SPRi 포럼
v2023.8.28

SPRi 소프트웨어정책연구소
Software Policy & Research Institute

생성AI 산업 생태계 현황

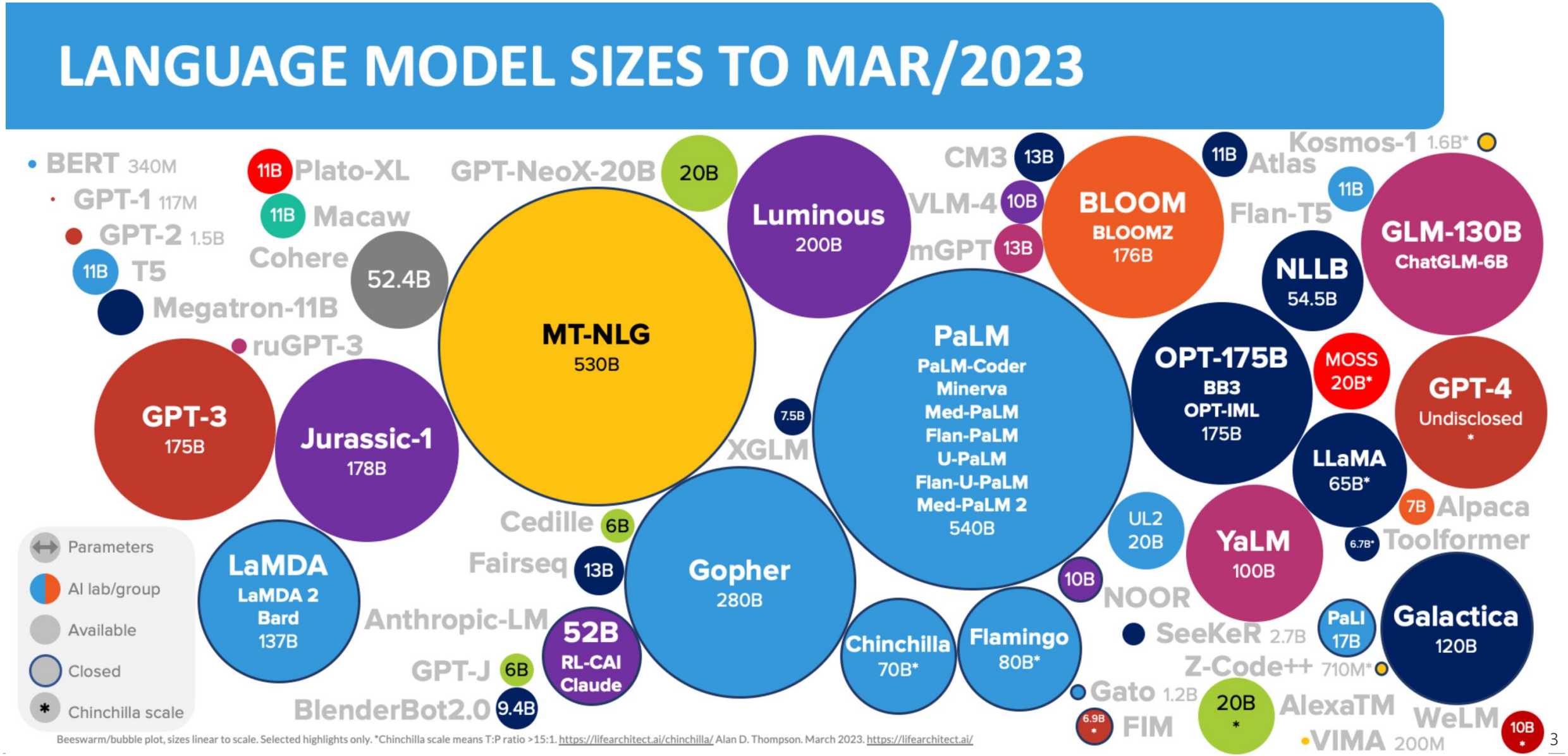
- Landscape of Generative AI Industry Ecosystem -

유재흥 책임연구원
소프트웨어정책연구소
2023.8.29(화)

인공지능 분야 주요 사건

NIA(2022.12.31), 현대 인공지능의 역사적 사건 및 산업 사회 변화 분석

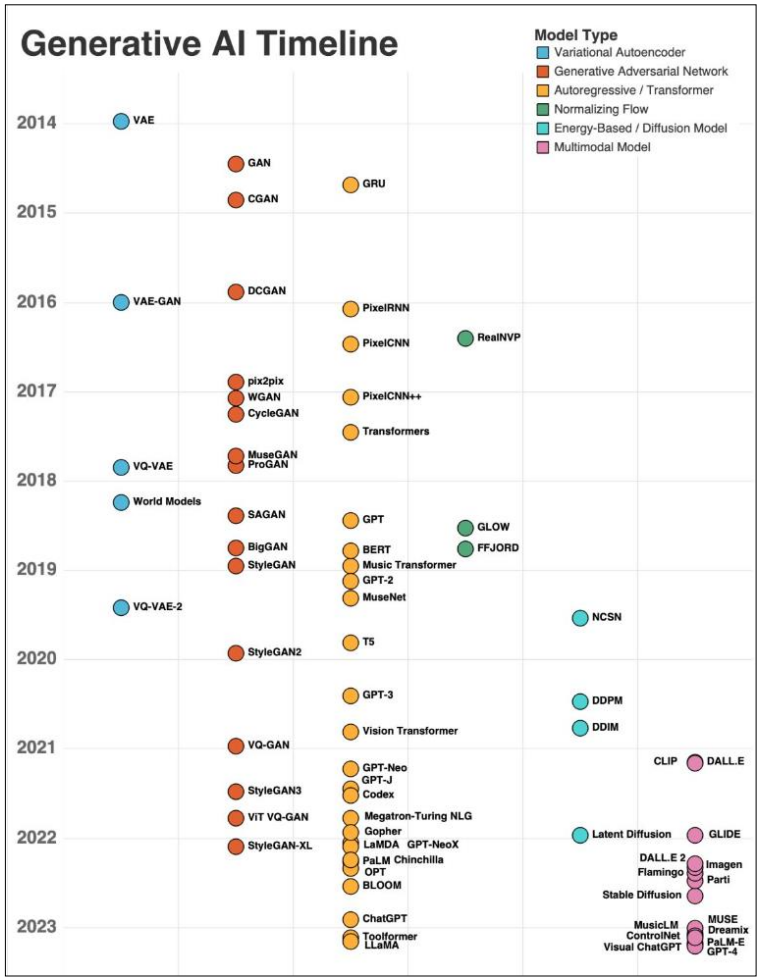
생성AI의 부상: 초거대 인공지능(LLM) 모델 개발 경쟁



생성AI의 부상: 2022년도 생성AI의 해

2022년 한해 ChatGPT 포함, 주요 언어, 음성, 비디오 등에서 주요한 생성 모델 등장

구분	모델	기업
언어	Chinchilla(7B, '22.3) PaLM(540B, '22.4), OPT-175B('22.5), BLOOM (176B, '22.7), ChatGPT(175B, '22.11), GPT-4('23.3), LLAMA (7B~65B, '23.2) PaLM(540B, '22.4)	Google DeepMind Google Meta HuggingFace OpenAI OpenAI Meta Google
이미지	Make-A-Scene('22.3), DALL-E2 '22.4), Imagen('22.5), Midjourney ('22.7), Stable Diffusion ('22.8)	Meta OpenAI Google Stability.AI Midjourney
오디오·비디오	Make-A-Video('22.9), MusicLM(오디오, '23.1), Gen-1(영상, '23.2), Gen-2(영상, '23.3),	Meta Google Runway Runway



Source: David Foster (linkedIn)

Source: SPRI

생성AI의 부상: 2023년도에도 계속

OpenAI GPT-4('23.5)이후 구글, 앤쓰로픽AI, 메타 등 초거대AI 모델 고도화

- 성능 강화 (멀티모달, 정확도), 신뢰성 및 안전성 강화

Google

PaLM 2

PaLM의 다음 세대 언어 모델인 PaLM 2를 소개합니다. PaLM 2는 향상된 다중 언어와 추론 능력, 그리고 코딩 능력을 갖춘 최첨단 언어 모델입니다.

· **다중언어:** PaLM 2는 100개 이상의 언어에 걸쳐 다국어 텍스트를 학습했습니다. PaLM 2의 다중언어는 그동안 해결하기 어려운 문제였던 다양한 언어의 미묘한 뉘앙스를 이해, 생성, 그리고 번역할 수 있으며, 이는 속담과 시, 수수께끼도 포함합니다.

· **추론 능력:** PaLM 2의 광범위한 데이터셋은 과학논문과 수학적 표현이 포함된 웹 페이지 등을 포함합니다. 따라서 논리와 상식적 추론, 수학 등에서 더욱 향상된 역량을 지니고 있습니다.

· **코딩:** PaLM 2는 공개되어 있는 방대한 양의 자연어와 소스 코드 데이터셋으로 사전 훈련되었습니다. 파이썬(Python)과 자바스크립트(JavaScript) 등 널리 사용되는 프로그래밍 언어에 뛰어날 뿐만 아니라, 프롤로그(Prolog), 포트란(Fortran), 베릴로그(Verilog)와 같은 언어에 특화된 코드를 생성할 수도 있습니다.

Google I/O ('23.5.11)

ANTHROP\C

Claude 2

Jul 11, 2023 • 4 min read

- 최대 10만 개의 토큰(약 7만 5,000단어)을 프롬프트로 지원
- 수백 쪽 분량의 기술 문서나 책 한 권을 편집하고 요약
- 초등학생을 위한 8,500개의 수학 문제 세트인 GSM8k에서 클로드 2는 88.0%의 정확도(이전에는 85.2%)를 달성
- 파이썬 코딩 벤치마크인 HumanEval 평가 56.0% → 71.2%
- 무료 이용
- 현재 미국, 영국에서만 이용 가능

Anthropic AI, '23.7.11

Meta AI

Llama 2 was trained on **40% more data** than Llama 1, and has double the context length.

Llama 2

MODEL SIZE (PARAMETERS)	PRETRAINED	FINE-TUNED FOR CHAT USE CASES
7B	Model architecture:	Data collection for helpfulness and safety:
13B	Pretraining Tokens: 2 Trillion	Supervised fine-tuning: Over 100,000
70B	Context Length: 4096	Human Preferences: Over 1,000,000

Meta, '23.7.18

생성AI의 부상: 2023년도에도 계속 Apple, Tesla 참전 예고 ('23.7.19)

美, 시총 'Magnificent 7' LLM 패권전쟁 (MS, Apple, Google, Nvidia, Amazon, Tesla, Meta)

- 애플 'AppleGPT' 개발 예고 - 아직 내부 적용(시리, 검색, 지도 앱 등), 자체 프레임워크 Ajax 기반, 내년 공개서비스 가능할 전망
- 테슬라 xAI 출범('23.7.12) – 안전한 초지능 개발, 머스크 사업(Tesla, SpaceX, Boring, Neurallink, Starlink, Twitter, Tesla Humanoid Bot)

Bloomberg

US Editi

● Live Now Markets Economics Industries Tech AI Politics Wealth Pursuits Opinion Businessweek Equality Gree

Technology
AI

Apple Tests 'Apple GPT,' Develops Generative AI Tools to Catch OpenAI

- Company builds large language models and internal chatbot
- Executives haven't decided how to release tools to consumers

FORTUNE

TECH · A.I.

Apple added $71 billion to its market value after news that it's been secretly building an 'Apple GPT' to rival OpenAI

BY CHLOE TAYLOR
July 20, 2023 at 9:17 PM GMT+9

ARTIFICIAL INTELLIGENCE

Generative AI: Apple's Secret Project 'Apple GPT' Emerges

머스크 인공지능 스타트업 'xAI' 출범

| "우주의 진정한 본질 이해 목표"…구글·오픈AI 출신 영입

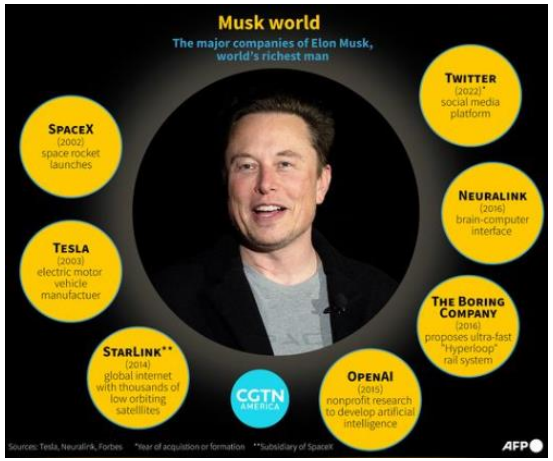
인터넷 | 입력 :2023/07/13 09:01 수정: 2023/07/13 10:52

AI Artificial intelligence Internet News Starlink technology World

Elon Musk's xAI: Exploring the Intersection of AI and Tesla's Vision

Elon Wants to Master the Universe With xAI

Musk's new AI quest poses a threat to Google's DeepMind.

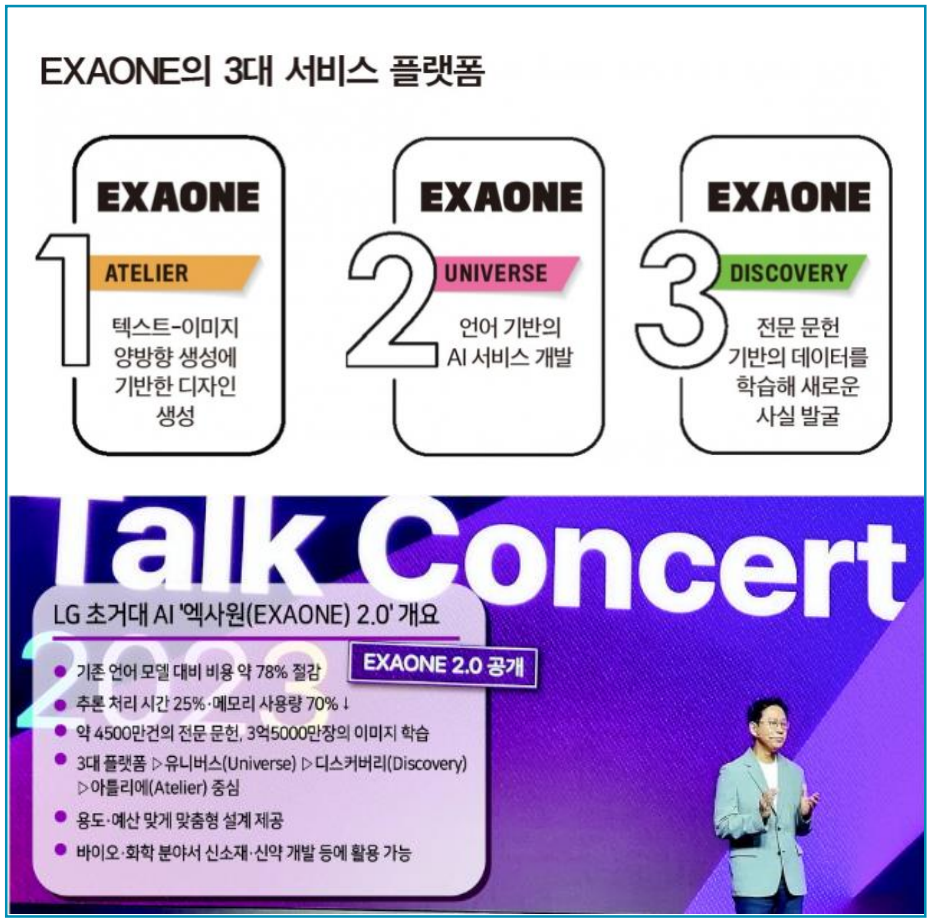


Source: Bloomberg, Fortune

생성AI의 부상: 2023년도에도 계속 국내 LG AI연구원 EXAONE 2 공개

LG 초거대 AI 모델: Exaone 공개 ('21.12) → Exaone 2 발표('23.7.19)

- Exaone1: 3천억개 파라미터, 6천억개의 말뭉치, 2.5억개의 이미지 학습
- Exaone2: 4,500만건의 전문 문헌, 3.5억장의 이미지 학습, 모델 개발 비용 절감



■ LG AI연구원, 19일 'LG AI 토크 콘서트'에서 '엑사원(EXAONE) 2.0' 첫 공개

□ △고품질 학습 데이터 △비용 효율성 △맞춤형 모델 설계 등 '엑사원 2.0'의 경쟁력 소개

□ '전문가 AI' 서비스 개발의 기반인 엑사원 3대 플랫폼 ①유니버스(Universe), ②디스커버리(Discovery), ③아틀리에(Atelier) 공개

① **엑사원 유니버스: 전문가용 대화형 AI 플랫폼**

- 전문성이 요구되는 분야의 질문에 대해 근거에 기반한 정확한 답변 생성하는 AI 플랫폼
- 이날 공개한 AI/머신러닝 분야를 시작으로 화학, 바이오, 제약, 의료, 금융, 특허 등 도메인 별 특화 서비스도 구축 진행 중

② **엑사원 디스커버리: 화학 및 바이오 분야 발전 앞당길 신소재·신물질·신약 개발 플랫폼**

- 멀티모달 AI 기술 활용해 전문 문헌의 텍스트뿐만 아니라 분자 구조, 수식, 차트, 테이블, 이미지 등 非텍스트 정보까지 데이터베이스화
- AI와 대화하며 ▲전문 문헌 검토 ▲소재 구조 설계 ▲소재 합성 예측까지 진행 가능, 1만회가 넘었던 합성 시행착오를 수십회로 줄이고 연구개발 소요 시간은 40개월에서 5개월로 단축 예상

③ **엑사원 아틀리에: 인간의 창의적 발상을 돕는 멀티모달 AI 플랫폼, 이미지를 언어로 표현하고 언어를 이미지로 시각화**

- LG AI연구원, 3분기 중 LG 계열사 디자이너 대상으로 서비스 오픈
- 연내 다양한 형태의 서비스 개발해 사업 모델 확장 계획

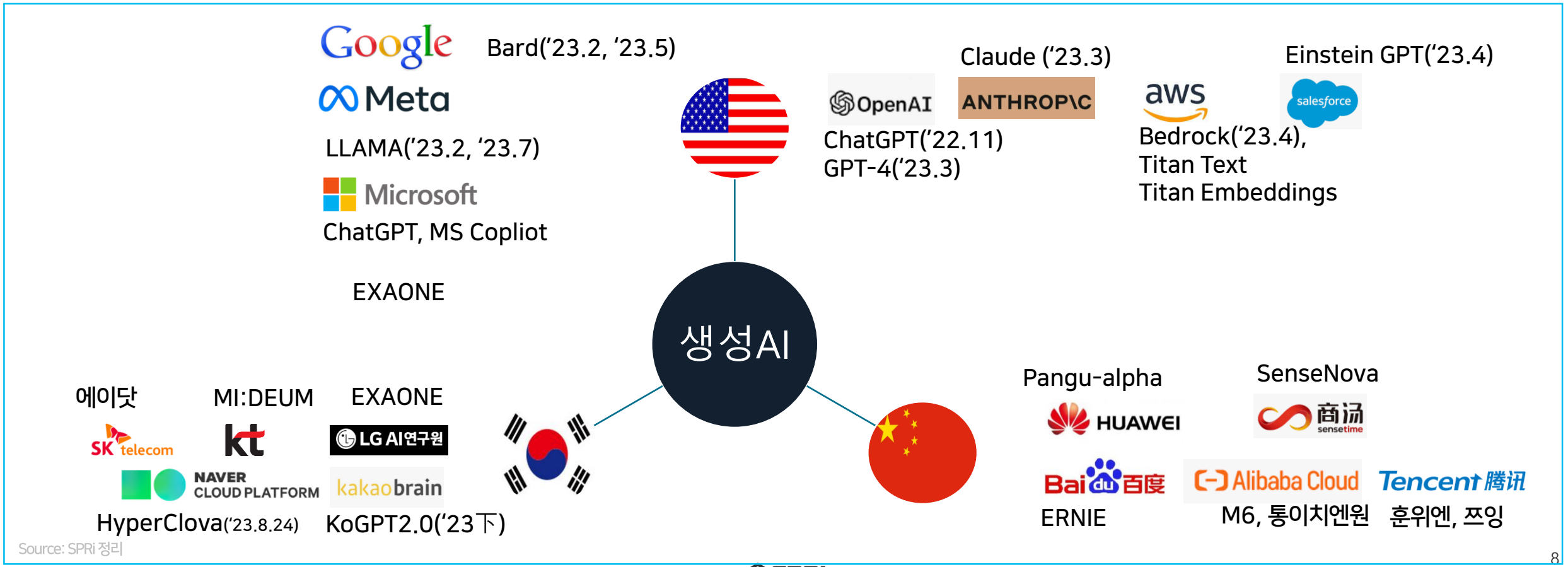
■ 배경훈 LG AI연구원장 "LG는 국내에서 유일하게 이중 언어 모델과 양방향 멀티모달 모델을 모두 상용화한 기업이며, 세상의 지식을 이해하고 발견하는 상위 1%의 전문가 AI를 개발하고 있다"며, ""다른 생성형 AI들과는 차별화된 고객 가치'를 창출하는 글로벌 경쟁력을 갖춘 AI 컴퍼니로 발전해 나갈 것"

Source: LG AI 연구원

생성AI의 부상: 2023년 본격 각축전

OpenAI社의 ChatGPT의 출시는 생성 AI 시대의 본격적인 도래를 알림

- ChatGPT 등장 이후 기존 빅테크와 스타트업에서 유사 언어모델중심의 AI생성 모델 경쟁 본격화
- Microsoft가 관련 시장을 주도하고 있는 가운데 구글, 메타, 아마존 등 미국 빅테크가 경쟁에 가세하고 있으며 중국 BAT를 포함한 AI기업들도 대응 서비스 출시, 국내에서도 네이버, 카카오, 통신사 등 대응 준비 중



Source: SPRi 정리

생성 AI의 부상: 경량 모델 연구

공개 모델을 활용해 비용효율적으로 LLM 개발 → 중소규모 조직, 도메인 특화 분야 적용

LLaMA

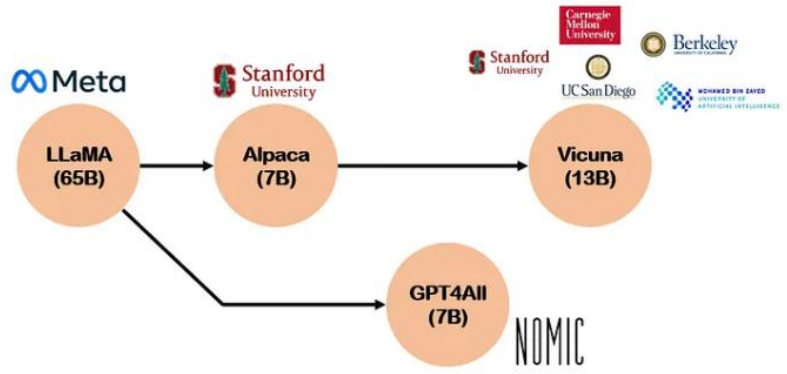
- META(구 Facebook)는 650억 개 이하의 파라미터를 갖는 경량 모델 LLaMA를 공개(`23.02.)

학습 및 경량화

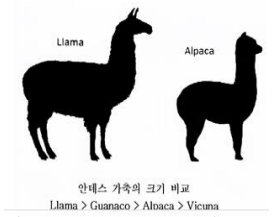
- LLaMA는 7B, 13B, 33B, 65B 등 **4가지의 크기**의 모델로 공개
- 7B는 1조 개의 토큰으로 훈련, 33B 및 65B는 1조 4천억 개의 토큰으로 훈련
- 훨씬 작은 규모지만, 데이터 학습 강화 및 파인 튜닝하여 더 큰 규모의 모델과 비교 가능한 효율적인 모델

성능

- 13B 모델이 일부 벤치마크에서 GPT-3(175B) 보다 뛰어나고, **65B 모델은 훨씬 더 큰 PaLM(540B)과 경쟁 가능**한 수준



LLaMA 계열 Chatbot의 계보도(by author)



안데스 가축의 크기 비교
Llama > Guanaco > Alpaca > Vicuna

▲ 과나코·라마·알파카·비쿠나 /조선일보 DB·위키피디아

GPT4ALL

- NomicAI는 GPT 버전의 경량화 생성AI 모델인 GPT4ALL을 Github에 공개(`23.04.)

학습 및 경량화 (LLAMA로 학습)

100만개의 문답 (GPT3.5로 수집) → 43만개의 문답 (경량화 → 파인튜닝)

전용 HW 없이, **노트북**에서도 AI 모델 개발 및 프로그램 **실행 가능**

비용절감 : 8시간학습 및 학습비용 174만원

- 모델 경량화 4일 소요
- 데이터 학습 8시간 소요
- GPU 비용 $800(약 107만원)
- 오픈AI API 비용 $500(약 67만원)

기존 ChatGPT 1일 운용비용: 10만달러 **(약 1.2억원)**

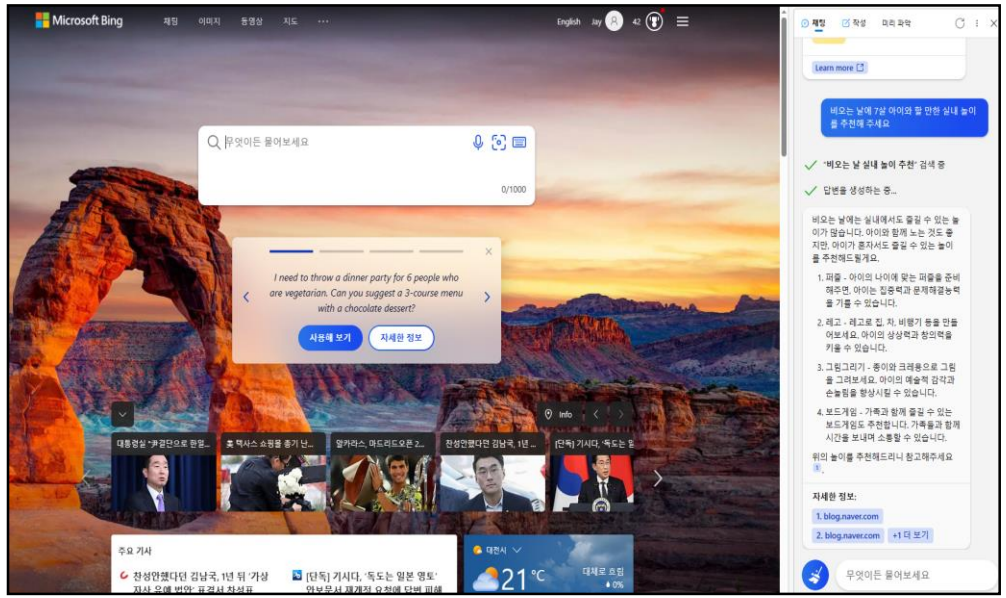
정보서비스 산업의 변화: 검색 시장의 패러다임 변화

ChatGPT의 등장으로 정보 서비스 시장의 패러다임: 검색 → 생성+검색으로 확장

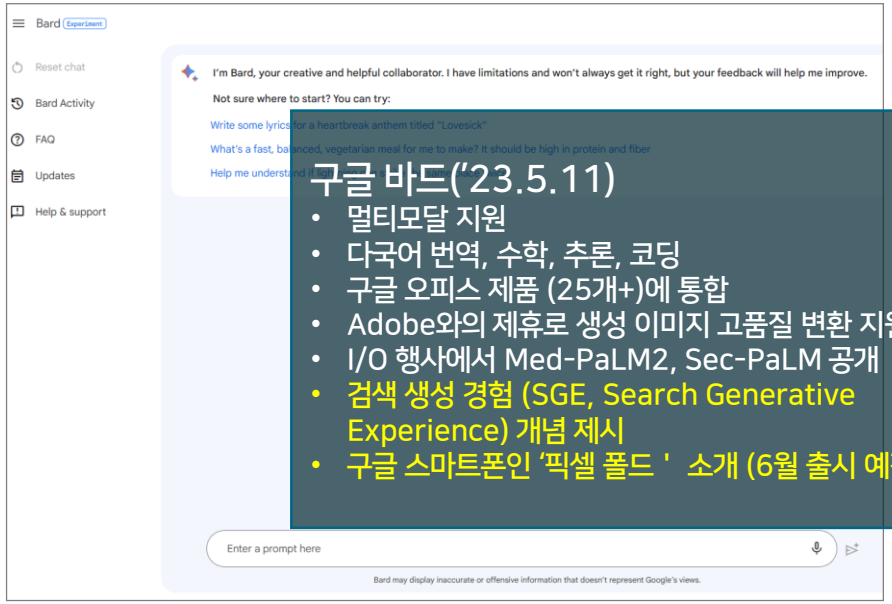
- MS는 자사 검색엔진(Bing)에 ChatGPT를 결합한 New Bing을 출시
- 구글은 ChatGPT 대항마로 구글 Bard 공개 '(23.3월 미국, 영국 우선 출시), 180개국 전면 출시 ('23.5.11, Google I/O 행사 후)

향후 정보 서비스는 빠르고 효율적인 정보 검색 → 신뢰할 만한 정보 생성과 결합

MS는 Bing 검색사이트와 챗GPT를 통합한 형태의 사용자 인터페이스 제공



구글 Bard는 별도 사이트로 제공 ('23.5월 초 기준)

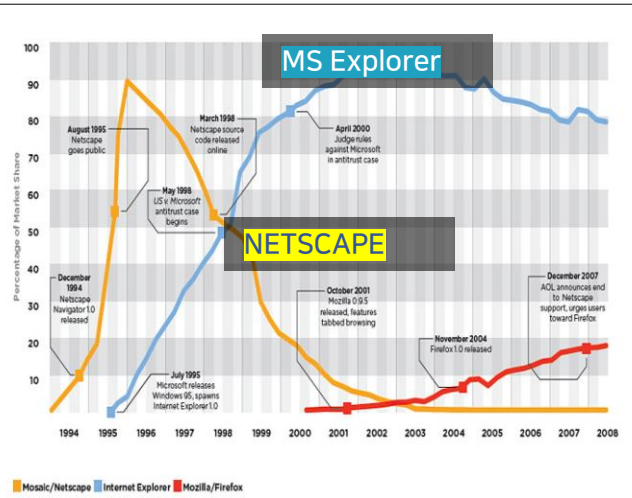
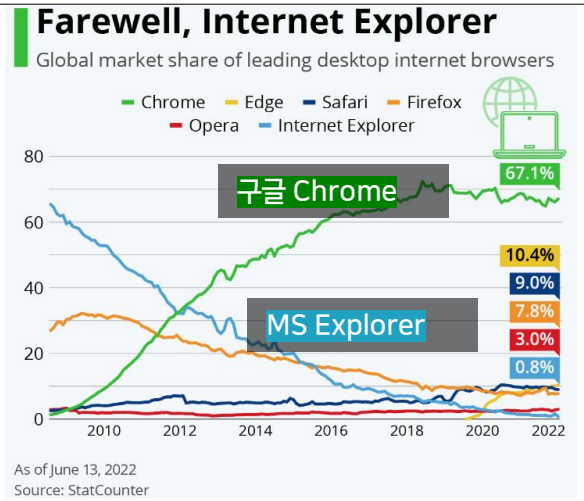
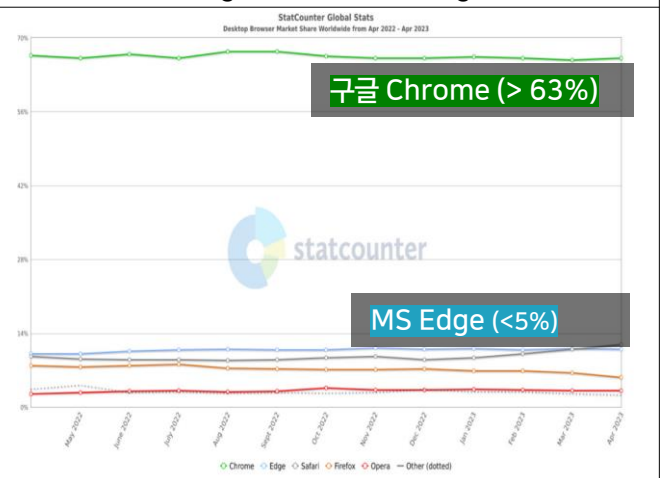


구글 바드('23.5.11)

- 멀티모달 지원
- 다국어 번역, 수학, 추론, 코딩
- 구글 오피스 제품 (25개+)에 통합
- Adobe와의 제휴로 생성 이미지 고품질 변환 지원
- I/O 행사에서 Med-PaLM2, Sec-PaLM 공개
- 검색 생성 경험 (SGE, Search Generative Experience) 개념 제시
- 구글 스마트폰인 '픽셀 폴드 ' 소개 (6월 출시 예정)

정보서비스 산업의 변화: 브라우저 전쟁 3라운드

과디지털 혁신의 과도기마다 브라우저 패권 경쟁 → 승자독식의 시장 지배력 확보

1라운드: PC-인터넷시대	2라운드: 모바일시대	3라운드: AI시대
NetScape vs. MS-Explorer	MS-Explorer vs. Google Chrome	Google Chrome vs MS-Edge
		
1990년대 마이크로소프트의 익스플로어 끼워팔기로 넷스케이프 추월	2000년대 구글의 안드로이드OS를 확산으로 모바일 브라우저 패권 확보	2023년 현재 ChatGPT의 결합을 통해 MS는 Edge 브라우저 시장 점유 확대 추진

자료: SPRi

산업(X) + Generative AI 융합: 생성AI 산업 혁신 기폭제

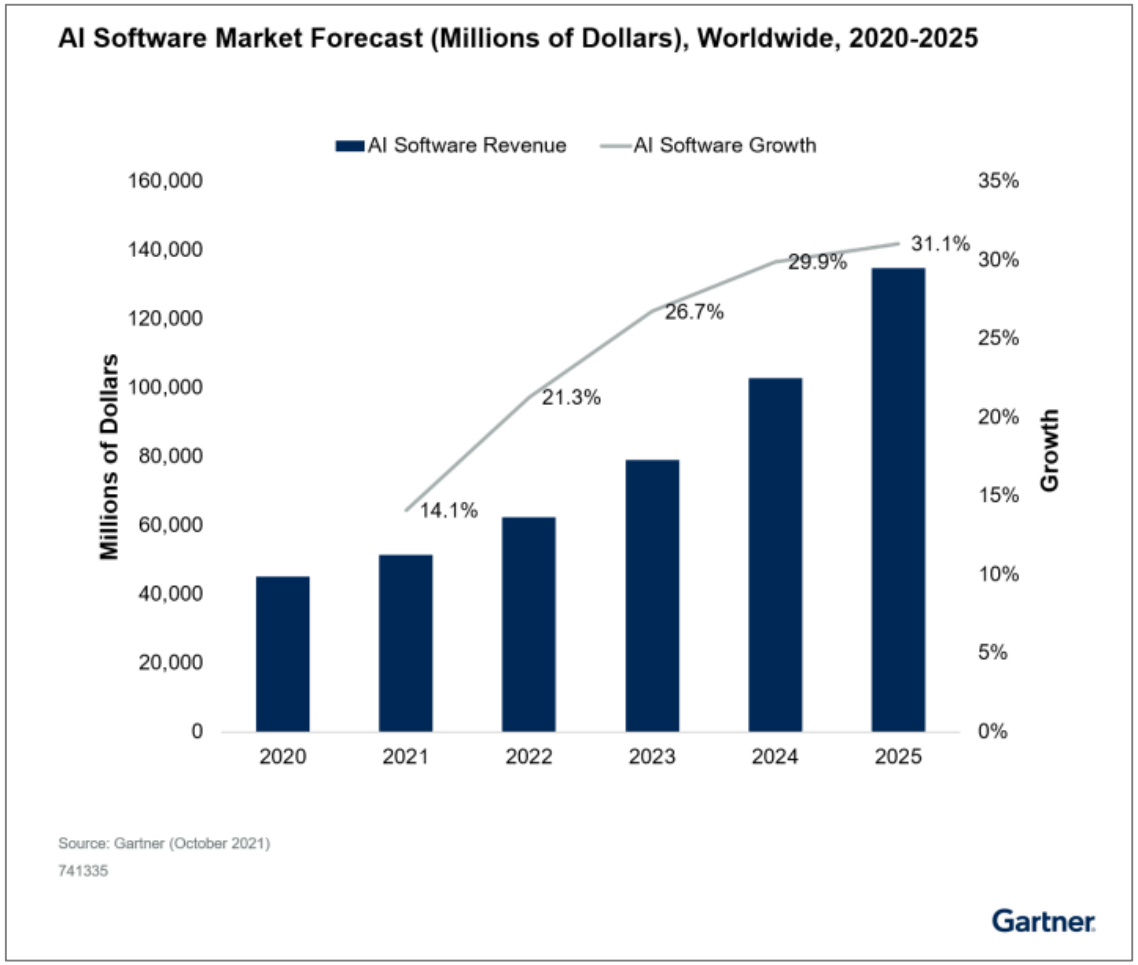
생성 AI는 산업 전반에 결합되어 산업의 혁신을 촉진할 기폭제 역할로 기대

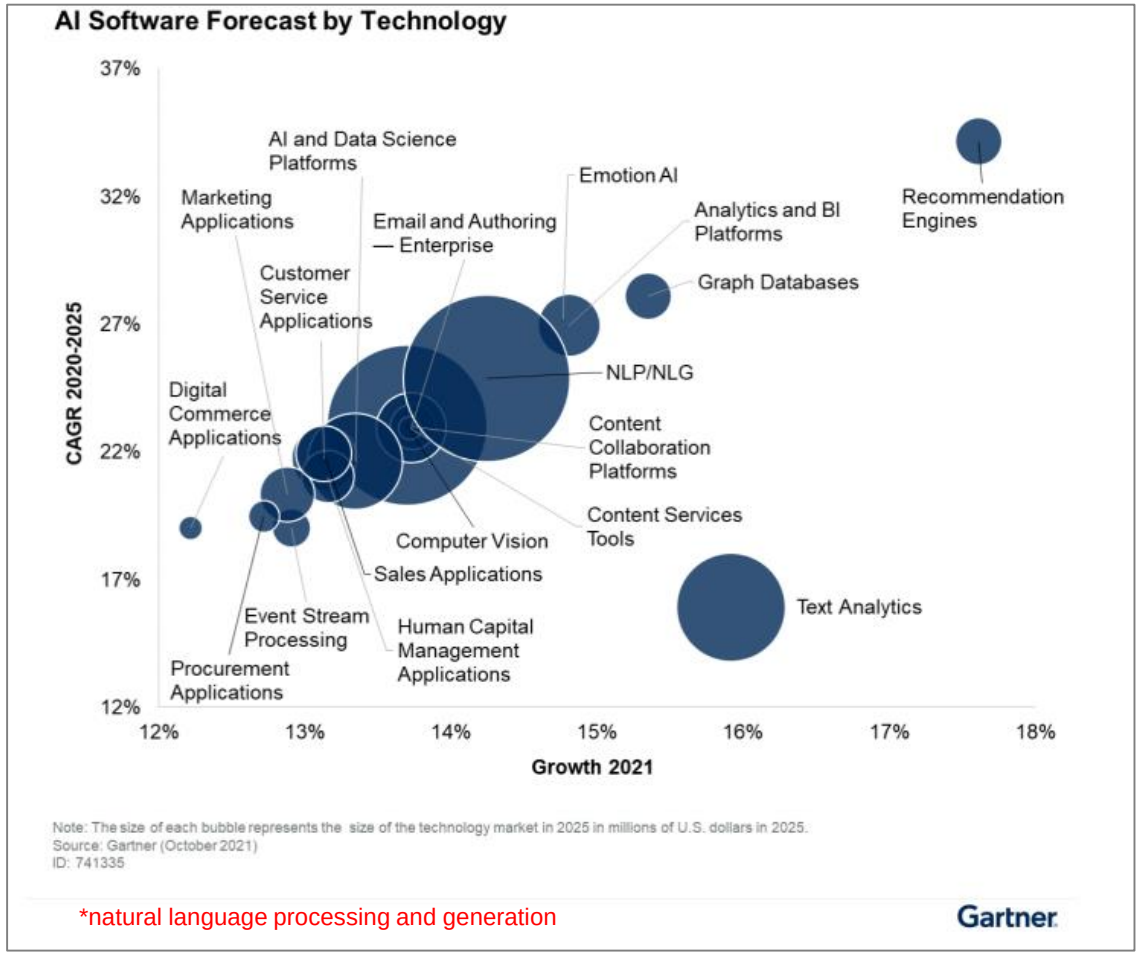
- 지난 3년간 VC들은 생성 AI 솔루션에 17억 달러 이상 투자, 생성 AI 시장의 급속성장을 예고 (Gartner, '23.1)
- 생성AI가 10년 후 세계 일자리 3억 개에 영향을 미치고, 글로벌 GDP를 7% 증가시킬 것 (Goldman Sachs, '23.1)

AI 시장: 생성AI 시장은 연평균 25% 성장

AI시장은 2025년 $134.8십억 달러에 이를 것으로 전망

- 5년 CAGR 24.5% vs. SW성장율('21) 13.2%





*natural language processing and generation

Source: Gartner(2021)

산업(X) + Generative AI 융합: 경제적 파급효과

GenAI는 연간 2.6~4.4조 달러의 잠재적 가치 생산 (총 63개의 use case 가정)

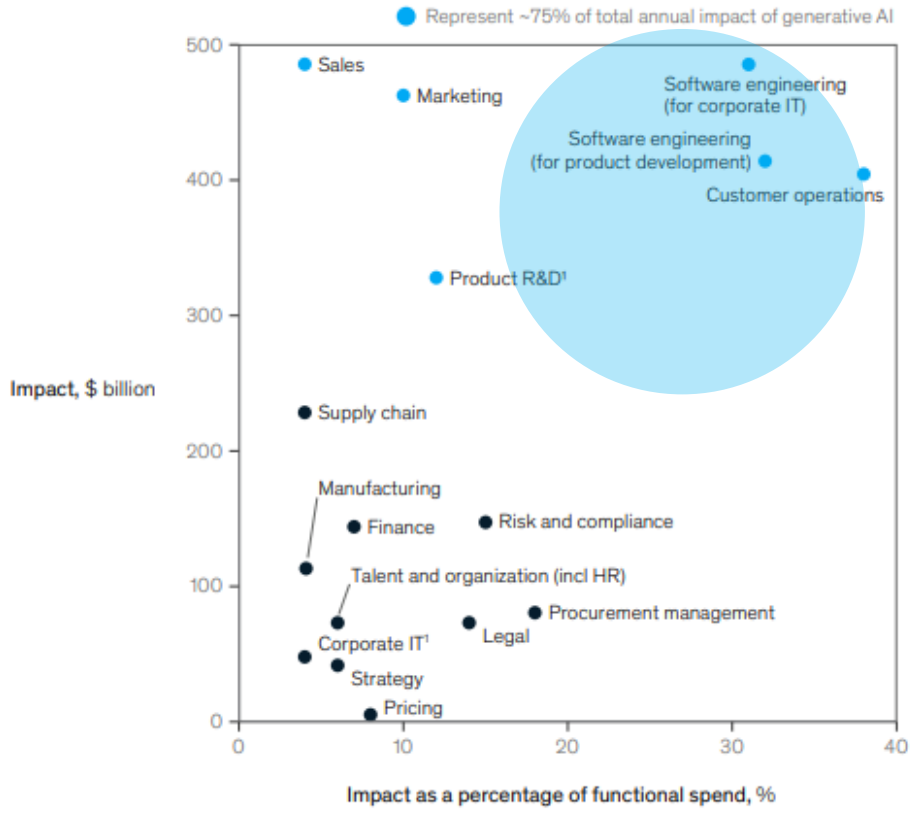
Generative AI use cases will have different impacts on business functions across industries.

Generative AI productivity impact by business functions[1]

Low impact — High impact

	Total, % of industry revenue	Total, $ billion	Marketing and sales 760–1,200	Customer operations 340–470	Product R&D 230–420	Software engineering 580–1,200	Supply chain and operations 290–550	Risk and legal 180–260	Strategy and finance 120–260	Corporate IT[2] 40–50	Talent and organization 60–90
Administrative and professional services	0.9–1.4	150–250									
Advanced electronics and semiconductors	1.3–2.3	100–170									
Advanced manufacturing[3]	1.4–2.4	170–290									
Agriculture	0.6–1.0	40–70									
Banking	2.8–4.7	200–340									
Basic materials	0.7–1.2	120–200									
Chemical	0.8–1.3	80–140									
Construction	0.7–1.2	90–150									
Consumer packaged goods	1.4–2.3	160–270									
Education	2.2–4.0	120–230									
Energy	1.0–1.6	150–240									
Healthcare	1.8–3.2	150–260									
High tech	4.8–9.3	240–460									
Insurance	1.8–2.8	50–70									
Media and entertainment	1.8–3.1	80–130									
Pharmaceuticals and medical products	2.6–4.5	60–110									
Public and social sector	0.5–0.9	70–110									
Real estate	1.0–1.7	110–180									
Retail[4]	1.2–1.9	240–390									
Telecommunications	2.3–3.7	60–100									
Travel, transport, and logistics	1.2–2.0	180–300									
		2,600–4,400									

Using generative AI in just a few functions could drive most of the technology's impact across potential corporate use cases.



Note: Impact is averaged.
[1]Excluding software engineering.
Source: Comparative Industry Service (CIS), IHS Markit; Oxford Economics; McKinsey Corporate and Business Functions database; McKinsey Manufacturing and Supply Chain 360; McKinsey Sales Navigator; Ignite, a McKinsey database; McKinsey analysis

Source: McKinsey('23.6)

산업(X)+생성AI: 성장통

생성 AI의 부정적 측면을 개선하기 위한 기술 향상 및 법·제도 개선 노력이 필요

(기술) 정확도 오류

- '21년 9월까지의 과거 학습데이터에 기반하므로 시의성 오류 존재
- 의료 분야에 활용 시 정확도 문제 해결 필요

※ OpenAI의 CEO인 Sam Altman은 중요한 업무를 ChatGPT에 의존하는 행위는 실수가 될 것이라고 우려를 표명

(산업) 빅테크 중심의 승자 독식, 일자리 감소 문제

- 대규모 데이터, 컴퓨팅 인프라, 인재, 자본력의 시너지에 비례하여 기존 글로벌 빅테크가 유리한 위치
- 생성AI는 노동자 대체가 가능하며, 전문직에도 영향

※ 英 DailyMail은 ChatGPT 등 생성AI가 5년내 현재 직업의 20%를 차지할 것

생성AI

(법·제도) 저작권 이슈

- AI 기반 창작물에 대한 학습 데이터 관련 저작권 이슈 지속 발생 및 소송 사례 다수 발생

※ Getty Image는 AI 이미지 생성 개발사 Stability AI에 저작권 침해 소송 제기, 성명문 발표

(보안) 개인정보 및 금융정보 유출 문제

- 생성AI 알고리즘에 주입된 민감한 개인 정보가 제3자에게 그대로 유출되어 금융사기·범죄에 악용 가능
- 통신, 금융 등 산업에서는 고객 개인정보를 다루는 특성상 보안 우려로 생성AI의 사용을 금지하는 추세

※ JP모건, BOA 등 월가 은행은 생성AI 기반 챗봇 사용을 금지 (Bloomberg)

新AI 생태계의 출현: AI가치사슬 통합화

AI반도체(Chips)-클라우드(Cloud)-AI플랫폼(Model)-애플리케이션(Applications)의 가치사슬로 연결

- 국내외 선도 AI 기업들은 저비용, 고성능을 위한 기술 최적화를 위해 AI 생태계의 수직적 통합 추진

초거대 AI생태계 현황 – 글로벌

글로벌 빅테크를 중심으로 투자 및 전략적 제휴를 통한,
초거대 AI 생태계(반도체-클라우드-AI플랫폼-애플리케이션)의 수직적 통합이 가속화

독자 AI서비스

이미지 영상 분야에서 주로 활용

- stability.ai: (서비스) Text-to-Image 서비스용 사용자 도구 DreamStudio제공('22.12~,beta) (기반모델) Stable Diffusion 2.0 ('22.11), StableDiffusion 2.1('22.12), SDXL 0.9 ('23.6)
- runway: (서비스) Text-to-Video, Image-to-Video 서비스 제공, 사용자 콘텐츠 제작을 위한 30개이상의 "AIMagic"도구가포함된 크리에이티브제품구축 (기반모델)Gen-1('23.2), Gen-2('23.3)

AI 앱· 서비스

OpenAI의 기반모델 중심의 문서 생성 및 업무 생산성 지원에서 잠재력 가시화

OpenAI GPT기반	CarMax, Viable, MEM, Agolia, CopyAI, Debuild.co 등 300개의 이상 AI App이 OpenAI社의 GPT-3 API를 기반으로 개발 ('21.3월 기준) ChatGPT이후 지난 '23년 3월 23일 공개한 ChatGPT 플러그인스토어에 등록된 Plugins는 약 629여개('23.7.7 기준)
MS-OpenAI(GPT-4)	MS의 다양한 오피스제품(Word, Excel, PowerPoint, Outlook 등)군에 ChatGPT 기능 도입한 'Microsoft 365 Copilot'발표('23.3.15), GPT-4통합깃허브 CopilotX 출시 예정('23.3월 계획 발표)

AI 플랫폼(모델)

빅테크기업의 자체 모델 개발과 스타트업 투자 병행으로 초거대 생성 AI 기술 주도

OpenAI GPT4.0('23.3.15)	Google PaLM('23.3.14)	Meta LLAMA('23.2.24)	Hugging Face BLOOM('22.7)	ANTHROPIC Claude('23.3.14)
- ChatGPT(GPT-3.5, 파라미터 1750억개)기반, 파라미터수는 비공개, 멀티모달 지원 - ChatGPT PLUS로 유료이용 - PluginStore 개시('23.3)	- 5,400억개 파라미터 - PaLM API와 개발 지원 도구 MakerSuite 공개 - Bard에 PALM2('23.5) 적용 (파라미터 3,400억개추정)	- 상대적으로 적은 파라미터 (기본형 650억개)로도 GPT-3보다 벤치마크 성능 우수 - OPT-175B에 이어 공개 - LLaMA2 공개('23.7)	- 1,000명의 연구자 커뮤니티가 참여한 프랑스 BigScience 프로젝트와 공동개발한 파라미터 1,760억개 의 초거대 언어모델로 깃허브에 공개	- ChatGPT 대항마 'Claude' 개발 - Google은 Anthropic에 약 3억 달러 투자 ('23.2) - Claude2 발표('23.7)

AI클라우드

마이크로소프트, 구글, 아마존AWS 등 AI서비스 주도권 확보를 위한 각축전

Microsoft, Azure	Open AI와의 독점적 계약으로 Open AI의 AI모델 훈련 및 서비스 배포에 활용	-MS Azure-OpenAI API서비스제공중이며, 메타, Hugging Face 등도 Azure활용, Meta의 LLAMA2 협업('23.7)
Google Google Cloud	구글은 자사 AI칩(TPU)을 기반으로 클라우드를 구축하고 AI 컴퓨팅 인프라 제공	- 구글자체언어모델(LaMDA, PaLM 등) 및 자회사 DeepMind의 AI모델(Gopher 등) 훈련
Amazon aws	클라우드 서비스 선두 기업으로 컴퓨팅 인프라 없는 AI모델 기업과 전략적 제휴	- Stability AI, Hugging Face 등 AI모델개발 특화 기업들에게 전략적 인프라제공 협력 -다양한 AI기반모델을 활용해 기업의 생성 AI모델 서비스를 지원하는 'Amazon BedRocks' 공개('23.4)

AI 반도체

엔비디아(Nvidia)의 독주가 지속, AI 반도체 스타트업 경쟁 가세

NVIDIA	AI 반도체 시장의 80%이상 점유, 가장 최신의 GPU기반 DGX H100 시스템 공급 시작('23.1), 생성AI 개발 지원 클라우드인 '엔비디아 AI 파운데이션' 공개('23.3)	AMD	GPU 시장 2위, 세계 최대 '프로그래머블반도체 (FPGA)' 업체인 자일링스 500억달러에 인수('22.2), 대규모AI모델개발 특화 GPU 'MI300X' 공개('23.6)
intel	이스라엘 AI반도체 하나바랩스 인수(20억달러)를 통해 AI반도체 시장에 진출, 2025년 AI 전용 반도체(GPU)인 '팔콘쇼어' 출시 계획 발표('23.5)	SambaNova SYSTEMS	SW와 HW가 통합된 AI 지원 반도체 지향형 美스타트업('17년설립), 인텔, 구글에서 1.5억달러의 투자유치, 학습과추론을동시에수행할수있는 DataScale SN30 출시('22.9),

※ 출처 : 소프트웨어정책연구소(2023)

초거대 AI생태계 현황 – 국내

네이버 하이퍼클로바(HyperClova) 플랫폼을 중심으로 선도적으로 국내 AI 생태계 구축 중이며 IT대기업이 기술력, 인프라를 바탕으로 경쟁에 합류

독자 AI서비스

LLM 기반 신규 AI서비스

Saltlux	· 기업용 LLM루시아(LUXIA) GPT('23.9)
KONAN TECHNOLOGY	· 13B 코난LLM 자체개발공개('23.8)
maum.ai	· 마음GPT-13B, AI휴먼 M3 · 초거대AI클라우드 플랫폼운영
allganize	· 알리(Alli) GPT3.5기반, 기업용 챗봇 서비스
NC	· 자체개발 LLM VARCO 고액('23.8) 1.3B, 6.4B, 13B 모델, 시나리오개발

AI 앱· 서비스

헬스케어, 레저, 금융, 광고 등 다양한 분야에서 활발하게 적용

챗GPT기반 서비스	업스테이지(Askup), 뤼튼(문서생성), 굿닥(건강 AI 챗봇), 마이리얼트립(숙소구매 등 여행플래너) , 라이너(보험챗봇)등
하이퍼클로바 기반 서비스	잡브레인(AI자소서생성), 라이팅젤(자소서 자동 완성, 소설 창작), 뤼튼(카피라이팅), 킵그로우(블로그 포스팅)

AI 플랫폼

초거대 AI 기술 수준 고도화 및 비영어권 시장 공략

HyperClova(Naver)	KoGPT(카카오브레인)	믿음(KT)	AI 엑사원(LG)	에이닷(SK)
- 파라미터 2,040억개 - GPT-3대비 한국어 데이터 6,500배 이상 학습 - 서치GPT개발 및 출시('23.하반기), - HyperClova X 공개 ('23.8.24)	- 파라미터 300억개 - KoGPT활용 특정 분야 전문 AI버티컬 서비스 출시 계획(~'23) - KoGPT2.0 출시계획('23 下)	- AI컨택센터와 기가지니 적용 - 에이센클라우드 연내출시예정 - 맞춤형 초거대AI모델제작 도구 '믿음 렛츠' 출시 준비 - 국내 AI 반도체 스타트업 리벨리온과 함께 GPU 팜 구축추진	- 국내 최대규모(파라미터 3천억개) - 한국어,영어 원어민 수준 구사 - 계열사 AI시범 적용 - 셔터스톡, 파슨스과 협력 - 이미지생성AI 캡셔닝AI공개('23.6) - EXAONE 2 공개('23.7)	- 파라미터 수백억개 - 계열사 및 자회사와 협력하여 자체 생태계 구축(예, 자체 AI생태계 '아이버스'에 탑재 검토) - SKT, Anthropic에 투자(1B)('23.8)

AI클라우드

AI 도입을 통한 글로벌 빅테크 추격

네이버클라우드	초거대 AI(하이퍼클로바)를 적용한 플랫폼 내 AI 기반 서비스를 API 형태로 제공	- 클로바스튜디오 내 튜닝기능 제공(1천여개 기업이 사용 신청, '23.2. 기준)
KT 클라우드	자체 개발 AI를 탑재하여 금융·e커머스·헬스케어 등의 분야로 사업 확장	- HW+SW의 풀스택 솔루션을 한 번에 제공하여 글로벌 경쟁력 제고
NHN클라우드	AI 기술을 접목한 특화 상품 제공 및 바우처 공급	- AI반도체기업 '퓨리오사AI ' 와 컨소시엄을 통해 AI 반도체 솔루션 개발 협력 TF 출범('22.12.)

AI 반도체

전통 반도체 제조기업과 AI플랫폼, AI반도체 스타트업과 협력체계 구축

SAMSUNG 삼성전자	네이버, 세미파이브와 협력하여 AI반도체 및 솔루션 개발 중	세미파이브 : AI 칩 생산 효율화 부문에서 협력, 퓨리오사 '워보이' 양산 시작('23.4) 네이버 : 하이퍼클로바 구동을 위한 AI반도체 솔루션 개발 추진('22.12.)
SK 하이닉스	AI반도체 스타트업(사피온, 파두, 알세미 등)과 협력하여 AI시장 진출 계획	사피온: AI 추론에 특화해 효율성 극대화한 AI 칩 X220 출시('20) , X300시리즈 '23년下 출시 알세미: SK하이닉스 사내벤처로 '20년 분사하여 AI 반도체 모델 솔루션 개발에 주력

※ 출처 : 소프트웨어정책연구소(2023)

新AI 생태계의 출현: AI플러그인스토어

인터넷, 모바일 패러다임으로 전환기마다 있었던 '웹/앱 생태계'의 확산 전략의 연장

인터넷 혁명기

과거

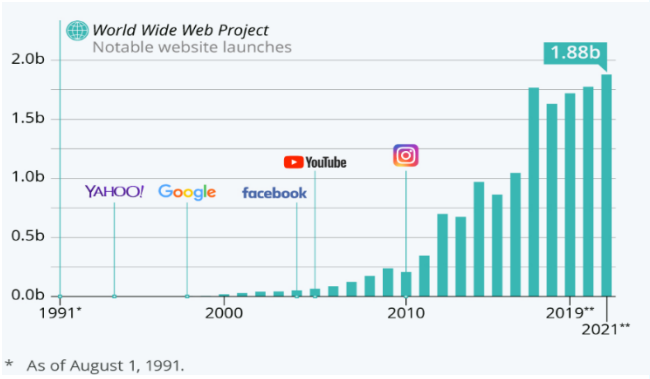
로컬 DB
파일시스템

현재

웹사이트
(하이퍼텍스트)

인터넷 웹

WWW 프로젝트 수: 0→18억8천만개



모바일 혁명기

과거

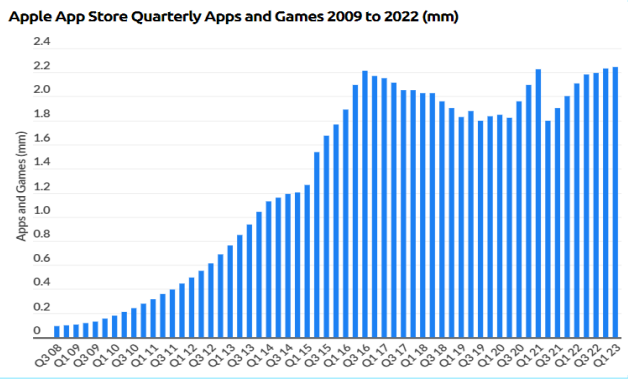
통신사 폐쇄형
(예: WIPI)

현재

AppStore
개발형 플랫폼
(3rd 파티 생태계)

모바일 앱 (애플 앱스토어)

앱 수: 0→218만4천여개



AI 혁명기

현재

서비스제공자 모델(OpenAI), 파인튜닝 모델

미래

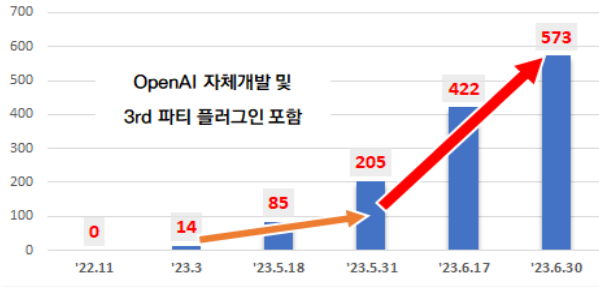
AI Plugin 개방형 플랫폼
(3rd 파티 생태계)

AI플러그인

3rd 파티 수: 14개('23.3.)→573개('23.6.)

(763('23.7.26)→856('23.8.28))

OpenAI ChatGPT 플러그인 수 ('23.6.30)



※ 출처 : 소프트웨어정책연구소(2023)

新AI 생태계의 출현: OpenAI 플러그인 생태계

파운데이션 모델의 한계(실시간,정확성 등)를 극복하고, 신뢰성 부담을 절감
기존 서비스 제공자들은 플러그인을 통해 생성AI 기능 활용을 극대화하고 서비스를 고도화

ChatGPT 플러그인 (예시)

- 익스피디아 (Expedia) 숙박·항공권 예약
- 피스컬노트 (FiscalNote) 정책·법률 정보 검색
- 인스타카트 (Instacart) 식료품 주문 배송
- 카약 (Kayak) 숙박·항공권 예약
- 클라나 쇼핑 (Klarna Shopping) 온라인 쇼핑
- 마일로 패밀리 (Milo Family) 가족 돌봄 서비스
- 오픈테이블 (OpenTable) 레스토랑 예약 서비스
- 샵(Shop) 온라인 쇼핑
- 스픽(Speak) 외국어 교육 앱 플랫폼
- 울프럼 (Wolfram) 컴퓨터 수치 및 통계 계산
- 재피어(Zapier) 업무 자동화 도구

864개(8.23)

OpenAI

플러그인 기능 제공 효과성

생성AI 플러그인 활용
어플리케이션·서비스·제품

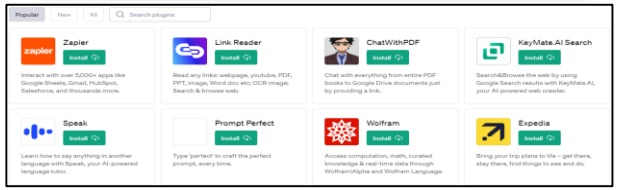
제공 서비스
성능 고도화
및 다양화

협력체계

외부유입
신뢰성 부담 절감

생성AI 플랫폼
(초거대 AI 파운데이션 모델)

플러그인 생태계 확산 배경



- ChatGPT가 제공하는 GPT API 활용
- 챗봇의 기능 극대화 및 서비스 결합
- 제공되는 서비스 다양화

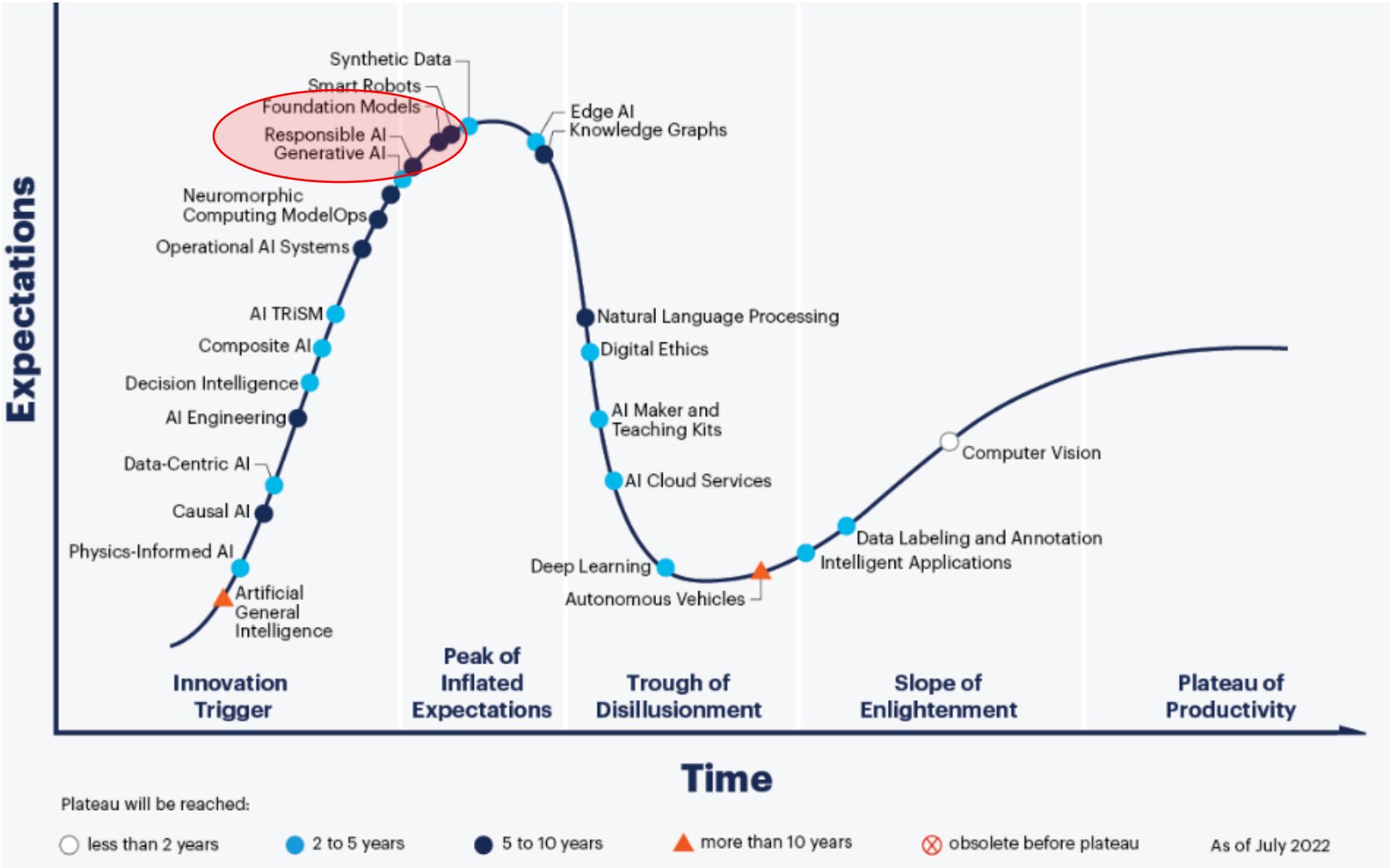
- 외부의 서비스를 ChatGPT에 적용
- 기존 한계점(실시간·정확성 결여 등)을 보완
- 맞춤형 솔루션을 제공 & 신뢰성 부담 절감

플러그인 생태계 확산

※ 출처 : OpenAI, ChatGPT plugins announcements

생성AI는 기대의 정점을 지나는 과정

올해를 지나며 환상이 거치고, 본격적인 USE CASE와 안정적 서비스 운영 방식의 모색 예상



Source: Gartner(2022)

요약 및 시사

- 패러다임
  - (일상으로 침투하는 GenAI) Generative Search Experience(GSE), Copliot, Subscription 모델
- 사업 전략
  - (전략적 제휴와 수직 통합화) 가치 사슬 전단계에서의 수직 통합화 및 전략적 제휴 강화, 경쟁과 공존의 전략 필요
  - (AI플러그인생태계) 태동하고 있는 AI판 앱스토어, 플러그인생태계의 주도권 확보, 신뢰성 검증의 효율화 및 수익 모델 고도화
  - (초거대 대안으로서 sLLM) 특수 도메인 타겟, 모델 개발 및 검증 효율성, Single Chip 활용 가능
  - (스타트업으로 성장 가능성) OpenAI, Anthropic, Cohere, AI12Lab, Stability.AI, HuggingFace의 성공 모델
  - (초거대 AI의 투자의 불균형) 초거대 투자, 설비투자 및 운영비용 측면에서 거대 기업 유리
  - (가치사슬에서 새로운 기회) 에너지 , 칩설계, AI인프라SW, 데이터 신디케이션, AI모델허브, 신뢰성 인·검증, 뉴디바이스
  - (오픈소스전략) 직접 개발 vs. 오픈소스 활용 사이의 전략적 선택 필요, Meta의 오픈소스 AI모델 전략의 의미
- 기술 신뢰성
  - 데이터 편향성, 공정성, 설명가능성, 투명성 확보
  - 생성AI의 실시간성 확보, 할루시네이션 이슈, 다국어 지원 문제 해결
  - 저비용 고성능 초경량 모델 개발 , 전력 소모 및 탄소 배출 대응
- 법제도 대응
  - (저작권) 학습데이터 저작권 침해, 생성물 저작권 인정, 기밀 데이터 유출
  - (불법데이터 유통) 가짜 뉴스(텍스트), 이미지 생성 및 유통에 따른 사회 혼란 방지
  - (규제준수) 글로벌 인공지능 규제(EU AI Act, WH AI Blueprint, UK AI Standard Roadmap) 동향 모니터링 및 준수 전략