

2026년
7호

KPF 미디어브리프

K P F Media Brief Webzine

뉴스 크롤러 신원과 목적 밝혀야

뉴욕주 의회, 스텔스 크롤러 금지 법안 통과

김선호 | 한국언론진흥재단 책임연구위원

- ✓ 언론사 사이트에 접근하는 크롤러의 식별 정보와 이용 목적 공개 의무화
- ✓ 식별 정보를 숨긴 채 손해를 끼친 '스텔스 크롤러'는 하루 최대 15,000달러 과징금 부과
- ✓ 국내에서도 크롤러 등록·접근 통제, 공동 탐지체계와 표준 라이선스를 결합한 뉴스 콘텐츠 보호 정책 논의 필요

* 스텔스 크롤러: 운영주체와 정보 수집 목적을 밝히지 않은 채 웹사이트의 콘텐츠를 자동으로 수집하는 소프트웨어



■ 스텔스 크롤러 금지 법안 제정 배경

생성형 인공지능과 AI 검색 서비스가 확산하면서 뉴스 콘텐츠의 생산·유통 질서가 크게 흔들리고 있다. 과거 검색엔진의 웹 크롤링은 언론사 기사를 색인한 뒤 이용자를 원문 사이트로 연결하는 방식이었다. 언론사는 콘텐츠 접근을 허용하는 대신 검색 유입과 광고·구독 수익을 얻을 수 있었다. 그러나 생성형 AI 서비스는 언론사 사이트에서 수집한 정보를 학습 데이터, 검색 결과의 근거 자료, 이용자 질문에 대한 요약 답변 등에 활용하면서도 원문 사이트로 충분한 트래픽을 돌려주지 않는다. 뉴스 콘텐츠를 수집하는 행위는 늘지만, 콘텐츠 생산자에게 돌아가는 경제적 보상과 이용자 유입은 오히려 감소하는 구조가 형성된 것이다.

문제는 일부 AI 사업자나 데이터 수집 업체가 자신의 신원과 크롤링 목적을 명확히 밝히지 않거나, 언론사가 robots.txt 등을 통해 표시한 접근 거부 의사를 우회해 콘텐츠를 수집한다는 데 있다. 일반 이용자의 브라우저나 검색엔진 크롤러로 가장하거나 IP 주소와 식별 정보를 수시로 변경하는 경우에는 언론사가 누가, 어떤 목적으로 콘텐츠에 접근하는지 파악하기 어렵다. 유료 구독자에게만 공개된 기사까지 수집되는 사례도 제기되고 있다.

이러한 문제에 대응해 미국 뉴욕주 의회는 2026년 6월 ‘스텔스 크롤러 금지 법안’(Stealth Crawler Prohibition Act)을 통과시켰다.¹⁾ 이 법안은 저작권 침해 여부를 직접 규정하기보다, 뉴스 콘텐츠에 접근하는 자동화된 크롤러의 신원과 이용 목적을 공개하도록 규정하고 이를 숨긴 크롤러가 언론사에 손해를 끼치는 행위를 금지한다. AI 시대 웹 크롤링의 투명성을 규제 대상으로 삼았다는 점에서 주목할 만한 법안이다. 법안은 뉴욕주 상·하원을 모두 통과했으며, 시행을 위해 뉴욕 주지사의 최종 서명만을 남겨둔 상태다.

■ 언론사 웹사이트 무단 크롤링의 문제점

데이터 거래 플랫폼인 톨비트(TollBit)의 2025년 2분기 조사에 따르면, 조사 대상 사이트에서 전체 방문 클릭 중 AI 봇·에이전트의 비중은 2분기 초 200건당 1건에서 2분기 말 50건당 1건으로 네 배 증가했다.²⁾ AI 봇의 요청 가운데 robots.txt의 접근 제한을 우회한 비율도 13.26%로, 2024년 4분기 3.3%의 네 배였다. 반면, AI 서비스는 전통적인 검색엔진과 비교해 언론사에 훨씬 적은 트래픽을 돌려주는 것으로 나타난다. AI가 기사 내용을 직접 요약해 제공하면 이용자는 원문을 클릭할 필요가 줄어들기 때문이다. 이른바 ‘제로 클릭’ 이용이 확산할수록 언론사는 취재비용을 부담하면서도 광고 노출, 구독 전환, 독자 데이터 확보의 기회를 잃는다. 이는 단순한 저작권 분쟁을 넘어 뉴스 생태계의 가치 배분 문제로 이어진다.

AI 크롤러는 용도에 따라 크게 세 가지로 구분할 수 있다. 첫째는 AI 기본모델의 사전학습이나 추가학습을 위한 학습용 크롤러, 둘째는 이용자의 질문에 답하기 위해 최신 웹 정보를 실시간으로 수집하는 검색용 크롤러, 셋째는 이용자를 대신해 정보를 탐색하고 작업을 수행하는 AI 에이전트용 크롤러다. 하나의 크롤러가 여러 용도로 사용되거나 수집된 데이터가 목적을 달리해 재사용되기 때문에 언론사가 최종적인 이용 목적을 파악하기 어렵다. 무단 크롤링에는 구체적으로는 다음과 같은 행위가 포함된다.

- 크롤러 운영자와 소프트웨어의 신원을 허위로 표시하거나 웹 브라우저로 가장하는 행위
- AI 학습, 검색 답변 생성, 데이터 판매 등 실제 수집 목적을 알리지 않는 행위
- robots.txt 접근 제한이나 IP 차단 등 언론사의 거부 의사를 우회하는 행위
- 유료 장벽이나 로그인 절차 뒤에 있는 기사까지 자동으로 수집하는 행위
- 짧은 시간에 대량 요청을 보내 서버와 보안 인프라에 과도한 부담을 주는 행위

1) <https://www.nysenate.gov/legislation/bills/2025/S9934/amendment/A>

2) https://tollbit.com/state-of-the-bots/q2-2025/?utm_source=chatgpt.com

- 제3의 데이터 브로커, 검색 결과 또는 웹 아카이브를 통해 차단된 콘텐츠를 우회 취득하는 행위
- 수집한 기사를 대체재가 될 정도로 길게 재현하거나 다른 SI 사업자에게 재판매하는 행위

무단 크롤링이 언론사에 미치는 피해는 크게 네 가지 항목으로 나뉘볼 수 있다. 첫째, 서버·네트워크 비용이 증가한다. 대규모 자동 요청을 차단하기 위해서는 요청을 일단 처리하고 분석해야 하므로 대역폭, 클라우드 이용료, 보안 솔루션 비용이 증가한다. 둘째, 광고와 구독 수익의 감소다. SI 서비스가 기사 핵심을 자체 화면에서 완결적으로 제공하면 언론사 사이트 방문과 유료 구독 가능성이 줄어든다. 콘텐츠 수집량과 원문 유입량이 일정 수준 비례했던 과거 검색 시대의 교환관계가 성립하지 않는 것이다. 셋째, 콘텐츠 라이선스 시장이 훼손될 우려가 있다. 크롤러의 신원을 숨기면 언론사는 이용을 허가하거나 거부할 대상을 특정할 수 없으며, 정당하게 사용료를 지급하는 사업자와 몰래 콘텐츠를 취득하는 사업자 사이에도 불공정 경쟁이 발생한다. 넷째, 언론사 브랜드를 약화시킨다. SI가 기사 내용을 부정확하게 재구성할 경우 이용자는 오류가 원 언론사의 보도에서 비롯된 것으로 오인할 수도 있다. 장기적으로는 신뢰할 수 있는 저널리즘과 출처가 불분명한 SI 정보의 구별이 어려워지고, 취재에 투자할 경제적 유인도 감소한다.

■ 스텔스 크롤러 금지 법안의 주요 내용

스텔스 크롤링을 금지하는 뉴욕주 법안은 일반사업자법 개정안으로서 제1750조부터 제1755조를 신설하는 내용으로 구성되어 있다. 이 법안이 이용 허락을 받지 않은 모든 무단 크롤링을 금지하는 것은 아니다. 무단 크롤링 중에서도 신원과 이용 목적을 밝히지 않는 ‘스텔스 크롤러’가 사이트 운영에 부담이나 경제적 손해를 일으켰다면 이 법안의 규제 대상이 될 수 있다.

1. 적용 대상

법안은 ‘크롤러’를 웹사이트나 기타 인터넷 정보원에 접근해 정보를 검색·스캔·색인·스크랩하는 소프트웨어로 폭넓게 정의한다. 온라인 크롤러, 스파이더, 페처, 클라이언트, 봇, 사용자 에이전트뿐 아니라 SI 에이전트와 이에 준하는 도구도 포함한다.

보호 대상인 ‘뉴스 정보원(covered news source)’은 신문에 국한되지 않으며, 인쇄매체, 텔레비전, 라디오, 네트워크, 케이블, 위성 및 디지털 출판·서비스의 웹사이트나 관련 정보원 가운데 다음 조건을 모두 충족해야 한다.

- 전통적인 언론사에 상응하는 공적 정보 제공 기능을 수행할 것
- 취재·편집·제작·유통에 상당한 노동·기술·비용을 투입해 독창적인 텍스트·음성·사진·영상 콘텐츠를 생산할 것
- 매월 한 차례 이상 콘텐츠를 새로 발행하거나 갱신하고 오류 정정·해명 절차를 갖출 것
- 뉴욕주 내 월간 활성 이용자·시청자·청취자 또는 구독자가 1,000명 이상일 것

이 정의는 언론사 보호 대상을 기업 규모나 매체 형식보다 ‘원본 콘텐츠 생산에 대한 투자’와 ‘편집·정정 체계’를 중심으로 설정했다는 점에서 의미가 있다.

2. 신원과 이용 목적의 공개 의무

뉴스 정보원에 접근하는 크롤러는 접근 시점 또는 그 이전에 신원과 이용 목적을 공개해야 한다. 구체적으로 사용자 에이전트 문자열에 다음 정보를 정확하게 표시해야 한다.

- 요청을 보내는 소프트웨어 제품의 명칭
- 소프트웨어의 버전
- 해당 소프트웨어를 운영하는 회사의 신원
- 수집한 콘텐츠가 사용될 수 있는 모든 용도와 목적

특히 단순히 “웹 크롤링”이라고 밝히는 것만으로는 충분하지 않고, AI 학습, 모델 미세조정, 검색 색인, 실시간 답변 생성, 에이전트 작업 등 콘텐츠의 구체적 이용 가능성을 언론사가 확인할 수 있는 형식으로 공개해야 한다. 이 의무를 이행하지 않는 크롤러가 법안상의 ‘스텔스 크롤러’다.

3. 금지되는 행위

법안은 스텔스 크롤러의 존재 자체를 일률적으로 처벌하지 않는다. 운영자가 스텔스 크롤러를 배포해 뉴스 정보원의 운영을 훼손·저해하거나 부담을 주는 경우, 또는 뉴스 정보원에 경제적 손해를 발생시키는 경우를 위법행위로 규정한다. 따라서 위반이 성립하려면 대체로 두 요소가 결합되어야 한다.

- 크롤러가 신원 또는 이용 목적 공개 의무를 이행하지 않았을 것
- 그 크롤링이 사이트 운영에 부담을 주거나 경제적 손해를 야기했을 것

이 구조는 공개된 웹 정보에 대한 모든 자동 접근을 금지하지 않으면서, 기만성과 피해가 결합된 행위를 규제하려는 절충적 접근이다. 그러나 ‘운영상 부담’과 ‘경제적 손해’의 범위가 넓어 실제 집행 과정에서 인과관계와 손해의 판단 기준이 쟁점이 될 수 있다.

4. 집행 절차와 제재

법안의 집행 권한은 뉴욕주 검찰총장(Attorney General)에게 부여된다. 검찰은 스텔스 크롤러 운영자가 법을 위반했거나 위반할 가능성이 있다고 판단할 만한 증거가 있는 경우, 법원에 소송을 제기할 수 있다. 검찰은 위법한 크롤링 행위를 중단시키기 위해 법원에 금지명령을 청구할 수 있으며, 본안 판결 전 피해 확산을 방지하기 위한 잠정적 구제 조치도 함께 요청할 수 있다.

법 위반이 인정되면 크롤러 운영자에게는 위반 행위별로 하루 최대 15,000달러의 과징금(civil penalty)이 부과될 수 있다. 이때 실제 피해의 발생과 규모를 입증할 필요는 없지만, 해당 스텔스 크롤링이 뉴스 사이트의 운영에 부담을 주거나 경제적 손해를 초래할 수 있는 행위였다는 점은 인정돼야 한다.

신원을 숨긴 크롤러 운영자를 추적하기 위한 별도의 절차도 마련됐다. 피해 언론사는 크롤러 운영자가 누구인지 알 수 없는 경우, 검찰이 소송을 제기하기 이전에 법원에 소환장 발부를 요청할 수 있다. 소환장을 받은 인터넷·통신 서비스 제공자는 보유한 범위에서 운영자의 신원을 확인할 수 있는 정보를 제공하고 관련 증거를 보존해야 한다.

다만, 이 법안은 피해 언론사에 손해배상을 직접 청구할 수 있는 권리를 별도로 부여하지는 않는다. 언론사는 익명 운영자의 신원을 확인하기 위한 소환장을 법원에 요청할 수 있지만, 법안에 따른 금지명령과 과징금 부과는 뉴욕주 검찰이 소송을 통해 청구한다. 또한, 과징금은 피해 언론사에 지급되는 손해배상금이 아니라 위법행위에 대해 주정부가 부과하는 제재금이다.

■ 스텔스 크롤링 금지 법안의 의의와 한계

이 법안의 가장 큰 의의는 뉴스 콘텐츠의 AI 이용을 둘러싼 문제를 저작권 침해 여부에만 의존하지 않고 접근의 투명성과 언론사 사이트의 안정적 운영을 중심으로 재구성했다는 데 있다. 생성형 AI 학습이 공정이용에 해당하는지에 대해서는 소송과 논쟁이 계속되고 있다. 반면 크롤러가 신원과 목적을 정확하게 밝혀야 한다는 원칙은 저작권 판단보다 명확하고 기술적으로도 구현 가능하다. 투명한 신원 표시는 언론사가 접근 허용, 차단, 라이선스 협상 또는 법적 대응 중 하나를 선택하기 위한 전제조건이 된다.

이 법안이 시행되면 AI 사업자와 언론사 간 정보 비대칭을 줄이는 효과도 기대할 수 있다. 현재 언론사는 서버 로그에서 대량 접근을 발견하더라도 실제 운영자와 이용 목적을 파악하기 어렵다. 법안은 크롤러 운영자가 스스로 신원과 모든 이용 목적을 공개하도록 함으로써 입증 부담의 일부를 사업자에게 이전한다.

그리고 이 법안은 뉴스 콘텐츠를 단순한 공개 웹 데이터가 아니라 지속적인 노동과 자본 투입을 통해 생산되는 경제적·공공적 자원으로 인정한다. 저널리즘의 보호는 언론사의 사적 이익에 그치지 않고, 신뢰할 수 있는 정보의 지속적 생산과 민주적 공론장 유지라는 공익에 연결된다는 것이 법안의 기본 취지다.

다만 몇 가지 한계도 있다. 우선 법안은 허락을 받지 않은 콘텐츠 이용 자체를 금지하지는 않는다. 크롤러가 신원과 목적을 공개한 상태에서 콘텐츠를 수집하는 경우에는 이 법안만으로 이를 막기 어렵고, 저작권법이나 계약법 등 다른 법리가 필요하다.

경제적 손해와 크롤링 사이의 인과관계를 어떻게 판단할 것인지도 불분명하다. 서버 부담은 로그 자료로 비교적 쉽게 확인할 수 있지만, 검색 유입 감소나 수익 감소가 어떤 크롤러 때문인지 특정하기 어려울 수 있다.

해외 사업자와 여러 서버와 IP 주소를 이용해 접속 요청을 분산하는 크롤러에 대한 법 집행 가능성도 문제다. 프록시 서버와 여러 데이터 중개업체를 거쳐 콘텐츠를 수집하면 실제 운영자를 특정하기 어렵다.

마지막으로, 지나치게 광범위하게 적용될 경우 검색엔진, 학술연구, 공익적 기록보존, 팩트체크 등 사회적으로 유익한 크롤링까지 위축시킬 가능성이 있다. 법안은 이용 목적을 공개하도록 하지만 공익적·비상업적 이용에 대한 예외는 두고 있지 않다.

■ 시사점

뉴욕주의 스텔스 크롤러 금지 법안은 AI 시대의 뉴스 콘텐츠 보호 정책이 AI 학습의 적법성을 둘러싼 저작권 논쟁에서 '누가, 어떤 목적으로, 어떤 방식으로 콘텐츠에 접근하는가'라는 접근 투명성의 문제로 확대되고 있음을 보여준다. 이 법안은 모든 무단 크롤링을 금지하지는 않지만, 신원과 목적을 숨긴 채 뉴스 사이트에 부담과 경제적 손해를 발생시키는 행위에 제재 근거를 마련했다는 점에서 큰 의미가 있다.

특히, AI 검색이 확대되고 있는 시점에서 뉴스 생태계의 핵심 쟁점은 AI의 뉴스 이용을 누가 통제하고, 가치를 어떻게 배분하며, 이용자가 원 출처와 연결될 수 있는가에 있다. 국내에서도 크롤러 신원·목적 공개, 용도별 접근 통제, 공동 탐지 인프라, 라이선스 표준 설계 등 종합적인 대응체계를 검토할 필요가 있다.

※ 이 글은 2026년도 한국언론진흥재단 미디어연구센터가 수행 중인 <AX 시대 저널리즘의 혁신 과제> 연구에 수록될 예정입니다.

2026년
7호

KPF 미디어브리프

K P F M e d i a B r i e f W e b z i n e

발행인 김효재

편집인 임석규

기 획 한국언론진흥재단 미디어연구센터

발행일 2026년 7월 23일

한국언론진흥재단 미디어연구센터

04520 서울특별시 중구 세종대로 124 프레스센터빌딩 13층

전화 (02) 2001-7743 팩스 (02) 2001-7740

www.kpf.or.kr

편집 (주)나눔커뮤니케이션

04034 서울특별시 마포구 잔다리로7길 16 교평빌딩 304호

전화 (02) 333-7136 팩스 (02) 333-7146

©한국언론진흥재단 미디어연구센터 2026

〈KPF미디어브리프〉는 한국언론진흥재단의 정부광고 수수료로 제작되었습니다.

자세한 내용은 QR코드를 통해 확인해 주세요.

