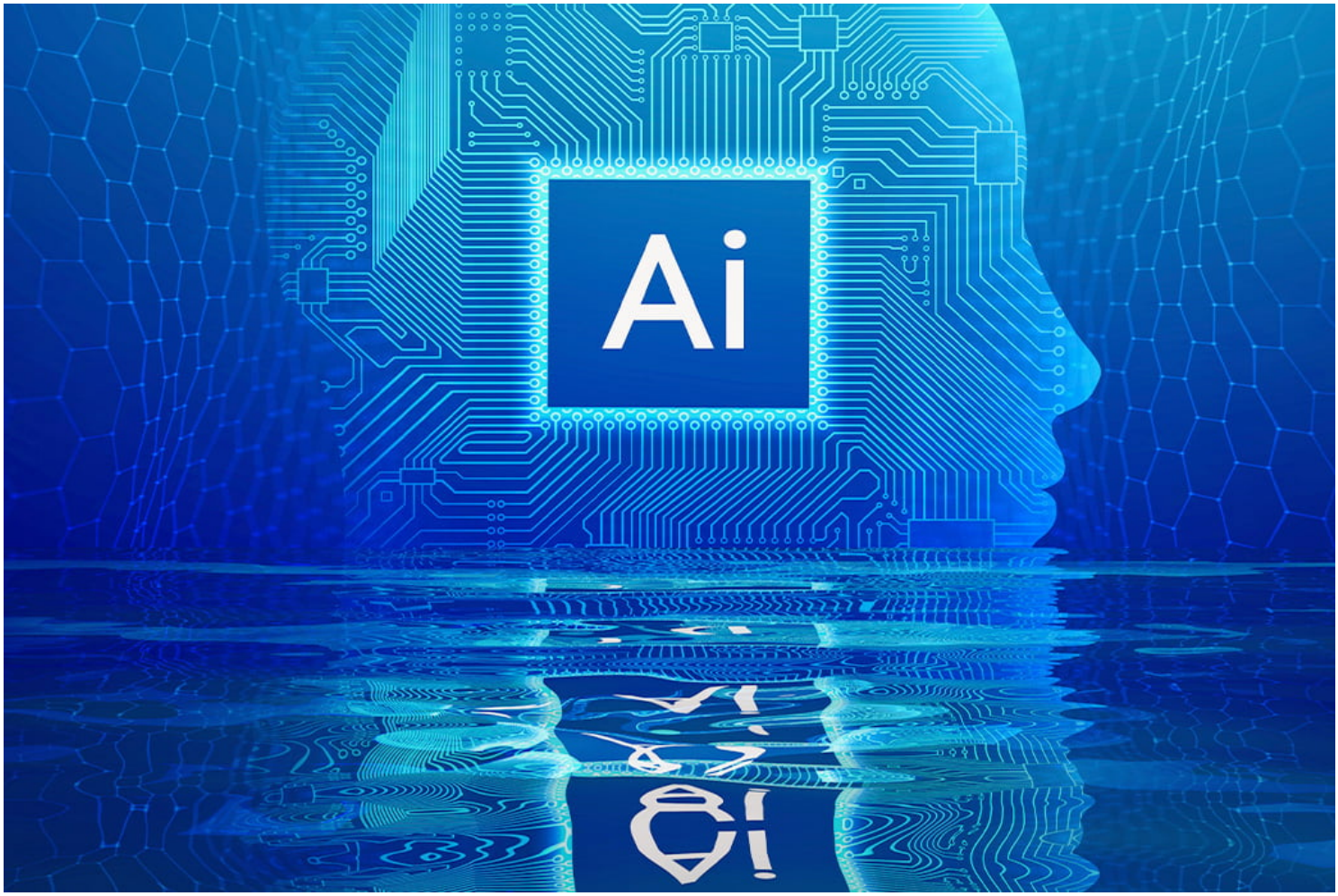
Newmedia of the world

뉴스의 공정 이용과 저널리즘 산업의 위기

글 이성규 (미디어스피어 대표)



갈등이 첨예해지고 있다. 갈등에 점점 날이 서고 있다. 오픈AI는 공정 이용을 강조하며 방어하고 있지만, 언론사들은 생존을 걸고 소송전을 시작하려는 움직임이다. 애초 IAC(인터액티브코퍼레이션)[1]를 위시한 미국 내 대형 언론사 간 연합이 소송의 포문을 열 것으로 예상됐지만, 오히려 뉴욕타임스가 좀 더 빠르게 나서는 형국이다[2]. 뉴욕타임스는 허락을 얻지 않고 자사 기사를 학습한 것과 관련, 지난 7월께부터 오픈AI와 보상 협상을 진행했지만 끝내 합의에 이르지는 못했다. 오히려 협상 결렬이 기폭제가 되어 빠르게 소송 단계로 넘어갈 태세다.

뉴욕타임스 CEO인 메러디스 코핏 레비안은 지난 6월 “이미 사용된 콘텐츠와 앞으로 모델 학습에 계속 사용될 콘텐츠에 대해 공정한 가치 교환이 이루어져야 한다”는 명확한 입장을 밝힌 바 있다. 뉴스 데이터를 학습 용도로 사용한 데 따른 보상을 기술 기업들로부터 받아내겠다는 확고한 의지를 드러낸 것이다. 뉴욕타임스는 8월 GPTbot을 차단하면서 더 이상의 웹스크래핑이 불가하도록 조치도 취했다[3]. 오픈AI 쪽이 만족스러운 협상안을 제안하지 않는 이상 법정 다툼이 불가피한 단계로 넘어가는 중이다.

뉴욕타임스와 오픈AI가 소송전을 벌이게 된다면 이는 기술 기업과 저널리즘 기업의 대리전이 될 가능성이 높다. 최종 판결이 나오기까지 지난한 시간이 흘러가겠지만, 이 과정에서 의미 있는 해석과 대가 산정 방정식이 도출될 가능성도 적지 않다. 핵심 쟁점이 무엇이냐에 따라 다른 언론사들이 뒤이은 소송전에 참여할지 여부를 판단할 수도 있다. 전 세계 언론사들이 두 기업이 법정에 서길 학수고대하는 이유이기도 하다.

쟁점은 두 가지로 좁혀지고 있다. 한 가지는 허락 없이 뉴스 데이터를 수집하고 저장한 학습용 데이터세트 구축을 공정 이용으로 볼 것인가 아닌가이다. 일반적으로 학습용 데이터세트는 웹상에 공개된 뉴스나 콘텐츠를 스크래핑한 뒤 기계가 학습 가능한 형태로 토큰화한 거대한 데이터 꾸러미를 지칭한다. 예를 들어 뉴욕타임스의 기사를 스크래핑한 다음 이를 적절한 단위로 쪼개거나 해체하고, 숫자 형태인 벡터로 변환해 저장한 저장소가 데이터세트다[4]. 미국신문협회 격인 뉴스미디어얼라이언스는 공정 이용에 해당하는 4가지 요소에 데이터세트는 부합하지 않는다는 의견을 낸 바 있다. 반면 오픈AI 쪽은 데이터세트를 ‘비표현적 복제’라며 공정 이용의 범주 안에 있다고 반박하고 있다.

또 다른 쟁점은 오픈AI 등 거대 언어모델이 생성한 정보가 뉴스 소비를 대체하고 언론 산업 전반을 위험에 빠뜨리는지 여부다. 현재 글로벌 언론사들은 이 점을 가장 위협적으로 느끼고 있다. 'ChatGPT-4'나 구글 '바드' 등은 사용자가 요청한 뉴스의 세부 내용 등을 요약해주거나 설명하는 데 꽤나 높은 성능을 발휘하고 있다. 이를 통해 과거나 최근 뉴스를 생성 AI를 통해 소비하는 게 가능해지는 상황이다. 이러한 소비 행태가 보편화되면 굳이 언론사 웹사이트를 방문해 뉴스를 소비할 동기가 사라질 수도 있다. 구글은 최근 개편한 구글 생성 AI 통합검색 화면에서 언론사 등으로 넘어갈 수 있는 링크를 점차 늘리고 있지만, 언론사들의 불안감을 불식시키기엔 여전히 부족해 보인다. 일각에선 검색을 통해 언론사로 유입되는 트래픽이 상당 부분 감소할 것이라고 예측하고 있다. 이를 염두에 둔 듯 뉴욕타임스 담당 변호사들은 "오픈AI가 신문사의 기사를 활용하여 뉴스 사건에 대한 설명을 뱉어내는 것은 공정 이용으로 보호되어서는 안 되며, 신문사의 보도를 대체할 위험이 있다"고 주장하고 있다[5].

현재 전 세계 저널리즘 산업은 생성 AI로 몸살을 앓고 있다. 생성 AI의 혁신성은 인정하면서도 이로 인해 파생될 산업의 위기를 걱정하고 있다. 가뜩이나 디지털 광고 수익을 빅테크 기업에 빼앗긴 상

황이어서, 새로운 수익원 발굴이 절실한 시점이다. 게다가 AI 기업에 대한 공정 이용이 폭넓게 인정이 되면, 이를 차단하기 위해 어쩔 수 없이 유료장벽을 올리고, 스크래핑 봇을 틀어막는 조치를 취할 수밖에 없다. 더 이상 광고 수익만으로는 고품질의 저널리즘 생산이 불가능해진 상황에서 구독료를 받고, 저작권 라이선스 수익을 얻을 수 있는 조건을 구축해 대응할 수밖에 없는 처지다. 저널리즘 산업의 지속가능성과 수익을 보전하기 위한 이러한 선택은 결과적으로 정보 소비의 빈익빈 부익부를 낳을 수밖에 없다. 고품질 정보에 접근할 수 있는 지불의 장벽들이 계속 높아질 수밖에 없어서다.

국내 언론 상황도 미국과 별반 다르지 않다. 네이버의 '클로바X'와 생성 AI 검색 서비스인 'Cue:[6]'가 공개되면서 언론사들의 불안감은 서서히 커지고 있다. 몇몇 분야에서 기대에 못 미치는 성능에 위안을 삼는 이들도 있다. 하지만 어떤 방식으로든 수용자들의 뉴스 소비에 영향을 미칠 수밖에 없다는 것이 언론계의 중론이다.

물론 미국 언론사들처럼 소송을 검토할 단계까지는 넘어가지 않고 있다. 미국과는 다른 뉴스 제공 계약관행이 긴 시간 이어져왔기 때문이다. 과거 뉴스를 생성 AI의 학습 데이터로 이용한 사실이 그간의 계약에 위반되는가를 놓고 논쟁이 이어지고도 있다. 해외 상황을 예의주시하면서 국내 기술기업들에게 적절한 보상을 요구하려는 물밑 움직임도 계속되는 중이다. 그 갈등이 수면 위로 올라올 가능성이 점차 높아지고 있다.

모든 갈등에는 최선은 아니더라도 최적의 타협안은 있기 마련이다. 수익 없는 저널리즘은 존재할 수 없지만, 고품질 저널리즘 없는 생성 AI도 존재하기 어렵다. 사용자들이, 시민이 그리고 이 사회가 그것을 필요로 해서다. 타협 지점은 바로 이 사이 어딘가에 놓여있다. 이를 인정하면 쉬워진다. 어쩌면 법리적 다툼 속에서 새 길이 찾아질지도 모른다. 그것이 서로를 인정하는 방법이라면 거쳐야 할 경로일지도 모르겠다.

우리 「저작권법」 제1조는 이렇게 말하고 있다. "저작자의 권리와 이에 인접하는 권리를 보호하고 저작물의 공정한 이용을 도모함으로써 문화 및 관련 산업의 향상발전에 이바지함." 어쩌면 이 한 줄 속에 기술 기업과 저널리즘 기업 간 최적의 타협 지점이 존재할지도 모른다.

참조

[1] https://www.iac.com

[2] https://www.semafor.com/article/08/13/2023/new-york-times-drops-out-of-ai-coalition

[3] https://www.theverge.com/2023/8/21/23840705/new-york-times-openai-web-crawler-ai-gpt

[4] https://developers.google.com/machine-learning/guides/text-classification/step-3?hl=ko

[5] https://www.npr.org/2023/08/16/1194202562/new-york-times-considers-legal-action-against-openai-as copyright-tensions-swirl

[6] https://cue.search.naver.com/

#생성형AI #공정이용 #오픈AI #뉴욕타임스 #클로바X #Cue