

2023 March Meeting of OpenChain-KWG

Legal Issues of AI Technologies

– Case Study: Getty Images v. Stability AI –

March 28, 2023

Jeongsook Park

Open Source Center, ETRI

Contents

- Introduction
- Legal Issues with Case Study
 - Getty Images v. Stability AI (Case 1:23-cv-00135-UNA)
 - Andersen et al v. Stability AI Ltd. et al (Case 3:23-cv-00201)
 - Legal Issues on AI Technology
- Open Source Compliance on AI
- Concluding Remarks

References

- [1] 그림 인공지능/논란 및 사건 사고, 나무위키
- [2] Getty Images(US), Inc. v. Stability AI, Inc., Docket No. 1:23-cv-00135-UNA
- [3] Andersen et al v. Stability AI Ltd. et al, Docket No. 3:23-cv-00201
- [4] 화우, “그림 생성 초거대 AI 모델에 대한 법적 이슈 – 최근 초거대 AI 모델에 대하여 제기된 미국 소송에서의 쟁점을 중심으로-”, 2023.3.3, <https://m.lawtimes.co.kr/Content/Article?serial=185810>.
- [5] 이유정, “인공지능 기술과 법“, 원내 전문가 초청 세미나, 2022.6.21.
- [6] 조연하, “인공지능의 콘텐츠 창작에서 저작물 이용의 법적 쟁점에 관한 고찰“, 사이버커뮤니케이션학보, 통권 제39권2호, 2022.6.
- [7] 김경애, “인공지능과 오픈소스 라이선스“, 공유저작물 및 오픈소스SW 라이선스 컨퍼런스, 2022.8.26.
- [8] 이철남, “인공지능 학습용 데이터(셋)와 공유저작물 관련 쟁점“, 공유저작물 및 오픈소스SW 라이선스 컨퍼런스, 2022.8.26.
- [9] G. Rajbahadur, et al, “Can I use this publicly available dataset to build commercial AI software? – A Case Study on Publicly Available Image Datasets”, arXiv:2111.02374v5 [cs.LG], 11 Apr 2022, <https://arxiv.org/abs/2111.02374>
- [10] 장학성, “상용 AI 서비스에 공개 Dataset을 사용해도 되나요?”, https://sktelecom.github.io/blog/2022/20220214_ai_dataset_copyright/
- [11] Bradley M.Kuhn, “If Software is My Copilot, When Programmed My Software?”, <https://sfconservancy.org/blog/2022/feb/03/github-copilot-copyleft-gpl/>, 2022.2.3.
- [12] OpenDatalogy Project: <https://github.com/OpenDatalogy/OpenDatalogy>
- [13] Community Data License Agreement, <https://cdla.dev>

1. Introduction

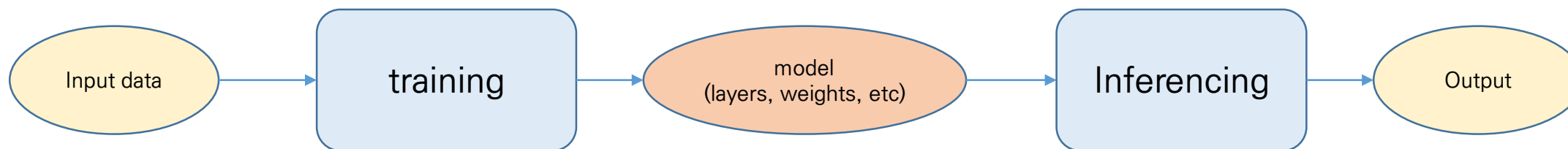
Introduction

- AI 기술은 매우 빠르게 발전 중

- 지적 존재와 관련되는 작업을 수행하는 디지털 컴퓨터 또는 컴퓨터 제어 로봇 능력

- 교육을 위해 많은 입력 데이터 필요, 저작권 침해 가능


- 학습 결과도 저작권 침해 대상일 수 있음



Motivation for Presentation

Wonjae Park 오전 11:29
요새 핫한 Generative AI 관련 소식입니다.
Getty Images가 미국에서도 소송을 진행하네요.
<https://www.theverge.com/2023/2/6/23587393/ai-art-copyright-lawsuit-getty-images-stable-diffusion>


▼ **The Verge**
Getty Images sues AI art generator Stable Diffusion in the US for copyright infringement
Getty says Stability AI stole 12 million images without permission
2월 7일 (208kB) ▼



3 2개의 댓글 23일 전 마지막 댓글

Wonjae Park 오전 11:30
원래는 Stability AI의 본사가 있는 영국에서 소송 진행했다 했었는데, Stable Diffusion Litigation (<https://stablediffusionlitigation.com/>)에 자국/영감? 을 받았는지 미국에서 추가로 소송 진행한다는 소식입니다

stablediffusionlitigation.com
Stable Diffusion litigation · Joseph Saveri Law Firm & Matthew Butterick
Stable Diffusion litigation (667kB) ▼











3

OpenChain-KWG slack

어마어마한 기술이 또 나왔네요.
게시자 [redacted] ETRI Bulletin board
게시일 2023-02-14 11:28
게시 조회수 645

이미지 생성쪽에 요즘 핫한 stable diffusion 관련된 기술인데요..
학습된 기존 모델(ckpt 파일 등)에 객체나 스타일을 추가할 수 있는 text inversion, dreambooth, lora 등등에 이어... 또다른 add-on 기술이 하나 나왔는데... 이번 것은 좀 강력합니다.
<https://github.com/lllyasviel/ControlNet>

Input (User Scribble)	Default	Automatic Prompt	User Prompt
			
			

"a turtle in river" "a masterpiece of cartoon-style turtle illustration"

Decided to dig!



<input type="checkbox"/> No	제목	부서	게시자	게시일	첨부	조회수
<input type="checkbox"/> 301454	나노코리아2023 심포지엄 (07/05~07/07)	광융합부품연구실	김기수	2023-02-14 20:15		532
<input type="checkbox"/> 301453	와글와글에 글 올려주신 Stable diffusion에 관한 소송이 발생했다는 소식입니다.	[8] 오픈소스센터	박정숙	2023-02-14 20:02		2007
<input type="checkbox"/> 301452	한국핀테크학회 2023 동계학술대회 (2/15~16) 개최	지능정보표준연구실	이원석	2023-02-14 15:00		633
<input type="checkbox"/> 301451	[지식정보서비스실] 2023년도 상용 이미지·아이콘 콘텐츠 제공 서비스 안내 (게티이미지뱅크)	지식정보서비스실	이수화	2023-02-14 11:28		645

ETRI Bulletin board

What to Expect from Litigation Case Studies

• 경영진 입장

- 법적 이슈 없도록 잘 대응하고 조심조심하자!

• 개발자 입장

- 그래서 Stable Diffusion은 써도 되나요?
- 법적 이슈가 없으려면 어떤 걸 조심하면 되나요?

• OSPO 입장

- AI는 오픈소스 컴플라이언스 측면에서 어떤 이슈가 있을 수 있나?
- 오픈소스 리스크를 없앨 수 있는 방법은?
- AI 관련 법은 어떻게 나아갈 것인가?

Purpose of Presentation

- AI 소송 사례 분석을 통한 관련 법률 이슈 이해하기
- AI에 대한 오픈소스 컴플라이언스 대응 방안 이해하기
- 기관별 대응 방안 이해하기



2. Legal Issues with Case Study

– Getty Images v. Stable Diffusion–

Painting AI Issues [1]

No.	유형	설명
1	대회 및 전시회 참여 논란	• MidJourney 예술대회 우승 사건, 암스테르담 국립미술관 아마추어 전시회 AI 그림 참여 논란
2	인간 작가 위장 및 사기	• 인간 커미션 위장, A 사용자의 허위 포트폴리오 사건, 텀블벅 프로젝트의 AI 사용 의혹, 서울코믹우프우드 AI 굿즈판매 논란
3	상업적 이용 논란	• 랜덤채팅의 그녀! AI 그림 논란, 승리의 여신: 니케 AI 일러스트 사용 의혹 및 사용 부인 논란, 백야극장 축전 AI 그림 논란
4	무단 학습 논란	• 그림학습에 대한 대중의 인식 오류, 김정기 그림 AI 학습 및 고인모독 논란, 트위치 스트리밍 그림 학습 사건, 라스트 오리진 일러스트레이터 AI 학습 사건, 일러스트 작가 화풍 AI 학습 후 인간작가 행세, 미드저니의 미키마우스 생성 논란, AI 학습 및 어그로성 댓글 사건 등
5	학습모델 도용 및 사기	• 텀블벅 AI 그림체 판매 및 사기 논란, Civitai 학습 모델 도용 사이트 등장
6	소송 관련	• 이미지 생성 AI 개발사 저작권 소송, 미국 일리노이 주 무단 개인정보 학습 집단 소송
7	인간작가를 향한 사이버불링	• AI 그림 플랫폼 이용자의 악성 어그로. 애니메이터 Tarrow 조롱 사건, 무분별한 AI 사용 의혹 제기
8	Img2img 기능 오용 및 도용	• Img2img 트롤링 사건, AI 사용자 간의 그림 img2img 도용사건, 웹소설 표지 일러스트 img2img 사건, 대규모 img2img 사건, @U_MASU 유저의 img2img 이용 사건, 일본 무명 버츄얼 유튜버 무단 상표 등록 사건
9	기타	• AI 그림 도용사건, AI 개발자 인터뷰 날조 논란, stable diffusion 선정적 그림 생성 및 화풍학습 너프에 대한 반발 , 텐센트 QQ 소세계 서비스의 윤리적 논란 및 의혹, 랜사업 서비스의 윤리적 논란, 중국의 AI 콘텐츠 검열 등
10	딥페이크	• 빌리 아일리시 딥페이크 포르노 사건, 프랑스 연금 개혁 시위 페이크 사진 논란, 튀르키예 지진 피해자 후원 사기, 트위치 스트리머 QTCinderella의 딥페이크 포르노 피해 및 사용 중지 호소 사건 등

Getty Images v. Stability AI

- **Stability AI의 Stable Diffusion**

- 독일 뮌헨대학교의 "잠재 확산 모델을 이용한 고해상도 이미지 합성 연구" 논문 기반의 text-to-image 생성 모델
- Stability AI와 Runway ML 등의 지원 받아 개발된 딥러닝 인공지능 모델
 - ✓ Stability AI는 방글라데시계 영국인 Emad Mostaque가 사비로 만든 회사
- 기존의 text-to-image 모델들과 달리 컴퓨터 리소스를 줄여 4GB 이하의 VRAM을 가진 컴퓨터에서 운영 가능
- 소스코드 공개, 수익창출 사용자 인터페이스 Dream Studio 제공

- **Try**

- <https://www.youtube.com/watch?v=zNJ1kXLUeQs>

- **Source code**

- <https://github.com/Stability-AI/stablediffusion/blob/main/LICENSE-MODEL>
- <https://compvis.github.io/taming-transformers/>
- Stable AI에서 오픈소스 머신러닝 전용 라이선스 새로 생성
 - ✓ Stable Diffusion으로 서비스할 경우 반드시 명시하고 파인튜닝한 모델들은 라이선스에 명시된 제한 용도에는 사용 금지

Stable Diffusion

• 이력

- V1.0('22.8.22.)
 - ✓ '22.10.7 NovelAI의 제로데이 해킹으로 소스코드 공개 * NovelAI의 이미지 제너레이터는 Stable Diffusion 기반임
 - ❖ 출처: <https://namu.wiki/w/NovelAI/%EC%9D%B4%EB%AF%B8%EC%A7%80%20%EC%A0%9C%EB%84%A4%EB%A0%88%EC%9D%B4%ED%84%B0>
 - ✓ Web UI의 저작권, 초상권 침해 및 인간 기본권 침해 등의 항의와 논란 이어짐 ('23.1.13, Case 3:23-cv-00201)
- V2.0('22.11.24.)
 - ✓ 학습 데이터 해상도가 512x512에서 768x768으로 높아졌고 OpenCLIP이라는 독자적인 텍스트 인코더 도입
 - ✓ 학습 데이터에서 성인물을 제거했고, 특정 창작자의 화풍을 모방하는 기능이 제대로 작동하지 않음. 화풍 문제는 새로운 텍스트 인코더인 OpenCLIP의 문 제점으로 밝혀짐
- V2.1
 - ✓ 신체 비율에 중요한 이미지는 제거하지 않은 채로 학습
- Riffusion('22.12.13.)
 - ✓ 음악도 생성 가능
 - ✓ 여러 음악의 스펙트로그램의 이미지를 학습시킨 뒤 AI가 생성한 스펙트로그램 이미지를 다시 음악으로 변환하는 원리
- V3.0('22.12.15.)
 - ✓ 원작자가 데이터셋에 있는 자신의 작품을 삭제할 수 있게 함
 - ❖ 헤브아이빈트레이너(HaveIBeenTrained)'에 접속해 텍스트나 파일을 입력: <https://haveibeenentrained.com/>
- 모델의 학습과정에서 저작권자의 동의없이 인터넷에 공개된 미술저작물이 수집된 사례가 발견되면서 소송 촉발. 상업적 목적의 활용도 요인이 됨
 - ✓ 저작권자들이 Stability AI, 미드저니, 디비언트아트를 상대로 캘리포니아주 북부지방법원에 집단 소송 제기('23.1.13, Case 3:23-cv-00201)
 - ✓ Getty Images가 영국('23.1.17.)뿐만 아니라(Stability AI 본사 위치) 미국 델라웨어 지방법원에 소송 제기('23.2.3, Case 1-23-cv-00135-UNA)

Claim Review – Case 1:23-cv-00135-UNA [2]

• 요약

➤ Getty Images

- ✓ 세계 최대 규모의 시각 자료 보유 업체로서, 디지털 미술 업계에서 사진, 그림을 공급하는 주요 공급자 중 하나임.
- ✓ 보유한 대다수 사진, 그림 자료에 대해 저작권을 보유하거나 독점 이용권이 있고, 자료들을 사용자에게 유료 라이선스를 구매해 이용하도록 하는 방식으로 사업함
- ✓ 각 사진과 그림 자료에는 제목, 캡션, URL 등 각종 메타데이터 정보가 연계되어 있고, 이를 DB화하고 검색 시스템을 구축하는데 상당한 시간과 노력을 투입하였음
- ✓ Getty Images 이용 약관에 라이선스 없이는 웹사이트 내 어떠한 자료도 다운로드, 복사, 재전송을 받을 수 없고, 데이터 마이닝, 로봇 등의 데이터 수집, 추출 방법에 사용될 수 없다고 명시되어 있어, 해당 자료들은 AI 학습 데이터로 사용될 수 없음

➤ Stability AI

- ✓ Stable Diffusion 모델은 독일 업체 'LAION'으로부터 제공 받은 5억 개에 이르는 학습 데이터로 학습이 이뤄지고, 해당 학습 데이터에 포함된 Getty Images의 자료는 적어도 1,200만 개에 이르는데, Getty Images는 이에 대해 허락한 바 없음
 - ✓ Stability AI는 Stable Diffusion 모델의 위 학습 과정에 관여하였고, 이후 OSS로 Stable Diffusion 모델을 공개한 당사자임. Stable Diffusion을 최초 공개한 이후, Stability AI를 포함한 기존 개발자들은 Stable Diffusion 모델을 개량하여 상위 버전의 모델도 계속하여 배포해 왔음
 - ✓ Getty Images 가 구축한 DB는 사진, 그림 자료와 그에 대응되는 정확한 문자 정보를 갖추고 있기 때문에 AI 모델 학습 데이터로 안정맞춤이고, 이는 Stable Diffusion 모델의 성능 향상에 크게 기여하였음
 - ✓ 실제로 Stable Diffusion 모델로 생성된 출력물에는 '게티 이미지와 유사한 그림'이나 'Getty Images의 상표, 워터마크 형체가 남아 있는 그림'이 발견되었음
- 최근 논문에서도 Stable Diffusion 모델이 학습 데이터로 사용된 사진, 그림을 거의 그대로 혹은 극히 유사한 형태로 재생산하는 사례를 보고한 바 있음
- Stability AI는 Stable Diffusion 모델에 기반하여 유료 그림 생성 서비스 플랫폼 Dream Studio을 운영하여 수익을 창출하고 있고, 이는 Getty Images의 서비스 수요를 강탈하는 효과로 이어짐
- Stability AI는 Getty Images의 동의없이 Stable Diffusion 모델을 OSS로 배포하여 제3자가 이를 접근하고, 개량하도록 허용함

Relevant Evidences (1)

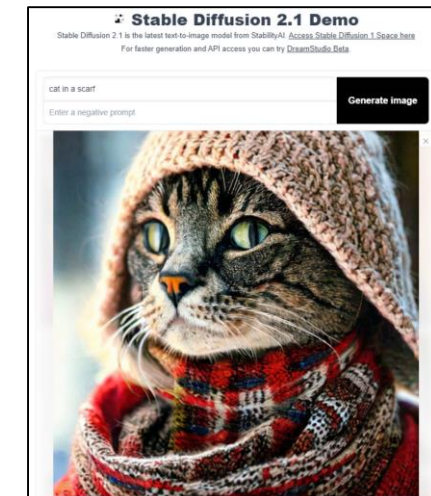
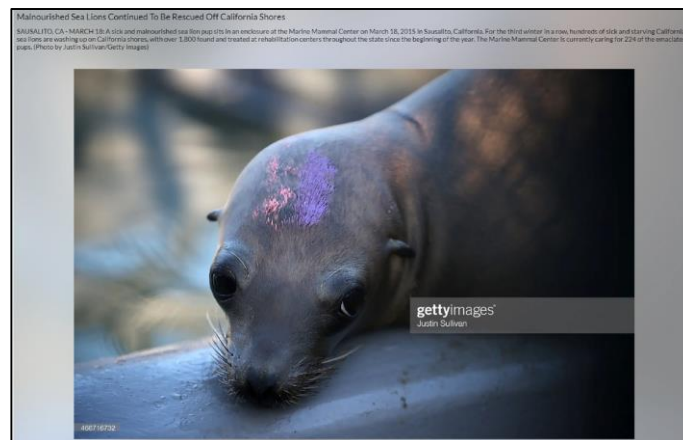
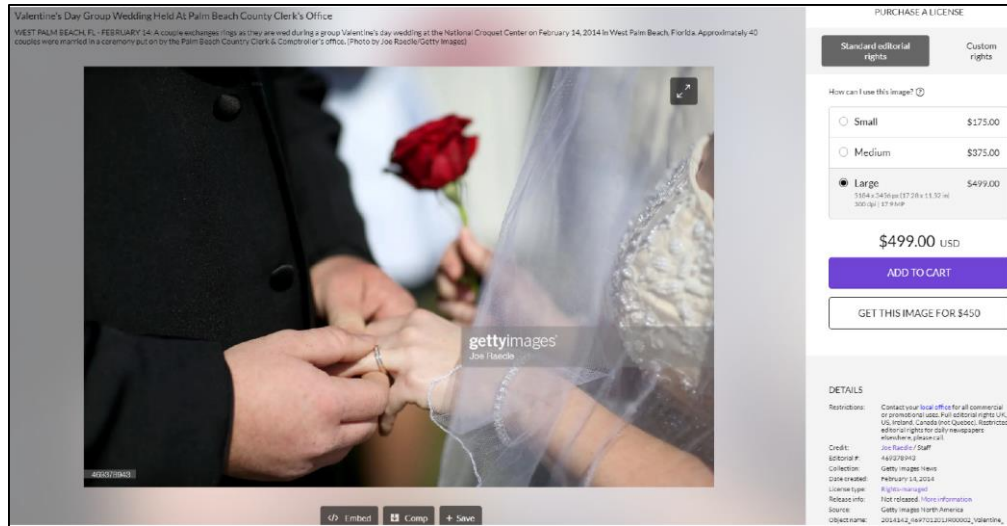
- <https://mbdrive.gettyimageskorea.com/guide/cost/license>

제한사항 – 추가 라이선스를 구매하지 않은 경우

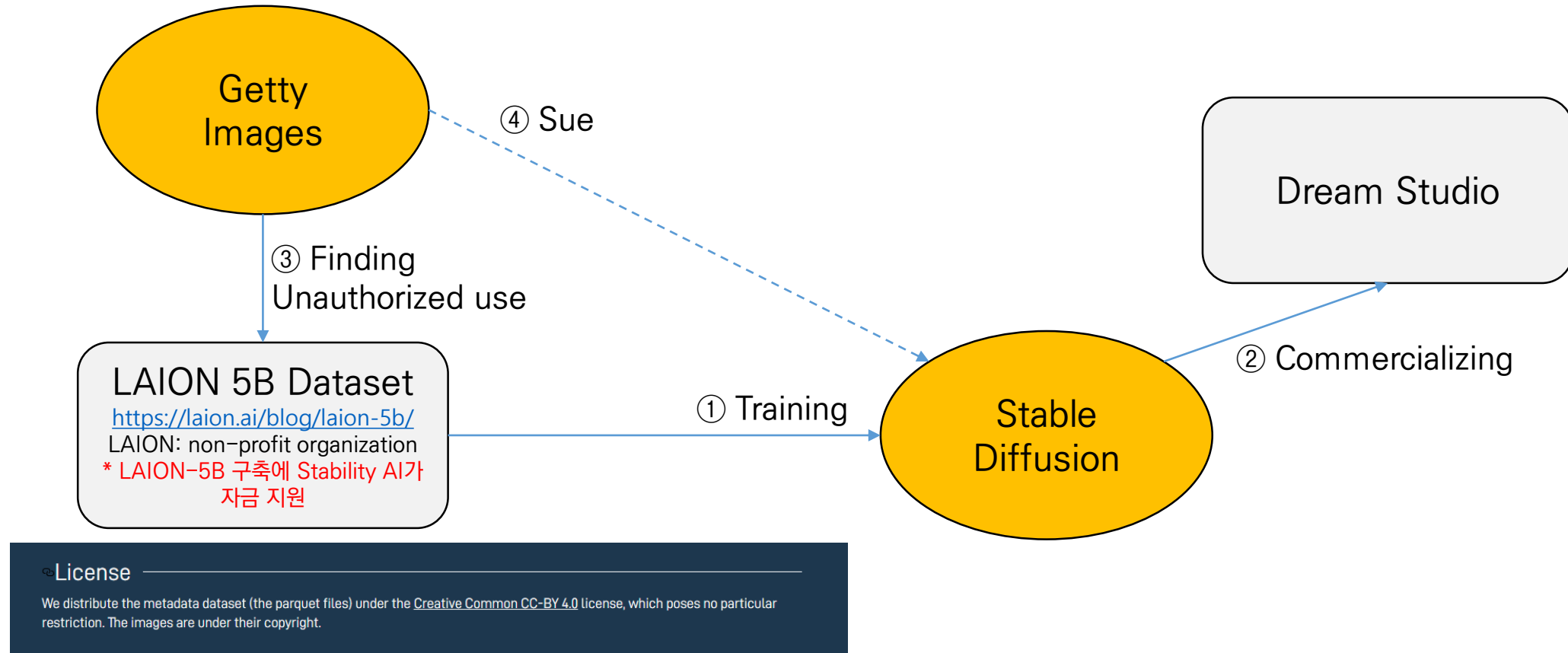
게티이미지코리아의 사전 서면 허가과 추가 라이선스 비용 지불하지 않는다면, 다음 사항은 허가되지 않는다.

- h. “주문형” 상품 금지. 게티이미지코리아와 사전에 서면 합의하지 않았다면, “주문형” 상품과 관련하여 콘텐츠를 사용할 수 없다. (예: 제 3자가 라이선스 이미지를 선택하여 특별주문제작 방식으로 맞춤제작을 하는 상품 등과 같이, 엽서, 머그컵, 티셔츠, 캘린더, 포스터, 휴대전화의 스크린세이버나 배경화면, 혹은 유사 상품들(zazzle.com, cafepress.com과 같은 상품의 주문 제작 대행 웹사이트를 통한 판매 포함))
- i. 전자 템플릿에 사용 금지. 게티이미지코리아와 사전에 서면 합의하지 않았다면, 재판매나 기타 배포를 목적으로 한 전자 및 디지털 템플릿에 콘텐츠를 사용할 수 없다. (예: 웹사이트 템플릿, 비즈니스 카드 템플릿, 전자 연하장 템플릿, 브로슈어 디자인 템플릿)
- j. 상표 또는 로고에 사용 금지. 게티이미지코리아와 사전에 서면 합의하지 않았다면, 상표, 디자인 마크, 상품명, 상호명, 서비스 마크, 로고에 일부 또는 전체의 콘텐츠를 사용할 수 없다. 추가적으로, 귀하는 어느 법령에서도 그러한 콘텐츠의 일부 또는 전체를 상표로 등록하는 것과 상표등록, 사전 사용, 영업권에 의존하여 제3자가 해당 콘텐츠 또는 유사한 콘텐츠를 사용하는 것을 막을 자격을 갖지 않는다. (당사, 당사의 고객 또는 해당 콘텐츠의 저작권자를 포함)
- k. 머신 러닝, AI, 또는생체 인증 기술 사용 금지. 게티이미지코리아와 사전에 서면 합의하지 않았다면, 콘텐츠(캡션 정보, 키워드, 연관 메타데이터 포함)를 머신 러닝, 인공 지능, 사람을 식별하기 위한 기술을 목적으로 하는 경우, 사용이 불가능하다. 추가적으로, 게티이미지코리아는 ‘초상권이 해결된 콘텐츠’에 대해서, 그러한 사용에 대한 승인을 획득하였음을 명시하거나 보증하지 않는다.
- l. 메타데이터 이용 금지. 게티이미지코리아와 사전에 서면 합의하지 않았다면, 콘텐츠와 별도로, 캡션 정보, 키워드, 관련 텍스트 등의 콘텐츠 연관 메타데이터를 사용할 수 없다. 또한 어떠한 제 3자도 그러한 콘텐츠 관련 정보(를) 접근 및 사용하도록 할 수 없다.
- m. 전자 제품 화면에 사용 금지. 커스텀 라이선스 없이 콘텐츠를 휴대폰, 태블릿, TV, 냉장고 등의 전자디스플레이의 배경화면으로 사용할 수 없다.
- n. 게티이미지코리아와 사전에 서면 합의하지 않은 경우, 이용자는 콘텐츠의 어떠한 항목도 판매를 목적으로 하는 변경 불가능한 디지털자산(예: NFT(대체불가능한토큰))과 연관지어 사용할 수 없다.

Relevant Evidences (2)



Factual Relationship



Claims by Getty Images

No.	적용 조항	위반사항	내용
1	17 U.S.C. § 101 <i>et seq.</i>	저작권 침해	<ul style="list-style-type: none"> Getty Images의 저작물을 승인없이 복제하고 파생 작업 생성하여 Getty Images의 저작권을 고의로 침해함
2	17 U.S.C. § 1202(a) (DMCA)	허위 저작권 관리 정보 제공	<ul style="list-style-type: none"> 고의로 Getty Images의 워터마크 수정 버전을 적용
3	17 U.S.C. § 1202(b) (DMCA)	저작권 관리 정보의 제거 및 변경	<ul style="list-style-type: none"> Getty Images의 워터마크/메타데이터를 의도적으로 제거/변경
4	Section 32 of the Lanham Act, 15 U.S.C. § 1114(1)	상표권 침해	<ul style="list-style-type: none"> Getty Images 마크 무단으로 사용으로 대중에게 혼돈 초래
5	Section 43(a) of the Lanham Act, 15 U.S.C. § 1125(a)	불공정 경쟁	<ul style="list-style-type: none"> Getty Images 마크 무단 사용으로 Getty Images의 명성과 선의 훼손
6	Section 43(c) of the Lanham Act, 15 U.S.C. § 1125(c), Section 3313 of the Delaware Trademark Act	상표 희석	<ul style="list-style-type: none"> Stability AI가 Getty Images 마크를 저품질 또는 기괴한 이미지에 사용하면 흐릿하거나 변색되어 마크의 품질이 희석됨
7	Delaware's Uniform Deceptive Trade Practices Act	델라웨어 기만 거래 관행법	<ul style="list-style-type: none"> 생성된 합성 이미지에 Getty Images 마크를 무단으로 사용

※ **상표 희석(Trademark dilution)**: 유명 상표와 동일 또는 유사한 상표를 사용함으로써 유명 상표의 제품 또는 서비스의 식별력을 떨어뜨리는 것

※ Lanham Act: 1946년에 제정된 미국 연방 상표법

※ DMCA(Digital Millennium Copyright Act): 1998년에 제정된 저작권법

[Tip] Legal Terms

• 17 U.S.C. § 101 et seq.

- U.S.C. : U.S. Code, 미국 법전(책)
 - ✓ 미국의 일반법과 영구법을 주제별로 체계화한 것으로, 번호가 앞에 옴
- § 101: 조항 의미
- Et seq.: 이하 참조
- 17 U.S.C. § 101 et seq.
 - ✓ 미국 법전 17번(저작권법)의 101조(저작권 정의) 이하 참조
- 17 U.S.C. § 1201~1205: DMCA(Digital Millennium Copyright Act)
 - ✓ 1996년 채택된 ‘WIPO 저작권조약(WCT)’과 ‘WIPO 실연 및 음반조약(WPPT)’의 이행, 그리고 디지털 시대의 새로운 저작권보호를 모색하기 위해 1998년 10월에 제정
 - ✓ 1) 온라인 저작권을 강화하고 이를 방해하는 기술개발을 불법으로 규정, 2) 저작물에 접근을 통제하는 DRM 기술의 우회를 목적으로 하는 기술, 장치 또는 서비스의 제공이나 배포 금지, 3) 저작권 침해가 실질적으로 이루어졌는지 여부를 불문하고 접근금지를 우회하는 행위를 처벌, 4) 인터넷에서 저작권 침해에 대한 벌칙 강화
- 참고: <https://www.govinfo.gov/>



• Cal. Civ. Code § 3344

- California Civil Code(캘리포니아 주 민법) 제 3344조

• Cal. Bus. & Prof. Code (California Business and Professions Code, 캘리포니아주 사업 및 직업법)

• Declaratory Relief (선언적 구제)

- 미국의 민사소송법 내 절차로 판사가 민사소송에서 당사자들의 권리, 의무, 책임 등을 선언하는 것. 보통의 경우 법원은 실제적 손해가 발생할 때까지 기다림
- 어떤 이행명령이나 조치, 배상을 명령하지는 않음. 법원은 선언적 판결을 내리는 것을 꺼려하나 헌법관련 중요 법리적인 문제의 경우 혼란을 막기 위해 선언적 판결을 내림
- 사례: 특허소송의 경우, 침해혐의자가 특허권자를 상대로 비침해, 무효, 권리 불행사의 확인을 구하는 소 또는 반소
 - ✓ 주로 침해혐의자가 유리한 관할확보를 위한 수단으로 사용

• Statutory Law(법정법, 성문법)

- 입법 기관에 의해 제정, 공포되어 문서화된 법

- 요약

- 
- ETRI 오픈소스센터
Open Source Center

Claims by Andersen et al

No.	적용 조항	위반사항	내용
1	17 U.S.C. §§ 106, <i>et seq.</i>	직접적인 저작권 침해	<ul style="list-style-type: none"> 피고의 액세스 권한은 저작권 소유자인 원고의 허락을 받지 못함
2	17 U.S.C. §§ 106, <i>et seq.</i>	대리 저작권 침해	<ul style="list-style-type: none"> 특정 아티스트의 이름으로 가짜 그림을 만들고 판매
3	DIGITAL MILLENNIUM COPYRIGHT ACT 17 U.S.C. §§ 1201–1205	디지털 밀레니엄 저작권 위반	<ul style="list-style-type: none"> 피고는 CMI를 제거하고 원고에게 승인을 받지 않음
4	STATUTORY RIGHT of PUBLICITY <i>Cal. Civ. Code § 3344</i>	캘리포니아 민법 퍼블리시티권 위반 (초상권)	<ul style="list-style-type: none"> 피고는 AI 제품에 원고의 이름을 동의없이 사용
5	COMMON LAW RIGHT of PUBLICITY Common Law	관습법 퍼블리시티권 위반	<ul style="list-style-type: none"> 피고의 상업적 이점을 강화하기 위해 고의로 원고의 이름과 신원을 사용
6	15 U.S.C. § 1125; <i>Cal. Bus. & Prof. Code</i> §§ 17200, <i>et seq.</i>; and Common Law	불공정 경쟁	<ul style="list-style-type: none"> 피고는 불법적인 비즈니스 관행에 참여(저작권 위반, 원고 권리 위반)
7	DECLARATORY RELIEF 28 U.S.C. § 2201(a) and <i>Cal. Code Civ. Proc. § 1060</i>	선언적 구제	

Summary of Complaints [4]

- Stable Diffusion 모델 학습에 원저작자의 허락을 받지 않음
- 모델 학습 결과를 상업적 사용에 활용함
 - Stability AI, MidJourney, DeviantArt
 - ✓ Stability AI의 Dream Studio
 - ✓ DeviantArt의 DreamUp
 - ✓ Midjourney의 Web App
 - 특정 작가를 넣으면(예: “in style of 하태임”) 해당 작가의 화풍으로 그림(fake)을 출력하여 이를 판매함
- Stable Diffusion은 오픈소스로 공개
 - 개인이 fine tuning할 수 있음
- 저작권, 상표권, 퍼블리시티권, 부정경쟁방지법 등의 쟁점



▲ 하태임 작가

AI Litigation Cases

No.	Name	Plaintiff	Defendant	Issue	Reason
1	OpenAI 소송	오픈소스 개발자	OpenAI의 Copilot	코딩 복제 저작권 침해	공개한 원데이터의 오픈소스 라이선스 의무사항 미준수 - 소송 제기 후 검증 작업 중(white paper) - MS에 불리한 결과 나옴
2	생성 AI 소송	Getty Images	Stability AI	이미지 복제 저작권 침해	Getty Images 라이선스 없이 이미지 무단 복제
3		Andersen et al	Stability AI, MidJourney, DeviantArt	이미지 복제 저작권 침해 및 퍼블리시티권 침해	원작자의 허락없이 이미지 무단 복제 및 상업적 활용
4	이루다 소송	서비스 이용자	스캐터랩의 AI 챗봇 이루다	개인정보 침해	사적인 SNS 대화를 이루다의 학습데이터로 사용
5	Stock Fish 소송	Stockfish	Chessbase	오픈소스 라이선스 미준수	DL 모델의 weigh값을 derivative works로 볼 것인가가 쟁점
6	다부스(미국 출신)	다부스	특허청	특허발명자가 AI임	발명자가 AI여서 출원 불인정(한국, 미국, 유럽, 영국, 독일, 호주) 우리나라는 저작권자, 특허권자 모두 사람만 대상

• 저작권자 인정 범위

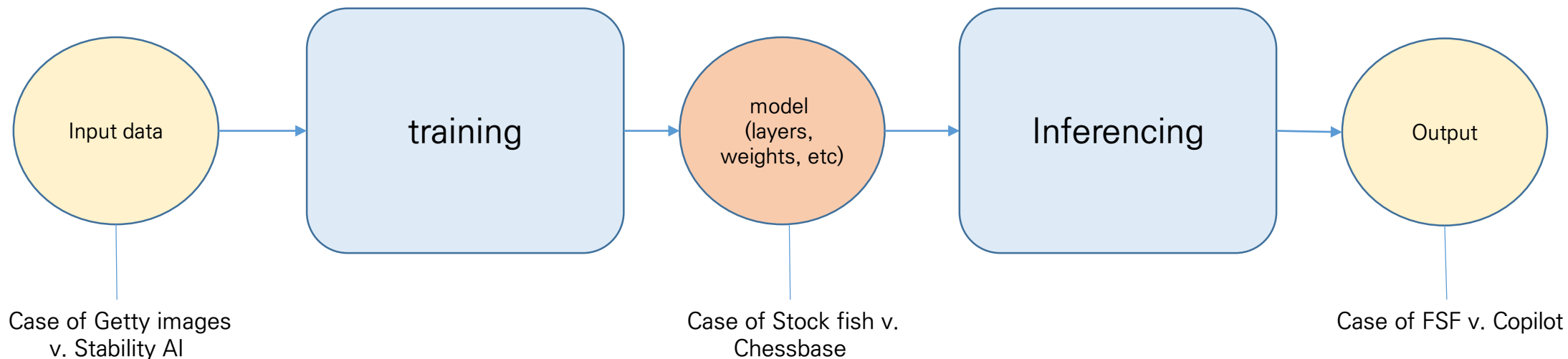
No.	법리	Features
1	미국	• 인간의 창작물만 저작물로 인정
2	EU	• 공정이용, 연구목적에 한해 데이터 마이닝을 제한
3	영국	• 저작물은 AI의 도움을 받은 사람이 만들 수 있음
4	일본	• 저작권은 인간의 창작물에 한정, 인간이 창작에 관여없이 AI에 의해 완전 자율적으로 작성된 콘텐츠는 저작권 없음
6	한국	• 인간이 만든 것만 저작물로 인정 (복제권, 전송권, 2차적 저작물 작성, 동일성유지권)

• 특허권자 불인정 사례

국가	AI 발명자 인정여부	진행 상황	국가	AI 발명자 인정여부	진행 상황
한국	(특허청) 불인정	행정소송 제기	영국	(항소법원) 불인정	대법원 계류중
미국	(항소법원) 불인정	대법원 계류중	독일	(항소법원) 불인정	대법원 계류중
유럽	(최종법원) 불인정	불인정 확정	호주	(대법원) 불인정	불인정 확정

세계 주요국들의 '다부스' 특허출원 관련 소송 진행 경과/사진제공=특허청

Legal Issues of AI Technology



저작권법, 콘텐츠산업법 상 보호받는 작업결과물이거나 부정경쟁방지법 상 상당한 투자나 노력으로 보호받을 수 있는 데이터 포함 가능

학습 모델을 GPL 파생 저작물인지가 쟁점

인공지능 학습과정에서 데이터 복제, 분석하게 되므로 저작권법상의 침해 문제 발생 가능

AI-related Content in Copyright Law [5]

- **공정이용**

- 저작물의 통상적인 이용방법과 충돌하지 아니하고 저작자의 정당한 이익을 부당하게 해치지 아니하는 경우

- **일시적 복제**

- 인공지능의 학습을 위한 저작권 복제가 일시적인 경우

- **면책규정**

- 빅데이터 활용 활성화를 위해 정보를 분석하는 과정에서 필요한 경우 저작물 등을 복제할 수 있도록 규정하는 내용의 저작권법 개정안 발의

- **제43조(정보분석을 위한 복제 · 전송)**

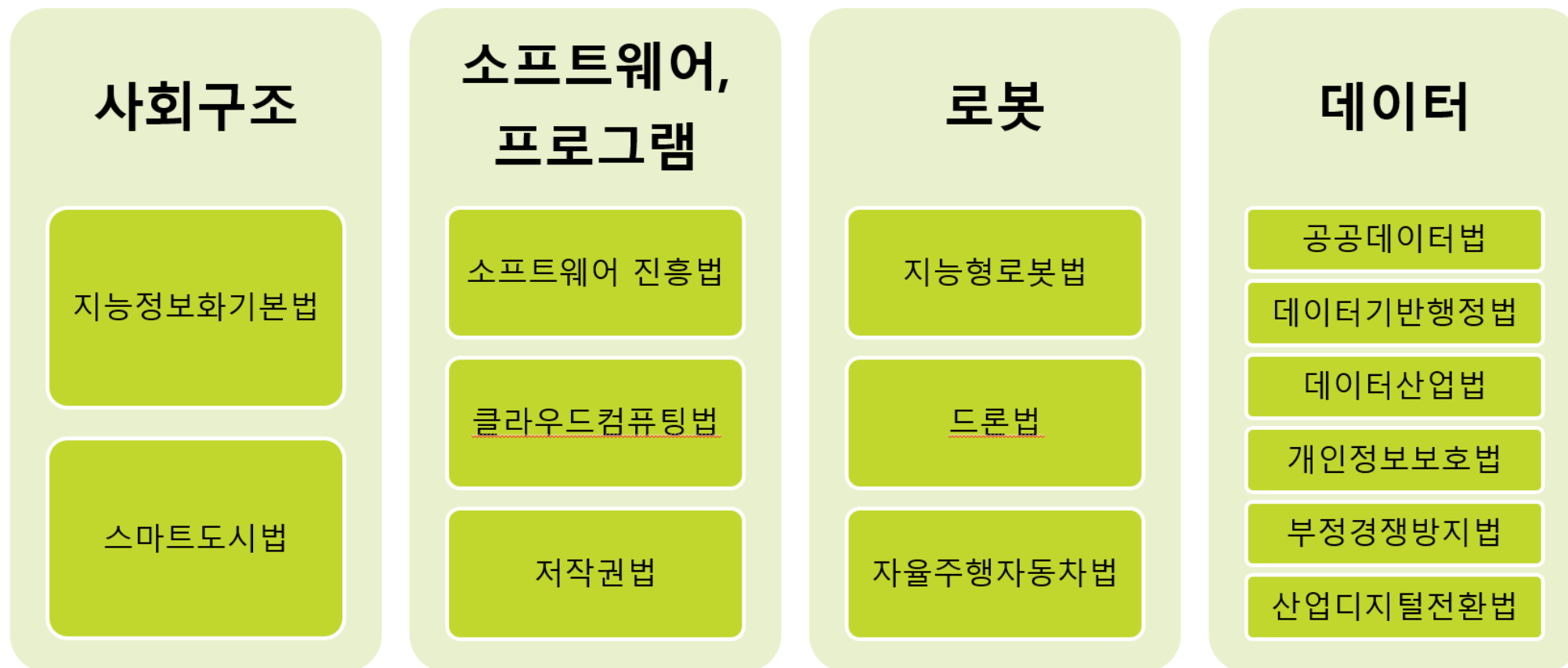
- ① 컴퓨터를 이용한 자동화 분석기술을 통해 다수의 저작물을 포함한 대량의 정보를 분석(규칙, 구조, 경향, 상관관계 등의 정보를 추출하는 것)하여 추가적인 정보 또는 가치를 생성하기 위한 것으로 저작물에 표현된 사상이나 감정을 향유하지 아니하는 경우에는 필요한 한도 안에서 저작물을 복제 · 전송할 수 있다. 다만, 해당 저작물에 적법하게 접근할 수 있는 경우에 한정한다.

- ② 제1항에 따라 만들어진 복제물은 정보분석을 위하여 필요한 한도에서 보관할 수 있다.

Protection of Works by AI [5]

- **필요성:** 개발자들에게 충분한 경제적 보상을 제공하여, 인공지능 기술 발전의 동기 제공
- **문제점:** 창작물의 영역이 인공지능의 결과물로 포화, 인간의 창작 동기를 감퇴시키고 결과적으로는 인류의 문화발전에 관한 공익을 저해
- **저작권법 제2조 제1호는 “저작물”을 “인간의 사상 또는 감정을 표현한 창작물”이라고 정의**
 - 인공지능의 창작물은 저작권법에 따른 저작물로 보호될 수 없음
- **인간이 인공지능과 상호 협력하면서 인공지능의 창작 행위에 기여한 경우**
 - 인간의 기여분이 인정된다면 인간의 저작권은 인정
- **업무상 저작물인 경우**
 - 인공지능과 회사 사이에 사용 관계에 의한 실질적인 지휘, 감독관계가 있다고 보기도 어렵고, 업무상 저작물도 인간이 창작한 저작물일 것을 전제로 하므로, 인공지능이 생성한 창작물에 대하여 업무상 저작물의 법리를 그대로 적용하기는 어려울 것임
- **인공지능 제품 그 자체는 제조회사의 “상당한 투자나 노력으로 만들어진 성과”에 해당, 인공지능이 생성한 창작물 또한 본질적으로 제조회사의 알고리즘 프로그램에 내재된 기술로 생성된 것으로 “상당한 투자나 노력으로 만들어진 성과”에 해당**
 - 부정경쟁방지법에 의한 보호 가능
- **경쟁자가 상당한 노력과 투자에 의하여 구축한 성과물을 상도덕이나 공정한 경쟁질서에 반하여 자신의 영업을 위하여 무단으로 이용함으로써 경쟁자의 노력과 투자에 편승하여 부당하게 이익을 얻고 경쟁자의 법률상 보호할 가치가 있는 이익을 침해하는 행위는 부정한 경쟁행위로서 민법상 불법행위에 해당 (대법원 2010. 8. 25.자 2008마1541 결정)**

Current Status of AI-related Laws in Korea [5]



AI-related Act Proposed by the National Assembly ('22.6.) [5][6]

인공지능

- 인공지능 연구개발 및 산업진흥, 윤리적 책임 등에 관한 법률안 (이상민 의원)
- 인공지능산업 육성에 관한 법률안 (양향자 의원)
- 인공지능 기술 기본법안 (민형배 의원)
- 인공지능 육성 및 신뢰기반 조성 등에 관한 법률안 (정필모 의원)
- 인공지능 교육 진흥법안 (안민석 의원)
- 인공지능에 관한 법률안 (이용빈 의원)

알고리즘

- 지능정보화 기본법 개정안 (이원욱 의원)
- 정보통신망법 개정안 (이원욱 의원)
- 정보통신망법 개정안 (류호정 의원)

딥러닝

- 저작권법 개정안

• AI 개발 과정에 개입 가능한 규제법

- 신용 정보의 보호 및 이용에 관한 법률
- 개인정보보호법
- 공정거래법

• 일반법을 마련하기보다 AI 개발 과정에 적용 가능한 기존법과 가이드라인을 적극 활용할 것을 권장

- Cf) EU는 2018년부터 법안 마련 작업을 단계적으로 마련

3. Open Source Compliance on AI

Open Source Compliance on AI

- AI 소스코드뿐만 아니라 학습데이터, AI 창작물의 법률 이슈도 함께 검토
 - 소스코드: 오픈소스 라이선스 준수 여부
 - 학습데이터: 저작권, 퍼블리시티권
 - AI 창작물: 저작권
- 학습데이터(★)
 - Publicly Available Datasets를 많이 활용 [7][9][10]
 - 메타데이터 라이선스와 데이터의 저작권은 data point별로 별도 존재
 - ✓ LAION-5B: 메타데이터는 CC BY-4.0, 이미지는 별도의 저작권 존재
 - ✓ 상업적 활용이 불가능할 수도 있음
 - 외부 공개 시 학습데이터의 법률 이슈도 검토 필요

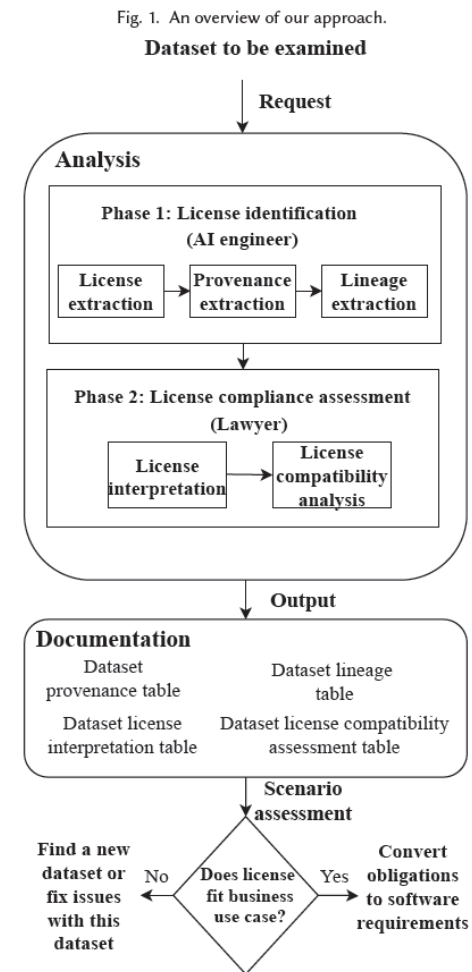
How to Treat Datasets (1)

• 데이터셋 라이선스의 어려움

- 해당 데이터셋에 대한 안전하고 정확한 라이선스 식별이 힘들
- 해당 데이터셋에 대한 라이선스가 유효한지 확인하는 것이 어려움
- Publicly available dataset에 적용된 라이선스는 일반적으로 모호하여 권리와 의무가 모호

• 논문의 접근 방법 [9], [10]

- Phase 1: License identification
 - ✓ Step 1: License extraction
 - ✓ Step 2: Provenance extraction
 - ✓ Step 3: Lineage extraction
 - ❖ Substep 1: 데이터셋 생성 프로세스 추적
 - ❖ Substep 2: 데이터 소스의 공식 출처 찾기
 - ❖ Substep 3: 시간적으로 적용 가능한 라이선스인지 확인
 - ❖ Substep 4: 데이터 소스에 대한 라이선스 식별
- Phase 2: License compliance assessment
 - ✓ Step 1: License interpretation
 - ✓ Step 2: License compatibility analysis



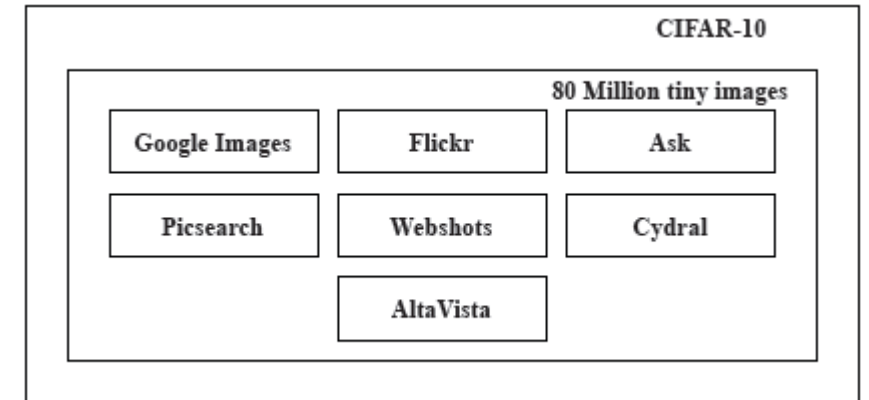
How to Treat Dataset (2)

• CIFAR-10의 Provenance 추출

Table 1. Recorded provenance details for CIFAR-10

Dataset-related details	Dataset name	Dataset version	Origin date	Origin
	CIFAR-10	N/A	2009	https://www.cs.toronto.edu/~kriz/cifar.html
	Description of dataset		Description of data collection process	
	The CIFAR-10 dataset consists of 60000 32x32 colour images in 10 classes, with 6000 images per class. There are 50000 training images and 10000 test images		The CIFAR-10 and CIFAR-100 are labeled subsets of the 80 million tiny images dataset. They were collected by Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton.	
	Downloaded outlet	Is outlet licensed?	Is dataset publicly available?	Additional notes
	N/A	N/A	Yes	This dataset is a subset of another dataset called 80 Million Tiny Images
License-related details	Where license was found		License location	License content
	Present on the official dataset website		https://www.cs.toronto.edu/~kriz/cifar.html	(not pasting content due to space)
Metadata	Hashcode		Size	Format
	MD5: c58f30108f718f92721af3b95e74349a (Python version)		163MB (Python version)	tar.gz

• Lineage 추출



How to Treat Dataset (3)

• CIFAR-10의 라이선스 해석

License metadata	Licensor		License name		Dataset name		Dataset version
	Alex Krizhevsky		Custom license		CIFAR-10		N/A
	Credit/Attribution Notice						
	Learning Multiple Layers of Features from Tiny Images, Alex Krizhevsky, 2009.						
	License validity period		Liability /Warranty		Designated third parties		Additional conditions
	N/A		N/A		Only by agreement		None
Data (standalone)	Access		Tagging		Distribute		Re-represent
Rights	✓		✓ (X)		✓ (X)		✓ (X)
Obligations	Cite paper		Cite paper		Cite paper		Cite paper
Data rights in conjunction with model	Bench- mark	Re- search	Publish	In- ternal Use	Commercialization		Model Reverse Engineer
					Out- put	Model	
Rights	✓	✓	✓	✓	✓ (X)	✓ (X)	✓
Obligations	Cite paper	Cite paper	Cite paper	Cite paper	Cite paper	Cite paper	Cite paper

• OpenDataology Project [12]

Cf) Publicly Available Dataset

No.	Dataset	Features	Classification
1	CIFAR-10	<ul style="list-style-type: none"> collected by Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton. labeled subsets of the 80 million tiny images dataset CIFAR-10: 60000 32x32 color images in 10 classes, 6000 images per class (50000 training images and 10000 test images) CIFAR-100: 100 classes containing 600 images each, 20 superclasses (500 training images and 100 testing images per class) https://www.cs.toronto.edu/~kriz/cifar.html 	Image classification
2	ImageNet	<ul style="list-style-type: none"> an image DB organized according to the WordNet hierarchy (lexical database for English) https://www.image-net.org/ WordNet: https://wordnet.princeton.edu/license-and-commercial-use 	Image
3	Cityscapes	<ul style="list-style-type: none"> new large-scale dataset that contains a diverse set of stereo video sequences recorded in street scenes from 50 different cities, with high quality pixel-level annotations of 5 000 frames in addition to a larger set of 20 000 weakly annotated frames. https://www.cityscapes-dataset.com/ 	Stereo video sequences
4	FFHQ (Flickr-Faces-HQ)	<ul style="list-style-type: none"> a high-quality image dataset of human faces License: individual images were published in Flickr by their respective authors under either Creative Commons BY 2.0, Creative Commons BY-NC 2.0, Public Domain Mark 1.0, Public Domain CC0 1.0, or U.S. Government Works license https://github.com/NVlabs/ffhq-dataset 	Image of human faces
5	VGGFace2	<ul style="list-style-type: none"> A dataset for recognizing faces across pose and age around 3.31 million images divided into 9131 classes https://www.robots.ox.ac.uk/~vgg/data/vgg_face/ 	Image of human faces
6	MS COCO (Common Objects in Context)	<ul style="list-style-type: none"> Microsoft 사에서 만든 4세 아이가 쉽게 인식할 수 있는 91개 개체 유형의 사진이 포함된 총 250만 개 레이블로 표시된 인스턴스 데이터셋으로 91개의 클래스와 328,000장의 이미지가 포함되어 있고 Object의 수는 2,500,000여 개 https://cocodataset.org/#home 	Image

- 그 외: Flickr, WiderFace (Amazon), Kinetics-400/600/700, Something-something (Qualcomm), Ego4D, HMDB, UCF101, ADE20K, COYO, AIHub, LAION-5B, HuggingFace 등
- <https://pytorch.org/vision/stable/datasets.html>
- <https://paperswithcode.com/datasets>

Open Source Compliance Strategy

- **오픈소스 사용 시 점검 사항**
 - 소스코드 라이선스
 - 데이터셋 라이선스
 - 데이터 포인트 라이선스
- **오픈소스로 공개 시 점검사항**
 - 소스코드 라이선스
 - ✓ 오픈소스 라이선스 적용: MIT, BSD-3-Clause, GPL-2.0 등
 - 데이터셋 라이선스
 - ✓ CDLA Permissive License V2.0
 - ✓ CDLA sharing v1.0
 - ✓ Open Use of Data Agreement v1.0
 - ✓ Computational Use of Data Agreement v1.0
 - ✓ CCL
 - Cf) Community Data License Agreement [13]
 - ✓ <https://cdla.dev>, LF 프로젝트
 - ✓ 개인, 조직 간에 공개적으로 데이터 액세스, 공유 및 사용을 가능하게 하는 협업 라이선스

4. Concluding Remarks

- 발표 요약

- Stable Diffusion 관련 소송사례 2건 분석
- AI 관련 법적 이슈 분석
- 오픈소스 컴플라이언스 대응 방안

- 논의 사항

- AI의 법적 지위 허용 여부 및 AI 법률 발전 방향
- 공정 이용(Fair use)의 범위
- 학습데이터셋 검증 방안

- 향후 계획

- 공개된 학습 데이터셋 분석

감사합니다

