LEGAL UPDATE

AI

Aug. 2024

미국 법원의 스크래핑, 크롤링, 데이터 이용 관련 판결 동향

- 최근 선고된 콘텐츠 제공 기업과 기술 기업 간 분쟁 사례를 중심으로 -

생성형 AI 개발을 위한 웹 크롤링・스크래핑이 대규모로 이루어지며 콘텐츠 제공 기업과 기술기업 사이의 갈등이 심화되고 있습니다. 특히 생성형 AI 모델이 활발하게 연구・개발되는 미국에서는 유사 분쟁 사례에 대한 지속적인 법원 판결이 선고되고 있는 바, 그 과정에서 학습데이터를 확보하기 위한 웹 크롤링・스크래핑의 위법성 여부가 다양한 관점에서 다투어지고 있습니다.

이에 따르면 AI 학습이나 서비스 제공을 위하여 크롤링・스크래핑을 하는 경우 단순히 저작권 문제만이 아니라 형사법적 쟁점, 라이선스 위반 등 다양한 쟁점에 대한 폭넓은 검토가 필요하다는 점을 알 수 있습니다. 국내 지식재산권 법제 내지 관련 담론이 미국의 선례를 많이 참고하여 이루어진다는 점을 고려하면 위와 같은 판결들은 향후 국내 분쟁 사례에서 참고할 수 있는 중요한 선례가 될 수 있습니다.

1. **Ryanair vs. Booking.com 사건 – 컴퓨터 사기 및 남용 방지법(CFAA) 위반 인정**
2. **GitHub Copilot 집단소송 기각 – 부당이득・손해배상, 디지털 밀레니엄 저작권법(DMCA) 위반, 라이선스 위반 모두 불인정**
3. **Kadrey vs. Meta Platforms 사건 – 저작권 침해 불인정(계속 중)**
4. **Thomson Reuters vs. Ross Intelligence 사건 – 저작권 침해 인정**
5. **시사점**

1. Ryanair vs. Booking.com 사건 (2024.07.19) – 컴퓨터 사기 및 남용 방지법(CFAA) 위반 인정[1]

2024년 7월 19일, 델라웨어 지방법원 배심원단은 Booking.com이 Ryanair 웹사이트를 불법적으로 스크래핑했다고 판결했습니다. 해당 사건에서 Booking.com은 예약 서비스 제공을 위하여 Ryanair 웹사이트를 지속적으로 스크래핑하였는데 배심원단은 이에 대하여 Booking.com이 미국 컴퓨터 사기 및 남용 방지법(CFAA)을 위반했다고 인정하였습니다. 배심원단은 판단의 전제로서 Booking.com이 "사기의 고의"를 가지고 있고, Ryanair가 이로 인해 경제적 손해를 입었다는 점을 인정하는 한편, Booking.com이 Ryanair에 대하여 제기한 명예훼손, 불공정 경쟁, 기만적 상거래 관행을 이유로 한 반소를 모두 기각하였습니다.

[1] https://corporate.ryanair.com/news/ryanair-wins-its-case-against-travel-giant-booking-com-in-delaware-court-ruling/

본 내용은 일반적인 정보의 제공만을 목적으로 발행되므로 이에 수록된 내용은 저희 법인의 공식적인 견해나 구체적 사안에 관한 법률의견이 아님을 밝혀드립니다.

2. GitHub Copilot 집단소송 기각 (2024.06.24) – 부당이득・손해배상, 디지털 밀레니엄 저작권법 (DMCA) 위반, 라이선스 위반 모두 불인정[2]

2024년 6월 24일, 샌프란시스코 연방 지방법원의 Jon S. Tigar 판사는 GitHub의 AI 코딩 도우미 'Copilot'에 대한 집단소송을 대부분 기각했습니다. 이 소송은 변호사이자 오픈 소스 프로그래머인 Matthew Butterick가 Joseph Saveri 법률 사무소와 함께 GitHub가 공개 코드 저장소의 코드를 무단으로 활용하여 Copilot을 훈련시켰다고 주장하며 제기되었습니다.

이 판결에서는 크게 (i) 부당이득 및 징벌적 손해배상, (ii) 디지털 밀레니엄 저작권법(DMCA) 위반, (iii) 오픈소스 라이선스 위반이 각 문제되었습니다. 판사는 (i)에 대하여 원고들이 본래의 코드가 생성 코드에서 동일하게 재현되었다는 것을 입증하지 못했으며, 그에 따라 부당이득과 징벌적 손해배상 청구를 뒷받침할 충분한 증거를 제공하지 못했다고 판단하여 원고의 청구를 기각하였습니다. Copilot이 코드에서 저작권 관리 정보를 제거하여 DMCA를 위반하였다는 (ii) 주장에 대하여, 판사는 Copilot이 제안한 코드가 개발자의 원본 작업과 충분히 동일하지 않다고 판결하여 해당 청구를 기각하였습니다. 한편, Copilot이 적절한 출처 표시 없이 코드를 사용하여 오픈 소스 라이선스를 위반했다는 취지의 (iii)에 대하여 판사는 별도의 기각 판결을 내리지 않았는데, 이는 향후 AI 개발에 있어 오픈소스 코드 사용에 대한 법적 논란이 지속될 수 있음을 시사합니다.

3. Kadrey vs. Meta Platforms 사건 (2023.11.20) – 저작권 침해 불인정(계속 중)[3]

2023년 11월 20일, Kadrey를 포함한 저자들이 Meta의 대규모 언어 모델 LLaMA가 자신의 저작물을 무단으로 이용해 생성형 AI를 트레이닝했다며 저작권 침해 소송을 제기한 사건에 대한 판결이 선고되었습니다. Meta는 해당 이용이 작품을 그대로 복사한 것이 아니라 그것을 토대로 생성형 AI를 트레이닝하여 새로운 결과물을 도출해내는 용도였다고 주장하며 자사의 저작물 이용이 변형적 이용(transformative use)에 해당한다는 공정이용 주장을 펼쳤습니다.

이에 대하여 원고는 LLaMA 언어모델 자체가 저서로부터 추출한 정보가 없으면 작동하지 않기 때문에 저작권을 침해하는 2차적 저작물(infringing derivative work)이라고 주장하였으나, 재판부는 본 사건이 단순히 저작물을 그대로 재생산하는 상황이 아니기에 저작권을 침해하는 2차적 저작물이라고 단정지을 수 없다고 보아, 저작권을 침해하는 2차적 저작물이 되기 위해서는 AI 결과물이 저작물과 유사해야 한다고 판시하였습니다.

다만 해당 사건에서 원고의 저작권 침해 및 공정 이용 관련 내용은 상기 소송에서 완전히 종결되지 못하였기에, 원고는 2023년 12월에 수정된 소장을 제출했습니다.

4. Thomson Reuters vs. Ross Intelligence 사건 (2023.09.25) – 저작권 침해 인정[4]

2023년 9월 25일, Thomson Reuters Enterprise Centre GmbH와 Ross Intelligence Inc. 간의 소송에 대한 판결이

---

2 https://www.infoworld.com/article/2515112/judge-dismisses-lawsuit-over-github-copilot-ai-coding-assistant.html; https://www.theregister.com/2024/07/08/github_copilot_dmca/

3 Ibid.

4 Hannah Albarazi, "These Are The High-Stakes AI Legal Battels To Watch In 2024", LexisNexis, 2024. 7. 2. 참고

본 내용은 일반적인 정보의 제공만을 목적으로 발행되므로 이에 수록된 내용은 저희 법인의 공식적인 견해나 구체적 사안에 관한 법률의견이 아님을 밝혀드립니다.

내려졌습니다. 이 사건에서는 Ross의 생성형 AI가 Westlaw 법률 리서치 데이터베이스로부터 허가를 받지 않고 무단으로 자료를 이용해 AI 알고리즘 트레이닝을 진행하여 문제가 되었습니다. Westlaw는 Ross의 생성형 AI를 이용하는 경쟁 법률 검색엔진이 188개의 헤드노트, 각주, 요약 포인트 등을 차용했다고 주장했으며, 경쟁사도 이에 대한 이의를 제기하지 않았습니다.

주요 쟁점은 Ross의 생성형 AI가 저작권으로 보호되는 자료를 트레이닝하는 데 이용하는 것이 저작권법상 "공정이용(fair use)"에 해당하는지 여부였습니다. 법원은 변형적 이용(transformative use)이 있었는지 여부에 대하여 Ross의 생성형 AI가 Westlaw 데이터의 언어 패턴만 학습하였는지 여부를 판단 기준의 하나로 고려하였고, 특히 해당 AI가 Westlaw 변호사 편집인들의 작업을 창의적으로 재생산한 경우 변형적 이용이 아니라고 보았습니다. 아울러, 법원은 변형적 이용의 판단에 저작물의 이용이 향후 저작물의 시장에 미치는 영향 등을 고려해야 한다고 보았습니다.

결과적으로, 법원은 침해 및 데이터 이용이 향후 저작물 시장에 미치는 요소에 있어서 사실관계 확인이 필요하다고 판단하여 피고의 공정이용 주장을 기각했습니다.

5. 시사점

근래 미국에서 선고되고 있는 판결들은 전통적인 저작권 침해 쟁점은 물론 사기 등 형사법적 쟁점까지 폭넓게 다루고 있습니다. 이에 따르면 AI 학습이나 서비스 제공을 위하여 크롤링・스크래핑을 하는 경우 단순히 저작권 문제만이 아니라 형사법적 쟁점, 라이선스 위반 등 다양한 쟁점에 대한 폭넓은 검토가 필요하다는 점을 알 수 있습니다.

국내 지식재산권 법제 내지 관련 담론이 미국의 선례를 많이 참고하여 이루어진다는 점을 고려하면 위와 같은 판결들은 향후 국내 분쟁 사례에서 참고할 수 있는 중요한 선례가 될 수 있습니다. 따라서 국내 AI 기업과 데이터 기반 서비스 제공업체 역시 데이터 수집 및 이용 과정에서 법적 리스크를 면밀히 검토하되 이러한 선례를 참고할 필요가 있으며, 필요한 경우 라이선스 확보나 명시적 동의 획득 등의 조치를 취해야 할 것입니다.

화우 AI센터는 AI와 관련한 지식재산, 개인정보, 정보보안, 공정거래, 제조물책임, 입법컨설팅, 쟁송 등 모든 법적 영역에서 축적된 경험과 노하우를 기반으로 기업 고객을 위한 최적의 솔루션을 안내해드리고 있습니다. AI와 관련하여 문의사항이 있으신 경우 언제든지 화우에 연락하여 주시기 바랍니다.

Contacts

이광욱 파트너변호사 **T.** (+82) 2 6003 7535 **E.** kwlee@yoonyang.com

이근우 파트너변호사 **T.** (+82) 2 6003 7558 **E.** klee@yoonyang.com

이수경 파트너변호사 **T.** (+82) 2 6182 8132 **E.** sgyi@yoonyang.com

정호선 변호사 **T.** (+82) 2 6182 8548 **E.** junghs@yoonyang.com

강석준 변호사 **T.** (+82) 2 6182 8505 **E.** kangsj@yoonyang.com

배종우 변호사 **T.** (+82) 2 6182 8745 **E.** jwbai@yoonyang.com