ISSUE REPORT I 2023.11.9. IS-165

생성 AI 산업 생태계 현황과 과제

The current status and challenges of the generative AI ecosystem

유재흥, 안성원, 안미소, 노재원

SPRi 소프트웨어정책연구소
Software Policy & Research Institute

이 보고서는 「과학기술정보통신부 정보통신진흥기금」에서 지원받아 제작한 것으로
과학기술정보통신부의 공식의견과 다를 수 있습니다.
이 보고서의 내용은 연구진의 개인 견해이며, 본 보고서와 관련한 의문 사항 또는
수정·보완할 필요가 있는 경우에는 아래 연락처로 연락해 주시기 바랍니다.

소프트웨어정책연구소
유재홍 책임연구원(jayoo@spri.kr)

CONTENT

요약문

ChatGPT의 등장과 함께 생성 AI 생태계 주도권 확보를 위한 경쟁이 갈수록 치열해지고 있다. 빅테크 기업들은 보다 저렴한 비용으로 고품질의 인공지능 서비스를 제공하기 위한 전략으로 생성 AI 생태계의 가치 사슬을 수직 통합하는 방향을 택하고 있다. 반도체-클라우드-AI 모델(플랫폼)-AI 애플리케이션으로 이어지는 가치 사슬 간의 시너지를 최대로 이끌어내기 위해 엔비디아(NVIDIA), 마이크로소프트(MS), 오픈AI(OpenAI) 등과 같은 주요 기업들 사이에 전략적 제휴가 이뤄지고 있으며, 구글(Google), 메타(Meta) 등의 선도 기업들은 독자적인 가치 사슬을 확대하기 위한 투자를 진행중이다. 국내 기업들도 인공지능 생태계 내에서 경쟁력을 확보하기 위해 이와 유사한 수직 통합 전략을 채택하고 있다. 퓨리오사AI, 리벨리온 같은 AI 특화 반도체 스타트업과 KT, 네이버클라우드 등의 클라우드 인프라 기업 간에 전략적 협력이 진행중이고, 업스테이지, 뤼튼 등의 애플리케이션 개발사들도 클라우드 인프라 업체와의 협력 또는 자체 생태계 확보를 위해 투자하고 있다. 향후 생성 AI를 필두로 AI 기반모델을 제공하는 글로벌 선도 기업간 인공지능 패권 경쟁의 승자가 시장을 독식할 가능성이 높아지고 있다. AI 안전성과 신뢰성을 확보하기 위한 연구 개발 투자, 공공부문에서의 AI 리터러시 및 일상화 전략을 통한 인공지능 시장 기반 구축, 그리고 국내 실정에 적합한 AI 법제 정비 등을 통해 국내 인공지능 산업과 기업 생태계의 육성을 위한 공공의 역할이 필요한 시점이다.

Executive Summary

With the advent of ChatGPT, the competition to secure a leading position in the generative AI ecosystem is becoming increasingly fierce. Big tech companies are opting to vertically integrate the value chain of the generative AI ecosystem as a strategy to provide high-quality artificial intelligence services at a more affordable cost. Strategic alliances are being formed among major companies such as NVIDIA, Microsoft (MS), and OpenAI to maximize the synergy between the value chains that extend from semiconductors to cloud, AI models (platforms), and AI applications. Leading companies like Google and Meta are also making investments to expand their own value chains. Domestic companies are adopting similar vertical integration strategies to secure competitiveness within the AI ecosystem. Strategic collaborations are underway between AI-specialized semiconductor startups such as FuriosaAI and Rebellions, and platform companies like KT and Naver Cloud. Application developers like Upstage and Wrtn technologies are also investing in collaborations with platform companies or in securing their own ecosystems. Moving forward, with generative AI leading the way, there is an increasing likelihood that the winners of the artificial intelligence hegemony competition, centered around global leading companies, will monopolize the market. In light of this, there is a pressing need for public sector involvement to foster the domestic AI industry and corporate ecosystem. This necessitates a strategic investment in research and development to bolster the safety and reliability of AI technologies. Concurrently, there is a need to lay a robust foundation for the AI market, achieved through enhancing AI literacy and implementing strategies that promote the widespread adoption of AI within the public sector. Additionally, it is imperative to update and refine AI legislation, ensuring it is aptly tailored to accommodate the unique conditions and requirements of the domestic landscape.

I. 논의 배경

■ 생성 인공지능(Generative AI)의 부상으로 전 산업에 걸친 디지털 혁신 전환 가속화 및 글로벌 인공지능(AI) 패권 경쟁 본격화

- AI 산업에서 MS, 구글 등 해외 빅테크(Big tech.) 기업들이 주도권을 확보한 가운데, 가치 사슬 강화를 위한 기업 간 합종연횡이 활발히 진행 중
  - AI반도체, 클라우드, 초거대AI모델, AI애플리케이션으로 이뤄진 AI 생태계 가치사슬을 엔비디아, MS, 구글 등 선도기업들이 수직 통합화
    * 엔비디아는 MS, 메타 등 빅테크와 전략적 협력, 구글은 AI인프라-클라우드-서비스에 이르는 풀스택 강화
  - 국내에서도 AI반도체 스타트업, AI애플리케이션 개발사들과 네이버, 통신사 등 인프라 및 AI모델 개발 기업과의 전략적 협력이 진행 중
    * (AI반도체 스타트업) 리벨리온, 퓨리오사AI (AI플랫폼) 네이버 하이퍼클로바X, LG 엑사원, KT 믿음 등

■ 글로벌 AI 경쟁이 점차 치열해지는 가운데, 생성 AI를 비롯한 AI 전반의 국가 경쟁력 확보를 위한 분석과 다양한 전략 추진이 필요

- AI 기술은 국가 핵심 안보 기술로 부상하고 있으며, 향후 글로벌 빅테크 기업에 의한 국내 시장 잠식 및 기술 종속 우려가 높아지고 있는 상황

- 글로벌 빅테크들은 거대 자본력을 기반으로 인프라 투자에 적극적이며, 이에 대항하여 국내 AI 생태계를 활성화하기 위한 전략 모색이 중요한 시점

■ 이 보고서에서는 국내·외 생성 AI 산업 생태계 현황을 분석하고, 우리나라가 AI 산업 주도권 확보를 위한 정책적 요소들을 도출하고자 함

- 생성 AI 산업의 정의와 범위, 산업의 발전 방향, 국내외 주요 기업들의 패권 경쟁의 지향점과 전략 비교를 통한 국내 산업 현황 분석

- 민간의 자본력과 기술력이 주도하고 있는 생성 AI산업 생태계에서 국내 기업들의 육성과 성장 지원을 위한 정부의 역할 모색

II. 생성 AI 산업 생태계 현황 분석

2.1 생성 AI의 부상

■ (개념) 생성 AI는 기존의 데이터를 기반으로 새로운 데이터 및 결과물을 생성하는 AI 기술

- 사용자의 입력을 기반으로 사용자가 원하는 결과를 유추해 텍스트, 이미지, 오디오 등 다양한 형태의 결과물을 만들어 내는 AI 알고리즘 (SPRI, 2023)[1]

■ (특징) 생성 AI는 데이터의 분포를 학습하고 이와 유사한 분포를 추정하여 데이터를 생성

- 주로 데이터의 진위여부를 판단하는 전통적 판별(discriminative) 모델과 달리 다양한 언어, 이미지, 영상 정보를 입력받아 추론 정보를 출력하는 생성 모델이 최근 비약적으로 발전
  - 대량의 데이터 학습 모델로 오토인코더(Auto-Encoder)[2], GAN(Generative Adversarial Network)[3] 등을 시작으로 2017년 이후 트랜스포머(Transformer) 아키텍처 등장 후 비약적 발전

- 기반 모델(Foundation Model)[4]은 대표적인 생성 AI 모델로, 대규모 데이터를 학습해 다양한 작업(질의응답, 정보추출, 번역, 객체 인식 등)에 적용할 수 있으며, 파인 튜닝(미세조정)을 거쳐 특정 작업에 특화 가능
  - 기반 모델의 매개 변수에 따라 모델의 규모가 상이하며 매개 변수의 크기에 따라 성능이 높아지는 것으로 보고되어 수천억 개에 이르기까지 규모가 점차 커지고 있는 상황
    * GPT-1(1.17억개, '18.6) → GPT-2(15억개, '19.2) → GPT-3(1,750억개, '20.6) → GPT-3.5(1,750억개, '22.11) → GPT-4(미공개, '23.3)

1 소프트웨어정책연구소, 이슈리포트 – 생성 AI의 부상과 산업 변화, 2023. 6

2 입력 데이터를 최대한 압축시킨 후, 데이터의 특징을 추출하여 다시 본래의 입력 형태로 복원하는 기계학습 기법으로 1990년대 초 제안됨

3 생성적 적대 신경망 모델은 진짜같은 이미지를 생성하기 위해 2014년 이안 굿페로우 등에 의해 제안된 개념으로 생성자와 판별자를 두고 서로 경쟁적으로 학습하는 과정에서 서로가 서로를 구분할 수 없는 단계까지 학습하는 기법

4 스탠포드 인간 중심 AI 연구소(HAI)의 기초 모델 연구 센터(CRFM)가 2021년 8월에 사용한 용어로 "기반 모델"은 광범위한 데이터를 이용해 사전 학습된 모델을 지칭

- 생성 AI는 방대한 양의 데이터로 사전 훈련된 기반 모델을 중심으로 의료·제조·교육·금융·미디어 산업 등 다양한 분야의 태스크 수행

- 그러나, 확률적 추론에 의한 결과 생성으로 발생하는 환각(hallucination), 학습 데이터에 따른 편향된 결과 도출 등 기술적 한계와 프라이버시, 저작권 등 사회적 이슈도 야기

■ (시장) 글로벌 생성 AI 시장은 향후 10년간 연평균 24% 이상 성장이 예상되며, 경제적 파급효과는 연간 2.6조 달러 이상*으로 전망

- 전 세계 생성 AI 시장 규모는 2023년 448억 9천만 달러에서 2030년 2,070억 달러로 전망되어, 2022년부터 2032년까지 연평균 24.4% 성장할 것으로 예상(Statista, 2023)[5]

  * △전체 AI 시장은 연평균 24.5% 성장해 2025년 약 1,350억 달러에 이를 것(Gartner 2023), △생성 AI가 연간 2.6~4.4조 달러의 잠재적 가치를 생산할 것으로 추정(McKinsey, 2023)

- 생성 AI 기술에 대한 산업계의 지출 비중은 2020년 1% 미만에서 2032년 12%로 성장하여, 향후 10년간 산업기술 분야에서 생성 AI의 영향력이 확대될 것으로 전망(Bloomberg, '23.6)[6]

  * △IDC는 2023년 전세계 생성AI 솔루션 지출 규모가 약 160억 달러, 2023-2027년간 연평균 73.7% 성장해 2027년 1,431억달러로 성장 전망(IDC, 2023)

■ (산업) 미디어 산업을 중심으로 제조, 금융, 의료 등 다양한 산업에서 확산 전망

- 생성 AI 시장은 2022년 기준 미디어 및 엔터테인먼트 부문이 34%로 가장 높을 것으로 예상되고, 자동차·운송산업, 금융업, 헬스케어 순으로 생성 AI 시장 점유율이 높을 것으로 추정(Precedence Research, '23.6)[7]

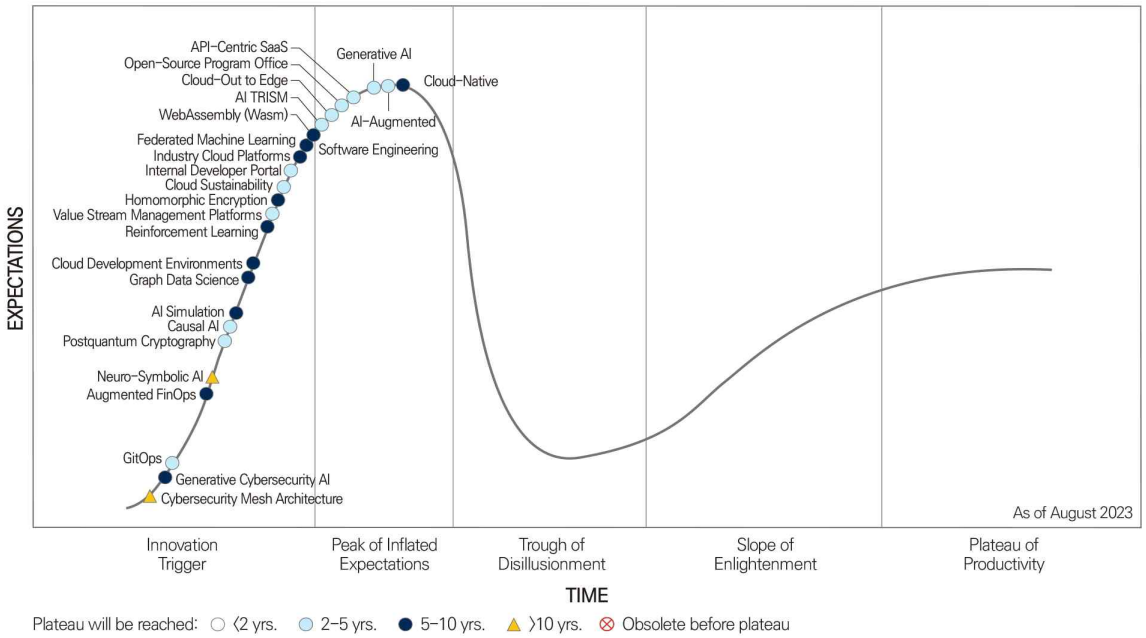
  - 생성 AI의 영향력은 SW엔지니어링 분야에서 가장 활발할 것으로 분석 (Mckinsey, '23.6)

- 한편, 가트너의 전망에 따르면 생성 AI는 올해 기대감의 정점에 달한 상태로 향후 2년에서 5년 내에 실질적인 혁신 성과가 나타날 것으로 예측

5 Stastia Market Insights, 2023

6 Bloomberg, Generative AI to Become a $1.3 Trillion Market by 2032, Research Finds, 2023.6.

7 https://www.precedenceresearch.com/generative-ai-market



※ 자료: Gartner, 2023.8

[그림 2-1] 2023년 가트너 이머징 기술 하이프 사이클('23.8)[8]

2.2 글로벌 생성 AI시장 양상 : 빅테크기업의 각축전

■ 챗GPT의 등장 이후, 생성 AI 시장은 글로벌 AI 기업들의 전장으로 변모

- OpenAI, MS, 구글이 중심이 되어 진행되던 생성 AI 경쟁이 메타, 아마존, 엔비디아, 애플 등 글로벌 빅테크 기업들의 참여로 각축전 양상
  - 주요 빅테크 기업들의 자체 AI 반도체 개발, 클라우드 확보, AI 모델과 서비스 또는 애플리케이션 생태계를 구축함으로써 생성 AI 산업의 주도권 확보 경쟁

- 국내 기업들도 글로벌 빅테크에 대응하기 위한 전략적 방안으로 기업 간 협력을 통한 수직 통합화를 적극 추진 중
  - AI 반도체 스타트업과의 협력으로 인프라 비용 절감, 한국어 특화 AI 모델을 통해 한국어 서비스 경쟁력 확보, 글로벌 AI서비스 제공 역량 강화 등 추진

8 Gartner, Gartner Places Generative AI on the Peak of Inflated Expectations on the 2023 Hype Cycle for Emerging Technologies, 2023.8.16

■ 다양한 AI 기업들이 각자의 영역에서 기회를 도모

● 기업들은 고유 역량(하드웨어(AI반도체), 클라우드, AI모델, 서비스)을 앞세워 전략적 투자 및 제휴로 생성 AI 가치 사슬[9]의 수직 통합화 추진([그림 2-2])

※ 자료: Andreeseen Horowitz 참고, 소프트웨어정책연구소 작성

[그림 2-2] 생성 AI 가치 사슬

2.3 해외 기업 동향

■ (AI 인프라) 엔비디아가 주도하는 AI반도체 시장과 스타트업의 가세

● **(엔비디아)** AI 학습에 필요한 GPU 시장의 90% 이상 점유한 것으로 추정

- 업계에서 1만 개 이상의 GPU 프로세서를 통해 초거대 AI 모델 구축을 지원하는 기업은 엔비디아가 유일한 것으로 추정

9 글로벌 투자 기업인 Andreeseen Horowitz는 인공지능 산업의 구조를 하드웨어(AI반도체), 클라우드, AI모델(폐쇄형, 개방형, 허브형), 애플리케이션(B2C, B2B)으로 구분하고 있으며 본 연구에서도 이를 참고하여 분석함

* 글로벌 4대 클라우드 데이터센터(AWS, MS-Azure, Google Cloud, AliCloud)에서 사용하는 AI 가속기의 97.4%가 엔비디아 제품으로 조사(Liftr, '23.5)

- 엔비디아의 핵심 경쟁력은 대규모 GPU 병렬 프로그래밍을 가능하게 하는 CUDA플랫폼으로 개발자들이 효율적으로 엔비디아의 하드웨어를 최대화하여 사용할 수 있도록 장기간 최적화된 소프트웨어 라이브러리와 툴을 제공

- AI알고리즘에 특화된 NPU 제품 중 엔비디아 GPU보다 나은 성능을 보이는 프로세서도 있으나 다수 프로세서의 연결을 통한 통신 속도에서는 엔비디아가 독보적이라는 평가

* 대부분의 반도체 회사가 사용하는 칩간 통신 플랫폼인 PCI5.0은 초당 128기가바이트(GB) 수준이나 엔비디아의 경우 NV링크 스위치라는 자체 인터페이스를 통해 초당 전송속도를 900GB 수준으로 상향[10]

● **(구글) 구글은 자체 AI칩(TPU) 기반의 클라우드를 구축하고 고도화**

- 구글 텐서 처리 장치(Tensor Processing Unit, TPU)는 구글에서 2016년 5월에 발표한 데이터 분석 및 딥러닝용 하드웨어로 구글 자체 텐서플로 SW를 이용

* 구글은 2015년에 내부적으로 TPU를 사용하기 시작했으며 2018년 판매를 시작했고, '22년 10월부터 제4세대 TPUv4를 구글 클라우드를 통해 제공하기 시작

- 구글은 오클라호마 데이터 센터에 TPU v4를 이용해 9엑사플롭스(Exa Flops)[11]의 연산 성능을 지원하는 세계 최대 머신러닝(ML) 클러스터를 구축한 것으로 알려짐

* 구글 TPUv4 Pod는 오클라호마 지역에서 칩마다 시간당 0.97~3.22달러로 제공되며 구글의 가장 작은 인스턴스 1년 약정 가격은 월 5,924달러 ('22.10)[12]

● **(메타) 자체 설계 AI 반도체를 공개하며 컴퓨팅 인프라 기술 확보**

- 메타는 AI 프로그램 구동을 위해 2020년부터 개발한 것으로 알려진 1세대 맞춤형 실리콘 '메타 트레이닝 및 추론 가속기(MTIA)'를 발표 ('23.5)

* MTIA는 병렬로 작동하는 회로 블록의 메시 구조로 구성되었으며 메타의 AI 프레임워크인 '파이토치(PyTorch)'를 사용해 최적화 소프트웨어를 실행

- MTIA는 AI 전용 반도체로 학습과 추론 동시 처리할 수 있으며 단일 칩에서는 GPU보다 최대 3배의 1와트당 초당 부동소수점연산 수에서 효율적으로 평가

- 현재 메타는 리서치슈퍼클러스터(RSC)에 엔비디아 제품 중심으로 컴퓨팅 인프라를 구축했으나 향후 의존도를 낮추며 2025년 실전 투입 계획[13]

10 조선비즈(2023.6.2.), 1만개 GPU가 하나로 움직인다
11 초당 100경(京) 번의 연산을 수행하는 단위로 페타플롭(Peta Flops, 초당 1,000조)보다 1천 배 빠름
12 AI타임즈(2022.10.12.), 구글 엑사스케일 4세대 TPU AI시스템 공개

● **(스타트업) 미국의 삼바노바를 비롯해 세레브라스, 그래프코어, 누비아, 헤일로, 웨이브컴퓨팅, 그록 등이 AI 반도체 시장에 참여**

- 삼바노바 시스템즈(SambaNova)는 '23년 3월 맞춤형 생성 AI 모델 구축 및 서비스 플랫폼인 '삼바노바 스위트(SambaNova Suite)*'를 공개

  * 삼바노바는 리눅스 기반 소프트웨어 스택 '삼바플로우(SambaFlow)'위에 작동하는 AI 전용 장비인 데이터스케일(DataScale) 시스템을 출시(SN30)('22.10) 했으며 8소켓 엔비디아 DGX A100보다 8배 더 빠르게 작동[14]

■ (AI 클라우드) 새로운 성장 모멘텀을 기대하는 클라우드 빅3

● **(MS Azure) 애저(Azure)-OpenAI 서비스로 AI 시장 주도권 확보**

- 생성 AI 모델 설계 및 실행에 특화된 새로운 클라우드 서비스를 추가*

  * 챗GPT 이후 급증한 고객사의 생성 AI, LLM 개발 및 연구 요구에 맞춰 4세대 인텔 제온프로세서와 엔비디아 H100 등으로 구성된 클라우드 인프라 확충

- 애저 클라우드 플랫폼에서 제공하는 AI 개발툴(Azure AI Studio)과 오픈AI 머신 러닝 모델을 사용해 기업들의 자체 코파일럿(Copilot) 개발 지원

● **(구글 GCP) 기계학습 플랫폼인 '버텍스 AI(Vertex AI)'로 생성 AI 모델 개발 지원**

- 구글의 LLM인 '팜2(PaLM2)'와 동영상 생성 도구 '이매진(Imagen)', 코드 생성 도구 '코디(Codey)' 등 생성 AI도구를 일반에 제공함으로써 개발자 유입

  * 구글 클라우드의 '버텍스 AI 모델 가든'에 접속해 모델 직접 사용, 미세조정 등 작업할 수 있으며 모델 가든에는 100개 이상의 기반 모델(Foundation model) 포함

- 구글 클라우드에서 생성 AI앱 빌더 (Generative AI App Builder)*를 지원해 개발자들이 업무용 생성 AI 앱을 만들거나 데이터를 분석할 수 있도록 지원[15]

  * 젠 앱 빌더는 복잡한 AI 모델 생성 및 관리, 배포 과정을 단계별로 자동화하고 단순화해 효율적으로 생성 AI를 활용할 수 있도록 오케스트레이션 기능을 지원

● **(아마존 AWS) 기업용 클라우드 서비스 '베드록'을 출시하며 경쟁 가세**

- 아마존은 '23년 4월 애플리케이션 프로그래밍 인터페이스(API)를 통해 아마존 및 AI 스타트업의 기반 모델을 사용할 수 있도록 하는 '베드록(bedrock)' 출시

13 DigitalToday(2023.5.19.), 메타 전용AI칩 개발 프로젝트 공개
14 AI타임즈(2022.9.16.) 스타트업 삼바노바 차세대 통합 AI플랫폼 '데이터스케일' 출시
15 ZDNET(2023.3.31.), 구글 비전문가도 간단히 생성AI 다루는 도구 공개

- 개발자는 베드록 서비스를 통해 프롬프트 기반 텍스트 생성, 이미지 생성 등 아마존의 기반 모델*을 자체 데이터로 훈련시켜 독자적인 모델 구축 가능

  * AWS는 자체 LLM인 '타이탄 텍스트(Titan Text)'와 '타이탄 임베딩스(Titan Embeddings)'를 함께 출시('23.4)

- AWS는 엔비디아와 협력해 생성 AI를 위한 클라우드 서비스 고도화하고, 허깅페이스, 스태빌리티AI 등과 제휴로 각사의 AI 기반모델*을 활용한 개발 환경 제공

  * 베드록에서 사용가능한 기반모델로는 AWS의 Titan을 포함해 AI21Labs의 Jurassic-2, Anthropic의 Claude2, Cohere의 Command and Embed, Stability.ai의 Stable Diffusion 등이 포함

■ (AI 모델) 빅테크와 스타트업이 경쟁하는 사적 모델과 공개 모델

● (OpenAI GPT-4) GPT-4 기반의 ChatGPT 서비스 제공하여 초기 주도권 확보

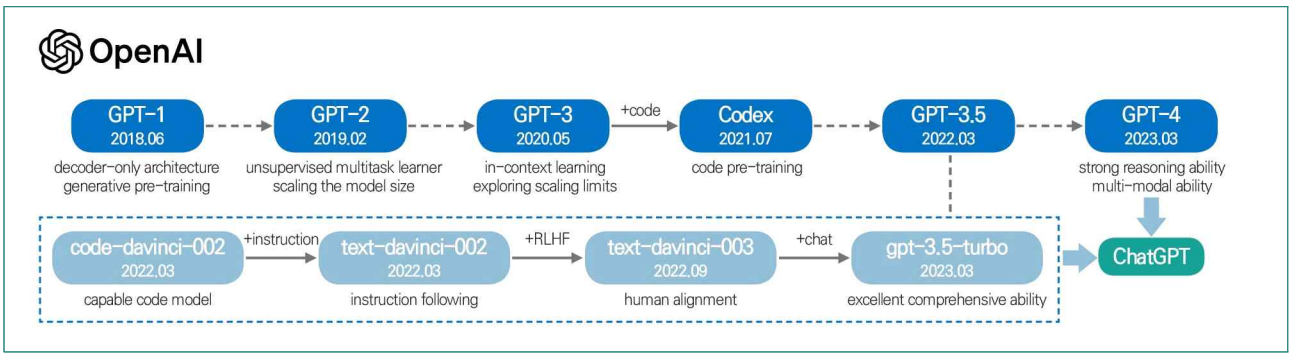
- OpenAI는 1,750억 개의 파라미터를 통해 학습한 초거대 언어모델 GPT-3.5을 기반으로 AI챗봇 서비스인 ChatGPT를 '22년 11월에 출시

  * 주간 활성 이용자 1억명, 포춘500대 기업의 92% 이상이 ChatGPT를 활용하는 것으로 보고(OpenAI, '23.11)

- '23년 3월부터 GPT-3.5에서 개량된 GPT-4 모델*을 기반으로 서비스 제공하고 있으며 11월 6일 첫 OpenAI 개발자 행사에서 GPT-4 터보** 모델 공개

  * GPT-4는 멀티모달 입력을 제공하고, 입력 토큰수가 약 4천 개에서 32천 개로 증가했으며, 언어 이해력이 향상(영어이해 70% → 85.5%)되고 할루시네이션 문제도 개선된 것으로 보고

  ** 2023년 4월까지 데이터 학습, 입력 토큰수 128천개로 확대, 이미지생성과 텍스트-음성 변환 지원, 입력 토큰은 $0.01로 GPT-4보다 1/3 수준으로, 출력 토큰은 $0.03으로 1/2 수준으로 이용자 비용을 낮춤



※ 자료: Zhao et al., (2023), A Survey of Large Language Models

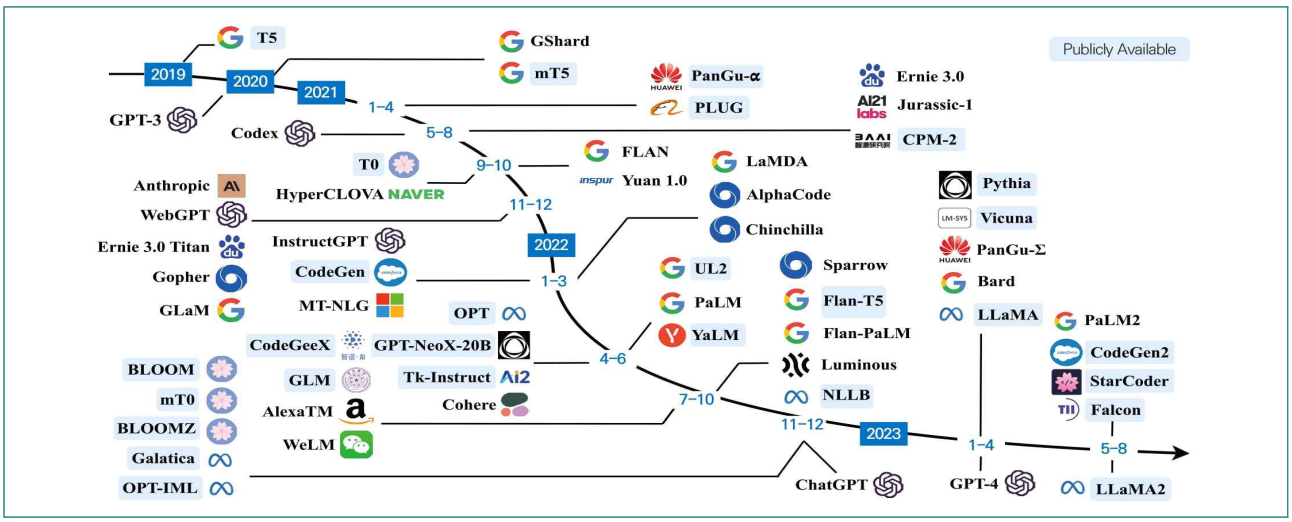
[그림 2-3] OpenAI의 LLM 개발 현황

● (구글 PaLM2) 구글 '바드(Bard)'의 기반 모델로 PaLM2 활용

- 구글은 LaMDA 기반의 챗봇 바드를 공개('23.2)한 후, '23년 5월 180개국에 전체 공개 서비스를 출시하면서 기반언어모델을 PaLM2로 변경

* PaLM2는 '22년 공개된 구글 PaLM의 업그레이드 모델로서 파라미터 3,400억 개 사용하여 100개 이상의 언어 학습, 상식 추론, 수학, 논리에서 우수한 평가를 받음

- **(기타)** Meta의 LLaMA, 허깅페이스의 BLOOM 등은 오픈소스로 공개
  - 메타는 생성 AI 모델 LLaMA(7B, 13B, 30B, 65B)('23.2), LLaMA2('23.7)[16]를 공개하며 연구자들이 자유롭게 활용토록 하였으며, 허깅페이스 역시 BLOOM(176B) 공개
    * 메타의 LLaMA는 상대적으로 적은 규모의 모델로 GPT3(175B)를 능가하는 성능을 나타내었고 스탠포드大 연구진을 LLaMA을 파인튜닝한 'Alpaca 7B'을 개발 공개



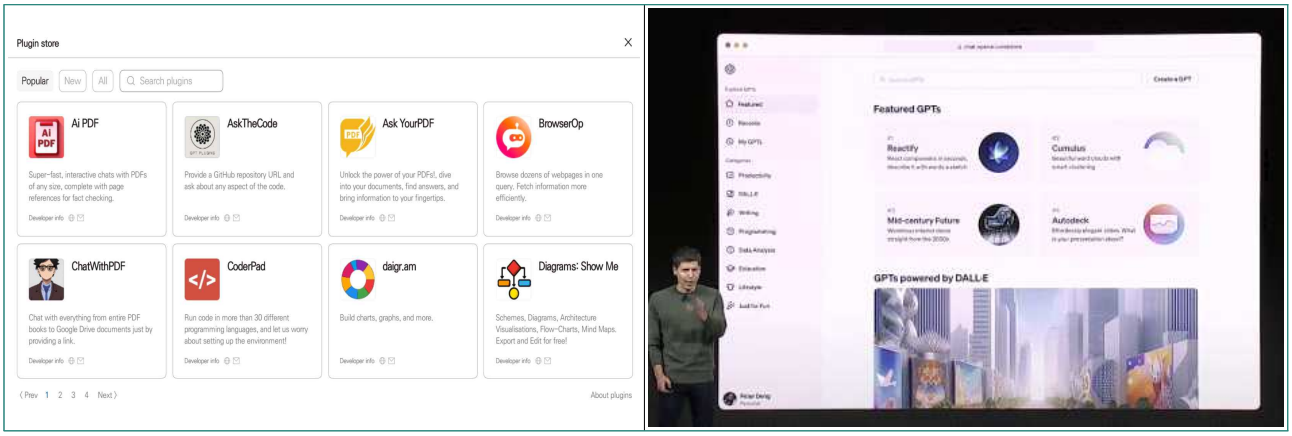
※ 자료: Zhao et al., (2023), A Survey of Large Language Models

[그림 2-4] 초거대 AI 모델 개발 현황

■ (AI 서비스) AI 플러그인 생태계 중심의 애플리케이션 시장 형성

- **(OpenAI)** 챗GPT에서 3rd party 앱의 기능을 연결해 활용할 수 있는 플러그인스토어(Plug-in Store)를 출시('23.3)하고 맞춤형 AI챗봇 마켓플레이스인 'GPT스토어' 출시 준비중('23.11)
  - 플러그인을 통해 챗GPT의 한계인 실시간 정보 검색, 계산, 외부 서비스와 상호 작용 등 일반인 대상의 서비스 경험 확대
    * OpenAI 자체적으로 Browsing with Bing(실시간 검색), Advanced Data Analytics(Code Interpreter), Dall·E 3 플러그인 제공을 하고 있으며 일부 플러그인은 유료가입자에 한해 이용 가능

16 2023년 7월 18일, LLaMA 업그레이드 모델 LLaMA2 공개(7B, 13B, 70B), LLaMA 대비 학습 토큰수는 2TB으로 40% 증가, 입력 토큰(Context length)도 2K → 4K로 확대, 무료로 상업적 활용이 가능하며 월 활성 사용자(MAU)기준 7억 명이 넘을 경우 메타와 별도 라이센스 계약 필요, Microsoft와 전략적 협력을 통해 MS Azure에서도 이용가능



※ 자료: (좌)Chat.Openai.com (총 Plug-ins 수는 '23.10.30 기준 1,030개), (우) OpenAI 개발자 행사에서 맞춤형 GPTs 소개('23.11.6)

[그림 2-5] OpenAI의 AI 플러그인과 GPTs 예시

- GPT스토어는 사용자들이 노코드 방식으로 개발*한 맞춤형 AI챗봇서비스를 사고 팔수 있는 마켓플레이스 플랫폼으로 11월 중 유료가입자를 대상으로 선보일 계획[17]

  * Assistants API와 함께 새로운 GPT 개발을 지원하는 사용자 정의 지시사항 및 코드 인터프리터(Code Interpreter), 외부 데이터 및 지식 검색(Retrieval), 함수호출(Function Calling) 기능을 활용해 구현

● **(MS) 코파일럿(MS Copilot) 애플리케이션 생태계 구축**

- 자사의 MS-Office 365 제품군과 생성 AI(챗GPT) 기능을 연동해 생산성을 지원하고 깃허브(Github)의 코파일럿(Copilot)을 통해 AI 페어 프로그래밍 지원

  * MS Office 365의 유료(월 30달러) 구독 정책 발표('23.7)하고 11월 제품 출시 예정

- 애저-오픈AI 서비스를 사용해 써드파티(3rd party) 개발자들이 AI앱을 개발하면 플러그인 표준과 상호운용성이 확보되어 다양한 MS의 제품군에서 활용할 수 있도록 함

● **(구글) 챗봇 '바드'와 써드파티 앱 통합 시사**

- 자사의 생성 AI 모델을 검색서비스, 생산성 앱(지메일, 구글 Docs, 유튜브, 구글맵 등)에 통합 계획 발표('23.9)

- 외부 써드파티 앱에 구글의 생성 AI 기능 통합 확장*계획으로, 카약, 오픈테이블, 인스타카드, 울프람 등 OpenAI 플러그인을 제공하는 기업들과도 협력 추진

  * (예시) 어도비의 생성 AI 모델 제품인 '어도비 파이어플라이(Adobe Firefly)'를 바드에 통합해 바드에서 이미지 생성 멀티모달 서비스 지원[18]

17 OpenAI, New models and developer products announced at DevDay, 2023.11.6
18 서울경제(2023.5.11.), 구글 '오픈AI 플러그인 저격, 생태계 확장 전략 맞불

2.4 국내 기업 동향

■ (AI 인프라) AI 반도체 스타트업들이 국내 플랫폼 업체들과 협력 추진

● **(리벨리온AI) KT와 전략적 투자를 통해 KT 인프라에 리벨리온 칩 탑재**

- 2020년 설립된 리벨리온은 KT로부터 300억원의 전략적 투자 유치하고 사업 협력 ('22.7)
  * KT는 리벨리온의 AI 반도체 '아톰(ATOM)'의 성능을 개선하고, 초거대 AI 모델에 최적화한 '아톰플러스(ATOM+)' 개발과 적용 추진
- 생성AI 시장을 목표로 삼성전자와 협력해 AI반도체 '리벨' 개발추진('23.10)
  * 삼성전자 파운드리 4나노공정을 이용하고 삼성의 차세대 고대역메모리(HBM3E)탑재 추진

● **(퓨리오사AI) 국내외 AI모델 개발사들과의 전략적 제휴 확대**[19]

- 2017년 창업한 퓨리오사AI는 카카오엔터프라이즈, 허깅페이스, LG AI연구원 등과 협력 추진
  * △(카카오엔터프라이즈) 카카오클라우드에 퓨리오사AI의 워보이(Warboy) NPU탑재 서비스 제공('23.2), △(허깅페이스) 차세대 AI반도체 개발 협력('23.2), △(LG AI연구원) 퓨리오사의 2세대 AI칩 레니게이드로 자사 초거대AI모델인 엑사원 기술 검증('23.6)
- 네이버클라우드에 AI 반도체(NPU) 공급사로 참여하면서 교육 플랫폼 기업 엘리스그룹의 GPU·NPU 클라우드 플랫폼 및 AI반도체 팜 구축* 추진('23.9)
  * 과학기술정보통신부가 2030년까지 3단계에 걸쳐 추진하는 'K-클라우드' 프로젝트의 1단계로 국산 NPU를 데이터센터에 적용하고 클라우드 기반 AI서비스까지 제공하는 실증 사업

● **(사피온) SK관계사들이 세운 스타트업으로 SK의 AI 사업 협력**

- 2021년 SK ICT 3사(SK텔레콤, 하이닉스, SK스퀘어)가 협력해 설립된 사피온은 국내 최초로 데이터센터용 반도체 X220(28나노 공정)을 2020년 양산해 공급
  * 2023년 8월 600억원 규모 시리즈A 투자 유치 완료로 AI반도체부터 SW, 서비스에 이르는 풀스택 서비스 추진
- LLM을 지원하는 X330*를 공개, TSMC 7나노 공정을 활용 이르면 연말 상용화 ('23.11)
  * AI 학습용으로도 가능하나 추론에 특화된 NPU로 전작 X220 대비 연산 속도 4배, 전력 효율 2배 이상 증가,
- 향후, 스마트폰 등 엣지 디바이스용 사피온X350은 내년 상반기 공개, 자율주행용 X340은 2026년 상반기 양산 예정, HBM을 탑재한 차세대 NPU X430*은 2025년 말 공개 목표[20]

19 동아일보(2023.6.15.) 국산 AI반도체 사업에서 과반이 '퓨리오사AI'선택

20 전자신문(2023.11.16.), 사피온, AI반도체 'X330'출시

■ (AI 클라우드) 글로벌 빅3(MS, 구글, AWS)에 대항해 네이버, KT 등 자체 클라우드 기반 구축

● **(네이버) 클라우드 비용 효율성 확보를 위해 하드웨어 기업과의 협력 강화**

- 네이버클라우드-삼성전자가 FPGA 형태로 AI 반도체 솔루션을 우선 개발하고 테스트를 통해 주문형반도체(ASIC)로 생산하는 협력 방식 추진 중[21]
- 인텔과 함께 고가의 GPU를 CPU로 대체하는 방식으로 AI 서버 구축해 운영 비용 절감[22]

* 네이버 플레이스 AI모델 인프라에 적용해 추론 프로세스는 GPU, 각종 전처리와 결과 후처리는 CPU와 CPU 최적화 소프트웨어(인텔의 파이토치 확장팩)를 통해 연 4억원 수준의 비용 절감 ('23.10)

● **(KT) 엔비디아에 대한 의존도가 높은 상황에서 국내 AI 반도체 스타트업과의 전략적 제휴를 통해 데이터센터 확충**

- KT는 리벨리온이 개발한 데이터센터용 AI반도체 '아톰'을 KT IDC에 적용하고, 자사 초거대 AI서비스 '믿음(Mi:dm)'에도 탑재 예정
- AI 인프라SW 전문기업인 '모레(MOREH)*'에 2021년 전략 투자를 진행하고 AI 학습용 클라우드 서비스, 초거대 AI 모델 개발에 협력

* 모레(MOREH)는 AI 개발자들이 기존에 존재하는 다양한 AI모델들을 코드변경 없이 엔비디아 GPU가 아닌 다른 GPU 및 AI프로세서들에서도 그대로 사용할 수 있도록 풀스택 솔루션 제공

■ (AI 모델) 대기업은 매개변수 천 억개 이상의 초거대모델, 중소기업은 상대적 경량 모델에 집중[23]

● **(초거대모델) 대규모 자본과 인프라가 요구되는 매개변수 수천억개 이상의 초거대 언어 모델은 LG, 네이버, KT 등 대기업이 중심이 되어 진행 중**

- LG AI 연구원은 2021년 12월 엑사원(매개변수 약 3,000억개)을 공개한 후 지난 7월 엑사원2 공개하고, 전문가용 챗봇(유니버스), 화학·바이오 등 산업 도메인 특화 서비스(디스커버리), 이미지 생성 및 캡셔닝 AI(아틀리에) 세 가지 종류의 플랫폼 발표('23.7)

* 4천500만 건의 전문 문헌과 3억 5,000만 장의 이미지를 학습, 고비용 문제를 해결하기 위해 AI 모델의 경량화, 최적화를 통해 언어 모델의 추론 시간은 25% 단축, 메모리 사용량은 70% 감소, 비용도 약 78% 절감

- 네이버클라우드는 2021년 5월 세계에서 세번째로 자체개발 초거대 AI모델인 '하이퍼클로바(매개변수 2,040억개)'를 공개한 후 올해 8월 진화 버전인 '하이퍼클로바X' 발표*

21 중앙일보(2023.4.10.), 국산 AI반도체로 세계클라우드 시장 진출, 또다른 한류

22 전자신문(2023.10.30.), 고가 GPU 대신 CPU로 AI를 인텔, 네이버와 AI서버 구축

23 초거대언어모델은 일반적으로 매개변수 수천억개 규모를 갖는 트랜스포머 모델을 의미하며, 경량모델은 개발 및 운영상의 비용효율성을 고려해 개발된 수십~수백억개 수준의 상대적으로 적은 매개변수 규모를 갖는 모델로 구분하여 지칭

* 대화형 AI 서비스 '클로바X'('23.8), 생성 AI 검색 '큐:'('23.9), 블로그 저작툴인 Clova for Writing 베타테스트('23.10), 보안성이 강화된 'Neurocloud for HyperClovaX'('23.10) 출시 등 자사 검색, 쇼핑, 광고 등 주요 서비스의 생성 AI 접목 확대 계획

- KT도 초거대AI 플랫폼 믿음(Mi:dm) 출시('23.10)

* 4종(베이직, 스탠다드, 프리미엄, 엑스퍼트)의 모델 발표, 70억 개의 파라미터를 갖는 모델은 연구개발용으로 무료 사용 가능하며 최대 2,000억 개 파라미터 탑재 모델 준비 중

● **(경량모델) 코난테크놀로지, 솔트룩스, 뤼튼, 업스테이지 등 생성 AI서비스 중소 개발사들은 상대적 경량화를 통해 비용효율성을 높인 자체 모델을 개발하거나 및 타사 기반모델을 활용**

- (코난테크놀로지) '코난LLM'이라는 비용효율성을 높인 자체 LLM을 개발하고 온프레미스 형태로 B2B, B2G 시장 공략 계획 발표('23.8)

* 20억건의 양질의 학습용 문서를 확보해 학습 토큰 4,920억 개(한국어 2,940억 개) 사용, 파라미터(매개변수) 131억 개, 410억 개 버전으로 발표, 연내 학습토큰 7,000억개를 사용한 후속 모델 발표 계획

- (솔트룩스) '루시아GPT'라는 자체 LLM을 개발하였으며 지식그래프 기술 등을 활용해 환각오류를 최소화해 온프레미스 B2B, B2G 시장에 진출 계획 ('23.9)

* 자체 축적 한글데이터 1TB 활용, 학습 토큰 5,000억 개 이상 사용해 70억 개(7B), 130억 개(13B), 200억 개(20B), 500억 개(50B) 모델 공개, 연내 100B 모델 출시 예정

- 뤼튼, 업스테이지 등은 네이버 하이퍼클로바, OpenAI의 ChatGPT, 오픈소스 LLM 등을 자체적으로 활용해 시험 서비스를 개발하거나 상용화 추진 중

■ (AI 서비스) AI 플러그인 생태계 중심의 애플리케이션 시장 형성

● **(네이버) 하이퍼클로바X 플랫폼에 플러그인 서비스인'스킬' 탑재 ('23.8)**

- 네이버클라우드의 '하이퍼클로바X'는 '스킬'로 명명한 OpenAI의 '플러그인' 기능과 유사한 써드파티 앱 연동 기능을 결합해 네이버의 주요 서비스(블로그, 쇼핑 등)의 정보 활용 계획

● **(뤼튼) 자사 AI챗봇과 연계한 플러그인 플랫폼인 '뤼튼2.0' 출시('23.4)**

- 뤼튼테크놀로지스는 GPT-4를 탑재한 '챗뤼튼'을 무료 공개하고, 외부데이터와 서비스를 연계하여 이용할 수 있는 플러그인서비스 '뤼튼2.0'을 공개

● **(기타) 생성 AI 기술을 활용한 다양한 스타트업들이 서비스 개발**

- 올거나이즈코리아-알리GPT('23.2), 업스테이지-아숙업('23.3), 엔씨소프트-바르코('23.8), 포티투닷-챗베이커 ('23.7) 개발

〈표 2-1〉 국내외 생성 AI 기업 현황

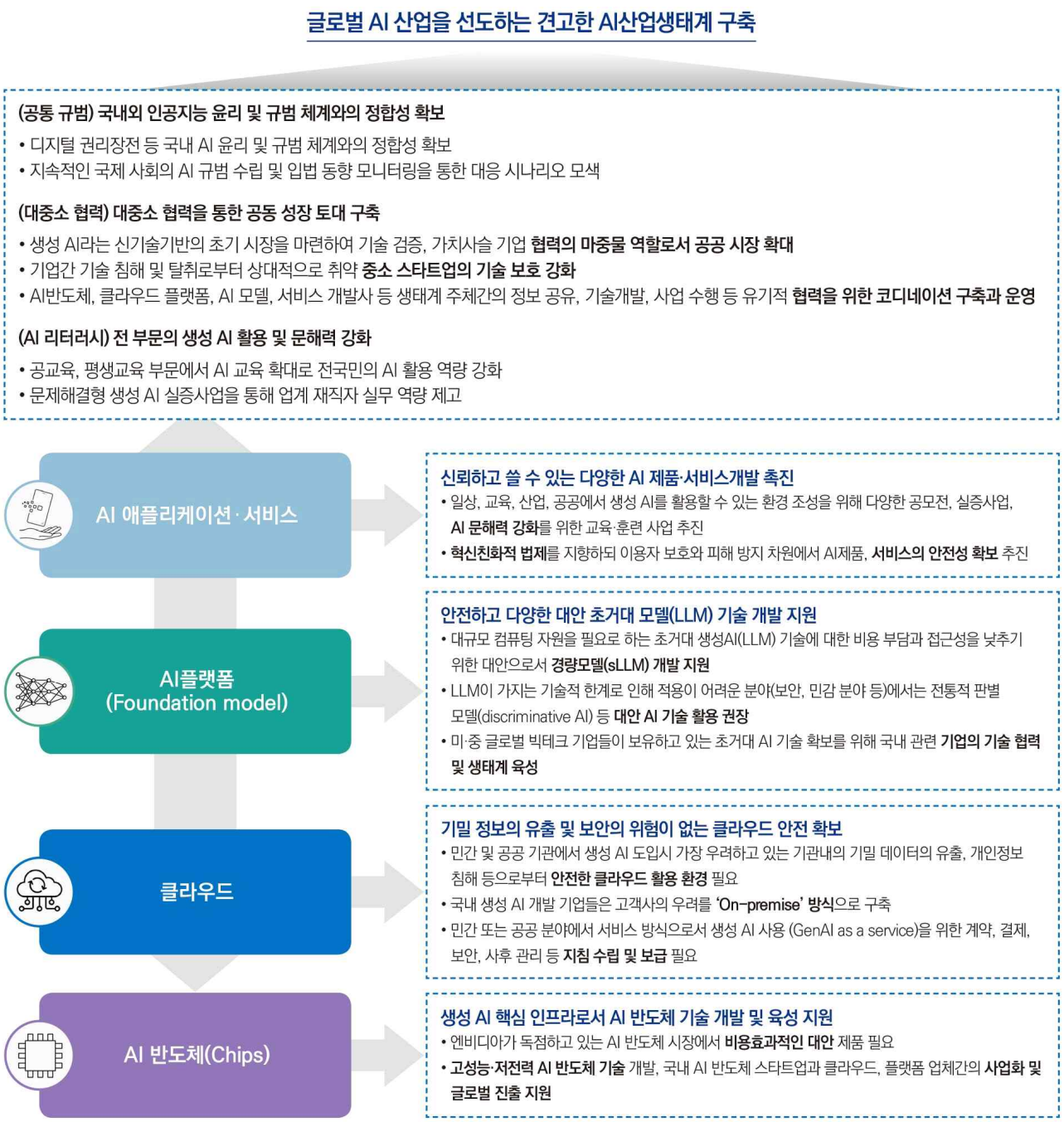
구분	기업	현황
애플리케이션 (서비스)	MS	• 자사 검색서비스 Bing에 ChatGPT 기능 연동('23.2) • MS Azure를 통해 Azure-OpenAI 서비스 제공 중이며 최근 Meta의 오픈소스 모델인 LLaMA2도 AI 모델 카탈로그에 등록해 서비스 • 빙챗에서 이미지생성 AI인 오픈AI의 Dalle-3 무료 제공('23.9), MS office 365 유료 버전 출시예정('23.11), MS Windows Copilot 출시 예정,
	구글	• 바드와 검색서비스, 생산성앱 통합 계획 발표('23.9.) - 이메일, Google Docs, Google Meet 등 다양한 생산성 도구에 자사 생성 AI인 Google Bard 결합, 구글 검색에 Bard 기능 통합 • 외부 앱의 바드기능 통합 확장 계획
	오픈AI	• 챗GPT 플러그인 서비스 출시('23.3.), ChatGPT Plus 유료 서비스('23.2) • ChatGPT 엔터프라이즈서비스('23.9) 출시, 코드 생성 AI Codex('21.11) • 이미지생성 AI Dalle-2('22.3), Dalle-3 출시('23.9) • 챗GPT 유료 서비스에 플러그인 서비스로 Dall*E 3 추가 • Assistants API를 통해 사용자 맞춤형 AI챗봇을 개발을 지원하고 사용자 개발 GPT마켓플레이스인 'GTP스토어' 출시 계획 발표('23.11)
	어도비	• Adobe Firefly 생성 AI 기능으로 이미지, 영상 생성 ('23.3)
	네이버	• 하이퍼클로바 플랫폼에 플러그인 서비스인 '스킬' 탑재 ('23.8) • 생성 AI 기반 검색 큐: 출시('23.9)
	뤼튼	• 카피라이팅, SNS 문구, 자기소개서, 책초안 작성 등 콘텐츠 생성 서비스 • 자사 AI챗봇과 연계한 플러그인 플랫폼인 '뤼튼2.0' 출시('23.4)
	몽키런	• MonkeyLearn:감정 및 의도 분석을 포함한 텍스트 분석 작업을 제공하는 코딩이 필요 없는 AI 도구
	카피AI	• 비즈니스를 위한 고품질 카피를 생성하는 AI 기반 카피라이터 • 긴 형식의 블로그 기사부터 어조 바꾸기 도구, 판매 문구 생성기
	재스퍼	• 2021년부터 광고 카피나 블로그 기사, SNS 게시물 등 50여가지 스타일의 글쓰기 생성 AI 서비스 출시
AI기반모델	구글	• PALM, PALM2, 연내 차세대 멀티모달 생성 AI '제미니' 출시 예정
	MS	• 멀티모달 대형언어모델 '코스모스-1' 공개, 1.6B개의 파라미터로 훈련('23.3) • 파이썬 코드 생성 모델 '파이-1'공개('23.6), 이를 업그래이드한 매개변수의 13억개의 멀티모달LLM 파이-1.5(phi-1.5) 오픈소스로 공개('23.11.3) • LLaMA기반 상대적 경량LLM(매개변수 13B) Orca 오픈소스 공개 ('23.6), LLaMA2기반 Orca2(매개변수 7B, 13B) 공개('23.11)
	오픈AI	• ChatGPT('22.11), ChatGPT-4('23.4), GPT-4 터보공개('23.11) • 연내 멀티모달 생성 AI 'GOBI' 출시 추정
	앤스로픽	• 신뢰성 확보에 초점을 Claude('23.3), Claude2 출시('23.7)
	아마존	• AI모델 허브 서비스 아마존 베드록 출시('23.4), 자사 Titan 보유
	xAI	• 일론 머스크가 2023년 7월에 설립한 AI 스타트업에서 '그록(Grok)' 출시('23.11) • 매개변수 330억개를 가진 Grok-0의 업그래이드 버전으로 트위터 데이터 학습
	메타	• LLaMA(7B, 13B, 30B, 65B) 출시('23.2), LLaMA2('23.7)
	허깅페이스	• 오픈소스 기반 모델 BLOOM(176B) 공개('22.6.)
	런웨이	• 텍스트-이미지, 영상 생성 AI 모델, Gen-1, Gen-2 출시('23.3)
	네이버클라우드	• 하이퍼클로바 모델 출시('21.5), 하이퍼클로바X 출시('23.8)

구분	기업	현황
	카카오브레인	• GPT-3 모델의 한국어 특화 모델인 KoGPT('21.11) 공개, 이미지 생성 모델 칼로1.0('22.12), 칼로2.0 공개('23.7), KoGPT2.0 연말 공개 계획
	LG AI연구원	• 엑사원1.0 공개('21.12), 엑사원2.0 공개('23.7) – 전문가용 챗봇 유니버스, 이미지 생성 및 캡셔닝 AI 아틀리에, 산업 도메인특화 디스커버리 플랫폼 발표
	코난테크놀로지	• 자체 초거대언어모델 '코난 LLM' 공개하고 13.1B와 41B 버전 제공('23.8.)
	솔트룩스	• '루시아GPT' 자체 개발('23.9.)
	엔씨소프트	• 자체 언어모델 바르코(VARCO) 출시('23.8) • 바르코 LLM 기반 이미지 생성툴, 텍스트 생성 및 관리툴, 디지털 휴먼 생성 툴로 구성된 '바르코 스튜디오' 서비스 소개
클라우드	구글	• 기계학습 플랫폼인 '버텍스AI'로 구글의 LLM인 '팜2'와 동영상 생성 도구 '이매진', 코드 생성 도구 '코디' 등 생성 AI도구를 일반에 제공 • 생성 AI 빌더 (Generative AI App Builder)를 지원해 개발자 지원 • 젠 앱 빌더는 오케스트레이션 기능을 지원
	MS	• 생성 AI 모델 설계 및 실행에 특화된 클라우드 서비스 개시 • Azure AI Studio와 오픈AI 머신러닝 모델을 사용해 코파일럿(Copliot) 개발 지원
	아마존	• AWS에 베드록이라는 AI모델 허브 구축해 이용자가 다양한 모델을 선택해 개발할 수 있도록 함 • 자사의 Titan모델을 비롯해 앤스로픽 Claude2, 코히어, Stable Diffusion 등의 모델 활용 가능
	네이버클라우드	• 하이퍼클로바X 기반 클라우드, 뉴로클라우드 포 하이퍼클로바를 통해 보안 민감한 B2B, B2G 생성 AI 서비스 제공
	KT	• 리벨리온과 전략적 제휴를 통해 AI반도체 '아톰'을 KT IDC에 적용하고, 개발중인 자사 초거대 AI서비스 '믿음'에도 탑재 예정 • '모레(MOREH)'에 전략 투자를 진행하고('21) AI 학습용 클라우드 서비스, 초거대 AI 모델 개발에 협력
하드웨어 (AI반도체)	엔비디아	• GPU 시장의 80%이상을 차지하며 부동의 시장 지배력 확보 • A100에 이어 고성능 H100 칩을 출시('23.5)했으며 공급 부족 상황 • H100보다 속도 향상된 H200(SK하이닉스의 고대역폭메모리(HBM3E) 탑재)을 공개했으며 내년 2분기 출시 예정('23.11) • AI 개발 프레임워크인 'CUDA'를 GPU 프로그래밍 생태계 장악, 전세계 1.5만 스타트업, 4만개 기업이 활용, 쿠다 누적 다운로드 4천만 상회 (2022년에만 2,500만건)
	구글	• 데이터 분석 및 딥러닝용 하드웨어 구글 TPU 기반 4세대 TPUv4를 구글 클라우드 통해 제공('22.4.) • TPU v4를 이용해 9엑사플롭스의 연산 성능을 지원하는 세계 최대 머신러닝(ML) 클러스터를 구축
	인텔	• 이스라엘 AI반도체 하나바랩스인수(20억달러)를 통해AI반도체 시장에 진출, 2025년 AI 전용 반도체(GPU)인 '팔콘쇼어' 출시 계획 발표('23.5) • AI 개발자 클라우드 출시('23.3), AI 개발 프레임워크인 '오픈비노' 제공
	메타	• 1세대 맞춤형 실리콘 '메타 트레이닝 및 추론 가속기(MTIA) 발표('23.5.), 학습과 추론 동시 처리할 수 있으며 단일 칩에서는 GPU보다 최대 3배의 1와트당 초당부동소수점연산 수에서 효율적으로 평가
	AMD	• GPU 시장 2위, 세계 최대 '프로그래머블반도체(FPGA)' 업체인 자일링스 500억달러에 인수('22.2), 대규모AI모델 개발 특화 GPU 'MI300X' 공개('23.6)

구분	기업	현황
	삼바노바	• SW와 HW가 통합된 AI지원 반도체 지향형 美스타트업('17년설립), 인텔, 구글에서 1.5억달러의 투자유치, 학습과 추론을 동시에 수행할 수 있는 DataScaleSN30 출시('22.9),
	퓨리오사AI	• 네이버클라우드와 AI 반도체 솔루션 개발 협력 TF 출범('22.12.) • 퓨리오사AI가 최근 AI 반도체 '워보이(WARBOY)' 개발을 마치고 4월부터 삼성전자 파운드리를 통해 양산 개시 ('23.4) • 네이버클라우드와 협력해 GPU·NPU 팜을 구축, AI 교육 플랫폼 기업 엘리스그룹의 데이터 센터 구축 추진('23.9)
	사피온	• 국내 최초로 데이터센터용 반도체 X220을 2020년부터 양산해 공급 • LLM 지원 X330 공개, TSMC 7나노 공정을 활용 이르면 연말 상용화('23.11) • 스마트폰 등 엣지디바이스용 X350 내년 상반기 공개, 자율주행용 X340 2026년 상반기 양산 예정, HBM 탑재 차세대 NPU인 X430은 2025년말 공개 목표
	리벨리온AI	• LLM추론 특화AI 반도체 설계, 아이온('21.11), 아톰 출시(KT 데이터센터에 초도 물량 납품, 현재 대규모 LLM 적용 차세대 AI반도체 리벨 개발중 '24년 출시 목표) • 삼성전자와 LLM 특화 고성능 반도체 '리벨' 개발 협력 ('23.10)
	MS	• 자체 개발 AI 그래픽처리장치(GPU) '마이아 100'과 고성능 컴퓨팅 작업용 중앙처리장치(CPU) '코발트 100' 공개 (OpenAI와 테스트 완료)('23.11) • 코발트 100은 'ARM 아키텍처'에 기반을 두어 전략 소비 효율 추구 • 두 칩 모두 TSMC가 제조하며, 자체 AI기반 소프트웨제품과 애저 클라우드 서비스에 활용 계획

※ 자료: 소프트웨어정책연구소 정리, 언론 및 가사 자료 종합 (2023.11.6. 기준)

III. 정책적 시사



[그림 3-1] 생성 AI 산업 생태계 강화를 위한 정책 시사점

■ 인공지능 가치 사슬별 기업 경쟁력 강화 지원

● (애플리케이션) 신뢰하고 쓸 수 있는 다양한 AI 제품·서비스개발 촉진

- 일상, 교육, 산업, 공공에서 생성 AI의 수요 저변 확대를 위해 다양한 공모전, 실증사업, AI 문해력 강화를 위한 교육·훈련 사업 추진

  * (예) 생성 AI 기술 고도화 챌린지, 생성 AI 활용 시나리오, 예술 작품 공모전, 공공 분야 생성 AI 활용 실증 사업 추진, 공공 도서관 및 교육 과정에서 AI 활용 교육 증진

- 생성 AI를 활용한 다양한 서비스 창출을 위해 혁신친화적 규제 환경을 지향*하되 이용자 보호와 피해 방지 차원에서 AI제품, 서비스의 안전성 확보** 추진

  * AI 개발 윤리지침 보급, 민간 업계 자율 규제안 마련 및 준수 촉진

  ** 생성 AI의 공공 도입을 위한 가이드라인, AI서비스에 대한 신뢰성 검·인증 체계, 생성 AI 저작권 가이드라인, AI위험성 또는 AI 잠재적 사회적 영향력 분석 연구 추진

● (AI모델) 안전하고 다양한 초거대 모델(LLM) 대안 기술 개발 지원

- 대규모 컴퓨팅 자원을 필요로 하는 초거대 생성AI(LLM) 기술에 대한 비용 부담*과 접근성을 낮추기 위한 대안으로서 경량모델(sLLM) 개발 지원

- 한편, 환각 오류, 실시간 정보 업데이트의 한계 등 LLM이 가지는 제약으로 인해, 적용이 어려운 분야에서는 전통적 판별 모델(discriminative AI) 등 대안 기술 활용 권장

  * 의료(영상진단), 제조(불량품식별), 금융(사기 탐지), 소매(이탈 고객 예측), 교통(자율주행) 등 정확한 판단이 요구되는 분야에서 기술 고도화 필요

- 초거대 AI 기술 확보를 위해 국내 관련 기업의 기술 협력 및 생태계 육성

  * 초거대 AI모델, 인프라, 애플리케이션 개발 기업을 연계해 민간·공공의 수요 기관의 실증 사업 기획 및 확산

- 생성 AI 개발자 및 중소기업에 기술 축적의 기회를 제공하고 상용화를 위한 비용 부담 완화를 위해 오픈소스 AI 모델의 활용 지원

  * 메타(Meta)의 LLaMA 등 공개 기반 모델을 활용한 다양한 파생 모델과 활용 사례 등장 [참고3]

● (클라우드) 기밀 정보의 유출 및 보안의 위험이 없는 클라우드 안전 확보

- 민간 및 공공 기관에서 생성 AI 도입시 가장 우려하고 있는 기관 내의 기밀 데이터의 유출, 개인정보 침해 등으로부터 안전한 클라우드 활용 환경 필요*

  * 보안 확보를 위해 외부망과 분리된 사내망, 또는 물리적으로 분리된 네트워크 환경 구축을 위한 공공 수요 존재

- 국내 생성 AI 개발 기업들은 고객사의 우려를 '온프레미스'방식으로 구축 추진*

  * 네이버클라우드 뉴로클라우드, 코난테크놀로지, 솔트룩스, 마음AI 등 국내 생성 AI 모델 개발 기업들은 데이터 유출을 원천 차단한 방식의 B2B, B2G 서비스 제공 계획

- 민간 또는 공공 분야에서 서비스 방식으로서 생성 AI 사용(GenAI as a Service)을 위한 계약, 결제, 보안, 사후 관리 등 지침 수립 및 보급 필요

● (하드웨어) 생성 AI 핵심 인프라로서 AI 반도체 기술 개발 및 육성 지원

- 엔비디아가 독주*중인 AI 반도체 시장에서 비용 효과적인 대안 제품 필요**

* 엔비디아는 GPU 시장의 80% 이상을 점유하며, GPU 관련 소프트웨어(CUDA) 생태계를 견고히 구축 중

** 모레(MOREH(韓)), 모듈라(Modular(美)) 등 엔비디아 칩만 지원하는 쿠다와 달리 AI 개발자가 AMD, 인텔 및 구글과 같은 다른 회사에서 설계한 칩에서 AI 모델을 쉽게 훈련하고 실행할 수 있는 소프트웨어를 개발 추진

- 고성능·저전력 AI 반도체 기술 개발, 국내 AI 반도체 스타트업과 클라우드, 플랫폼 업체간의 사업화 및 글로벌 진출 지원 요구

■ 생성 AI 산업 생태계의 통합 경쟁력 강화 기반 구축

● (공통 규범) 생성 AI 기술 및 서비스 개발 시 국내외 AI 윤리 및 규범 체계와의 정합성 확보

- 디지털 권리장전 등 국내 AI 윤리 및 규범 체계와의 정합성 확보

* △(AI규범) AI윤리기준('20.12), 디지털 권리장전('23.9), △(AI전략) AI국가전전략('19.12), 신뢰할 수 있는 인공지능 실현전략('21.5), 초거대 AI 경쟁력 강화방안 ('23.4), 전국민 AI 일상화 실행 계획('23.9)

- 국제 사회의 인공지능 규범 정립 및 입법 동향*을 지속적으로 모니터링하고 대응 방안 모색

* 백악관 신뢰할 수 있는 AI에 관한 행정명령('23.10), EU의회의 인공지능법(안) 채택 ('23.6), G20 책임있는 AI 공동선언('23.9), G7 히로시마프로세스('23.5), 백악관 AI 권리장전 청사진('22.10), OECD AI 권고안 ('20.8)

● (대중소 협력) 대중소 협력을 통한 AI생태계 참여 기업들의 공동 성장 토대 구축

- 생성 AI라는 신기술기반의 초기 시장을 마련하여 기술 검증, 가치사슬 기업간 **협력의 마중물 역할로서 공공 시장 확대***

* 과기정통부-디지털플랫폼정부위원회는 〈민간의 첨단 초거대 인공지능 활용 지원〉 사업 추진 중[24]

- 기업간 기술 침해 및 탈취로부터 취약한 **중소 스타트업의 기술 보호 강화***

* 침해된 중소기업의 기술 또는 경영상의 정보 중 'SW 및 프로그램 파일'이 38.5%, 데이터 관련 정보(30.8%), 아이디어 및 제안서(23.1%) 순으로 나타남 (중소기업벤처부 2023[25])

- 민간 중심의 **협력 코디네이션 체계 구축 및 운영**을 통해 AI 생태계 참여 주체간의 정보 공유, 공동기술개발, 파트너십 구축 촉진 필요

24 디지털플랫폼정부위원회, 중소·벤처기업, 공공기관·지자체 대상 초거대 인공지능 활용지원, 2023.07.03.

25 중소기업벤처부, 2023 중소기업 기술보호 수준 실태조사 보고서, 2023.08

참고1 글로벌 AI 생태계 현황

독자 AI서비스

이미지 영상분야에서 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 1('23.7)
runway	(서비스) Text-to-Video, Image-to-Video 서비스 제공, 사용자 콘텐츠 제작을 위한 30개 이상의 "AI Magic" 도구가 포함된 크리에이티브 제품 구축 (기반모델) Gen-1('23.2), Gen-2('23.3)

AI 앱·서비스

OpenAI의 기반모델 중심의 **문서 생성 및 업무 생산성 지원**에서 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 1,030여개('23.10.30 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-4 통합깃허브 Copliot X 출시 예정('23.3 월 계획 발표), MS Office 365에 생성AI 기능 통합한 서비스 11월 본격 출시 예정

AI 플랫폼(모델)

빅테크기업의 자체 모델 개발과 **스타트업 투자 병행**으로 초거대 생성 AI 기술 주도

OpenAI GPT4.0('23.3.15)	- ChatGPT(GPT-3.5, 파라미터 1,750억개)기반, 파라미터수는 비공개, 멀티모달 지원 - ChatGPT PLUS로 유료이용 - PluginStore 개시('23.3) - GPT-4 Turbo 공개('23.11) - Dalle-*E 3, TTS 지원 - GPT Store 출시계획('23.11) - 사용자 GPT 개발 도구인 Assistants Playground와 Assitants API 공개('23.11)
Google PaLM('23.3.14)	- 5,400억개 파라미터 - PaLM API와 개발 지원 도구 MakerSuite 공개 - Bard에 PALM2('23.5) 적용(파라미터 3,400억개추정)
Meta LLAMA('23.2.24)	- 상대적으로 적은 파라미터 (기본형 650억개)로도 GPT-3보다 벤치마크 성능 우수 - OPT-175B에 이어 공개 - LLaMA2 공개('23.7)
Hugging Face BLOOM('22.7)	- 1,000명의 연구자 커뮤니티가 참여한 프랑스 BigScience 프로젝트와 공동개발한 파라미터 1,760억개 의 초거대 언어모델로 깃허브에 공개
ANTHROPIC Anthropic Claude('23.3.14)	- ChatGPT 대항마 'Claude' 개발 - Google은 Anthropic에 약 3억달러 투자 ('23.2) - Claude2 발표('23.7) - 구글 20억달러 추가투자('23.10)

AI클라우드

마이크로소프트, 구글, 아마존AWS 등 AI서비스 주도권 확보를 위한 각축전

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

AI 반도체

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

NVIDIA 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) - H100보다 속도 향상된 H200(SK하이닉스의 고대역폭메모리(HBM3E) 탑재)을 공개했으며 내년 2분기 출시 예정('23.11)
AMD AMD	- GPU 시장 2위, 세계 최대 '프로그래머블반도체 (FPGA)' 업체인 자일링스 500억달러에 인수('22.2) - 대규모AI모델 개발 특화 GPU 'MI300X' 공개('23.6)
intel intel	- 이스라엘 AI반도체 하나바랩스 인수(20억달러)를 통해 AI반도체 시장에 진출 - 2025년 AI 전용 반도체(GPU)인 '팔콘 쇼어' 출시 계획 발표('23.5)
Microsoft Microsoft	- 자체 개발 AI 그래픽처리장치(GPU) '마이아 100'과 고성능 컴퓨팅 작업용 중앙처리장치(CPU) '코발트 100' 공개 (OpenAI와 테스트 완료)('23.11) - 코발트 100은 'ARM 아키텍처'에 기반을 두어 전력 소비 효율 추구 - 두 칩 모두 TSMC가 제조하며, 자체 AI기반 소프트웨제품과 애저 클라우드 서비스에 활용 계획
SambaNova SambaNova	- SW와 HW가 통합된 AI 지원 반도체 지향형 美스타트업('17년설립), 인텔, 구글에서 1.5억달러의 투자유치 - 학습과 추론을 동시에 수행할 수 있는 DataScale SN30 출시('22.9)

※ 출처 : 소프트웨어정책연구소(2023.11)

참고2 국내 AI 생태계 현황

독자 AI서비스 — LLM 기반 신규 AI서비스

솔트룩스 Saltlux	- 기업용 LLM루시아(LUXIA) GPT('23.9)
코난테크놀로지 KONAN TECHNOLOGY	- 자체 초거대언어모델 '코난 LLM' 공개('23.8.)
마음AI maum.ai	- 마음GPT-13B, AI휴먼 M3 - 초거대AI클라우드 플랫폼운영
Allganize allganize	- 알리(Alli) GPT3.5기반, 기업용 챗봇 서비스
NC nc	- 자체개발 LLM VARCO 공개('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(모델) — 초거대 AI 기술 수준 고도화 및 비영어권 시장 공략

HyperClovaX(네이버클라우드)	- 자체 개발 초거대AI모델인 하이퍼클로바 공개('21.5): 매개변수 2,040억개, GPT-3대비 6,500배 많은 한국어 데이터 세트 학습 - 하이퍼클로바의 진화 모델인 하이퍼클로바X 공개('23.7): 대화형 서비스 '클로바X', 생성형 AI 검색 큐(CUE:), 및 네이버 검색, 쇼핑, 광고 등에 접목
KoGPT(카카오브레인)	- GPT-3 모델의 한국어 특화 모델인 KoGPT('21.11) 공개 - 이미지 생성 모델 칼로1.0('22.12) - 칼로2.0 공개('23.7), KoGPT2.0 연말 공개 계획
믿음(KT)	- 초거대AI모델 경량모델부터 초대형모델까지 4종 출시 ('23.10.31) - 70억 파라미터 모델은 무료개방 - 'KT믿음 스튜디오'를 통해 AI모델 및 서비스개발환경제공
엑사원(LG AI연구원)	- 엑사원 공개('21.12): EXAONE(EXpert Ai for EveryONE)은 약 3,000억개의 매개변수 기반 초거대 AI모델 - 엑사원2.0 발표('23.7): 전문가용 대화용 플랫폼(유니버스), 이미지 생성 및 캡셔닝 AI(아틀리에), 멀티모달 AI 기반 산업 도메인 특화 지식 발견 플랫폼(디스커버리) 포함
에이닷(SKT)	- 파라미터 수백억 개 - 계열사 및 자회사와 협력하여 자체생태계 구축(예, 자체 AI생태계 '아이버스'에 탑재 검토) - SKT, Anthropic에 투자(1억 달러)('23.8)

AI클라우드 — AI 도입을 통한 글로벌 빅테크 추격

네이버클라우드	- 초거대 AI(하이퍼클로바)를 적용한 플랫폼 내 AI 기반 서비스를 API 형태로 제공 - 보안 및 데이터 기밀성이 필요 기업/기관 고객을 대상으로 '뉴로클라우드 for 하이퍼클로바X'별도 제공
KT 클라우드	- 자체 개발 AI를 탑재하여 금융·e커머스·헬스케어 등의 분야로 사업 확장 - HW+SW의 풀스택 솔루션을 한 번에 제공하여 글로벌 경쟁력 제고 - 국내 AI반도체 스타트업 '리벨리온'의 아톰을 적용한 NPU 인프라 구축

AI 반도체 — 전통 반도체 제조기업과 AI플랫폼, AI반도체 스타트업과 협력체계 구축

삼성전자 SAMSUNG 삼성전자	- 네이버클라우드, 리벨리온와 협력하여 AI반도체 및 솔루션 개발 중 - 리벨리온 : 삼성전자 파운드리 4나노공정 이용 AI반도체 '리벨' 공동 개발추진('23.10~) - 네이버클라우드 : 하이퍼클로바 구동을 위한 AI반도체 솔루션 개발 추진('22.12.~)
SK 하이닉스 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
사피온 SAPEON	- 국내 최초로 데이터센터용 반도체 X220을 2020년부터 양산해 공급 - LLM 지원 X330 공개, TSMC 7나노 공정을 활용 이르면 연말 상용화('23.11) - 스마트폰 등 엣지디바이스용 X350 내년 상반기 공개, 자율주행용 X340 2026년 상반기 양산 예정, HBM 탑재 차세대 NPU인 X430은 2025년말 공개 목표
퓨리오사AI FURIOSA	- 네이버클라우드와 AI 반도체 솔루션 개발 협력 TF 출범('22.12.) - 퓨리오사AI가 최근 AI 반도체 '워보이(WARBOY)' 개발을 마치고 4월부터 삼성전자 파운드리를 통해 양산 개시 ('23.4) - 네이버클라우드와 협력해 GPU·NPU 팜을 구축, AI 교육 플랫폼 기업 엘리스그룹의 데이터 센터 구축 추진('23.9)
리벨리온 rebellions_	- LLM추론 특화AI 반도체 설계, 아이온('21.11), 아톰 출시(KT 데이터센터에 초도 물량 납품, 현재 대규모 LLM 적용 차세대 AI반도체 리벨 개발중 '24년 출시 목표) - 삼성전자와 LLM 특화 고성능 반도체 '리벨' 개발 협력 ('23.10)

※ 출처 : 소프트웨어정책연구소(2023.11)

참고3 오픈소스 기반 생성 AI 활용 예시와 오픈소스 모델의 특징

◇ 메타의 라마(LLaMA) 모델 공개 후 오픈소스 기반 모델의 활용 확대

- 메타는 OPT-175B('22.5), LLaMA('23.2), LLaMA2('23.7) 등을 오픈소스로 공개하였고, 특히 LLaMA2는 연구뿐 아니라 상업적 용도로도 활용 가능
- 개발자, 연구자를 중심으로 오픈소스 모델의 활용이 확산되는 가운데 LLaMA는 Alpaca, Vicuna, GPT4ALL 등으로 파생되어 특화 서비스 개발에 적용
- 최근 메타의 LLAMA2를 기반으로 음악·오디오 생성 AI '오디오크래프트', 언어 음성번역 '심리스M4T', 코딩 특화 '코드라마(Code LLAMA)'를 공개하며 다양한 파생 모델을 공개 중

◇ 오픈소스 LLM의 활용 사례

사례	기반 모델	내용
업스테이지	LLAMA2	• Upstage/LLAMA-2-70-B-Instruct 모델 • 허깅페이스가 운영하는 오픈LLM 리더보드에서 GPT-3.5 (71.9)를 제치고 72.3점으로 최고 성능 평가('23.8.1)
파인드(Phind)	LLAMA2	• 코딩 특화 AI 모델, HumanEval[26] 평가에서 GPT-4 성능 추월('23.9) • (Phind-Code LLAMA-34B-v2)
스탠포드 알파카	LLAMA-7B	• 600달러와 훈련 데이터 5만2000개로 챗GPT와 비슷한 성능을 구현하는 '알파카 7B'를 개발 ('23.3) • 80GB A100 클라우드 처리 컴퓨터 8대에서 약 3시간 훈련 • '알파카 7B'의 성능에 대해 이메일이나 소셜 미디어의 글 작성, 생산성 도구 등 다양한 분야에서 GPT-3.5와 비교한 결과 알파카는 90개 항목에서, GPT는 89개 항목에서 성능이 상대보다 앞선 것으로 보고
고려대학교	Polyglot-KO	• 오픈소스 한국어 LLM에 양질의 한국어 데이터셋을 추가학습한 구름LLM (KULLM) 깃허브 공개('23.6) • KoAlpaca, KoVicuna와 같은 오픈소스 한국어 LLM에 비해 우수하며 GPT-4 (100점기준)대비 71.1 성능기록
Stability AI	LLaMA2-70B	• 합성 데이터셋을 포함한 소규모 데이터셋으로 미세조정한 FreeWilly2 모델 비상업적 라이센스로 공개('23.7) • 다양한 벤치마크 테스트에서 GPT-3.5 기반의 ChatGPT를 능가하는 성능을 기록

◇ 오픈소스 LLM의 특성

장점	(비용) 개발 진입장벽 개선을 통한 AI 접근성 향상, 개발 속도 향상 및 지속적 개선, 훈련 및 학습 비용 절감, 특정 도메인 빠른 적용 가능, (신뢰성) 공개에 따른 기술의 투명성, 유연성, 가시성 향상, (협업) 지식 공유 커뮤니티 활성화 - "오픈소스 AI는 AI 도구에 민주적으로 접근할 수 있게 하고 공평한 경쟁의 장을 만들어준다. 이는 혁신을 촉진시키는 효과를 가져올 수 있다"(주커버그, '23.9.13[27])
단점	(규제) 개인정보 등 민감 정보 유출 우려, 무분별한 AI 활용 및 악의적 사용으로 인한 사회적 피해, 책임 주체의 불분명, 저작권에 따른 상용 서비스에 적용 제한, (품질) 유지 관리 주체 부재에 따른 품질 저하 우려, 별도 데이터셋 구축 및 파인튜닝 필요, 범용적 적용이 어려움 - "AI가 오픈소스를 통해 잘못된 정보나 유독물질을 퍼뜨릴 수 있다"(빌게이츠, '23.9.13) - "일부 오픈소스는 매우 훌륭하지만 일부는 미래에 우리가 원하는 방향으로 되지 않을 수 있다. 모델에 대한 면밀한 평가가 필요하다"(샘 알트만, '23.9.13)

26 OpenAI HumanEval 데이터세트 벤치마크는 코드 생성 모델 관련 164개 문제로 구성

27 척슈머(Chuck Schumer) 미국 민주당 상원원내대표가 개최한 AI 인사이트 포럼(비공개), 2023.9.13

참고문헌

국외문헌

- Andreeseen Horowitz, Who Owns the Generative AI Platforms, 2023.1.
- iBloomberg, Generative AI to Become a $1.3 Trillion Market by 2032, Research Finds, 2023.6.
- Gartner, Gartner Places Generative AI on the Peak of Inflated Expectations on the 2023 Hype Cycle for Emerging Technologies, 2023.8.16.
- McKinsey Global Institute (2023.6), The economic potential of generative AI: The next productivity frontier
- IDC, GenAI Implementation Market Outlook: Worldwide Core IT Spending for GenAI Forecast, 2023-2027, 2023.10.
- OpenAI, New models and developer products announced at DevDay, 2023.11.6.
- Stastia Market Insights, 2023
- PrecedenceReserach, Generative AI Market (By Component: Software, Services; By Technology: Generative Adversarial Networks (GANs), Transformers, Variational Auto-encoders, Diffusion Networks; By End-Use: Automotive & Transportation, BFSI, Media & Entertainment, IT & Telecommunication, Healthcare, Others) - Global Industry Analysis, Size, Share, Growth, Trends, Regional Outlook, and Forecast 2023-2032, 2023.7.
- The White House, FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence, 2023.10.30.
- Zhao et al.,(2023.9.10.), A Survey of Large Language Models, Arxiv

국내문헌

- 소프트웨어정책연구소(2023.6), 이슈리포트 – 생성 AI의 부상과 산업 변화
- 중소기업벤처부(2023.8), 2023 중소기업 기술보호 수준 실태조사 보고서
- 디지털플랫폼정부위원회, 중소·벤처기업, 공공기관·지자체 대상 초거대 인공지능 활용지원, 2023.07.03.
- 한국소프트웨어산업협회, 한국소프트웨어산업협회, 초거대AI추진협의회 공식 출범, 2023.6.29.
- AI타임즈(2022.10.12.), 구글, 엑사스케일 4세대 TPU AI 시스템 공개
- Digital Today(2023.5.19.), 메타, 전용 AI칩 개발 프로젝트 공개...2025년 실전 투입
- ZDNET(2023.3.31.), 구글, 비전문가도 간단히 생성 AI 다루는 도구 공개
- 조선비즈(2023.6.2.), “1만개 GPU가 하나로 움직인다”… 챗GPT 시대 엔비디아 독주의 비밀
- 서울경제(2023.5.11.), 구글 “오픈AI 플러그인” 저격…생태계 확장 전략 맞불
- 동아일보(2023.6.15.), 국산 AI 반도체 사업에서 과반이 ‘퓨리오사AI’ 선택
- 중앙일보(2023.4.10.), 국산 AI반도체로 세계 클라우드 시장 진출, 또다른 한류

주 의

이 보고서는 소프트웨어정책연구소에서 수행한 연구보고서입니다.
이 보고서의 내용을 발표할 때에는 반드시
소프트웨어정책연구소에서 수행한 연구결과임을 밝혀야 합니다.

생성 AI 산업 생태계 현황과 과제

The current status and challenges of the generative AI ecosystem

경기도 성남시 분당구 대왕판교로 712번길 22 글로벌 R&D 연구동(A) 4층
Global R&D Center 4F 22 Daewangpangyo-ro 712beon-gil, Bundang-gu, Seongnam-si, Gyeonggi-do

www.spri.kr

ISSN 2733-6336