

[영국] 상원 통신 및 디지털위원회 LLM 관련 의견수렴 보고서

한국저작권위원회 법제연구팀
유혜정 연구원

1. 개요

영국 상원 내 통신 및 디지털위원회(Communications and Digital Committee, 이하 '디지털위원회')는 향후 3년간 대형언어모델(Large Language Model, 이하 'LLM')이 야기할 수 있는 기회와 위험에 대응하기 위해 정부가 해야 할 일들을 파악하는 것을 목적으로, 2023년 7월 7일부터 9월 5일까지 산업, 기술, 권리자 등 다양한 이해관계자들의 의견을 수렴했음.¹⁾ 그리고 2024년 2월 2일, 해당 의견수렴의 주요 내용을 요약하고 이에 대한 위원회의 의견을 추가한 'Large Language models and generative AI' 보고서를 공개했음.

2. 의견수렴 주요 내용

(1) 이용자

스테빌리티 AI는 텍스트-데이터 마이닝(Text Data Mining, 이하 'TDM') 수행 전 이용허락을 받는 것은 AI 개발을 억제하기 때문에 AI 개발의 중심지를 AI에 친화적인 국가로 이동시킬 것이라고 주장함. 구글 딥마인드는 AI가 저작물을 직접적으로 복제한다면 이를 방지하기 위한 적절한 조치를 취할 것이라고 언급했음.

마이크로소프트는 누구에게나 읽고, 배우고, 이해할 수 있는 권리가 있어야 하고, 영국 저작권법은 이를 위해 기술을 도구로써 이용하는 것을 허용하는 예외를 포함하고 있기 때문에 AI 학습에 문제가 없다고 주장함. Aleph Alpha는 기술적으로 AI가 학습한 원본의 특정 단어나 문장을 추적할 수 없다고 주장함.

(2) 권리자

게티이미지는 옵트아웃 방식²⁾에 대해 "사후에 용서를 구하는 것"이라고 비판했으며, "저작권법의 사전 허락 원칙에 근본적으로 반대하는 것"이라고 강력하게 주장했음.

Publishers' Licensing Services³⁾는 AI가 이미 학습한 것들을 되돌리는 것이 쉽지 않기 때문에 옵트아웃 방식이 실용적이지 않다고 언급함. Publishers' Association⁴⁾은 공개적으로 알려진 book3 데이터셋⁵⁾이 명백한 저작권 침해의 증거라고 주장함.

1) 서면 119건, 대면 13회

2) 저작물 사용에 대해 묵시적 이용허락을 받았다고 간주하되 추후 저작자가 원치 않을 시 이용허락을 철회할 수 있게 하는 방식

3) 도서, 잡지, 저널 등의 출판사를 대표하는 영국의 비영리기관으로, 출판사들의 국내외 수익을 대리하는 집중관리단체

4) 도서, 저널, 이북 등의 출판사를 회원사로 두고 출판 무역 진흥을 목적으로 하는 조직

5) 약 19만 6,640권의 도서 데이터를 포함하고 있는 오픈 소스 AI 학습 데이터셋으로 저작권자의 허락을 받지 않은 도서들이 다수 포함되어 있음

(3) 학계

Andres Guadamuz⁶⁾박사는 책의 텍스트가 수많은 토큰들로 변환되고 최종 모델은 원 학습데이터의 통계적인 표현(statistical representations)만을 포함하고 있기 때문에, LLM이 반드시 저작물을 '갖고 있는 것(hold)'은 아닐 수 있다고 주장함. 반면에 Hayleigh Bosher⁷⁾박사는 영국 저작권법상 일시적 또는 부수적 복제를 허용하는 예외 범위가 좁기 때문에 LLM은 해당 예외에 적용되지 않는다고 주장했음.⁸⁾ 또한 Hayleigh 박사는 LLM이 명백하게 상업적인 목적으로 저작물을 이용하기 때문에 인간이 저작물을 학습하는 것과 비교해서는 안 될 것이며, 저작권법은 기술이 어떻게 작동하는지 파악할 것이 아니라 미래에 어떻게 적용될 수 있는지에 기반해야 한다고 언급함.

3. 보고서 주요 내용

위원회의 보고서는 영국이 AI 분야에서 세계적 경쟁력을 확보하기 위해서는 LLM에 대한 긍정적인 비전이 필요하기 때문에 AI 스타트업 지원, 인프라 구축 및 기술 개발 등이 필요하며, 영국만의 독자적인 LLM 개발을 고려해야 할 필요가 있다고 주장함.

위원회 보고서는 저작권과 관련해서 필요하다면 정부가 법률 제·개정을 통해 대형 기술 회사들이 저작권자의 허락 또는 보상 없이 저작물을 사용하여 발생하는 분쟁을 종결시켜야 한다고 주장함. 구체적으로는 저작권자가 AI 학습데이터를 확인할 수 있도록 해야 하고, AI 기업은 웹 크롤링 대상을 밝힐 것을 요구하고 있음. 또한 기업들이 라이선스 취득에 대한 보상을 지불할 수 있도록 라이선싱 단체 및 데이터 저장소 소유자와 협력하여 양질의 데이터셋에 투자하고, 공정한 관행 형성을 위해 조달 시장(Procurement market)을 이용할 것을 방안으로 제시함.

상원 보고서는 2024년 초에 공개될 예정인 AI 저작권 행동강령(The government's code of practice on copyright and AI)⁹⁾이 저작권자의 권리를 온전하게 행사할 수 있도록 해야 한다고 밝혔음. 만약 해당 행동강령이 2024년 봄까지 발표되지 않는다면, 정부는 반드시 법률 제·개정을 포함한 분쟁해결 방법을 준비해야 한다고 주장함. 현재의

6) 영국 서식스 대학교(Sussex University) 지식재산권법 교수

7) 영국 브루넬 법학대학교(Brunel Univeresity Law School) 부학장 및 지식재산권법 교수

8) 영국 저작권법 28A조

제28A조(일시적 복제물의 제작)

컴퓨터프로그램이나 데이터베이스를 제외한 어문저작물 또는 연극, 음악 또는 미술 저작물, 발행물의 판면배열, 녹음물 또는 영화의 저작권은 순간적이거나 부수적이며, 기술적 과정의 필수적이고 본질적 부분이고 다음을 가능하게 하는 것이 유일한 목적인 일시적 복제물의 제작으로 인하여 침해되지 아니한다-

- (a) 중개에 의한 제3자 사이의 네트워크상 저작물의 송신; 또는
- (b) 저작물의 합법적인 사용;

그리고 독립적인 경제적 중요성을 가지고 있지 않아야 한다.

9) 영국 지식재산청(Intellectual Property Office)은 2023년 6월 29일부터 AI 산업 내 저작자, 기업체, 학계 등을 대표하는 이해관계자들과 함께 워킹그룹을 구성하여 AI 저작권 분야 내 자율적인 행동강령 발표를 준비 중

저작권법이 저작자에게 충분한 보호를 제공하고 있는지 확인하고, 만약 불명확하다면 법률 개정을 통해 저작권 원칙이 미래에 대비하고 기술 중립적일 수 있는 방안을 마련해야 한다고 밝힘.

4. 시사점

이번 의견수렴은 AI 산업 내 저작물의 이용자인 기업과 권리자간의 이해관계가 상반된다는 것을 확인한 보고서로, 국내에서 논의되고 있는 AI 산업 내 각 이해관계자들의 의견과 비슷한 양상을 보임. 영국 의회는 영국 정부에 대하여, 이해갈등 조정을 위해 법률 제·개정 등을 통해서라도 조기에 원만한 상생 방안을 마련하도록 촉구한 것으로 보임. 우리와 마찬가지로 영국에서도 AI-저작권 문제를 해결하기 위한 다양한 고민을 하고 있는 것으로 해석할 수 있음.

참고자료

《링크1》 <https://committees.parliament.uk/work/7827/large-language-models/>