17억 (117M)	BookCorpus (7천만 단어)	최초의 GPT 모델, 사전학습(Pretraining) + 파인튜닝(Finetuning) 프레임워크 제안	제한된 성능, 학습 데이터와 크기 한계
GPT-2	2019	15억 (1.5B)	8백만 개 웹페이지 (WebText)	본격적인 "언어 생성" 성능 시연, 다수 과제에서 few-shot 가능	일부 위험성 우려로 초기에 부분 공개
GPT-3	2020	1,750억 (175B)	다양한 웹/문서 데이터	Few-shot/Zero-shot 성능 대폭 향상, 별도 fine-tuning 없이도 높은 성능	고비용, 비효율적인 추론, 사실 오류 가능성

자료: IBK투자증권, 주: BookCorpus: 자연어처리(NLP) 연구에서 널리 사용되는 대규모 영어 텍스트 데이터셋

표 34. 제로샷과 퓨샷

구분	제로샷 (Zero-shot) 학습	퓨샷 (Few-shot) 학습
제공되는 예시 수	0개 (예시 없음)	소수 (보통 1~5개)
작동 방식	사전 학습된 방대한 지식과 일반화 능력에 전적으로 의존	사전 학습된 지식 기반
주요 특징	가장 효율적 (추가 데이터 불필요) 모델의 뛰어난 일반화 능력 발휘	제로샷보다 향상된 성능 기대 특정 형식이나 스타일을 따르는 작업에 특히 유용
활용 시점	일반적인 질문 응답	특정 형식화된 출력 필요 시

자료: Google, 언론 보도 종합, IBK투자증권

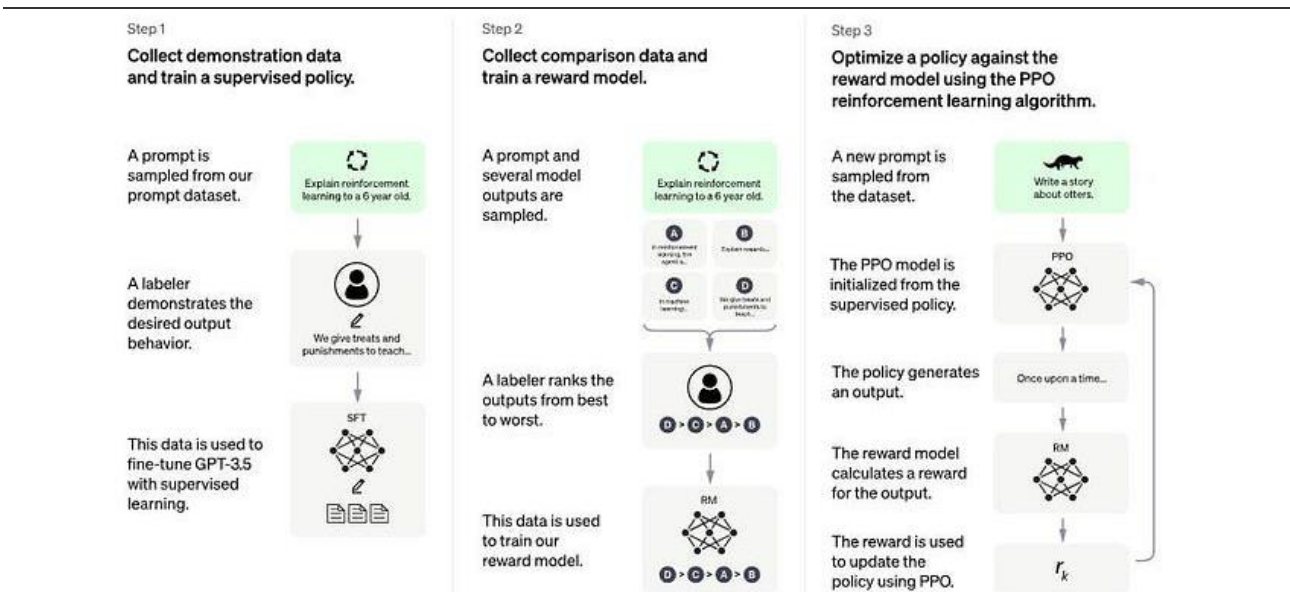
GPT-3.5, 인간 피드백 기반
정교화 진행

GPT-3.5 시리즈 (2021~2022): 2021년부터 2022년까지는 GPT-3를 기반으로 인간 피드백 강화학습(RLHF: Reinforcement Learning from Human Feedback)을 통해 모델을 미세조정하여 사용자의 지시에 더 잘 따르고 유해한 출력을 줄이는 GPT-3.5 시리즈를 개발했다. RLHF는 AI가 여러 가지 답변을 내놓으면 사람들이 직접 "이 답변이 더 좋아", "이건 틀렸어", "이건 좀 위험한 내용이야"와 같이 평가하고 순위를 매긴다. 이렇게 모아진 사람의 피드백을 바탕으로 AI는 '어떤 답변이 좋은 답변인지'를 학습하게 된다. 이를 통해 AI는 단순히 데이터를 학습하는 것을 넘어, 사람의 의도를 더 잘 이해하고, 안전하며 유용한 결과물을 만들어내는 방향으로 진화할 수 있게 된다.

GPT-3.5 적용한 ChatGPT,
1억 돌파 신기록

ChatGPT (2022): 2022년 11월, GPT-3.5 기반의 대화 최적화 모델인 ChatGPT를 출시하며 전 세계적인 센세이션을 일으켰다. ChatGPT는 출시 두 달 만에 사용자 1억 명을 돌파하며 AI 기술에 대한 대중의 폭발적인 관심을 불러일으켰고, 기술이 소비자에게 직접적으로 다가가는 전환점이 되었다.

그림 72. InstrcutGPT의 RLHF(Reinforcement Learning from Human Feedback)



자료: OpenAI, IBK투자증권

그림 73. 사용자 1억명 도달 시기 비교

RANK	ONLINE PLATFORM	FOUNDED	TIME TAKEN
1	Threads	2023	5 Days
2	ChatGPT	2022	2 MONTHS
3	TikTok	2017	9 MONTHS
4	Instagram	2010	2.6 YEARS
5	WhatsApp	2009	3.6 YEARS
6	Snapchat	2011	3.8 YEARS

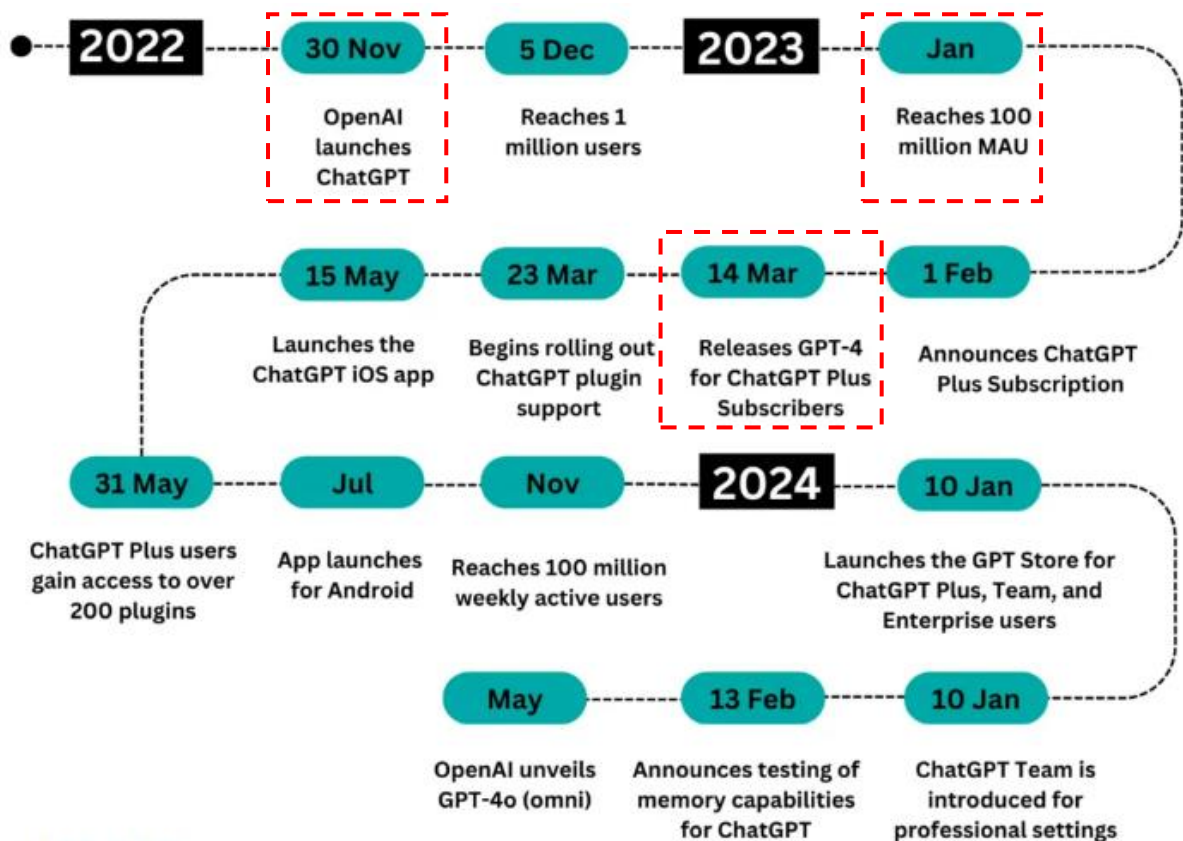
자료: IBK투자증권

향후 전망:

OpenAI, 에이전트 AI · 슈퍼
어시스턴트 · 인프라 전략 시동

OpenAI는 최근 전략의 핵심을 ‘슈퍼 어시스턴트’와 ‘에이전트형 AI’ 구현에 두고 있다. ChatGPT는 단순한 챗봇을 넘어 사용자의 복잡한 요청을 이해하고 외부 시스템과 연동해 실행까지 수행하는 자동화 에이전트로 진화하고 있으며, 이를 위해 Operator, Deep Research 등 다양한 기능형 에이전트를 차례로 출시했다. 이러한 에이전트 중심 전략은 MCP(Model Context Protocol) 도입을 통해 외부 도구 및 시스템과의 표준화된 연결을 지향하며, 향후 오픈 웨이트 언어모델 공개를 통해 개발자 생태계를 적극적으로 확장할 계획이다. 이와 동시에 OpenAI는 SoftBank, Oracle 등과 함께 ‘Stargate’ 프로젝트를 추진해 2029년까지 5천억 달러 규모의 초대형 AI 인프라를 구축할 예정이다.

그림 74. OpenAI ChatGPT History



자료: IBK투자증권

표 35. OpenAI의 실적 추정

연도	매출액(십억달러)	순손실(십억달러)
2023	1.6	-3 ~ -5
2024	3.7	약 -5
2025(E)	12.7	-11 ~ -14

자료: IBK투자증권

Meta

메타, RoBERTa 기반
오픈소스 전략 강화

초기 NLP 연구와 대화모델 (2018~2020): 메타(Meta, 구 Facebook)는 2019년 구글이 발표한 BERT(Bidirectional Encoder Representations from Transformers, 2018) 모델의 사전학습 방식과 하이퍼파라미터를 개선한 RoBERTa(Robustly Optimized BERT Approach)를 공개하며 연구 역량을 입증했다. RoBERTa는 BERT와 동일한 모델 아키텍처를 기반으로 하지만, 더 많은 데이터와 긴 학습 시간, 마스킹 방식 변경 등을 통해 성능을 크게 향상시킨 개량형 모델이다. 메타의 가장 두드러진 전략은 '오픈소스 중심' 접근이다. 실제로 RoBERTa, BlenderBot뿐만 아니라 이후 공개한 LLaMA 시리즈, PyTorch, Segment Anything 등 대부분의 AI 연구성과를 오픈소스로 배포하며 생태계 확대와 연구 협력을 유도하고 있다.

저사양 친화 LLM 공개로
연구 생태계 주도

OPT (2022): 2022년에는 GPT-3에 대응하여 자체적으로 구현한 OPT(Open Pre-trained Transformer)를 공개하며 학계와 업계에 모델 접근을 허용함으로써 "대형 언어 모델 연구의 투명성"을 강조했다.

LLaMA 1 (2023): LLaMA 1은 650억 파라미터 모델이 GPT-3(1,750억 개)보다 적은 파라미터로도 경쟁력 있는 성능을 보였고, 연구 목적으로 가중치를 공개하여 저사양 하드웨어에서도 모델 실행이 가능하도록 지원했다. 상업화를 지양하고, 더 작고, 더 빠르고, 연구자 친화적인 모델을 지향한 점에서 당시 Google/OpenAI의 전략과 뚜렷이 대비되며, 이후 LLaMA 2와 3를 거쳐 오픈소스 생태계의 핵심 모델 계보를 형성하게 된다.

표 36. Meta의 RoBERTa와 구글의 BERT 비교

항목	BERT (Google, 2018)	RoBERTa (Meta, 2019)
모델 구조	Transformer (Encoder 기반)	동일 (BERT와 동일한 아키텍처)
하이퍼파라미터 튜닝	제한적 튜닝	다양한 하이퍼파라미터 실험 및 최적화 적용
성능 (GLUE, SQuAD 등)	당시 최고 수준	동일 모델 구조로 BERT 성능을 상회함을 입증
오픈소스 여부	공개 (TensorFlow 기반)	공개 (PyTorch 기반)
대표 논문	"BERT: Pre-training of Deep Bidirectional Transformers..."	"RoBERTa: A Robustly Optimized BERT Pretraining Approach"

자료: Google, 언론 보도 종합, IBK투자증권

표 37. LLaMA1과 구글의 PaLM, OpenAI의 GPT-3.5 비교

항목	Meta - LLaMA 1	Google - PaLM	OpenAI - GPT-3.5
출시연도	2023년 2월	2022년 4월	2022년 11월 (ChatGPT 서비스 기준)
파라미터 수	7B / 13B / 33B / 65B	8B / 62B / 540B	약 6B ~ 175B (GPT-3는 175B)
특징	고효율 학습 설계, 소형 모델도 강력	초거대 모델 기반, 고성능 지향	대화 최적화, ChatGPT 상용 서비스화
공개 여부	가중치 공개 (연구 목적 한정)	비공개	비공개 (API 제공만)
전략 방향성	오픈소스 중심, 연구 생태계 확대	내부 활용 중심, Bard 백엔드로 발전	상업화 중심, API · ChatGPT 구독 모델

자료: Google, 언론 보도 종합, IBK투자증권

LLaMA 2~3.1, 오픈소스 LLM 성능 한계 돌파

LLaMA 2 (2023): 2023년 7월, Meta가 발표한 LLaMA 2는 상업적 사용까지 허용하는 파격적인 오픈소스 라이선스로 배포되어 업계에 큰 반향을 일으켰고, Microsoft는 이에 발맞춰 LLaMA 2를 자사의 Azure 클라우드 및 Windows 개발 환경에 최적화하여 함께 제공하기 시작했다. Meta는 이를 통해 오픈소스 생태계 확장에 더욱 박차를 가했으며, Microsoft는 OpenAI 중심의 폐쇄형 생태계를 유지하면서도 LLaMA 2를 병행 지원함으로써 AI 플랫폼의 중립성과 확장성을 확보하려는 전략적 균형을 꾀했다.

LLaMA 3 및 3.1 (2024): 2024년에는 이전 세대 대비 다양한 영역에서 GPT-4급 성능을 목표로 개발된 LLaMA 3 및 3.1을 공개했다. 특히 LLaMA 3.1 405B는 "프론티어 수준(frontier-level)의 최초 오픈소스 모델"로 불리며, 성능 면에서 당시 가장 앞선 폐쇄형 모델들에 필적하는 수준에 도달했다. LLaMA 3.1은 상업적 이용이 가능한 '오픈소스' 모델임에도 불구하고, 구글의 Gemini나 OpenAI의 GPT-4와 같은 비공개 최상위 모델들과 어깨를 나란히 할 만큼 뛰어난 성능을 보여주었다는 것을 강조하는 표현이다. 이는 오픈소스 AI의 기술적 한계를 뛰어넘어 AI 모델 최상위 성능을 달성했다는 점에서 AI 생태계에 큰 의미를 가진다.

향후 전망:

Meta, 슈퍼AI 목표로 인재·인프라 집중 투자

2025년 들어 최근 출시한 Llama 4가 혁신성이 부족하고 타 LLM모델 대비 열위를 보인다는 평가를 받았다. 이에 따라 저크버그 CEO는 AI 전략을 대대적으로 강화하며, '슈퍼인텔리전스(Superintelligence)' 개발을 목표로 새로운 연구 조직인 Meta Superintelligence Labs(MSL)를 설립했다. 이 조직은 Scale AI의 창업자 알렉산드르 왕을 영입해 AI 최고책임자로 임명하고, 오픈AI 및 구글 등 경쟁사 출신 인재들을 대규모로 영입하는 등 업계 최고의 인재 확보에 집중한다. Llama 시리즈 등 오픈소스 기반의 대형 언어모델을 지속적으로 발전시키는 한편, AI가 통합된 스마트 글래스, 광고 자동화, Meta AI 챗봇 등 다양한 서비스에 AI를 적극 적용하고 있다. 2025년에는 AI 인프라 확장에도 막대한 투자를 단행해, 13억 개 이상의 GPU와 1GW 이상의 컴퓨팅 파워를 확보하고, 600억~650억 달러의 자본 지출을 계획 중이다. 이러한 전략은 단순한 모델 개발을 넘어, 새로운 수익원 창출, 그리고 Meta 플랫폼 전반에 AI를 심층적으로 통합하는 데 초점을 맞춘다.

표 38. LLaMA3 및 경쟁사 서비스

범주	LLaMA 3 (Meta AI)	Google (Gemini, PaLM)	OpenAI (GPT4)
출시/주력	LLaMA 3 (2024년 4월)	Gemini Pro/Ultra (2024년 초/예상), PaLM 2	GPT4, GPT4 Turbo
접근 방식	오픈 소스 (상업적 이용 가능)	Google Cloud AI (API), Bard/Gemini	OpenAI API (유료), ChatGPT (무료/유료)
주요 특징	연구 및 커스터마이징에 유리 효율성 및 다양한 모델 크기 튜닝 용이	다중 모달리티 (텍스트, 이미지, 영상) Google 생태계 연동 최신 정보 접근 (검색 연동)	고품질 텍스트 생성 및 추론 범용성 및 다양한 앱 통합
성능 수준	상위권 (오픈 소스 중 최고 수준)	최상위권 (특히 다중 모달리티)	최상위권 (텍스트 기반 최고 수준)
단점/제한	상업 모델 대비 지원/책임 부족	최상위 모델 접근 제한, 비용 발생	비용 발생, 블랙박스 모델, 환각 현상 가능

자료: Google, 언론 보도 종합, IBK투자증권

NAVER

HyperCLOVA X, 한국어 특화
멀티모달 AI

HyperCLOVA (2021): NAVER는 2021년 HyperCLOVA를 공개하며 본격적으로 초대규모 AI 모델 경쟁에 뛰어들었다. HyperCLOVA는 한국어 데이터 중심의 GPT-3 계열 모델로, 약 2,040억 파라미터 규모를 자랑하며 NAVER의 핵심 서비스인 검색, 커머스, 광고 시스템에 활용 가능한 높은 수준의 언어 이해 및 생성 능력을 탑재했다. 이는 국내 환경에 최적화된 AI 모델 개발의 중요성을 강조하는 전략이었다.

HyperCLOVA X (2023): 2023년에는 HyperCLOVA의 차세대 모델인 HyperCLOVA X를 출시하며 본격적인 챗봇 및 멀티모달 AI 경쟁에 합류했다. HyperCLOVA X는 AI 질문응답 챗봇 'CLOVA X'와 생성형 검색 서비스 'Cue:'로 분화되어 사용자들에게 직접 서비스 형태로 제공되기 시작했다. 이 모델은 한국어를 중심으로 다국어를 지원하는 구조를 가지고 있으며, 외부 API를 통해 다양한 외부 기업 및 개발자들과의 연결성을 확대하고 있다. 또한 내부적으로는 NAVER의 핵심 서비스인 검색, 쇼핑, 파파고 번역, 광고 최적화 등 기존 서비스에 LLM 기술을 접목하여 기능 고도화를 이루고 있다.

HyperCLOVA X,
검색·커머스 AI로 진화

기술적 강점 및 인프라: NAVER HyperCLOVA X의 핵심 기술 특징은 단연 한국어 특화 모델이라는 점이다. 국내 웹, 커머스, SNS 등 방대한 한국어 데이터를 기반으로 고품질의 파인튜닝(fine-tuning)을 거쳐 한국어에 대한 이해와 생성 능력이 매우 뛰어나다. 2024년부터는 외부 플러그인 및 툴 통합을 통해 활용성을 더욱 넓히고 있다. NAVER는 이러한 대규모 모델 학습을 위해 자체 슈퍼컴퓨터 인프라인 '하이퍼스케일 AI 컴퓨팅 센터 (GAK 세종)'를 구축하여 모델 학습 비용과 시간을 효율적으로 관리하고 내재화했다. LLM을 자사의 핵심 비즈니스인 검색과 커머스를 강화하는 수단으로 활용하고 있으며, 'Cue:'를 통해 키워드 검색을 넘어 사용자의 복잡한 의도를 이해하고 문장 단위의 정보 요약 및 추천을 제공함으로써 검색 경험을 진화시키고 있다. 네이버 쇼핑과 스마트스토어 광고 등과도 긴밀하게 연계되어 비즈니스 시너지를 창출하고 있다. 전략적으로는 B2B 중심의 AI 플랫폼을 지향하고 있으며, HyperCLOVA X는 외부 API 형태로 제공되어 대기업은 물론 중소기업, 공공기관도 쉽게 활용할 수 있도록 설계되었고, 'NAVER Cloud'와 통합된 AI+클라우드 비즈니스 모델을 추진 중이다.

표 39. HyperCLOVA, Gemini, GPT 비교

항목	HyperCLOVA (2021)	HyperCLOVA X (2023)	Google Gemini 2.5 Pro (2025)	OpenAI GPT-4 / GPT-4o (2024)
출시 시점	2021년 5월	2023년 Team NAVER 컨퍼런스	Gemini 1.0 발표 2023년 12월, 2.5 Pro는 2025년 초 등장	GPT-4 출시 2023년 3월, GPT-4o 2024년 5월
파라미터 수	약 2,040억	공식 미공개	모델별 비공개, 멀티엑스퍼트 · Mixture-of-Experts 구조 사용	GPT-4 추정 1.8T, 전문가 추정 1T 이상
학습 데이터	한국어 560B 토큰 중심, 영어 포함	한국어 · 영어 포함 총 6조 토큰, reasoning-focused 사전학습 중심	대규모 멀티모달 데이터 (텍스트/이미지/음성/비디오)	공개 안 됨. 멀티모달+음성/비전 데이터 사용
언어/문화 특화	한국어 in-context few-shot 우수	한국어 · 한국문화에 최적화, reasoning/KOCULT 강화	다국어 기반, 여러 모달리티 동시처리 최적화	영어 중심, 다국어에 강함. GPT-4o로 비영어 성능 개선
멀티모달 입력		현재 텍스트 중심 (추론 강화형 THINK 기반)	텍스트 · 코드 · 이미지 · 음성 · 비디오 입력력 모두 지원	GPT-4: 텍스트 · 이미지 입력, GPT-4o는 음성+비전 실시간 처리
주요 강점	한국어, 한문화 컨텍스트, prompt 최적화 연구	한국어 reasoning 벤치마크 최상위 · 문화적 맥락 이해 · 비용 대비 효율적 inference	멀티모달 + 복합 추론 최강 · 코딩, 시뮬레이션 생성 능력 탁월	학습 · 전문가 벤치마크 우수 · GPT-4o 는 실시간 음성+비전 멀티태스킹 지원

자료: IBK투자증권, 주: BookCorpus: 자연어처리(NLP) 연구에서 널리 사용되는 대규모 영어 텍스트 데이터셋



Company Analysis

기업명	투자의견	목표주가
NAVER (035420)	매수(유지)	320,000원
카카오 (035720)	매수(유지)	72,000원
알파벳 (GOOGL US)	—	—
메타 (META US)	—	—
마이크로소프트 (MSFT US)	—	—
아마존 (AMZN US)	—	—
에스티로더 (EL US)	—	—

매수 (유지)

목표주가(상향) 320,000원
현재가 (7/14) 249,000원

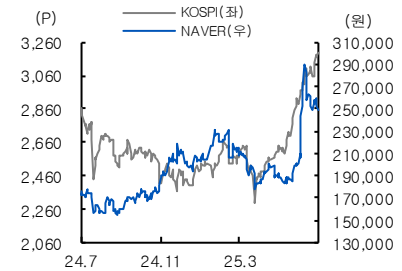
KOSPI (7/14)	3,202.03pt
시가총액	39,451십억원
발행주식수	158,437천주
액면가	100원
52주 최고가	290,500원
최저가	155,000원
60일 일평균거래대금	286십억원
외국인 지분율	45.0%
배당수익률 (2025F)	0.4%

주주구성	
국민연금공단	9.24%
BlackRock Fund Advisors 외 13인	6.05%

주가상승률	1M	6M	12M
상대기준	12%	-4%	27%
절대기준	24%	23%	42%

	현재	직전	변동
투자 의견	매수	매수	-
목표주가	320,000	300,000	-
EPS(25)	11,831	11,098	▲
EPS(26)	14,414	13,320	▲

NAVER 상대주가 (%)



NAVER (035420)

글로벌 AI경쟁력으로 저평가를 벗어날 때

목표주가 32만원으로 상향, 투자 의견 매수 유지

NAVER의 목표주가는 12개월 FW EPS를 7% 상향 부분을 반영하여 산정했다. 신정부의 경기 부양 의지가 크고 대규모 추경과 민생회복 지원금으로 인해 3분기 실적 개선 가능성이 커졌다. 동사의 핵심 인력이 정부 인사에 발탁된 만큼 향후 AI 및 소상공인 기반 커머스 전략에서 시너지가 기대된다.

2분기 실적 컨센서스 상회 전망

2분기는 매출액 2,9조원(YoY +11.4%, QoQ +4.4%), 영업이익 5,257억원(YoY +11.2%, QoQ +4.0%)를 기록하여 컨센서스에 부합할 것으로 추정된다. 2분기 서치플랫폼 등의 광고 매출이 AdVoost 도입 효과가 반영되면서 시장 기대치를 상회할 것으로 예상된다. 한편 커머스의 스마트스토어 수수료 인상 효과가 반영되면서 2분기 수익성 개선에 기여할 것으로 분석된다. 콘텐츠는 일본의 글든위크 효과로 웹툰 매출이 전분기 대비 상승할 것으로 예상된다.

차세대 추론형 AI 모델, ‘하이퍼클로바X 씽크’ 공개

2025년, 기존 대형 언어모델(LLM)을 한 단계 진화시킨 차세대 추론형 인공지능 모델 ‘하이퍼클로바X 씽크(HyperCLOVA X Think)’를 공개했다. 이 모델은 단순한 응답 생성형 LLM을 넘어, 인간처럼 생각의 과정을 거쳐 답변을 도출하는 능력을 갖춘 것이 핵심이다. 특히 한국어 처리 성능에서 업계 최고 수준의 결과를 기록했다.

또한, 동사는 정부가 모집 중인 ‘독자 AI 파운데이션 모델’ 프로젝트에 유력한 참가 기업으로 평가된다. 한글에 강점을 가지는 자체 LLM을 기반으로 자사의 다양한 서비스와 연계, 통합시키면서 AI모델이 실제 실적 개선으로 이어질 가능성이 높다.

(단위:십억원,배)	2023	2024	2025F	2026F	2027F
매출액	9,671	10,738	11,915	13,122	14,400
영업이익	1,489	1,979	2,265	2,625	2,982
세전이익	1,481	2,322	2,500	3,023	3,453
지배주주순이익	1,012	1,923	1,875	2,284	2,609
EPS(원)	6,180	11,913	11,831	14,414	16,466
증가율(%)	33.4	92.8	-0.7	21.8	14.2
영업이익률(%)	15.4	18.4	19.0	20.0	20.7
순이익률(%)	10.2	18.0	15.7	17.4	18.1
ROE(%)	4.4	7.9	7.0	7.9	8.3
PER	36.2	16.7	21.0	17.3	15.1
PBR	1.6	1.2	1.4	1.3	1.2
EV/EBITDA	18.2	11.4	11.9	10.6	8.9

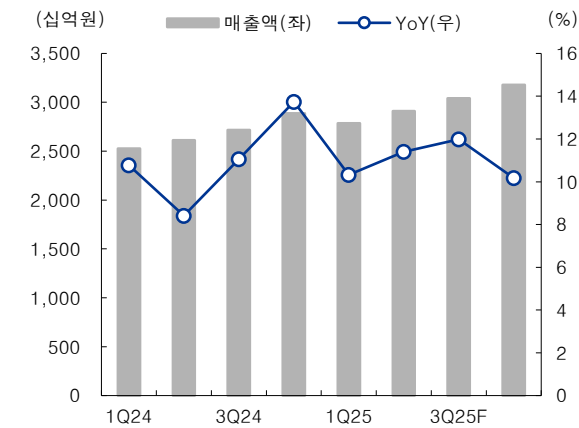
자료: Company data, IBK투자증권 예상

표 40. NAVER 실적 추정

(십억원)	1Q24	2Q24	3Q24	4Q24	1Q25	2Q25F	3Q25F	4Q25F	2024F	2025F	2026F
매출액	2,526	2,610	2,716	2,886	2,787	2,908	3,041	3,179	10,738	11,915	13,122
QoQ	-0.4%	3.3%	4.0%	6.3%	-3.4%	4.4%	4.6%	4.6%			
YoY	10.8%	8.4%	11.1%	13.7%	10.3%	11.4%	12.0%	10.2%	11.0%	11.0%	10.1%
서치플랫폼	905	978	998	1,065	1,013	1,104	1,149	1,205	3,946	4,470	4,909
커머스	703	719	725	775	788	791	816	869	2,923	3,265	3,647
핀테크	354	368	385	401	393	405	431	453	1,508	1,682	1,859
콘텐츠	446	420	463	467	459	466	487	486	1,796	1,899	2,019
클라우드	117	125	145	178	134	141	158	165	564	599	688
영업이익	439	473	525	542	505	526	603	631	1,979	2,265	2,625
OPM	17.4%	18.1%	19.3%	18.8%	18.1%	18.1%	19.8%	19.8%	18.4%	19.0%	20.0%
QoQ	8.3%	7.6%	11.1%	3.2%	-6.8%	4.0%	14.7%	4.6%			
YoY	32.9%	26.8%	38.2%	33.7%	15.0%	11.2%	14.8%	16.4%	32.9%	14.4%	15.9%
세전이익	660	467	653	541	580	613	692	616	2,500	2,500	3,023
QoQ	50.3%	-29.2%	39.8%	-17.1%	7.2%	5.6%	12.9%	-10.9%			
YoY	466.3%	11.0%	29.5%	-17.1%	-12.1%	31.1%	5.9%	-10.9%	68.8%	30.2%	-92.9%
순이익(지배)	511	338	520	554	425	463	522	465	1,923	1,875	2,284
NIM	20.2%	13.0%	19.2%	19.2%	15.2%	15.9%	17.2%	14.6%	17.5%	15.7%	-0.4%
QoQ	62.7%	-33.7%	53.8%	6.4%	-23.3%	8.9%	12.9%	-10.9%			
YoY	827.4%	26.4%	38.5%	76.5%	-16.8%	36.7%	0.3%	-16.1%	90.0%	-2.5%	21.8%

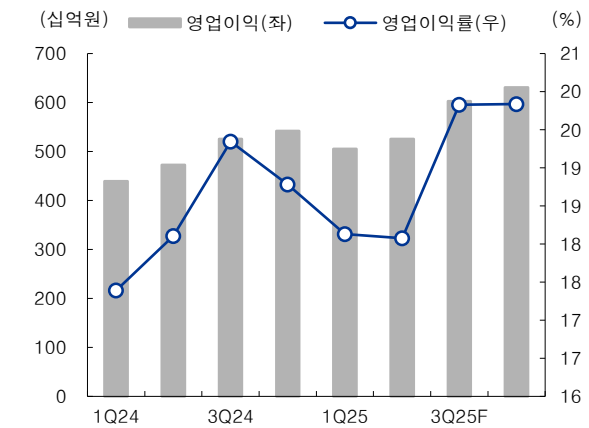
자료: NAVER, IBK투자증권 추정

그림 75. NAVER의 매출액 및 YoY 추이



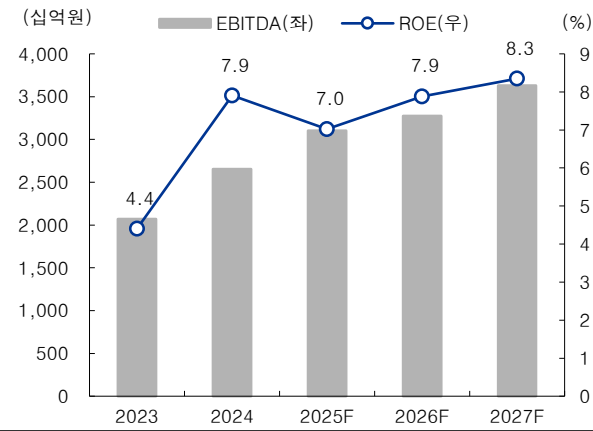
자료: NAVER, IBK투자증권 추정

그림 76. NAVER의 영업이익 및 영업이익률 추이



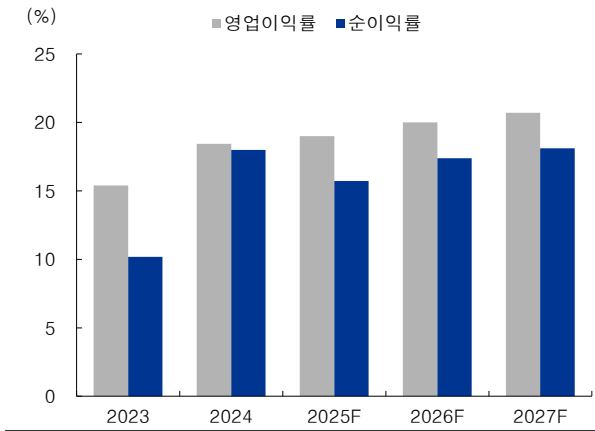
자료: NAVER, IBK투자증권 추정

그림 77. NAVER의 EBITDA 및 ROE 추이



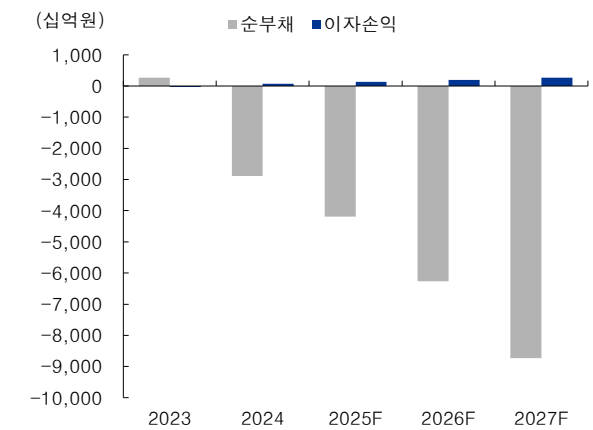
자료: Quantiwise, IBK투자증권

그림 78. NAVER의 영업이익률 및 순이익률 추이



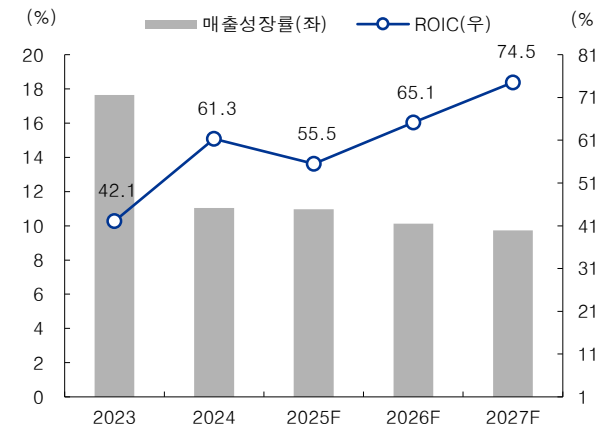
자료: Quantiwise, IBK투자증권

그림 79. NAVER의 순부채 및 이자손익 추이



자료: Quantiwise, IBK투자증권

그림 80. NAVER의 매출성장률과 ROIC 추이



자료: Quantiwise, IBK투자증권

정부의 소버린 AI 개발의 핵심 파트너

정부 GPU 임차 사업 수주, AI 인프라 역량 인정받아

네이버, K-AI용
GPU 공급 주도

네이버클라우드는 2025년 7월, 과학기술정보통신부와 정보통신산업진흥원이 주관하는 ‘AI 컴퓨팅 자원 활용 기반 강화’ 사업에서 GPU 공급을 위한 우선 협상 대상으로 선정되었다. 본 사업은 민간이 보유한 고성능 GPU를 AI 모델 개발 기업에 클라우드 방식으로 제공해, 국가 AI 생태계 전반의 연산 자원 부족 문제를 해소하는 데 목적이 있다. 네이버클라우드는 ‘2트랙’ 사업자로 참여하며, 엔비디아의 최신 고성능 GPU인 H100/H200 1,000장 또는 B200 500장을 2025년 하반기부터 약 11개월간 공급할 계획이다.

K-AI 전략과 연계, 한국형 파운데이션 모델 개발 본격화

해당 GPU 자원은 과기정보통신부가 추진하는 ‘독자 AI 파운데이션 모델 프로젝트 (K-AI)’에 활용될 예정이다. 이 프로젝트는 한국형 LLM 생태계 구축을 목표로, 최대 5개 민간 컨소시엄에 연산 자원과 데이터를 집중 지원하는 국가 전략 사업이다.

그림 81. 과기정보통신부의 ‘독자 AI 파운데이션 모델 프로젝트 (K-AI)’

2 추진 방향



자료: 과학기술정보통신부, IBK투자증권

최근 AI 전략 및 서비스 현황

1. '온 서비스 AI(On-Service AI)' 전략 선언

전 서비스에 AI 심층 도입

2024년 진행된 팀네이버 통합 컨퍼런스인 DAN24에서 전 서비스에 AI를 녹여내는 것이 목표로 2025년을 'AI 서비스 적용의 해'로 선언했다. 자사 핵심 서비스 전반에 AI 기술을 심층적으로 도입하고 있다. 이를 '온서비스 AI' 전략이라 명명하고 있으며, 단순한 AI 기능 추가를 넘어 각 서비스의 본질적 경쟁력을 AI로 끌어올리는 것을 목표로 하고 있다.

2. 주요 AI 서비스 및 기능

2-1. 검색 및 정보 요약

검색 영역에서는 생성형 AI 기반 'AI 브리핑' 기능을 통해 사용자의 질의 의도를 파악하고, 핵심 정보를 요약 제공한다. 기존 키워드 기반 검색과 달리 요약된 답변과 함께 출처 및 관련 링크까지 제공해 정보 탐색 효율을 높이고 있다.

2-2. 지도·공간 서비스

지도 서비스에는 실시간 AI 분석 기반의 'AI 지도'를 적용하고 있다. 사용자의 위치, 교통 상황, 목적 등을 고려해 최적 경로를 제공하며, 실시간 맞춤형 추천과 상황별 안내 기능을 탑재했다. 기존에는 전체 이용자 데이터를 기반으로 평균 도착 시간을 안내했지만, 현재는 사용자별 운전 습관을 반영한 정교한 예측이 가능해졌다.

2-3. 쇼핑·스토어

AI 기반 이미지 인식 기술을 활용해 상품 이미지에서 자동으로 설명 문구를 생성하는 기능이 도입되었다. 또한, 배경음악 삽입, 동영상 소음 제거 등 콘텐츠 편집 기능도 함께 제공하며, 사용자 맞춤형 상품 추천 알고리즘도 고도화하고 있다. 이를 통해 판매자와 소비자 모두에게 효율적인 쇼핑 경험을 제공하고 있다.

네이버 D2SF, 패션 특화 AI 스타트업 '스튜디오랩'에 투자

– 2025년 2월, 네이버의 스타트업 투자 조직 D2SF는 멀티모달 AI를 활용한 이커머스 콘텐츠 자동화 솔루션을 개발한 스타트업 '스튜디오랩'에 33억 원 규모의 프리시리즈A 투자를 진행했다고 밝혔다

– 스튜디오랩은 자체 개발한 패션 특화 AI 모델을 기반으로, 쇼핑몰 상세 페이지를 자동으로 생성해주는 솔루션 'GENCY'를 운영 중이다. 사용자는 제품 사진 몇 장만 업로드하면, 해당 제품에 최적화된 마케팅 문구, 썸네일, 레이아웃, 세부 디자인을 15~30초 내에 생성할 수 있다. 하이퍼클로바X 기반 '네이버플러스 스토어', AI 쇼핑 기능 강화

2025년 3월, 네이버는 초대규모 언어모델 하이퍼클로바X를 기반으로 한 AI 쇼핑앱 ‘네이버플러스 스토어’에 신규 기능을 추가했다. 새롭게 도입된 기능은 ‘AI 쇼핑 가이드’와 ‘발견’이다.

2-4. 광고·콘텐츠 생성

2025년 6월 네이버는 인공지능(AI)을 활용해 쇼핑 광고의 효율을 높이고, 광고 운영의 복잡함을 줄여주는 ‘애드부스트 쇼핑(ADVoost Shopping)’을 오픈 베타 형태로 공개했다. ‘애드부스트’는 네이버가 지난해 선보인 AI 기반 광고 기술 솔루션 브랜드로, 정교한 데이터 분석과 타겟 확장, 광고 소재 자동 생성, 쇼핑·검색 연동 광고 운영 등 광고주의 성과 극대화와 편의성을 동시에 추구하는 것이 핵심이다. ‘애드부스트 쇼핑’은 네이버 쇼핑 플랫폼 내 판매자들을 위한 전용 자동화 솔루션이다. 광고 캠페인의 생성과 운영, 상품 연동 및 소재 추천, 게재 위치 결정과 사용자 노출까지 전 과정을 AI가 자동으로 처리한다.

그림 82. 네이버의 애드부스트 쇼핑 베타



자료: 네이버, IBK투자증권

2-5. 에이전트

네이버는 2026년 검색 AI 에이전트 ‘AI 탭’을 출시할 예정이다. ‘AI 탭’은 기존 통합검색과는 분리된 전용 페이지로 제공되며, 사용자의 질문 맥락을 깊이 있게 이해하고, 단순 정보 제공을 넘어 실제 행동까지 돕는 것을 목표로 하고 있다.

예를 들어, 사용자가 “5살 아이와 제주도 갈 만한 곳 추천해줘”라고 입력하면, 네이버는 과거 대화 이력이나 서비스 이용 데이터를 바탕으로 적합한 장소를 선별해 제안한다. 이어 사용자가 특정 장소를 선택하고 여행 코스를 요청하면, 네이버 지도 기반으로 최적 동선을 안내하고, 숙소 주변의 추천 장소와 예약 가능한 식당 등의 정보까지 연계해 제공한다. 이는 단순한 질의응답을 넘어, 사용자의 의도에 따라 실제 행동을 유도하고 지원하는 AI 에이전트형 검색 환경 구축을 지향하는 것으로, 네이버는 이를 통해 AI 시대에 맞는 검색 패러다임 전환을 본격화하고 있다.

3. 최신 LLM 및 경량 모델

‘쌍크’, 경량화해
오픈소스로 배포

‘하이퍼클로바X 쌍크’는 2025년 공개한 차세대 추론형 인공지능(AI) 모델이다. 기존의 대형 언어모델(LLM)보다 한 단계 진화해, 인간처럼 ‘생각의 과정’을 거쳐 답변을 도출하는 능력을 갖췄다는 점이 특징이며 특히 한국어 처리 능력에서 업계 최고 수준의 성과를 보였다. 서울대 언어학자가 주관한 한국어 벤치마크 테스트 ‘KoBALT-700’에서 48.9점을 기록하며, 엑사원 딥(33점), QwQ-32B(32.4점) 등 주요 경쟁 모델들을 크게 앞섰다. 이미지 기반 정보까지 추론할 수 있는 멀티모달 처리 능력도 초기 실험에서 의미 있는 성과를 보였다. 과학이나 수학 시험 문제처럼 그래프와 도표를 포함한 이미지 기반 데이터를 입력하면, 이를 해석해 정답을 유추하는 데 성공했다.

네이버는 6월 30일 하이퍼클로바X 쌍크의 개발 완료 및 기술 리포트 발표를 공식화했으며, 향후 해당 모델의 경량화 버전을 오픈소스로 배포할 계획이다. 또한 자사 대화형 AI 플랫폼인 클로바 X(CLOVA X)를 통해 순차적으로 적용 및 서비스에 활용할 예정이다.

그림 83. 네이버의 HyperCLOVA X SEED



자료: 네이버, IBK투자증권

4. AI 인프라 확장

모로코에 500MW
AI 데이터센터 구축

자체 데이터센터 및 클라우드 인프라를 바탕으로 AI 모델의 안정적 운영과 확장성을 확보하고 있다. 특히 ‘소버린 AI’ 전략을 통해 한국은 물론 일본, 동남아, 중동 등 국가별 규제 및 문화에 맞춘 현지화된 AI 생태계 구축을 추진하고 있다.

모로코 AI 데이터센터 구축

- 2025년 6월 네이버는 북아프리카 모로코에 인공지능(AI) 데이터센터를 구축하는 글로벌 컨소시엄에 참여한다고 밝혔다. 이번 프로젝트는 엔비디아, AI 인프라 전문 기업 넥서스 코어 시스템즈, 글로벌 투자사 로이드 캐피탈 등과 함께 추진되며, 네이버는 클라우드 플랫폼의 운영 주체로 참여한다.

- 데이터센터는 최대 500MW 규모의 재생에너지 기반 슈퍼컴퓨팅 인프라로, 1단계로는 연내 40MW급 설비에 엔비디아 GB200 GPU를 탑재해 연내 구축을 완료할 계획이다. 이후 단계적으로 인프라를 확장해, 전체 규모를 국내 '각 세종'(270MW)의 약 1.8배 수준인 500MW급으로 확대한다는 방침이다.

- 이번 사업은 네이버의 ‘소버린 AI’ 전략을 본격화하는 사례로, 데이터 저장부터 처리, 운영까지 현지에서 독립적으로 수행하는 인프라 구조를 통해 데이터 주권에 민감한 유럽·중동·아프리카(EMEA) 지역 내 신뢰도 제고가 가능할 것으로 기대된다.

NAVER (035420)

포괄손익계산서

(십억원)	2023	2024	2025F	2026F	2027F
매출액	9,671	10,738	11,915	13,122	14,400
증가율(%)	17.6	11.0	11.0	10.1	9.7
매출원가	0	0	0	0	0
매출총이익	9,671	10,738	11,915	13,122	14,400
매출총이익률 (%)	100.0	100.0	100.0	100.0	100.0
판매비	8,182	8,758	9,651	10,497	11,418
판매비율(%)	84.6	81.6	81.0	80.0	79.3
영업이익	1,489	1,979	2,265	2,625	2,982
증가율(%)	14.1	32.9	14.4	15.9	13.6
영업이익률(%)	15.4	18.4	19.0	20.0	20.7
순금융손익	-79	54	247	397	471
이자손익	-26	69	137	191	265
기타	-53	-15	110	206	206
기타영업외손익	-195	144	-9	0	0
중속/관계기업손익	267	145	-3	0	0
세전이익	1,481	2,322	2,500	3,023	3,453
법인세	496	390	627	741	846
법인세율	33.5	16.8	25.1	24.5	24.5
계속사업이익	985	1,932	1,873	2,282	2,607
중단사업손익	0	0	0	0	0
당기순이익	985	1,932	1,873	2,282	2,607
증가율(%)	46.3	96.1	-3.0	21.8	14.2
당기순이익률 (%)	10.2	18.0	15.7	17.4	18.1
지배주주당기순이익	1,012	1,923	1,875	2,284	2,609
기타포괄이익	-261	699	717	0	0
총포괄이익	724	2,631	2,591	2,282	2,607
EBITDA	2,071	2,653	3,100	3,275	3,630
증가율(%)	11.0	28.1	16.9	5.6	10.8
EBITDA마진율(%)	21.4	24.7	26.0	25.0	25.2

투자지표

(12월 결산)	2023	2024	2025F	2026F	2027F
주당지표(원)					
EPS	6,180	11,913	11,831	14,414	16,466
BPS	142,887	160,694	176,174	189,523	204,926
DPS	1,205	1,130	1,130	1,130	1,130
밸류에이션(배)					
PER	36.2	16.7	21.0	17.3	15.1
PBR	1.6	1.2	1.4	1.3	1.2
EV/EBITDA	18.2	11.4	11.9	10.6	8.9
성장성지표(%)					
매출증가율	17.6	11.0	11.0	10.1	9.7
EPS증가율	33.4	92.8	-0.7	21.8	14.2
수익성지표(%)					
배당수익률	0.5	0.6	0.4	0.4	0.4
ROE	4.4	7.9	7.0	7.9	8.3
ROA	2.8	5.2	4.8	5.5	5.9
ROIC	42.1	61.3	55.5	65.1	74.5
안정성지표(%)					
부채비율(%)	47.4	41.4	37.2	35.5	34.3
순차입금 비율(%)	-1.8	-12.8	-16.2	-21.6	-27.3
이자보상배율(배)	11.7	18.9	21.4	24.8	28.1
활동성지표(배)					
매출채권회전율	18.9	22.2	24.3	25.5	26.7
재고자산회전율	824.3	587.5	514.8	650.7	733.3
총자산회전율	0.3	0.3	0.3	0.3	0.3

*주당지표 및 밸류에이션은 지배주주순이익 및 지배주주지분 기준

재무상태표

(십억원)	2023	2024	2025F	2026F	2027F
유동자산	7,028	9,375	11,287	13,544	16,229
현금및현금성자산	3,576	4,196	5,548	7,390	9,610
유가증권	808	2,860	2,853	3,088	3,343
매출채권	491	478	502	527	553
재고자산	15	22	25	16	23
비유동자산	28,710	28,793	29,158	29,276	29,474
유형자산	2,742	2,910	2,479	2,445	2,475
무형자산	3,446	3,657	3,695	3,731	3,775
투자자산	21,231	20,910	21,654	21,660	21,667
자산총계	35,738	38,168	40,445	42,820	45,704
유동부채	6,306	6,092	7,155	7,316	7,643
매입채무및기타채무	0	0	0	0	0
단기차입금	333	135	142	147	153
유동성장기부채	443	200	1,820	1,820	1,820
비유동부채	5,194	5,075	3,811	3,911	4,029
사채	1,656	2,007	394	394	394
장기차입금	994	863	905	905	905
부채총계	11,500	11,167	10,965	11,227	11,672
자배주주지분	23,206	25,460	27,912	30,027	32,468
자본금	16	16	16	16	16
자본잉여금	1,243	1,423	1,460	1,460	1,460
자본조정등	-607	-639	-634	-634	-634
기타포괄이익누계액	-1,990	-1,306	-586	-586	-586
이익잉여금	24,544	25,965	27,657	29,772	32,212
비지배주주지분	1,032	1,541	1,567	1,565	1,564
자본총계	24,238	27,001	29,479	31,593	34,032
비이자부채	6852	6997	6754	7010	7449
총차입금	4,648	4,170	4,211	4,217	4,223
순차입금	-441	-3,462	-4,766	-6,837	-9,306

현금흐름표

(십억원)	2023	2024	2025F	2026F	2027F
영업활동 현금흐름	2,002	2,590	2,332	2,744	3,050
당기순이익	985	1,932	1,873	2,282	2,607
비현금성 비용 및 수익	1,326	1,003	878	253	176
유형자산감가상각비	528	609	836	650	648
무형자산상각비	54	65	0	0	0
운전자본변동	361	152	-480	19	2
매출채권등의 감소	43	-3	-502	-25	-26
재고자산의 감소	-2	-13	-3	9	-8
매입채무등의 증가	0	0	0	0	0
기타 영업현금흐름	-670	-497	61	190	265
투자활동 현금흐름	-950	-1,340	-716	-1,167	-1,278
유형자산의 증가(CAPEX)	-641	-554	-590	-616	-678
유형자산의 감소	7	32	4	0	0
무형자산의 감소(증가)	-51	-24	-32	-36	-43
투자자산의 감소(증가)	535	839	106	-6	-7
기타	-800	-1633	-204	-509	-550
재무활동 현금흐름	-110	-770	-313	265	447
차입금의 증가(감소)	218	105	0	0	0
자본의 증가	0	0	0	0	0
기타	-328	-875	-313	265	447
기타 및 조정	-90	139	49	0	1
현금의 증가	852	619	1,352	1,842	2,220
기초현금	2,724	3,576	4,196	5,548	7,390
기말현금	3,576	4,196	5,548	7,390	9,610

매수 (유지)

목표주가(상향) 72,000원
현재가 (7/14) 58,300원

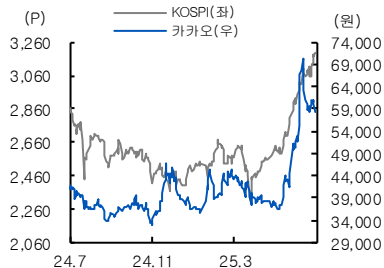
KOSPI (7/14) 3,202.03pt
시가총액 25,769십억원
발행주식수 442,014천주
액면가 100원
52주 최고가 70,400원
최저가 32,800원
60일 일평균거래대금 221십억원
외국인 지분율 27.8%
배당수익률 (2025F) 0.1%

주주구성
김범수 외 92 인 24.30%
국민연금공단 6.40%

	1M	6M	12M
주가상승률			
상대기준	2%	23%	23%
절대기준	13%	58%	38%

	현재	직전	변동
투자 의견	매수	매수	-
목표주가	72,000	52,000	-
EPS(25)	966	870	▲
EPS(26)	1,148	1,117	▲

카카오 상대주가 (%)



카카오 (035720)

한국형 Agentic AI 구현이 기대된다

목표주가 72,000원으로 상향, 투자 의견 매수 유지

카카오의 목표주가는 12개월 FW SPS(주당매출액)에 2022년(OpenAI의 ChatGPT가 출시)이후 NAVER와 카카오의 PSR 상단값인 Target PSR 3.8배를 적용하여 산출했다. 국내 AI산업에 정책적 지원이 예상되고 3분기부터 경기 회복 가능성이 높아진 만큼 광고 중심의 실적 개선 가능성도 높은 것으로 판단한다. 동사는 카카오톡, 모빌리티, 페이, 멜론 등 전 국민 대상 생활밀착형 서비스 다수 보유하고 있기 때문에 NEXT AI로 기대되는 Agentic AI의 개척자로서 역할이 기대된다

2분기 영업이익은 컨센서스 소폭 하회 전망

2분기는 매출액 1.96조원(YoY -2.2%, QoQ +5.2%), 영업이익 1,191억원(YoY -11.1%, QoQ 13.0%)를 기록하여 영업이익이 시장 기대치를 소폭 하회할 것으로 추정된다. 2분기 광고형 매출은 5월부터 소비자 지수가 반등하는 등 전분기 대비 개선될 것으로 예상하지만 커머스는 소폭 하회할 것으로 예상된다. 또한 신사업과 뮤직이 전분기 대비 +10% 이상 호조를 보일 것으로 전망된다. 비용은 디어유가 연결로 편입되면서 인건비가 소폭 상승하고 콘텐츠 부문 매출 호조로 매출 연동비 역시 증가할 것으로 전망한다.

오픈AI 협력, '한국형 슈퍼 AI 에이전트' 연내 출격

오픈AI와의 협력을 통해 연내 '한국형 슈퍼 AI 에이전트 플랫폼'을 출시할 계획이다. 이 플랫폼은 카카오톡을 비롯한 다양한 카카오 서비스와 긴밀하게 연계되어 사용자가 한 번의 명령으로 복합적인 기능을 실행할 수 있도록 설계될 예정이다. 하반기 시범 서비스(CBT) 종료 후 11월경 정식 공개될 가능성이 높다.

(단위:십억원,배)	2023	2024	2025F	2026F	2027F
매출액	7,557	7,872	7,971	8,682	9,357
영업이익	461	460	548	723	931
세전이익	-1,648	-3	607	747	1,023
지배주주순이익	-1,013	55	427	507	713
EPS(원)	-2,276	124	966	1,148	1,613
증가율(%)	-174.7	-105.5	675.7	18.9	40.5
영업이익률(%)	6.1	5.8	6.9	8.3	9.9
순이익률(%)	-24.0	-2.1	6.4	6.5	8.3
ROE(%)	-10.3	0.6	4.1	4.7	6.2
PER	-23.9	306.9	60.4	50.8	36.1
PBR	2.5	1.7	2.4	2.3	2.2
EV/EBITDA	20.3	12.8	22.9	20.2	16.3

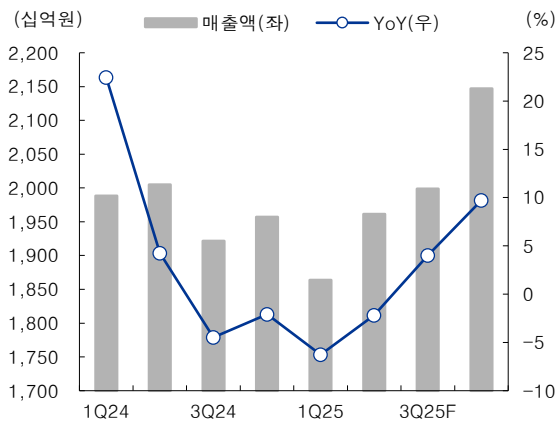
자료: Company data, IBK투자증권 예상

표 41. 카카오 실적 추이

(십억원)	1Q24	2Q24	3Q24	4Q24	1Q25	2Q25F	3Q25F	4Q25F	2024	2025F	2026F
매출액	1,988	2,005	1,921	1,957	1,864	1,961	1,999	2,147	7,872	7,971	8,682
QoQ	-0.5%	0.8%	-4.2%	1.9%	-4.8%	5.2%	1.9%	7.4%			
YoY	22.5%	4.2%	-4.5%	-2.1%	-6.3%	-2.2%	4.0%	9.7%	0.0%	1.3%	8.9%
플랫폼부문	955	955	943	1,047	993	1,025	1,022	1,126	3,901	4,166	4,533
특비즈	516	508	501	557	553	549	537	610	2,082	2,249	2,448
포털비즈	85	88	76	83	74	77	72	79	332	302	308
플랫폼 기타	354	360	366	407	366	399	413	437	1,487	1,615	1,776
콘텐츠부문	1,033	1,050	978	910	871	936	977	1,021	3,971	3,805	4,150
스토리	227	216	219	203	213	216	211	213	864	852	895
게임	243	233	234	163	145	143	160	200	873	648	791
미디어	95	90	55	74	75	72	60	78	313	285	299
뮤직	468	511	471	470	438	506	547	530	1,920	2,020	2,165
영업이익	120	134	131	75	105	119	153	171	460	548	723
OPM	6.0%	6.7%	6.8%	3.9%	5.7%	6.1%	7.7%	8.0%	5.8%	6.9%	8.3%
QoQ	-25.2%	11.4%	-2.6%	-42.2%	39.7%	13.0%	28.6%	11.5%			
YoY	92.2%	18.5%	5.0%	-53.1%	-12.4%	-11.1%	17.4%	126.5%	-0.1%	19.2%	31.8%
세전이익	141	165	136	-444	193	171	209	34	-3	607	747
QoQ	흑전	17.2%	-17.7%	적전	흑전	-11.1%	22.0%	-83.9%			
YoY	117.0%	47.6%	20.6%	적지	37.1%	4.0%	54.1%	흑전	적지	흑전	23.0%
순이익(지배)	74	101	106	-226	172	114	126	15	55	427	507
NM	3.7%	5.1%	5.5%	-11.5%	9.2%	5.8%	6.3%	0.7%	0.7%	5.4%	5.8%
QoQ	흑전	37.4%	4.4%	적전	흑전	-33.7%	10.9%	-88.0%			
YoY	232.1%	81.1%	407.1%	적지	133.1%	12.5%	19.4%	흑전	흑전	673.1%	18.7%

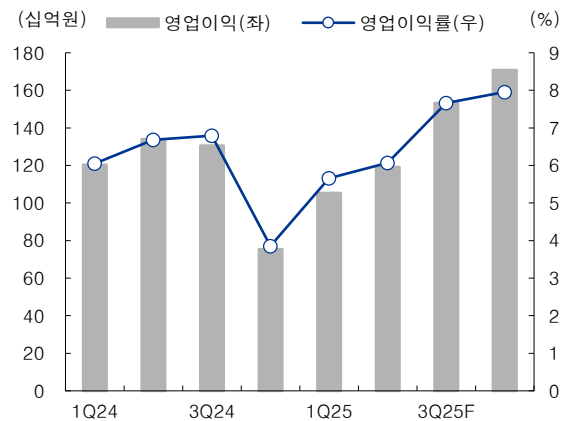
자료: 카카오, IBK투자증권

그림 84. 카카오의 매출액 및 YoY 추이



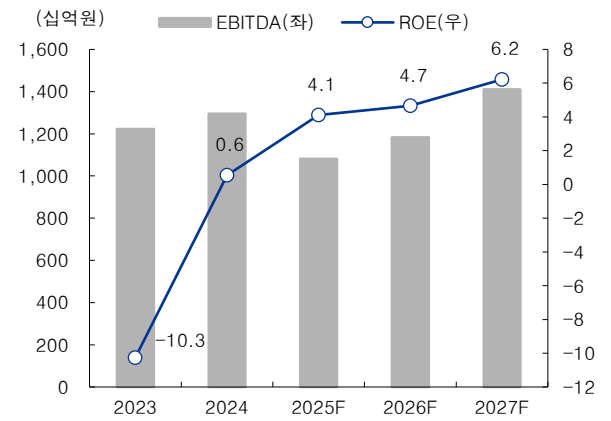
자료: 카카오, IBK투자증권 추정

그림 85. 카카오의 영업이익 및 영업이익률 추이



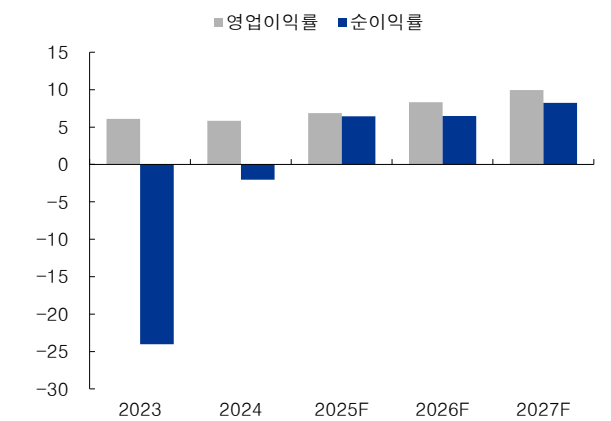
자료: 카카오, IBK투자증권 추정

그림 86. 카카오의 EBITDA 및 ROE 추이



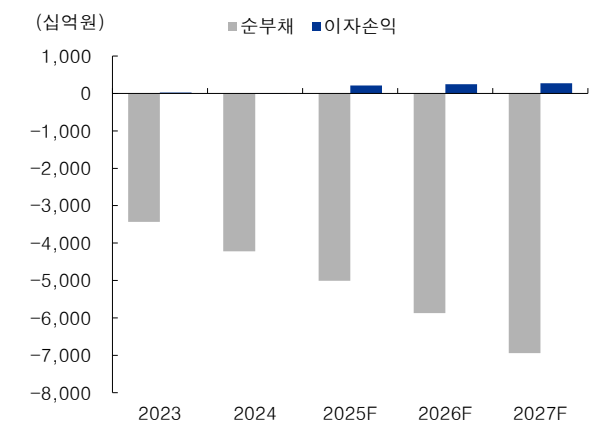
자료: Quantiwise, IBK투자증권

그림 87. 카카오의 영업이익률 및 순이익률 추이



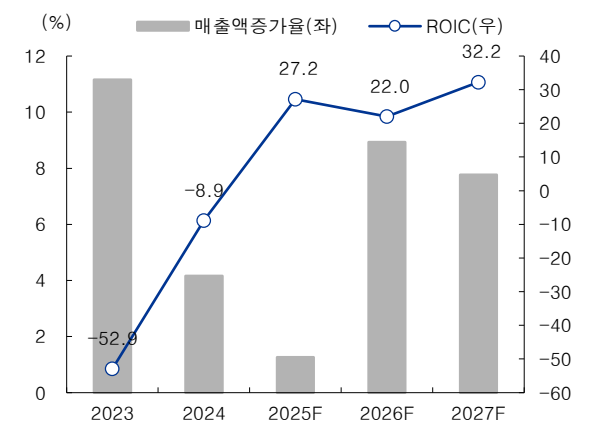
자료: Quantiwise, IBK투자증권

그림 88. 카카오의 순부채 및 이자손익 추이



자료: Quantiwise, IBK투자증권

그림 89. 카카오의 매출성장률과 ROIC 추이



자료: Quantiwise, IBK투자증권

1. AI 전략

AI 대중화,
카카오 국내 집중 전략

올해를 ‘AI 대중화 원년’으로 설정하고, 자사 핵심 플랫폼에 AI 기술을 본격적으로 접목하는 전략을 추진 중이다. 올해 AI 전략의 핵심을 국내 시장에 두고, 해외 진출보다는 국내 사용자 경험의 고도화에 집중하고 있다. 카카오톡, 카카오페이, 카카오프렌즈 등 주요 생활 플랫폼에 AI 기술을 접목해 서비스의 품질과 사용 편의성을 높이는 데 주력 중이다. ‘AI 대중화’를 주요 방향으로 설정하고, 누구나 일상에서 손쉽게 AI를 활용할 수 있는 환경을 구축하는 데 초점을 맞추고 있다.

2. AI 서비스

2-1. AI 에이전트(오픈AI)

오픈AI와 손잡은 카카오,
에이전트 개발 박차

2025년 2월 ChatGPT 개발사인 오픈AI와 파트너십을 체결하고, 공동 AI 에이전트 개발에 착수한 바 있다. 2025년 11월 출시 목표를 공식화했다. 이번 제휴를 통해 카카오는 AI 기반의 커뮤니케이션 기술 고도화와 함께, 자사 서비스 전반에 걸친 AI 확산 및 사용자 경험 개선이 기대된다. 또한 국내 시장 중심의 AI 생태계 내 입지를 강화하는 동시에, 글로벌 AI 기술 기업과의 전략적 협력을 통해 기술력 내재화 및 중장기 플랫폼 경쟁력 제고가 가능할 것으로 판단된다.

그림 90. 카카오 정신아 대표와 오픈AI CEO 샘 알트먼



자료: 조선일보, IBK투자증권

2-2. AI 지도

AI 기반 장소 추천 기능인 ‘AI메이트 로컬’을 통해 사용자가 원하는 조건을 자연어로 입력하면, 위치, 가격, 메뉴, 주차 가능 여부 등 다양한 기준에 맞춰 맛집을 추천해주는 서비스를 운영 중이다. 향후 대화형 AI 에이전트가 지도와 결합되면 장소 추천을 넘어 예약, 결제, 길 안내 등 사용자의 목적을 한 번에 해결하는 통합형 서비스로 진화할 수 있다.

2-3. 생성형 AI 검색

AI가 문맥을 추론해 최적의 답변을 요약·정리하는 생성형 AI 검색 서비스를 준비 중이다. 카카오톡, 카카오톡스토리 등 콘텐츠 플랫폼에도 AI 기반 추천, 자동 요약, 이미지·음성 생성 등 다양한 AI 기능을 적용하고 있다.

2-4. 카카오톡빌리티

2025년 5월 카카오톡빌리티는 ‘구글 I/O 2025’에서 카카오 T 쿱·배송 서비스에 적용된 ‘AI 주소 자동 붙여넣기’ 기능을 소개했다. 사용자가 메시지에서 주소, 수신인, 연락처 등이 포함된 긴 텍스트를 복사한 후 카카오 T 쿱·배송 화면을 열면, AI가 필요한 정보만 자동 추출해 입력하는 기능이다. 해당 기능은 지난 2월 적용 이후 서비스 접수 소요 시간을 24% 단축, 신규 이용자의 접수 완료율도 13.39%p 향상시키며 실제 사용성 개선 효과를 입증했다. 이와 함께, 카카오톡빌리티는 이용자의 사용 패턴과 예약 이력을 분석해 이동수단 및 목적지 자동 추천, 서비스 흐름 최적화 등 자체 AI 기반 기능도 지속 고도화하고 있다.

3. AI 인프라 - 남양주 디지털 허브 구축

남양주에 6,000억 규모
디지털 허브 구축

- 6월 경기도청에서 열린 투자 협약식(MOU)에 경기도 남양주 왕숙 도시첨단산업단지 내 약 6,000억 원 규모의 디지털 허브를 조성하기로 공식 발표했다. 이는 안산 데이터센터에 이은 카카오의 두 번째 자체 데이터센터로, 2029년 준공을 목표로 한다.

그림 91. 카카오-남양주시, ‘디지털 허브’ 조성 위한 업무협약 체결



자료: 카카오, IBK투자증권

카카오 (035720)

포괄손익계산서

(십억원)	2023	2024	2025F	2026F	2027F
매출액	7,557	7,872	7,971	8,682	9,357
증가율(%)	11.2	4.2	1.3	8.9	7.8
매출원가	0	0	0	0	0
매출총이익	7,557	7,872	7,971	8,682	9,357
매출총이익률 (%)	100.0	100.0	100.0	100.0	100.0
판매비	7,096	7,411	7,422	7,960	8,426
판매비율(%)	93.9	94.1	93.1	91.7	90.1
영업이익	461	460	548	723	931
증가율(%)	-19.1	-0.1	19.2	31.8	28.7
영업이익률(%)	6.1	5.8	6.9	8.3	9.9
순금융손익	9	24	144	121	141
이자손익	27	3	215	251	276
기타	-18	21	-71	-130	-135
기타영업외손익	-2,122	-541	76	0	0
종속/관계기업손익	4	54	-161	-98	-49
세전이익	-1,648	-3	607	747	1,023
법인세	168	159	94	183	251
법인세율	-10.2	-5,300.0	15.5	24.5	24.5
계속사업이익	-1,817	-162	513	564	772
중단사업손익	0	0	0	0	0
당기순이익	-1,817	-162	513	564	772
증가율(%)	-270.3	-91.1	-417.0	9.9	36.9
당기순이익률 (%)	-24.0	-2.1	6.4	6.5	8.3
지배주주당기순이익	-1,013	55	427	507	713
기타포괄이익	102	359	117	0	0
총포괄이익	-1,715	197	630	564	772
EBITDA	1,222	1,296	1,081	1,183	1,410
증가율(%)	6.1	6.1	-16.5	9.4	19.2
EBITDA마진율(%)	16.2	16.5	13.6	13.6	15.1

투자지표

(12월 결산)	2023	2024	2025F	2026F	2027F
주당지표(원)					
EPS	-2,276	124	966	1,148	1,613
BPS	21,948	22,858	24,067	25,154	26,707
DPS	61	68	61	61	61
밸류에이션(배)					
PER	-23.9	306.9	60.4	50.8	36.1
PBR	2.5	1.7	2.4	2.3	2.2
EV/EBITDA	20.3	12.8	22.9	20.2	16.3
성장성지표(%)					
매출증가율	11.2	4.2	1.3	8.9	7.8
EPS증가율	-174.7	-105.5	675.7	18.9	40.5
수익성지표(%)					
배당수익률	0.1	0.2	0.1	0.1	0.1
ROE	-10.3	0.6	4.1	4.7	6.2
ROA	-7.5	-0.6	2.0	2.1	2.7
ROIC	-52.9	-8.9	27.2	22.0	32.2
안정성지표(%)					
부채비율(%)	81.7	84.8	83.0	85.8	87.7
순차입금 비율(%)	-24.7	-30.3	-34.2	-38.7	-43.6
이자보상배율(배)	2.8	2.3	0.0	0.0	0.0
활동성지표(배)					
매출채권회전율	14.2	13.6	29.6	0.0	0.0
재고자산회전율	50.6	77.3	245.6	0.0	0.0
총자산회전율	0.3	0.3	0.3	0.3	0.3

*주당지표 및 밸류에이션은 지배주주순익 및 지배주주지분 기준

재무상태표

(십억원)	2023	2024	2025F	2026F	2027F
유동자산	9,988	10,959	11,478	12,525	13,801
현금및현금성자산	5,389	6,145	6,936	7,774	8,827
유가증권	1,770	1,584	1,763	1,908	2,066
매출채권	622	538	0	0	0
재고자산	139	65	0	0	0
비유동자산	15,192	14,814	15,315	15,671	16,082
유형자산	1,337	1,286	1,264	1,308	1,375
무형자산	5,769	5,137	5,545	5,720	5,923
투자자산	5,365	5,726	5,647	5,671	5,697
자산총계	25,180	25,773	26,793	28,196	29,883
유동부채	7,564	8,633	6,969	7,543	8,165
매입채무및기타채무	242	227	0	0	0
단기차입금	1,053	1,458	1,523	1,648	1,784
유동성장기부채	100	0	0	0	0
비유동부채	3,757	3,197	3,372	3,513	3,672
사채	0	0	0	0	0
장기차입금	846	328	378	378	378
부채총계	11,321	11,830	12,155	13,021	13,963
지배주주지분	9,764	10,141	10,638	11,119	11,805
자본금	45	44	44	44	44
자본잉여금	8,840	8,911	8,904	8,904	8,904
자본조정등	8	-34	-33	-33	-33
기타포괄이익누계액	-1,050	-724	-619	-619	-619
이익잉여금	1,922	1,943	2,342	2,822	3,509
비지배주주지분	4,095	3,802	4,000	4,056	4,115
자본총계	13,859	13,943	14,638	15,175	15,920
비이자부채	7591	8326	8468	9208	10014
총차입금	3,730	3,504	3,687	3,813	3,949
순차입금	-3,430	-4,225	-5,011	-5,869	-6,944

현금흐름표

(십억원)	2023	2024	2025F	2026F	2027F
영업활동 현금흐름	1,341	1,250	1,685	1,267	1,453
당기순이익	-1,817	-162	513	564	772
비현금성 비용 및 수익	3,001	1,462	506	436	387
유형자산감가상각비	489	576	389	315	330
무형자산상각비	271	260	144	145	150
운전자본변동	248	-4	541	16	17
매출채권등의 감소	-108	-20	551	0	0
재고자산의 감소	43	0	76	0	0
매입채무등의 증가	0	0	0	0	0
기타 영업현금흐름	-91	-46	125	251	277
투자활동 현금흐름	-1,780	10	-945	-1,022	-1,111
유형자산의 증가(CAPEX)	-575	-380	-227	-360	-396
유형자산의 감소	14	32	6	0	0
무형자산의 감소(증가)	-135	-87	-149	-320	-352
투자자산의 감소(증가)	-39	-16	-29	-24	-26
기타	-1045	461	-546	-318	-337
재무활동 현금흐름	1,020	-521	7	594	712
차입금의 증가(감소)	199	225	50	0	0
자본의 증가	0	0	-10	5	7
기타	821	-746	-33	589	705
기타 및 조정	-28	17	19	-1	-1
현금의 증가	553	756	766	838	1,053
기초현금	4,836	5,389	6,170	6,936	7,774
기말현금	5,389	6,145	6,936	7,774	8,827

Not Rated

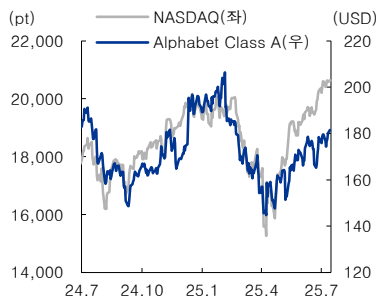
산업 분류(GICS) 소프트웨어
상장거래소 Nasdaq

현재가 (7/14) 181.56달러
시가총액(USD) 12,121억 달러
시가총액(원) 1,766,997십억 원
52주 최고가 207.05억 달러

주주구성
The Vanguard Group, Inc. 8.72%
Page (Lawrence E) 6.59%

	1M	6M	12M
주가상승			
상대기준	3.9%	-4.3%	-1.9%
절대기준	-2.4%	-12.7%	-14.1%

알파벳 상대주가 (%)



알파벳 (GOOGL US)

AI 패러다임 전환의 선도 기업

Gemini, 제품 내재형 AI로 진화

Gemini는 동사의 AI 서비스로, 2025년부터 본격적으로 상용화되었다. 동사의 대부분 서비스에 적용되기 시작하여 검색(AI Overview)에서는 사용자의 질문에 응답을 제공한다. Nest 카메라, Google Home 앱에서는 상황을 인식하여 반응하고, Google Vids, Docs, Classroom 등 서비스에도 통합되고 있다.

최근 경량화된 Gemini 모델을 자사 하드웨어에 내재화 중이다. 픽셀(Pixel) 스마트폰에서 음성 인식, 스마트 카메라, 실시간 요약 기능을 온디바이스로 구현했다. 클라우드 의존도를 낮추는 동시에 배터리 효율성을 제고했다. 이러한 제품 내재형(Product-embedded) AI 전략은 Gemini를 단순 AI 서비스가 아닌, 실시간 반응형 AI Agent로 진화시킬 것이다.

Physical AI를 통한 인터페이스 확장

2025년 5월 Google I/O(개발자 컨퍼런스)에서 Gemini 기반 XR 스마트클래스 시제품을 공개했다. 해당 디바이스는 삼성전자, Xreal, Warby Parker, 젠틀몬스터 등과의 협력 하에 개발되었으며, 실시간 번역, 내비게이션, 미디어 제어, 질의응답 등 기능을 수행한다. Android XR 플랫폼에서 구동되는 이 기기는 초저지연 추론, 시선 추적 기반 UI 설계를 바탕으로, 사용자와의 물리적 상호작용을 실시간 처리하는 Physical AI의 시발점이 될 것이다.

공격적인 AI 인프라 투자 단행

2025년 약 750억 달러의 자본 지출(CapEx)을 통해 AI 인프라 확장에 박차를 가하고 있다. 주로 북미 지역 데이터센터 확장, 서버 최적화, 구글 독자 개발 AI 가속기(TPU) 확보에 쓰일 예정이다.

2Q에도 안정적인 성장세 예상

2025년 2분기 매출액은 938.2억 달러로 전년 동기 대비 +10.7%, 전분기 대비 +4.0% 증가할 것으로 예상된다. 영업이익은 309.1억 달러로 전년 동기 대비 12.7%, 전분기 대비 +1.0% 증가가 전망된다. 전반적으로 견고한 실적 성장세가 유지될 것으로 보인다.

(단위:백만달러,배)	2020	2021	2022	2023	2024
매출액	182,530	257,640	282,840	307,390	350,020
영업이익	41,220	78,710	74,840	84,290	112,390
세전이익	48,080	90,730	71,330	85,720	119,820
지배주주순이익	40,270	76,030	59,970	73,800	100,120
EPS(원)	3	6	5	6	8
증가율(%)	19	91	-19	27	39
영업이익률(%)	23	31	26	27	32
순이익률(%)	22	30	21	24	29
ROE(%)	19	32	24	27	33
PER	30	26	19	24	24
PBR	5	8	4	6	7
EV/EBITDA	19	20	12	17	17

자료: Company data, IBK투자증권

구글의 Physical AI 도약 준비

네스트 카메라 x Gemini: 더 똑똑한 홈 모니터링

네스트, 문맥 이해하는
AI 카메라로 진화

구글 네스트 카메라에 Gemini AI가 통합되면서 홈 모니터링 시스템이 진화했다. Gemini의 멀티모달 능력 덕분에 카메라는 단순 움직임 감지를 넘어, 문맥을 심층적으로 이해하게 되었다. 예를 들어, 네스트 카메라는 사용자에게 "개가 정원에서 땅을 파고 있다"고 알람을 보내기도 하고, 사용자가 "아이들이 오늘 오후에 뒷마당에서 놀았니?"와 질문을 하면 원하는 녹화 기록을 빠르게 찾아낼 수 있다.

스마트 홈 자동화도 가능해졌다. Gemini는 사용자의 복잡한 자연어 명령을 이해하여 여러 기기를 동시에 제어할 수 있다. 예시로, 네스트 카메라가 집 안 움직임이 없는 것을 감지하고, 사용자의 스마트폰 위치가 집 밖으로 벗어난 것을 확인하면, Gemini가 자동으로 "집을 비우신 것 같아요. 현관문 잠금 상태를 확인하고, 조명을 모두 끈 후, 로봇 청소기를 작동할까요?"라고 제안할 수 있다. 또한, 네스트 도어벨 카메라가 방문객을 감지하면, Gemini는 "손님이 오신 것 같아요. 현관 조명을 켜고, 거실 커튼을 열어두고, 잔잔한 음악을 틀어둘까요?"라고 제안할 수 있다.

그림 92. Nest 카메라



자료: Google, IBK투자증권

구글 홈 x Gemini: 스마트 중앙 제어 시스템

구글홈,
Gemini로 맞춤형 루틴 제공

구글 홈(Google Home, 구글의 인공지능 기반 스마트 스피커)에 Gemini가 통합되면서, 훨씬 더 지능적이고 개인화된 중앙 관리 시스템으로 변모했다.

- Gemini는 사용자의 패턴, 시간, 날씨, 의 정보까지 종합적으로 분석하여 최적의 자동화 루틴을 제안한다. 예를 들어, 외출 준비 중임을 감지하면 자동으로 조명을 끄고 문을 잠그는 등의 맞춤형 제안을 제공한다.
- Gemini는 여러 제조사의 스마트 기기를 연동하는 허브 역할을 한다. 이는 사용자가 여러 앱을 오갈 필요 없이 한 번에 관리할 수 있도록 해준다.

그림 93. Google Home



자료: Google, IBK투자증권

XR 글래스 x Gemini: 초저지연 추론 디바이스

Gemini XR 글래스,
실시간 AI 구현

2025년 5월 Google I/O에서 공개된 Gemini 기반 XR 스마트글래스는, 구글이 실감형 인터페이스(Embodied Interface) 시장에서 재도전에 나섰음을 상징한다. 과거 Google Glass의 실패를 넘어선 이번 플랫폼은, 삼성전자, Xreal, Warby Parker, 젠틀몬스터 등과의 전략적 파트너십을 통해 설계되었다. 해당 디바이스는 실시간 번역, 질의응답, 내비게이션, 미디어 재생 등 다양한 기능을 Gemini 기반으로 수행하며, Android XR 플랫폼 상에서 작동한다.

이 XR 글래스는 초저지연 추론(Low Latency Inference), 시선 추적 기반 UX, 프라이버시 보존형 설계 등 고도화된 기술 스택을 요구하며, 구글의 AI 기술력이 총체적으로 집약된 결과물이다. 이는 실시간 물리적 상호작용을 필요로 하는 Physical AI의 전형적 사례로, 구글은 이를 통해 공간 인식형 컴퓨팅(Spatial Computing), 스마트 내비게이션, 증강 작업지원 등 다양한 산업군에서 새로운 사용자 경험을 창출할 계획이다.

그림 94. XR 스마트글래스



자료: Google, IBK투자증권

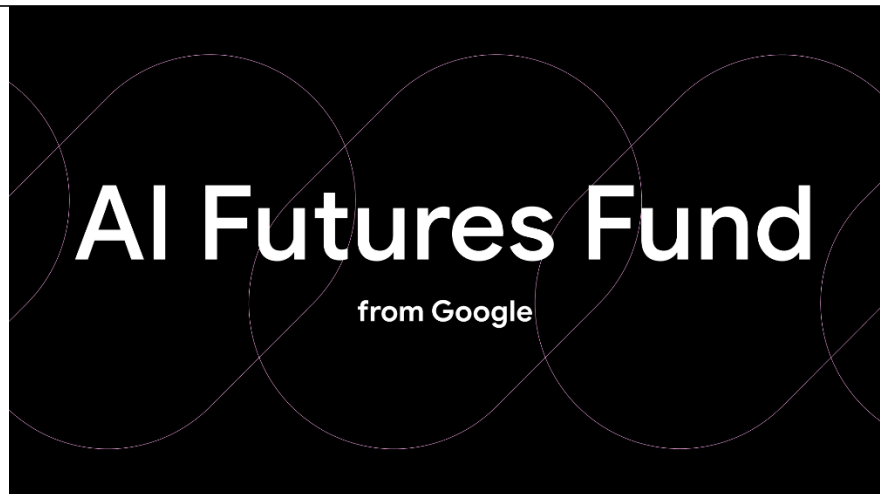
구글의 적극적 선제적 투자

AI 인프라 및 인재 투자

AI 인프라에 750억 달러 투자

구글은 2025년 약 750억 달러의 CapEx를 통해 AI 및 클라우드 인프라 확장을 가속화하고 있다. 주요 투자 항목은 북미 지역 데이터센터 증설, 서버 최적화, 맞춤형 AI 가속기 칩(TPU) 확보 등으로 구성된다. 이는 전년 대비 약 43% 증가한 수치이다. 한편, DeepMind와 협력해 대규모 AI 툴킷을 산업 전반에 배포하고 있으며, 'AI Futures Fund'를 통한 생성형 AI 스타트업 투자, 'AI Opportunity Fund'를 통한 100만 명 이상의 AI 기초 교육 무상 제공 등을 통해 AI 생태계의 확장성과 사회적 수용성을 동시에 확보하려 하고 있다. 인재 확보 측면에서는 Character.AI 등 외부 연구조직 출신의 핵심 인재를 재영입해 내부 역량을 제고하고 있다.

그림 95. XR 스마트글래스

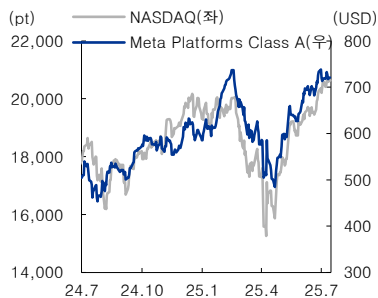


자료: Google, IBK투자증권

Not Rated

산업 분류(GICS)	소프트웨어
상장거래소	Nasdaq
현재가 (7/14)	720.92달러
시가총액(USD)	18,126억 달러
시가총액(원)	2,642,470십억 원
52주 최고가	747.9달러
주주구성	
The Vanguard Group, Inc.	8.76%
Fidelity Management & Research Company LLC	6.01%
주가상승	1M 6M 12M
상대기준	5.6% 21.3% 44.5%
절대기준	-0.8% 12.9% 32.3%

메타 상대주가 (%)



메타 (META US)

AI 오픈소스 진형의 만형

강력한 AI 모델, Llama 생태계 확장

동사는 Llama 시리즈를 중심으로 AI 생태계 확대에 집중하고 있다. 최근 공개된 Llama 4는 텍스트뿐 아니라 이미지, 오디오 등 다양한 데이터를 통합적으로 처리하는 멀티모달 기능이 특징이다. Scout과 Maverick이라는 두 모델을 통해 긴 문맥 처리, 효율성, 범용성까지 동시에 개선했다. 특히 Llama를 오픈소스로 공개하며 개발자 커뮤니티와의 시너지를 극대화하고 있다. 2024년 3월 기준 Llama 다운로드 수는 10억 회를 돌파했으며, 같은 해 8월 Llama 3.1도 오픈소스로 공개됐다.

이러한 생태계 확대는 커뮤니티 기반의 확산 전략을 중시하고 있음을 보여준다. Hugging Face, AWS, Azure 등 주요 AI 플랫폼과의 통합도 적극 추진 중이다. 동사는 오픈소스를 기반으로 학습, 튜닝, 배포까지 전 주기를 사용자에게 개방하고 있다.

AI 클래스를 통한 현실 확장

Ray-Ban과 협업한 스마트 글래스를 통해 AI 기술의 실생활 적용을 가속화하고 있다. 이 제품에는 동사의 AI 챗봇이 내장돼 있으며, 메시지 송수신, 인터넷 검색, 이미지 생성 등의 기능을 지원한다. 최근에는 시각 인식, 실시간 번역 기능이 추가되면서 현실과 AI의 경계가 더욱 좁아지고 있다.

사용자가 특정 사물을 바라보면 AI가 이를 인식하고 정보를 제공하며, 시각장애인을 위한 'Be My Eyes' 기능도 탑재되어 있다. 언어 번역은 영어, 스페인어, 프랑스어, 이탈리아어 간에 실시간으로 가능하다. 스마트 메모리, 음성 명령 등도 꾸준히 고도화되고 있으며, 동사는 AI 클래스를 통해 탈스크린 기반의 새로운 컴퓨팅 패러다임을 실현하고 있다.

외형 성장에 집중

매출액은 446.1억 달러(YoY +14.2%, QoQ +5.4%)로 증가가 예상되나, 영업이익은 168.9억 달러(YoY +13.7%)로 전분기 대비 -3.8% 감소할 전망이다. 대규모 투자가 감행되는 만큼 외형 성장에 집중하는 구간으로 해석된다.

(단위: 백만달러, 배)	2020	2021	2022	2023	2024
매출액	182,530	257,640	282,840	307,390	350,020
영업이익	41,220	78,710	74,840	84,290	112,390
세전이익	48,080	90,730	71,330	85,720	119,820
지배주주순이익	40,270	76,030	59,970	73,800	100,120
EPS(원)	3	6	5	6	8
증가율(%)	19	91	-19	27	39
영업이익률(%)	23	31	26	27	32
순이익률(%)	22	30	21	24	29
ROE(%)	19	32	24	27	33
PER	30	26	19	24	24
PBR	5	8	4	6	7
EV/EBITDA	19	20	12	17	17

자료: Company data, IBK투자증권

오픈소스 생태계 확장

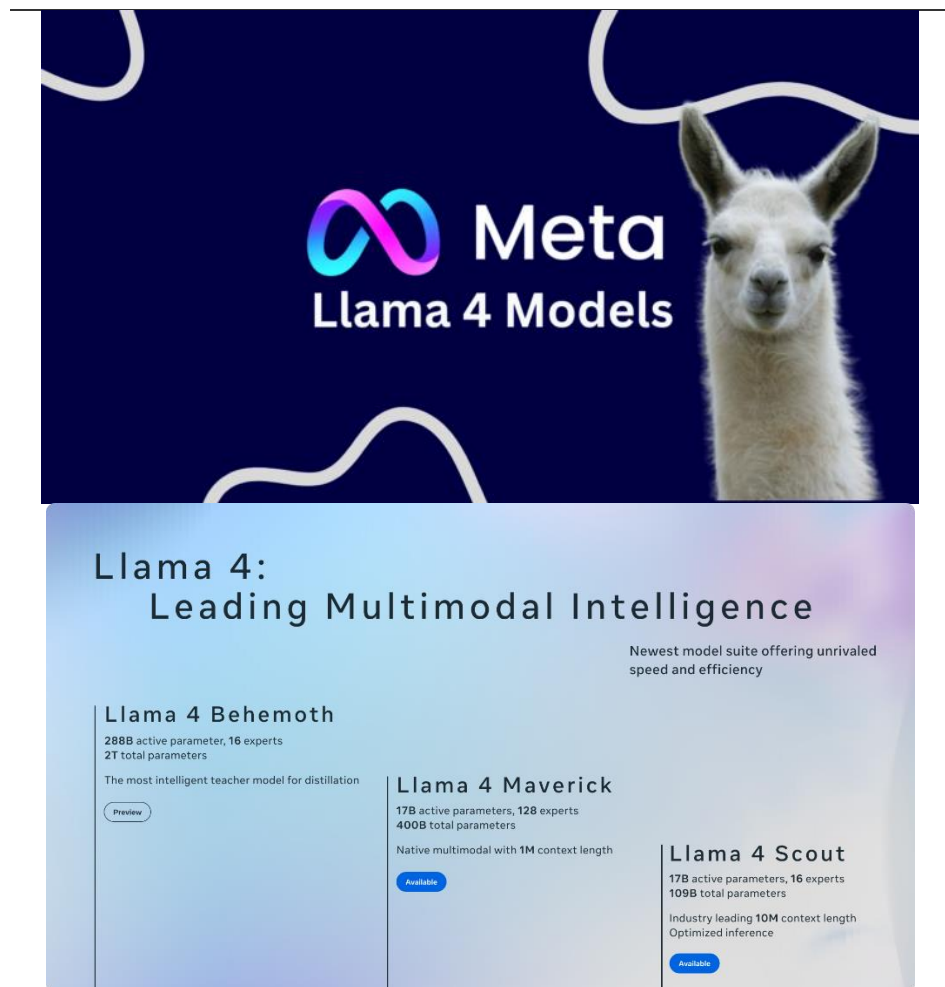
Llama 4 Scout와 Maverick

Llama 3.1로 오픈소스
생태계 가속

메타는 인공지능(AI) 분야에서 자사의 대규모 언어 모델(LLM)인 Llama 시리즈를 중심으로 강력한 생태계를 구축하고 있다. 특히 지난 4월 공개된 Llama 4는 텍스트뿐만 아니라 이미지, 오디오 등 다양한 데이터를 처리하고 통합하는 멀티모달 기능을 강화하였다. Llama 4는 Scout와 Maverick 두 가지 모델을 선공개하며 긴 문맥 처리와 효율성, 광범위한 작업에서의 높은 성능을 동시에 추구한다.

더욱이 메타는 Llama 모델을 오픈소스로 지속 제공하며 전 세계 개발자들이 AI 개발에 참여하고 혜택을 누릴 수 있도록 독려하고 있다. 이는 오픈AI나 구글이 자사 모델을 비공개로 유지하는 것과는 차별화된 전략으로, 지난 3월 Llama 모델 다운로드 수가 10억 회를 돌파하며 오픈소스 전략의 성공 가능성을 입증하였다. 2024년 8월에는 Llama 3.1을 오픈소스로 공개하며 생태계 확장에 대한 의지를 재확인하였다.

그림 96. 메타의 Llama 4 모델



자료: Meta, IBK투자증권

AI 글래스로 인공지능-현실 연결

스마트 글래스 출시

레이밴 · 오클리 · 프라다,
스마트 글래스 3종 출동

Ray-Ban Meta 스마트 글래스를 통해 AI 기술의 실생활 적용을 선도하고 있다. 이 스마트 글래스에는 메타의 AI 챗봇이 통합되어 WhatsApp 등에서 메시징에 접근하고, 실시간 인터넷 액세스와 이미지 생성 기능을 갖춘 멀티태스킹 챗봇 기능을 제공한다. 특히 최근 업데이트를 통해 AI 비디오 분석 및 실시간 번역 기능이 추가되었다. 사용자가 글래스를 통해 특정 물체를 바라보면 AI가 질문에 답변하거나, 공원에서 나무의 종류를 알려주는 등 증강현실 기반의 정보 제공이 가능하다. 또한 영국에서는 영어, 스페인어, 이탈리아어, 프랑스어 간의 실시간 음성 번역 기능이 도입되기 시작하여 언어 장벽을 허무는 데 기여할 것으로 기대된다. 스마트 메모리, 핸드프리 음성 명령, Be My Eyes 네트워크 연결을 통한 자원봉사자 도움 등 다양한 AI 기능이 지속적으로 업데이트되며, 메타는 AI 글래스를 통해 현실과 AI를 매끄럽게 연결하는 비전을 실현하고 있다.

오클리-메타 AI 스마트 글래스 'HSTN'

오클리-메타 AI 스마트 글래스 'HSTN'은 특히 스포츠 활동에 최적화된 AI 글래스로 주목받고 있다. 오클리의 고유한 디자인과 메타의 첨단 AI 기술이 결합된 이 제품은 렌즈 내에 디스플레이가 없으면서도 다양한 스마트 기능을 제공한다. 프레임에 내장된 카메라, 마이크, 오픈형 스피커를 통해 스마트폰과 연동하여 음악 감상, 통화, 그리고 메타 AI와의 음성 대화가 가능하다. 특히 AI는 사용자 시야를 기반으로 실시간 정보 제공이나 언어 번역까지 수행할 수 있다. 3K 초고화질 카메라를 탑재하여 활동 영상을 선명하게 기록할 수 있으며, IPX4 등급의 생활 방수를 지원하여 운동 중에도 걱정 없이 사용할 수 있다. 배터리 사용 시간은 기존 레이밴 메타 글래스의 2배인 최대 8시간으로 늘었으며, 충전 케이스 사용 시 최대 48시간까지 사용 가능하다. 디자인은 오클리의 다양한 프레임과 렌즈 조합으로 구성되며, 도수 렌즈 호환도 가능하여 사용자 편의성을 높였다.

프라다-메타 AI 스마트 글래스

한편, 프라다와의 협업을 통한 스마트 글래스는 메타가 웨어러블 기술을 패션 업계 전반으로 확장하려는 첫 번째 사례가 될 것으로 보인다. 아직 구체적인 출시 일정과 상세 기능은 공개되지 않았지만, 명품 브랜드 프라다의 디자인 아이덴티티와 메타의 AI 기술이 결합될 것으로 예상되어 기대를 모으고 있다. CNBC 보도에 따르면, 프라다 안경은 두꺼운 템플(다리 부분)로 유명하여 마이크와 칩 등 스마트 안경에 필요한 부품을 수용하기에 이상적인 것으로 평가된다. 이는 프라다 스마트 글래스가 기존 레이밴이나 오클리 모델과는 또 다른 차원의 디자인 및 기능적 특징을 가질 수 있음을 시사한다. 메타는 이러한 패션 브랜드와의 협력을 통해 AI 스마트 글래스를 대중적인 패션 아이템이자 일상생활의 필수 웨어러블 기기로 자리매김하려는 전략을 펼치고 있다.

그림 97. 메타X레이벤 스마트 글래스



자료: Meta, IBK투자증권

그림 98. 6월 20일 공개된 메타X오클리 스마트 글래스



자료: Meta, IBK투자증권

메타의 AI 인프라 투자

AI 인프라 투자와 인재 확보

Scale AI에 140억 달러 투자
AI 인프라 투자도 가속

메타는 AI 경쟁력을 확보하기 위해 공격적인 투자를 단행하고 있다. 2024년 6~7월에는 Scale AI에 약 140억 달러를 투자했고, CEO 알렉산더 왕 등 핵심 인력을 일부 영입했다. 이는 메타의 외부 투자 중 최대 규모로, 내부 모델 학습 인프라를 자급화하려는 전략으로 해석된다. 투자 직후 오픈AI와 구글은 Scale AI와의 협력을 중단했다.

인프라 투자도 병행 중이다. 새 AI 데이터센터 건설을 위해 약 286조 원(2,000억 달러)을 조달하고 있으며, 미국과 유럽에 AI 슈퍼컴퓨터도 구축하고 있다. GPU는 NVIDIA H100과 B200 기반으로 구성되고 있다. 오픈AI 출신 연구자 영입에도 속도를 내고 있으며, 최고 1억 달러에 달하는 보상 패키지를 제시하고 있다.

그림 99. Scale AI CEO, 알렉산더 왕



자료: Google, IBK투자증권

그림 100. 미국 일리노이주 디캘브 205만㎡ 부지에 조성된 메타의 데이터센터

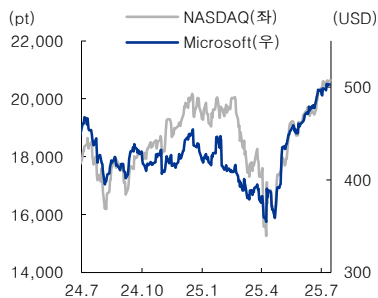


자료: 연합뉴스, IBK투자증권

Not Rated

산업 분류(GICS)	소프트웨어
상장거래소	Nasdaq
현재가 (7/8)	503.02달러
시가총액(USD)	37,387억 달러
시가총액(원)	5,450,343십억 원
52주 최고가	506.78달러
주주구성	
The Vanguard Group, Inc.	9.26%
BlackRock Institutional Trust	4.99%
주가상승	1M 6M 12M
상대기준	5.9% 21.0% 10.9%
절대기준	-0.5% 12.6% -1.3%

마이크로소프트 상대주가 (%)



마이크로소프트 (MSFT US)

윈도우와 오피스는 LMM과 최상의 공합

Copilot 솔루션으로 차별화

동사는 LLM(대규모 언어 모델) 분야에서 독보적인 경쟁력을 확보하고 있다. ChatGPT 개발사인 OpenAI에 대한 지속적인 투자를 통해 GPT-4와 같은 최신 LLM을 자사 제품에 광범위하게 통합하고 있다. 이는 Microsoft 365 Copilot, Microsoft Security Copilot, GitHub Copilot 등 다양한 코파일럿 솔루션으로 이어져 사무 생산성, 보안, 개발 등 광범위한 분야에서 AI 활용을 극대화한다. 또한, Azure AI 모델 카탈로그를 통해 다양한 산업 파트너들과 협력하여 금융 서비스, 소매, 헬스케어 등 산업별 맞춤형 AI 모델과 솔루션을 제공하며 차별화된 가치를 창출하고 있다. Rezo.ai와의 파트너십을 통해 LLM 효율성을 높이고, Skeleton-of-Thought와 같은 연구를 통해 LLM의 추론 능력 및 응답 시간 개선에 힘쓰는 점도 주목할 만하다.

AI 수익 모델 정착

Azure 클라우드를 통해 AI 모델을 구축하고 배포하며, Microsoft 365, Dynamics 365 등 기존 소프트웨어에 AI 기능을 통합하고 있다. GE Aerospace, Starbucks 등 글로벌 기업들이 동사의 Azure OpenAI Service 및 Microsoft 365 Copilot을 활용하여 비즈니스 효율성을 높이고 있다.

윈도우 PC로 Physical AI 역량 견비

Copilot+ PC는 온디바이스 AI 기능을 활용하여 기존 PC보다 더 빠르고 효율적으로 AI를 활용하도록 설계되었다. 특히 퀄컴의 스냅드래곤 X 엘리트 (Snapdragon X Elite) 및 스냅드래곤 X 플러스(Snapdragon X Plus) 프로세서를 탑재하여, NPU(신경망 처리 장치)를 통해 클라우드 연결 없이도 AI 작업을 수행할 수 있다.

AI 수익화로 2Q 실적도 맑음

매출액은 738억 달러로 전년 동기 대비 14.0%, 전분기 대비 +5.3% 증가할 것으로 예상된다. 영업이익은 321.4억 달러로 전년 동기 대비 +15.2%, 전분기 대비 +0.5% 증가가 전망된다. GE Aerospace, Starbucks 등 글로벌 기업이 Azure OpenAI Service 및 Microsoft 365 Copilot을 활용하고 있는 만큼 이익 개선세 지속을 예상한다.

(단위: 백만달러, 배)	2020	2021	2022	2023	2024
매출액	182,530	257,640	282,840	307,390	350,020
영업이익	41,220	78,710	74,840	84,290	112,390
세전이익	48,080	90,730	71,330	85,720	119,820
지배주주순이익	40,270	76,030	59,970	73,800	100,120
EPS(원)	3	6	5	6	8
증가율(%)	19	91	-19	27	39
영업이익률(%)	23	31	26	27	32
순이익률(%)	22	30	21	24	29
ROE(%)	19	32	24	27	33
PER	30	26	19	24	24
PBR	5	8	4	6	7
EV/EBITDA	19	20	12	17	17

자료: Company data, IBK투자증권

AI 솔루션 혁신 전략

차별화된 LLM 솔루션 제공

LLM 기반 코파일럿(Copilot)
전방위 확산

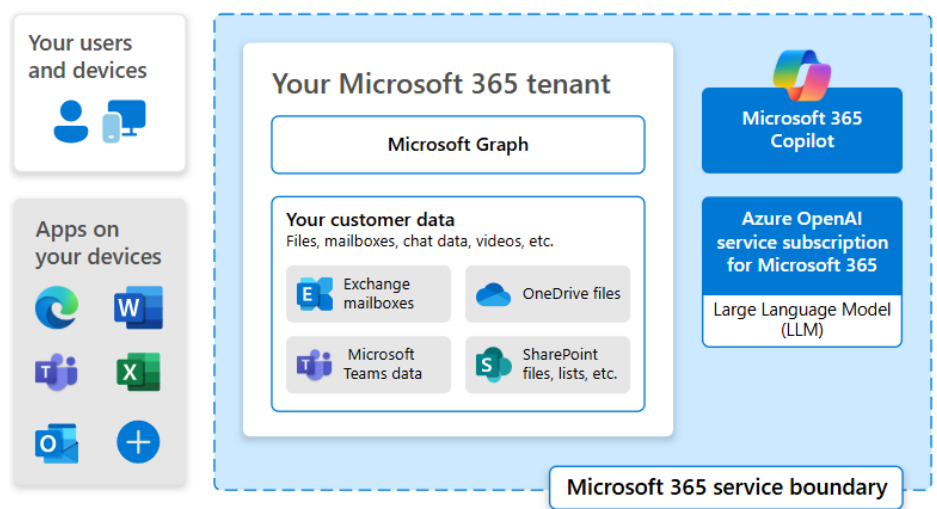
마이크로소프트는 LLM(대규모 언어 모델) 분야에서 독보적인 경쟁력을 확보하며 AI 혁신을 선도하고 있다. ChatGPT 개발사인 OpenAI에 대한 전략적 투자로 GPT-4와 같은 최신 LLM을 자사 제품에 광범위하게 통합했다. 이는 다양한 코파일럿 솔루션으로 이어져 사무 생산성, 보안, 개발 등 전반적인 분야에서 AI 활용을 극대화했다.

Microsoft 365 Copilot은 일상적인 사무 업무의 생산성을 혁신하고, Microsoft Security Copilot은 복잡한 사이버 보안 위협에 대한 방어력을 강화하며, GitHub Copilot은 개발자들의 코딩 효율을 극대화했다. 이처럼 코파일럿은 사무 생산성, 보안, 개발 등 광범위한 분야에서 AI 활용을 극대화하여 사용자들에게 전례 없는 효율성과 지능적인 업무 환경을 제공하고 있다.

나아가 마이크로소프트는 Azure AI 모델 카탈로그를 통해 다양한 산업 파트너들과 협력하며 금융, 소매, 헬스케어 등 산업별 맞춤형 AI 모델과 솔루션을 제공한다. 특히 의료 분야에서는 환자 데이터 분석 및 진단 보조에 LLM을 활용하여 의료진의 효율성을 높였다.

또한, 마이크로소프트는 LLM의 효율성과 성능 개선을 위한 연구 개발에 끊임없이 투자해왔다. Rezoive.ai와의 파트너십으로 LLM의 비즈니스 적용 사례를 발굴하고 효율성을 높이며, Skeleton-of-Thought와 같은 혁신적인 연구를 통해 LLM의 추론 능력 및 응답 시간 개선에 힘쓰고 있다. 이러한 지속적인 노력은 마이크로소프트가 AI 기술의 한계를 넓히고 미래 AI 표준을 제시하려는 강력한 의지를 보여준다고 볼 수 있다.

그림 101. 마이크로소프트의 365 Copilot 기술 구조



자료: 마이크로소프트, IBK투자증권

클라우드 및 소프트웨어 생태계 확장

강력한 AI 시너지 창출

코파일럿+ PC로
온디바이스 강화

마이크로소프트는 강력한 클라우드 및 소프트웨어 생태계를 바탕으로 AI 시너지를 창출하며 시장 지배력을 공고히 하고 있다. 그 핵심은 바로 애저(Azure) 클라우드다. 애저는 전 세계 수많은 기업이 AI 모델을 구축하고 배포하는 데 쓰이는 주요 인프라로, 마이크로소프트는 이곳에 방대한 데이터와 컴퓨팅 자원을 제공한다. 특히 최신 GPU와 AI 칩에 꾸준히 투자하며 컴퓨팅 성능을 극대화하고 있어, AI 모델 학습과 운영에 필요한 탄탄한 기반을 제공하고 있다.

Microsoft 365, Dynamics 365 같은 핵심 소프트웨어 제품에 AI 기능을 통합해 사용자 경험을 혁신하고 있다. 예를 들어, 팀즈(Teams) 같은 협업 도구에 AI 요약 및 번역 기능을 넣어 글로벌 소통을 돕고, 다이내믹스 365 코파일럿(Dynamics 365 Copilot)으로 고객 관계 관리나 전사적 자원 관리 효율을 크게 높이고 있다.

최근 선보인 코파일럿+ PC(Copilot+ PC)는 마이크로소프트 AI 전략의 또 다른 중요한 축이다. 이 PC들은 온디바이스(On-device) AI 기능을 강화해 클라우드 연결 없이도 AI 작업을 수행하도록 설계되었다. 이는 사용자 프라이버시를 높이고, 네트워크 지연 없이 즉각적인 AI 기능을 제공하며 개인 컴퓨팅 경험을 한 단계 끌어올린다. 퀄컴, 인텔 등과의 긴밀한 하드웨어 파트너십을 통해 AI 전용 NPU(신경망 처리 장치)를 탑재한 기기를 늘리며, 소프트웨어와 하드웨어의 유기적인 결합으로 AI 생태계를 더욱 공고히 하는 중이다.

그림 102. 마이크로소프트의 새로운 Copilot+ PC (좌: Surface Pro, 우: Surface Laptop)



자료: 마이크로소프트, IBK투자증권

100조 원 규모 AI 데이터 센터 투자

전방위적 AI 투자로 미래 경쟁력 확보

AI 데이터센터에 100조 원 투입

AI 시대의 주도권을 확고히 잡기 위해 막대한 투자를 단행하고 있다. 특히, 2025 회계연도에만 AI 데이터센터 구축에 800억 달러(약 100조 원) 이상을 투자할 계획이다. 이 투자는 단순히 시설을 늘리는 것을 넘어, AI 모델 학습과 클라우드 기반 애플리케이션 배포에 필수적인 핵심 인프라를 압도적인 규모로 강화하려는 전략적 움직임이다. 이는 빠르게 발전하는 AI 기술에 발맞춰 필요한 컴퓨팅 자원을 선제적으로 확보하고, 전 세계 고객들에게 고성능 AI 서비스를 안정적으로 제공하려는 마이크로소프트의 청사진을 담고 있다.

더불어, AI 기술 교육을 제공하여 미래 AI 시대에 필요한 인재를 선제적으로 확보하고 있다. 이러한 전방위적인 투자는 단순한 기술 개발을 넘어, AI 생태계 전반을 아우르며 미래 경쟁력을 확보하려는 장기적인 비전을 드러낸다.

그림 103. 마이크로소프트의 AI 데이터센터



자료: 마이크로소프트, IBK투자증권

Not Rated

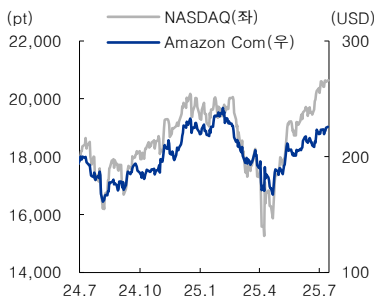
산업 분류(GICS) 소매(유통)
상장거래소 Nasdaq

현재가 (7/14) 225.69달러
시가총액(USD) 23,960억 달러
시가총액(원) 3,492,913십억 원
52주 최고가 242.52억 달러

주주구성
Bezos (Jeffrey P) 8.50%
The Vanguard Group, Inc. 7.80%

주가상승	1M	6M	12M
상대기준	6.4%	3.6%	16.0%
절대기준	0.0%	-4.7%	3.9%

아마존 상대주가 (%)



아마존 (AMZN US)

Physical AI 실현이 가장 빠른 AI 테크 기업

자체 AI 에이전트 공개

동사가 2025년 3월 말 공개한 '노바 액트(Nova Act)'는 동사의 범용인공 지능(AGI) 연구소에서 개발된 첫 AI 에이전트이다. 이 모델은 사용자 명령 없이도 쇼핑, 예약 등 다양한 작업을 자율적으로 수행하며, 기존 경쟁사 모델들을 능가하는 성능을 보여주고 있다.

노바 액트를 자사 음성 비서 '알렉사 플러스'에 통합하여 서비스 경쟁력을 강화할 계획이다. 또한, 2조 개 파라미터 규모의 초대형 LLM '올림푸스(Olympus)' 개발에도 집중 투자하고 있어, LLM 시장의 주요 플레이어로서 입지를 더욱 공고히 할 것으로 예상된다. 이러한 LLM 기술 고도화는 클라우드 서비스인 AWS의 성장뿐만 아니라, 이커머스, 디바이스 등 전 사업 부문에 걸쳐 혁신을 가져올 핵심 동력이다.

물류 및 배송 AI 개척

100만 대 이상의 로봇을 물류 센터에 배치하고 AI 파운데이션 모델 '딥플리트(DeepFleet)'를 활용하여 로봇 움직임을 최적화하고 있다. '키바(Kiva)' 로봇(현재는 아마존 로보틱스)처럼 선반을 통째로 옮기는 로봇부터, 물품을 집어 올리는 로봇 팔, 자율 이동 로봇 등 다양한 형태의 로봇이 물류 센터에서 활약하고 있다. 이를 통해 물류 센터 내 이동 시간을 10% 단축하고 비용을 절감하는 성과를 거두고 있다.

배달 기사용 스마트 글래스는 운전자에게 정확한 길 안내와 상세한 배송 정보를 제공하여 배달 시간을 단축하고 오류를 줄이는 데 기여하고 있다. 또한, 이커머스 부문에서도 제품 특징에 대한 AI 음성 요약 테스트하는 등 고객 편의성을 높이는 AI 기술을 적극적으로 도입하고 있다. 이러한 실생활 서비스 AI 접목은 동사의 핵심 사업 경쟁력을 강화하고 미래 성장을 견인할 것이다.

외형 확장에 주목

매출액은 1,619억 달러로 전년 동기 대비 +9.4%, 전분기 대비 +4.0% 증가할 것으로 예상된다. 영업이익은 166.1억 달러(YoY +13.2%)로 관세 리스크가 잔존하여 전분기 대비로는 -9.8% 감소할 것으로 추정한다. 그럼에도 불구하고 외형 확장에 주목할 필요가 있다고 판단된다.

(단위: 백만달러, 배)	2020	2021	2022	2023	2024
매출액	182,530	257,640	282,840	307,390	350,020
영업이익	41,220	78,710	74,840	84,290	112,390
세전이익	48,080	90,730	71,330	85,720	119,820
지배주주순이익	40,270	76,030	59,970	73,800	100,120
EPS(원)	3	6	5	6	8
증가율(%)	19	91	-19	27	39
영업이익률(%)	23	31	26	27	32
순이익률(%)	22	30	21	24	29
ROE(%)	19	32	24	27	33
PER	30	26	19	24	24
PBR	5	8	4	6	7
EV/EBITDA	19	20	12	17	17

자료: Company data, IBK투자증권

Physical AI 완비 기업

아마존, 앤트로픽 최대주주 등극

앤트로픽 최대주주 등극
Physical AI 청사진

아마존은 AI 스타트업 앤트로픽(Anthropic)에 총 80억 달러를 투자하며 최대주주 지위를 확보했다. 초기 2023년 9월에 12.5억 달러를 투자한 데 이어, 2024년 3월에는 27.5억 달러, 같은 해 11월에는 40억 달러를 추가로 투자했다. 누적 투자액은 경쟁사인 구글의 약 30억 달러를 크게 웃돌며, 아마존이 앤트로픽의 가장 강력한 전략적 파트너이자 자본 후원자임을 입증했다.

앤트로픽, Physical AI 구현 준비 중

앤트로픽의 클로드(Claude)는 주로 언어 모델로서 텍스트 기반의 상호작용과 추론에 강점을 보여왔지만, 최근에는 피지컬 AI(Physical AI) 역량을 강화하며 그 활용 범위를 물리적 세계로 확장하려는 움직임을 보이고 있다. 이는 단순히 텍스트를 이해하고 생성하는 것을 넘어, 로봇이나 다른 물리적 시스템과 상호작용하여 실제 환경에서 작업을 수행하고 문제를 해결하는 능력을 의미한다.

앤트로픽은 클로드 모델을 로봇 공학이나 자율 시스템에 통합하여, AI가 복잡한 물리적 작업을 계획하고 실행하며, 예기치 않은 상황에 유연하게 대처할 수 있도록 연구하고 있다. 이러한 피지컬 AI로의 확장은 클로드가 단순한 대화형 AI를 넘어 산업 자동화, 물류, 제조 등 다양한 현실 세계 애플리케이션에 적용될 수 있는 잠재력을 열어줄 것으로 기대된다.

AWS 중심의 전략적 시너지

아마존은 단순한 자본 투자에 그치지 않고, 자사의 클라우드 플랫폼인 AWS를 앤트로픽의 주요 AI 인프라로 제공하고 있다. 앤트로픽은 AWS의 AI 전용 칩(Trainium, Inferentia)을 활용해 Claude 모델을 학습·배포하고 있으며, 이는 아마존 클라우드의 매출 확대와 AI 경쟁력 강화에 핵심적인 역할을 한다. 아마존은 이번 투자를 통해 AI 인프라 생태계 전반에 걸쳐 주도권을 넓혀가고 있다.

LLM 시장 리더십 강화

고도화된 LLM 기술

AI 에이전트 '노바 액트' 공개
초대형 LLM '올림푸스' 개발 중

인공지능 분야의 핵심 동력인 대규모 언어 모델(LLM) 개발에 전례 없는 박차를 가하고 있다. 2025년 3월 말, 아마존의 범용인공지능(AGI) 연구소에서 개발된 첫 AI 에이전트인 '노바 액트(Nova Act)'가 공개되었다. 이 모델은 사용자 명령을 기다리지 않고도 쇼핑 목록 작성, 레스토랑 예약, 복잡한 문서 생성 등 다양한 작업을 스스로 판단하고 자율적으로 수행하는 능력을 보여주었다. 아마존은 노바 액트가 스크린 텍스트 인식 및 복잡한 상호작용 테스트에서 이미 경쟁사의 주요 AI 에이전트 모델들을 능가하는 성능을 입증했다고 강조한다. 이러한 탁월한 자율성은 아마존의 핵심 음성 비서인 '알렉사 플러스'에 통합되어 서비스 경쟁력을 획기적으로 강화할 계획이다.

나아가 아마존은 LLM의 규모와 역량을 한층 더 끌어올리는 데 집중 투자하고 있다. 2조 개에 달하는 파라미터를 가진 초대형 LLM '올림푸스(Olympus)' 개발에 막대한 자원을 투입하고 있는 것으로 알려져 있다. 올림푸스 개발은 아마존이 단순한 클라우드 인프라 제공자를 넘어, 고성능 AI 모델의 핵심 개발자로서 입지를 확고히 하려는 전략적 움직임으로 해석될 수 있다. 이러한 LLM 기술 고도화는 아마존의 클라우드 서비스인 AWS의 성장 동력을 제공할 뿐만 아니라, 이커머스 플랫폼의 고객 경험 혁신, 디바이스 기능 향상 등 전 사업 부문에 걸쳐 혁신을 가져올 핵심 기반이 될 것이다. 장기적으로 아마존은 LLM을 통해 더욱 개인화되고 지능적인 서비스를 제공하며 사용자 생태계를 더욱 공고히 할 것으로 기대된다.

그림 104. 아마존의 AI 기반 음성비서 '알렉사 플러스'가 탑재된 스피커 '에코'



자료: 아마존, IBK투자증권

AI 물류 로봇으로 Physical AI 현실화

로봇 자동화

물류센터에 로봇
100만 대 가동 중

- 아마존은 현재 100만 대가 넘는 로봇을 전 세계 물류 창고에 배치했다. 덕분에 재고 관리, 물품 이송, 분류, 포장 등 주문 처리의 모든 과정에서 자동화가 가능해졌다.
- '키바(Kiva)' 로봇(현재는 아마존 로보틱스)처럼 선반을 통째로 옮기는 로봇부터, 물품을 집어 올리는 로봇 팔, 자율 이동 로봇 등 다양한 형태의 로봇이 아마존 물류 센터에서 활약하고 있다.

AI 기반 로봇 운영 시스템 구축

딥플리트 도입으로
로봇 운영 최적화

- '딥플리트(DeepFleet)' 도입: 아마존은 2024년 6월, 물류 로봇 운영의 효율성을 극대화하기 위해 생성형 AI 파운데이션 모델인 '딥플리트'를 발표했다. 이 시스템은 물류 네트워크 전반에 걸쳐 로봇들의 움직임을 실시간으로 최적화하고 조정하여, 로봇 이동 시간을 10% 단축하고 운영 비용을 절감하고 있다.
- '프로테우스(Proteus)' 로봇: '프로테우스'는 아마존이 개발한 자율 이동 로봇이다. 물류 창고 내에서 반복적이고 육체적으로 힘든 작업을 대신한다. 이 로봇은 좁은 공간에서 무거운 물건을 옮기거나, 인간 작업자와 협력하여 효율을 높이는 등 다기능을 수행하며, 창고 효율성을 최대 25% 향상시키고 인건비 절감 및 손상 클레임을 줄이는 효과를 가져왔다.
- '벌컨(Vulcan)' 로봇: '벌컨'은 사람의 촉각과 유사한 기능을 가진 로봇 팔이다. 다양한 크기와 모양의 물건을 선반에서 정확하게 집어 올릴 수 있어, 기존 로봇이 처리하기 어려웠던 섬세한 작업을 가능하게 하여 물류 자동화의 범위를 더욱 확장했다.

그림 105. 아마존의 딥플리트(DeepFleet)



자료: 아마존, IBK투자증권

25년 자본 지출 1,000억 달러 증액

AI 인프라 투자 확대

AI 인프라 등 자본 지출
1,000억 달러로 증액

2025년에는 AI 투자를 포함한 총 자본 지출을 1,000억 달러(약 145조 6,300억 원)로 대폭 증액할 계획을 발표했다. 이 막대한 자금 중 상당 부분은 핵심 사업인 AWS(Amazon Web Services)의 AI 인프라 확충에 집중적으로 투입될 예정이다. 특히, 고성능 AI 모델 학습과 추론에 필요한 GPU 클러스터 및 데이터센터 확장에 우선순위를 두는 모습이다.

지역별로도 대규모 투자가 이어지고 있다. 펜실베이니아주에 200억 달러를 투자하여 대규모 AI 서비스 확장 및 데이터센터 2곳을 신규 구축하고, 노스캐롤라이나주에도 100억 달러를 투자하여 AI 및 클라우드 인프라를 확장하고 있다. 이러한 지역별 투자는 분산된 고성능 컴퓨팅 환경을 구축하여 전 세계 고객들에게 안정적이고 빠른 AI 서비스를 제공하기 위함이다.

더불어 아마존은 자체 AI 칩 개발에도 박차를 가하고 있다. 고성능 AI 학습용 칩인 '트레이니움(Trainium)'과 추론용 칩인 '인페렌시아(Inferentia)' 개발을 통해 외부 반도체 의존도를 낮추고 AI 워크로드에 최적화된 하드웨어 생태계를 구축하고 있다.

또한, AI 모델 개발 및 배포를 위한 관리형 서비스인 '베드록(Bedrock)' 플랫폼 구축에도 집중 투자하고 있다. 베드록은 기업들이 다양한 파운데이션 모델을 활용하여 자체 AI 애플리케이션을 쉽게 개발하고 배포할 수 있도록 지원하며, 이는 AWS 고객들이 AI를 쉽게 도입할 수 있는 강력한 유인책이 되고 있다. 이는 단순히 현재의 AI 수요를 충족시키는 것을 넘어, 미래의 폭발적인 AI 성장에 선제적으로 대비하는 전략적 포석이다.

그림 106. 아마존의 트레이니움(Trainium) 2

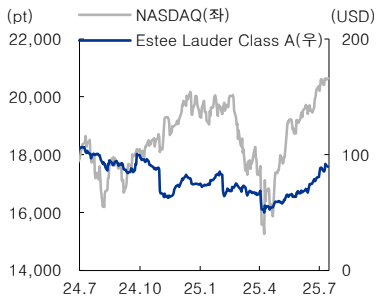


자료: 아마존, IBK투자증권

Not Rated

산업 분류(GICS)	필수소비재
상장거래소	New York
현재가 (7/14)	89,59달러
시가총액(USD)	322억 달러
시가총액(원)	46,985십억 원
52주 최고가	103,44달러
주주구성	
The Vanguard Group, Inc.	11.30%
Fidelity Management & Research Company LLC	5.60%
주가상승	1M 6M 12M
상대기준	33.1% 21.3% -13.3%
절대기준	26.7% 12.9% -25.5%

에스티로더 상대주가 (%)



에스티로더 (EL US)

산업내 AI 생산성 혁신 우수 사례

오픈AI 및 마이크로소프트와 협업

동사는 2024년 말부터 OpenAI 및 Microsoft와의 협업을 통해 사내 GPT Lab을 본격 가동하고 있다. 이를 기반으로 약 240여 개의 커스텀 GPT를 도입하여 브랜드별 소비자 인사이트 분석, 제품 기획, 트렌드 리포팅 등의 업무를 자동화하고 있으며, Microsoft Azure 기반의 AI 에이전트 'ConsumerIQ'를 활용해 80년치 내부 자료 검색·분석도 가능해졌다. 이 같은 사내 LLM 도입은 조직 생산성의 구조적 변화를 가져왔다.

AI로 마케팅·공급망 효율화

동사는 'Trend Studio' 플랫폼 내에서 트렌드 탐지, 가상 체험, 카피라이팅, 추천 알고리즘까지 통합적으로 운영하고 있다. Adobe Firefly와의 연동을 통해 디지털 자산 제작도 자동화하고 있다. 반복적인 콘텐츠 생성 작업을 줄이며, 캠페인 론칭 속도를 단축하고 브랜드별 마케팅 대응력을 높이고 있다.

2025년 상반기, 동사는 공급망 전체에 걸쳐 AI 기반 수요 예측 및 재고 배치 시스템을 도입했다. 주요 공장에는 디지털 트윈을 적용한 IBP(통합 비즈니스 계획) 프로세스가 적용되어, 실시간 예측 및 시뮬레이션 기반 재고 운영이 가능해졌다. 이는 글로벌 뷰티 시장의 수요 불확실성 확대 속에서 비용 효율성을 유지하는 핵심 요인으로 평가된다.

버추얼 파운데이션 체험

AI와 AR(증강현실)을 결합한 뷰티 테크 서비스를 제공하고 있다. iMatch Virtual Shade Exper는 AI를 기반으로 사용자의 피부색을 분석하여 적합한 파운데이션 색상을 추천하는 가상 체험 도구이다. 이는 매장에 직접 가지 않고도 집에서 편안하게 다양한 화장품을 시험해 볼 수 있게 한다. AI 및 AR 기반의 '가상 체험(Virtual Try-On)' 기술은 동사의 매출 증대(최대 2.5배 증가)와 반품률 감소(8% 이상 감소)에 기여하고 있다.

(단위: 백만달러, 배)	2020	2021	2022	2023	2024
매출액	14,290	16,220	17,740	15,940	15,610
영업이익	610	2,620	3,170	1,510	970
세전이익	1,050	3,330	3,040	1,400	770
지배주주순이익	680	2,870	2,390	1,010	390
EPS(원)	2	8	7	3	1
증가율(%)	-61	318	-16	-57	-61
영업이익률(%)	4	16	18	9	6
순이익률(%)	5	18	14	6	3
ROE(%)	16	57	41	18	7
PER	101	41	39	70	98
PBR	17	19	16	13	7
EV/EBITDA	25	32	22	30	18

자료: Company data, IBK투자증권

뷰티 AI 전략 강화

사내 GPT랩 가동

맞춤형 GPT 240개 도입
생산성 혁신

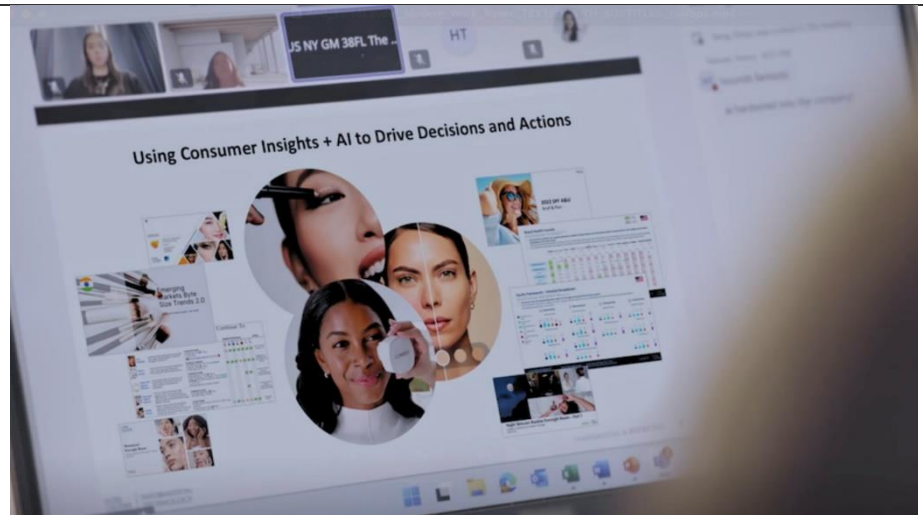
AI 기술을 전사적인 전략의 핵심으로 삼아 조직의 생산성과 의사결정 체계를 혁신하고 있다. 2024년 말부터 오픈AI(OpenAI) 및 마이크로소프트(Microsoft)와의 협력을 통해 사내 GPT 랩을 본격 가동하며, 이를 기반으로 약 240여 개의 맞춤형 GPT(Custom GPTs)를 도입했다. 이 커스텀 GPT들은 각 브랜드별 소비자 인사이트 분석, 제품 기획, 트렌드 리포팅 등의 업무를 자동화하여 의사결정 속도를 획기적으로 단축하고 있다.

애저 기반 AI 에이전트 도입

ConsumerIQ 도입으로
의사결정 자동화

특히 마이크로소프트 애저(Microsoft Azure) 기반의 AI 에이전트 'ConsumerIQ'를 활용하여 방대한 내부 자료를 신속하게 검색하고 분석하는 능력을 확보했다. 과거 몇 주가 걸리던 데이터 분석 작업이 몇 초 만에 가능해지면서, 에스티로더는 시장의 변화와 소비자 니즈를 실시간으로 파악하고 이에 맞는 신제품을 더욱 빠르게 출시할 수 있게 되었다. 이러한 사내 LLM 도입은 AI 도구의 내재화를 통해 조직 생산성의 구조적인 변화를 가져오며, 임직원들이 반복적인 업무에서 벗어나 더욱 전략적이고 창의적인 업무에 집중할 수 있도록 지원하고 있다.

그림 107. 에스티로더의 ConsumerIQ 툴 개발 과정



자료: 에스티로더, IBK투자증권

제조업 생산성 부스트 업

마케팅 콘텐츠 제작 효율화

Adobe Firefly 연동으로
콘텐츠 제작 자동화

마케팅 분야에서 생성형 AI를 중심으로 한 효율화 전략을 추진하고 있다. '트렌드 스튜디오' 플랫폼 내에서 트렌드 탐지, 가상 체험, 카피라이팅, 추천 알고리즘까지 통합적으로 운영하며, 이는 고객들에게 더욱 개인화되고 몰입감 있는 경험을 제공하는 핵심 동력이 되고 있다.

더 나아가, 에스티로더는 어도비 파이어플라이(Adobe Firefly)와의 연동을 통해 디지털 자산 제작의 자동화 수준을 높이고 있다. 이는 연간 수십만 건에 달하는 마케팅 에셋 생산을 신속하고 효율적으로 수행하며, 크리에이티브 팀이 전략적인 아이디어 구상과 캠페인 기획에 더욱 집중할 수 있도록 돕는다. 반복적인 콘텐츠 생성 작업을 줄임으로써 캠페인 론칭 속도를 단축하고 브랜드별 마케팅 대응력을 높이는 것은 물론, 고정비 부담이 높은 뷰티 업종에서 Gen AI 기반의 마케팅 자동화는 비용 구조 개선과 시장 대응 속도 및 품질 향상 측면에서 중장기적인 경쟁력을 제고하는 중요한 수단으로 작용한다.

공급망 최적화

AI 기반 수요예측·재고 최적화

AI 기술을 공급망 전반의 최적화에도 적극적으로 적용하고 있다. 2025년 상반기부터 공급망 전체에 걸쳐 AI 기반의 수요 예측 및 재고 배치 시스템을 도입하여 글로벌 뷰티 시장의 수요 불확실성에 대응하고 비용 효율성을 유지하는 데 핵심적인 역할을 하고 있다. 또한, 주요 공장에는 디지털 트윈(Digital Twin)을 적용한 IBP(통합 비즈니스 계획) 프로세스가 적용되어, 실시간 예측 및 시뮬레이션 기반의 재고 운영이 가능해졌다. 이는 갑작스러운 시장 변화에도 유연하게 대응하고, 최적의 재고 수준을 유지하여 운영 효율성을 극대화할 것으로 예상된다.

그림 108. 어도비 파이어플라이 생성형 AI로 에스티로더 디지털 마케팅 제작



자료: 에스티로더, IBK투자증권

뷰티AI와 AR의 만남

버추얼 파운데이션 체험

AI로 파운데이션 색상
정확히 추천

버추얼 파운데이션 체험 도구 'iMatch Virtual Shade Expert'는 고객들이 실제로 제품을 바르지 않고도 자신에게 맞는 파운데이션 색상을 찾을 수 있도록 돕는 혁신적인 기술이며, 증강 현실(AR)과 인공지능(AI)을 결합하여 구현된다.

- 실시간 피부 톤 분석: 사용자가 스마트폰이나 컴퓨터 카메라를 통해 자신의 얼굴을 비추면, AI 알고리즘이 얼굴의 미세한 피부 톤, 언더톤(피부 밑 색조), 그리고 현재 조명 상태까지 정밀하게 분석한다. 에스티로더는 방대한 피부 톤 데이터베이스를 기반으로 이 분석의 정확도를 높였다.
- 가상 파운데이션 적용: AI가 분석한 피부 톤에 맞춰 사용자가 선택한 에스티로더 파운데이션 제품의 색상과 질감(예: 매트, 글로우), 커버 강도(예: 가볍게, 두껍게)를 얼굴 위에 실시간으로 증강 현실 그래픽으로 오버레이한다. 사용자가 고개를 움직이거나 표정을 바꿔도 파운데이션이 얼굴에 자연스럽게 밀착되어 있는 것처럼 매우 사실적으로 보이도록 한다.
- 정밀한 색상 추천: 단순히 여러 색상을 보여주는 것을 넘어, AI는 사용자의 고유한 피부에 가장 잘 어울리는 파운데이션 색상을 정확히 추천한다. 이는 수만 가지의 조합 속에서 최적의 선택을 도와준다.
- 다양한 비교 및 체험 기능: 고객은 여러 파운데이션 색상을 빠르고 쉽게 번갈아 가며 체험할 수 있으며, 화면 분할 기능을 통해 파운데이션 적용 전/후를 비교해볼 수도 있다. 이는 온라인 구매 시 발생할 수 있는 색상 선택의 어려움을 크게 줄여준다.

그림 109. 에스티로더의 'iMatch Virtual Shade Expert'



자료: 에스티로더, IBK투자증권

Compliance Notice

동 자료에 게재된 내용들은 외부의 압력이나 부당한 간섭없이 본인의 의견을 정확하게 반영하여 작성되었음을 확인합니다.

동 자료는 기관투자자 또는 제3자에게 사전 제공한 사실이 없습니다.

동 자료는 조사분석자료 작성에 참여한 외부인(계열회사 및 그 임직원등)이 없습니다.

조사분석 담당자 및 배우자는 해당종목과 재산적 이해관계가 없습니다.

동 자료에 언급된 종목의 지분을 1%이상 보유하고 있지 않습니다.

당사는 상기 명시한 사항 외 고지해야 하는 특별한 이해관계가 없습니다.

종목명	담당자	담당자(배우자) 보유여부			1%이상	유가증권	계열사	공개매수	IPO	회사채	중대한	M&A
		수량	취득가	취득일	보유여부	발행관련	관계여부	사무취급		지급보증	이해관계	관련
해당 사항 없음												

투자 의견 안내 (투자기간 12개월)

종목 투자 의견 (절대 수익률 기준)			
매수 15% 이상	Trading Buy (중립) 0%~15%	중립 -15%~0%	축소 -15% 이상 하락
업종 투자 의견 (상대 수익률 기준)			
비중 확대 +10% ~	중립 -10% ~ +10%	비중 축소 ~ -10%	

투자 등급 통계 (2024.07.01~2025.06.30)

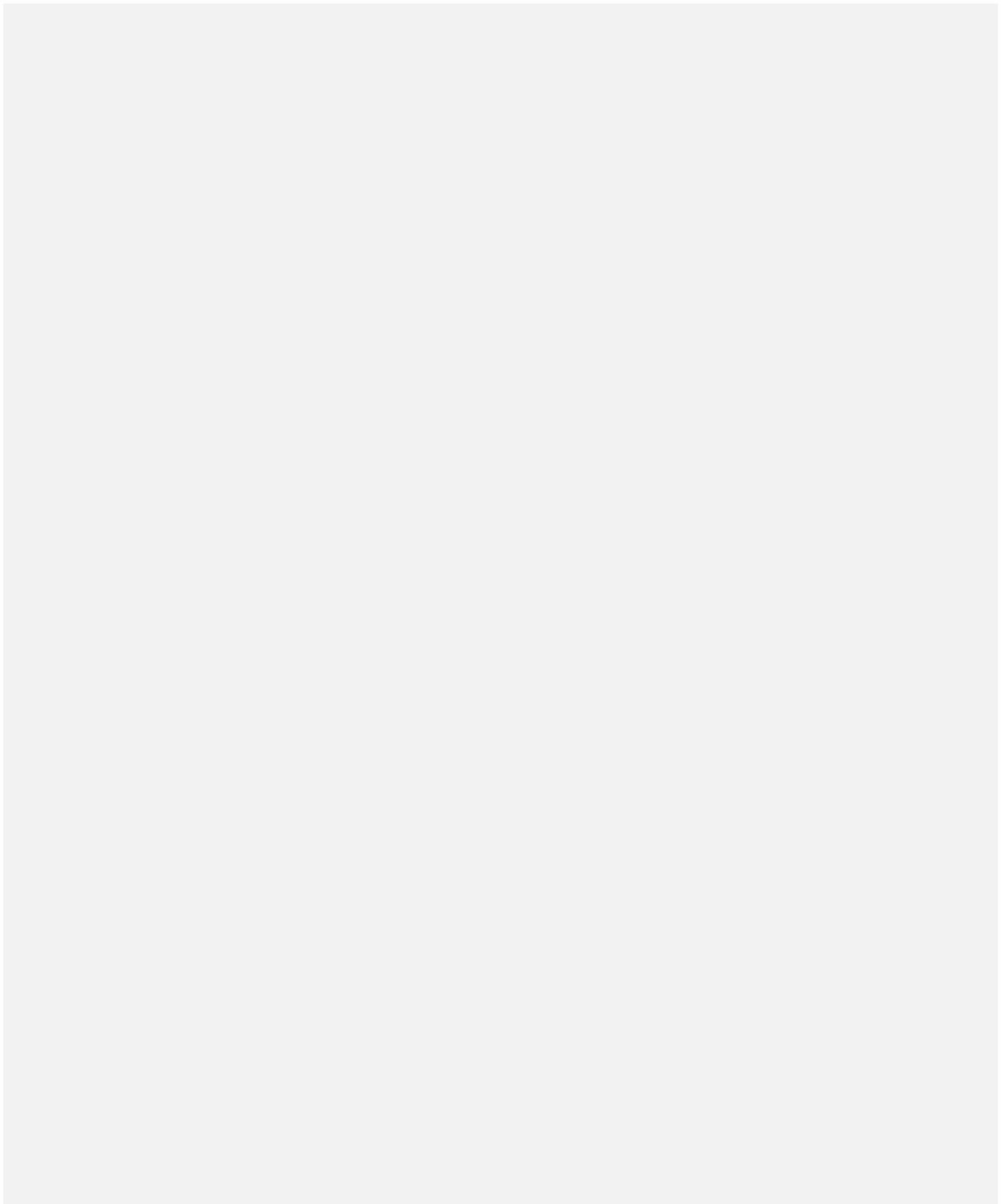
투자 등급 구분	건수	비율(%)
매수	132	89.8
Trading Buy (중립)	11	7.5
중립	4	2.7
매도	0	0

최근 2년간 주가 그래프 및 목표주가(대상 시점 1년) 변동 추이

(▲) 매수, (■) Trading Buy (중립), (●) 중립, (◆) 축소, (■) Not Rated / 담당자 변경

NAVER	추천일자	투자 의견	목표가(원)	과리율(%)	
				평균	최고/최저
	2023.05.31	매수	300,000	-32.65	-21.83
	2024.05.31	1년 경과	300,000	-43.96	-40.10
	2024.11.08	매수	260,000	-21.81	-10.77
	2025.02.07	매수	300,000	-29.53	-3.17
	2025.07.15	매수	320,000		
카카오	추천일자	투자 의견	목표가(원)	과리율(%)	
				평균	최고/최저
	2022.11.29	매수	79,000	-28.22	-10.25
	2023.09.04	매수	75,000	-42.11	-34.60
	2023.11.09	매수	65,000	-28.95	-6.00
	2024.11.09	1년 경과	65,000	-39.65	-28.31
	2025.02.20	매수	59,000	-30.46	-23.05
	2025.05.08	매수	52,000	-2.62	35.38
	2025.07.15	매수	72,000		

Note



Note



IBKS Research Center

성명	직급	담당업종	전화	이메일
용대인	전무(부문장)	총괄	6915-5400	daeinyong@ibks.com
이승훈	상무대우(본부장)	인터넷/게임	6915-5680	dozed@ibks.com

투자분석부

변준호	연구위원	Strategy	6915-5670	ymaezono@ibks.com
정용택	수석 Economist	Economy	6915-5701	ytjeong0815@ibks.com
김인식	연구위원	자산배분/ETF	6915-5472	kds4539@ibks.com
정형주	연구위원	채권/크레딧	6915-5654	hj.jeong@ibks.com
조경진	연구위원	해외주식	6815-5464	ckjins@ibks.com
권순호	연구원	Quant	6915-5667	snowkonn@ibks.com

기간산업분석부

이동욱	연구위원	화학/정유	6915-5671	ttreestump@ibks.com
남성현	연구위원	유통·식자재/지주	6915-5672	rockrole@ibks.com
김유혁	연구위원	미디어/엔터	6915-5673	yuhyuk.kim@ibks.com
이현욱	연구원	자동차/2차전지	6915-5659	hwle1125@ibks.com
오지훈	연구원	조선/기계	6915-5662	jihoonoh@ibks.com

혁신기업분석부

이건재	연구위원	소재·부품·장비/스몰캡	6915-5676	geonjaelee83@ibks.com
김운호	연구위원	IT/반도체	6915-5656	unokim88@ibks.com
김태현	연구위원	음식료/유틸리티/통신	6915-5658	kith0923@ibks.com
정이수	연구위원	제약/바이오	6915-5677	ysjeong306@ibks.com
조정현	연구원	건설/부동산	6915-5660	controlh@ibks.com
강민구	연구원	IT/디스플레이	6915-5473	kmg@ibks.com

“국민과 중소기업에 필요한 참 좋은 IBK투자증권”



서울특별시 영등포구 여의도동 국제금융로 6길 11
대표번호 02-6915-5000
고객지원부 1588-0030, 1544-0050

영업부	02) 6915-2626	IBK WM센터 대구	053) 752-3535
강남센터	02) 2051-5858	IBK WM센터 광주	062) 382-6611
강남역 금융센터	02) 532-0210	IBK WM센터 일산	031) 904-3450
분당센터	031) 705-3600	IBK WM센터 판교	031) 724-2630
IBK WM센터 역삼	02) 556-4999	IBK WM센터 평촌	031) 476-1020
IBK WM센터 목동	02) 2062-3002	IBK WM센터 천안	041) 569-8130
IBK WM센터 강남	02) 2057-9300	IBK WM센터 부산	051) 741-8810
IBK WM센터 한남동	02) 796-8500	IBK WM센터 창원	055) 282-1650
IBK WM센터 중계동	02) 948-0270	IBK WM센터 울산	052) 271-3050
IBK WM센터 반포자이	02) 3481-6900	IBK WM센터 시화공단	031) 498-7900
IBK WM센터 동부이촌동	02) 798-1030	IBK WM센터 남동산단	032) 822-6200