The Transactions of the Korea Information Processing Society
TKIPS Vol.14, No.6, pp.468~479

eISSN: 3022-7011
https://doi.org/10.3745/TKIPS.2025.14.6.468

Evaluating The Openness of Impactful AI Models with A Focus on LLMs

Kil-Won Jeon† · Hyun-Jun Han†† · Kang-Won Lee†††

ABSTRACT

Generative AI models are increasingly driving technical innovations and making societal impacts. Recognizing their significance, governments(e.g, EU AI Act) and technical communities (e.g., OSI) demand a higher degree of openness and transparency in AI development. Open sourcing AI models enables deeper scrutiny of their inner workings, accelerates innovation, and mitigates potential risks. Although numerous AI models are marketed as "open source," many fall short of the traditional standards of openness. Moreover, despite recent efforts, there is currently no comprehensive framework for characterizing the openness of AI models. In this paper, we propose a novel framework to quantify the degree of openness in AI models. We apply our framework to evaluate several high-impact models, including models developed by Korean companies, and investigate the relationship between openness and performance of AI models with a focus on large language models (LLMs).

Keywords : Open Source AI Model, AI Model Openness, Large Language Model, Openness Evaluation Framework

영향력 있는 인공지능 모델의 개방성 평가 – 대형 언어 모델 중심으로

전 길 원† · 한 현 준†† · 이 강 원†††

요 약

생성형 AI 모델은 점점 더 많은 기술 혁신을 이끄는 것과 동시에 사회에 큰 영향을 미치고 있다. 이러한 중요성을 인식하여, 정부(예: EU AI Act)와 기술 커뮤니티(예: OSI)는 AI 개발에서 더 높은 수준의 개방성과 투명성을 요구하고 있다. AI 모델을 오픈 소스로 공개함으로써 내부 처리 과정을 보다 면밀히 검증할 수 있으며, 혁신 속도를 가속화하고 잠재적 위험을 완화할 수 있다. 많은 AI 모델이 "오픈 소스"로 홍보되고 있으나, 전통적인 오픈 소스 기준에는 미치지 못하는 경우가 많다. 게다가, 최근의 노력에도 불구하고 현재까지 AI 모델의 개방성을 체계적으로 특성화하는 종합적인 프레임워크는 아직 존재하지 않는다. 본 논문에서는 AI 모델의 개방성 정도를 정량화할 수 있는 새로운 프레임워크를 제안한다. 이 프레임워크를 활용하여 한국 기업이 개발한 모델을 포함한 영향력이 높은 주요 모델들을 평가하고, 개방성과 AI 모델 성능 간의 상관관계를 대형 언어 모델(LLMs)을 중심으로 조사한다.

키워드 : 오픈 소스 AI 모델, AI 모델 개방성, 대형 언어 모델, 개방성 평가 프레임워크

1. 서 론

대형 언어 모델이 등장한 이후로 인공지능 서비스의 성능이 꾸준하게 증가하며 대형 언어 모델(LLM)에 대한 관심과 오픈 소스 대형 언어 모델(Open Source LLM)에 대한 관심 또한 증가하고 있다(Fig. 1). 오픈 소스 언어 모델뿐만 아니라, 오픈 소스 AI 공개 프로젝트의 수가 2022년에 비해 2023년에 248% 증가하는 등[1] 오픈 소스 AI 전반에 대한 전반적인 관심이 증가하고 있다. 하지만 구체적으로 "어느 정도 개방한 모델이 오픈 소스 AI 모델"이라는 정의는 불분명하다. 최근 이

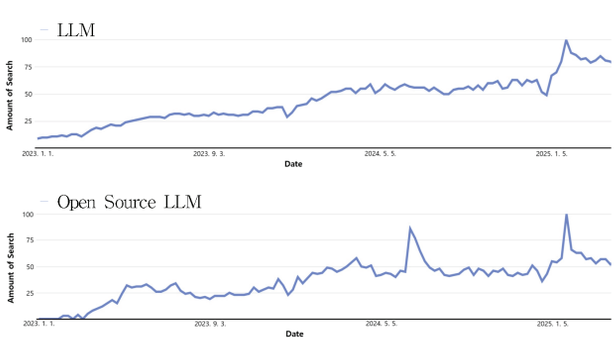


Fig. 1. Google Trend about LLM, Open Source LLM (2023.01.01.~2025.04.10.).

※ 이 논문은 2024년도 세종대학교 교내 연구비 지원에 의한 논문입니다.
† 준 회 원 : 세종대학교 지능기전공학부 스마트기기공학전공 학사
†† 준 회 원 : 세종대학교 컴퓨터공학과 학사
††† 정 회 원 : 세종대학교 컴퓨터공학과 교수
Manuscript Received : April 16, 2025
Accepted : May 11, 2025
* Corresponding Author : Kang-Won Lee(kangwon.lee@sejong.ac.kr)

를 해결하기 위한 다양한 연구가 진행되었는데, Liesenfeld et al.[2]는 14가지 항목으로 개방성(openness)을 구분하는 프레임워크를 제시한 후, 이를 기반으로 오픈 소스를 구분하는 방식을 제안하였다. Liu et al.[3]는 7가지 항목을 기준으로 모델들을 평가하였고, 공개형 모델 구축 프레임워크를 제안하였다. Eiras et al.[4]는 Elo점수, Google Scholar 인용 수, HuggingFace 다운로드 수를 이용해 평가할 모델을 선정하고, 개방성을 5단계로 구분하였다. White et al.[5]는 17개의 구성 요소를 이용하여 3개로 구분하는 모델 개방성 프레임워크(MOF)를 제안하였다.

최근 1-2년 동안 인공지능 모델의 개방성에 관한 연구들이 활발히 이루어졌으나, 각각 조사한 방식과 제안 내용이 상이하며, 오픈 소스 단체(OSI)에서 최근 발표한 내용[6]이 반영되지 않았다는 문제가 존재한다. 또한, 기존 연구에선 국내 AI 모델들이 소외되어 국내 모델의 개방성에 대한 객관적 평가가 이루어지지 않았다.

본 연구의 기여는 다음과 같다:

(1) AI 모델의 개방성을 측정하기 위해 모델의 기본 요소, 접근성 및 재현성, 훈련 방법, 데이터를 종합적으로 평가하는 개방성 평가 프레임워크를 제안한다.
(2) 제안된 프레임워크를 통해 현재 가장 영향력 있는 개방형, 비개방형 인공지능 모델들을 선정하고, 이들을 한국 자체 인공지능 모델들과 함께 평가한다.
(3) 평가 결과를 통해 기존에 개방 모델이라고 널리 알려있는 Llama와 같은 모델이 사실 다른 모델들에 비해 개방성이 높지 않다는 것을 발견한다. 또 개방성과 성능의 상관관계를 분석한다.

본 논문의 구성은 다음과 같다. 2장에서는 오픈 소스 AI의 개방성과 관련된 선행 연구들을 고찰한다. 3장과 4장에서는 각각 연구 대상 모델 선정 방법론과 개방성 평가 프레임워크의 구성 요소를 제시한다. 5장에서는 선정된 모델들에 대한 개방성 평가 결과를 분석하고, 6장에서는 AI 모델 개방성의 이점과 한계점에 대해 논의한다. 마지막으로 7장에서는 본 연구의 결론과 향후 연구 방향을 제시한다.

2. 기존 연구

기존의 오픈 소스 AI 관련 주요 연구들은 다양한 평가 기준으로 AI 모델의 개방성을 분석하였다.

Liesenfeld et al.[2]는 오픈 소스 AI의 개념을 새롭게 정의하였다. 이 논문에서는 14가지 항목으로 개방성(openness)를 구분하는 프레임워크를 제시하였고, 스스로 오픈 소스 또는 오픈 액세스라고 주장하는 모델을 46개를 선정하여 위 항목별로 평가하였다. 또한 기업들이 과학적 검토와 법적 노출을 회피하기 위해 제한적인 공개만 하고서 오픈 소스라고 주장하는 행태(open washing)를 비판하였다.

Liu et al.[3]는 자체 개발한 Amber와 CrystalCoder의 완전 개방 모델로 제안했으며 2021-2023년 사이 공개된 대형 언어 모델들의 개방성이 점차 감소하고 있음을 시사하였다. 이들은 사전학습 코드, 하이퍼파라미터, 중간 체크포인트, 옵티마이저 상태, 데이터 소스 등 7가지 항목을 기준으로 모델의 재현성과 투명성을 이분법적으로 평가하였고, 이를 바탕으로 기존에 개방성이 부족했던 대형 언어 모델들과 차별화된 공개형 모델 구축 프레임워크를 제안하였다.

Eiras et al.[4]는 Elo 점수, Google Scholar 인용 수, Hugging Face 다운로드 수를 이용해 평가할 45개의 대형 언어 모델을 선정하였고, 개방된 정도를 1-5단계로 나누었다. 코드, 데이터, API의 공개 여부를 종합적으로 분석하고, 라이선스까지 포함한 다차원 평가 프레임워크를 제시하였다.

White et al.[5]는 오픈 소스라고 주장되지만 실제로는 코드나 데이터가 충분히 공개되지 않는 Open-Washing 현상을 지적하며, MOF (Model Openness Framework)를 제안하였다. 이 프레임워크는 17개의 구성 요소로 나누어져 있으며, 이를 'Open Science', 'Open Tooling', 'Open Model'의 3단계로 구분해 각 단계에 17개의 구성 요소를 매핑하여 모델의 개방성 단계를 구분한다.

Tarkowski[7]는 오픈 소스 AI와 데이터 거버넌스를 중심으로 다루고 있으며, AI 개발에 대한 데이터 관리와 책임감 있는 공유의 중요성을 강조하였다. 또한 데이터를 모두에게 제약 없이 개방하는 방식보다 데이터의 특성에 따라 맞춤형 공유 및 관리 체계를 구축하는 데이터 커먼즈 모델 도입과 소수의 전문가나 AI 개발자만 결정하는 것이 아닌 여러 이해관계자의 참여하는 거버넌스 체계를 제안하였다. 마지막으로 데이터 준비, 라이선스, 데이터 관리자의 중요성, AI 시스템의 환경 지속 가능성, 공정한 보상, 정책 등 안전한 AI 생태계 구축에 필요한 핵심 요소들에 대해 설명하였다.

Stanford HAI[8]는 공개 모델과 비공개 모델의 예와 각 조직의 입장을 소개하였다. 또한 2024년부터 대형 언어 모델 비공개 모델과 공개 모델의 벤치마크 점수 추세를 조사하였다. 공개 모델과 비공개 모델 각 선두 모델의 성능 차이가 2024년 1월 8% 였던 것에 반해 2025년 2월에는 1.7%로 성능 차이가 좁혀지고 있는 것을 보여주었다.

Bommasani et al.[9](2024)는 2023년 10월 주요 10개의 기업 개발자들에게 100개의 평가 항목으로 모델 투명성 지수(Foundation Model Transparency index)를 조사하였는데 평균 점수가 37점으로 모델들의 제한된 정보만 공개하는 것으로 나타났다. 6개월 이후 14명의 개발자들을 대상으로 조사한 결과 평가 점수가 기존에 비해 21점 향상된 58점의 결과가 나왔다. 또한 100개의 평가 항목 중 13가지 대표 평가 항목을 선정하여 14개의 모델의 개방성 변화율을 백분율로 나타내었다.

OSI (Open Source Initiative)는 오픈 소스의 정의와 평가 기준을 마련하기 위해 1998년 설립된 이후, 개발자 및 사용자 커뮤니티에 신뢰할 수 있는 기준을 제공해 왔다. OSI가 제시한 OSD (Open Source Definition)에서는 오픈 소스로 인정받기 위해 충족해야 하는 기준을 제시한다(소스코드 공개, 2차 저작물 허용, 라이선스 정책 등).

최근 인공지능의 발전으로 AI 시스템 구성 요소들의 개방성이 중요시되면서 OSI는 OSAID (Open Source AI Definition) v1.0을 제시하였다. OSAID는 OSD와 동일한 아이디어로 오픈 소스 AI로 인정받기 위해 4가지 자유 조건이 사용자에게 제공되어야 하는데(Table 1), 사용자가 제약 없이 모델을 사용, 수정, 검사, 연구를 할 수 있어야 한다. 또한 Table 2은 OSI에서 제시한 모델 개방성의 체크리스트로서 모델의 개방성을 이전보다 구체적으로 세분화 하였다. 컴포넌트를 데이터, 코드, 모델 3가지로 구분하여 세부 컴포넌트를 정의하였고 각 컴포넌트에 적용되는 법적 틀을 마련하였다[6].

기존의 인공지능 개방성 연구는 특정 영역 (코드, 가중치)의 개방성에 집중한 경향을 보이며, 개방성 레벨도 이분법, 오분법 등으로 다양하다. 이를 극복하기 위해 OSI에서 제시한 포괄적인 개방성 정의를 반영하여 공개모델의 재사용과 연구를 허용하는지에 대한 체계는 아직 잡혀있지 않다. 또한 기존 연구는 영미권의 개방 모델 위주로 진행이 되어 있어 개방성을 표방하지 않는 상업 서비스와 국내에서 개발된 인공지능 모델에 대한 연구 또한 진행되지 않았다. 본 연구에서는 이러한 문제를 종합적으로 개선하고자 한다.

Table 1. Open Source AI Definition (OSAID) Announced by OSI

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.

Table 2. Model Openness Checklist Provided by OSAID

Component	Detail of Component	Legal Framework
Data	Dataset, Paper, Technical report, Data card	Available under OSI-approved terms
Code	Data preprocessing, Inference code, Train/Validation/Test code, Supporting libraries and tools	Available under OSI-approved license
Model	Model architecture, Model parameter	Available under OSI-approved license/terms

3. 모델 선정 방법론

본 연구에서는 영향력 있는 대형 언어 모델을 대상으로 조사를 진행하며, 연구 대상 모델 선정은 두 가지 접근법을 통해 이루어진다. 첫째, 오픈 소스로 공개된 언어 모델의 경우 HuggingFace 플랫폼에 등재되어 있기에, 해당 플랫폼의 지표를 활용하여 영향력을 평가하고 선정한다. 둘째, 비공개 대형 언어 모델의 경우 시장 점유율과 사용자 성장 추세를 분석한 산업 보고서를 기반으로 선정한다.

3.1 오픈 소스 대형 언어 모델

영향력이 큰 오픈 소스 대형 언어 모델은 사용자의 관심도가 높고 사용성과 확장성이 우수한 모델로 정의할 수 있다. 이러한 관점에서 본 연구에서는 HuggingFace 플랫폼의 좋아요 수(관심도), 다운로드 수(사용성), 자식 모델 수(확장성)를 주요 지표로 활용한다. 이들 지표는 각각 로그 변환(1) 후 Min-Max 정규화(3)를 적용하여 종합 점수를 산출한다.

선정 기준

1) 자식 모델의 수(x_{ci}): 해당 모델을 기반으로 파인튜닝하거나 변형하여 제작한 새로운 모델을 HuggingFace에 업로드한 수치를 의미한다. 이는 HuggingFace의 모든 텍스트 생성 모델 중 원본 모델의 이름이 포함된 모델들을 집계하여 산출한다.

2) 모델의 좋아요 수(x_{li}): 모델이 HuggingFace에 업로드된 이후 사용자들이 좋아요 버튼을 클릭한 총 횟수를 의미한다. 각 사용자는 특정 모델에 대해 한 번의 좋아요만 표시할 수 있다.

3) 전 달의 모델 다운로드 수(x_{di}): 사용자들이 해당 모델을 이용하기 위해 직전 한 달 동안 다운로드한 총 횟수를 나타낸다. 개별 사용자는 여러 번의 다운로드가 가능하다. HuggingFace는 업로드 이후 전체 기간의 누적 다운로드 수를 집계하지 않으며, 이러한 방식은 전체 기간 중 가장 다운로드가 많은 모델을 파악할 수 없다는 한계가 있으나, 신규 모델에 대한 불이익을 줄이는 특징이 있다.

이를 바탕으로 자식 모델 기준의 점수 S_{ci}를 다음과 같이 계산한다.

$$y_{ci} = \ln(x_{ci} + 1) \tag{1}$$

$$Y_c = \{y_{ck} | k = 1, ..., n\} \tag{2}$$

$$S_{ci} = \frac{y_{ci} - \min(Y_c)}{\max(Y_c) - \min(Y_c)} \tag{3}$$

(1)-(3)의 과정을 x_{li}, x_{di}에도 동일하게 적용하여 S_{li}, S_{di}를

구한다.

$$S_i = S_{ci} + S_{li} + S_{di} \quad (4)$$

이렇게 구한 종합점수 S_i를 이용하여 내림차순 정렬한 뒤, 상위 15개의 모델을 선정한다.

위의 방법을 사용하여 최종 선정된 모델은 다음과 같다: Llama3(Meta), DeepSeek-R1(Deepseek), Llama2(Meta), GPT-2(OpenAI), Mistral(Mistral AI), Mixtral(Mistral AI), Qwen2.5(Alibaba), Phi-3(Microsoft), QwQ(Alibaba), DeepSeek-V3(Deepseek), StarCoder(HuggingFace, ServiceNow), opt(Meta), Phi-2(Microsoft), Gemma(Google), Gemma-2(Google)이다.

3.2 비공개 대형 언어 모델

ChatGPT와 같이 오픈 소스로 공개되지 않았으나 대형 사용자 기반을 보유한 모델들은 사회에 지대한 영향을 미친다. 따라서 본 연구는 이러한 영향력 있는 비공개 언어 모델들도 포함하여 종합적인 평가를 진행한다.

미국 IT 기업 FirstPageSage가 발표한 대형 언어 모델 시장 점유율 분석 보고서[10]에 따르면, 시장 점유율 1위는 GPT-3.5와 GPT-4를 기반으로 하는 ChatGPT로 59.7%를 차지한다. 2위는 GPT-4를 활용하는 Microsoft Copilot으로 14.4%의 점유율을 보유하며, 3위는 Gemini 모델을 사용하는 Google Gemini로 13.5%를 차지한다. 4위는 Mistral 7B와 Llama 2를 기반으로 하는 Perplexity로 6.2%의 점유율을 보이고, 5위는 Claude 3 모델을 활용하는 Claude AI로 3.2%를 차지한다. 6위는 Grok 2와 Grok 3을 사용하는 Grok으로 0.8%의 시장 점유율을 기록한다.

본 연구에서는 이 중 자체 모델을 보유하지 않은 서비스(Copilot, Perplexity)를 제외하고, 독자적인 모델을 개발한 GPT, Gemini, Claude, Grok을 최종 조사 대상으로 선정한다.

3.3 국내 언어 모델

국내 언어 모델의 경우, 그 수가 많지 않아 사용자 수가 많을 것으로 예상되는 대기업 모델 위주로 선정하였다. 현재 존재하는 자체 제작 국내 언어 모델로는 LG의 Exaone, 카카오의 Kanana, Naver의 HyperCLOVAX, KT의 mi:dm, SKT의 A.X, 삼성의 Gauss가 있어 이들을 조사 대상으로 선정하였다.

4. 개방성 평가 방법론

본 연구에서는 앞서 제시한 모델 선정 방법론에 따라 선별된 영향력 있는 대형 언어 모델들의 개방성을 종합적으로 평가한다. 평가는 총 네 가지 주요 항목으로 구성되며, 이는 모델 기본 개방성, 훈련 방법론 개방성, 데이터 개방성, 그리고 재현성 및 접근성이다. 이러한 다차원적 평가를 통해 각 모델의 개방성 수준을 체계적으로 분석한다.

4.1 모델 기본 개방성

모델 기본 개방성은 가중치, 코드, 라이선스, 논문, 아키텍처, 토크나이저의 여섯 가지 요소를 통해 평가하며, 이는 모델과 관련된 기본 정보의 공개 수준을 측정하는 항목이다. 각 평가 요소에 대한 세부 내용은 다음과 같다:

1) **가중치**: 학습된 모델의 가중치를 사용자가 접근하고 활용할 수 있도록 공개하였는지 평가하는 지표이다.

2) **코드**: 모델의 학습 및 구현에 사용된 코드의 공개 여부를 평가하는 지표이다. 이는 OSI (Open Source Initiative)에서 정의한 오픈 소스의 핵심 요건인 소스 코드 공개 원칙의 충족 여부를 판단한다.

3) **라이선스**: 모델이 제공하는 라이선스 정책의 개방성을 평가한다. 라이선스는 모델의 사용, 수정, 재배포, 상업적 활용 등 오픈 소스 원칙과 관련된 다양한 요소를 포함한다.

4) **논문**: 모델 개발에 적용된 연구 방법론의 공개 여부를 평가하는 지표이다.

5) **아키텍처**: 모델의 구조적 설계에 대한 공개 여부를 평가한다. 이는 모델의 수정, 재배포 및 재현성 검증을 위해 중요한 요소이다.

6) **토크나이저**: 모델의 학습 및 추론 과정에서 사용되는 토큰화 도구의 공개 여부를 평가하는 항목이다. 토크나이저는 모델 성능에 중대한 영향을 미치므로[11], 모델의 재현성 평가와 후속 연구를 위해 이에 대한 공개 수준을 평가한다.

4.2 접근성 및 재현성

접근성 및 재현성은 하드웨어, 소프트웨어, API의 세 가지 항목을 통해 평가하며, 이는 모델의 재현을 위한 요구사항과 접근성 관련 요소의 공개 수준을 측정하는 항목이다. 각 평가 요소에 대한 세부 내용은 다음과 같다:

1) **하드웨어**: 모델 학습 과정에서 요구되는 하드웨어 사양의 공개 여부를 평가하는 지표이다. 대형 언어 모델은 규모가 커질수록 성능이 향상되는 특성으로[12] 인해 상당한 양의 하드웨어 자원을 필요로 한다. 따라서 재현성 확보와 후속 연구 지원을 위해 사용된 하드웨어의 양과 종류에 대한 정보 공개 수준을 평가한다.

2) **소프트웨어**: 모델 학습 과정에서 활용된 소프트웨어 스택의 공개 여부를 평가하는 지표이다. 모델 개발 과정에서는 다양한 소프트웨어 라이브러리가 사용되며, 이에 대한 정확한 정보는 재현성과 후속 연구에 필수적이다.

3) **API**: 모델을 직접 다운로드하지 않고도 활용할 수 있는 API 제공 여부를 평가하는 지표이다. 무료로 접근 가능한 API가 존재한다면, 이는 오픈 소스의 핵심 원칙인 '누구나 사용할 수 있어야 한다'는 조건을 충족한다고 볼 수 있다.

Table 3. Basic Model Openness Evaluation Criteria

	Open	Semi-Open	Closed
Weights	Model weights are publicly available without any permission required	Model weights are available after obtaining permission	Model weights are not publicly available and cannot be used
Code	The entire code used for training and implementing the model is publicly available	Part of the code used for training and implementing the model is publicly available	The code used for training and implementing the model is not publicly available
License	No restrictions on model use, modification, redistribution, or commercial use (e.g., MIT, Apache, etc.)	One or more restrictions exist on model use, modification, redistribution, or commercial use (e.g., Llama, Google, etc.)	Three or more restrictions exist on model use, modification, redistribution, or commercial use (No corresponding license exists)
Paper	Official paper or technical report exists	Website or blog post exists	No related documentation exists
Architecture	The model's structure and hyperparameters are fully disclosed (e.g., number of transformer layers disclosed, etc.)	The model's structure is disclosed (e.g., use of transformer, etc.)	No information related to the model's structure is disclosed
Tokenizer	The name of the tokenizer used is explicitly disclosed (e.g., SentencePiece tokenizer, etc.)	A downloadable and usable tokenizer exists (e.g., tokenizer registered on HuggingFace)	No information related to the tokenizer is disclosed and it cannot be used

4.3 훈련 방법론 개방성

훈련 방법론 개방성은 사전학습, 파인튜닝, 강화학습의 세 가지 요소를 통해 평가하며, 이는 학습된 모델을 재현하기 위한 훈련 방법론의 공개 수준을 측정하는 항목이다.

4.4 데이터 개방성

데이터 개방성은 총 네 가지 항목으로 평가하며, 이는 사전학습, 파인튜닝, 강화학습(RLHF(Reinforcement Learning with Human Feedback)와 DPO(Direct Preference Optimization) 포함), 그리고 데이터 필터링이다. 이는 모델

Table 4. Accessibility and Reproducibility Evaluation Criteria

	Open	Semi-Open	Closed
Hardware	Full disclosure of the quantity and type of hardware required for training the model (e.g., 1920 H100 GPUs, etc.)	Disclosure of the type of hardware required for training the model (e.g., H100, TPU, etc.)	No disclosure of hardware requirements
Software	Full disclosure of the software specifications required for training the model	Partial disclosure of the software specifications required for training the model (e.g., PyTorch, Pathway, etc.)	No disclosure of software specifications
API	Public API exists	Currently not public, but planned for future release	No API exists

Table 5. Training Methodology Openness Evaluation Criteria

	Open	Semi-Open	Closed
Pre-training	Detailed disclosure of the training methodology and process applied during training, sufficient to reproduce the trained model	Partial mention or explanation of some training methods applied	No explanation of training methodology
Fine-tuning			
Reinforce Learning (including DPO)			

Table 6. Data Openness Evaluation Criteria

	Open	Semi-Open	Closed
Pre-training	Full disclosure of the quantity and types of data used in the model training process	Brief disclosure of the types of data used in the model training process	No disclosure of data used in the model training process
Fine-tuning			
Reinforce Learning (including DPO)			
Data Filtering	Full disclosure of data filtering methodology and content being filtered	Partial disclosure of data filtering methodology or content being filtered	No disclosure of data filtering methodology and content being filtered

학습에 사용된 데이터에 관한 정보의 공개 수준을 측정하는 항목이다.

사전학습, 파인튜닝, 강화학습 항목에서는 각 학습 단계에서 활용된 데이터에 대한 정보 공개 여부를 평가한다. 데이터 필터링 항목에서는 학습 데이터 정제 과정에 적용된 기술과 정제 내용에 대한 공개 수준을 평가한다.

Table 7은 이전 연구들이 본 연구에서 조사하는 개방성 요소들을 포함하는 정도를 나타낸 것이다. 표에서 볼 수 있듯이, 본 연구는 기존 연구들의 프레임워크가 다루지 못했던 개방성 요소들까지 포괄적으로 조사한다. 표기 방식으로는 ○는 해당 카테고리 요소의 2/3 이상이 포함된 경우, ×는 전혀 포함되지 않는 경우, 그리고 그 외의 경우는 △로 표시한다.

예를 들어, Liensenfeld et al.의 연구는 코드, 가중치, 데이터, 라이선스, 논문, 아키텍처, API를 다루고 있다. 이는 본 연구의 모델 기본 개방성 기준 중 5개 요소가 포함되므로 ○로 표기하며, 접근성 및 재현성 기준 중에서는 1개 요소만 포함되므로 △로 표기한다. 훈련 방법론 개방성에 관련된 요소는 전혀 포함하지 않기 때문에 ×로 표기하고, 데이터 개방성 관련 요소는 모두 포함하므로 ○로 표기한다.

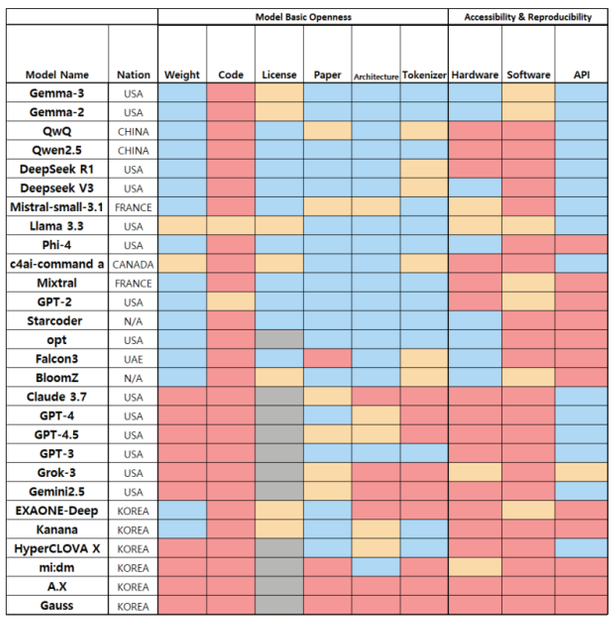
Table 7. Openness Framework Coverage by Previous Research

	Model Basic Openness	Accessibility and Reproducibility	Training Methodology Openness	Data Openness
Ours	○	○	○	○
OSD	△	△	×	×
OSAID	○	△	×	○
Liesenfeld et al. [2]	○	△	×	○
Liu et al. [3]	△	×	△	○
Eiras et al. [4]	△	×	×	○
White et al. [5]	○	△	×	○

5. 개방성 평가 결과

본 연구에서는 앞서 제시한 모델 선정 방법론에 따라 영향력 있는 언어 모델들과 국내 대형 언어 모델을 선정하여 조사한 결과(Fig. 2, Fig. 3)를 바탕으로 개방성 순위(Fig. 4)를 산출하였다. 본 절에서는 프레임워크 조사 결과와 개방성 점수를 기반으로 오픈 소스 언어 모델의 개방성에 관한 분석 결과를 제시한다.

5.1 프레임워크 조사 결과 분석



*국가 정보가 'N/A'로 기재된 모델은 여러 국가의 개발자들이 공동으로 기여한 프로젝트
*라이선스가 'N/A'로 기재된 모델은 공개하지 않아 라이선스 정책이 따로 없는 언어 모델

Fig. 2. Model Basic Openness, Accessibility, and Reproducibility Survey Results

5.1.1 모델 기본 개방성 측면(Fig. 2)

1) 분석 결과, Llama3.3, GPT-2를 제외한 모든 조사 대상 모델이 소스 코드를 공개하지 않는 것으로 나타난다. Llama3.3, GPT-2 역시 완전한 코드 공개가 아닌, 실행 불가능한 일부 코드만을 제공하고 있다. 여기서 소스 코드는 모델 학습 및 훈련에 사용된 코드를 의미하며, HuggingFace에 가중치 활용을 위해 공개된 포함하지 않는다.

2) 라이선스 측면에서는 조사 대상 모델의 약 40%의 모델이 사용에 제약이 1개 이상 존재하는 라이선스를 채택하고 있으며, 특히 국내 개방 모델들은 사용 제약이 있는 라이선스를 사용하고 있는 것으로 확인된다.

3) 논문 측면에서는 해외 비공개 모델들이 자체 연구 내용을 담은 논문을 공개하는 경향을 보인 반면, 국내 비공개 모델의 경우 Naver의 HyperCLOVA X외에 연구 내용을 공개하는 모델은 없는 것으로 나타난다. 추가로 주목할 만한 점은 오픈 소스를 표방하는 Falcon3가 본 연구 시점(2025년 4월)까지 논문을 공개하지 않고 있다는 점이다.

4) 아키텍처 공개 측면에서는 다수의 개방형 모델이 정보

를 공개하고 있으나, Mistral의 경우 정확한 구조 파악이 불가능한 수준의 제한적 정보만을 제공하고 있다. 또한, 비공개 모델들은 대체로 아키텍처를 공개하지 않는 경향을 보이나, OpenAI는 예외적으로 논문을 통해 아키텍처 관련 정보를 제공하고 있다.

5) 토크나이저의 경우, 50% 이상의 오픈 소스 언어 모델이 구체적인 토크나이저 사양을 공개하지 않고 있으며, 대신 트랜스포머 라이브러리를 통해 사용가능한 토크나이저만을 제공하고 있다.

5.1.2 접근성 및 재현성 측면(Fig. 2)

1) 하드웨어 측면에서는 오픈 소스 언어 모델 중 40% 미만이 사용한 GPU 유형 및 수량을 공개하고 있으며, 그 외의 모델은 관련 정보를 전혀 제공하지 않는 것으로 나타난다.

2) 소프트웨어 환경의 경우, 오픈 소스 언어 모델의 약 30%만이 사용한 소프트웨어를 일부 공개하고 있으며, 전체 소프트웨어 스택을 완전히 공개한 모델은 존재하지 않는다.

3) API 측면에서는 약 55%의 오픈 소스 모델이 API를 함께 제공하고 있다. 비개방 모델의 경우에도 70%가 API를 제공하거나 제공 예정인 것으로 확인된다. 국내 모델의 경우 Naver의 HyperCLOVA X를 제외하고는 API를 제공하지 않는다.

5.1.3 훈련 방법론 개방성 측면(Fig. 3)

1) 모든 모델들의 훈련 방법이 사전학습, 파인튜닝, 강화학습을 모두 포함하는 것은 아니다. 예를 들어, GPT-2의 경우에는 파인튜닝이나 강화학습을 진행하지 않고 사전학습만 진행한다. 이런 경우, 파인튜닝과 강화학습은 'N/A'로 표기한다.

		Training Methodology Openness			Data Openness			
Model Name	Nation	Pre-training	Fine-Tuning	Reinforcement learning / DPO	Pre-training	Fine-Tuning	Reinforcement learning / DPO	Data Filtering
Gemma-3	USA							
Gemma-2	USA							
QwQ	CHINA							
Qwen2.5	CHINA							
DeepSeek R1	USA							
Deepseek V3	USA							
Mistral-small-3.1	FRANCE							
Llama 3.3	USA							
Phi-4	USA							
c4ai-command a	CANADA							
Mixtral	FRANCE							
GPT-2	USA							
Starcoder	N/A							
opt	USA							
Falcon3	UAE							
BloomZ	N/A							
Claude 3.7	USA							
GPT-4	USA							
GPT-4.5	USA							
GPT-3	USA							
Grok-3	USA							
Gemini2.5	USA							
EXAONE-Deep	KOREA							
Kanana	KOREA							
HyperCLOVA X	KOREA							
mi:dm	KOREA							
A.X	KOREA							
Gauss	KOREA							

Fig. 3. Training Methodology Openness and Data Openness Survey Results

2) 훈련 방법론 측면에서는 BloomZ를 제외한 오픈 소스 언어 모델 중 사용한 사전학습, 파인튜닝, 강화학습 방법론을 완전히 공개한 모델이 존재하지 않는다. 더욱이 오픈 소스를 표방함에도 학습 방식을 전혀 공개하지 않는 모델도 확인된다.

5.1.4 데이터 개방성 측면(Fig. 3)

1) 데이터 측면에서는 학습에 사용된 데이터셋을 정확하게 명시하고 전체 공개한 모델은 BloomZ 외에는 존재하지 않는 것으로 나타난다.

2) 데이터 필터링 방식의 경우, 데이터의 다른 항목에 비해 상대적으로 많은 모델이 전체 공개를 하고 있는 것으로 확인되었다. BloomZ의 경우 기존 공개 데이터셋을 활용하여 별도의 필터링을 적용하지 않았기 때문에 해당 정보를 공개하지 않은 것으로 추정된다.

이러한 분석 결과는 오픈 소스 언어 모델이 표방하는 개방성과 실제 제공되는 정보 사이에 상당한 괴리가 존재함을 시사한다.

5.2 개방성 평가 점수

본 연구에서는 앞서 제시한 프레임워크 조사표(Fig. 2, Fig. 3)를 기반으로 개방성 평가를 수행한다. 각 항목에 대해 완전 개방(1점), 준개방(0.5점), 비개방(0점)으로 점수화하여 정량적 분석을 진행한다.

	Rating			
Model Name	Ours	OSAID Liesenfeld [4] White [13]	Liu [3]	Eiras [1]
Gemma-3	5	6	7	9
Gemma-2	1	4	7	9
QwQ	19	20	19	19
Qwen2.5	6	3	2	2
DeepSeek R1	9	9	6	7
Deepseek V3	3	2	2	2
Mistral-small-3.1	16	15	19	19
Llama 3.3	8	11	11	13
Phi-4	4	1	1	1
c4ai-command a	11	13	12	9
Mixtral	14	8	7	6
GPT-2	11	9	14	13
Starcoder	7	6	10	8
opt	13	14	15	17
Falcon3	15	16	17	18
BloomZ	1	5	4	4
Claude 3.7	22	21	22	21
GPT-4	25	22	24	22
GPT-4.5	26	26	27	27
GPT-3	18	16	13	12
Grok-3	22	22	26	24
Gemini2.5	22	24	23	24
EXAONE-Deep	21	18	17	15
Kanana	10	12	5	4
HyperCLOVA X	16	18	15	16
mi:dm	19	24	19	24
A.X	27	27	24	22
Gauss	28	28	27	27

Fig. 4. Openness Rank

오픈 소스에 대한 다양한 관점을 반영하기 위해 Table 7을 기반으로 ○는 가중치 1, △는 가중치 0.5, ×는 가중치 0으로 중복을 포함하여 총 6가지 가중치 체계를 적용한다.

각 개방성 영역별로 항목 점수의 합을 해당 영역의 항목 수로 나누어 정규화된 평균값을 도출한다. 이는 영역별 항목 수 차이로 인한 가중치의 중복 적용을 방지하기 위함이다. 이후 각 영역의 평균값에 5.2.1에서 언급한 가중치를 적용하여 최종 종합 순위(Fig. 4)를 산출한다. OSAID, Liensenfeld, White는 가중치가 같아 통합하여 나타내었다.

5.3 분석 결과

1) 일반적으로 오픈 소스로 널리 알려진 Llama 시리즈는 예상과 달리 어떠한 평가 방식에서도 상위 5위 안에 진입하지 못하며, 본 논문을 제외한 다른 논문의 기준에서는 10위를 벗어나는 것도 확인할 수 있다. 이는 오픈 소스 언어 모델에 대한 사회적 인식과 실제 개방성 간의 괴리를 시사한다.

2) Llama에 비해 상대적으로 개방 모델이라는 인식이 낮은 Gemma-2, Gemma-3 모델이 모든 평가 방식에서 일관되게 Llama보다 높은 순위를 기록한다. 이는 대중적 인식과 실제 개방성 수준 간의 불일치를 보여주는 사례이다.

3) 일반적인 개방성 순위 상위권에는 BloomZ, Phi-4, Deepseek V3, Qwen2.5, starcoder 등이 위치한다. 이들 모델은 다양한 개방성 측면에서 균형 있는 정보 공개를 실천하고 있는 것으로 보인다.

4) 국내 오픈 소스 모델 중에서는 Kakao의 Kanana 모델이 가장 높은 순위를 기록하며, 본 논문을 제외한 다른 기준에서는 4, 5등의 순위를 기록하기도 한다.

5) 국내 EXAONE-Deep 모델의 경우, 오픈 소스로 공개되었음에도 불구하고 실제 개방성 순위는 21위로 모든 기준에서 해외 비공개 모델인 GPT-3(18위)보다 낮게 나타났다. 이는 단순히 가중치 공개만으로는 진정한 의미의 오픈 소스라고 보기 어렵다는 점을 시사한다.

6) 중국의 QwQ모델은 오픈 소스로 분류됨에도 불구하고, 가중치 외 대부분의 정보를 비공개하여 개방성 순위가 19위에 불과하다. 반면, 국내 비공개 모델인 Naver의 HyperCLOVA X는 연구 관련 정보를 적극적으로 공개하여 모든 기준에서 오픈 소스 모델인 QwQ보다 높은 16위를 달성한다.

7) 가중치 체계에 따라 모델들의 순위가 크게 변동하는 현상이 관찰된다. 예를 들어, Phi-4는 본 연구의 종합 평가 기준에서는 4위를 기록하나, 다른 모든 개별 평가 기준에서는 1위를 차지한다. 오픈 소스 모델을 평가하는 방식에 따라 기준이 크게 변할 수 있음을 시사한다.

이러한 분석 결과는 오픈 소스 언어 모델의 개방성이 단순히 이분법적으로 판단할 수 없는 복합적인 특성을 지니고 있으며, 다양한 측면에서의 종합적 평가가 필요함을 보여준다. 또한, 오픈 소스로 분류되는 모델들 간에도 실제 개방성 수준에 상당한 차이가 존재한다는 점은 오픈 소스 언어 모델에 대한 보다 정교한 분류 체계의 필요성을 시사한다.

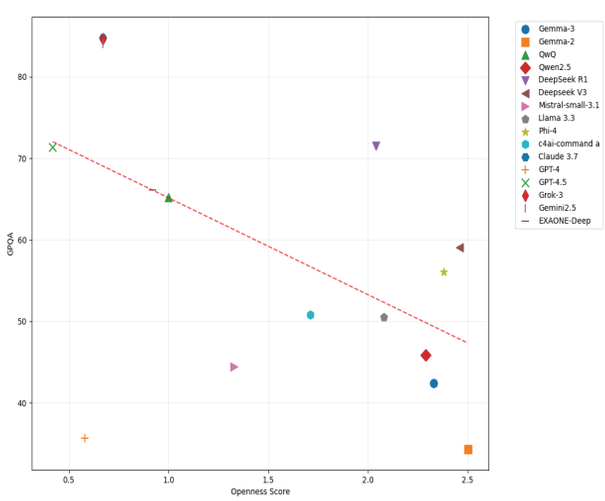


Fig. 5. Correlation between GPQA and Openness Score

5.4 개방성 점수와 성능과의 관계

본 연구에서는 언어 모델의 개방성 수준과 실제 성능 간의 관계를 체계적으로 분석한다. 분석에는 GPQA, MMLU, Elo 세 가지 주요 벤치마크를 활용하며, X축은 모델의 개방성 점수를, Y축은 각 벤치마크에서의 성능 점수를 나타낸다.

5.4.1 GPQA 벤치마크와 개방성의 관계(Fig. 5)

GPQA(Graduate-Level Google-Proof Q&A) 벤치마크에서는 모델의 개방성과 성능 간에 뚜렷한 음의 상관관계가 관찰된다($y = -9.17x + 74.71$, Pearson Correlation:-0.55). 구체적으로:

1) 가장 높은 개방성 점수를 기록한 모델(그래프 우측 끝)이 가장 낮은 GPQA 성능을 보인다.

2) 반대로, 가장 낮은 개방성 점수를 가진 모델(그래프 좌측 끝)이 가장 우수한 GPQA 성능을 기록한다.

이러한 결과는 GPQA와 같은 고난도 벤치마크에서 우수한 성능을 보이는 모델들은 그들의 기술적 우위를 보호하기 위해 낮은 수준의 개방성을 유지하는 것으로 해석할 수 있다.

5.4.2 MMLU 벤치마크와 개방성의 관계(Fig. 6)

반면, MMLU(Massive Multitask Language Understanding) 벤치마크에서는 개방성과 성능 간에 뚜렷한 상관관계가 관찰되지 않는다($y = -1.01x + 75.41$, Pearson Correlation:-0.08):

1) 개방성이 가장 높은 모델과 가장 낮은 모델 간의 MMLU 성능 차이가 미미하게 나타난다.

2) 전반적으로 MMLU 점수는 개방성 수준과 무관하게 분포하는 양상을 보인다.

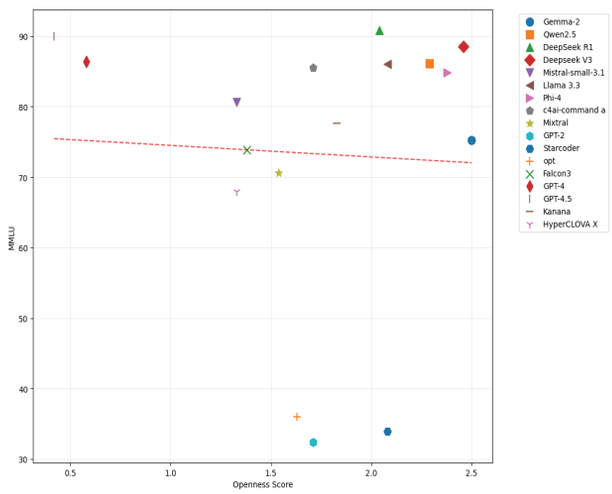


Fig. 6. Correlation between MMLU and Openness Score

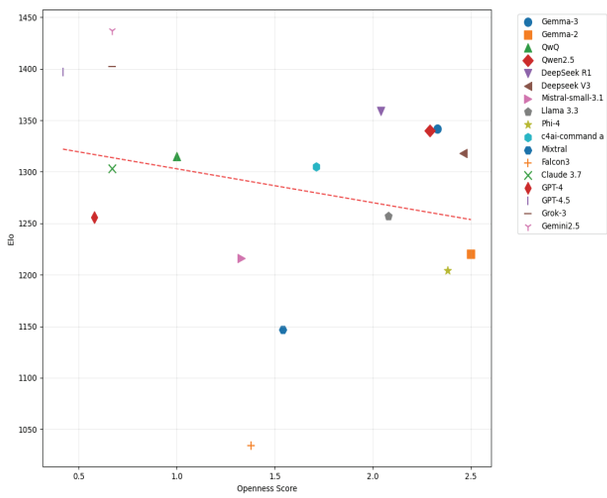


Fig. 7. Correlation between Elo and Openness Score

이러한 결과는 MMLU 벤치마크의 난이도와 관련이 있는 것으로 보인다. MMLU는 GPQA에 비해 난이도가 낮은 편으로, 최근의 대형 언어 모델들은 개방성 수준과 상관없이 이 벤치마크에서 대체로 높은 성능을 달성한다. 실제로 최신 연구에서는 MMLU 대신 MMLU-Pro와 같은 심화 버전의 벤치마크를 활용하기도 하나, 해당 지표를 다룬 모델이 많지 않아 본 연구에서는 MMLU를 기준으로 분석을 진행한다.

5.4.3 Elo 벤치마크와 개방성의 관계 (Fig. 7)

Elo 벤치마크에서는 GPQA와 MMLU의 중간 정도의 상관관계가 관찰된다($y = -7.31x + 1302.18$, Pearson Correlation:-0.24):

1) 개방성과 성능 간에 약한 음의 상관관계가 존재하지만, GPQA에서처럼 뚜렷하게 나타나지는 않는다.

다른 벤치마크에 비해 데이터 포인트들이 추세선 주위에 더 밀집되어 분포하는 경향을 보인다.

Elo 점수는 모델 간 직접 비교를 통해 상대적 성능을 평가하는 방식으로, 다양한 작업과 난이도를 포괄하는 종합적인 지표이다. 이러한 특성으로 인해 개방성과 성능 간의 관계가 보다 일관된 패턴으로 나타나는 것으로 해석할 수 있다.

5.4.4 종합 분석

세 가지 벤치마크(GPQA, MMLU, Elo)를 통해 AI 모델의 개방성과 성능 간의 관계를 종합적으로 분석한 결과, 다음과 같은 주요 패턴이 확인된다:

1) 난이도에 따른 상관관계 차이: 고난도 작업(GPQA)에서는 개방성과 성능 간에 뚜렷한 음의 상관관계가 나타나는 반면, 상대적으로 난이도가 낮은 작업(MMLU)에서는 그러한 상관관계가 약화되는 경향이 있다. 이는 기술적 난이도가 높은 영역일수록 비공개 모델이 경쟁 우위를 가지고 있음을 시사한다.

2) 개방성과 성능의 트레이드오프: 모든 벤치마크에서 추세선의 계수가 음수로 나타난 점은 현재 AI 생태계에서 개방성과 성능 사이에 일정 수준의 트레이드오프 관계가 존재함을 의미한다. 이는 최고 성능의 모델을 개발하는 기업들이 경쟁 우위를 유지하기 위해 핵심 기술 요소를 비공개로 유지하는 경향과 일치한다.

3) 데이터 분포의 의미: Elo 벤치마크에서 다른 벤치마크에 비해 데이터 포인트들이 추세선에 상대적으로 더 가깝게 분포하는 것은 종합적 성능 비교 평가에서 개방성의 영향이 보다 일관되게 나타남을 의미한다.

이러한 종합 분석 결과는 AI 모델의 개방성과 성능 간의 관계가 단순한 이분법적 구도를 넘어, 작업의 난이도, 평가 방식, 기술 성숙도 등 다양한 요인에 의해 영향을 받는 복잡한 현상임이 보여진다. 또한 현재의 기술 환경에서는 최고 성능을 추구하는 모델과 높은 수준의 개방성을 추구하는 모델 사이에 일정한 간극이 존재하며, 이는 AI 생태계의 다양성과 균형을 위한 정책적, 기술적 고려가 필요함을 시사한다.

6. AI 모델 개방성의 필요성 및 도전적인 문제

6.1 개방성의 장점

오픈 소스 AI는 모델의 내부 구조와 학습 과정을 투명하게 공개함으로써 사용자의 신뢰성을 높여주는 효과가 있다. 실제로 한 조사에 따르면 AI 애플리케이션의 신뢰도 조사에서 39%의 응답자만이 신뢰한다고 답했으며, 학습데이터 공개가 오픈 소스 AI의 기반이 되어야 한다는 설문조사 결과가 있다 [36, 37].

이는 AI에 대한 사회 전반의 신뢰도가 아직 낮다는 것을 의

미한다. 이러한 상황에서 오픈 소스 AI는 투명성을 통해 AI 모델이 의료, 법적 의사결정 등 높은 신뢰성을 요구하는 작업에 활용될 가능성을 높인다[4]. 또한 학습데이터를 공개함으로써 학습 과정에서 벤치마크 데이터가 학습데이터로 유입되는 '벤치마크 오염' 문제를 파악할 수 있다[3, 38, 39, 40].

오픈 소스 AI는 모델 개발의 초기 비용을 낮춤으로써 스타트업이나 연구 기관의 진입장벽을 낮추어, AI 산업 전반에 긍정적인 효과를 준다[4]. 또한, 오픈 소스 AI를 기반으로 빠른 프로토타입 제작이 가능하며, 다수의 산업에서 사용자가 인공지능을 접목시킴으로써, AI 산업이 다양한 분야에서 발전될 수 있다[14].

다수의 제 3자가 다방면으로 테스트하고, 피드백을 제공함으로써 오픈 소스 AI 모델이 다양한 운영 조건에 대한 적응성과 완결성을 확보할 수 있다. 또한, 오픈 소스 AI 모델은 최적화, 경량화, 파인튜닝을 통해 성능 개선이 되며, 알고리즘상의 버그나 보안의 취약점을 발견하여 안전성을 높일 수 있다. 다양한 국가와 문화의 사용자들은 피드백을 통해 오픈 소스 AI 모델의 편향성을 줄이는데 기여한다[4].

오픈 소스 AI 커뮤니티는 단순히 파라미터 수와 컴퓨팅 자원을 확장하는 것보다 적응성과 계산 효율성에 중점을 두고 있다. 실제로 다수의 오픈 소스 대형 언어 모델들은 특정 작업에 맞추어 파인튜닝을 하며 발전하고 있다. Llama-2는 Grouped Query Attention을 통해 메모리 오버헤드를 줄이는 동시에 생성 언어의 품질을 높이고, Mistral-7B는 비교적 작은 매개변수 규모에도 불구하고 여러 언어 벤치마크에서 GPT-3.5보다 성능이 뛰어넘는 성과를 이루었다[42].

6.2 개방성의 도전적 문제

동시에 오픈 소스 AI는 여러 리스크와 도전과제가 존재한다. 빅테크 기업의 AI 모델 개발은 막대한 자본과 시간이 투자되는데, AI 모델을 오픈 소스로 공개하면 경쟁사가 쉽게 사용할 수 있게 되어 기업 경쟁력이 약화된다. 이러한 현실적인 이유로 일부 기업들은 초기 모델을 오픈 소스로 배포하여 의미있는 개방성을 지향하다가, 파라미터만 공개하는 전략을 채택하기도 한다[2, 43]. 이는 위 평가 결과처럼, 개방성과 모델 성능이 트레이드 오프 관계에 놓인 원인 중 하나가 된다.

또한 불특정 다수가 개발한 모델이 사회 또는 타인에게 피해를 끼칠 경우, 책임의 주체를 명확하게 정하기 어렵다는 문제가 있다.

모델의 재사용과 수정이 쉬워 모델을 악의적으로 변형시키거나, 혹은 오픈 소스 AI 커뮤니티에 올라온 다양한 비 검열 모델들이 악의적인 개인이나 조직에 의해 오남용되기 쉽다. 사진, 글, 영상을 조작하여 가짜뉴스를 생성하거나 음란물을 생성하는 사례가 발생하고[4], 인종 차별적인 발언이나 선동, 자해같이 극단적인 행동을 권유, 옹호하는 발언을 하는 챗봇 등으로 악용될 수 있다.

모델이 오픈 소스로 공개되어 있다 하더라도 대형 모델일 경우 실행시키거나 파인튜닝 과정에 많은 컴퓨팅 자원이 요구된다. 따라서 오픈 소스의 접근성 부분이 여전히 제한적이라는 도전과제가 존재한다.

이를 해결하기 위해 AI 모델을 제한적으로 공개함으로써 오픈 소스 모델의 장점을 취하면서 문제점을 피할 수 있는 방법 또한 고민되어야 한다. 예를 들어, AI 모델과 훈련 데이터를 협약을 맺은 일부 책임감 있는 외부 연구자에게만 기술 개방을 하여 연구, 감시를 할 수 있도록 하는 방식을 도입할 수 있다. 또한 기존 오픈 소스 라이선스와 새롭게 만들어질 라이선스에 윤리적인 사용 목적만 허용하는 조건을 명시하는 것을 의무화하며, 지속적인 오픈 소스 커뮤니티를 활성화 하여 성능 발전과 모델 경량화에 도모하고, 오픈 소스 AI 거버넌스를 체계적으로 운영함으로써 기술 발전과 윤리적인 책임을 함께 도모하는 방법을 모색해야 할 것이다.

7. 결론 및 향후 과제

본 연구에서는 기존에 제시된 다양한 척도를 고려하고, OSI에서 제시한 OSAID를 준용하여 모델의 기본 개방성, 접근성 및 재현성. 훈련 방법론, 데이터 개방성을 포함하는 종합적 개방성 평가 프레임워크를 제안하였다. 이를 통해 많은 언어 모델들이 스스로 '개방형' 또는 '오픈 소스' AI 모델이라고 지칭하지만, 이러한 사실이 실제 개방성 수준을 일관되게 반영하지 않는다는 점을 확인하였다. 또한 국내 모델들이 해외 모델들에 비해 개방성이 상대적으로 낮다는 문제점을 발견하였다. 인공지능 모델을 개방하는 데는 실질적인 어려움이 존재하지만, 개방을 통해 얻을 수 있는 장점이 많기 때문에 SOTA(State Of The Art) 모델의 개방을 지속적으로 장려하는 방법을 찾아나가는 것이 필요하다.

본 연구는 이에 대한 시작으로 대형 언어 모델에 대한 개방성에 초점을 맞췄다. 향후 이미지 생성 모델, 멀티모달 AI, 에이전트 모델 등 다양한 AI 모델의 개방성도 조사하는 것이 필요하다. 또한 본 연구는 연구자가 직접 논문, 오픈 소스 커뮤니티(e.g., HuggingFace, Github), 홈페이지나 블로그를 참고하여 수작업을 통해 AI 모델의 개방성을 조사하고 평가하였다. 지속적으로 새로운 인공지능 모델들이 출현하고 있으므로, 이들의 개방성을 자동으로 분석할 수 있는 시스템을 개발하여, 다양한 AI 모델의 개방성을 지속적으로 평가하고 현행화하는 리더보드 시스템을 구축하는 것이 필요하다.

References

[1] HyperLAB, Cooperative Competitiveness—Open Source and AI [Internet], https://hyperlab.hits.ai/blog/opensource.

[2] A. Liesenfeld and M. Dingemanse, "Rethinking open source generative AI: open-washing and the EU AI Act,"

in *ACM FAccT*, 2024.
[3] Z. Liu et al., "LLM360: Towards fully transparent open-source LLMs," *arXiv preprint arXiv:2312.06550*, 2023.
[4] F. Eiras et al., "Risks and opportunities of open-source generative AI," *arXiv preprint arXiv:2405.08597*, 2024.
[5] M. White et al., "The Model Openness Framework: Promoting Completeness and Openness for Reproducibility, Transparency, and Usability in Artificial Intelligence," *arXiv preprint arXiv:2403.13784*, 2024.
[6] OSI, The Open Source AI Definition – 1.0 [Internet], https://opensource.org/ai/open-source-ai-definition.
[7] A. Tarkowski, "Data governance in open source AI: Enabling responsible and systemic access," *Open Future*, 2025.
[8] HAI, Artificial Intelligence Index Report 2025 [Internet], https://hai.stanford.edu/ai-index/2025-ai-index-report.
[9] R. Bommasani et al., "The Foundation Model Transparency Index v1.1: May 2024," *arXiv preprint arXiv:2407.12929*, 2024.
[10] FirstPageSage, Top Generative AI Chatbots by Market Share – March 2025 [Internet], https://firstpagesage.com/reports/top-generative-ai-chatbots/.
[11] S. Choo and W. Kim, "A study on the evaluation of tokenizer performance in natural language processing," *Taylor & Francis*, 2023.
[12] OpenAI, "Language Models are Few-Shot Learners," *arXiv preprint arXiv:2005.14165*, 2020.
[13] A. Kamath et al., "Gemma 3 Technical Report," *arXiv preprint arXiv:2503.19786*, 2025.
[14] M. Riviere et al., "Gemma 2: Improving Open Language Models at a Practical Size," *arXiv preprint arXiv:2408.00118*, 2024.
[15] Qwen Team, QwQ: Reflect Deeply on the Boundaries of the Unknown [Internet], https://qwenlm.github.io/blog/qwq-32b-preview/.
[16] A. Yang et al., "Qwen2.5 Technical Report," *arXiv preprint arXiv:2412.15115*, 2024.
[17] D. Guo et al., "DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning," *arXiv preprint arXiv:2501.12948*, 2025.
[18] A. Liu et al., "DeepSeek-V3 Technical Report," *arXiv preprint arXiv:2412.19437*, 2024.
[19] DeepSeek, DeepSeek-V3-0324 Release [Internet], https://api-docs.deepseek.com/news/news250325.
[20] A. Grattafiori et al., "The Llama 3 Herd of Models," *arXiv preprint arXiv:2407.21783*, 2024.
[21] M. Abdin et al., "Phi-4 Technical Report," *arXiv preprint arXiv:2412.08905*, 2024.
[22] Open AI, Language Models are Unsupervised Multitask Learners [Internet], https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf.
[23] R. Li et al., "StarCoder: may the Source be with you!," *arXiv preprint arXiv:2305.06161*, 2023.
[24] S. Zhang et al., "OPT: Open Pre-trained Transformer Language Models," *arXivpreprint arXiv: 2205.01068*, 2022.
[25] A. Ahmadian et al., "Command A: An Enterprise-Ready Large Language Model," *arXiv preprint arXiv:2504.00698*, 2025.
[26] N. Muennighoff et al., "Cross Lingual Generalization through Multitask Finetuning," *arXiv preprint arXiv:2211.01786*, 2022.
[27] Anthropic, Claude 3.7 Sonnet System Card [Internet], https://assets.anthropic.com/m/785e231869ea8b3b/original/claude-3-7-sonnet-system-card.pdf
[28] OpenAI, "GPT-4 Technical Report," *arXiv preprint arXiv:2303.08774*, 2023.
[29] OpenAI, GPT-4.5 System Card [Internet], https://cdn.openai.com/gpt-4-5-system-card-2272025.pdf.
[30] T. Brown et al., "Language Models are Few-Shot Learners," *arXiv preprint arXiv:2005.14165*, 2020.
[31] x.ai, Grok 3 [Internet], https://x.ai/news/grok-3.
[32] Google DeepMind, Gemini 2.5: Our most intelligent AI model [Internet], https://blog.google/technology/google-deepmind/gemini-model-thinking-updates-march-2025/.
[33] LG AI Research, "EXAONE Deep: Reasoning Enhanced Language Models," *arXiv preprint arXiv: 2503.12524*, 2025.
[34] Kanana LLM Team, "Kanana: Compute-efficient Bilingual Language Models," *arXiv preprint arXiv: 2502.18934*, 2025.
[35] NAVER Cloud HyperCLOVA X Team, "HyperCLOVA X Technical Report," *arXiv preprint arXiv:2404.01954*, 2024.
[36] A. Lawson et al., "2023 Open Source Generative AI Survey Report," Technical Report, *Linux Foundation*, 2023.
[37] N. Gillespie et al., "Trust in artificial intelligence: A global study," Technical Report, *University of Queensland and KPMG*, 2023.
[38] T. Wei et al., "Skywork: A more open bilingual foundation model," arXiv preprint arXiv:2310.19341, 2023.
[39] K. Zhou et al., "Don't make your LLM an evaluation benchmark cheater," *arXiv preprint arXiv:2311.01964*, 2023.

[40] A. Matton et al., “On leakage of code generation evaluation datasets,” *arXiv preprint arXiv:2407.07565*, 2024.
[41] M. Habibi, “Open sourcing GPTs: Economics of open sourcing advanced AI models,” *arXiv preprint arXiv:2501.11581*, 2025.
[42] J. Manchanda et al., “The open-source advantage in large language models (LLMs),” *arXiv preprint arXiv:2412.12004*, 2024.
[43] D. G. Widder, S. West and M. Whittaker, “Open (For Business): Big Tech, concentrated power, and the political economy of open AI,” *SSRN preprint 4543807*, 2023.

전 길 원

https://orcid.org/0009-0002-0049-1711
e-mail : kilwon.jeon00@gmail.com
2020년~현 재 세종대학교 지능기전공학부 스마트기기공학전공 학사
관심분야 : LLM, AI Agent, 클라우드 컴퓨팅, 인공지능, 오픈소스 소프트웨어

한 현 준

https://orcid.org/0009-0008-5129-0067
e-mail : super010119@gmail.com
2021년~현 재 세종대학교 컴퓨터공학과 학사
관심분야 : 인공지능, AI Agent

이 강 원

https://orcid.org/0000-0002-3025-4699
e-mail : kangwon.lee@sejong.ac.kr
1992년 서울대학교 컴퓨터공학과 학사
1994년 서울대학교 컴퓨터공학과 석사
2000년 일리노이 주립대학교 전산학 박사
2000년~2007년 IBM 왓슨 연구소 research staff member
2008년~2014년 IBM 왓슨 연구소 research manager
2014년~2021년 SK 텔레콤 종합기술원 기술원장
2022년~2024년 SK온 DT부문장
2024년~현 재 세종대학교 컴퓨터공학과 교수
관심분야 : 컴퓨터 네트워크, 클라우드 컴퓨팅, 인공지능, 블록체인, 오픈소스 소프트웨어