미국

미국에서의 AI 저작권 소송(4)

고려대학교 법학전문대학원/교수
이대희

1. Millette 관련 소송

• 유튜버 David Millette은 유튜브 콘텐츠를 학습데이터로 사용한 OpenAI, Google, Nvidia를 상대로 캘리포니아 북부 연방지방법원에서 다음의 집단소송을 제기하였다.

① Millette v. OpenAI, Inc. (5:24-cv-04710, 2024.8.2.)
② Millette v Google (5:24-cv-04708, 2024.8.2.)
③ Millette v Nvidia (5:24-cv-05157, 2024.8.14.)

• 위의 소송 중에서 ②와 ③은 취하되고, ①은 관할집중 결정에 의하여 뉴욕 남부 연방지방법원(S.D.N.Y.)으로 이송되어, Millette v. OpenAI, Inc.(1:25-cv-03297) 케이스로 진행되고 있다.[1)]

• 위 ①, ②, ③ 케이스의 사실관계와 주장사항은 모두 동일하다.[2)]

(1) 사실관계

① 원고들은 유튜브 사용자이자 영상 제작자들이다.

② OpenAI의 학습데이터를 구성하는 많은 부분은 원고가 생성하고 업로드한 영상물이고, OpenAI는 전사(transcribe, 음성·영상에 포함된 말을 그대로 텍스트로 옮기는 행위)하여 모델 학습과 미세조정에 사용하였다.

(2) 원고주장

원고들은 저작권 침해가 아닌 주법(州法)상 부당이득·원상회복·불공정 경쟁을 청구원인으로 삼았는데, 이는 '전사'의 저작권법상 지위가 모호하여 공정이용 항변이 받아들여질 위험을 회피하기 위한 것으로 보인다.

2. Thomson Reuters Enterprise Centre GmbH v. ROSS Intelligence Inc.

• 델라웨어 연방지방법원(765 F.Supp. 3d 382 (D.Del. 2025))

• 원고 Thomson Reuters는 법률 플랫폼 Westlaw를 소유하고 있고, 피고 Ross Intelligence는 AI를 이용하여

1) 이송된 법원의 사건명으로는 In re OpenAI, Inc. Copyright Infringement Litigation(1:25-md-03143)인데, 이 사건명에는 여러 케이스가 포함되어 있고 그 중의 하나가 Millette v. OpenAI 케이스이다.

2) S.D.N.Y. 이송 및 Millette 케이스에 대해서는 이대희, 미국에서의 AI 저작권 소송(2) (저작권동향 2026.2.11.) 참조.

법학 연구 검색엔진을 제공하고 있다.

- 이 케이스는 델라웨어 연방지방법원에서 제기되어(1:20-cv-00613, 2020.5.6.), 제1심 판단이 이루어졌고(2025.2.11.), 현재 제3연방 항소법원(3rd Cir.)에서 항소심이 진행중이다.

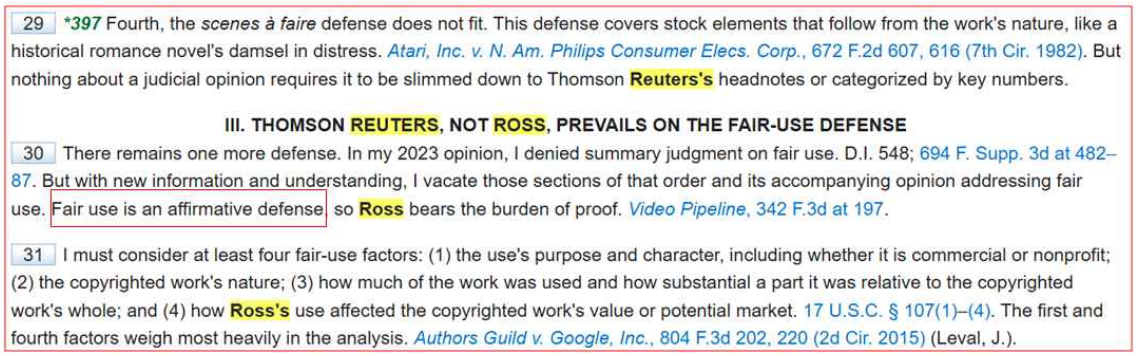
(1) 사실관계

Westlaw의 경쟁자인 Ross는 AI 검색툴을 학습시키기 위하여 법률 질문·응답 데이터베이스를 필요로 하였고, 이에 따라 Westlaw에 콘텐츠 이용허락을 요청하였으나 Westlaw는 거절하였다. Ross는 LegalEase로부터 Bulk Memos라는 학습데이터를 획득하였는데, Bulk Memos는 법학 질문과 이에 대한 적절한 응답과 그렇지 못한 응답을 편집한 것이었다. LegalEase는 Westlaw의 헤드노트(headnote)를 이용하여 25,000개의 질문-답변 세트를 만들어냈는데, 변호사가 할 법한 질문에 대하여 판결문에서 직접 인용한 문장을 답변으로 제공하였고, Westlaw의 헤드노트를 복제하거나 붙여넣기를 한 것은 아니었다. 최종적으로 Ross는 Bulk Memos를 학습데이터로 변환하여 Westlaw와 경쟁하는 제품을 제작하였다.

(2) 헤드노트

이 케이스에서 쟁점이 된 대상은 Westlaw가 제공하는 3,384개의 헤드노트에 대한 것이다. Westlaw는 판결문, 법령, 논문 등을 제공하는데 판결문의 경우, 판결 본문이 시작되기 전에 개요(synopsis), 판결주문(holding), 헤드노트 및 키번호(key number)를 제공한다. 헤드노트는 특정 판례의 법률 요점을 요약한 문장인데, 판례에 따라 몇백 개에 해당하는 것도 존재하다. 키번호는 법률 요점을 법률 주제 체계에 따라 분류한 번호인데, Westlaw가 10만여 개의 법률 쟁점에 따라 부여하고 헤드노트와 함께 제공된다. 본 대상 케이스도 45개의 헤드노트가 존재하는데, 예컨대 [30]은 다음과 같다.

위의 이미지는 West Key Number System에 따라 분류한 West headnote인데, 이것은 Thomson Reuters Enterprise Centre GMBH v. Ross Intelligence Inc. 케이스의 [30]번에 해당하는 헤드노트이다. 헤드노트 번호인 "[30]"을 클릭하면 다음과 같이 헤드노트의 내용에 해당하는 판례 부분으로 연결되어 해당 주제에 대한 판결의 특정 부분을 쉽게 찾아볼 수 있다.



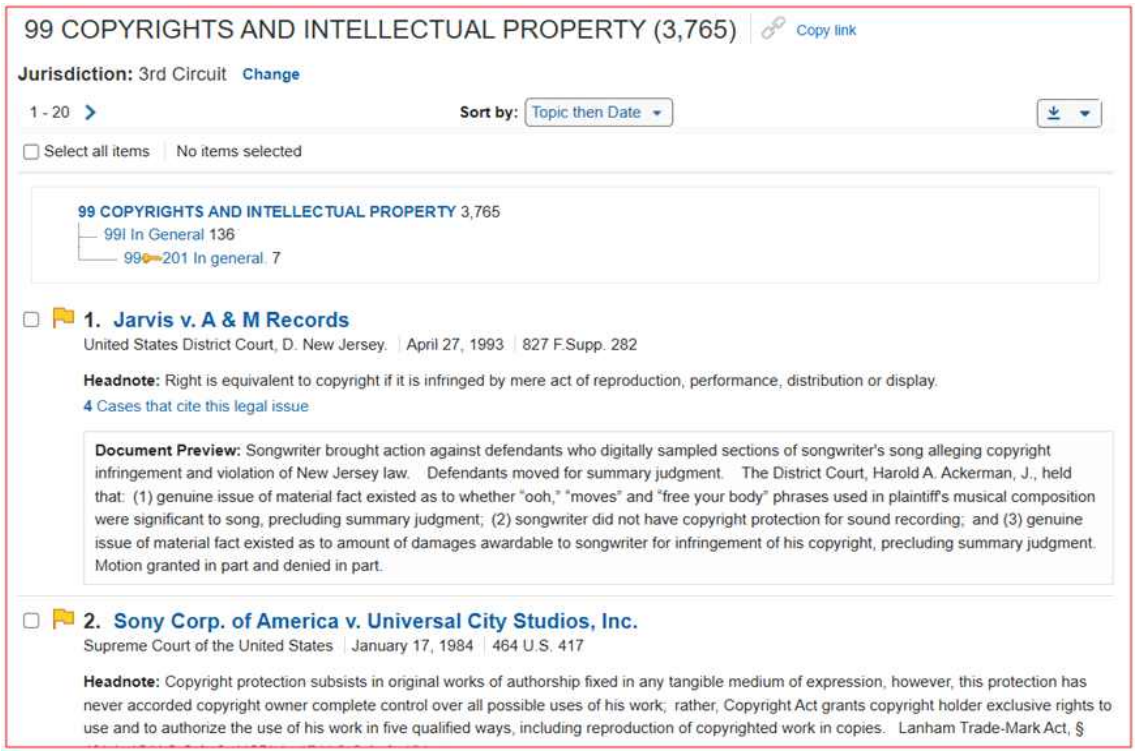
29 *397 Fourth, the *scenes à faire* defense does not fit. This defense covers stock elements that follow from the work's nature, like a historical romance novel's damsel in distress. *Atari, Inc. v. N. Am. Philips Consumer Elecs. Corp.*, 672 F.2d 607, 616 (7th Cir. 1982). But nothing about a judicial opinion requires it to be slimmed down to Thomson **Reuters's** headnotes or categorized by key numbers.

III. THOMSON REUTERS, NOT ROSS, PREVAILS ON THE FAIR-USE DEFENSE

30 There remains one more defense. In my 2023 opinion, I denied summary judgment on fair use. D.I. 548; 694 F. Supp. 3d at 482–87. But with new information and understanding, I vacate those sections of that order and its accompanying opinion addressing fair use. Fair use is an affirmative defense so **Ross** bears the burden of proof. *Video Pipeline*, 342 F.3d at 197.

31 I must consider at least four fair-use factors: (1) the use's purpose and character, including whether it is commercial or nonprofit; (2) the copyrighted work's nature; (3) how much of the work was used and how substantial a part it was relative to the copyrighted work's whole; and (4) how **Ross's** use affected the copyrighted work's value or potential market. 17 U.S.C. § 107(1)–(4). The first and fourth factors weigh most heavily in the analysis. *Authors Guild v. Google, Inc.*, 804 F.3d 202, 220 (2d Cir. 2015) (Leval, J.).

위 첫 이미지에서 "[30]" 다음에 나오는 "Copyrights and Intellectual Property" 부분을 클릭하면 다음과 같이 나타난다.



99 COPYRIGHTS AND INTELLECTUAL PROPERTY (3,765) Copy link

Jurisdiction: 3rd Circuit **Change**

1 - 20 > **Sort by:** Topic then Date

Select all items | No items selected

99 COPYRIGHTS AND INTELLECTUAL PROPERTY 3,765
- 99I In General 136
  - 99k201 In general 7

1. Jarvis v. A & M Records
United States District Court, D. New Jersey. | April 27, 1993 | 827 F.Supp. 282

Headnote: Right is equivalent to copyright if it is infringed by mere act of reproduction, performance, distribution or display.
4 Cases that cite this legal issue

Document Preview: Songwriter brought action against defendants who digitally sampled sections of songwriter's song alleging copyright infringement and violation of New Jersey law. Defendants moved for summary judgment. The District Court, Harold A. Ackerman, J., held that: (1) genuine issue of material fact existed as to whether "ooh," "moves" and "free your body" phrases used in plaintiff's musical composition were significant to song, precluding summary judgment; (2) songwriter did not have copyright protection for sound recording; and (3) genuine issue of material fact existed as to amount of damages awardable to songwriter for infringement of his copyright, precluding summary judgment. Motion granted in part and denied in part.

2. Sony Corp. of America v. Universal City Studios, Inc.
Supreme Court of the United States | January 17, 1984 | 464 U.S. 417

Headnote: Copyright protection subsists in original works of authorship fixed in any tangible medium of expression, however, this protection has never accorded copyright owner complete control over all possible uses of his work; rather, Copyright Act grants copyright holder exclusive rights to use and to authorize the use of his work in five qualified ways, including reproduction of copyrighted work in copies. Lanham Trade-Mark Act, §

위 이미지는 "Copyrights and Intellectual Property"라는 주제의 판결이 제3연방 항소법원에서 3,765개가 검색된다는 것을 알려주고 있다(물론 제3연방 항소법원이 아니라 다른 법원으로 변경하거나 확대하는 것도 가능하다).

위 첫 이미지 오른쪽 부분에 표시되어 있는 다음의 정보들은 Westlaw가 저작권 및 지적재산권을 대주제부터 계층적으로 세부 법률 분야로 연결하는 것을 보여주고 있다.

99 Copyrights and Intellectual Property
99XVI Actions and Judicial Proceedings
99XVI(C) Evidence
99XVI(C)2 Presumptions, Inferences, and Burden of Proof
99k1017 Defenses and Permitted Uses
99k1020 Fair use

"99 COPYRIGHTS AND INTELLECTUAL PROPERTY"의 대분류에서 "99k1020 Fair use"까지 세부 분류까지의 항목을 클릭하면 각 항목에 해당하는 제3연방 항소법원의 판결이 검색된다.

이 케이스에서 쟁점은 헤드노트를 이용한 것에 관한 것인데, 헤드노트는 판례의 법률적 요점을 요약한 것으로서 편집자들이 작성하고 판결문의 원문을 그대로 사용하기도 한다. 위의 [30]에 해당하는 헤드노트는 "Fair use is an affirmative defense on which a copyright-infringement defendant bears the burden of proof." 부분이다. 헤드노트는 판결문의 원문을 사용하거나 비교적 짧은 문장이므로 저작물성이 부인될 수 있는데, 헤드노트가 항상 이런 식으로 구성되는 것은 아니므로 저작물성이 항상 부인된다고 하기 어렵다. ~~위~~ 위 [30]의 헤드노트는 16개의 낱말로 구성되어 있지만, [44]는 109개의 낱말로 구성되어 있다.[3)]

(3) 법원의 판단: 공정이용 여부

1. 첫째, 헤드노트는 장문의 판결문을 정교하게 축소시킨 핵심적인 법률 요점으로서, 판결문의 일부분을 추출, 종합, 설명하여 독창성(originality)을 드러낼 수 있고, 자체적으로 저작물로 성립하거나 편집저작물에 해당한다. 둘째, Westlaw가 자료를 정리하는 방법인 키번호 시스템은 독창성을 충족하기 위한 최소한의 불꽃(spark)을 가지는 것으로서 독창성이 있다.

2. 침해에 대하여 고의나 과실이 없다거나(innocent infringement), 저작권 남용, 아이디어와 표현의 합체(merge), 필수장면(scènes à faire) 등의 항변을 받아들이지 않는다.

3. 공정이용의 첫째 요소와 관련하여, Ross의 이용은 상업적이고, 변형적이지 않고, 원피고는 경쟁관계에 있으며, (컴퓨터프로그램에 있어서) 중간과정의 복제(intermediate copying)에 대하여 공정이용이 인정되는 경우가 있지만 이 사건은 컴퓨터프로그램을 복제하는 것이 아니며 아이디어에 접근하기 위하여 표현을 복제해야 할 필요가 있는 코드는 존재하지 않는다. 따라서 첫째 요소는 Ross에게 불리하게 작용한다.

Westlaw의 헤드노트는 편집적 판단이 필요하지만 창의성이 높지 않고, 키번호 시스템은 사실적 편집물에 해당하

3) [44] Copyrights and Intellectual Property Adaptations and derivative works The fourth factor of the fair-use defense, namely the effect of the defendant's use of copyrighted material on the copyrighted work's value or potential market, favored legal-research platform in its copyright infringement suit against competitor for having allegedly infringed platform's copyrighted material through competitor's development of a legal-research product driven byartificial intelligence (AI) that was trained using platform's copyrighted annotations, or headnotes, of judicial opinions and its numbering system reflecting a taxonomy for organizing its headnotes, where competitor intended to compete with platform by developing a substitute in platform's original market for legal research, and competitor's conduct affected a potential derivative market for training legal AI.

고, Ross는 헤드노트를 공중에게 공개한 것이 아니므로, 둘째 및 셋째 요소는 Ross에게 유리하게 작용한다.

시장으로는 법률검색 플랫폼 시장과 (최소한) 법률 AI 학습데이터 시장이 존재하는데, Ross는 Westlaw와 경쟁하는 대체시장을 개발하려고 하였고, Ross가 잠재적 학습데이터 시장이 존재하지 않는다거나 이러한 시장이 영향을 받지 않는다는 것을 입증하지 못하였고, Ross에 의하여 공익이 제공되는 것도 아니다. 따라서 넷째 요소는 Ross에 유리하게 작용하지 않는다.

(4) 결론

원고는 저작권 침해 및 기여침해를 구하는 약식판결(summary judgment)을 청구하였고, 피고는 공정이용 항변을 구하는 약식판결을 청구하였는데, 법원은 원고의 청구를 인용하였다.

(5) 소송경과 및 항소심

이 소송은 2020.5.6. 제기되었는데, 원고 Thompson Reuter는 법원이 일정한 자료 부분에 대하여 저작권 침해를 인정하고, 피고의 공정이용 항변을 배제할 것을 구하는 약식판결을 청구하였고, 피고 Ross는 원고의 헤드노트를 자신이 사용하는 것이 공정이용에 해당한다는 것을 구하는 약식판결을 청구하였다. 법원은 2023.9.25. 판결(694 F.Supp.3d 467 (D.Del. 2023))을 통하여, 원고의 저작권 침해 및 피고의 공정이용 항변에 대한 약식판결 청구를 기각하였다. 법원은 약식판결 청구와 관련된 중요한 사실관계에 대하여 다툼이 존재한다고 판단하여 기각하였고, 이에 따라 심리(trial)가 이루어지고 배심원 판단이 이루어질 예정이었다. 그런데 헤드노트 등의 저작권, Ross에 의한 실제 복제 여부, 변형적 이용 여부 등 많은 쟁점이 존재하였던 상황에서, 심리에 의하지 않고 약식판결 청구만을 기각하였으므로 이 소송의 판결이 확정된 것이 아니었고, 중간판결(interlocutory order)이 내려진 것뿐이었다. 미국 '민사소송규칙(FRCP)'에 의하면, 판결이 확정되기 전까지는 이러한 중간판결을 수정할 수 있는데(§54(b)), 앞서 상세히 살펴본 판단(2025.2.11.)이 수정한 판단이다.

판결이 확정되어야 항소하는 것이 원칙이지만, 최종 판결 전에 특정 쟁점에 대해서만 항소하는 것이 허용되는데(FRCP §1292(b), 중간항소, interlocutory appeal), Ross는 중간항소를 청구하였고 법원이 이를 수용하였다. Ross는 헤드노트가 독창성 요건을 충족하지 못하는지 여부와 Westlaw 헤드노트의 0.076%를 사용한 것이 변형적인지 여부에 대하여 항소 허가 청구를 하였으므로, 항소심에서는 이들 쟁점에 대하여 심리될 예정이다.[4] 항소심 판단이 나올 때까지 연방지방법원에서 진행되었던 나머지 사항에 대한 심리도 연기된 상태이다.

3. Advance Local Media LLC v Cohere Inc.

- 뉴욕 남부 연방지방법원 (1:25-cv-01305, 2025.2.13.)
- 원고 Advance Local Media 등[5]은 출판사이고, 피고 Cohere는 AI 모델을 개발, 운영, 이용허락하는 사업을 운

4) Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc, 25-2153, (2025.6.24., 3rd Cir.).

5) Advance Magazine Publishers Inc. d/b/a Condé Nast; The Atlantic Monthly Group; Forbes Media; Guardian News & Media Limited; Insider, Inc.; Los Angeles Times Communications; The McClatchy Company; Newsday; Plain Dealer Publishing Co.; Politico LLC; The

영하고 있다.

(1) 사실관계 (원고 주장)

피고는 Command Family(Command, Command R, and Command R+ 등)라는 LLM AI 시스템을 학습시키기 위하여, 원고 저작물(뉴스 기사 등)을 복제하여 학습, 실시간 이용, 검색증강생성(RAG) 기능을 이용한 결과물 생성을 위하여 사용하고 있다. 피고는 원고 저작물을 있는 그대로(vervatim copies) 제공하거나, 상당히 발췌하여 제공하거나, 원저작물을 대체하는 요약의 형태로 결과물을 제공하고 있다. 피고는 원고 저작물을 복제하여 학습 데이터셋을 구축하고, 실시간으로 사용하고, 원고들의 기사를 결과물로 제공하고, 원고 명의로 가짜 기사를 제공하여, 저작권과 상표권을 침해하고 있다.

(2) 위반 주장

1. 직접 저작권 침해

피고는 원고의 이용허락이나 동의를 받지 않고 원고 저작물을 복제, 배포, 전시, 2차적저작물을 작성함으로써, 원고의 복제권, 2차적저작물작성권, 배포권, 전시권을 직접 침해하고 있다. 이러한 침해는 원고의 저작권을 무시하거나 고의에 의한 것이고, 이로 인하여 원고는 상당한 회복 불가능한 손해를 입었고 앞으로도 입게 될 것이다.

2. 2차적 저작권 침해

첫째, Command를 사용하는 주체와 이용허락을 받은 주체는 원고 저작물을 복제, 배포, 전시, 2차적저작물을 작성함으로써 저작권을 직접 침해하고 있는데, 피고는 이러한 직접침해를 인식하고 있었고 침해에 실질적으로 기여함으로써, 기여침해 책임이 있다.

둘째, 피고는 원고 저작물을 복제, 상당한 발췌, 원고 저작물을 대체하는 요약을 제공하고 있고 이러한 서비스를 적극 홍보함으로써, 사용자들의 직접침해를 유인(induce)하였다.[6)]

셋째, 피고는 자신의 AI 모델에 의한 침해행위를 감독하고 통제할 법적 권리와 실질적 능력을 가지고 있는데도 합리적 조치를 취하지 않음으로써 대위침해 책임을 부담한다.

4. Pierce v. Photobucket, Inc.

- 콜로라도 연방지방법원 (1:24-cv-03432, 2024.12.11.)
- 피고 Photobucket은 온라인 사진 저장서비스를 제공하고 있었는데, 현재 가지고 있는 130억 장 이상의 사진을 이용하여 생체인식 및 생성형 AI 등 다양한 용도로 사용하고 있다. 원고 Mac Pierce 등은 이러한 사진에 담겨있는 인물의 주체이다(집단소송).

Republican Company; Toronto Star Newspapers Limited; Vox Media. Condé Nast는 Vogue, The New Yorker, GQ, Vanity Fair, Wired, Bon Appétit, Architectural Digest 등을 발행하고 있다.

6) 미국 판례법상 2차적(간접) 침해책임은 대체로 기여침해와 대위침해로 나뉘는데, Metro-Goldwyn-Mayer Studios Inc. v. Grokster 케이스(545 U.S. 913 (2005)) 이후 유인침해가 독립적인 간접침해로 인정되는 경향이 있다.

(1) 사실관계 (원고 주장)

피고가 130억 장의 사진을 생체인식 및 생성형 AI 용도로 이용하거나 제3자에게 이용허락하는 것은 유효한 동의를 얻지 않은 것이고, 동의를 얻기 위하여 속임수와 강압을 사용하였다.

(2) 위반 (원고 주장)

1. 피고의 행위는 프라이버시 및 기망 금지에 관한 주법 위반, 계약 위반, 타인의 재산을 자신의 것처럼 처분하는 불법행위(conversion), 부당이득, 공동 불법행위(civil conspiracy)에 해당한다.

2. 퍼블리시티권 침해

원고 등 집단의 얼굴 구조(face geometry), 생체 식별자(biometric identifier), 모습(likeness)을 동의 없이 상업적으로 이용함으로써, 캘리포니아(Cal. Civ. Code §3344(a)), 뉴욕(Civil Rights Law §§50, 51), 버지니아(Va. Code §8.01-40(A)), 퍼블리시티권을 인정하는 커먼로(common law)를 위반하였다.[7)]

3. 저작권관리정보 규정 위반

피고 Photobucket 및 미확인 피고들은 사진저작물의 저작자 명칭, 이용 조건, 제목, 날짜 및 기타 식별정보 등 저작권 관리정보(CMI)를 제거하거나, 제거되었다는 것을 인지하고서 배포함으로써, CMI 제거 및 배포금지에 관한 규정을 위반하였다(미 저 §1202(b)(1), (3)).

(3) 법원의 판단

법원은 2026.3.10. Photobucket 이용약관에 포함된 중재조항의 존재와 적용 가능성을 전제로 일부 원고의 분쟁은 법원이 아닌 중재에서 해결되어야 한다고 판단하였고, 동시에 원고들이 제기한 금전적 손해배상 청구에 대해서는 연방법원이 이를 판단할 관할권이 부족하다는 이유로 각하하였다. 따라서 DMCA 위반이나 저작권 침해 등 실체적 위법 여부에 대해서는 판단에 이르지 않았으며, 사건 전반에 대해 더 이상 법원에서 진행할 절차가 없다고 보아 이를 최종 판결로 종결하기보다는 필요 시 재개할 수 있는 상태로 절차적으로 종료(administratively closed)하였다. 따라서 향후 중재 절차가 끝나거나, 관할 문제를 보완해서 다시 제기하거나, 추가 사정이 발생하면 당사자가 소송 재개를 신청하여 사건을 다시 진행할 수 있는 상태이다.

5. In Re Mosaic LLM Litigation[8)]

- 캘리포니아 북부 연방지방법원 (3:24-cv-01451-CRB)

---

7) 현재 미국에서 퍼블리시티권은 연방 차원에서 보호되지 않고 주 차원에서 보호되는데, 주 차원에서는 주에 따라 성문 입법이나 판례에 의하여 보호되고 있다. 캘리포니아주는 민사법(Civil Code)을 통하여 성명, 음성, 서명, 사진, 모습(likeness)을 보호하고(§§3,344), 뉴욕주는 민권법(Civil Rights Laws, CVR)을 통하여 성명, 초상(portrait), 사진, 모습, 목소리를 보호하고(§§50, 51), 민사 구제 및 절차법(Title 8.01)을 통하여 성명, 초상, 사진을 보호하고 있다(Va. Code §8.01-40).

8) 이 케이스 명칭은 In re... 식으로 시작하는데 "...에 관한 사건"을 의미한다. In re는 서로 대립하는 당사자 구조가 아니라 분쟁의 중심이 재산(property)이거나(예컨대 파산), 반드시 상대방을 필요로 하지 않는 경우(입양, 성명 변경 등), 병합소송 등에서 사용된다.

(1) 소송 경과

이 케이스는 캘리포니아 북부 연방지방법원에서 제기된 O'Nan, et al. v. Databricks Inc., et al. 케이스(3:24-cv-01451-CRB, 2024.3.8.)와 Makkai, et al., v. Databricks, Inc., et al. 케이스(3:24-cv-02653-CRB, 2024.5.2.)가 병합(2024.12.2.)된 케이스이다. 이 케이스들은 원고 Stewart O'Nan과 Rebecca Makkai 등이 LLM(MPT 및 DBRX)을 제작한 Mosaic ML, Inc.와 이를 배포한 Databricks, Inc.를 상대로 제기한 집단소송이었다.

두 케이스가 병합된 후 수정된 소장이 제출되었는데(2025.6.27.), 원고들은 첫째, 피고의 MPT-7B 및 MPT-7B 모델이 Books3 데이터셋으로 학습시켰고, 자신들의 서적이 Books3에 포함되어 있고 따라서 Mosaic ML은 저작권을 직접 침해하였다고 주장하였다. 둘째, Databricks의 DBRX 모델은 MosaicML의 침해행위에 기반하여 이를 확장한 것으로서 역시 저작권을 침해하였다고 주장하였다. 이에 따라 원고들은 피고들이 여러 차례에 걸친 Books3 데이터셋의 복제에 의하여 저작권을 직접 침해하였고, MosaicML을 인수한 Databricks가 대위침해 책임이 있다고 주장하였다.

(2) 사실관계 (원고 주장)(수정 소장)

수정된 소장은 MPT 모델의 학습데이터 이용에 관한 사실관계를 보다 구체화하고, 해당 데이터셋이 MosaicML에서 확보·보관된 이후, Databricks가 MosaicML을 인수 합병하는 과정에서 동일한 개발 조직 및 모델 개발 흐름 하에서 DBRX 모델의 개발 및 학습 과정에도 활용되었음을 주장하는 방향으로 보강하였다. 이에 의하여 기존 소장에서 기각된 DBRX 관련 직접침해 주장에 대한 사실적 기반을 보완하였다.

(3) 위반 주장

원고는 첫째, 저작물 복제에 의한 MosaicML 및 Databricks의 저작권 직접침해, 둘째, MosaicML의 침해에 대한 Databricks의 대위책임, 셋째, MosaicML 및 사용자에 의한 기여 및 유인책임이 있다고 주장하였다.

(4) 피고 주장

피고는 원고의 저작물이 해당 AI 모델의 학습 데이터(training dataset)에 실제로 포함되었다는 것이 입증되지 않았고, AI 모델의 개발 과정에서 일부 데이터가 사용되었을 가능성이 있다는 것만으로는 최종 모델이 해당 저작물을 기반으로 만들어졌다고 볼 수 없고, 특정 AI 모델에 대한 침해를 주장하려면 해당 모델이 실제로 원고의 저작물을 사용했다는 사실이 개별적으로 입증하여야 하므로, 원고의 청구가 기각될 것을 주장하였다.

6. In re Google Generative AI Copyright Litigation

- 캘리포니아 북부 연방지방법원(5:23-cv-03440-EKL)

• Leovy et al. v. Google LLC 케이스(5:23-cv-03440-EKL, N.D.C.A. 2023.7.11.)와 Zhang et al v. Google LLC et al. 케이스(5:24-cv-02531-EKL, N.D.CA. 2024.4.26.)를 병합(2024.10.28.)한 케이스이다.
• 원고 Steve Almond 등은 시각예술가 및 저작자로서 Google 및 Alphabet을 상대로 집단소송을 제기하였다.

(1) 사실관계(원고 주장)

1. Google은 상업적 AI 사업을 창출하기 위하여 수백만 개 등록 저작물의 저작권을 고의로 침해하였다. Google은 자사의 AI 모델인 Gemini(멀티 모달)와 Imagen(텍스트 입력하면 이미지 생성)을 학습시키는데 원고의 저작물을 사용하였다. AI 모델을 다양한 기반 제품(검색, 클라우드, Google Docs 등)에 통합하고 있는데 이러한 제품들은 원고 저작물로 AI 모델을 학습시킴으로써 창출할 수 있었다. Google은 여러 차례에 걸쳐 저작물의 복제물을 제작하였는데, 복제물은 ①데이터 수집과 ②학습과정에서 제작되어 ③AI 모델의 파라미터의 구조 안에 영구적으로 내재화(incorporate)되었다.

2. 학습데이터

(ⅰ) Gemini

Google은 LaMDA라는 LLM을 이용하여 Bard(Gemini 전신)를 개발하였는데, LaMDA는 Google의 데이터셋인 Infiniset(1.56조 개 단어 규모의 대규모 인터넷 콘텐츠 말뭉치, C4 데이터셋 포함)으로 학습·개발되었는데, C4는 Common Crawl 데이터셋을 필터링한 버전이다. Infiniset은 공개 온라인 커뮤니티(Reddit, Twitter 등)에서의 대화 50%, C4 데이터셋 12.5%, 컴퓨터프로그램의 소스코드 및 개발 관련 문서(GitHub, Stack Overflow 등) 12.5%, Wikipedia 12.5%, 영어 웹 문서 6.5%, 영어 이외 웹 문서 6.5%로 구성되어 있다. C4는 저작권 침해 복제물을 포함한 방대한 양의 저작권 자료를 포함하고 있는데, 전자책 해적시장으로 유명한 b-ok.org와 최소 27개의 기타 해적 시장의 데이터를 포함하고 있고, C4에는 세 번째로 큰 출처인 Scribd.com은 6천만 권의 전자 서적 및 오디오 서적이 포함되어 있다. Google의 Infiniset은 Bard를 학습시킨 기본 데이터셋이고, 원고 Almond 등의 저작물이 Infiniset 및 C4에 포함되어 있다.

(ⅱ) Imagen

Google은 LAION 데이터셋의 콘텐츠를 이용하여 Imagen을 구축하였는데, LAION은 AI 모델 학습에 사용되는 대규모 데이터셋을 공개적으로 제공하는 프로젝트이다. LAION-400M과 LAION-5B는 약 4억 개 및 58억 개의 이미지-텍스트 쌍을 기반으로 구성되어 있는데, 이미지 파일 자체를 포함하지 않고, 각 이미지에 대응하는 URL, 캡션, 유사도 점수, NSFW(부적절 콘텐츠 여부) 플래그, 이미지 크기 정보 등의 메타데이터로 이루어져 있다. AI를 학습시키려면 데이터셋에 포함된 URL을 통하여 실제 이미지를 별도로 다운로드받아야 한다. Google이 메타데이터에 포함된 URL로부터 실제 이미지를 다운로드하는 단계에서 이미지 복제물이 만들어지고, 이후 모델 학습 과정에서도 해당 이미지 데이터가 사용되면서 추가적인 복제가 이루어진다. 원고 Zhang 등의 저작물은 LAION-400M 및 LAION-5B(전자는 후자의 부분 집합임)에 존재하므로 Google은 원고의 저작권을 침해한 것에 해당한다.

(2) 위반 주장

1. 저작권 직접침해: Google

Google은 학습 데이터셋을 수집·구성하는 과정에서의 복제, 모델 개발단계에서의 사전 학습 및 학습에서의 여러 차례 복제, 데이터셋 수집·구성 및 사전학습·학습과정에서의 2차적저작물 작성 등에 의하여, 원고의 복제권, 배포권, 2차적저작물작성권, 전시권을 침해하였다.

2. 대위침해: Alphabet

Alphabet은 Google의 침해행위를 감독할 권리와 능력을 가지고 있고, 이러한 침해행위로부터 직접적인 금전적 이익을 얻게 됨으로써, Google의 직접침해에 대하여 대위침해 책임이 있다.

7. Dow Jones & Company, Inc. v. Perplexity AI, Inc.

- 뉴욕 남부 연방지방법원 (1:24-cv-07984, 2024.10.21.)
- 원고 Dow Jones & Company, Inc. 및 NYP Holdings, Inc.(모기업은 News Corporation)은 The Wall Street Journal과 the New York Post 등을 출판하는 주체이고, Purplexity는 사용자의 질의에 대하여 응답하는 AI 서비스를 제공하고 있다.

(1) 사실관계(원고 주장)[9)]

1. Perplexity는 ①RAG 데이터베이스(RAG 인덱스)에 포함시키기 위하여 원고의 웹페이지에 있는 저작물(뉴스 기사)을 대규모로 복제하고, ②질의에 따른 답변(결과물)으로 원고의 저작물을 전부 또는 일부 그대로 복제하거나, 바꾸어 쓰거나 요약하여 제공하고 있다.

2. RAG은 AI 응용프로그램이 특정의 제한된 데이터나 콘텐츠에 집중하여 사용자의 질의에 응답하도록 하는데, 이를 위하여 RAG 인덱스는 LLM이 학습된 이후에도 새로운 콘텐츠를 수집·저장한다. AI 응용프로그램이 사용하는 인덱스된 데이터베이스는 질의에 대한 응답을 제공하기 위하여 필요한 특정 정보에 대한 모든 콘텐츠를 저장하고 있다.

3. 결과물을 제공하기 위하여 RAG 인덱스에서 선택된 원본 콘텐츠는 사용자의 질의와 함께 LLM의 Context window에 다시 복제되고, LLM은 결과물 내에서 원본 콘텐츠를 복제하거나 2차적저작물을 작성한다.

4. Perplexity는 생성한 결과물(응답 포함)의 분석 등을 위하여 별도의 데이터베이스에 저장한다.

(2) 원고 위반 주장

1. RAG 인덱스 입력에 따른 저작권 침해

Perplexity는 RAG에 사용할 데이터베이스에 입력하기 위하여 웹크롤러를 이용하여 원고의 저작물을 복제하였는데, 이는 저작권 침해에 해당한다.

2. 사용자 질의에 대한 결과물 생성에 따른 침해

9) 소장은 2024.10.21. 제출되었는데 이후 2025.1.18. 수정된 소장이 제출되었음.

질의에 대한 결과물은 원고 결과물을 있는 그대로 또는 거의 동일하게 복제하거나, 요약, 축약, 파생콘텐츠 등의 형태로 제공되는 것으로서 모두 저작권 침해에 해당한다. 또한 결과물을 생성하는 과정에서 원고의 저작물을 추가적으로 복제하는데, 여기에는 Perplexity가 사업목적상 결과물을 저장하는 것을 포함한다.

(3) 피고 주장

첫째, Perplexity의 답변은 해당 정보가 있는 인터넷상의 출처에 대한 링크를 포함하고 있고, 이용자가 출처 정보에 접근하기 위해서는 해당 링크를 클릭하여야 하므로, Perplexity가 제공하는 서비스는 전통적인 검색엔진과 동일하다. 또한 이용자들이 정보가 있는 웹페이지를 우회하지 않도록 하기 위한 안전장치들을 두고 있다. 이러한 행위들은 저작권으로 보호되지 않는 사실과 아이디어를 이용하는 것에 해당한다.

둘째, 검색가능한 데이터베이스를 구축하기 위하여 저작물을 수집하는 행위는 수십 년 동안 법원이 인정한 것으로서, Perplixity가 인터넷상에서 콘텐츠를 인덱스하여 답변 엔진을 작동하기 위하여 내부처리 과정에서 이를 사용한 행위는 문제되지 않는다.[10)]

8. DOE 1 v. GitHub, Inc.

- 캘리포니아 북부 연방지방법원 (4:22-cv-06823, 2022.11.3.)
- 원고 J. Doe 1 및 J. Doe 2[11)]

(1) 소송 이력

이 소송은 ①원고의 집단소송 제기(2022.11.3.), 피고의 기각 청구, 법원의 부분 인용 및 기각 판결(2023.5.11.), ②원고의 1차 수정 소장 제출(2023.6.8.), 피고의 기각 청구, 법원의 부분 인용 및 기각 판결(2024.1.22.), ③원고의 2차 수정 소장 제출(2024.1.25.), 피고의 기각 청구, 법원의 부분 인용 및 기각 판결(2024.6.24.), ④원고의 중간 항소 청구(2024.7.24.), 법원의 항소 허가 및 나머지 부분에 대한 심리 정지(2024.9.27.) 등을 거쳐, ⑤현재 항소심(9th Circuit) 계류 중이다.

- 원고들은 자신들이 저작권 이익(copyright interest)을 가지고 있는 자료(소스코드 등)를 오픈소스 라이선스인 MIT License 등에 따라 GitHub(SW 개발자들이 SW를 개발하기 위한 코드(code) 저장·협업 플랫폼)에 게재하였고, 피고는 Github, MS Corporation, OpenAI이다.

(2) 사실관계 및 위반 주장(원고 주장)

1. 피고들은 GitHub에 게재되어 있는 대량의 코드 데이터를 AI 툴인 Copilot 및 Codex의 학습에 사용하였는데, 이들 코드는 MIT License 등과 같은 오픈소스 라이선스가 적용된 자료이다. 이러한 라이선스는 코드를 사용할 경

10) 답변의 나머지 사항들은 모두 원고의 주장이 입증되지 않는다는 식으로 되어 있음.

11) 원고는 Doe 1으로 되어 있는데, 신원을 보호하거나, 특정되지 않은 당사자를 지정하거나, 집단소송의 대표가 익명을 원할 때 사용되는 명칭이다. 남자는 John Doe, 여자는 Jane Doe로 표시하고, 여러 명일 경우 Doe 1, Doe 2 식으로 표시한다.

우 저작자 표시(attribution), 저작권 고지, 라이선스 조건의 이행 등을 요구한다. 그러나 피고들이 개발한 Copilot은 Codex를 기반으로 하여 사용자가 입력하는 것에 따라 코드 블록을 생성하면서, 해당 코드가 이러한 라이선스 조건을 준수하지 않도록 하였다. 특히 Copilot의 결과물은 원본 코드에 포함되어 있는 저작권 귀속 등이 제거되어 제공되며, 이러한 코드는 마치 Copilot이 자체적으로 생성한 것처럼 제공된다. 또한 Copilot은 종종 특정 오픈소스 저장소의 코드와 동일하거나 추적 가능한 형태의 코드를 재현함에도 불구하고, 출처나 권리 정보를 포함하지 않은 상태로 결과물을 배포하고 있다. 이에 따라 원고들은 피고들이 저작권 관리정보(CMI) 규정, 연방 상표법, 캘리포니아주의 부정경쟁법 및 개인정보보호법, 오픈소스 라이선스와 관련한 계약 등을 위반하였다고 주장하였는데, CMI에 관한 부분만을 살펴보기로 한다.

2. CMI

피고들이 사용한 자료에는 저작권 고지, 제호, 저작자 명칭, 저작권자 명칭, 이용허락조건, CMI를 나타내는 번호나 기호 등 CMI를 포함하고 있다.[12] 피고들은 Github에 업로드된 코드를 Copilot에 포함시킴으로써 CMI를 제거·변경하였고, CMI가 제거·변경되었다는 것을 알고서 CMI 및 자료 복제물을 배포하였고, 저작권 침해를 유도한다는 것 등을 알고 있었고, CMI가 제거됨으로써 사용자들이 저작권을 침해하지 않고 자료를 이용할 수 없었으며, Copilot의 결과물을 통하여 허위의 CMI를 제공·배포하여 저작권 관리규정을 위반하였다.

(3) CMI 위반에 대한 법원의 판단

1. 법원은 피고들이 결과물 생성 과정에서 코드에서 CMI를 제거하도록 의도적으로 프로그램을 설계하였고, 고의성(scienter)[13]을 가지고 있었고, 변경된 CMI의 배포(1202(b)(2))에 관해서는 충분히 주장하지 않았다고 판단하였다. 이에 따라 법원은 피고가 1201(b)(1) 및 (3)을 위반하였다는 원고 주장에 대하여 피고들이 기각할 것을 청구한 것에 대하여 기각을 거부하였고, 1201(b)(2) 위반 주장에 대해서는 원고가 소장을 수정할 수 있도록 허가하였다(2023.5.11.).

2. CMI 규정은 ①CMI의 삭제·변경을 금지하고(1202(b)(1)) ②삭제·변경된 저작물, 저작물의 복제물, 음반의 배포 등을 금지하고(1202(b)(3)), ③CMI 자체의 배포 등을 금지하고 있다(1202(b)(2)). ①과 ②의 경우에는 부착되었던 CMI가 삭제·변경될 저작물을 대상으로 하는데, 이 규정이 AI가 생성한 결과물에도 적용되는가, 곧 '결과물이 수정된 형태의 저작물인 경우'에도 적용되는지 문제된다. 이 케이스에서 원고들이 수정된 소장에서 주장한 것은 Copilot은 '동일한 복제물(identical copy)'이 아니라 원본 자료를 수정한 결과물을 제공하였다는 것이었다. 따라서 피고는 동일한 복제물이 아니므로 피고가 1202(b)(1) 및 1202(b)(3)을 위반했다는 원고의 주장은 성립할 수 없다고 주장하였다. 법원은 저작물이 동일하지 않으면 CMI 규정이 적용되지 않는다고 하여 피고들의 주장을 인용하였는데, 다만 원

12) 미국 저작권법은 저작권 관리규정(CMI)의 위반 요건으로 첫째, 저작권자의 허락이나 법률에 의하지 아니하고 ㉮저작권관리정보를 고의로 제거·기타 변경하거나, ㉯제거·변경되었다는 것을 알면서 그 제거·변경된 상태의 저작권관리정보를 배포·수입하거나, ㉰제거·변경되었다는 것을 알면서 저작물·저작물의 복제물·음반을 배포·수입·공연하는 것을 금지하고 있다. 둘째, CMI의 제거, 변경, 배포행위가 저작권 침해를 유도·가능하게·촉진·은폐하게 되다는 것에 대한 고의(손해배상의 경우에는 과실로 충분)가 있어야 한다. §1202. 이 규정 위반요건으로는 2중의 주관적 인식(double scienter)가 필요한데, 곧 ①CMI를 제거·변경했다는 사실이나 제거·변경된 상태라는 사실에 대한 인식과 ②이러한 행위에 의하여 저작권 침해가 유도되는 것 등에 대한 인식(또는 과실)이다.

13) 고의성(scienter)은 부정행위(wrongdoing)에 대한 의도(intent)나 인지를 나타내는 것으로서, 불법성을 인지하거나 진실을 무모할 정도로(reckless) 무시하여 행동하였다는 것을 나타내는 정신상태를 의미한다. 주로 사기 사건이나 의도가 중요한 요소인 사건에 책임을 입증하기 위하여 필요한 요건이다.

고들이 소장을 수정할 수 있도록 허가하였다(2024.1.22.).

3. 법원은 수정된 소장에서도 원고들이 CMI에 관한 동일성 요건을 충족하지 못하였다고 하여 이전 판단과 동일하게 판단하였고, 이에 따라 원고의 1201(b) 위반 주장을 기각하였다(2024.9.27.).

9. Vacker v. ElevenLabs, Inc.

- 델라웨어 연방지방법원 (1:24-cv-00987, 2024.8.29.)
- 원고 Karissa Vacker와 Mark Boyett는 성우이고, Brian Larson 등은 성우들이 내레이션한 저작물의 저작(권)자 및 출판사이다. 피고 ElevenLabs, Inc.는 자신의 웹사이트와 API를 통하여 텍스트-음성 변환 서비스를 제공하면서 텍스트를 합성 오디오 내레이션으로 생성할 수 있도록 한다. 또한 피고의 시스템은 AI를 이용하여 음성을 복제·생성하고 다국어 음성-대-음성으로 번역하는 서비스를 제공하고 있다.

(1) 합의에 의한 AI 분쟁의 첫 종결

당사자들은 2025.8.18. 조정에 의하여 합의에 이르게 되었고, 소송은 2025.11.6. 취하되었다. 당사자들이 합의한 내용은 공개되지 않았지만, 이 케이스는 미국에서 제기·진행되고 있는 AI 소송에서 합의에 의하여 종결된 첫 번째 케이스이다. Bartz v. Anthropic PBC 케이스(3:24-cv-05417, N.D.CA. 2024.8.19.)는 공정이용에 관한 판결이 나온 이후 당사자들이 합의하였고 이에 따라 합의 내용에 대하여 법원이 승인하였고 이를 이행하는 절차가 진행되고 있지만, Vacker v. ElevenLabs, Inc. 케이스에서는 소송이 진행되는 동안 당사자들이 합의한 이후 소송을 취하한 것이어서, 양자는 차이점이 있다.

(2) 사실관계(원고 주장)

원고들은 성우, 저작물의 저작자, 출판사이고, 성우가 저작물을 내레이션하고 내레이션의 결과물인 오디오북을 발행·판매하고 있다. 이 사건에서는 서적인 어문저작물 및 내레이션에 의하여 작성되는 음향저작물(sound recording)과 퍼블리시티권의 보호대상인 '음성'이 관계된다. 피고는 Vacker와 Boyett의 오디오북 내레이션을 바탕으로 AI를 학습시켰고, 가공인물인 Bella와 Adam이라는 이름으로 서비스를 제공하는데 이들의 음성은 원고 Vacker와 Boyett의 음성과 실질적으로 유사하다. 피고는 원고의 음성을 이용하여 텍스트-음성 변환, 음성 복제, 다국어 텍스트-음성 변환, 음성-대-음성 번역, 더빙, API 제공, 목소리의 특성을 조절하는 인터페이스 등의 서비스를 제공하고 있다.

(3) 위반(원고 주장)

1. 퍼블리시티권 및 모습·동일성의 부당사용 등 인격권 침해[14]

피고는 원고 Vacker(텍사스주 주민)와 Boyett(뉴욕주 주민)의 음성이나 음성의 복제물을 이용하여 서비스를 제공함으로써, 이들의 모습(likeness)과 동일성을 고의에 의하여 부당하게 이용하였고 따라서 퍼블리시티권 등을 침해하였다.[15]

14) 프라이버시 침해와 부당이득도 주장하고 있지만, 퍼블리시티권만 살펴보기로 한다.

2. 기술적 보호조치 규정 위반

첫째, 원고들의 오디오북은 이용허락받은 복제물만 읽을 수 있도록 함으로써 접근을 통제하고 있는데, 피고가 학습을 위하여 오디오 파일을 준비하기 위해서는 파일을 복호화하고 DRM 보호를 무력화하여야 하는데, 이는 접근통제 무력화 금지규정(§1201(a)(1))의 위반에 해당한다.

둘째, 피고는 DRM 보호를 제거하기 위한 절차를 만들고, 오디오 및 텍스트를 보호되지 않는 포맷으로 변환시켰는데, 이는 접근통제 무력화 도구의 거래금지규정(§1201(a)(2))의 위반에 해당한다.

3. 저작권 관리정보 규정 위반

원고들의 오디오북 내레이션에는 저작물, 저작자, 내레이터, 저작권자 등의 정보를 식별하는 메타데이터 등의 CMI가 포함되어 있다. 피고는 학습데이터로 사용하기 위하여 원고의 디지털 오디오 파일에서 CMI를 제거·변경하였는데, 이는 CMI의 제거·변경 금지 및 제거된 저작물 등의 배포·수입·공연 금지규정(§1202(b)(1), (3))의 위반에 해당한다.

15) 이 사건에서는 피고가 AI를 학습시키기 위하여 특정인의 음성을 이용하였고, 특정인의 음성으로 들리는 소리를 만들어내는 서비스를 제공하고 있다. 따라서 이 사건에서는 일반적인 퍼블리시티권 외에 특정인의 음성·모습이라고 쉽게 식별할 수 있는 AI 생성 표현물인 디지털 모사물(digital replica)과 관계된다. 원고가 퍼블리시티권 침해에 대한 구체적인 법률, 조문, 판례를 언급하지 않고 원고(성우)가 주민으로 있는 뉴욕주 성문법 및 텍사스주의 퍼블리시권에 관한 입법 및 커먼로를 위반하였다고만 주장하고 있다. 뉴욕주는 민권법(Civil Rights Laws, CVR)을 통하여 성명, 초상(portrait), 사진, 모습, 목소리를 보호하고(§§50, 51), 사자(死者)에 대하여 디지털모사권(digital replica right)을 인정하고 있다(§50-(f)). 텍사스주는 사자의 성명, 음성, 서명, 사진, 모습을 성문 입법에 의하여 보호하고 있는데(§§26.001-26.015), 생존한 사람의 성명이나 모습 등 동일성(identity)은 판례에 의하여 보호되고 있다.

참고자료

- https://www.courtlistener.com/docket/69921640/millette-v-openai-inc/
- https://www.courtlistener.com/docket/17131648/thomson-reuters-enterprise-centre-gmbh-v-ross-intelligence-inc/
- https://www.courtlistener.com/docket/69636122/advance-local-media-llc-v-cohere-inc/
- https://www.courtlistener.com/docket/69456658/pierce-v-photobucket-inc/
- https://www.courtlistener.com/docket/68325564/onan-v-databricks-inc/
- https://www.courtlistener.com/docket/67599029/in-re-google-generative-ai-copyright-litigation/
- https://www.courtlistener.com/docket/69280523/dow-jones-company-inc-v-perplexity-ai-inc/
- https://www.courtlistener.com/docket/65669506/doe-1-v-github-inc/
- https://www.courtlistener.com/docket/69111793/vacker-v-elevenlabs-inc/