

ISSUE



오픈소스시 개념 및 글로벌 오픈소스 모델 동향

· 권영환 소프트웨어정책연구소 산업정책연구실 책임연구원 | younghwan.kwon@spri.kr

EXECUTIVE SUMMARY

최근 오픈소스(Open Source) 생태계가 인공지능(AI) 기술 혁신과 인공지능 전환(AI Transformation)의 기술적 기반을 제공하면서 오픈소스AI가 크게 주목받고 있다. 주요 딥러닝 프레임워크 28개 중 25개가 오픈소스 기술로 AI 개발 편의성을 향상시키고 모델 대형화를 촉발시켰으며, 트랜스포머(LLM 기반 구조), BERT(사전 학습 패러다임 확립) 등의 오픈소스 모델들은 AI 기술의 빠른 확산을 가능하게 해 오픈소스AI라는 새로운 개념이 등장하였다.

하지만, 오픈소스AI에 대한 명확한 개념이 존재하지 않아 발생하는 혼동을 해결하고자 OSI(Open Source Initiative)는 오픈소스AI 정의(Open Source AI Definition)를 발표하였고, 리눅스 재단은 모델 개방성 프레임워크(Model Openness Framework)를 발표하였다. 이들은 공통적으로 오픈소스AI를 4가지 자유(사용, 연구, 수정, 공유)가 허용된 AI로 정의하고 공개 범위를 데이터, 모델, SW(코드)로 구분하고 있다. 두 개념 사이에 일부 차이점이 있지만, 이들은 오픈소스AI와 오픈소스 모델에 대한 명확한 개념을 제시한다는 점에서 큰 의미가 있다.

최근 오픈소스AI 동향을 보면 깃허브(GitHub)의 AI 관련 프로젝트와 허깅페이스(Hugging Face) 중심의 오픈소스 모델 개발이 빠르게 증가하고 있다. 2024년 기준 깃허브의 AI 프로젝트 수는 432만 개이며, 2025년 12월 기준 허깅페이스의 오픈소스 모델 수는 225만 개를 넘어섰다. 그리고 기업의 89%가 AI 개발 과정에 오픈소스 기술을 활용하고 있으며, 63%가 오픈소스 모델을 활용하고 있는데, 그 이유는 혁신(67%), 시장 표준(67%), 생산성(50%), 개발 비용 절감(49%) 때문이다.

Epoch AI의 유명 AI 모델(Notable AI Models) 분석에서 2018년부터 오픈소스 모델 공개가 증가하여 2018년 이후 유명 AI 모델의 47.3%가 오픈소스 모델이었다. 오픈소스 모델 공개를 주도하는 국가는 미국과 중국이며, 개발 참여 기관 다수는 산업계와 학계였다. 공개되는 모델의 주요 유형은 언어 모델, 비전 모델, 멀티모달이었다. 오픈소스 모델 공개 확산의 이면에는 미·중의 AI 주도권 경쟁이 있으며, 초기 오픈소스AI 생태계가 미국 기업들(메타, 구글 등)이 주도하였다면, 2025년부터 딥시크(DeepSeek), 알리바바 같은 중국 기업들이 우수한 성능과 낮은 비용 기반의 오픈소스 모델을 적극 공개하며 영향력을 빠르게 확산하고 있다. 이 기업들은 오픈소스 모델과 연계한 자사 제품·서비스의 지능화를 통해 고객 만족도와 생산성 향상을 추진하고 추가 수익 창출을 위해 오픈소스 모델 연계 신규 서비스 출시를 전략적으로 추진하고 있다.

이러한 글로벌 오픈소스AI 동향을 보면, AI 3대 강국 도약이라는 국가적 과제 추진을 위해서 오픈소스AI 생태계의 전략적 가치에 대한 재인식이 필요하다. 미국과 중국 기업에게 글로벌 오픈소스AI 생태계는 기술 확보(초기 딥시크의 라마(Llama) 모델 구조 활용, 라마의 MoE(딥시크) 수용) 수단, 우수한 기술의 개방형 검증을 통한 기술 신뢰성 확보 수단, 기술 공개 및 무료 사용 허용을 통한 생태계 저변 확대 수단으로 활용되고 있다. 따라서, AI 3대 강국으로 도약하기 위해서는 전략적 오픈소스AI 활용이 중요해지고 있다. 우선 오픈소스AI는 전략적 기술 원천으로서 가치가 있기 때문에 AI 기술 주권 확보를 위해 선진 오픈소스AI 기반 R&D 추진을 바탕으로 원천 AI 기술 역량 내재화 및 산업 특화 AI 기술 확보가 필요하다. 그리고 범국가 AI 대전환을 위한 전략적 기술 도구로서 가치가 있으므로 오픈소스AI 활용 확산을 위한 기반(생태계) 조성 및 인재 양성이 필요하다.

EXECUTIVE SUMMARY

Recently, the open source ecosystem has been drawing significant attention as it provides the technological foundation for innovation in artificial intelligence(AI) and for AI transformation(AX, AI Transformation). Among 28 major deep learning frameworks, 25 are open source technologies, which have improved the ease of AI development and accelerated the scaling-up of models. In addition, open source models such as the Transformer(the architecture underlying LLMs) and BERT(which established the pre-training paradigm) have enabled the rapid diffusion of AI technology, giving rise to the new concept of Open Source AI.

However, in order to address the confusion caused by the absence of a clear concept of Open Source AI, the Open Source Initiative(OSI) published the Open Source AI Definition, and the Linux Foundation released the Model Openness Framework(MOF). Both frameworks commonly define Open Source AI as AI that permits four freedoms(use, study, modify, and share) and distinguish the scope of disclosure into data, the model, and software(code). Although there are some differences between the two, they are meaningful in that they present clearer concepts of Open Source AI and open models.

Looking at recent Open Source AI trends, AI-related projects on GitHub and open model development centered on Hugging Face are increasing rapidly. As of 2024, the number of AI projects on GitHub reached 4.32 million, and as of December 2025, the number of open source models on Hugging Face surpassed 2.25 million. In addition, 89% of companies use open source technologies in the AI development process, and 63% use open models. The main reasons are innovation(67%), market standards(67%), productivity(50%), and development cost reduction(49%).

According to Epoch AI's analysis of Notable AI Models, releases of open source models have increased since 2018, and 47.3% of notable AI models released after 2018 were open source. The countries leading open model releases are the United States and China, and the majority of participating organizations are from industry and academia. The primary types of released models are language models, vision models, and Multimodal models. Behind the expansion of open model

releases lies the U.S.-China competition for AI leadership. While the early Open Source AI ecosystem was led largely by U.S. companies (such as Meta and Google), since 2025 Chinese companies such as DeepSeek and Alibaba have been rapidly expanding their influence by actively releasing open models that combine strong performance with low cost. These companies are pursuing higher customer satisfaction and productivity by enhancing the intelligence of their products and services in conjunction with open models, and they are strategically driving the launch of new services linked to open models in order to generate additional revenue.

Considering these global trends in open-source AI, there is a need to reassess the strategic value of the open-source AI ecosystem in advancing the national objective of becoming one of the world's top three AI powers. For U.S. and Chinese firms, the global open-source AI ecosystem is being leveraged as a means of securing technological capabilities(e.g., early DeepSeek's use of the LLaMA model architecture and the adoption of LLaMA's Mixture-of-Experts(MoE) approach by DeepSeek), as a mechanism for establishing technological credibility through open and transparent validation of high-quality technologies, and as a tool for broadening the ecosystem base by disclosing technologies and allowing free use. Accordingly, strategic utilization of open-source AI is becoming increasingly important for achieving top-three AI status.

First, because open-source AI holds value as a strategic source of technology, it is necessary to internalize core AI capabilities and secure industry-specific AI technologies by promoting advanced open-source-AI-based R&D to strengthen AI technological sovereignty. Second, because open-source AI also has value as a strategic technological instrument for a nationwide AI transformation, it is essential to build the foundations(ecosystem) for expanding the use of open source AI and to cultivate specialized human resources.

I 배경

○ 최근 오픈소스(Open Source) 생태계가 인공지능(AI) 기술 혁신과 인공지능 전환(AI Transformation)*의 기술적 기반을 제공하면서 오픈소스 생태계의 AI 기술들이 크게 주목받고 있음

* AI를 활용한 새로운 부가가치(생산성, 효율성, 신규 비즈니스 등) 창출 과정이 소프트웨어(SW) 산업을 넘어 다양한 산업 분야로 확산되는 산업 혁신 현상을 의미함¹

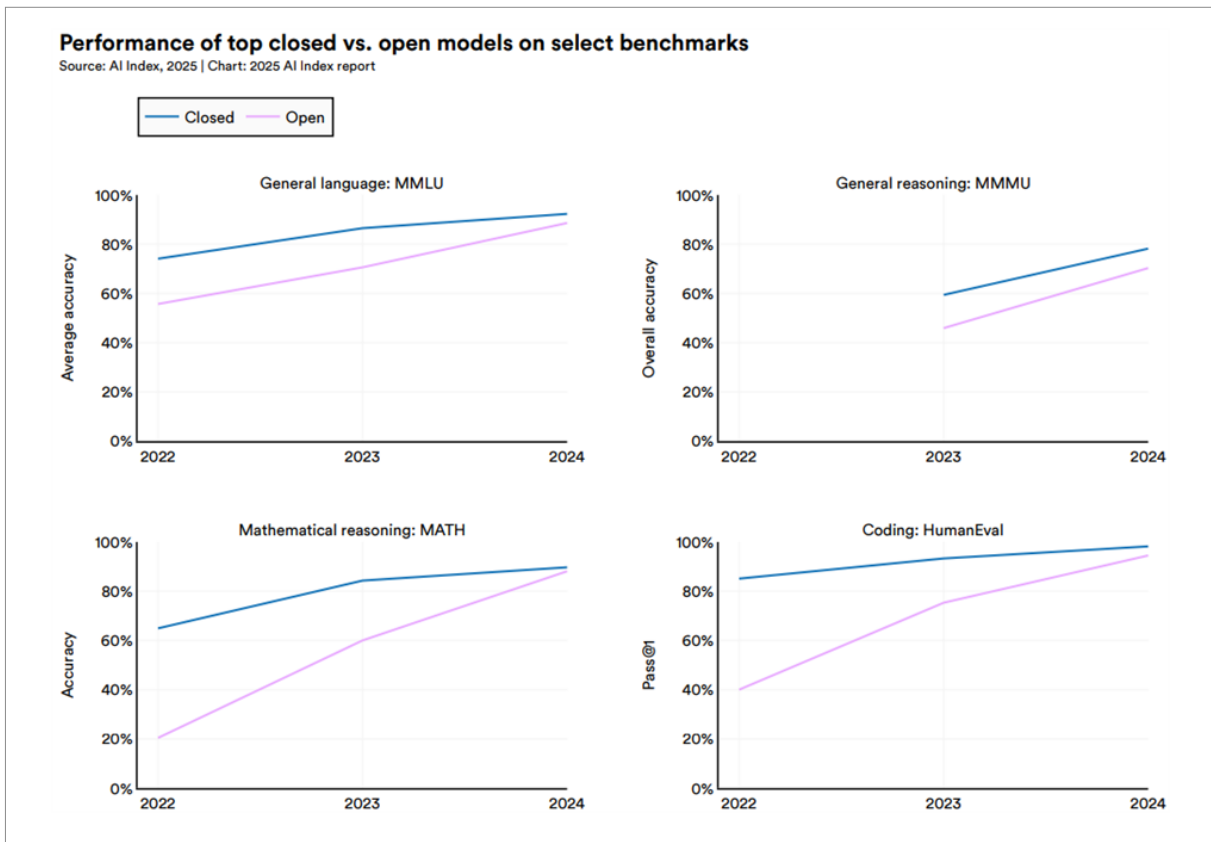
- SW 기술·산업 혁신을 주도하던 오픈소스 생태계에서 딥러닝 프레임워크와 새로운 모델 구조를 공개하면서 AI 기술 혁신과 확산을 주도하고 있음
 - 다양한 오픈소스 기술들(리눅스 커널(Linux Kernel), 자바스크립트(JavaScript), 안드로이드(Android), 쿠버네티스(Kubernetes) 등)이 개방형 혁신을 통해 폐쇄형 SW 기술을 대체하며 시장 지배적 영향력*을 발휘함
 - * 운영체제 분야의 리눅스 커널, 웹 분야의 자바스크립트, 클라우드 분야의 쿠버네티스 등은 각각의 분야에서 독과점적인 영향력을 발휘하는 사실상 시장 표준 기술임
 - 주요 딥러닝 프레임워크의 대부분(28개 중 25개)²은 오픈소스 기술로 AI 기술 접근성을 개선하였고, AlexNet(CNN 구조), 트랜스포머(LLM 기반 구조), BERT(사전 학습 패러다임 확립), GPT-2(트랜스포머 디코더) 등의 혁신적인 모델*들이 공개되면서 AI 기술 혁신이 빠르게 확산됨
- 최근 우수한 성능을 가진 많은 모델이 AI 기술의 핵심(모델 구조, 가중치, 학습 코드 등)을 공개하면서 상용 AI 서비스의 대안으로 주목받음
 - OpenAI의 GPT(Generative Pre-trained Transformer)가 주도하던 생성 AI 시장에서 라마(Llama), 젤마(Gemma), 딥시크(DeepSeek), 큐웬(Qwen) 등과 같이 GPT와 유사한 성능*의 모델들이 공개되고 있음
 - * Chatbot Arena 리더보드 기준 GPT 대비 1.7% 격차가 있었으며, 특히 범용 언어 영역(MMLU), MATH, Coding (HumanEval) 분야에서 매우 근사한 성능을 보임³
 - 핵심 기술 요소를 공개한 모델을 오픈소스 모델이라 하며, AI 제품·서비스 개발 및 운영에 필요한 상용 AI 라이선스 및 서비스 비용을 절감할 수 있는 방안으로 주목받고 있음

¹ LandingAI, <https://landing.ai/case-studies/ai-transformation-playbook>, 2025.06.10. 방문

² Wikipedia, "Comparison of deep learning software", https://en.wikipedia.org/wiki/Comparison_of_deep_learning_software, 2025.12.25. 방문

³ Stanford HAI, "AI Index 2025", 2025.

[그림 1] 폐쇄형 AI 모델과 오픈소스 모델의 기술 격차 사례



출처: Stanford HAI, "AI Index 2025", 2025.

○ 오픈소스AI*의 영향력이 확대되면서 AI 3대 강국 도약을 위한 글로벌 오픈소스AI 현황 분석을 통한 새로운 정책 추진이 중요해지고 있음

* 오픈소스 생태계의 AI 기술들이 늘어나고 확산됨에 따라 이를 줄여서 오픈소스AI라고 함

- 2장에서 오픈소스AI 영향력 확대와 중요성, 개념 정립의 필요성을 제시하고, OSI(Open Source Initiative)의 오픈소스AI 정의(Open Source AI Definition)와 리눅스 재단의 모델 개방성 프레임워크(Model Openness Framework)를 통해 최신 오픈소스AI 개념을 소개함
- 3장에서 최근 주목받는 주요 글로벌 오픈소스AI 동향으로 전 세계 오픈소스 모델 현황 및 미·중 기업의 오픈소스 모델 경쟁 현상을 소개함
- 4장은 결론으로 본고를 요약하고 AI 3대 강국 도약을 위한 오픈소스AI에 대한 정책적 시사점을 제시함

II

주요 오픈소스AI 개념 소개⁴

1. 오픈소스AI의 부상과 중요성

○ 과거 SW 기술·산업 혁신을 선도하던 오픈소스 생태계가 AI 기술 혁신과 확산의 기반이 되었음

- 오픈소스 생태계에서 개발된 다양한 SW 기술들은 SW 생태계 외연 확장의 동력으로 작용하며 기술·산업 혁신을 선도하였음
 - 1980년대 태동한 초기 오픈소스 생태계는 SW 호환성 향상 및 협업 활성화를 위한 개발자 중심 생태계로 출발하였지만, 이후 운영체제, DB, 웹 분야에서 상용 SW의 대안*으로 주목받으며 인터넷 산업 활성화에 크게 기여하였음
 - * 전 세계 서버 시장(450억 달러 규모)에서 리눅스와 오픈소스 웹서버 비중은 모두 70% 이상⁵이며, DB 분야의 오픈소스 DB 비중은 약 50%⁶임
 - 2010년대 전후부터 빅데이터, 클라우드, AI 기술 혁신을 주도하고 산업 맞춤형 플랫폼 개발 협업이 활성화되면서 플랫폼 산업 확산 및 산업 디지털 전환의 기반을 제공하였음
- 2020년대 들어 오픈소스 생태계는 깃허브와 허깅페이스를 중심으로 AI 분야의 오픈소스 개발 협업이 빠르게 증가하며 기술 혁신 및 확산의 중심이 되었음
 - 2025년 깃허브* 개발자 수는 1억 8천만 명으로 증가하였고 그 중심에는 430만 개 이상의 AI 프로젝트들**이 있으며, 누적 스타 수가 1,764만 개 이상으로 주목받음^{7,8}
 - * 2025년 깃허브는 3,600만 명의 개발자 증가, 4,320만 건의 풀리퀘스트(기여, +23% 증가), 약 10억 건의 커밋(코드 변경 승인, +25.1% 증가), 6.3억 개의 프로젝트 수로 증가함
 - ** 주요 프로젝트로 vLLM(고처리량 LLM 추론 엔진), Ollama(로컬 모델 실행 및 관리 도구), Transformers(허깅페이스의 모델 구조), RAGFlow(Retrieval-Augmented Generation(RAG) 엔진) 등이 있음

⁴ OSI(Open Source Initiative)에서 오픈소스AI 정의(OSAID, Open Source AI Definition)를 규정함에 따라 OSAID와 다른 정의를 포괄하고 용어 혼동을 피하기 위해 정의와 다른 용어를 사용함

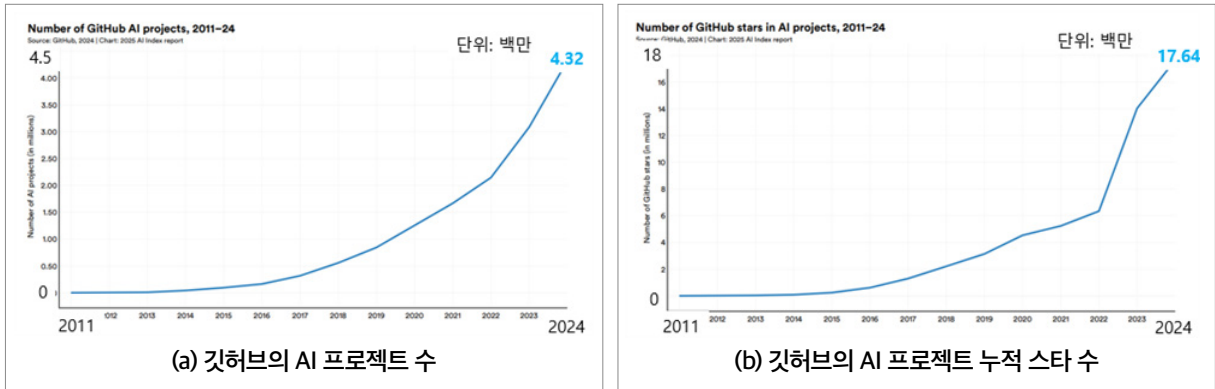
⁵ Gitnux, "Report 2025 - Server Statistics", 2025.04.29.

⁶ DB Engines, "Popularity of open source DBMS versus commercial DBMS", 2025.11.20.

⁷ GitHub, "Octoverse 2025", 2025.10.28.

⁸ Stanford HAI, "AI Index 2025", 2025.

[그림 2] 깃허브의 AI 프로젝트 수와 관련 프로젝트들의 누적 스타 수의 증가(2011-2024)



출처: Stanford HAI, "AI Index 2025", 2025.

- 2025년 허깅페이스* 개발자 수는 500만 명 이상으로 약 224만 개의 공개 모델과 56만 개의 공개 데이터셋을 기반으로 AI 모델 개발 협업이 가장 활발히 이루어지고 있는 세계 최대 오픈소스 머신러닝 개발 플랫폼임⁹

* 2016년 설립된 AI 스타트업으로 깃허브에서 개발되는 Transformers를 기반으로 최신 AI 모델의 효율적 개발, 데이터셋 공유, 강력한 엔비디아 추론 기능을 제공하고 있으며, 라마, 딥시크, 큐웬 등의 유명 오픈소스 모델이 공개·확산되는 플랫폼임

[그림 3] 기술 관점의 오픈소스 생태계의 확장

시대	1980년대 (태동)	2000년 전후 (성장)	2010년 전후 (고도화)	2020년 전후 (확산)	2025년 현재 (확장)
SW 규모	시스템SW의 구성요소	시스템SW	플랫폼SW	산업 맞춤형 플랫폼	AI = SW + 데이터
주요 오픈소스	GNU 프로젝트 (X.11 - 윈도우 환경, Gdb - 디버거, Gcc - 컴파일러, Bash - SW 제어 환경, 등)	리눅스 커널 (운영체제, 1991), MySQL (DB, 1995), 아파치 HTTP 서버 (웹 서버, 1995), 모질라 (브라우저, 1998)	아파치 하둡 (빅데이터, 2006), 안드로이드 (모바일, 2008), 오픈스택 (클라우드, 2010), Automotive Grade Linux (커넥티드카, 2012), 텐서플로우 (인공지능, 2015)	Fintech Open Source (2017, 금융), Academy Software (2018, 미디어), LF Networking (2018, 통신/5G), LF Energy (2018, 에너지), OS-Climate (2020, 기후 온난화), Green Software (2021, 친환경 SW)	메타 (라마), 미스트랄AI, 딥시크, 알리바바 (큐웬), 바이두 (여니), OpenAI (gpt-oss)
목적/역할	SW 개발 호환성 제공	상용SW 대안	SW/IT 신기술 개발 협력	산업 디지털 혁신 및 지구 문제 대응	AI 혁신
산업 분야	SW 개발자	SW 산업	ICT 산업	전 산업	신산업(AI)
IT 생태계 파급 효과	SW개발 저변 확대 (C, 윈도우 환경, 등)	인터넷 산업 활성화	플랫폼 비즈니스 확산	산업 디지털 전환	AI 전환
핵심 개념	자유 소프트웨어	오픈소스(SW)	엔터프라이즈 오픈소스	...	오픈소스AI

출처: SPRi 작성

⁹ Hugging Face, <https://huggingface.co/>, 2025.11.21. 방문

○ 오픈소스 생태계가 AI 기술 혁신과 확산을 촉발시킨 주요 계기는 오픈소스AI 프레임워크와 허깅페이스의 등장임

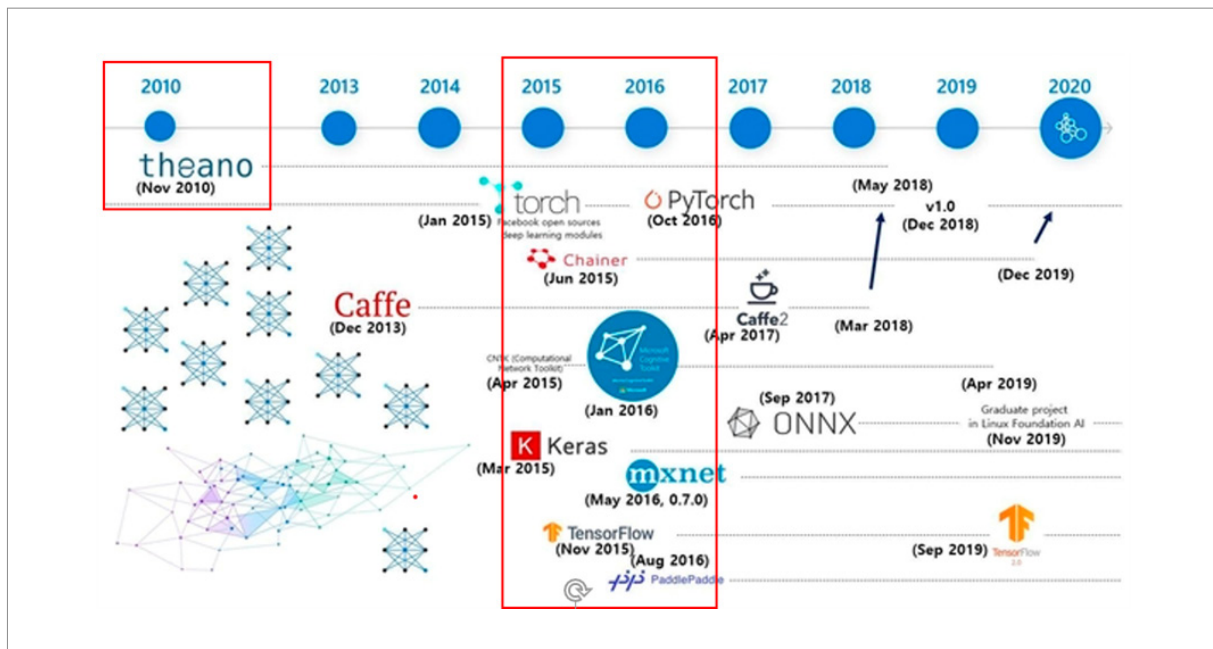
- 오픈소스 생태계가 AI 기술 혁신의 중심이 된 배경에는 2010년대에 활발히 개발된 오픈소스AI 프레임워크*들이 있으며, 이들은 AI 개발 환경을 표준화하고 편의성을 향상시키며 AI 기술 혁신(딥러닝 전환, 모델 대형화)과 확산의 토대가 되었음

* AI 개발을 위한 기본 인프라로 표준화된 필수 기능(자동 미분, 텐서 계산, GPU/TPU 연산 등)과 대규모 학습 환경(분산 연산 및 병렬화, GPU/TPU 클러스터 관리 등)을 API 형태로 제공하여 AI 개발 난이도를 완화하며 AI 확산의 원동력이 됨

- 테아노(2010년)는 최초의 오픈소스 딥러닝 프레임워크로 딥러닝 환경 구축 편의성을 향상시켰으며, 텐서플로우(2015년)와 파이토치(2016년)는 대규모 학습 기능의 대중화·표준화를 통해 AI 기술 확산과 혁신을 촉발*함

* 최근 몇 년간 8개 연구 저널에서 수집한 데이터 분석 결과, 텐서플로우와 파이토치를 사용하는 논문 비중은 80% 수준임¹⁰

[그림 4] AI 확산 및 혁신을 촉발한 오픈소스AI 프레임워크 기술 변화

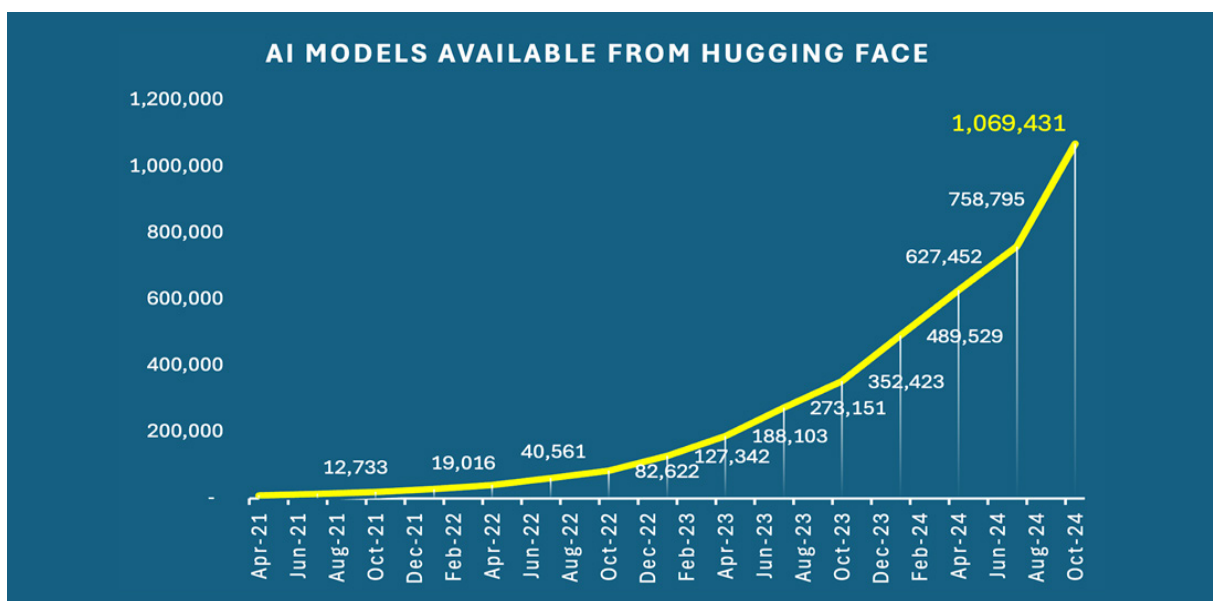


출처: IT조선, “누구나 개발자 #1” ②오픈소스AI 개발 도구가 애저클라우드와 만났을 때”, 자료 SPRI 보완

¹⁰ Ryan O'Connor, “PyTorch vs TensorFlow in 2023”, Speech&Text, 2021.12.14.

- Epoch AI 보고서는 AI 학습량을 기반으로 2010년 이전을 딥러닝 이전 시대(2배/20개월 증가), 2010년 이후를 딥러닝 시대(2배/6개월 증가), 2015년/2016년부터 대규모 시대(딥러닝 시대보다 2~3배 증가)로 구분함¹¹
- 허깅페이스는 2016년 설립되어 머신러닝 모델 개발·배포를 위한 라이브러리(Transformers 등)를 오픈 소스로 개발하여 서비스를 제공하는 AI 스타트업임
 - 깃허브에서 개발된 허깅페이스의 트랜스포머스(Transformers)는 급성장하는 오픈소스 프로젝트(스타 수 14만 개 이상)로 크게 주목받았으며, 이 증가 속도는 AI 분야 타 프로젝트(텐서플로우, 파이토치 등) 대비 매우 빠른 속도임¹²
 - 허깅페이스 플랫폼에서 제공되는 AI 개발 환경(트랜스포머스, 텐서플로우/파이토치 연계, 엔비디아 칩 가속 기능 등)으로 AI 개발 편의성이 다시 향상되며, 허깅페이스에서 개발되는 모델 수가 2021년 1만 개 이하에서 2024년 100만 개 이상으로 급증함¹³

[그림 5] 허깅페이스의 AI 모델 수의 증가



출처: Security Boulevard, "Hugging Face Has Become a Malware Magnet", 2024.10.23.

¹¹ Epoch AI, "Compute trends across three eras of machine learning", 2022.05.02.

¹² Security Boulevard, "Hugging Face Has Become a Malware Magnet", 2024.10.23.

¹³ Star History, <https://star-history.com>, 2025.11.25. 방문

○ 특히, 오픈소스는 중소기업의 AI 개발에 더 큰 영향력을 발휘하고 있으며, 그 이유는 비용 절감, 생산성 향상 같은 기업 경쟁력과 직결되기 때문임¹⁴

- 기업의 AI 개발 과정에서 오픈소스 활용률은 89%일 정도로 오픈소스가 크게 영향력을 발휘하며, 특히 중소기업의 오픈소스AI 활용률은 78%로 대기업보다 11% 더 높음
 - 89%의 기업들이 AI 개발에 오픈소스를 활용하고 있으며, 63%가 오픈소스 모델을 현재 활용 중이고 76%는 오픈소스AI 활용을 확대할 것으로 응답하였음
 - 종사자 수가 10명에서 249명인 중소기업의 오픈소스AI 활용률은 78%로 가장 높았으며, 상대적으로 1만 명 이상인 대기업의 활용률은 67%로 가장 낮았음
- 기업들이 오픈소스를 활용하는 핵심 이유는 최신 기술 도입, 기술 저변(호환성) 확보, 효율성이며, 특히 80% 이상의 응답자들이 생산성 향상 및 비용 절감 효과가 크다고 응답함
 - AI 분야에서 오픈소스를 활용하는 주요 이유는 혁신(67%)과 시장 표준 및 호환성(67%)이 가장 높은 응답률을 보였고, 이어서 생산성(50%), 제품 개발 비용 절감(49%)으로 나타났음
 - 그리고, 상세 질문에서 개별 장점들의 동의 여부에 관한 질문에는 생산성 향상(83%), SW 소유 비용 절감(82%), SW 품질 향상(82%), 타기업 종속 회피(79%), 기업 업무 환경 개선(77%), 혁신 가속화(77%) 순으로 나타났음

○ 세계적으로 국가 안보 및 경쟁력 강화를 위한 소버린 AI(Sovereign AI)의 필요성에 공감하는 비율이 79%이며, 소버린 AI 확보 방안으로 오픈소스 협력이 필요하다는 응답이 94%였음¹⁵

- 지역적으로 소버린 AI의 필요성은 미국(86%), 유럽(83%), 아시아-태평양(79%) 순으로, 이를 위해 82%의 조직에서 자체적인 AI 솔루션을 개발하고 있다고 응답함
 - 소버린 AI가 필요한 이유는 데이터 유출 방지 및 사생활 보호를 위한 데이터 제어(72%)가 가장 높았으며, 이어서 국가 간 분쟁으로 인한 전략 자산(AI) 위험성 예방을 위한 국가 보안(69%), AI가 국가 산업 경쟁력과 직결됨으로 인한 경제적 경쟁력 확보(48%) 순으로 응답이 많았음
 - 자체 오픈소스AI 개발은 미국(90%), 유럽(86%)에 비해 아시아-태평양(72%)이 적었으며, 조직 규모 면에서는 종사자 수가 1천 명 이상인 대기업(92%)에서 높았으며, 50~1천 명(71%), 1~49명(62%) 순으로 낮아짐

¹⁴ Linux Foundation & Meta, "The Economic and Workforce Impact of Open Source AI", 2025.05.

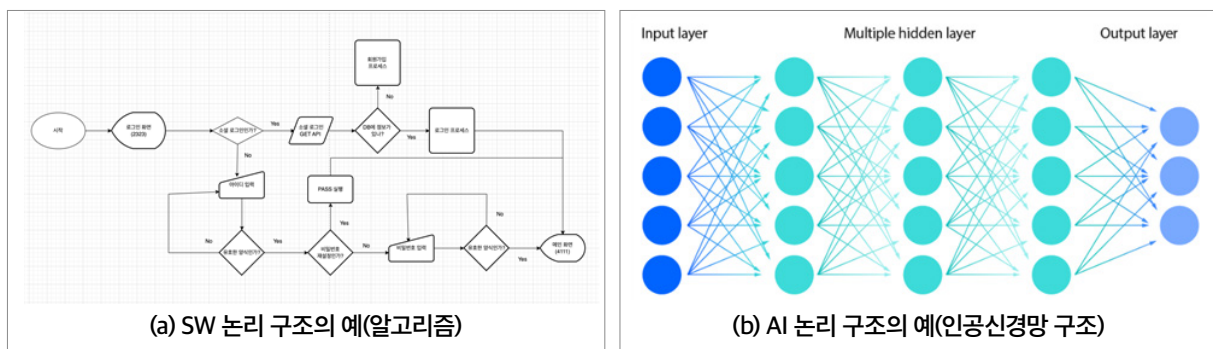
¹⁵ Linux Foundation, "The State of Sovereign AI", 2025.08.

- 소버린 AI 확보 방안으로 많은 응답자(94%)가 오픈소스 협업이 필요하다고 응답하였으며, 특히 조직 규모가 작을수록 오픈소스 협업이 중요하다고 인식함
 - 독자적인 소버린 AI 확보 수단으로 오픈소스 협업이 필요하다고 응답한 이유는 기술의 핵심인 소스코드, 웨이트, 데이터가 공개되어 있어서 모델 웨이트·구조 접근성(84%), 코드 검사·수정(79%), 학습 방법 투명성(76%), 특정 기업 종속성 회피(69%), 파인튜닝 가능(69%) 때문임
 - 특히 종사자 수가 작은 조직일수록 오픈소스 협업이 중요하다고 인식하였는데, 그 이유는 자원(인적, 자금 등) 부족으로 독자적인 AI 역량(인프라, 데이터, 인력, 모델 등) 확보가 어렵기 때문임
- 소버린 AI 확보를 위한 글로벌 오픈소스 협업 대상의 핵심은 파운데이션 모델(59%)과 데이터(59%)이며, 핵심 협업 방안으로는 오픈소스 프로젝트 기여(59%), AI 표준(45%) 제정, 책임 있는 AI(Responsible AI) 원칙 수립이었음
 - 가장 가치 있는 글로벌 협업 대상은 베이스 및 파운데이션 모델(59%)과 데이터 자원 및 데이터셋(59%)에 이어 도구 및 플랫폼 개발(39%), HW와 연산 인프라(38%) 순서였음
 - 그리고, 응답자들이 참여하고 싶은 글로벌 협업의 형태는 오픈소스AI 프로젝트 및 도구 기여(59%), AI 시스템의 기술 표준 제정(45%), 책임 있는 AI 원칙 및 사례 협력(45%), 공통 평가 벤치마크 및 지표 개발(40%) 순서였음

2. 오픈소스AI 개념 정립의 필요성

- 오픈소스 생태계의 AI 기술의 영향력이 확대되면서 오픈소스AI라는 새로운 용어가 등장하였지만, SW와 AI의 기술적 차이 때문에 오픈소스AI에 대한 명확한 기준이 필요해짐

[그림 6] SW와 AI의 논리 구조 차이



출처: UIUX Information

출처: crone_hi.log

- AI는 SW의 일부이지만, 기술적으로 AI는 데이터를 분석하여 도출한 통계적 계층의 연속인 인공지능망 구조를 기반으로 하므로, 사람의 논리를 구현하는 알고리즘으로 구성된 SW와 기술적 차이가 존재
 - SW는 사람이 현상을 분석하여 도출한 논리(규칙, 절차 등)를 구현한 알고리즘을 기반으로 함
 - AI는 사람이 구현한 통계적 방법론 알고리즘이 현실 데이터를 학습·분석하고 통계 및 확률적 규칙과 패턴을 추출하여 이를 인공지능망 형태로 구조화한 논리 구조임
- 기존 오픈소스 정의는 SW(알고리즘)를 기반으로 정의되어 있어 AI의 새로운 기술적 요소(학습 데이터, 인공지능망 구조의 가중치(웨이트) 등)를 포괄하지 못하기 때문에 오픈소스의 재현성을 확보하기 위한 새로운 개념 정의가 필요함
 - OSI*에서 2000년에 정의한 OSD(Open Source Definition)는 알고리즘의 구현물인 소스코드를 중심으로 정의되어 10가지 기준(① 무료 재배포, ② 소스코드, ③ 파생 저작물, ④ 저자 소스코드의 무결성, ⑤ 개인 또는 단체 차별 금지, ⑥ 활용 분야 차별 금지, ⑦ 라이선스 배포, ⑧ 특정 제품 제한 금지, ⑨ 다른 소프트웨어 차별 금지, ⑩ 기술 중립)을 제시함¹⁶
 - * 2000년 전후에 자유소프트웨어(Free Software) 정신을 기반으로 오픈소스 개념을 공식적으로 정의하여 오픈소스 생태계 기반을 마련한 비영리 단체¹⁷
 - AI는 SW와 달리 새로운 기술적 요소(데이터, 사람이 도출하지 않은 논리 구조-모델 구조와 가중치 등)가 요구되는데, 이 새로운 요소들은 기존 오픈소스 정의에 포함되지 않기 때문에 오픈소스AI에 대한 자의적 해석으로 인한 혼동*이 발생함
 - * 메타의 라마, 트위터의 Grok, MS의 Phi-2 등은 데이터, 소스코드, 모델을 모두 공개하지 않고 일부만 공개하여 투명성 및 재현성에 제약이 있음에도 오픈소스로 홍보되면서 일부에서 해당 모델들을 오픈소스와 구분하기 위해 오픈워싱(Open washing)이라고 함
 - 따라서, 오픈소스AI의 구성요소와 공개 조건을 명확히 함으로써 오픈소스AI의 자율성, 투명성, 재현성, 재사용, 협력적 개선을 보장하고, 오픈소스AI에 대한 명확한 기준을 제시하기 위해 정확한 개념 정립이 필요해짐

3. OSI의 오픈소스AI 정의(Open Source AI Definition)¹⁸

○ OSI는 약 2년간의 논의를 거쳐 AI 시스템 특성을 고려한 새로운 오픈소스AI 정의(OSAID, Open Source AI Definition) v1.0을 2024년 10월에 발표함

¹⁶ Open Source Initiative, "Open Source Definition", <https://opensource.org/osd>, 2025.06.17. 방문

¹⁷ Open Source Initiative, "About the Open Source Initiative", <https://opensource.org/about>, 2025.06.17. 방문

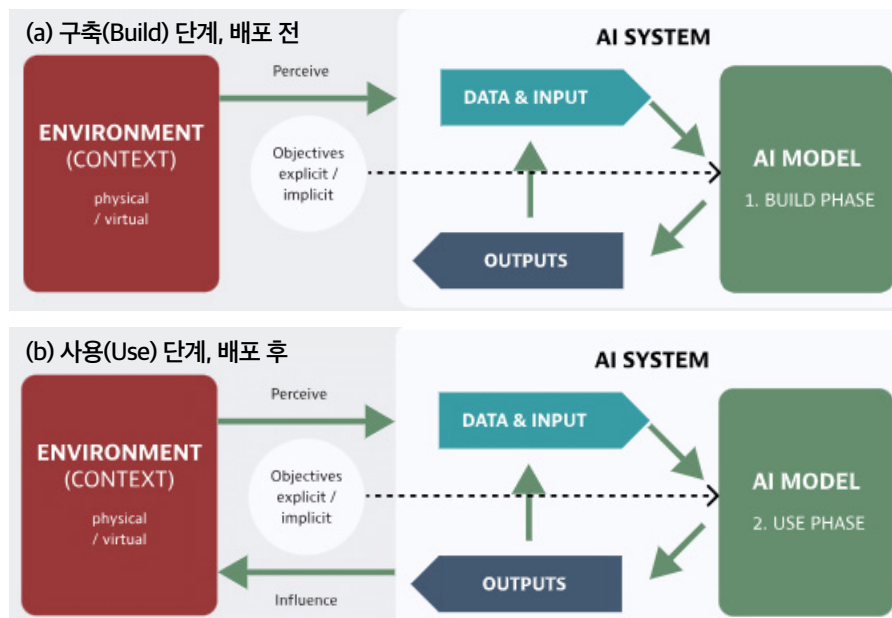
¹⁸ Open Source Initiative, "OSAID(Open Source AI Definition) v1.0", 2024.10.28. 발표

- 오픈소스AI 정의의 AI 시스템은 OECD(Organization for Economic Co-operation and Development)의 AI 원칙(Principles)*에서 활용된 개념을 기반으로 완전한 기능 구조와 그 구조 내부의 개별 요소 모두를 포괄함
 - * 2019년 OECD에서 신뢰성 있는 AI 기준(원칙)을 제시하기 위해 발표된 국제 권고안으로, 전 세계 70여 개 국가에서 AI 정책 수립을 위해 활용되고 있음
- OECD AI 원칙의 AI 시스템은 “명시적 또는 암묵적 목적을 위해 입력된 내용을 바탕으로 예측, 콘텐츠, 추천, 또는 물리적 또는 가상 환경에 영향을 미칠 수 있는 의사결정과 같은 출력을 생성하는 방법을 추론하는 기계 기반 시스템”으로 정의됨¹⁹
- OSI에서 정의된 오픈소스AI는 4가지 자유(사용, 연구, 수정, 공유)*가 보장된 AI 시스템으로 구체적 사례로 머신러닝 시스템 수정을 위한 구성요소를 제시함
 - * 이는 자유SW의 4가지 자유(실행, 연구/변경, 복제/배포, 환원)²⁰과 매우 유사함

오픈소스AI 정의 v1.0의 4가지 요구사항(자유)

- Use the system for any purpose and without having to ask for permission.
- Study how the system works and inspect its components.
- Modify the system for any purpose, including to change its output.
- Share the system for others to use with or without modifications, for any purpose.

[그림 7] OECD의 AI 시스템 개념



출처: OECD, "OECD AI Principles Overview", <https://oecd.ai/en/ai-principles>

¹⁹ OECD, "OECD AI Principles Overview", <https://oecd.ai/en/ai-principles>, 2025.06.17. 방문

²⁰ GNU 운영체제, "자유 소프트웨어란 무엇인가?", <https://www.gnu.org/philosophy/free-sw.ko.html>, 2025.11.20. 방문

- 즉, 오픈소스AI는 목적(상용화, 연구 등)과 대상에 대한 ① 차별과 허가 없이 사용하고, ② 작동 및 구성 요소 검사 방법을 연구하며, 출력 변경을 포함한 ③ 시스템을 수정하고, 수정 여부와 관계없이 ④ 공유할 수 있는 AI 시스템을 의미함
- 오픈소스AI 정의의 적용 사례로 머신러닝 시스템을 수정하기 위한 선호 형태*로 3가지 구성요소(데이터 정보, 코드, 매개변수)를 아래 표와 같이 구체적으로 제시함

* OSAID v1.0에서는 수정을 위한 구성요소만 제시하지만, 실질적으로 해당 구성요소들은 사용, 연구, 공유에도 적용됨

[표 1] 머신러닝 시스템 수정을 위해 선호되는 형태

구성요소	개요	공개되는 정보의 예	제공 조건
데이터 정보	동등한 시스템 구축을 위해 시스템 학습에 사용된 데이터에 대한 충분한 상세 정보	(1) 학습 데이터에 대한 완전한 설명(공유할 수 없는 데이터 포함), 데이터 출처, 범위 및 특성, 수집·선택 방법, 라벨링 절차, 데이터 처리·필터링 방법 (2) 공개적으로 사용 가능한 모든 학습 데이터 목록 및 획득 장소 (3) 제3자로부터 얻을 수 있는 모든 학습 데이터 목록 및 유료를 포함한 획득 장소	OSI 승인 조건*
코드	학습 및 실행에 사용되는 완전한 소스코드로 데이터 처리, 필터링 방식, 학습 방식에 대한 전체 명세서	사용된 데이터 처리·필터링 코드, 인수·설정을 포함한 학습 코드, 검증·테스트, 토큰라이저 및 하이퍼파라미터 검색 코드와 같은 지원 라이브러리, 추론 코드 및 모델 아키텍처	OSI 승인 라이선스
매개변수	가중치 또는 기타 구성 설정과 같은 모델 매개변수	학습의 주요 중간 단계의 점검 포인트와 최종 최적화 상태	OSI 승인 조건*

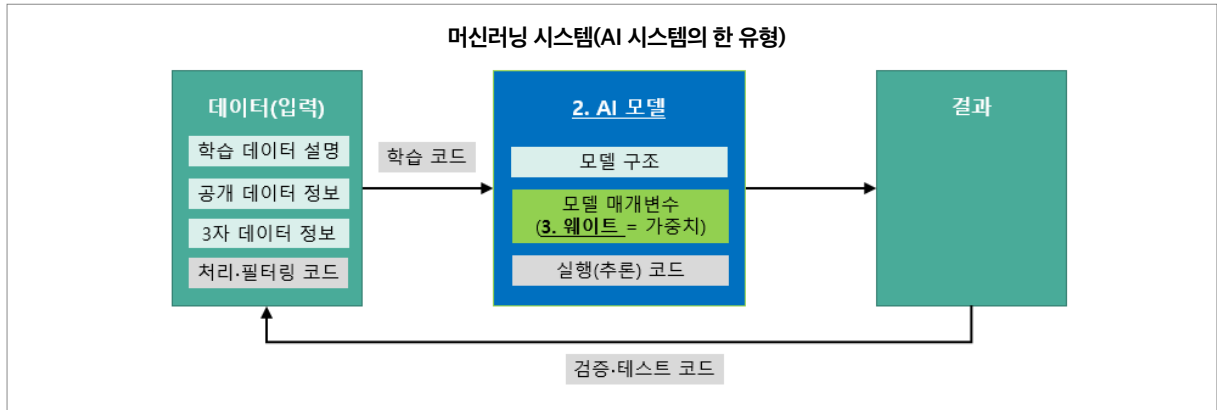
출처: Open Source Initiative, Open Source AI Definition v1.0(2024) 내용 SPRi 번역

* 데이터와 매개변수에는 OSI 승인 라이선스를 적용하기 어렵기 때문에 법적으로 모호한 조건(Term)이라는 용어를 활용함

○ 오픈소스AI를 AI 시스템 구성요소의 기술적 범주에 따라 3가지 유형(① AI 시스템, ② AI 모델, ③ AI 웨이트)의 AI로 구분함

- 오픈소스AI 시스템은 가장 포괄적인 범주로 AI 시스템 전체에 적용되며, AI 시스템의 하위 구성요소에 따라 독립적 동작이 가능한 AI 모델, AI 웨이트(가중치)를 별도 규정함
 - AI 모델은 모델 구조, 모델 매개변수(가중치 포함), 모델 실행을 위한 추론 코드로 구성된다고 정의됨
 - AI 웨이트(가중치)는 주어진 입력으로부터 출력을 생성하기 위한 모델 구조와 연계된 학습된 매개변수 집합을 의미함

[그림 8] 오픈소스AI 정의의 3가지 유형의 AI(① AI 시스템, ② AI 모델, ③ AI 웨이트)



출처: SPRi 작성

- AI 모델에 대응하는 오픈소스 모델과 AI 가중치에 대응하는 오픈소스 웨이트를 별도로 정의하며, 해당 구성요소들의 공개와 함께 추가 공개 조건*을 제시함

* 해당 매개변수(가중치)를 도출하는 데 사용된 데이터 정보와 코드(학습)를 의미함

- 따라서, 오픈소스 모델은 모델 구조, 매개변수(가중치), 실행(추론) 코드와 더불어 매개변수 생성에 활용된 학습 데이터와 학습 코드가 같이 공개되어 있어야 함
- 오픈소스 웨이트*는 매개변수와 매개변수 실행에 필요한 모델 구조 정보(모델 구조와 연계된 매개변수)와 함께 매개변수 생성에 활용된 학습 데이터와 학습 코드가 같이 공개되어야 함

* 현실적으로 통용되는 오픈웨이트 개념과의 차이점(학습 데이터 및 코드 미공개)과 한계점(재현성 부족, 데이터 불투명성, 규제 장벽, 제한된 협력)에 대한 정보를 2025년에 추가²¹하며 구분하고 있음

[표 2] OSI의 오픈웨이트와 오픈소스AI 비교

특징	오픈웨이트	오픈소스AI
가중치 및 편향	공개	공개
훈련 코드	미공유	전체 공유
중간 점검 포인트	보류	공개 권고
훈련 데이터셋	미공유/미공개	공개
훈련 데이터 구성(정보)	부분/미공개	전체 공개

출처: Open Source Initiative, "What are Open Weights?" 내용 SPRi 번역

²¹ Open Source Initiative, "What are Open Weights?", <https://opensource.org/ai/open-weights>, 2025.12.26. 방문

○ 오픈소스AI 정의는 오픈소스AI의 가치를 재인식시키고 정확한 기준을 처음 제시하였지만, 엄격한 기준과 개념적 정의로서의 모호성에 대한 비판들이 제기됨

- 24개 단체/기업과 100명 이상의 개인이 AI의 개방형 혁신을 위한 공식적 논의의 첫걸음으로 오픈소스AI 정의(v1.0)의 중요성과 가치를 공개 지지함²²
 - 주요 지지 선언 단체로 모질라 재단, 이클립스 재단, 수세, 오픈인프라 재단, code.gouv.fr(프랑스 오픈소스 공유 허브), OSG-JP(Open Source Group Japan), KAIYUAN(중국 오픈소스 협회) 등과 같은 비영리 단체*가 대부분임
 - * 미국은 대학(프린스턴, 조지워싱턴, 조지아공대, 카네기 멜론 등)을 중심으로, 한국은 국가 연구기관(ETRI, AI안전연구소)을 중심으로 개인적 지지 선언이 있었음
 - AI의 개방형 혁신(투명성/안전, 경쟁 촉발, 활용 확산)을 가속화하기 위한 기술적 구성요소(데이터, 모델, 코드, 웨이트 등) 도출 및 기준(4가지 자유)에 대해 다수가 지지하고 있음
- 하지만, 기존 오픈소스 모델들의 오픈소스AI 정의 미충족과 개념적 정의의 한계성(법적 기준 모호, 현실성 부족 등)에 대한 비판들이 제기됨
 - 기존 오픈소스 모델로 알려진 것 중에 일부만 OSI의 오픈소스AI 정의를 충족하고 있으며, 글로벌 AI 선도기업이 공개한 모델 중 다수*는 조건을 충족하지 못하고 있음²³
 - * 메타의 라마, 구글의 젤마, X의 Grok, 미스트랄AI의 미스트랄, 딥시크 등은 데이터 정보, 관련 모든 코드 등 중 일부만 공개하거나, 상업적 활용 금지 또는 제한적 사용을 허가하고 있기 때문에 오픈소스 모델이 아닌 오픈 웨이트 모델로 구분되고 있음
 - 이외에도 학습 데이터의 미공개 허용(개인정보, 보안 등), 데이터 공개의 법적 기준 모호성(라이선스라는 용어 미사용), 현실성 부족(기존 오픈웨이트 모델과 불일치), 보안 및 악용 증가 등 일부 이슈들에 대한 향후 OSI 대응*이 중요해지고 있음
 - * 1998년 정의된 오픈소스 정의도 용어(Open vs Free), 카피레프트 vs 퍼미시브(Copyleft vs Permissive) 논란, 특허 보복 조항 등의 다양한 논란 속에서 오픈소스 정의의 일부 문구가 개정되면서 정립되었으며, 이미 FAQ 및 Open Weights 정보를 추가로 제공하며 대응 중

²² Open Source Initiative, "The Open Source AI Definition 1.0 Endorser organizations", <https://opensource.org/ai/endorsements>, 2025.06.17. 방문

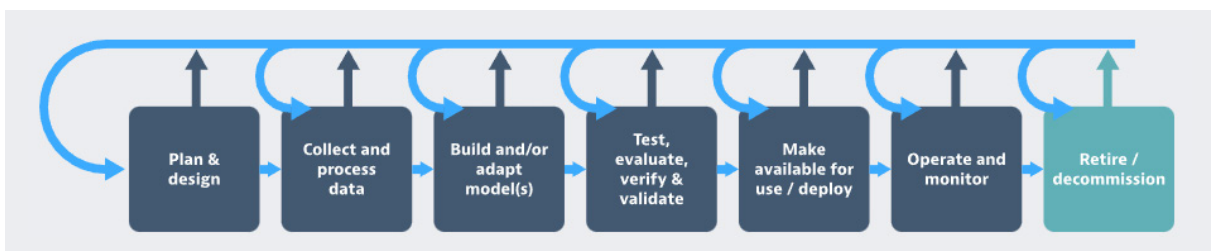
²³ Open Source Initiative, "The Open Source AI Definition 1.0 FAQ", <https://opensource.org/ai/faq#why-do-you-require-training-code-while-osd-2-doesn-t-require-compilers>, 2025.06.17. 방문

4. 리눅스 재단의 모델 개방성 프레임워크(Model Openness Framework)²⁴

○ 리눅스 재단은 2024년 12월에 머신러닝 모델의 개방성과 완전성을 평가하기 위한 분류 체계인 모델 개방성 프레임워크를 발표함

- 모델 개방성 프레임워크는 개념 중심의 OSI의 오픈소스AI 정의와 달리 모델 개발 수명 주기 관점에서 모델 구성요소들과 적용된 라이선스를 기반으로 개방성을 실질적으로 평가하기 위한 프레임워크임
 - AI 모델은 반복적인 데이터 학습을 통해 기계(컴퓨터)가 모델 내부의 최적 웨이트(통계/확률값)를 도출하는 방식으로 개발되기 때문에 OECD 아래 그림과 같이 AI 시스템 생명주기를 정의하고 있음²⁵

[그림 9] OECD의 AI 시스템 수명 주기(Lifecycle)



출처: OECD, "AI System lifecycle"

- 모델 개방성 평가를 위해 3가지 형태(코드, 데이터, 문서)로 구분되는 총 17개의 구성요소를 도출하였으며, 형태에 따라 적용되는 개방형 라이선스*들을 구체적으로 규정함
 - * (소스)코드 형태로 공개되는 6개 구성요소들은 오픈소스 라이선스가 적용되고, 데이터 형태로 공개되는 6개 구성요소들은 오픈데이터 라이선스가 적용되며, 문서 형태로 공개되는 5개 구성요소들은 오픈 콘텐츠 라이선스가 적용됨
- 17개 구성요소들의 개방 정도에 따라 3단계(클래스 1: 오픈 사이언스 모델, 클래스 2: 오픈 툴링 모델, 클래스 3: 오픈 모델)로 모델 개방성을 평가함
 - 개별 구성요소의 개방성은 적용된 개방형 라이선스(OSI 오픈소스AI 정의 v1.0의 4가지 자유 - 사용, 연구, 수정, 공유가 적용된)로 판단 가능

²⁴ Linux Foundation, "The Model Openness Framework Whitepaper", 2024.12.17.

²⁵ OECD, "AI System lifecycle", 2025.11.20. 방문

- 모델 개방성을 3단계(클래스 1: 오픈 사이언스 모델, 클래스 2: 오픈 툴링 모델, 클래스 3: 오픈 모델)로 구분하며, 수준별 모델 활용 가능 범위*를 구체적으로 제시함

* 선언적 의미를 가지는 OSI의 오픈소스AI 정의와 달리 오픈소스 모델 사용자를 위한 활용 가능 범위를 구체적으로 제시하고 있음

[그림 10] 모델 개방성 프레임워크(MOF)의 17개 구성요소



출처: 리눅스 재단 자료(2024) SPRI 보완

[표 3] 3단계 개방성 수준과 활용 범위

MOF 개방 수준	구성요소 개방 조건	유형	활용 범위
클래스 3. 오픈 모델 (Open Model)	1. 모델 구조	코드	• 제약 없는 사용 범위 (접속, 사용, 변경, 재배포) • 제품 및 서비스 개발 • 미세 조정 및 정렬 • 모델 최적화
	2. 최종 모델 매개변수	데이터	
	3. 기술보고서 또는 연구 논문	문서	
	4. 평가 결과	문서	
	5. 모델 카드	문서	
	6. 데이터 카드	문서	
	7. 모델 결과 샘플(선택)	데이터/코드	
클래스 2. 오픈도구 모델 (Open Tooling Model)	• 클래스 3의 모든 구성요소		• 학습 과정의 이해 • 벤치마크 결과 검증 • 추론 최적화
	8. 훈련, 검증 및 테스트 코드	코드	
	9. 추론 코드	코드	
	10. 평가 코드	코드	
	11. 평가 데이터	데이터	
	12. 지원 라이브러리 및 도구	코드	
클래스 1. 오픈 사이언스 모델 (Open Science Model)	• 클래스 2와 3의 모든 구성요소		• 모든 단계의 분석과 검사 • 유사 모델의 재구축 • 데이터 탐색 및 실험
	13. 연구 논문	문서	
	14. 데이터셋	데이터	
	15. 데이터 전처리 코드	코드	
	16. 중간 단계 모델 매개변수	데이터	
	17. 모델 메타데이터(선택)	데이터	

출처: 리눅스 재단 자료 SPRI 번역

○ 선언적인 OSI의 오픈소스AI 정의와 달리 주요 모델들의 개방성을 실질적으로 평가하기 위한 모델 개방성 도구(Model Openness Tool)와 오픈소스 모델에 적용 가능한 개방형 라이선스 정보들을 같이 제공함²⁶

- 모델 개방성 도구는 모델 개방성 프레임워크의 실질적 활용을 위해 구성요소별 공개 수준을 정량화하여 모델 개방성을 실질적으로 평가*하는 도구이며, 현재 베타 버전이 제공되고 있음

* 단순 개방성 유형(클래스) 구분이 아니라 개방성 수준을 정량적으로 평가함으로써 모델 간 개방성 수준 비교가 가능함

[그림 11] 모델 개방성 도구(Model Openness Tool) 평가 결과의 예

Name	Organization	Classification	Last updated	Badge
CrystalChat	LLM360	Unclassified	2025-11-12	Class III - Open Model In progress (60%)
Granite-4.0-H-Micro	IBM	Class III - Open Model	2025-11-12	Class II - Open Tooling Model In progress (60%) Class III - Open Model Qualified
Granite-4.0-H-Micro-Base	IBM	Class III - Open Model	2025-11-12	Class II - Open Tooling Model In progress (60%) Class III - Open Model Qualified
Granite-4.0-H-Small	IBM	Class III - Open Model	2025-11-12	Class II - Open Tooling Model In progress (60%) Class III - Open Model Qualified
Granite-4.0-H-Small-Base	IBM	Class III - Open Model	2025-11-12	Class II - Open Tooling Model In progress (60%) Class III - Open Model Qualified
Granite-4.0-H-Tiny	IBM	Class III - Open Model	2025-11-12	Class II - Open Tooling Model In progress (60%) Class III - Open Model Qualified

출처: 리눅스 재단, “Model Openness Framework”

- 또한, 오픈소스 모델에 적용 가능한 개방형 라이선스 약 700개를 수집하여 OSI 승인, 자유SW 재단 인증 여부 정보를 별도로 제공함으로써 모델 사용자에게 위험성을 직접적으로 판단할 수 있음

[그림 12] 모델 개방성 도구(Model Openness Tool)의 개방형 라이선스 정보 형태

Name	License ID	Content Type	OSI Approved	FSF Lib
3D Slicer License v1.0	3D-Slicer-1.0		No	No
3dfx Glide License	Glide		No	No
Abstyles License	Abstyles		No	No
Academic Free License v1.1	AFL-1.1	code	Yes	Yes
Academic Free License v1.2	AFL-1.2	code	Yes	Yes
Academic Free License v2.0	AFL-2.0	code	Yes	Yes
Academic Free License v2.1	AFL-2.1	code	Yes	Yes
Academic Free License v3.0	AFL-3.0	code	Yes	Yes
Academy of Motion Picture Arts and Sciences BSD	AMPAS		No	No

출처: 리눅스 재단, “Model Openness Framework”

²⁶ 리눅스 재단, Model Openness Framework, “Models”, <https://mot.isitopen.ai/models>, 2025.11.20.

○ 모델 개방성 프레임워크는 OSI의 오픈소스AI 정의와 달리 공개 항목과 사용자를 위한 정보들을 구체적으로 제시하고 있음

- 오픈소스AI 정의와 모델 개방성 프레임워크는 모두 자유SW의 4가지 자유를 기반으로 오픈소스AI 개념을 정의하고 기술적 관점에서 각각의 3가지로 구분하는 공통점이 있음
- 하지만, 적용 범위, 공개 항목의 구체성, 적용 라이선스, 유형 구분, 평가 방식에서 서로 차별성이 있음
 - 오픈소스AI 정의는 AI 시스템 전반을 대상으로 규정하여 적용 대상 및 공개 항목(예: 비머신러닝 시스템)이 모호해지는데, 개방형 모델 프레임워크는 적용 범위를 AI 모델로 한정함으로써 공개 항목을 17개로 세분화하고 적용 가능한 라이선스를 구체화함
 - 오픈소스AI 정의는 기준 충족 여부와 기술적 범주로 유형을 구분하는데, 모델 개방성 프레임워크는 활용 가능 범위를 기준으로 3가지(오픈 사이언스 모델, 오픈도구 모델, 오픈 모델)로 구분하고 평가 점수를 통한 유연한 구분이 가능함

[표 4] 오픈소스AI 정의와 모델 개방성 프레임워크 비교

구성요소	(OSI)오픈소스AI 정의	(리눅스 재단)모델 개방성 프레임워크
적용 대상 및 범위	• AI 시스템 전체 - 데이터, 데이터 정보, 모델, 가중치, 학습/추론 코드, 파이프라인, 비 ML 시스템 포함 가능 - 3가지 개념 간 공개 항목의 경계가 다소 모호	• AI 모델로 한정 - 모델 구성요소 중심으로 실무적 평가
공개 항목의 구체성	• 공개해야 할 범주 중심 (데이터 정보, 파라미터, 코드 등) - “무엇을 공개해야 하는가”는 원칙적 수준 - 항목 단위의 구체적 체크리스트는 없음	• 모델 구성요소 17개 항목을 명확히 제시 - 아키텍처, 파라미터, 학습·추론 코드, 전처리, 평가 데이터, 모델 카드·데이터 카드 등) - 항목당 공개 방식과 조건이 상세히 제시됨
적용 라이선스	• 소스코드에만 OSI 승인(OSI-approved) 라이선스 적용 • 데이터·모델에는 라이선스 용어 미사용 - Term(조항, 약관)이라는 모호한 용어 사용	• 유형별 권장 라이선스 제시 - 소스코드: OSI 승인 라이선스 - 데이터: OSI 승인 혹은 오픈 데이터 라이선스 - 문서: 오픈 콘텐츠 라이선스
유형 구분	• 기준 충족 여부(Yes/No) 판단 - 오픈소스AI 여부 판단 • 기술적 구성요소 기반 판단 - 시스템 → 모델 → 웨이트	• 모델의 공개 항목에 따라 3개 유형(Class)로 분류: - 클래스 1: Open Science Model - 클래스 2: Open Tooling Model - 클래스 3: Open Model • 유형에 따른 활용 가능 범위 제시
평가 방식	• 이분법적 평가(충족/불충족) - 명시적 점수 체계 없음 - 규범·정책 중심 평가	• 체크리스트 기반 등급 평가(Class 1~3) 가능 • 공개 항목을 정량화하여 유연한 등급화 (예: 클래스 2-In progress(90%) 가능) - 기술별 개방도 비교 가능

출처: SPRI 작성

III

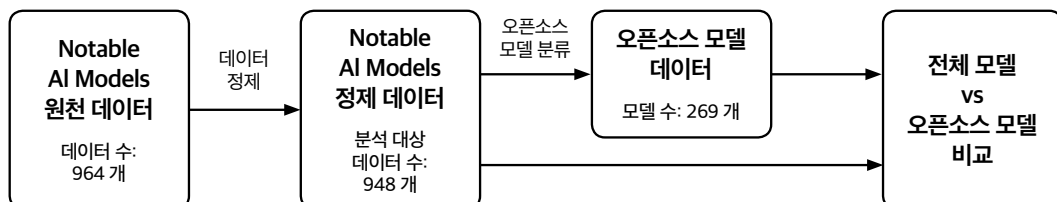
글로벌 오픈소스 모델 동향

1. 글로벌 오픈소스 모델(오픈웨이트) 현황

○ 오픈소스 모델 증가 현상과 현황을 실증적으로 검증하기 위해 Epoch AI의 Notable AI Models (유명 AI 모델) 데이터를 분석함

- Epoch AI는 AI 관련 기술 동향, 경제적·사회적 영향력을 분석·연구하는 비영리 연구기관으로 AI 선도 기업 및 정부 기관과 공동 연구를 수행²⁷
 - OpenAI, 구글 딥마인드, xAI, Anthropic 등의 글로벌 선도 AI 기업들과 공동 연구를 수행하고, 영국 과학기술혁신부와 협업을 추진함
- Notable AI Models²⁸는 전 세계 AI 모델 중 5가지 조건(① 공인 벤치마크의 최첨단 성능 개선, ② 인용 수 1,000개 이상, ③ 기술 발전의 중요성, ④ 학습 비용, ⑤ 중요한 활용) 중 하나 이상을 충족한 AI 모델 정보를 제공하는 자료를 대상으로 분석함
 - 본고에서 분석한 Notable AI Models의 6월 초 데이터를 다운로드하여 일부 데이터 수정 및 삭제* 후 분석함
 - * 총 데이터 964개에서 일부 데이터 손상(필드 값 위치 오류, 데이터값 손상 등)에 따라 최대한 수정 후, 분석이 불가능한 데이터는 삭제하여 최종 948개를 분석함
 - 정제된 유명 AI 모델 정보에서 완화된 오픈웨이트 기준을 적용한 오픈소스 모델(오픈웨이트) 269개를 선별하여 전체 모델과 오픈소스 모델 현황을 비교 분석함

[그림 13] Notable AI Models 기반 오픈 모델 현황 분석 개요



출처: SPRi 작성

²⁷ Epoch AI, <https://epoch.ai/about>, 2025.11.25. 방문

²⁸ Epoch AI, <https://epoch.ai/about>, 2025.11.25. 방문

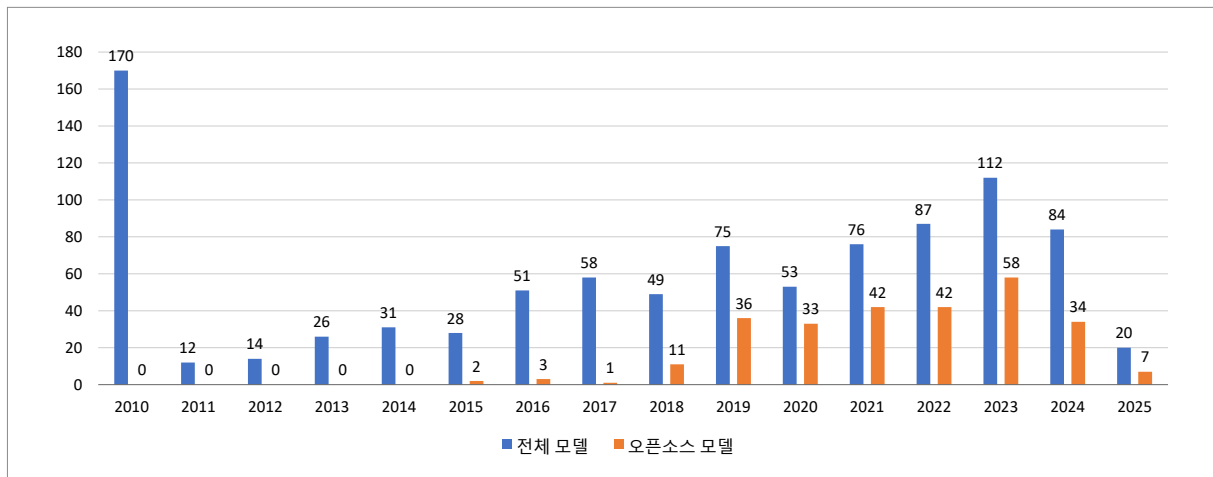
○ 오픈소스 모델 분류를 위해 엄격한 OSI의 오픈소스AI 정의 대신 시장 중심이 완화된 오픈웨이트 개념을 적용하여 현실적인 오픈소스AI 기술 동향을 분석

- OSI의 오픈소스AI 정의는 투명성과 완벽한 재현을 위해 4가지 자유 적용을 엄격하게 규정하고 있으나, 현실적으로 많은 기업이 이 기준을 적용하지 않고 있음
- 따라서, 시장 영향력이 큰 오픈웨이트 기준을 적용하고 오픈웨이트 모델을 포함한 글로벌 오픈소스 모델 동향을 분석하여 산업 관점의 기술 변화를 파악하고자 함

○ 도출된 오픈소스 모델(269개)과 전체 모델(948개)을 비교 분석(개발 연도, 참여 기관 국가, 참여 기관 유형, 모델 유형, 활용 분야, 데이터셋)함

- 유명 AI 모델의 공개 시점을 연도별로 비교 분석

[그림 14] 전체 모델과 오픈소스 모델의 연도별 현황 비교



출처: Notable AI Models 자료 SPRi 분석

- 가장 빨리 공개된 모델은 1950년에 벨 연구소(Bell Labs)에서 공개한 Theseus^{29*}이며, 2015년 오픈소스 모델인 SSD^{30**}가 처음 공개되었으며, 2018년 이후 오픈소스 모델의 비중이 크게 증가하면서 2018년 이후 오픈소스 모델 비중은 47.3%임

²⁹ MIT Technology Review, "Mighty mouse", 2018.12.19.

³⁰ Wei Liu 외 6인, "SSD: Single Shot MultiBox Detector"; arXiv, "Computer Vision, and Pattern Recognition", submitted on 8 Dec. 2015.

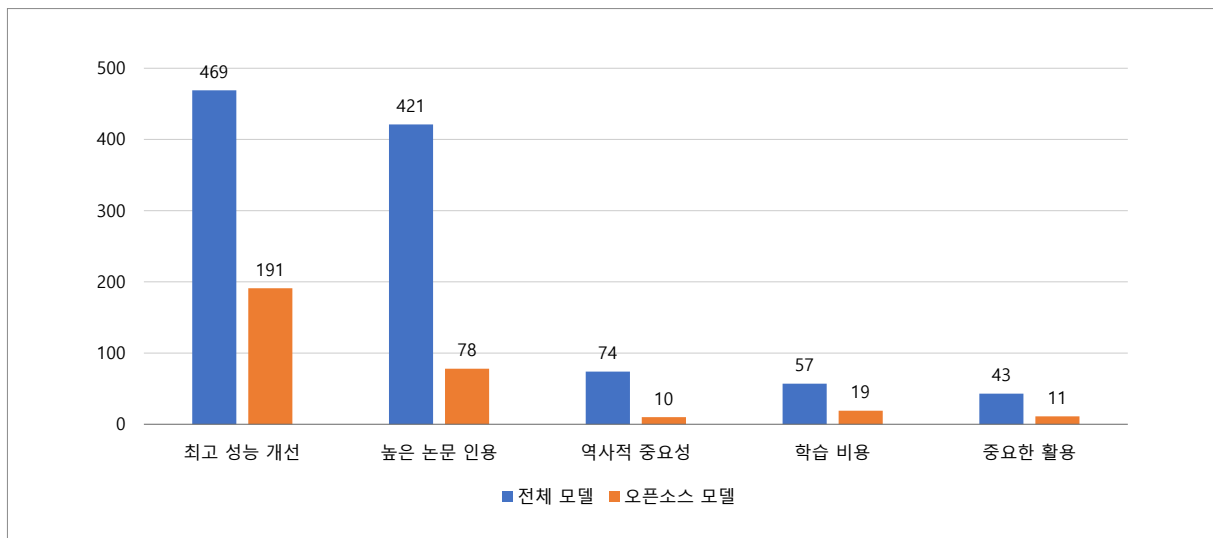
* 미국 벨 연구소의 클로드 섀넌(Claude Shannon)이 1950년에 제작한 세계 최초의 기계 학습 로봇 쥐로, 그리스 신화의 테세우스를 모티브 삼아 25칸짜리 미로를 탐색하고 학습(기억)하여 문제(미로 탈출)를 푸는 전기 기계식 로봇

** Single Shot MultiBox Detector의 약어이며, 2015년 논문으로 공개된 딥러닝 기반 오픈소스 객체 탐지 모델로 주 참여 기관은 노스캐롤라이나 대학과 구글임

- 오픈소스 모델이 가장 많이 공개된 해는 2023년이며 전체 모델 112개 중 약 절반인 58개의 모델(51.8%)이 공개되었고, 비중으로는 2020년에 64.2%(전체 53개 중 33개)로 가장 큰 비중을 차지하였음
- Notable AI Models 데이터는 유명 모델(5가지 조건)을 충족시켜야 하기 때문에 최근 공개(24년 이후)된 오픈소스 모델들이 반영되지 않았을 것으로 추정되며, 이를 감안하면 매년 공개되는 모델 수와 오픈소스 모델 수는 지속 증가할 것으로 예측됨

• 유명 AI 모델의 주요 선정 기준 비교 분석

[그림 15] 전체 모델과 오픈소스 모델의 유명 AI 모델 선정 기준(이유) 비교



출처: Notable AI Models 자료 SPRI 분석

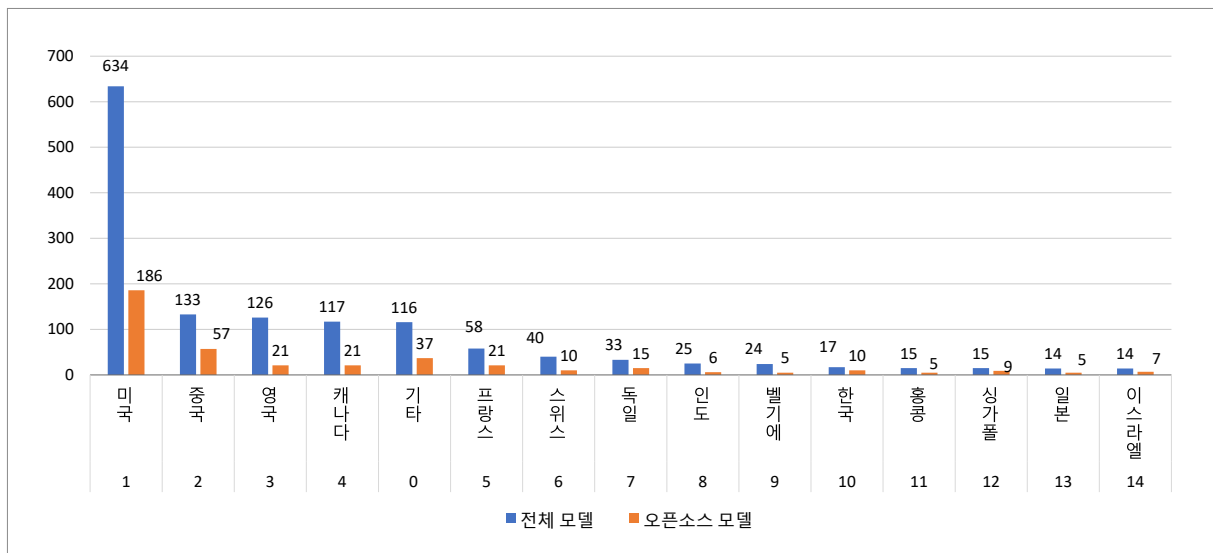
- 전체 모델 기준으로 유명 AI 모델에 가장 많이 선정된 기준은 최고 성능 개선(SOTA(State of the Art) improvement)이 469개로 전체 모델의 49.5%를 차지하였으며, 그다음은 논문 인용(1,000개 이상)이 421개, 이어서 역사적 중요성, 학습 비용*, 중요한 활용 순서임

* 학습 비용은 최근 추가된 선정 기준으로, 과거보다 학습량이 크게 증가하면서 발생한 비용 증가 이슈를 감안한 것으로 추정됨

- 오픈소스 모델도 전체 모델과 유사하게 최고 성능 개선이 가장 많은 191개로 전체 오픈소스 모델의 71.3%를 차지하였고, 이어서 논문 인용(78개), 학습 비용(19개), 중요한 활용(11개), 역사적 중요성(10개) 순서임
- 전체 모델과 오픈소스 모델의 선정 기준을 비교할 때, 최고 성능 개선에서 오픈소스 모델이 차지하는 비중이 40.7%로 오픈소스 모델의 가장 중요한 특징이 성능으로 판단되며, 이어서 학습 비용으로 오픈소스 모델 비중은 33.3%임

• AI 모델 개발 참여 기관의 소속 국가에 대한 비교 분석 결과

[그림 16] 전체 모델과 오픈소스 모델 개발 참여 기관의 주요 국가 비교



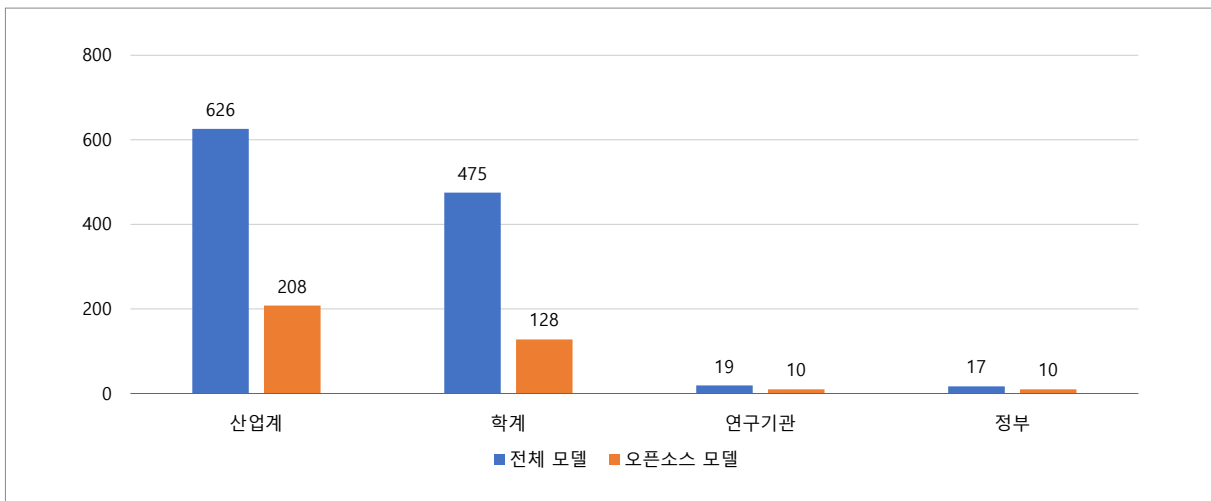
출처: Notable AI Models 자료 SPRI 분석

- 전체 모델 기준으로 미국이 634개로 압도적으로 가장 많았으며, 이어서 중국(113개), 영국, 캐나다, 프랑스, 스위스, 독일, 인도, 벨기에, 한국(10위, 17개) 순서임
- 오픈소스 모델도 미국이 가장 많은 186개로 1위였으며, 이어서 중국(57개), 기타(3개), 영국(21개), 캐나다(21개), 프랑스(21개), 독일(15개), 한국(8위, 10개*), 스위스, 싱가포르 순으로, 우리나라는 유명 오픈소스 모델 순위가 전체 모델보다 높음
- * 독자AI 파운데이션 모델 사업으로 2025년 하반기에 7개의 새로운 오픈소스 모델이 추가됨에 따라 전체 모델 수는 24개, 오픈소스 모델 수는 17개로 증가함
- 주요국 현황에서 오픈소스 모델 비중이 가장 높은 국가는 우리나라(58.82%)로 오픈 모델 연구가 가장 활발하였으며, 가장 낮은 국가는 영국으로 16.67%였음

- AI 모델을 연구하는 기관 유형에 대한 비교 분석 결과

- 산업계가 전체 모델(626개)과 오픈소스 모델(208개) 모두 가장 활발히 참여하고 있으며, 이어서 학계가 전체 모델(475개)과 오픈소스 모델(128개)로 분석되며 AI 기술 혁신을 산업계와 학계가 주도하고 있는 것으로 판단됨
- 연구기관과 정부는 참여 모델 수는 부족하나 오픈소스 모델 비중이 각각 52.63%와 58.82%로 높은 비중을 차지하고, 공공 R&D 분야에서 연구 결과물 확산 및 검증 강화를 위해 오픈소스 모델 비중이 높은 것으로 판단됨

[그림 17] 전체 모델과 오픈소스 모델의 개발 참여 기관 유형 비교

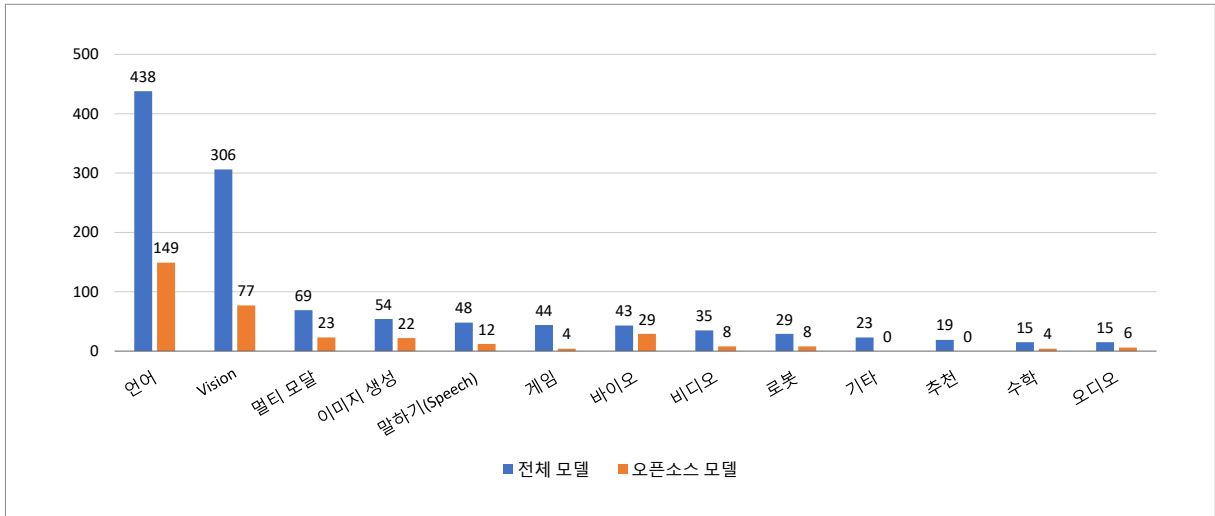


출처: Notable AI Models 자료 SPRi 분석

- AI 모델 유형에 대한 비교 분석 결과

- 전체 모델 기준으로 가장 많이 발표되는 모델 유형은 언어 모델(438개)이며, 이어서 비전(Vision, 306개), 멀티모달(69개), 이미지 생성(54개), 스피치(Speech, 48개), 게임(44개), 바이오(43개), 비디오(35개), 로봇(29개) 순서임
- 오픈소스 모델도 전체 모델과 유사하게 언어 모델(149개)이 가장 많았으며, 이어서 비전(Vision, 77개), 바이오(29개), 멀티모달(23개), 이미지 생성(22개), 스피치(Speech, 12개) 순서임
- 오픈소스 모델 비중이 가장 높은 분야는 67.44%를 차지하는 바이오 분야로 과학계에서 오픈소스 모델 연구가 활발히 이루어지고 있음을 알 수 있으며, 가장 낮은 분야는 게임 분야로 9.1%로 낮은 활성도를 보였음

[그림 18] 전체 모델과 오픈소스 모델의 주요 모델 유형 비교

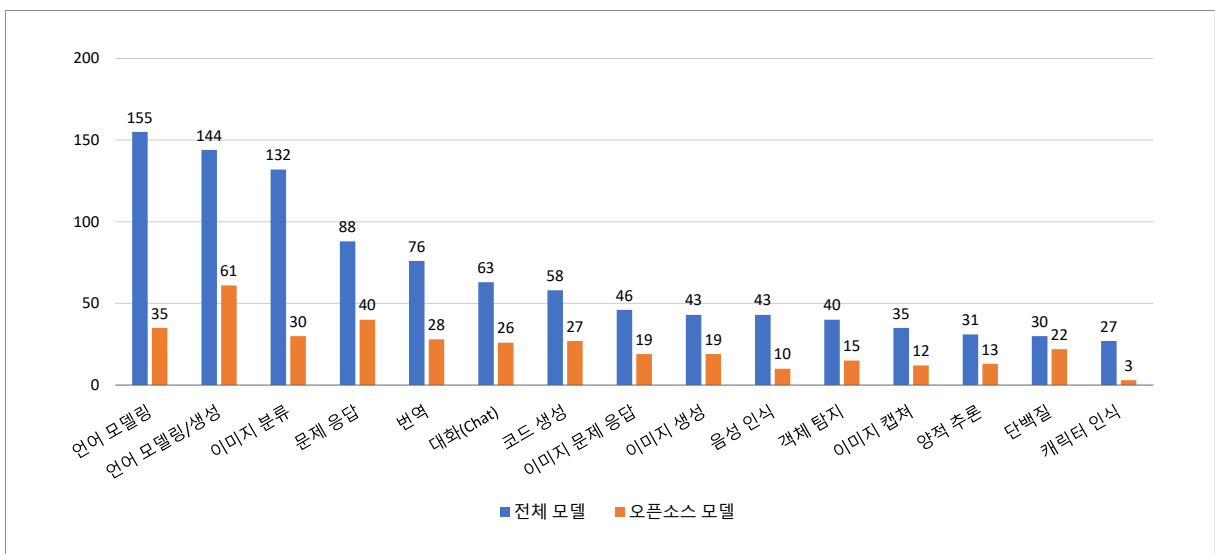


출처: Notable AI Models 자료 SPRI 분석

• AI 모델의 활용 분야에 대한 비교 분석 결과

- 전체 모델 기준으로 가장 많이 활용되는 분야는 언어 모델링(155개)이며, 이어서 언어 모델링/생성(144개), 이미지 분류(132개), 문제 응답(88개), 번역(76개), 대화(63개), 코드 생성(58개), 이미지 문제 응답(46개) 순으로 인간의 인식 및 표현과 관련된 분야(언어, 이미지, 문제 응답, 번역, 대화 등)의 AI 모델 활용이 가장 많음

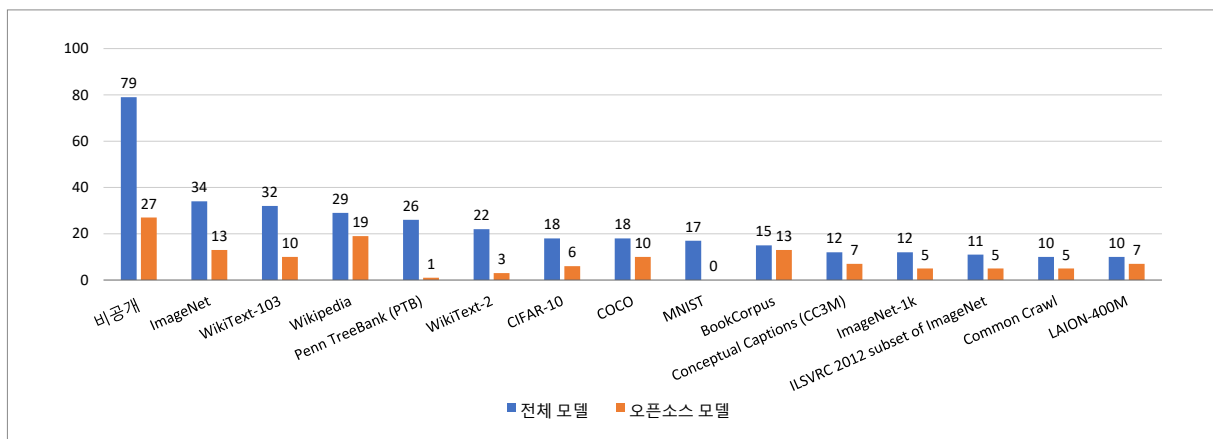
[그림 19] 전체 모델과 오픈소스 모델의 주요 활용 분야 비교



출처: Notable AI Models 자료 SPRI 분석

- 오픈소스 모델은 언어 모델링/생성(61개)이 가장 많았으며 문제 응답(40개), 언어 모델링(35개), 이미지 분류(30개), 번역(28개), 코드 생성(27개), 대화(26개), 단백질(22개), 이미지 문제 응답(19개), 이미지 생성(19개) 순서이며, 전체 모델과 비교할 때 인식보다는 소통 및 문제 해결과 같은 실용적 분야 활용이 보다 활발함
 - 오픈소스 모델 비중이 가장 높은 분야는 바이오에 해당되는 단백질 분야로 73.33%를 차지하며 과학계에서 오픈소스 모델 개발이 활발함을 알 수 있으며, 가장 낮은 분야는 언어 모델링 분야로 22.58%를 차지하며 오픈소스 모델은 모델링보다 언어 생성과 같은 실용 분야의 기술 개발이 활발하다고 판단됨
 - AI 모델의 학습에 사용되는 공개 데이터셋 비교 분석 결과
 - 전체 모델 기준으로 비공개 데이터의 활용 빈도가 79개로 가장 많았으며, 이어서 ImageNet(34개), WikiText-103(32개), Wikipedia(29개), Penn TreeBank(19개), WikiText-2(22개), CIFAR-10(18개), COCO(18개) 순으로 활용 빈도가 높았음
 - 오픈소스 모델은 학습 데이터를 공개해야 함에도 비공개 데이터가 가장 높은 빈도인 27개를 차지하였으며, 이어서 Wikipedia(19개), ImageNet(13개), BookCorpus(13개), WikiText-103(10개), COCO(10개), Conceptual Captions(7개) 순서로 전체 모델과 다소 차이가 있음
 - 주요 학습 데이터셋 중에서 오픈소스 모델 비중이 가장 높은 데이터셋은 65.52%로 가장 높은 활용 비중을 보인 Wikipedia이며, MNIST*는 활용률 0%로 오픈소스 모델에서 학습에 활용되지 않았음
- * 초기 단계의 딥러닝 연구에서 활용된 데이터로 단순한 데이터(손글씨 숫자)³¹ 특성상 학습 효용성이 없기 때문에 사용되지 않은 것으로 판단됨

[그림 20] 전체 모델과 오픈 모델에 활용된 주요 데이터셋 비교



출처: Notable AI Models 자료 SPRI 분석

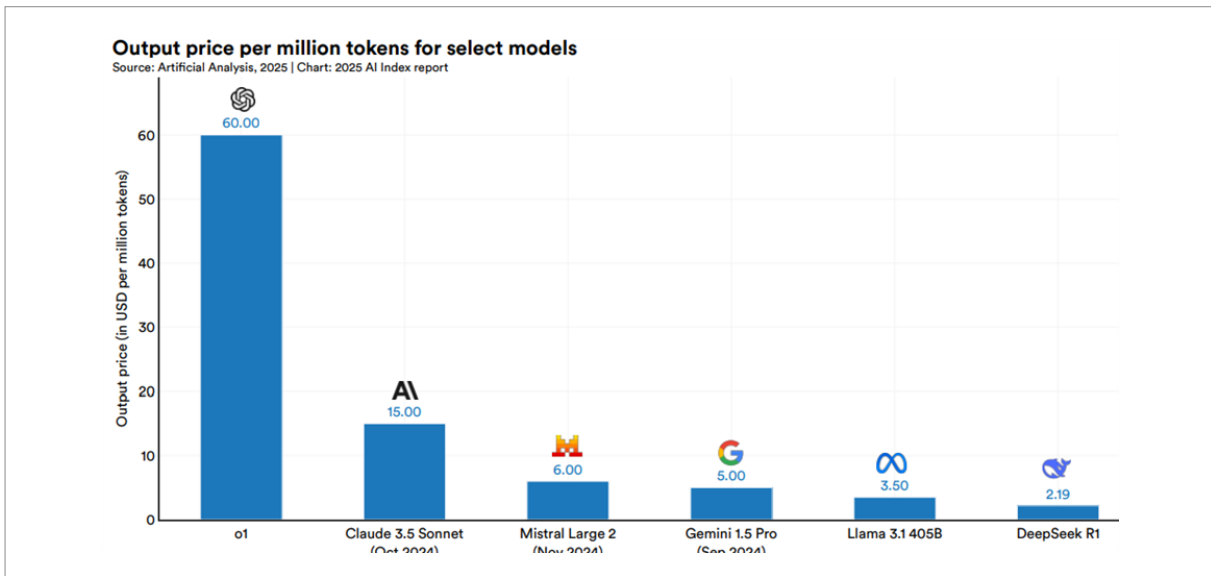
³¹ Wikipedia, "MNIST 데이터베이스", https://ko.wikipedia.org/wiki/MNIST_데이터베이스, 2026.01.19. 방문

2. 마·중 기업의 오픈소스 모델 경쟁

○ 글로벌 오픈소스 모델은 빠른 기술 혁신과 저렴한 비용으로 주목받으면서 최근 마·중 기업들의 AI 기술 주도권 경쟁 수단이 되고 있음

- 오픈소스 모델의 성능이 빠르게 향상되고 비용 측면에서도 상용 서비스인 OpenAI의 ChatGPT 대비 최대 약 1/30 수준으로 낮아지면서 AI 서비스를 구축하려는 많은 기업으로부터 큰 관심을 받고 있음³²

[그림 21] 상용 AI 서비스 대비 오픈소스 모델의 낮은 비용



출처: Stanford HAI, "AI Index 2025", 2025.

- 초기 오픈소스 모델은 미국 기업들을 중심으로 성장하였으나, 최근 딥시크, 알리바바 등의 중국 기업의 오픈소스 모델들이 주목받으며 중국 모델들이 빠르게 확산되고 있음³³
 - 중국의 오픈소스 모델들이 2025년 7월을 기준으로 누적 다운로드 수가 3억 4,900만으로 미국 오픈소스 모델의 다운로드 수(3억 3,400만)를 추월하였으며, 그 중심에는 알리바바의 큐웬과 딥시크가 있음
 - 2025년 2월부터 전 세계 월간 오픈소스 모델 도입 비중에서 중국의 오픈소스 모델(47%)들이 미국(37%)을 추월하였으며, 2025년 8월에는 중국의 오픈소스 모델 비중(64%)이 미국(30%)의 2배를 기록함

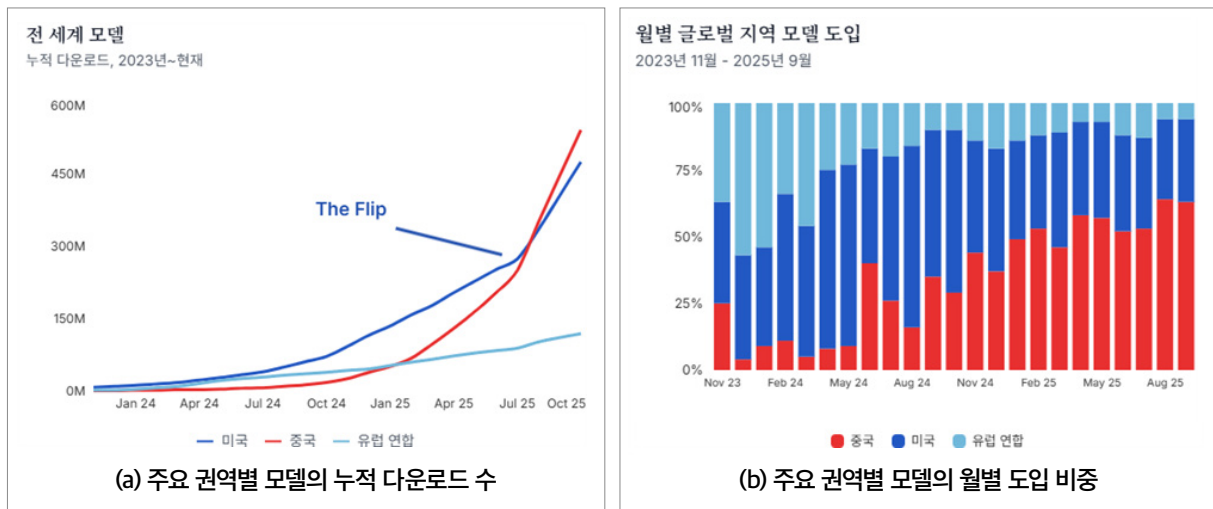
³² Stanford HAI, "AI Index 2025", 2025.

³³ ATOM Project, "State of AI Report", 2025.12. 방문

- 이에 미국에서 AI 리더십 유지를 위한 정부의 오픈소스AI 개발 및 생태계 활성화를 위한 민간 중심의 ATOM 프로젝트*가 추진되고 있음

* 미국의 AI 경쟁력이 중국에 뒤처질 것을 우려하여 오픈소스 모델 개발 선도, 전방위 AI 응용 개발 지원을 촉구하기 위한 프로젝트로 네이션 램버트 주도로 시작되어 2025년 12월 기준 학계, 기업계, 오픈소스 생태계 등을 중심으로 지지받음

[그림 22] 글로벌 생태계에서 중국 AI 모델의 부상

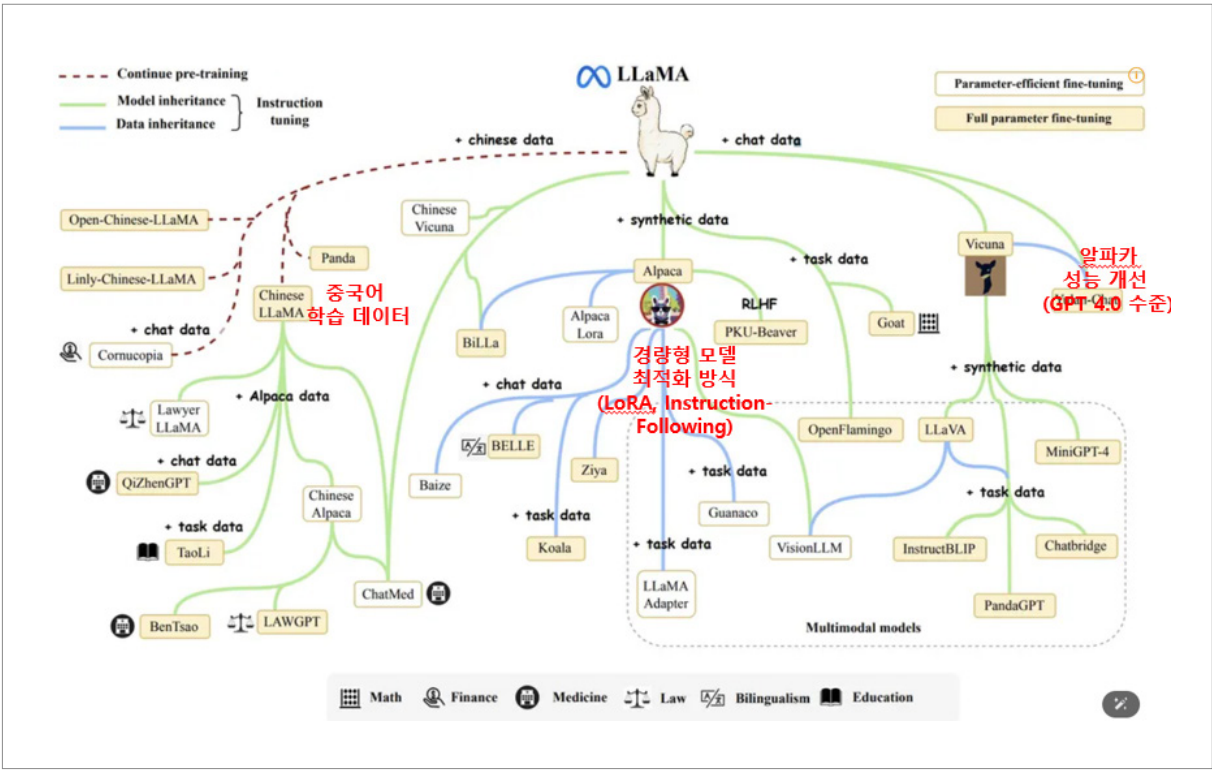


출처: ATOM Project, <https://www.atomproject.ai/>, 2025.12. 방문

○ 오픈소스AI 생태계를 성장시킨 주요 미국 기업으로 메타와 구글이 있으며, 이들은 오픈소스 전략을 통해 생태계 창출 및 영향력 확대를 추진함

- 메타는 ChatGPT에 대항하기 위해 2023년 2월부터 오픈웨이트 모델인 라마를 꾸준히 공개하면서 오픈소스AI 생태계 확산을 주도하고 있음³⁴
 - 메타는 라마 공개 초기부터 학습 데이터의 일부 공개, 선택적 모델 웨이트를 공개하였고, 3.1부터 월간 7억 사용자 이하의 상업적 활용을 허가하며 수익화를 고려한 전략적 생태계 확장(파생 모델* 수가 6만 5천 개 이상)을 위해 공개 정책을 추진하고 있음
 - * 라마의 파생 모델들은 모델 구조 활용, 웨이트 파인 튜닝, 재학습 등을 통해 개발된 새로운 모델(구조 변경, 웨이트 변경 등)을 의미하며, 대표적으로 Alpaca(스탠포드), Vicuna(UC 버클리), 딥시크 등이 있음

³⁴ Meta, Llama Homepage(<https://www.llama.com/>), 2025.11.26. 방문



출처: 나무위키 Llama 페이지 자료 SPRi 보완

- 초기부터 파라미터 수 기준 8B, 70B, 405B 모델을 지속적으로 공개하며 성능(추론 성능, MoE(Mixture of Experts) 도입, 멀티모달 등) 향상을 통해 4.0 버전에서 상용 AI 서비스인 제미니(Gemini) 2.0, GPT-4o보다 우수한 성능을 제공³⁵
 - 메타는 라마 기술을 활용한 자체 SNS 플랫폼(페이스북, 인스타그램 등)의 메타 AI 서비스(월간 활성 사용자 수 10억 명, 챗봇/광고 추천 등) 유료화하여 수익 창출을 시도³⁶하고 있으며 별도로 라마 자체 수익화*도 추진 중
- * 라마 자체 수익화를 위해 API 서비스 구축을 통한 서비스 수익 창출을 시도 중이며, 사업화 확장을 위한 별도의 파트너 생태계 구축 및 스타트업 지원 프로그램을 추진하고 있음

35 Meta, “The Llama 4 herd: The beginning of a new era of natively Multimodal AI innovation”, 2025.04.05.

36 Mirackle AI, “나도 모르게 쓰고 있었네” … 매달 10억 명이 사용한다, 메타 AI 전성시대”, 2025.05.29.

[그림 24] 메타의 라마 관련 파트너(수익화) 생태계



출처: 메타 라마 홈페이지

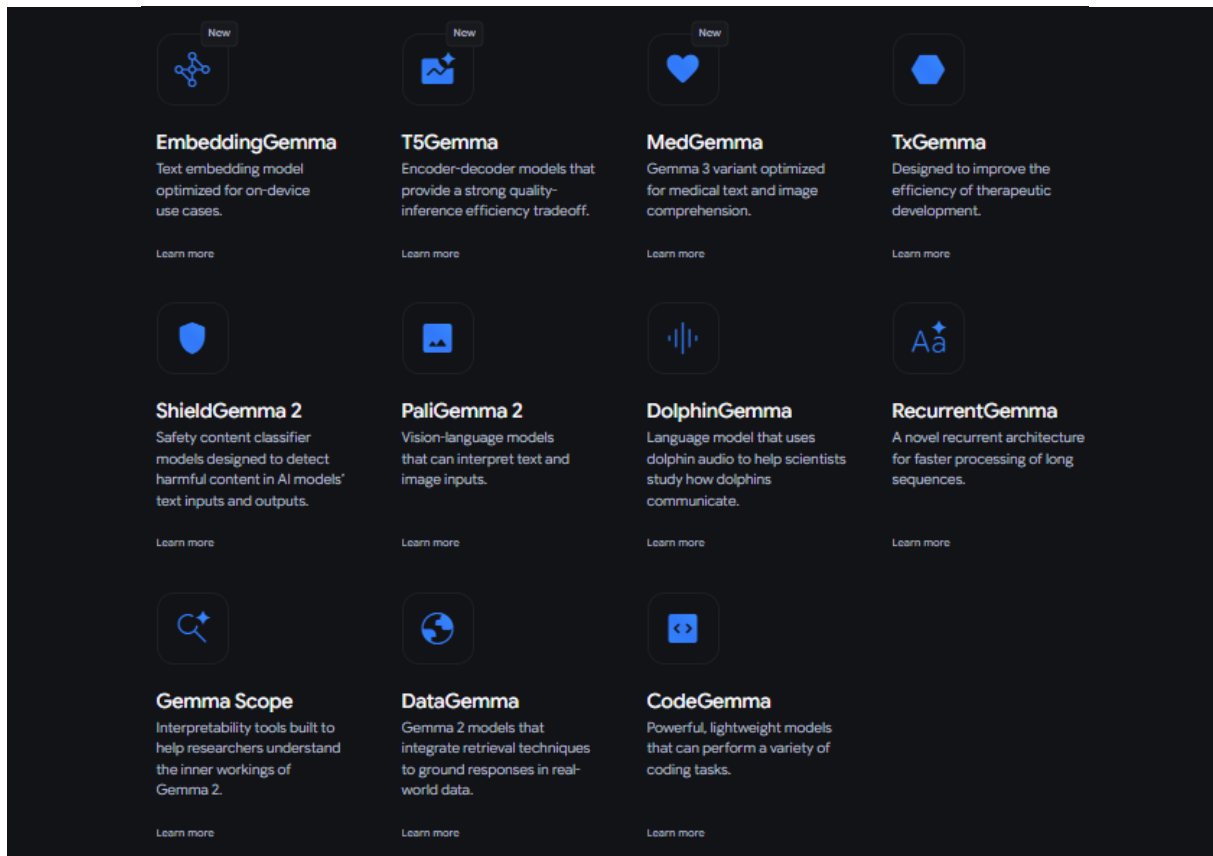
- 구글은 2016년 알파고 및 여러 오픈소스 모델을 통해 AI 기술들을 선제적으로 공개하였으며 최근 메타의 오픈소스 전략에 대응하고자 상용 AI 서비스인 Gemini의 경량 모델인 젬마를 2024년 2월부터 공개함^{37,38}
 - BERT(언어 모델 구조 혁신, 2018)*, T5(인코더-디코더 통합, 2020), PaLM(구글 LLM, 2021) 등 다양한 오픈소스 모델 공개를 통해 기술적 성과를 공개 검증하고 기술 완성도 향상을 전략적으로 추진하여 2023년부터 ChatGPT 대응 수단으로 Gemini 서비스를 유료로 제공함
 - * 기존 단방향의 언어 모델 구조를 양방향 사전 학습이 가능한 트랜스포머 구조로 최초 반영하였으며 사전 학습 이후 파인튜닝이 가능해지고 LLM 시대의 기술적 기반을 제공한 구글의 AI 모델
 - 메타 라마 중심으로 오픈소스 AI 생태계가 확장되면서 구글은 2024년 2월 Gemini의 경량형 모델인 젬마를 공개하며 라마를 견제하고 여러 파생 모델(EmbeddingGemma-온디바이스, MedGemma-의료용, TxGemma-의학연구, PaliGemma-이미지, ShieldGemma-안전성 평가, CodeGemma-코드 생성 등)을 공개하며 생태계 확장 수단으로 활용
 - 특히 구글은 자체 클라우드 서비스(Google Cloud)에서 젬마 인프라(GPU 등) 제공 및 AI 개발 서비스(Vertex AI)*를 통한 파인튜닝 편의성 제공 등을 제공하고, 수익화를 추진하기 위한 로스 리더(Loss Leader, 미끼 상품) 전략을 추진 중임
 - * 구글에서 제공하는 AI 개발 환경으로 AI 솔루션 개발 자동화, 개발 속도 향상, 오류 최소화, 확장성을 제공하고 있으며, 사용량과 고급 기능을 유료화함³⁹

³⁷ 구글코리아 블로그, “구글의 최첨단 오픈 모델 ‘젬마(Gemma)’를 공개합니다”, 2024.02.21., <https://blog.google/intl/ko-kr/products/explore-get-answers/-gemma-open-models-kr/>

³⁸ Google DeepMind, “Gemma”, <https://deepmind.google/models/gemma/>, 2025.11.30. 방문

³⁹ Guru, “Vertex AI란 무엇인가? 정의, 용도 및 기타 [2025]”, 2025.07.13.

[그림 25] 다양한 젤마 파생 모델들



출처: 구글 젤마 홈페이지

○ 2025년 1월 딥시크-R1 모델 공개 이후 최근 중국 기업의 오픈소스 모델들이 주목받고 있으며, 딥시크와 알리바바 등은 적극적인 오픈소스 전략을 통해 뛰어난 기술력을 국제적으로 증명하며 글로벌 영향력을 확대 중

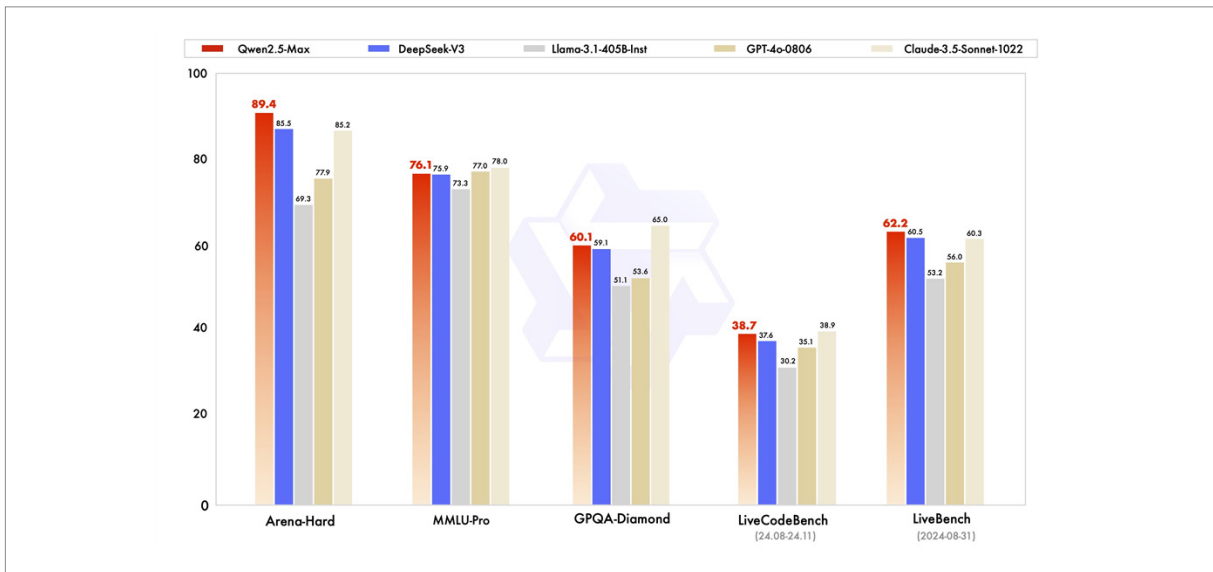
- 중국 AI 스타트업인 딥시크는 2025년 1월 딥시크-R1 모델을 공개하며 중국 AI 기술력을 전 세계에 증명한 대표 사례로 이후 중국 클라우드 업체뿐만 아니라 미국 클라우드 업체의 서비스로 제공되며 중국 AI 기술의 글로벌 진출을 선도함⁴⁰
 - 딥시크-R1은 모델 및 관련 코드의 공개 전략*취하며, OpenAI의 o1 모델과 유사한 추론 능력, 낮은 학습 및 추론 비용**에 대한 공개 검증을 통해 주목받으면서 파라미터 규모 중심의 트렌드에서 알고리즘의 중요성을 각인시킴

⁴⁰ 한컴테크, “DeepSeek-R1 기술 분석”, 2025.03.10.

* 메타(메타 라이선스), 젤마(Gemma Terms of Use)와 달리 MIT 라이선스를 채택하여 OSI 오픈소스AI 정의를 충족하고, 모델 사용자의 저작권 위험도가 낮음(학습 데이터의 저작권 위험은 판단 불가)

** 그룹별 동시 학습 구조, 정확도 보상, 지식 종류, 일부 파라미터 활성화 같은 다양한 SW 기술을 적용하여 학습 및 추론 비용을 낮췄음에도 ChatGPT-4o보다 좋은 성능 혹은 유사한 성능을 보이면서 AI 비용 절감 가능성을 제시함

[그림 26] 큐웬(알리바바)과 딥시크의 주요 벤치마크 결과



출처: Qwen Blog, Qwen2.5-Max: Exploring the Intelligence of Large-scale MoE Model, 2025.01.

- 이후 애플 앱스토어에서 자체 API 서비스가 ChatGPT를 넘어 1위를 차지했으며, 중국 클라우드 서비스(알리바바, 바이두, 텐센트 등)와 통신사들의 딥시크 모델 통합⁴¹과 글로벌 시장 진출*을 추진 중

* AWS, 애저, GCP 등 글로벌 클라우드 서비스의 딥시크 모델 제공⁴², 사우디 아람코 데이터 센터의 딥시크 모델 설치⁴³ 등과 같이 중국을 넘어 글로벌 진출 기회를 얻고 있음

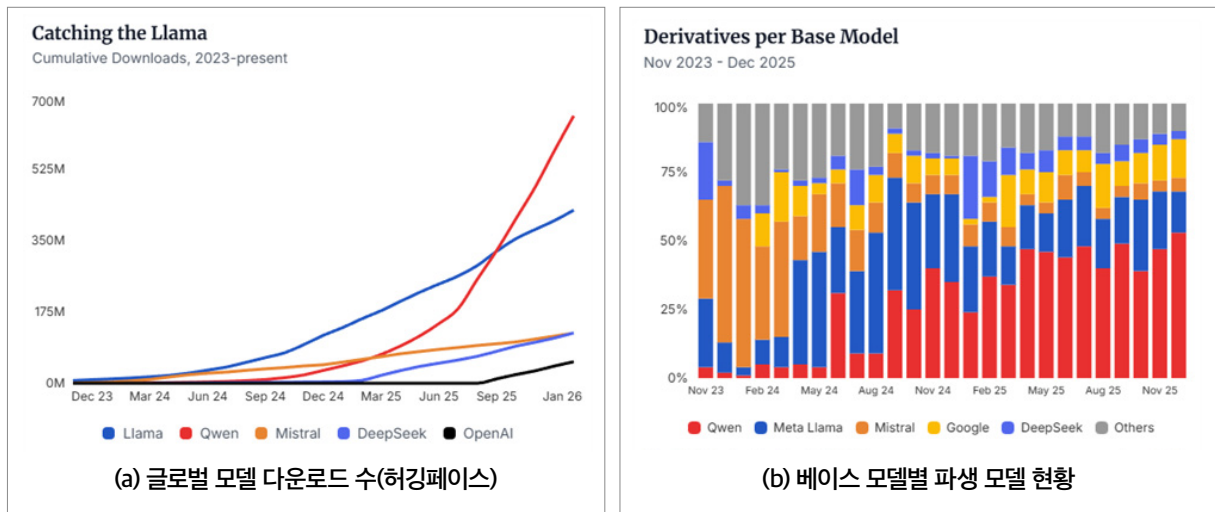
- 중국의 최대 클라우드 서비스 기업인 알리바바도 자체 모델인 큐웬을 2023년 8월부터 공개하고 있으며, 2024년 11월부터 아파치 라이선스를 적용하여 상업적 활용 제약을 풀면서 2025년 전 세계에서 가장 빠르게 확산되는 모델이 됨

⁴¹ China Daily, "DeepSeek's success an opportunity for global AI cooperation", 2025.03.10.

⁴² IT Daily, China Daily, "돌풍 일으킨 中 '딥시크', 빅테크 클라우드 AI 플랫폼에 안착", 2025.02.03., <http://www.itdaily.kr/news/articleView.html?idxno=230712>

⁴³ 조선비즈, "사우디 아람코도 딥시크 선택' ... 美 AI 독주 흔드는 中 'AI 굴기'", 2025.07.02., <https://biz.chosun.com/it-science/ict/2025/07/02/4CDQM7EJHNCRDEGAUYWYK2BAE/>

[그림 27] 주요 오픈소스 모델들의 허깅페이스 누적 다운로드 현황



출처: ATOM Project, <https://www.atomproject.ai/>, 2025.12. 방문

- 알리바바는 300개 이상의 모델 공개를 통한 빠른 기술 혁신, 아파치 라이선스를 통한 무료 활용 허용, 우수한 성능 등으로 2025년 12월 기준 가장 주목받고 있는 오픈소스 모델*로 부상하였음^{44, 45}

* 전 세계에서 4억 회 이상의 다운로드 및 14만 개 이상의 파생 모델이 있을 정도로 널리 활용되고 있으며, 최근 공개한 큐웬3-VL은 높은 정확도와 함께 일부 벤치마크에서 GPT-5 및 Gemini 2.5 프로보다 우수한 기록을 보임

- 알리바바는 큐웬을 활용하여 고객 응대 자동화(챗봇, AI 상담, 콘텐츠 생성), SW 개발 자동화(코드 생성, 디버깅, 테스트 케이스 생성 등) 등으로 비용 절감 및 업무 효율성·생산성을 향상⁴⁶시키고, 주력 사업 경쟁력 강화(이커머스 개인화 추천, 지능형 검색, 배송 최적화, 클라우드 AI 기능 향상 등)를 통한 매출 확대를 추진

- 2024년에 큐웬을 채택한 기업이 9만 개*가 넘어섰으며, 그 배경에는 MaaS(Model-as-a-Service) 제공, 모델 스튜디오(개발 환경) 제공⁴⁷ 등이 있으며, 2025년에는 글로벌 확산 지원을 위해 530억 달러 투자 계획*을 발표함⁴⁸

* 신규 AI 모델, 개발 플랫폼, 인프라 업그레이드를 통해 전 세계 AI 혁신을 가속화하며, 큐웬을 포함한 다양한 오픈소스 모델 제공, AI 기능의 플랫폼화, 애플리케이션 개발 플랫폼 제공 등이 포함됨

⁴⁴ CIO, “성능 높고 비용 낮다” 알리바바의 무료 멀티모달 AI 모델 ‘큐웬3 옴니’ … 분석가가 본 향후 전망”, 2025.09.24.

⁴⁵ tech42, “알리바바 ‘큐웬3-VL’ 기술보고서 공개 … ‘GPT-5·제미니AI 성능 추월’”, 2025.12.01.

⁴⁶ 노코드 버블 개발: 마르코의 콘솔로그, Qwen 2.5-Max: 알리바바의 AI 혁신과 미래 전망, 2025.02.01.

⁴⁷ Alibaba, Qwen LLM Tops 90,000 Enterprise Clients in First Year, 2024.05.09.

⁴⁸ 인공지능신문, “알리바바, 글로벌 고객 AI 역량 강화 나서 … 첨단 AI 모델, 도구 및 인프라 업데이트 등 해외 고객에 본격 서비스”, 2025.04.09.

IV 결론

○ 오픈소스 생태계가 AI 혁신·확산을 선도하면서 오픈소스AI 개념이 새로이 정립되고 있으며, 오픈소스 모델을 중심으로 오픈소스AI가 빠르게 확산되고 있음

- 오픈소스 기술들이 AI 기술 혁신과 확산*을 주도하면서 엄격한 오픈소스AI 개념이 새로 정의되었으나, 기업들은 수익 창출을 위해 모델(구조+매개변수)을 선택적으로 공개하는 오픈웨이트를 선호하고 있음
 - * 오픈소스AI 프레임워크 기술(텐서플로우, 파이토치)이 딥러닝 확산, 모델 대형화를 촉발시켰고, 오픈소스 트랜스포머스 기반의 허깅페이스가 오픈소스 모델 개발 및 활용을 확산시킴
 - AI는 SW에 데이터와 모델(구조+매개변수)이 추가되기 때문에 오픈소스AI는 모든 구성요소 공개를 통해 사용, 연구, 수정, 공유에 대한 제약이 없는 AI 기술임
 - 하지만, AI 선도기업들은 수익 창출을 위해 엄격한 오픈소스AI 개념을 충족하지 않는 오픈웨이트 모델 공개(학습 데이터와 SW인 학습·추론 코드를 비공개)와 사용 조건 제약을 통해 상용 AI 활용을 유도하는 오픈소스AI 사업화(로스 리더) 전략*을 추진함
 - * 사용자 수 제약, 비상업적 활용만 허용하는 등 제약 조건 부여, 개발 편의성·안전성 향상을 유료 지원(개발 환경인 클라우드 인프라 활용 유도 포함)하면서 무료 오픈소스AI(미끼)를 통한 수익화 유도 전략임
- Epoch AI의 유명 AI 모델 분석 결과, 오픈소스 모델*이 2018년 이후 빠르게 증가하고 있으며, 미·중 기업을 중심으로 우수한 성능의 모델들이 공개되고 있음
 - * OSI의 오픈소스AI 정의와 리눅스 재단의 모델 개방성 프레임워크가 최근(24년 10월/11월) 발표되었고, 이들의 엄격한 기준을 충족한 모델이 많지 않기 때문에 보편적인 오픈웨이트 모델(모델 공개)을 기준으로 분석함
 - 유명 AI 모델 데이터 기준, 오픈소스 모델은 2018년 이후 본격적으로 증가하였으며, 2018년 이후 발표된 유명 AI 모델의 47.3%는 오픈소스 모델임
 - 유명 오픈소스 모델 공개는 미·중 기업들이 주도하며, 이들은 우수한 성능(오픈소스 모델의 71.0%)을 기반으로 언어와 비전 모델을 중심으로 언어 모델링/생성, 문제 응답, 이미지 분류, 번역, 코드 생성, 대화(Chat) 분야에서 주로 활용되고 있음

○ 기존 오픈소스 모델은 미국 기업의 주도로 성장하였으나, 최근 중국 기업의 우수한 모델들이 다수 공개되며, 미·중 간의 AI 주도권 경쟁 수단으로 활용됨

- 메타와 구글 등 미국 기업들이 오픈소스 모델 공개를 포함한 오픈소스AI 생태계 형성을 주도하며 AI 신산업을 선도하였으며, 최근 중국 모델 부상에 따른 대응을 위해 오픈소스 모델 개발 협업 추진이 논의되고 있음
 - 미국의 글로벌 테크 기업들은 AI 핵심 기술(프레임워크, 모델 등)을 공개하며 AI 기술 주도권 확보*를 통해 AI 신산업 창출과 성장을 선도하였음
 - * 구글(텐서플로우, 트랜스포머, BERT, PaLM, 젤마 등), 메타(파이토치, 라마 등), 허깅페이스(트랜스포머스 등)를 공개하며 전 세계 AI 기술 발전을 선도함
 - 메타*와 구글은 AI 산업 주도권 확보를 위해 자체 기술을 라마와 젤마로 공개하고 있으며, 이를 활용하여 주력 서비스의 고객 만족도 향상 및 수익화를 동시에 추진*함
 - * 메타는 라마 기술을 활용한 메타 AI 서비스를 통해 고객 만족도 향상, 유료화를 통한 수익 창출과 라마 자체 수익화(API 서비스 제공 등)를 추진하고 있음
 - 글로벌 AI 주도권 유지를 위해 오픈소스 모델 개발 협업 활성화를 위한 민간 중심의 ATOM(America Truly Open Model) 프로젝트가 추진되고 있음⁴⁹
 - * 중국의 오픈소스 모델의 영향력 확대(25년 7월 기준 허깅페이스 누적 다운로드 수 역전)에 대응하고자 1만 개 이상의 GPU와 1억 달러 지원을 위한 공공·민간 협력을 촉구하는 프로젝트로 학계, 기업, 연구기관을 중심으로 추진되고 있음
- 2025년 1월 딥시크 공개 이후, 알리바바 쿼엔 등 중국 모델들이 우수한 성능과 오픈소스 라이선스를 기반으로 빠르게 확산되며 글로벌 비즈니스 협업 기회를 창출하고 있음
 - 딥시크와 알리바바는 우수한 중국 기술력을 전 세계에 공개 입증하였으며, OSI 승인 오픈소스 라이선스(MIT, 아파치)로 자유로운 모델 활용을 허용하면서 빠르게 확산됨
 - * 딥시크와 쿼엔(알리바바)의 준수한 성능과 낮은 학습 비용으로 2025년 7월 기준 허깅페이스 누적 다운로드 수 기준으로 중국 모델이 미국 모델을 역전하였음
 - 빠른 기술 확산과 미국 AI 기술의 대안으로 주목받으면서 중국 및 글로벌 비즈니스 기회를 적극적으로 창출*하고 있음
 - * 중국 내 클라우드 서비스 기업 및 통신사 협업뿐만 아니라 글로벌 클라우드 서비스인 AWS, 애저 같은 글로벌 클라우드 서비스 및 아람코 데이터 센터에 중국 모델이 적용되면서 오픈소스 모델 기반 중국 AI 기업의 해외 시장 진출이 가속화되고 있음

⁴⁹ ATOM Project, <https://www.atomproject.ai/>, 2025.12. 방문

[표 5] 미·중 기업의 오픈소스AI 동향

	메타	구글	딥시크	알리바바
주요 오픈소스AI	파이토치('16), 라마('23)	텐서플로우('15), BERT('18), T5('20) PaLM('21), 젤마('24)	딥시크('23)	큐웬('24)
목적	• OpenAI 견제 • 생태계 확장 • 기술 발전 가속화	• 기술 검증(경량) • 파생 모델(온디바이스, 의료, 코드 생성 등)	• 기술 검증 (낮은 비용, 성능)	• 기술 검증(성능) • 생태계 확장
내부 활용 (기존 제품)	• SNS 서비스 지능화 • 부가 서비스 개발	• 검색, 클라우드, 안드로이드 등 기존 제품·서비스 지능화	• 딥시크 API 서비스	• 고객 응대·SW 개발 자동화 • 이커머스/배송 최적화
AI 수익화	• 상용 AI 서비스 • 편의성·안전성 향상 제품군	• 상용 AI 서비스 (Gemini) 연계 • 클라우드 서비스 수익화	• 상용 AI 서비스 • 기술 공급 (자동차, 클라우드)	• 상용 AI 서비스 • 클라우드 서비스 수익화
주요 라이선스 및 정책	• 라마 라이선스 (OSI 미승인) • 월간 사용자 수 7억 명 이상 사용 불가 • 멀티모달 모델 EU 내 사용 불가	• 젤마 사용 조건 (OSI 미승인) • 구글 경쟁 제품/서비스 사용 불가	• MIT 라이선스 (OSI 승인)	• 아파치 라이선스 (OSI 승인)

출처: SPRi 작성

○ 미·중 간의 AI 기술 경쟁으로 최신 AI 기술 확보(내재화)에 유리한 환경이며, 오픈소스AI는 개방형 검증, 생태계 영향력 확보를 위한 전략적 수단임

- 미·중 간의 오픈소스 기반 AI 기술 패권 경쟁은 우수한 기술들이 공개되면서 해외 선진 AI 기술 내재화를 통한 AI 경쟁력 강화를 추진하기에 유리한 환경임
 - 미·중의 AI 선도기업들은 기술·산업 주도권 확보를 위해 최신 AI 기술을 전략적으로 공개*하고 있으며, 특히 비상업적(연구) 목적 활용을 모두 허용하고 있음
 - * 최신 모델 구조, 학습 방식(코드), 추론 방식(코드)들이 공개되어 연구 및 역설계를 통해 선진 AI 기술 노하우를 빠르게 내재화할 수 있는 긍정적 환경이 조성되고 있으며, 기존 모델을 재학습, 파인튜닝, 경량화를 통한 (자주적인) 산업 특화 모델 개발이 가능함

- 따라서, 최신 오픈소스AI 기술 내재화(단순 활용이 아닌 기술 원리 이해)를 통한 AI 기술력 향상 및 글로벌 경쟁력을 갖춘 한국어 우위 AI 모델 개발을 추진하여 해외 상용 AI 서비스 의존도 절감* 및 국가 주력 산업의 AI 전환이 필요함

* 최근 일본에서 디지털 적자 문제(수십조 원으로 추정)가 주목받고 있으며, 향후 우리나라도 자체 AI 기술력이 부재할 경우에 과다한 AI 서비스 비용 이슈가 발생할 가능성⁵⁰이 크며, 종속성 탈피를 위해서는 기술 자주성 확보가 중요함

- AI 기술의 오픈소스 전략은 개방형 검증을 통한 기술 신뢰성 확보와 영향력 확대(사용자 저변 확보)를 위한 전략적 수단임

- 중국 기업들은 미국 주도의 생태계에서 오픈소스 전략(우수한 모델 공개 → 투명한 검증 → 신뢰성 확보 → 활용 확산)을 통해 빠르게 영향력을 확대*하고 있음

* 우수한 성능의 오픈소스AI는 기술 공개를 통해 효과적인 개방형 검증 및 레퍼런스 확보가 가능하지만, 폐쇄형 상용 AI 기술은 개방형 검증이 불가능하므로 기술 신뢰성 확보 및 레퍼런스 확보를 위해 더 많은 홍보 활동(무료 서비스 제공 등)이 필요할 수 있음

- 따라서, 우수한 성능의 AI 기술 개발 후, AI 기술 공개(오픈소스화) 및 개방형 기술 검증은 기술 신뢰성 확보* 및 영향력 확대(기술 확산)를 위한 전략적 필요 사항*임

* 기술적 난이도가 높은 AI 기술은 공개만으로 기술 노하우 이전이 어려우며, 사용자 참여 기반의 공개 검증을 통한 기술 신뢰성 확보 및 투명하고 공정한 우수 기술력 입증에 유리하고, 검증된 기술은 사용자 중심으로 확산을 가속화할 수 있음

○ 따라서, AI 3대 강국 도약을 위해서는 전략적 오픈소스AI 활용을 통한 선진 오픈소스AI 기반 R&D 추진 및 오픈소스AI 활용 확산 기반 조성이 필요

- AI 기술 주권 확보를 위해 오픈소스AI를 전략적 기술 원천*으로 활용하여 선진 오픈소스AI 기반 R&D 추진을 통해 원천 AI 기술 역량 내재화 및 산업 특화 AI 기술 확보 필요

* 오픈소스AI는 기술을 공개하여 4가지 자유(사용, 연구, 수정, 공유)를 허용하기 때문에 최신 기술 확보 및 내재화에 유리하며, 지적재산권 침해 위험이 낮은 원천 기술로 전략적 가치(개발 속도 향상, 기업 종속성 완화 등)가 있음

⁵⁰ 매일경제, “AI費 비용 충격 오나 … 빅테크, 기업용 요금 줄인상”, 2025.09.02.

- 오픈소스AI 기반 R&D를 강화하여 최신 오픈소스AI 기술 내재화 및 성능 향상을 효율적으로 추진*하여 기술 자주성 확보** 및 해외 기업·기술 의존도 완화를 통한 AI 기술 강국 도약 기반 마련

* [그림 9]와 같은 AI 수명 주기((계획·설계(Plan&design)부터 종료·해체(Retire·decommission)까지)의 대부분 기술이 오픈소스로 공개되어 있기 때문에 해당 기술 내재화 및 성능 개선(안전성 향상, 최적화, 에너지 절감 향상 등)을 효율적으로 추진 가능

** AI는 복잡한 시스템 기술로 기술 독자성(Technological Originality)보다 기술 자주성(Technical Sovereignty), 독자 기술의 위험성(시장 고립, 비효율성 등) 극복과 더불어 자주적인 기술 결정권 확보(비효율 요소 기술 대체, 개선 등)가 중요하며 통제 가능한 외부 기술(오픈소스) 확보를 통한 공급망, 지정학적 위험 대응이 필요함

- 우수한 범용 오픈소스AI 기술을 기반으로 산업별 특화 AI 기술로 재설계 및 선택적 기능 고도화(재설계, 불필요 기술 제거, 특화 기술 추가, 성능 개선 등)를 추진*하여 R&D 효율성 제고 및 성과 확산 가속화 가능

* 오픈소스 생태계는 이미 우수한 오픈소스 기술을 기반으로 통신·네트워크(LF Networking, O-RAN), 자동차(Automotive Grade Linux, Eclipse SDV 등), 제조(LF Edge), 금융(Fintech Open Source), 미디어(Academy Software), 에너지(LF Energy) 분야 같은 산업 특화 기술 개발을 적극 추진하고 있음

- 범국가 AI 기반 대전환을 위한 전략적 기술 도구*로서 오픈소스AI 활용 확산을 위한 기반(생태계) 조성 및 인재 양성이 필요

* 오픈소스AI는 기술을 공개하여 자유로운 사용, 수정, 공유를 허용하기 때문에 지적재산권 침해 위험이 낮아 AI 전환을 위한 기술 인프라 및 도구로서 전략적 가치(비용 절감, 개발 속도 향상 등)가 있으며, 전 세계 기업의 89%가 AI 개발을 위해 오픈소스 기술을 활용함

- 공개 기술로 자유로운 사용이 가능한 오픈소스AI 활용을 확산하기 위해 전략적 활용 기반을 구축*하여 효율적인 AI 대전환과 국가 AI 경쟁력 강화를 추진

* 재정적·인적 자원이 부족한 중소기업들의 오픈소스AI 활용 역량 제고를 위한 컨설팅 제공, 오픈소스AI 활용 기업 발굴 및 육성, 오픈소스AI 기술 실증 제공, 오픈소스AI 커뮤니티 활동 및 협업 지원 등

- 오픈소스AI 전문 인력 양성*을 통해 선진 기술 내재화 및 산업 현장 중심의 활용 역량을 강화하여 범국가 AI 대전환의 효율성을 제고하고 지속가능성을 확보

* 오픈소스 모델 경량화 및 최적화, 오픈소스AI 프레임워크 활용 역량, 오픈소스 기반 데이터 처리 및 학습 등의 역량을 가진 전문 인력 발굴 및 양성을 위한 해커톤·개발자 대회 개최 및 오픈소스AI 기술 교육 프로그램 제공 등

참고문헌

1. 국내문헌

- GNU 운영체제, “자유 소프트웨어란 무엇인가?”, <https://www.gnu.org/philosophy/free-sw.ko.html>, 2025.11.20. 방문
- Wikipedia, “MNIST 데이터베이스”, https://ko.wikipedia.org/wiki/MNIST_데이터베이스, 2026.01.19. 방문
- Mirack AI, “나도 모르게 쓰고 있었네’ … 매달 10억 명이 사용한다, 메타 AI 전성시대”, <https://www.mk.co.kr/news/it/11329443>, 2025.05.29.
- 구글코리아 블로그, “구글의 최첨단 오픈 모델 ‘젬마(Gemma)’를 공개합니다”, <https://blog.google/intl/ko-kr/products/explore-get-answers/-gemma-open-models-kr/>, 2024.02.21.
- Guru, “Vertex AI란 무엇인가요? 정의, 용도 및 기타 [2025]”, 2025.07.13.
- 한컴테크, “DeepSeek-R1 기술 분석”, 2025.03.10.
- IT Daily, China Daily, “돌풍 일으킨 中 ‘딥시크’, 빅테크 클라우드 AI 플랫폼에 안착”, <http://www.itdaily.kr/news/articleView.html?idxno=230712>, 2025.02.03.
- 조선비즈, “사우디 아람코도 딥시크 선택’ … 美 AI 독주 흔드는 中 ‘AI 굴기’”, <https://biz.chosun.com/it-science/ict/2025/07/02/4CDQM7EJHNCRDEGAUYYWK2BAE/>, 2025.07.02.
- CIO, “성능 높고 비용 낮다’ 알리바바의 무료 멀티모달 AI 모델 ‘큐웬3 옴니’ … 분석가가 본 향후 전망”, 2025.09.24.
- tech42, “알리바바 ‘큐웬3-VL’ 기술보고서 공개 … ‘GPT-5·제미니’ 성능 추월”, 2025.12.01.
- 노코드 버블 개발: 마르코의 콘솔로그, Qwen 2.5-Max: 알리바바의 AI 혁신과 미래 전망, 2025.02.01.
- 인공지능신문, “알리바바, 글로벌 고객 AI 역량 강화 나서 … 첨단 AI 모델, 도구 및 인프라 업데이트 등 해외 고객에 본격 서비스”, 2025.04.09.

2. 국외문헌

- LandingAI, <https://landing.ai/case-studies/ai-transformation-playbook>, 2025.06.10. 방문
- Wikipedia, “Comparison of deep learning software”, https://en.wikipedia.org/wiki/Comparison_of_deep_learning_software, 2025.12.25. 방문
- Stanford HAI, “AI Index 2025”, 2025.
- Gitnux, “Report 2025 - Server Statistics”, 2025.04.29.
- DB Engines, “Popularity of open source DBMS versus commercial DBMS”, 2025.11.20.
- GitHub, “Octoverse 2025”, 2025.10.28.
- Hugging Face, <https://huggingface.co/>, 2025.11.21. 방문
- Ryan O'Connor, “PyTorch vs TensorFlow in 2023”, Speech&Text, 2021.12.14.
- Epoch AI, “Compute trends across three eras of machine learning”, 2022.05.02.
- Security Boulevard, “Hugging Face Has Become a Malware Magnet”, 2024.10.23.
- Star History, <https://star-history.com>, 2025.11.25. 방문
- Linux Foundation & Meta, “The Economic and Workforce Impact of Open Source AI”, 2025.05.

- Linux Foundation, “The State of Sovereign AI”, 2025.08.
- Open Source Initiative, “Open Source Definition”, <https://opensource.org/osd>, 2025.06.17. 방문
- Open Source Initiative, “About the Open Source Initiative”, <https://opensource.org/about>, 2025.06.17. 방문
- Open Source Initiative, “OSAID(Open Source AI Definition) v1.0”, 2024.10.28. 발표
- OECD, “OECD AI Principles Overview”, <https://oecd.ai/en/ai-principles>, 2025.06.17. 방문
- Open Source Initiative, “What are Open Weights?”, <https://opensource.org/ai/open-weights>, 2025.12.26. 방문
- Open Source Initiative, “The Open Source AI Definition 1.0 Endorser organizations”, <https://opensource.org/ai/endorsements>, 2025.06.17. 방문
- Open Source Initiative, “The Open Source AI Definition 1.0 FAQ”, <https://opensource.org/ai/faq#why-do-you-require-training-code-while-osd-2-doesn-t-require-compilers>, 2025.06.17. 방문
- Linux Foundation, “The Model Openness Framework Whitepaper”, 2024.12.17.
- OECD, “AI System lifecycle”, 2025.11.20. 방문
- Linux Foundation, Model Openness Framework, “Models”, <https://mot.isitopen.ai/models>, 2025.11.20.
- Linux Foundation, Model Openness Framework, “Licenses”, <https://mot.isitopen.ai/licenses>, 2025.11.20.
- Epoch AI, <https://epoch.ai/about>, 2025.11.25. 방문
- MIT Technology Review, “Mighty mouse”, 2018.12.19.
- Wei Liu 외 6인, “SSD: Single Shot MultiBox Detector”; arXiv, “Computer Vision, and Pattern Recognition”, submitted on 8 Dec. 2015.
- Stanford HAI, “AI Index 2025”, 2025.
- ATOM Project, “State of AI Report”, 2025.12. 방문
- Meta, Llama Homepage(<https://www.llama.com/>), 2025.11.26. 방문
- Meta, “The Llama 4 herd: The beginning of a new era of natively Multimodal AI innovation”, 2025.04.05.
- Google DeepMind, “Gemma”, <https://deepmind.google/models/gemma/>, 2025.11.30. 방문
- China Daily, “DeepSeek’s success an opportunity for global AI cooperation”, 2025.03.10.
- Alibaba, “Qwen LLM Tops 90,000 Enterprise Clients in First Year”, 2024.05.09.