

오픈소스 AI의 IPR 쟁점

IPR Issues on Open-source AI

류시원

전남대학교 법학전문대학원



CONTENTS

- 01 오픈소스 AI
- 02 개발 단계 쟁점
- 03 배포 단계 쟁점
- 04 사용 단계 쟁점
- 05 정리





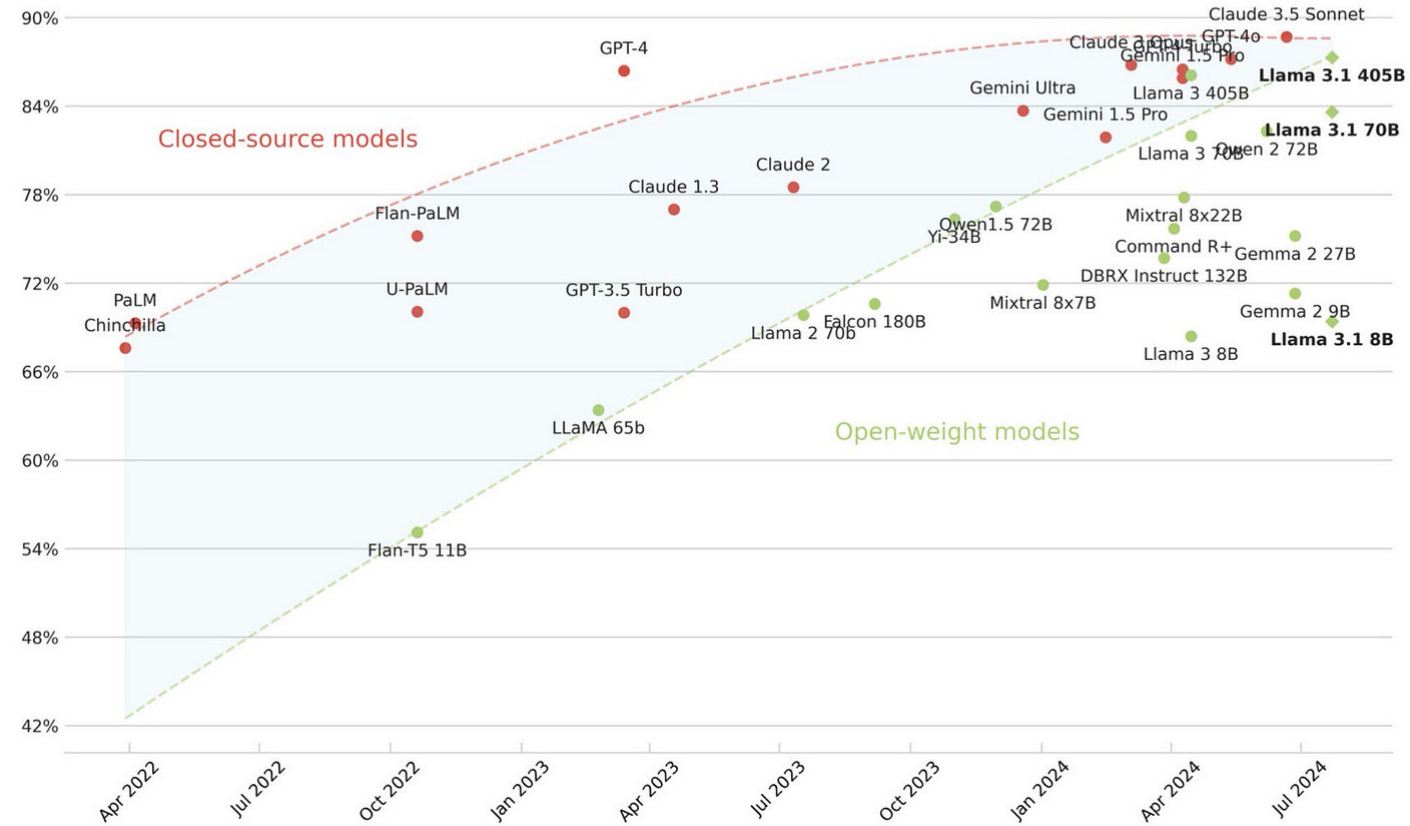
01

오픈소스 AI



@maximelabonne

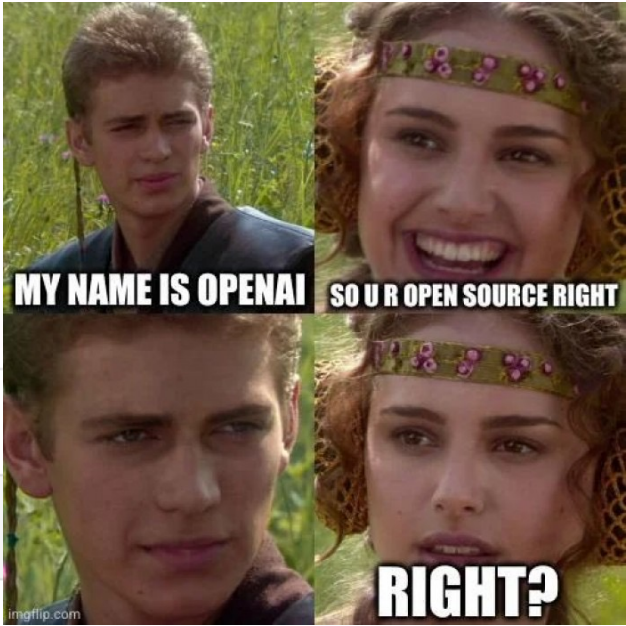
MMLU (5-shot)





오픈소스 AI

	gpt-oss-120b	gpt-oss-20b	OpenAI o3	OpenAI o4-mini
추론 및 지식				
MMLU	90	85.3	93.4	93
GPQA Diamond	80.9	74.2	77	81.4
Humanity's Last Exam	19	17.3	24.9	17.7
경쟁 수학				
AIME 2024	96.6	96	91.6	93.4
AIME 2025	97.9	98.7	88.9	92.7



(source: reddit.com)

(source: OpenAI)



오픈소스 AI

- Open Source AI Definition 1.0 (by OSI)

An Open Source AI is an AI system made available under terms and in a way that grant the freedoms¹ to:

- 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.

(source: OSI)

- 공개 요구사항

- 데이터 정보 : 시스템 학습에 사용된 데이터에 대한 충분히 상세한 정보
- 코드 : 시스템 학습 및 실행에 사용되는 소스코드 일체
- 파라미터 : 가중치, 기타 설정값들과 같은 모델 파라미터



오픈소스 AI

- Open Source AI Definition 1.0 (by OSI)

- Meets Definition?

(source: oss.kr)

Freedom to ...	GPT-o3	LLaMA 3.1	Mixtral 8x7B	Stable Diff. 3.5	DeepSeek-R1
Use	X	△	O	△	O
Study	X	△	△	△	△
Modify	X	△	O	△	O
Share	X	△	O	△	O

- Meets Requirements?

Elements	GPT-o3	LLaMA 3.1	Mixtral 8x7B	Stable Diff. 3.5	DeepSeek-R1
Data info.	X	X	X	X	X
Code	X	△	O	△	O
Parameters	X	△	O	△	O



오픈소스 AI

- Open Weight Definition (by OSA)
 - “Bridge the gap” between the closed and open source AI models by
 - **Protecting** Freedom to Use, Share
 - **NOT Protecting** Freedom to Study, Modify
 - 주요 내용
 - Must include weights and allow redistribution of weights
 - Must allow modifications and derived works
 - No discrimination against persons, groups, fields of endeavor
 - License must be attached to the weights and redistributed
 - License must not restrict other software
 - License must be technology-neutral



오픈소스 AI

• How 'OPEN' are LLMs?

Project	Availability					Documentation					Access				
	Open code	LLM data	LLM weights	RL data	RL weights	License	Code	Architecture	Prompt	Paper	Modelcard	Datasetcard	Package	API	
OpenAI GPT-4o	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	12.5
BLOOMZ	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	12.0
AmberChat	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	10.0
Open Assistant	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	9.5
OpenChat 3.5 7B	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	9.0
Pythia Chat Base 7...	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	8.5
Cerebras GPT 111...	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	8.0
RedPajama-INICITE...	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	7.5
dolly	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	7.0
Tulu V2 DPO 70B	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	6.5
MPT-30B Instruct	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	6.0
ixix	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	5.5
NeuralChat 7B	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	5.0
Vicuna 13B v1.3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	4.5
minChatGPT	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	4.0
ChatRWKV	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	3.5
BELLE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	3.0
GeLij Ultra 7B	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	2.5
Phi 3 Instruct	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	2.0
WizardLM 13B v1.2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	1.5
Airoboros L2 70B G...	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	1.0
ChatGLM 6B	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.5
Mistral 7B Instruct	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.0
WizardLM-7B	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.0

Mistral NeMo Instruct	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	5.0
Owen 1.5	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	4.5
StableVicuna-13B	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	4.0
Falcon-40B-Instruct	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	3.5
UltraLM	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	3.0
Yi 34B Chat	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	2.5
Koala 13B	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	2.0
Llama 3.1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	1.5
Mixtral 8x7B Instruct	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	1.0
Stable Beluga 2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.5
Stanford Alpaca	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.0
Falcon-180B-chat	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.0
Gemma 7B Instruct	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.0
Orca 2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.0
Command R+	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.0
LLaMA2 Chat	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.0
Nanbeige2-Chat	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.0
Llama 3 Instruct	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.0
Solar 70B	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.0
Xwin-LM	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.0
ChatGPT	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.0

(source: opening-up-chatgpt.github.io)



오픈소스 AI

- EU AI Act

- 고위험 AI, 범용 AI 모델의 의무 중 일부를 오픈소스 AI에 대해서는 면제
 - Art. 25(4) : "... This paragraph shall not apply to third parties making accessible to the public tools, services, processes, or components, other than general-purpose AI models, under a free and open-source licence."
 - Art. 53(2) : "The obligations set out in paragraph 1, points (a) and (b), shall not apply to providers of AI models that are released under a free and open-source licence that allows for the access, usage, modification, and distribution of the model, and whose parameters, including the weights, the information on the model architecture, and the information on model usage, are made publicly available. ..."

➔ '오픈소스' AI에서 documentation 부실화 위험 더 크다는 지적도 있음

([A Frankenstein-like approach: open source in the AI act – Open Future](#))



오픈소스 AI

- 소프트웨어와 AI의 차이

- 복합적 구성요소

- 코드, 아키텍처, 학습데이터, 파라미터, 문서, ...

- IP 보호의 중첩 적용

- 저작권, 데이터베이스권, 특허, 영업비밀, ...

- 학습데이터/파라미터 ≠ 코드

License	Number of HF models*	OSI recognized?	Permissive copyright?	Commercial use?
Apache 2.0	97,421	Yes	Yes	Yes
MIT	42,831	Yes	Yes	Yes
Open Rail Family	27,919	No	Yes	Yes
CreativeML – Open Rail	18,631	No	Yes	Yes
CC-BY-NC 4.0	7,081	No	Yes	No
Llama2	5,375	No	Yes	Yes, with exceptions
CC-BY-4.0	3,840	No	Yes	Yes
OpenRAIL ++	2,379	No	Yes	Yes
AFL 3.0	2,377	Yes	Yes	Yes
CC-BY-NC-sa 4.0	2,108	No	No	No
CC-BY-SA 4.0	1,547	No	No	Yes
GPL3	1,483	Yes	No	Yes

*As of 5/6/2024

(source: mena.io)



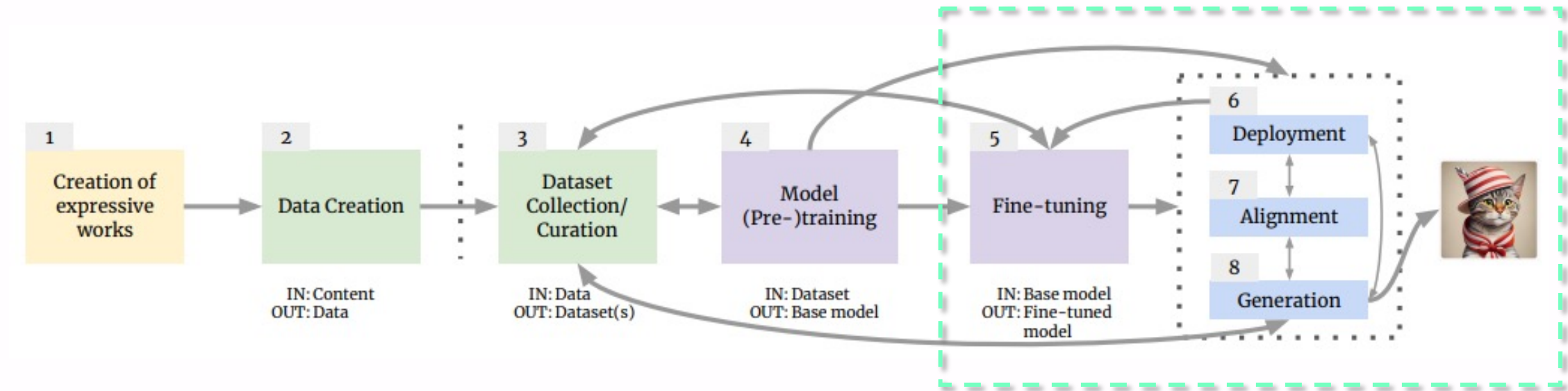
02

개발 단계 쟁점

02

개발 단계 쟁점

- AI 개발 단계



(source: Lee et al, 2023)

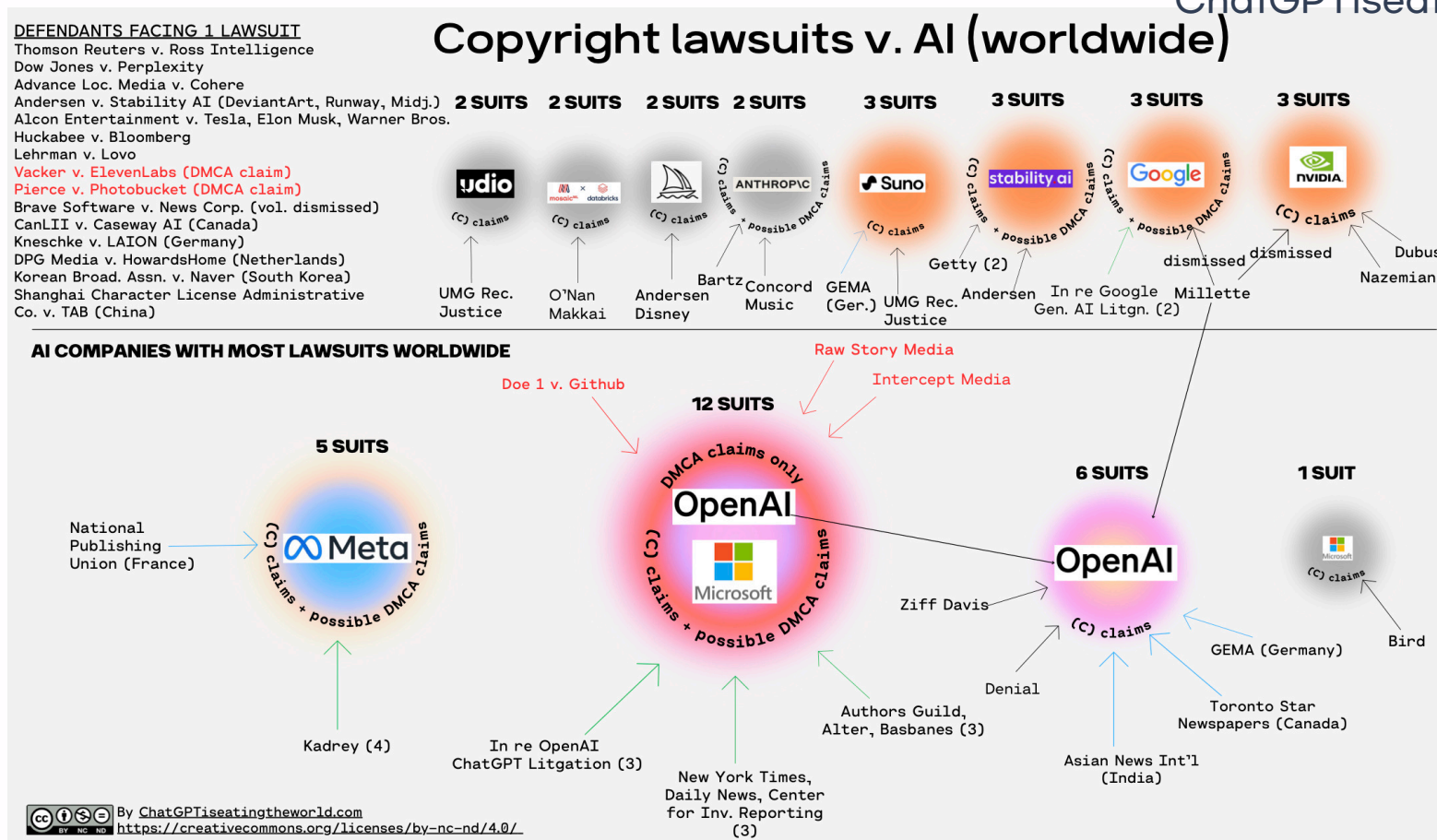


개발 단계 쟁점

- 법적 이슈: 학습데이터

- 저작권, 개인정보, 초상권/퍼블리시티권, ...

(source: ChatGPTiseatingtheworld.com)





개발 단계 쟁점

- 법적 이슈: 학습데이터

- EU CDSM Directive Arts. 3, 4 (TDM exception)

(독일 함부르크 지방법원 Kneschke v. LAION)

- Fair Use

(미국 Kadrey v. Meta Platforms; Bartz v. Anthropic PBC)

구분	판단 대상	목적과 성격: 변형성	시장 영향		공정이용 여부
엔쓰로픽	학습	○	동일 결과물	X	○
			경쟁 결과물	X	
			이용허락 시장	X	
	디지털화	○	중립		X
	해적 행위	X	○		
메타	다운로드	○	정보 역류	X	○
	+		이용허락 시장	X	
	학습		시장 희석	X	

US judge preliminarily approves \$1.5 billion Anthropic copyright settlement

By Blake Brittain

September 26, 2025 7:48 AM GMT+9 · Updated 3 hours ago



(source: 이대희, 저작권위원회 이슈리포트)



개발 단계 쟁점

- 법적 이슈: 학습데이터

- Doe 1 v. Github

Doe 1 v. Github, Inc., in the U.S. District Court for the Northern District of California, No 4:2022cv06823 (filed 11/03/22) (proposed class action). [[DOCKET](#)] [[COMPLAINT](#)] [[FIRST AMENDED COMPLAINT](#)] [computer programs] [[on appeal to Ninth Circuit](#)]

- Authors of computer programs
- Defendants' AI model: LLM ([Codex](#))
- Counts: (1) DMCA s. 1202 claim for removal of CMI, (2) breach of contract open-source license violations, (3) breach of contract by Github, (4) intentional interference with prospective economic relations Cal. common law, (5) negligent interference with prospective economic relations Cal. common law, (6) unjust enrichment, (7) unfair competition Cal. law, (8) negligence.

(source:
ChatGPTiseatingtheworld.com)



개발 단계 쟁점

- 법적 이슈: 학습데이터

- CCL 콘텐츠를 학습데이터로 사용한 경우 (Creative Commons의 입장)



Attribution (BY)

All CC licenses require attribution to the creator(s) of the licensed material. For AI model training, attribution could be a simple link to the source of the dataset used to train the model. Where retrieval-augmented generation (RAG) or other methods are available, providing attribution to the CC-licensed work tied to the particular model output with a link to the source is ideal.



ShareAlike (SA)

CC BY-SA and CC BY-NC-SA require that adaptations be shared under the same license. If AI models or outputs are based on ShareAlike content and they will be shared publicly, following the ShareAlike condition would require AI developers to use the same CC license as the original works.



NonCommercial (NC)

CC BY-NC and CC BY-NC-SA give permission for NonCommercial uses only. If AI training data includes the NonCommercial restriction, then following the NC restriction would require that all stages, from copying the data during training to sharing the trained model, must not be for commercial gain.



개발 단계 쟁점

- 법적 이슈: 오픈웨이트 모델 사용 조건
 - 이용 목적, 재배포 등 조건이 라이선스마다 다름 (e.g., Apache 2.0, MIT, OpenRAIL ...)
 - Component별 라이선스 조건이 서로 다를 수 있음 (e.g., DeepSeek)
 - Llama 3.3 Community License의 경우 (https://www.llama.com/llama3_3/license/)
 - 웹사이트, 블로그 포스트, 제품 설명서 등에 “Built with Llama” 표시
 - 재배포 시 “Llama 3.3 is licensed under the Llama 3.3 Community License, Copyright © Meta Platforms, Inc. All Rights Reserved.” 고지
 - 월간 활성 이용자수 7억 명 초과 시 Meta의 별도 라이선스 필요

cf. 다른 모델의 개선활동에 사용 금지 (Llama 3까지 있었으나, 3.1부터 삭제)



03

배포 단계 쟁점



배포 단계 쟁점

- 모델 배포에 의한 저작권 침해?

COPYRIGHT TRENDS REPORT

저작권 동향 제13호

2025. 9.

영국 Getty v Stability AI 항소심, 저작권 쟁점 변화

University of Colorado 연구원
이나라

2025년 6월, 영국 왕립 법원(Royal Courts of Justice) 내 항소법원 민사부(Court of Appeal, Civil Division)에서 열린 항소심에서는 그간 Getty가 주장하던 네 가지 청구가 바뀐 양상을 보였다.

① 저작권 침해 쟁점 변화: Getty는 Stable Diffusion 모델 자체가 영국 저작권법상 '침해 복제물 (infringing article)'에 해당한다는 쪽으로 주장을 전환했다. 즉, 해당 AI 모델이 저작권 침해의 산물이라는 점을 들어, 영국 내에서 이를 보유하거나 대중에게 배포하고 판매하는 행위 자체가 저작권 침해의 간접 침해(secondary infringement)라고 강조하였다.



배포 단계 쟁점

- 모델 배포에 의한 저작권 침해?

2. Memorization

The extent to which models retain or “memorize” training data, which would then travel with the model in subsequent distributions, was disputed by commenters. Some AI companies asserted that “[t]here is no copy of the training data — whether text, images, or other formats — present in the model itself.”¹⁰⁸ OpenAI characterized contrary arguments as based on “a common and unfortunate misperception of the technology,” and argued that model weights are just “large strings of numbers” that reflect “statistical relationship[s]” among the training tokens.¹⁰⁹

But others pointed to “numerous examples” of models generating “verbatim, near identical, or substantially similar outputs,” arguing that they can “embody the expressive works they were trained on.”¹¹⁰ News/Media Alliance stated that “regardless of the exact technical processes employed,” such behavior “has the same effect as memorization and retention.”¹¹¹

(source: USCO, Copyright and AI Report,
Part 3)



배포 단계 쟁점

- 모델 배포에 의한 저작권 침해?
 - Cooper & Grimmelmann (2025) 견해
 - 암기는 back-end 현상 / front-end 현상인 추출(extraction), 역류(regurgitation)와 구별됨
 - 암기가 기술적 실재라면, 학습데이터를 암기하고 있는 모델은 미디어 파일과 동등한 사본 취급 가능
 - 인간의 지각, 동시대 기술에 의한 완벽한 재현가능성은 사본성과 관련하여 요구되지 않음 (ancient papyrus scroll)
 - 미국 저작권법은 'copies'를 장래 개발될 기술까지 고려하여 개방적으로 정의함
 - 모델은 실제로 비결정론적으로 작동하지 않음 (pseudo-random sequence)
 - 실제 소송에서도 모델을 '콜라주 기계', '압축된 사본' 등으로 주장하는 경우가 있음 (Sarah Anderson v. Stability AI)



배포 단계 쟁점

- Disclaimer of Warranty vs. Implied Warranty

- GPL 등 오픈소스 라이선스는 대개 포괄적 보증 부인(disclaimer) 조항 포함함

15. Disclaimer of Warranty.

THERE IS NO WARRANTY FOR THE PROGRAM, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE PROGRAM "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE PROGRAM IS WITH YOU. SHOULD THE PROGRAM PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION.

[GPL 3.0]

[CISG Art. 42]

1) The seller must deliver goods which are free from any right or claim of a third party based on industrial property or other intellectual property, of which at the time of the conclusion of the contract the seller knew or could not have been unaware, provided that the right or claim is based on industrial property or other intellectual property:

- (a) under the law of the State where the goods will be resold or otherwise used, if it was contemplated by the parties at the time of the conclusion of the contract that the goods would be resold or otherwise used in that State; or
- (b) in any other case, under the law of the State where the buyer has his place of business.

(2) The obligation of the seller under the preceding paragraph does not extend to cases where:

- (a) at the time of the conclusion of the contract the buyer knew or could not have been unaware of the right or claim; or
- (b) the right or claim results from the seller's compliance with technical drawings, designs, formulae or other such specifications furnished by the buyer.



04

사용 단계 쟁점



사용 단계 쟁점

- 이용자의 책임 vs. 개발자의 책임
 - AI 산출물에 대한 이용자의 책임과 권리는 동전의 양면
 - 이용자에게 책임 부담시키는 경우 많음
 - OpenAI 이용약관

귀하의 콘텐츠. 귀하는 본 서비스에 인풋(“인풋”)을 제공할 수 있고 이를 바탕으로 본 서비스로부터 아웃풋(“아웃풋”)을 수령할 수 있습니다. 인풋과 아웃풋을 총칭하여 “콘텐츠”라고 합니다. 귀하의 콘텐츠가 관련 법률 또는 본 약관을 위반하지 않게 하는 등 콘텐츠에 대한 책임은 귀하에게 있습니다. 귀하는 본 서비스에 인풋을 제공하는데 필요한 모든 권리, 면허 및 허가를 보유하고 있음을 진술 보장합니다.

콘텐츠 소유권. 관련 법률상 허용되는 한도 내에서, 귀하와 OpenAI 중 귀하가 (a) 인풋에 대한 귀하의 소유권을 유지하고 (b) 아웃풋을 소유합니다. 당사는 아웃풋에 대한 당사의 권리, 소유권 및 이권이 있다면 그 모두 이로써 귀하에게 양도합니다.



사용 단계 쟁점

- 이용자의 책임 vs. 개발자의 책임

- 생성 AI 이용자의 직접책임 : 침해유발적 프롬프트 입력 등
- 생성 AI 개발자의 간접책임
 - 기여책임(contributory liability) : 침해행위 인식 + 침해행위 유도, 야기 or 실질적 기여
 - 대위책임(vicarious liability) : 침해행위에 대한 실질적 감독/통제 권한/능력 + 침해행위로부터 명백하고 직접적인 재정적 이익 획득

➔ 암기 현상 억제 위한 filtering 등 조치, 유사 콘텐츠 생성능력 강화하는 fine-tuning 자제, ...



05

정리



정리

- '오픈소스 AI'에 관한 통일된 정의 부재
 - 소프트웨어와 AI 모델의 차이
- 개발 단계
 - 저작권 소송 증가, 라이선싱 deal 확산 경향
 - 적법한 경로로 데이터 수집, AI 학습 이용에 관한 제한 있는지 확인
 - 오픈웨이트 모델 라이선스 조건 유의
- 배포 단계 : 생성 AI 모델의 '암기' 현상에 관한 논의 증대
- 사용 단계 : 생성 AI 이용자 책임이 개발자 책임으로 유입되지 않도록 조치

감사합니다

