

WORLD MEDIA TRENDS

2025 해외 미디어 동향

2025 WORLD MEDIA TRENDS

한국언론진흥재단
Korea Press Foundation



해외 미디어 시장의
최신 트렌드 및 주요 이슈

2025 해외 미디어 동향

기획	한국언론진흥재단 산업분석팀
총괄	이상현(한국언론진흥재단 산업분석팀 팀장)
책임	남유원(한국언론진흥재단 산업분석팀 차장) 김수지(한국언론진흥재단 산업분석팀 과장)
집필·감수	김태균(연합뉴스 기자) 이성규(블루닷에이아이 대표) 신현규(글리터컴퍼니 대표) 이동욱(인프런 CTO) 윤창수(서울신문 기자) 김수지(한국언론진흥재단 과장)
내부자문	전창영(한국언론진흥재단 미디어연구센터 선임연구위원)
발행인	김효재
편집인	남정호
발행일	2025년 12월 31일
디자인	아르떼203 03998 서울 마포구 성미산로10길 12 1층 T. 02-323-4893~4 F. 070-4157-4893
인쇄	(사)한국장애인유권자연맹 인쇄사업부 04623 서울 중구 서애로5길 12-17 T. 02-325-1585 F. 02-325-1586

한국언론진흥재단

04520 서울특별시 중구 세종대로 124 프레스센터빌딩 13층
T. 02-2001-7752 F. 02-2001-7740
www.kpf.or.kr

* 이 책에 실린 내용은 본 재단의 공식 견해가 아닌 필자의 연구결과임을 밝힙니다.

© 한국언론진흥재단, 2025

ISSN 2799-8290 (Print)
ISSN 2799-8304 (Online)

'2025 해외 미디어 동향'은 한국언론진흥재단의 정부광고 수수료로 제작되었습니다.
자세한 내용은 QR코드를 통해 확인해 주세요.



2025 해외 미디어 동향

2025 WORLD MEDIA TRENDS

해외 미디어 시장의
최신 트렌드 및 주요 이슈

<해외 미디어 동향> 시리즈는

세계 미디어 시장의 최신 흐름과 주요 이슈를 국내에 신속히 전달하고,
언론사 경영 전략 수립에 참고할 수 있는 아이디어를 제공하기 위해
2008년부터 발간되고 있습니다.

보고서 전문은 한국언론진흥재단 홈페이지(www.kpf.or.kr)의

미디어정보-정기간행물 코너에서 무료로 열람할 수 있습니다.

1호

AI 플랫폼과 언론사의 제휴 동향

01 서론	007
02 인터넷에 공개된 뉴스는 공짜?	009
03 저작권의 맹점과 이중적 태도	012
04 AP통신 사례로 본 AI와 언론의 첫 공식 계약	015
05 악셀 슈프링어와 파이낸셜타임즈의 사례	017
06 AI와 언론사의 제휴 확대: 24년 5월의 사례	019
07 오픈AI 외 주요 플랫폼의 계약 동향	022
08 저작권을 둘러싼 소송 격화	024
09 AI 제휴의 편중 현상	027
10 국내 언론과 AI 플랫폼 간 협상 및 소송 현황	030
11 국내 언론의 과제는?	033
12 결론 및 제언	036

2호

뉴스 유료화의 새로운 시도와 퇴보

01 들어가며: 구독 피로 시대, 뉴스 유료화는 어디로	039
02 뉴스 유료화 회의론 속 언론의 세 가지 대응 모델	042
03 유료 구독 성장 전략의 전환: 번들링 모델의 부상	046
04 뉴스 유료화에서 발 뺀 언론사들	056
05 페이지월 외 대안의 수익모델 찾아나선 언론사	063
06 AI, 뉴스 유료화에 어떤 영향 미칠까	070
07 나가며: 뉴스 유료화 성장과 쇠퇴 사이, 그래도...	077

해외 미디어 시장의 최신 트렌드 및 주요 이슈

3호

‘AI 네이티브 언론’ 전환 로드맵 A to Z

01 들어가며: ‘AI 네이티브 기업’의 등장	081
02 AI 네이티브 언론 구축 가이드	086
03 언론사 조직 차원의 AI 기술 도입 전략: 기술적 선택지	105
04 나가며: 일단 해 봅시다	117

4호

AI 시대, 저널리즘의 새 판을 짜다: DataJConf 2025 탐방기

01 들어가며: DataJConf 2025를 가다	121
02 생성형 AI가 조직을 뒤흔들다: 뉴스룸 ‘재건축’	125
03 뉴스룸을 바꾸는 작은 실험들	132
04 뉴스룸의 AI 리터러시	142
05 국경·언어·조직을 넘어: 데이터 기반 협업 저널리즘	146
06 AI 시대, 독자와 언론의 새로운 관계를 설계하라	153
07 나가며: DataJConf 2025가 남긴 것	162

1호

AI 플랫폼과 언론사의 제휴 동향

01 서론	007
02 인터넷에 공개된 뉴스는 공짜?	009
03 저작권의 맹점과 이중적 태도	012
04 AP통신 사례로 본 AI와 언론의 첫 공식 계약	015
05 악셀 슈프링어와 파이낸셜타임즈의 사례	017
06 AI와 언론사의 제휴 확대: 24년 5월의 사례	019
07 오픈AI 외 주요 플랫폼의 계약 동향	022
08 저작권을 둘러싼 소송 격화	024
09 AI 제휴의 편중 현상	027
10 국내 언론과 AI 플랫폼 간 협상 및 소송 현황	030
11 국내 언론의 과제는?	033
12 결론 및 제언	036

김태균 연합뉴스 기자

연세대 정치외교학과를 졸업하고 한국과학기술원(KAIST) 경영대에서 정보경영 전공으로 석사학위(MS)를 받았다. 2005년 연합뉴스에 입사해 정보과학부(현 테크부), 사회부, 국제뉴스부 등에서 기사를 썼으며, 데이터·AI전략팀장을 역임했다. 저서로는 《AI 시대의 저널리즘》(공저, 2019), 《AI 뉴스 제작 방법》(공저, 2021), 《챗GPT와 생성 AI 전망》(공저, 2023), 《생성 AI와 데이터, 돈, 사람》(공저, 2024) 등이 있다. (이메일 tae@yna.co.kr)

01

서론

인공지능(AI)의 머신러닝에는 전산 자원과 데이터가 쓰인다. 이 두 요소는 필수적이다. 공부하는 학생으로 비유하자면 전산 자원은 뇌의 에너지, 데이터는 학습 교재에 해당한다. 아무리 학구열에 불타도 교재가 없으면 ‘진도’를 나갈 수 없다. 특히 챗GPT와 제미니(Gemini) 등 주목받는 거대언어모델(이하 LLM. 언어 중심의 생성 AI)은 데이터의 중요성이 더 크다. 고성능의 LLM을 만들려면 고품질의 언어 데이터가 핵심이다. LLM 업체에서는 논리적이고 사실 관계가 명확한 장문의 텍스트를 선호한다.¹⁾ 이 때문에 뉴스는 출간된 책과 함께 LLM에 가장 유용한 데이터로 평가된다. 그렇다면 뉴스업계는 2022년 챗GPT 혁신 이후 LLM 개발사로부터 수시로 ‘리브콜’을 받아왔을까? 현실은 그렇지 않다. 실제 주요 LLM 업체가 언론사와 계약해 정식으로 뉴스 데이터의 사용권을 얻은 경우는 AI 개발의 최전선인 서구에서도 손에 꼽을 정도로 적다. 오히려 법적 공방이 치열해지는 추세다. 언론사들이 뉴스를 도용당했다며 LLM 업체에 소송을 거는 경우가 잇따르는 것이다. LLM 업체는 뉴스를 저작물로 쓴 것이 아니라 단순 학습한 것에 불과해 저작권 침해가 아니라고 항변한다. 뉴스의 학습 데이터로서의 가치에는 이견이 없지만, 실제 현장에서는 정당한 보상을 받기 어려운 상황이 벌어지고 있다.

AI 기업과 국외 주요 언론사와의 협업 뒤편에는 이런 아수라장이 숨겨져 있다. 단순히 제휴 소식이 잇따른다고 ‘뉴스가 AI 시대의 데이터 상품이 될 수밖에 없겠구나’라고 안도할 수 없다. 국내 상황은 더 어렵다. AI 업체와 언론사의 대형 제휴 성과가 전무한 실

1) AI is setting off a great scramble for data, The Economist, 2023.08.13.
<https://www.economist.com/business/2023/08/13/ai-is-setting-off-a-great-scramble-for-data>

정이다. KBS 등 지상파 방송사는 올해 1월, 최대 AI 개발사인 네이버를 상대로 뉴스 저작권 문제와 관련된 손해배상 청구소송을 제기했다.²⁾ 서구보다 한국의 생성 AI 업계는 규모가 작기 때문에, 앞으로 뉴스 제휴 협상은 상당한 난항이 예상된다. 이번 기획은 이에 따라 먼저 왜 뉴스 데이터의 가치를 인정 받기 어려운지 원인을 분석하고, 이런 환경에서 어떻게 국외에서 제휴가 ‘용하게’ 성사될 수 있었는지를 짚는다. 이런 계약은 순전히 ‘선의’를 주고받은 결과가 아니다. 법적 대응과 로비 등 언론사 측의 강경 압박이 협상을 끌어내는 데 적지 않은 도움이 됐다. 이 때문에 이 글에서는 AI 업계에 대한 소송전 설명에 대해서도 적지 않은 비중을 할애할 계획이다. 이런 양면적 상황을 이해하는 것이 우리 언론의 나아갈 길을 찾는 데 더 도움이 될 것으로 본다.

지금의 상황은 2000년초 네이버 등 포털 태동기 때와 비교해서 훨씬 언론사에 불리하다. 초창기 포털은 트래픽을 유인할 볼거리가 턱없이 부족했고, 이 때문에 최소한 뉴스를 주력 콘텐츠 상품으로 먼저 인정했다. 전제로 등의 대가를 언론사에 적극적으로 제시했다. 반대로 생성 AI 업계는 뉴스 데이터를 상품으로 인정하지 않으려는 경향도 적지 않다. 애초에 대가를 지불할 대상이 아니라는 것이다. 앞서 지적한 것처럼 AI 개발에는 전산 자원과 데이터가 다 필요하다. 그런데 전산 자원을 마련하는 것 하나만도 엄청난 돈이 들어간다. 이 상황에서 데이터 부담까지 무거워질 것을 좋아할 기업은 없다. 뉴스 데이터의 가치 산정이 선례가 없고, 무리하게 과금하면 신기술 개발이 위축될 수 있다는 주장까지 나온다.³⁾ 뉴스를 AI 학습 데이터로 팔아본 경험이 없던 언론사들이 AI 업계의 이런 논리에 제대로 대처할 역량이 부족한 것은 큰 문제다. 이런 상황에서 우리가 봐야할 것은 해외 언론사가 AI 기업과의 제휴에서 얼마를 받았는지 등의 당장의 결과가 아니다. 어떤 고민과 맥락 속에서 이런 계약이 성사됐는지 그 배경을 살펴보는 게 더 중요하다. 흠친 교재로 하는 공부는 한계가 있다. AI 업계도 학습 데이터 생산자와의 공존이 꼭 해결해야 할 과제다. 좋은 교재가 더 유입이 안되면 AI의 진화에도 제동이 걸린다. 인간 데이터 생산자와 AI가 혜택을 주고 받으며 함께 성장하는 ‘선순환’ 상황이 오기를 기원한다. 그런 미래를 앞당기려면 우리 언론계의 지혜와 노력이 필요하다. 이 글이 이 여정에 작은 도움이라도 됐으면 한다.

2) “AI학습에 왜 우리 기사 쓰나” 방송사들 네이버에 소송, 홍윤지, 법률신문, 2025.01.18.
<https://www.lawtimes.co.kr/news/204853>

3) “생성 AI 뉴스 콘텐츠 보상, 데이터 중심으로 이뤄져야”(종합), 오규진·최평천, 연합뉴스, 2023.10.05.
<https://www.yna.co.kr/view/AKR20231005065951001>

인터넷에 공개된 뉴스는 공짜?

다수의 텍스트 뉴스는 인터넷에서 누구나 쉽게 볼 수 있다. 최근 수년간 뉴스 유료화에 대한 시도가 잇따르면서 여러 매체에 ‘페이월(paywall)’(유료 구독자 외의 열람을 막는 디지털 장벽) 빗장이 걸렸지만 ‘기사는 공기나 물처럼 흔하다’는 명제를 바꾸기에는 역부족이다. 이는 네이버, 구글, 페이스북 등 인터넷 플랫폼이 20년 넘게 만든 뉴스 생태계의 결과다. 많은 언론사는 포털에 기사를 노출하며 전재료를 받거나, 검색엔진이나 소셜미디어를 통해 유입되는 트래픽에 광고를 붙여 수익을 낸다. 이 때문에 이들의 기사는 누구나 클릭 1~2번이면 볼 수 있어야 한다. 이렇게 인터넷에 공개된 기사는 크롤링(또는 웹 스크래핑)이라는 과정을 통해 제삼자가 대량으로 수집할 수 있다. 저인망 그물로 물고기를 닦치는 대로 끌어가는 것과 비슷하다. 크롤링은 기술적 난도와 비용이 낮다. 사실상 누구나 할 수 있다. 반대로 기사 저작권자가 이를 기술적으로 차단하기는 쉽지 않다. 문을 함부로 걸어잠그면 ‘도용 혐의자’가 아닌 선의의 독자도 기사를 보는 데 불편을 겪을 수 있다. 적잖은 국내외 언론사는 기사 말미에 ‘해당 기사를 AI 훈련에 쓰는 것을 금지한다’는 문구를 붙인다. 이는 상징적 선언이다. 기술적으로 크롤링을 못하도록 금지했다는 뜻이 아니다. 사후 법적 문제를 물을 수도 있다고 경고만 하는 것이다.

[그림 1] 뉴스 데이터는 물처럼 공짜로 쓸 수 있을까?



출처: 챗GPT 제작

AI 개발자들은 큰 유혹을 느낄 수밖에 없다. 도덕적 해이를 유발하기 쉬운 구조다. 데이터를 정식 계약해 입수하는 것은 시간과 돈이 들어간다. 반면 양질의 텍스트 기사 데이터는 사실상 누구나 무한대로 가져갈 수 있게 인터넷에 열려 있다. 몰래 기사를 크롤링해 학습 데이터로 써도 적발이 어렵다. 무단 크롤링을 했다는 ‘디지털 자취’를 숨기고 지우는 기법이 많기 때문이다. IT 업계에서는 특정 업체의 내부 전산 기록을 압수 수색하거나 업무에 직접 관여한 개발자가 ‘양심 선언’을 하지 않는 이상, 무단 크롤링의 단서를 찾는 것조차도 쉽지 않을 것이라는 견해가 많다. 크롤링한 데이터를 학습했다는 사실도 입증이 매우 까다롭다. LLM은 학습 데이터를 자기 인공 신경망 내 가중치(AI 능력 발현에 필요한 복잡한 수치 설정)를 조정하는 데만 쓸 뿐 이를 따로 저장하지 않는다. 즉, 학습이 끝나면 ‘교재’는 녹아서 없어진다. 만약 LLM이 ‘가나다’라는 텍스트를 학습했다고 가정하자. 이 학습의 결과로 AI는 ‘카낫닷’ ‘캣나라’ ‘카나다카’ 등 원 학습 텍스트와 비슷하거나 약간씩 다른 문구를 만들어낼 수 있는 능력을 갖게 된다. 원리대로라면 학습한 자료를 그대로 표출하지 않는다(물론 여기에도 예외는 있다. 자세한 설명은 후반부 ‘저작권을 둘러싼 소송 격화’ 챕터 참조).

저작권자로선 답답한 노릇이다. 무단 학습의 의혹을 제기할 수는 있어도 ‘스모킹건’ 증거를 찾기가 정말 어렵다. “귀사 LLM이 우리 기사와 엇비슷한 텍스트를 생성한다. 데

이터를 무단으로 긁어간 것 아니냐”라고 항의해도 AI 업체 측이 “도용했다는 증거가 있느냐. 우연의 일치로 엇비슷한 텍스트가 생성됐을 수 있다”고 얼마든지 반박할 수 있다. AI 개발사는 크롤링 데이터의 사용과 관련해 운신의 폭도 넓다. 자신이 직접 ‘점점하게’ 무단 크롤링을 할 필요가 없다. 시중에 이미 개인과 비영리 기관이 장기간 전 세계 웹사이트를 긁어 만든 ‘개방 데이터’가 많기 때문이다. 오픈AI가 세계 최고 LLM인 챗GPT를 훈련하는데 활용한 무료 초거대 자료 ‘커먼크롤(common crawl)’이 대표적 예다. 이런 데이터를 쓰고 나중에 문제가 된다면 책임을 외부에 전가해도 된다. 미국 언론계 일각에서는 AI 업계의 무단 학습 행태를 ‘깨고 털기(smash and grab)’ 범죄에 빗대기도 한다. 이 범죄는 말 그대로 가게 유리창을 깨고 물건을 가져가는 행위다. AI 학습 때문에 개발사들이 웹 곳곳을 돌며 타인의 데이터를 절취하는 ‘무정부 상태’가 계속된다고 언론인들은 한탄한다. 오픈AI나 구글의 경우 자사 LLM의 개발을 위해 세계 모든 웹데이터를 몽땅 크롤링해 쓴 것 아니냐는 의심을 공공연히 받는다. 이런 거대 AI 개발사들이 가장 싫어하고 기파하는 질문 중 하나는 ‘구체적으로 어떤 데이터를 훈련에 썼느냐’다. 이들은 지금 최고의 혁신기관 찬사와 함께 우리 시대의 최대 ‘깨고 털기’범이 아니냐는 비판을 함께 받는다.

[그림 2] LLM 학습의 활성화로 AI 업계에서는 웹에 있는 타인의 데이터를 임수해 마음대로 쓰는 사례가 늘고 있다. 이를 사실상 거리외의 상점을 깨고 들어가 물건을 절취하는 행위와 다르지 않다고 보는 시각이 팽팽하다.



출처: 챗GPT 제작

저작권의 맹점과 이중적 태도

도덕적 질타를 하는 것과 법적 책임을 묻는 것은 전혀 다른 이야기다. 당장 우리는 AI 개발사에 기사 무단 도용에 대해 손해배상과 관련자 처벌을 요구할 수 있을까? 현행 저작권법에서는 쉽지 않다. 저작권법은 ‘저작물을 학습하고 활용하는 주체는 사람’이라는 명제 아래 만들어졌다. 저작물을 대량으로 학습해 사람과 비슷한 콘텐츠를 만들어내는 AI 시스템은 이 법이 만들어질 당시 상상 바깥의 존재였다. AI의 머신러닝에 저작물을 쓰는 행위에 관해서는 금지 규정이 아예 없다. 심지어 AI 업체 측에 유리하게 해석될 대목까지 있다. ‘공정이용(Fair Use)’이란 조항이다. 공정이용은 교육, 창작, 학술 등 특별한 상황에서는 저작물을 써도 저작권자의 허락을 얻거나 대가를 지불할 의무를 제한적으로 면해준다. 사람이 종전 저작물을 탐구하고 이를 토대로 재창작을 하면 문화가 융성해지고 사회 구성원 모두에게 ‘윈윈’이 되기 때문에 도입된 규정이다. 특히 공정이용에는 ‘변형적 이용(transformative use)’이라는 조건이 있다. 원 저작물을 복사 재현하는 것이 아니라 이에 힌트만 얻어 아예 새로운 것을 만들어낸다면 공정이용에 해당한다는 것이다. 앞서 설명한 것처럼 LLM은 학습 텍스트를 토대로 비슷하면서도 사실 다른 결과물을 만들어내는 것이 본질이다. 연합뉴스 김태균 기자의 기사를 학습해도 다음 기사를 쓸 때는 매번 문체와 내용이 슬쩍슬쩍 바뀌고, 기자 이름도 ‘김태균’ ‘김길동’ 식으로 계속 변형된다. AI의 출력이 변형적 이용이라고 주장할 여지가 충분하다. 이 때문에 AI 개발사는 처음부터 AI의 저작물 활용은 명백한 공정이용이며, 원 저작자에게 댓가를 줄 수 없다고 못박는 경우가 적잖다.

이런 주장은 과거 관련 판례도 있다. 구글은 2005년 작가조합(Authors Guild)에 저작권 소송을 당했다. 2002년 ‘구글 책 검색’이라는 서비스가 화근이었다. 종이 책은 애초 디지털

텔화가 되어 있지 않아 검색이 되지 않는다. 구글은 검색 가능한 지식과 정보의 범위를 대거 넓히겠다는 대의 아래 책의 원문을 대량으로 스캔하는 작업을 시작했다. 디지털화한 책의 본문은 자사의 데이터베이스에 넣어두고, 검색 결과값에는 책 본문의 극히 일부(스니펫)만 잘라 맛보기로 보여줬다. 저작권자들은 기겁했다. 권리 보유자의 허가나 대가 지불 없이 책을 무단 복제해 자사 서비스에 활용했으니 명백한 저작권법 위반이라는 성토가 쏟아졌다. 구글은 소송에서 이런 책의 활용이 공정이용에 해당한다고 일관되게 주장했다. 스캔한 책 원문 일부만 ‘변형적’으로 짧게 보여준 데다, 지금껏 검색되지 않았던 도서에 대한 관심을 불러일으켜 출판업계에도 이득이 됐다는 취지였다. 미국 법원은 1, 2심에서 구글의 손을 잇달아 들어줬고 2016년 대법원에서 같은 취지의 판결이 확정됐다.⁴⁾ 이 당시에는 AI가 전혀 이슈가 아니었다. 그렇지만 원문을 그대로 표출하지 않고 IT 서비스의 품질을 위해 저작물을 쓰는 것이 공정이용으로 인정됐다는 사실은 AI 업계에 큰 영향을 줬다. 검색을 위해 책을 스캔하는 것과 AI 훈련을 위해 인터넷의 기사를 긁어오는 것이 다를 것이 없다는 주장에 힘이 실리는 것이다.

AI 업계에는 저작권 문제와 관련해 두 가지 반응이 엇갈린다. ‘법적으로 우리는 압도적으로 유리하다’는 견해가 대세지만 ‘그렇지만 걱정된다’는 신중론도 팽팽하다. 이런 상반된 속내를 잘 보여주는 사례가 있다. 페이스북 운영사인 메타에서 AI 개발진이 2003년 당시 저작권 문제에 관해 주고 받은 대화록이 올해 초 법원 서류로 공개됐다.⁵⁾ 소송은 미국의 판타지 소설가 크리스토퍼 골든 등의 작가들이 메타를 상대로 2023년 7월에 제기한 것으로, 메타가 자사 LLM인 라마(Lama)를 학습시키면서 도서 데이터를 무단으로 썼는지를 다룬다. 특히 메타는 이 소송에서 러시아에서 개발된 불법 도서 공유 서비스인 ‘립젠’(Library Genesis, LibGen)에서 AI 훈련 데이터를 입수했다는 의혹을 받고 있다. 원고 측이 공개한 자료를 보면 메타의 AI 개발 실무진들은 “해적판 자료를 회사 업무 노트북으로 다루는 것이 영 마음이 내키지 않는다” “립젠 같은 불법 데이터를 쓴다는 사실이 언론에 알려지면 나중에 우리 회사가 저작권 문제에 관해 규제 당국과 협상할 때 불리해지는 것 아니냐” 등 우려를 쏟아낸 것으로 나타났다. 이들은 립젠 데이터를 써도 되는지 사주인 마크 저커버그 메타 최고경영자(CEO)에게까지 문의해 ‘OK’ 답변을 들었다고도 했다. 다

4) 구글 북스, 11년만에 ‘면죄부’ 받았다, ZD넷, 2016.04.19. <https://zdnet.co.kr/view/?no=20160419085755>

5) Mark Zuckerberg gave Meta's Llama team the OK to train on copyrighted works, filing claims, Wiggers. Tech Crunch, 2025.01.09. <https://techcrunch.com/2025/01/09/mark-zuckerberg-gave-metas-llama-team-the-ok-to-train-on-copyrighted-works-filing-claims/>

른 대화록에서 메타 개발자들은 정식 계약으로 도서 데이터를 입수하는 것이 시간이 너무 걸린다고 한탄하며, 무단 책 데이터가 법적 위험성이 있어도 이미 수많은 스타트업이 이런 자료를 쓰고 있을 것이라고 애써 안도하는 모습도 보였다.⁶⁾ 저작권 이슈에 관해 ‘허락보다는 (사후) 용서를 구하는 기조로 가야 하고, 책 확보 문제에 대해서도 수시로 경영진의 최종 판단을 구하며 법적 리스크와 관련해 더 대담해질 필요가 있다’는 발언도 나왔다. 결국 ‘공정이용’ 방어막을 과시해도 저작권 문제는 AI 개발사 내부에서도 내심 ‘뜨거운 감자’일 수밖에 없을 것으로 보인다.

[그림 3] 생성 AI는 종전의 저작권법을 혼란에 빠뜨리고 있다.



출처: 챗GPT 제작

6) Court filings show Meta staffers discussed using copyrighted content for AI training, Wiggers, Tech Crunch, 2025.02.21. <https://techcrunch.com/2025/02/21/court-filings-show-meta-staffers-discussed-using-copyrighted-content-for-ai-training/>

AP통신 사례로 본 AI와 언론의 첫 공식 계약

오픈AI와 AP통신의 데이터 사용권 계약은 2023년 7월13일 발표됐다. 이는 LLM 개발사가 세계 주요 언론사와 맺은 첫 계약이다. 당시는 LLM의 무단 데이터 활용에 관해 미국 언론사의 항의가 들끓던 때였다. 폭스뉴스 등이 속한 언론 그룹 ‘뉴스코프(Newscorp)’는 아예 주요 AI 업체에 대해 강경 소송을 경고했고 뉴욕타임스와 세계신문 협회 등의 성토가 거셌다. 이 과정에서 LLM 선두주자인 오픈AI가 어떤 회사와 첫 협약을 맺을지는 상징적 의미가 클 수밖에 없었다. AP통신은 당시 언론계의 예측을 벗어나는 첫 주자였다. 세계 최대 뉴스통신사로서 규모와 위상이 남달랐지만, AI 업계에 대한 저작권 공세 측면에서는 대외적으로 적극적인 모습을 보이지 않았기 때문이다. 계약은 양측이 맺은 상호비밀유지협약(NDA) 때문에 알려진 내용이 제한적이다. 오픈AI는 1985년도까지의 과거 기사 아카이브에 대한 사용권을 얻었다. AP통신은 이에 대해 1회성 금전 대가와 AI 기술 사용권을 받았다. 양사는 사용료의 구체적 액수와 관해서는 언급을 피했다. AI 기술 사용권은 AI 업체들이 많이 제안하는 혜택으로, 챗GPT API 등 자사의 유료 AI 서비스를 일정 기간 또는 용량만큼 언론사 웹사이트나 연구개발(R&D) 등에 쓸 수 있게 해주는 것이 골자다. AP통신의 AI 업무를 총괄하는 어니스트 쿡(Ernest Kung) AI 프로젝트 매니저는 2023년 11월 국내에서 열린 컨퍼런스에서 오픈AI의 기술 지원으로 장문의 기사를 짧게 요약하는 기술과 다국어 고객 채팅 서비스 등의 연구를 추진하고 있다고 밝힌 바 있다.

[그림 4] 오픈AI와의 라이선스 계약을 알리는 AP통신의 기사 화면



▶ 출처: 웹사이트 캡처

국내외 언론계와 AI 업계의 반응을 종합해보면 이 계약은 언론사와의 데이터 사용권 계약의 가능성을 타진하는 ‘파일럿’ 성격이 강해 보인다. 먼저 계약의 주 대상인 기사 아카이브는 인터넷에 게재된 바 없는 수십년 전 뉴스들로, 최신 기사를 계속 공급하는 피드(데이터 경로)는 계약 대상에서 아예 제외됐던 것으로 전해진다. 대가 지급 방식도 수백만 달러(수십억 원) 수준의 단발성 금액 지급에 그친 것으로 전해진다. 통상 언론사에 가장 유리한 계약 방식으로는 LLM의 매출 발생에 따라 지속적으로 데이터 대가를 받는 ‘로열티’ 방식이 꼽히는 만큼 AP 내부적으로 아쉬움이 남을 수 있는 대목이다. AP통신을 첫 파트너로 택한 배경도 관측이 분분하다. AP통신이 미국의 다른 대형 언론사와 비교해 AI 업계에 상대적으로 온건한 자세를 보인 데다, 수많은 지역 언론에 뉴스를 공급하는 비영리 협동조합이라는 AP의 특성상 파트너십의 공익적 이미지를 살리려는 의도도 있었을 거란 풀이가 나온다.

악셀 슈프링어와 파이낸셜타임즈의 사례

2023년 12월13일 유럽의 거대 언론 그룹 악셀 슈프링어는 오픈AI와 데이터 계약을 체결했다고 밝혔다.⁷⁾ 악셀 슈프링어는 독일의 유명 대중지 '빌트'와 일간지 '디벨트'를 보유하고 있으며 미국의 유력 정치 매체 '폴리티코'와 경제 뉴스 '비즈니스 인사이더'의 소유주로 인지도가 높다. 이 계약은 데이터 제공 대상이 세분화됐다는 점에서 주목을 받았다. 악셀 슈프링어가 보유한 매체 기사의 '요약본'을 챗GPT 답변에 노출시키고, 챗GPT가 원 기사의 웹사이트 링크도 인용할 수 있게 했다. 산하 매체의 기사를 오픈AI의 LLM 훈련용 데이터로도 제공한다. 이에 대한 대가로 악셀 슈프링어는 3년 동안 약 수천만 유로(수백억 원)의 금전 보상과 챗GPT 기술 사용권을 얻어낸 것으로 전해진다. 특히 흥미로운 대목은 챗GPT가 기사 원문을 그대로 노출시키는 것은 허용하지 않은 것이다. 즉 짧은 요약이나 링크만 보여주고, 원 기사는 뉴스 웹사이트에 들어가야 볼 수 있도록 했다. 이는 LLM 서비스가 뉴스 매체를 대체하는 위험성을 방지하려는 포석으로 풀이된다.

2024년 4월29일에는 오픈AI와 영국 파이낸셜타임스(FT)의 계약 소식이 발표됐다. FT는 고품격 경제 정보의 대명사라는 가치를 앞세워 온라인에 무료 기사를 거의 풀지 않고 엄격한 유료화 정책을 고수한 곳이다. 이렇게 보수적인 성향의 FT가 발빠르게 오픈AI와 거래를 성사시켰다는 소식은 국내외 언론계에 큰 놀라움을 안겨줬다. 계약의 골자는 악셀 슈프링어와 비슷하다. FT는 챗GPT가 자사 기사의 요약, 일부 인용문, 링크만 쓸 수

7) Axel Springer and OpenAI partner to deepen beneficial use of AI in journalism, 2023.12.13.
<https://www.axelspringer.com/en/ax-press-release/axel-springer-and-openai-partner-to-deepen-beneficial-use-of-ai-in-journalism>

있게 했다.⁸⁾ 훈련용 데이터 제공에 관한 언급은 없었다. 여기에 대한 대가는 공식적으로 발표된 것은 없지만 월스트리트저널(WSJ)은 협의에 참여한 소식통을 인용해 FT가 매년 500만에서 1천만 달러(약 73억~146억 원)를 받을 것으로 보도했다.⁹⁾ 언론계에서는 보상금 외에도 악셀 슈프링어 사례 때처럼 AI 기술 사용권이 '패키지'로 지급됐을 것이란 관측이 많다. 실제 FT는 계약을 발표하며 앞서 FT 구성원들이 모두 오픈AI의 기술을 쓸 수 있도록 챗GPT 기업용 상품에 가입했다고 발표했다. 해당 데이터 계약에 할인이나 기간 연장 등의 대가성 혜택을 얻어줬을 수 있다는 것이 업계의 예측이다.

[그림 5] 악셀 슈프링어와 오픈AI의 제휴 협약 발표문. 이 계약은 처음으로 LLM 업체에 제공하는 데이터가 세분화됐다는 점에서 주목을 받았다



출처: 웹사이트 캡처

8) Financial Times announces strategic partnership with OpenAI, 2024.04.29. https://aboutus.ft.com/press_release/openai

9) OpenAI, WSJ Owner News Corp Strike Content Deal Valued at Over \$250 Million, The Wall Street Journal 2024.05.22. <https://www.wsj.com/business/media/openai-news-corp-strike-deal-23f186ba>

AI와 언론사의 제휴 확대: 24년 5월의 사례

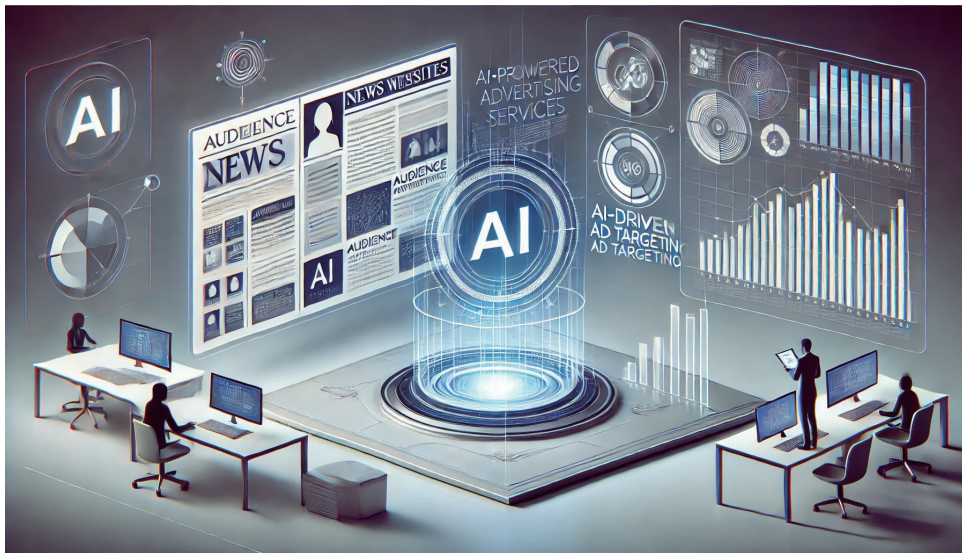
2024년 5월은 오픈AI와 언론사들의 계약 체결이 잇따랐던 달이다. 이 달 7일에는 미국의 언론사 '닷대시 메레디스'(Dotdash Meredith, 이하 DDM)와의 협약이 공개됐다.¹⁰⁾ DDM은 매거진 '피플'과 온라인 경제 매체 '인베스토피디아' 등 40여 개의 미디어 브랜드를 보유한 곳이다. 계약은 챗GPT에 DDM 측의 콘텐츠와 관련 링크를 노출시키고 LLM 훈련 데이터를 오픈AI에 제공하는 것이 골자다. DDM은 특히 계약의 대가로 자사의 차세대 맞춤형 광고 시스템 '디싸이퍼(D/Cipher)'에 오픈AI의 기술을 활용키로 했다. 디싸이퍼는 고성능 AI를 써서 당장의 콘텐츠 소비 성향만으로 독자의 성향을 정확히 예측해 최적의 광고를 골라준다. '쿠키(cookie)' 등 사전 개인 정보를 아예 쓰지 않기 때문에 계속 들끓는 사생활 침해 논란에서 자유로운 것이 핵심 장점이다. 언론사의 주요 수익원인 광고 시스템을 고도화하고자 오픈AI와의 계약을 활용한 것이라 흥미롭다.

서구 언론계의 대표적 '투견' 뉴스코프는 5월22일 오픈AI와의 계약 체결을 발표했다. 뉴스코프는 미국 월스트리트저널(WSJ), 영국 더타임스 및 더선 등 매체를 거느린 언론 재벌로 일찍부터 LLM의 데이터 도용에 대해 강력한 법적 대응을 예고한 바 있다. 뉴스코프는 AI 시대 이전부터 자사 지적재산권과 관련한 소송을 '본업' 못지 않게 열심히 해온 곳으로 유명하다. 22일 당일의 WSJ 보도에 따르면 뉴스코프와 오픈AI의 계약 기간은 5년이

10) Dotdash Meredith Announces Strategic Partnership with OpenAI, Bringing Iconic Brands and Trusted Content to ChatGPT 2024.05.07. PRNEWswire <https://www.prnewswire.com/news-releases/dotdash-meredith-announces-strategic-partnership-with-openai-bringing-iconic-brands-and-trusted-content-to-chatgpt-302138231.html>

고 계약 금액은 2억 5천만 달러(3천 643억 원)에 달한다.¹¹⁾ 이는 직접 보상금과 오픈AI 기술사용권을 다 합친 것으로, 지금까지 알려진 언론사와 생성 AI 업체간 계약금 중 최고액이다. 뉴스코프의 로버트 톰슨 최고경영책임자(CEO)는 이에 대해 “프리미엄 저널리즘에는 프리미엄(웃돈)이 붙는다는 것을 보여준 협약”이라고 자평했다. 콘텐츠 라이선스 구조는 FT나 악셀 슈프링어 등의 사례와 거의 같다. 즉 뉴스코프의 뉴스를 챗GPT 검색에 노출시킬 수 있게 하고, 기사를 훈련 데이터에 쓸 수 있는 권한을 제공하는 조건이다. 단, 뉴스코프는 최신 기사를 실시간으로 챗GPT가 보여주는 것은 허용하지 않고, 출고 뒤 일정 시간 뒤에야 노출이 가능할 수 있도록 한 것으로 전해진다. 자사 플랫폼을 대체해 챗GPT가 속보 등 실시간 뉴스를 공급하지 못하도록 못을 박은 것이다.

[그림 6] 데이터 제공의 대가로 LLM 업체가 언론사가 쓸 수 있는 AI 광고 시스템 등의 개발을 돕는 경우도 나오고 있다.



출처: 챗GPT 제작

오픈AI는 2024년 5월 29일 보кс 미디어(Vox Media), 디애틀란틱(The Atlantic)과 동시에 라이선스 협약을 맺었다. 보кс 미디어는 2000년대에 급부상한 인터넷 언론사로 테크놀로지 뉴스 웹사이트 ‘더버지(the verge)’와 종합 인터넷 매체인 ‘보кс(vox)’ 등을 갖고 있다. 디애틀란틱은 구독자가 100만 명이 넘는 미국의 오랜 잡지사로 장문의 탁월한 분석

11) 주 8과 같은 기사(OpenAI, WSJ Owner News…)

기사로 온라인 공간에서 인지도가 높다. 이 두 곳의 계약 내용은 엇비슷하다. 오픈AI에 기사를 제공하는 주요 대가로 AI 기반의 언론사 서비스를 개발하겠다는 것이다. 복스 미디어는 특히 자사의 선물 추천 웹사이트인 '더 스트래지스트 기프트 스카우트(The Strategist Gift Scout)'에 오픈AI 기술을 탑재해 개인화 정확도를 대거 높일 계획이라고 전했다.¹²⁾ 디애틀란틱도 최신 AI 기반의 뉴스 상품을 선보이는 '실험 공간' 역할을 할 '애틀란틱 랩'을 만든다고 밝혔다.

12) Vox Media and OpenAI Form Strategic Content and Product Partnership, Vox Media, 2024.05.29.
<https://www.voxmedia.com/2024/5/29/24166483/vox-media-openai-strategic-content-and-product-partnership>

오픈AI 외 주요 플랫폼의 계약 동향

오픈AI와 함께 LLM 개발의 선두주자인 구글은 언론사와의 뉴스 데이터 계약 사례가 알려진 것이 거의 없다. 유일하게 확인된 사례는 2025년 1월15일 발표한 AP통신과의 제휴다.¹³⁾ 이 계약에서 구글은 자사의 LLM 서비스인 '제미나이' 앱이 AP통신 최신 기사를 검색한 결과를 노출하는 권리를 얻었다. 이 대가로 AP통신이 어떤 보상을 받는지는 양측 모두가 함구해 드러난 바가 없다. 단, 계약 발표를 했던 AP통신 측 수장이 수석 부사장 겸 '최고 수익 책임자(chief revenue officer)'라는 사실을 볼 때 이번 거래가 무상 제공은 아닐 것으로 추정된다. 오픈AI와의 계약 때처럼 금전 보상과 기술 사용권을 다 챙겼을 공산이 크다. 앞서 구글은 2024년 4월 뉴스코프와 라이선싱 계약을 맺고 뉴스에 대한 대가로 연 500만~600만 달러(약 72억~86억 원)를 지급하기로 했다는 보도가 나온 바 있으나¹⁴⁾, 뉴스코프 측은 이후 이 기사의 내용을 부인했다.¹⁵⁾ 앤트로픽(Anthropic), 메타, xAI 등 그 외 주요 LLM 개발사는 상황이 더 박하다. 이들은 지금까지 언론사와의 제휴 계약을 맺었다는

13) Working with The Associated Press to provide fresh results for the Gemini app, Google Blog, 2025.01.15.
<https://blog.google/products/news/associated-press-gemini-app/>

14) Google to pay up to \$6 mln to News Corp for new AI content, The Information reports, Reuters, 2024.04.30.
<https://www.reuters.com/technology/google-pay-up-6-mln-news-corp-new-ai-content-information-reports-2024-04-30/>

15) News Corp Denies Reported AI Content Licensing Deal with Google, PYMNTS, 2024.04.30.
<https://www.pymnts.com/artificial-intelligence-2/2024/news-corp-denies-reported-ai-content-licensing-deal-with-google/>

실례는 물론이고 이를 추진하려 한다는 루머도 나온 것이 없다.¹⁶⁾

[표 1] LLM 개발사와 언론사와의 뉴스 데이터 제휴 현황

계약 발표일	LLM 개발사	언론사	특이점
2023.07.13	오픈AI	AP통신	LLM 업계의 첫 언론사 제휴
2023.12.13	오픈AI	악셀 슈프링어	데이터 제공 대상의 세분화
2024.04.29	오픈AI	파이낸셜타임스	연 500만~1천만 달러 지급
2024.05.07	오픈AI	닷대시 메레디스	언론사 서비스 개발에 대한 기술지원
2024.05.22	오픈AI	뉴스코프	역대 최대 금액의 보상
2024.05.29	오픈AI	복스 미디어	언론사 서비스 개발에 대한 지원
2024.05.29	오픈AI	디애틀랜틱	언론사 서비스 개발에 대한 지원
2025.01.15	구글	AP통신	AI가 실시간 뉴스 검색 결과를 노출하는 것을 허용

▶ 출처: 외신 및 필자 정리

16) 단, 생성형 AI 검색 업체인 '퍼플렉시티'는 언론사와의 제휴를 활발하게 한다. 미국 타임지와 영국의 인디펜던트지 등 매체 10여 곳과 계약을 맺은 것으로 알려졌다. 한국에서도 올해 3월 매일경제와 이데일리과 첫 파트너 협약을 체결했다. 그러나 퍼플렉시티는 자사 LLM 개발은 거의 하지 않고 주로 오픈AI 등 타사의 LLM 모델들을 가져와 검색 서비스에 활용하는 곳이다. 순수 LLM 개발사라고 보긴 어렵다. 퍼플렉시티가 계약하는 사용권은 기사를 AI 학습 데이터가 아니라 검색의 재료로 쓰겠다는 취지다. 업계에서는 이처럼 외부 LLM 모델들을 재'포장'해 사업을 하는 업체들을 '래퍼(wrapper)'라고 부른다. 이번 지면에서는 혼동을 줄이고자 래퍼들과의 제휴에 대해서는 따로 언급하지 않았다. LLM 시장이 초창기라 래퍼의 역할과 위상도 아직은 불분명하다. 향후 시장 성숙도에 따라 래퍼와 언론사의 계약 현황을 따로 정리할 기회가 있기를 기대한다.

저작권을 둘러싼 소송 격화

서구 언론계에서는 LLM 개발사와의 소송전이 한창이다. 한편에서는 계약서를 두고 악수하고 다른 곳에서는 칼을 갈며 소장을 쓴다. 대표적 사례로는 뉴욕타임스(NYT)가 꼽힌다. NYT는 2023년 12월 오픈AI와 마이크로소프트(MS)가 기사 저작권을 위반했다는 혐의로 소송을 제기했다. 이 업체들이 자사 기사를 무단으로 활용해 챗GPT와 MS 코파일럿 같은 AI 제품을 만들었다는 주장이다.¹⁷⁾ NYT는 이런 침해 행위에 대한 손해배상을 못하겠다면 해당 AI 제품들을 몽땅 폐기하라며 초강수를 뒀다. 소송에서 NYT가 부각한 문제는 '역류'(regurgitation) 현상이었다.¹⁸⁾ 앞서 설명한 것처럼 LLM은 원본을 학습해도 이론상 이를 똑같이 재현하지 않는다. 엇비슷한 '다른 것'을 창작해낸다. 그러나 NYT는 챗GPT 등이 끈질기게 질문을 물으면 뉴욕타임스의 기사 원문을 출력한다는 사실을 찾아냈다. '이런 내용의 기사를 써봐' '더 자세하게 이런 내용을 부각해봐' 식의 요청이 쌓이면 AI가 인터넷에서 많이 인용되던 장문의 유명 NYT 기사들을 글자 하나 안 틀리고 뱉어낸다. 영단어 regurgitation은 바다새나 펭귄이 새끼를 먹이려고 앞서 삼켜둔 생선을 고스란히 게워내는 모습을 묘사할 때 쓰는 말이다. 애초 소화를 시켜 전혀 딴 것을 만들어내야 하는데, 학습 자료를 원본 그대로 토해내면 '변형적 이용'이라는 주장을 펼 여지가 없어진다. 저작물을 무단으로 복사 및 재현하는 행위니 현행 저작권법에서 바로 문제를 삼을 수 있다. 오픈AI 등은 이를 매우 드물게 발생하는 '버그'이며 NYT가 악의적인 질문을 거듭

17) New York Times Sues Microsoft and OpenAI, Alleging Copyright Infringement, The Wall Street Journal, 2023.12.27. https://www.wsj.com/tech/ai/new-york-times-sues-microsoft-and-openai-alleging-copyright-infringement-fd85e1c4?mod=article_inline

18) NYT Complaint_Dec2023, 2023.12.27. https://nytc-o-assets.nytimes.com/2023/12/NYT_Complaint_Dec2023.pdf

해 자사 LLM을 고장내고 이를 소송 근거로 삼았다고 반발했다.

역류는 NYT 소송 만의 쟁점이 아니다. AI와 저작권 문제 전체를 좌지우지할 계기가 될 수 있다. 이 현상은 2023년 사진 판매 서비스 ‘게티이미지’가 이미지 생성 AI 업체 ‘스테빌리티 AI(Stability AI)’에 저작권 소송을 걸었을 때도 중요하게 언급됐다. 스테빌리티 AI가 게티이미지의 사진을 무단 학습했고, 이후 원본 대부분을 본뜬 합성 사진을 만든다는 얘기다. 게티이미지는 인터넷에 유통되는 사진에는 ‘워터마크’라는 반투명 로고를 넣어둔다. 스테빌리티 AI의 AI는 이 그림 속 워터마크도 다시 만들어내 ‘원본 역류’가 아니냐는 의혹을 샀다. AI 업계에서는 역류가 일시적 버그인 만큼, LLM 등 생성 AI의 근원적 문제가 아니라고 주장한다. 그러나 실제 이런 역류가 현 생성 AI의 불가피한 현상이라는 것이 밝혀진다면 상황은 180도 바뀐다. AI 업체들은 현행 저작권법 체계 아래에서도 훈련 데이터에 대한 저작권 보상을 해야 한다. 또 역류 현상을 알면서 기술 개발과 AI 훈련을 강행했다는 사실이 인정된다면 언론사 등에 대규모 손해배상을 해야 할 공산도 커질 전망이다. NYT 및 게티이미지 소송은 모두 2025년 현재도 결론이 나지 않았다. 역류 현상에 대한 판정은 이 두 판결에서 중요한 관전 포인트가 될 것이다.

[그림 7] ‘역류’의 예시. 챗GPT가 출력한 NYT의 ‘뉴욕시 택시 영업권 탐사보도’ 기사를 원본 기사와 대조한 결과값. 몇몇 문장과 단어를 제외한 대부분 대목이 동일하다.

Output from GPT-4: exempted it from regulations, subsidized its operations and promoted its practices, records and interviews showed. Their actions turned one of the best-known symbols of New York — its yellow cabs — into a financial trap for thousands of immigrant drivers. More than 950 have filed for bankruptcy, according to a Times analysis of court records, and many more struggle to stay afloat. “Nobody wanted to upset the industry,” said David Klahr, who from 2007 to 2016 held several management posts at the Taxi and Limousine Commission, the city agency that oversees medallions. “Nobody wanted to kill the golden goose.” New York City in particular failed the taxi industry, The Times found. Two former mayors, Rudolph W. Giuliani and Michael R. Bloomberg, placed political allies inside the Taxi and Limousine Commission and directed it to sell medallions to help them balance budgets and fund key initiatives. During that period, much like in the mortgage lending crisis, a group of industry leaders enriched themselves by artificially inflating medallion prices. They encouraged medallion buyers to borrow as much as possible and ensnared them in interest-only loans and other one-sided deals that often required borrowers to pay hefty fees, forfeit their legal rights and give up most of their monthly incomes. When the market collapsed, the government largely abandoned the drivers who bore the brunt of the crisis. Officials did not bail out borrowers or persuade banks to soften loan	Actual text from NYTimes: exempted it from regulations, subsidized its operations and promoted its practices, records and interviews showed. Their actions turned one of the best-known symbols of New York — its signature yellow cabs — into a financial trap for thousands of immigrant drivers. More than 950 have filed for bankruptcy, according to a Times analysis of court records, and many more struggle to stay afloat. “Nobody wanted to upset the industry,” said David Klahr, who from 2007 to 2016 held several management posts at the Taxi and Limousine Commission, the city agency that oversees cabs. “Nobody wanted to kill the golden goose.” New York City in particular failed the taxi industry, The Times found. Two former mayors, Rudolph W. Giuliani and Michael R. Bloomberg, placed political allies inside the Taxi and Limousine Commission and directed it to sell medallions to help them balance budgets and fund priorities. Mayor Bill de Blasio continued the policies. Under Mr. Bloomberg and Mr. de Blasio, the city made more than \$855 million by selling taxi medallions and collecting taxes on private sales, according to the city. But during that period, much like in the mortgage lending crisis, a group of industry leaders enriched themselves by artificially inflating medallion prices. They encouraged medallion buyers to borrow as much as possible and ensnared them in interest-only loans and other one-sided deals that often required them to pay hefty fees, forfeit their legal rights and give up most of their monthly incomes.
---	---

▶ 출처: 뉴욕타임즈 소장

뉴스코프의 핵심 매체 월스트리트저널(Wsj)과 뉴욕포스트는 2024년 10월 AI 검색 플랫폼(기반 서비스)인 ‘퍼플렉시티’에 저작권 소송을 냈다.¹⁹⁾ 퍼플렉시티는 오픈AI 연구원 등이 창업한 스타트업으로, 사용자가 질문을 입력하면 AI가 내용을 검색해 체계적인 ‘모범 답변’을 만들어주는 서비스를 한다. 퍼플렉시티는 ‘검색 공룡’ 구글의 위상을 위협할 유망주로 유명하다. Wsj와 뉴욕포스트는 퍼플렉시티의 AI가 답변을 만들며 자신들의 뉴스 콘텐츠를 무단으로 활용했다고 주장했다. 타사 기사를 몰래 가져와 자신의 AI 검색의 전반적 품질을 높인 만큼 대규모 ‘프리라이딩(무임승차)’에 해당하고, 퍼플렉시티의 차세대 검색이 뉴스를 요약해 언론 대체재 역할을 하기 때문에 ‘고객 강탈’을 저지른다는 주장이다. 복스 미디어, 디애틀란틱, 폴리τικο 등 매체들은 올해 2월 캐나다의 유명 LLM 개발사 ‘코히어(Cohere)’가 자사 기사를 AI 훈련 데이터로 도용했다며 소송전을 시작했다. 이 언론사들은 모두 앞서 설명한 것처럼 오픈AI와는 뉴스 사용권 계약을 맺은 바 있다. LLM 업체에 대해 제휴와 소송 전술을 모두 쓰는 셈이다. 코히어는 현대 LLM의 기초가 된 트랜스포머(Transformer) 기술을 개발한 연구자 등이 세운 스타트업으로, 세계적으로 개발 역량을 인정 받는 곳이다. 복스 미디어 등은 코히어의 LLM가 자사 기사를 고스란히 기억하고 있으며, 이를 토대로 원본과 동일한 기사나 이를 대체하는 대안 콘텐츠를 쏟아낸다고 주장했다. NYT의 ‘역류’ 주장과 비슷한 논리다. 이도 결국 저작물의 무단 복제 및 재현을 중대 위반 사항으로 보는 현행 저작권법에 호소하는 논리로 읽힌다.

19) Wall Street Journal, New York Post Sue AI Startup Perplexity, Alleging ‘Massive Freeriding’ Wall Street Journal, 2024.10.21. https://www.wsj.com/business/media/wall-street-journal-new-york-post-sue-perplexity-ai-c5d9554d?mod=article_inline

AI 제휴의 편중 현상

지금까지 언급한 국외 저작권 계약과 소송 사례를 살펴보면 한가지 흐름을 읽을 수 있다. LLM 업체들과 계약을 맺은 언론사들은 대다수가 대형 업체들이다. AP통신이나 FT 등 역사가 오래되고 보유 콘텐츠가 많은 곳이거나 뉴스코프와 악셀 슈프링어처럼 여러 매체를 거느린 미디어 그룹사가 계약 기회를 사실상 독식했다. 중소형 언론사는 디에틀란틱이 유일했고 지역 매체가 계약을 한 사례는 전무했다. 왜 이런 상황이 일어났을까? 이는 LLM 개발사들의 현실적 판단이 작용한 결과로 풀이된다. LLM 훈련에는 고품질 데이터가 ‘매우 많이’ 필요하다. 이 때의 ‘많이’란 사람이 일일이 직접 읽어보기 불가능한 정도의 분량을 뜻한다. AI의 학습은 인간의 학습과 스케일이 전혀 다르다. 개별 언론사와의 데이터 제휴는 보상 재원이 들어가는 데다 사전 논의, 본 협상, 계약서 작성, 세부 조항 협상 등 작업 때문에 거래 비용도 크다. 이 때문에 각 계약에 가장 이익을 최대화할 수 있는 곳은 이미 대량의 우량 데이터를 쌓아 놓은 거대 언론사가 될 수 밖에 없다. 디에틀란틱은 특이한 예외다. 디에틀란틱은 다른 매체가 많이 다루지 않는 특정 주제를 세밀히 파고드는 장문 분석 기사의 세계적 명가다. 각 주제를 ‘어떻게 이 정도로 꼼꼼히 길게 잘 쓸까’하는 찬탄이 나오게 다룬다. 예컨대 미국의 유력지 워싱턴 포스트의 경영 내용을 다룬 작년 6월자 기사²⁰⁾는 분량이 A4지 기준으로 20장에 육박한다. 이렇게 희귀 주제에 관한 장문의 심층 텍스트는 AI 훈련에 별도의 값어치가 있다. LLM이 해당 주제에 대해 ‘아무 말 대잔치’를 하지 못하도록 막고, 더 깊이 있게 사고할 수 있도록 해주는 길잡이 역할을 하기

20) The Real Story of the Crisis at The Washington Post, The Atlantic, 2024.06.27.

<https://www.theatlantic.com/ideas/archive/2024/06/washington-post-jeff-bezos-william-lewis/678799/>

때문이다. 디에틀란틱은 각 콘텐츠의 양(기사 길이)과 질이 전체 콘텐츠 수량의 열세를 극복한 특이 사례로 꼽힐 만하다. LLM 개발사의 눈 앞에 모든 언론사는 ‘평등하지’ 않다.

제휴와 소송의 상반된 상황에서 알 수 있듯 LLM 개발사들은 뉴스 데이터의 상품성에 대해 일관성 없는 태도를 보인다. 어떤 때는 돈을 주고 사용권을 얻어야 할 상품으로 대했다가 다른 때에는 인터넷에서 가져다 써도 저작권상 문제가 없는 ‘공정이용’ 대상물이라고 주장한다. 이런 ‘오락가락’ 행보는 저작권법 등 현행 법제에서 AI 훈련에 관한 규범이 없다는 현실을 볼 때 단순 기업의 잘못이라고만 볼 순 없다. 그러나 LLM 개발사들이 각 언론사를 매번 다른 원칙으로 대하며 ‘각개돌파’ 전술을 편다는 사실이 필자는 우려스럽다. 오픈AI와 계약을 체결한 언론사들은 인터넷에서페이월 등 방식으로 자사 기사를 가리고, 대외적으로 기사의 지적재산권을 강하게 주장했던 곳들이다. FT와 월스트리트저널은 구글링으로 찾을 수 있는 기사 대부분이 유료고객 페이월로 잠겨 있는 것으로 ‘악명’(?)이 높다. AP통신은 많은 기사를 인터넷에 공개하지만, 사업적으로는 기사 재산권 보호에 매우 열심이다. 특히 AP통신은 2021년 챗GPT 열풍 이전부터 AI 업계에 대해 ‘사용권 계약 필수’ 원칙을 물밑으로 알려온 것으로 알려져 있다. 앞서 AP통신은 2010년대 미국 헤지펀드에 AI 훈련 데이터로 기사를 팔았던 경험이 있었다. 이 때문에 경영진이 일찍부터 뉴스 데이터의 상품성을 잘 알고 있었다는 것이 미국 언론인들의 전언이다. LLM 개발사들은 인터넷으로 기사를 가져가기 힘들거나, 무단 사용을 문제삼을 공산이 큰 ‘강자’만을 대상으로 계약 제휴를 하고 있는 것으로 보인다. 나머지 상황에 대해서는 개별 소송에 대응만 하며 뉴스를 물처럼 공짜로 쓸 수 있는 현 상황을 관망한다. 이를 토대로 하면 다음과 같은 추측이 가능하다. LLM 훈련에 엄청난 양의 고품질 데이터가 필요한데 왜 뉴스 저작물 계약 사례는 이처럼 적을까? 애초 운동장은 LLM 개발사에 유리하게 기울어져 있다. 자금과 로비력이 막강한 AI 빅테크 회사로서는 시간도 이들의 편이다. 언론사와의 계약은 소수의 불가피한 상황에 선별적으로 하면 된다. 각 언론사의 제휴 소식은 언론계 전체로 봐선 딱히 ‘굿 뉴스’가 아니다.

국내 언론과 AI 플랫폼 간 협상 및 소송 현황

국내에서는 LLM 개발사가 언론사와 뉴스 사용권 계약을 정식 발표한 사례가 아직 없다. 미국 스탠퍼드대 기초모델연구센터(CRFM)가 집계한 세계 주요 LLM 현황을 보면, 한국에서는 네이버와 KT 등 모두 6개 업체가 10개의 LLM을 개발한 것으로 파악된다.²¹⁾ 이중 개발 규모가 가장 큰 네이버는 자사 모델인 ‘하이퍼클로바’와 ‘하이퍼클로바X’에 네이버 뉴스의 기사 데이터를 대거 쓴 것으로 파악돼 언론사들의 보상 요구가 잇따랐다.²²⁾ 네이버는 한국신문협회와 각 언론사 등과 협상을 벌였으나 데이터 가격 산정이 어렵고 국외 선례도 아직 적다는 등 이유로 네이버가 난색을 표하면서 실질적 성과가 나오지 못했다. 이 밖에도 다른 국내 LLM 개발사들이 비공개로 몇몇 언론사의 뉴스 데이터 라이선스를 산 것 같다는 추측이 나왔으나, 공식적으로 확인된 바는 없다. 신문협회는 결국 법제화 미비로 AI 업계에서 뉴스 저작권을 보장 받지 못하고 있다며 올해 3월 국회에 ‘저작권법’과 ‘AI 기본법’ 개정안 의견을 정부와 국회에 제출했다.²³⁾

21) 답시크 노리는 ‘K-AI 모델’은?...네이버·LG·트웹브랩스 등 꼽혀, SBS뉴스, 2025.02.06.
https://news.sbs.co.kr/news/endPage.do?news_id=N1007972668&plink=COPYPASTE&cooper=SBSNEWSEND

22) 네이버, AI 모델 학습에 사용된 뉴스 콘텐츠 보상 검토... “언론계와 협의할 것”, 조선비즈 2024.10.07.
<https://biz.chosun.com/it-science/ict/2024/10/07/HIYYF4BAABDNRI7XAT2VXYMCHE/>

23) 한국신문협회 “AI 시대 맞아 뉴스 저작권 보호해야” 한국경제 2025.03.04.
<https://www.hankyung.com/article/202503040954i>

[표 2] 국내에서 개발된 주요 LLM

회사명	LLM 명칭
코난테크놀로지	Konan LLM
KT	믿음(Midm)
LG AI 연구원	엑사원 2.0
모레	MoMo
네이버	하이퍼클로바X
네이버	하이퍼클로바
엔씨소프트	바스코-LLM
SK텔레콤	A.X
트웹브랩스	Marengo 2.6*
트웹브랩스	Pegasus-1*

▶ 출처: 미국 스탠퍼드대 기초모델연구센터(CRFM), 연합뉴스

* 트웹브랩스의 Marengo 2.6과 Pegasus-1 모델은 영상을 자연어로 처리하는 '멀티모달 기반의 언어모델'임.

소송전은 국내에서도 진행 중이다. 대표적인 사례는 KBS, MBC, SBS 등 지상파 3사다. 이 방송사들은 올해 1월 네이버가 AI 학습에 자사 기사를 무단으로 썼다며 저작권 침해 및 부정경쟁방지법 위반 혐의로 손해배상 청구소송을 냈다.²⁴⁾ 앞서 국내 방송사들의 연합체인 한국방송협회가 2023년 12월 네이버에 지상파의 뉴스, 영상, 오디오 콘텐츠를 AI 학습에 이용하는 것에 대한 보상을 요구했으나, 협의가 공전하자 지상파 3사가 먼저 실력 행사에 나선 것이다. 신문협회는 올해 2월 네이버를 공정거래위원회에 제소한다고 밝혔다.²⁵⁾ 입점 언론사의 동의를 받지 않고 아무런 대가 없이 네이버 뉴스의 데이터를 대거 AI 활용에 쓴 것은 거래상 우월 지위를 남용한 불공정 행위에 해당한다는 것이 제소의 취지다. 네이버 뉴스는 디지털 뉴스 생태계에 압도적 위상을 차지하는 만큼, 콘텐츠공급자(CP)인 언론사는 네이버와의 거래를 끊으면 인지도 추락과 매출 감소 등의 불이익을 겪을 수밖에 없다. 이런 유리한 위치에서 네이버가 불리한 약관을 강요하고 CP의 저작물을 연구 및 서비스 개선이라는 명분 아래 약탈했다는 것이 신문협회의 주장이다. 신문협회는 크롤링 등 기법으로 한국 뉴스를 무단 입수한 것으로 추정되는 오픈AI와 구글 등 국외

24) “AI 학습에 기사 무단 사용”...지상파 3사, 네이버에 저작권 소송, 연합뉴스, 2025.01.14.
<https://www.yna.co.kr/view/AKR20250114042500005>

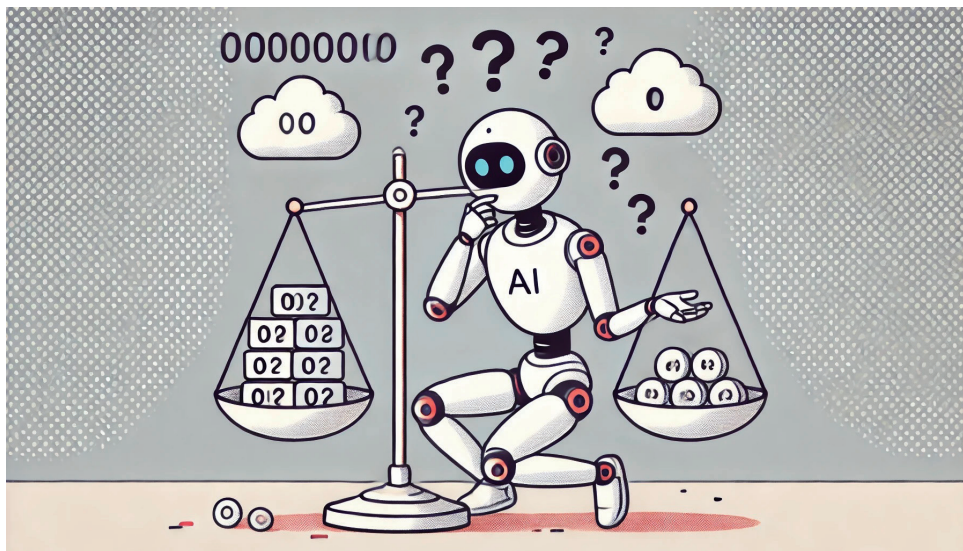
25) 신문협 “네이버, 뉴스 무단이용 위법” 공정위 제소. 기자협회보. 2025.02.17.
<https://www.journalist.or.kr/news/article.html?no=57887>

LLM 개발사에 대해서도 제소를 하겠다는 입장이라, 이 법정 공방의 범위는 더 넓어질 것으로 보인다.

국내 언론의 과제는?

국내 언론계 일각에서는 물밑 협상과 소송이 계속되며 자연스럽게 외국처럼 LLM 업계와 언론사 간의 제휴 발표가 잇따를 것이라는 기대감이 크다. 그러나 이런 변화를 위해서는 언론사들이 먼저 풀어야 할 난제가 많다. 앞서 지적한 것처럼 LLM 개발사들의 속내는 복잡다단하다. 이들의 선의에만 의존할 수 없다. 가장 우선적으로 언론사들이 해결해야 할 문제는 데이터의 적정 가치 산정이다. LLM 학습에 양질의 데이터가 대량으로 필요한 것은 엄연한 사실이지만, 구체적으로 개별 데이터가 LLM의 품질 향상에 얼마나 기여를 하는지에 관해서는 학계와 업계에서 정립된 산출법이 없다. 2023년 말 NYT 소송이 시작될 당시 오픈AI가 한 항변은 이 현실을 잘 보여주는 실례다. 이들은 “NYT 기사처럼 콘텐츠 출처(source) 별로 쪼개 보면 각 출처의 콘텐츠가 우리 LLM의 학습에 의미있는 기여를 했다고 말하기 어렵다. NYT의 콘텐츠가 향후 학습에 충분히 영향력을 가질 것으로 볼 수 없다”고 주장해 NYT를 격분케 만들었다. 각 데이터의 기여도와 가치를 측정하기 어렵다는 사실을 고지식(?)하게 실토해 저작권자와의 관계를 망가뜨린 것이다. 이런 측정법의 개발을 LLM 업계가 먼저 알아서 해주길 바라는 건 무리다. 다들 LLM 성능 경쟁과 개발 효율 개선에 몰두하는 ‘전쟁터’에서 데이터 가격 산정을 고민하는 것은 사치일 수 있기 때문이다. 이런 연구는 오히려 언론계가 적극적으로 지원하고 주도해야 한다. 예컨대 공공연구기관이나 데이터 업체와의 공동 연구를 진행할 수 있다.

[그림 9] LLM 학습에 쓰이는 개별 데이터의 기여도와 값어치를 객관적으로 측정하지 못한다는 현실은 데이터 제후에 큰 걸림돌이 되고 있다.



출처: 챗GPT 제작

개별 언론사가 보유한 기사 데이터가 LLM 개발사가 원하는 규모에 못 미치는 상황도 대책이 시급하다. 국내 AI 개발자들 사이에선 ‘한 매체가 제시한 뉴스 데이터가 LLM 학습에 필요한 규모의 100분의 1이나 1천분의 1에도 못 미쳐 난감했다’는 전언이 간혹 들린다. LLM의 학습은 인간의 가늠을 뛰어넘는 분량의 자료가 필요하다. LLM 개발사는 이 때문에 통상 여러 언론사의 기사를 묶은 ‘번들’ 구매를 희망한다. 그러나 공동 판매는 성사가 쉽지 않다. 각 사의 이해관계가 다르기 때문이다. 이런 약점을 노려 중소형 언론사의 기사를 헐값에 매집해 LLM 업계에 팔고 수익을 챙기는 ‘악탈적’ 중간업자가 득세할 수 있다는 우려도 나온다. 이 때문에 LLM에 최적화한 공동 뉴스 ‘아카이브’를 설립하는 논의가 필요해 보인다. 언론사의 데이터 판매 역량이 부족한 것은 또 다른 난관이다. AI 훈련용 데이터도 상품이다. 상품은 결국 어떻게 제대로 개발하느냐에 따라 가치가 천차만별이다. 예컨대 LLM 학습에 쉽게 쓸 수 있도록 정제가 되어 있고, 특정 이슈에 따른 학습이 가능하도록 내용 분류가 이뤄져야 한다. 데이터가 유용한 주석(annotation)을 잘 갖추고 있어야 하고, 여러 고객에게 팔 수 있도록 다층적이고 유연한 가격 체계를 제시할 수 있어야 한다. 이런 데이터 상품의 개발 실력은 국내 언론사가 열세일 수밖에 없다. 데이터 판매업을 한 경험이 있는 곳이 드물다. 이런 역량을 외부에서 ‘수혈’하는 방안을 고민해야 한다.

진짜 어려운 질문은 따로 있다. ‘일단 LLM 데이터의 판매를 성사시키고 난 뒤에는 어

떻게 해야 하느냐?’는 것이다. 계약을 맺었다고 샴페인을 터뜨릴 처지가 아니다. 1회성 판매로 매출이 발생한 뒤 LLM 업체들이 ‘당신네 자료는 필요없다’며 문전 박대를 할 수 있다. LLM 산업의 성장에 따라 계속 뉴스 데이터의 대가를 받을 조치가 필요하다. LLM 서비스의 매출 일부를 배분 받는 방식이나 수년 단위의 계약 등을 추진해야 한다. LLM 업체에 학습 자료를 제공하는 데이터 공급자(DP·Data Provider) 역할 만이 옳은 길인지도 사실 의문이다. 2000년대 언론사들이 네이버와 다음 등 포털에 콘텐츠공급자(CP·Content Provider)로 대거 참여했을 때의 경험이 ‘반면교사’가 될 수 있다. 인지도가 오르고 추가 매출이 들어와 포털 입점은 명백한 호재였지만, 시간이 지나자 주요 언론사들은 ‘포털 종속’의 흑독함을 실감했다. DP 사업도 당장 돈을 벌 수 있겠지만 중국엔 뉴스 주도권과 독자와의 접점을 잃는 ‘LLM 종속’의 관문이 될 수 있다. 서구 언론계에선 이 때문에 LLM 데이터 공급을 ‘악마와의 계약’이라고 비판하는 이들도 많다. AP통신은 올해 1월 구글과의 데이터 협약을 발표하는 기사의 말미에 다음과 같이 한 연구자의 발언을 덧붙였다. “데이터 공급은 언론사가 AI 기업에 콘텐츠 소비와 수익창출에 관한 통제권을 넘겨주는 계기가 될 수 있다. 독자들과의 직접적 관계를 강화하는 것 대신 저널리즘을 값싼 소비재로 전락시키는 플랫폼에 원재료를 공급하는 상황이 우려스럽다”²⁶⁾

[그림 10] 언론계 일각에서는 LLM 서비스가 언론을 대체할 수 있는 만큼, LLM용 데이터를 제공하는 행위 자체가 전략적으로 옳지 않다는 지적도 나온다.



출처: 챗GPT 제작

²⁶⁾ Google signs deal with AP to deliver up-to-date news through its Gemini AI chatbot, Associated Press, 2025.01.16.
<https://apnews.com/article/google-gemini-ai-associated-press-ap-0b57bcf8c80dd406daa9ba916adacfaf>

결론 및 제언

AI 업계는 항상 급변한다. 올해 1월 중국에서 고효율 LLM ‘딥시크(DeepSeek)’가 공개되며 세계 AI 산업이 크게 흔들렸다. 딥시크는 오픈AI 등 서구 주요 업체의 개발비와 비교해 10분의 1도 되지 않는 예산으로 챗GPT 등 주요 모델과 유사한 성능을 구현했다. ‘MoE’(Mixture of Experts) 등 AI 업계에서 연구됐던 여러 효율화 기법을 기발하게 집대성한 성과였다. 딥시크의 등장으로 AI 업계에서는 물량공세를 펴는 기존의 전략이 ‘과잉 투자’가 아니냐는 우려가 들불처럼 번졌다. 딥시크 발표 이후 AI 칩 대표주인 미국 엔비디아는 올해 1월 27일 하루에만 주가가 17% 폭락하며 시가총액 846조 원이 사라졌다. 딥시크가 LLM 개발에서 ‘대마불사’ 신화를 깨고 고효율 소형 모델의 장점을 부각하면서 AI 투자 규모가 크게 위축할 것이라는 관측은 대세가 됐다. 그러나 현실은 반대로 흘러갔다. 저렴한 가격으로 딥시크 같은 LLM을 여럿 만들 수 있다는 자신감에 힘이 실리면서 세계 곳곳에서 AI 개발이 더 활발해졌고, 중국과의 기술 패권 경쟁이 격화한 여파로 중주국 미국에선 AI 투자가 더 늘었다. 메타, 아마존, 알파벳, 마이크로소프트 등 4개 AI 선도 업체가 올해 2월 공개한 한 해 자본 지출 규모는 3천 200억 달러(466조 원)으로 작년 자본 지출액인 2천 300억 원보다 40%가 증가했다.²⁷⁾ 서론에서 언급한 것처럼 AI 개발에는 전산자원과 학습 데이터 두 요소가 항상 함께 쓰인다. AI 투자가 늘어나면 고품질 데이터에 대한 수요는 따라 오르고, 뉴스를 찾는 LLM 업계의 발걸음은 여전히 바빠질 것으로 현직 개발자들은 설명한다. 딥시크 같은 고효율 모델도 무(無)에서 생긴 것이 아니다. 모델을

27) 딥시크 충격에도... 4개 빅테크 올해 AI 투자 40% 확대, 연합뉴스, 2025.02.09.
<https://www.yna.co.kr/view/AKR20250209001800091>

만들려면 학습을 해야 하고 교재가 필요하다.

그러나 해외 사례를 토대로 관측하자면, 우리 언론사가 앞으로 가야 할 길은 마냥 순탄하진 않을 것이다. LLM 업체는 개별 제휴와 소송 대응을 모두를 염두에 두고 복잡미묘한 고심을 반복할 전망이다. AI 기업들이 먼저 뉴스 데이터의 상품성을 인정하고 안정적인 가격을 제시할 가능성은 필자가 단언컨대 0에 수렴한다. 이들에게 기사는 사용권을 사야 할 수 있고, 동시에 공정이용 규정에 따라 공짜로 퍼 쓸 수도 있는 ‘변화무쌍’한 존재다. 뉴스의 값어치 인정을 위해 뛰어야 할 곳은 결국 언론사들이다. 물도 수질을 따져 잘 포장하면 생수로 팔린다. AI 시대에 기사를 가격표가 제대로 붙은 ‘재화’로 탈바꿈시키기 위해 여러 노력이 필요하다. LLM 업체와의 계약이 한국에서 본격화해도 안심은 금물이다. 뉴스 데이터의 희소성이 점점 떨어지면 단순 공급자의 위상은 추락할 위험성이 있다. 앞서 제휴 사례에서 언급한 미국 정치 매체 ‘폴리티코’는 자사 유료 서비스 ‘폴리티코 프로’²⁸⁾에서 올해 3월 자체 개발한 LLM 도구를 출시했다. 폴리티코 프로의 주요 고객은 워싱턴의 정관계 종사자들인데, 이들을 위해 AI가 손쉽게 정책 보고서 등을 만들어준다. 이 LLM 도구는 폴리티코의 정책 기사를 학습해 글의 질과 정확도를 높였다. 이처럼 LLM 업체와의 파트너십을 맺은 해외 언론사들은 자사 기사 데이터의 가치를 보호하는 동시에 AI 기업으로 진화하고자 안간힘을 쓰고 있다. AI를 자사의 매출 및 가치 창출에 적극적으로 활용하는 길을 찾는 것이다. 이는 해외 제휴 사례에서 우리가 가장 눈여겨 봐야 할 대목이다.

²⁸⁾ POLITICO Unveils Policy Intelligence Assistant: An AI Powered Suite for Policy Professionals, POLITICO, 2025.03.06. <https://www.politico.com/blogs/politico-press/2025/03/06/politico-unveils-policy-intelligence-assistant-an-ai-powered-suite-for-policy-professionals-00215482>

2호

뉴스 유료화의 새로운 시도와 퇴보

01	들어가며: 구독 피로 시대, 뉴스 유료화는 어디로	039
02	뉴스 유료화 회의론 속 언론의 세 가지 대응 모델	042
03	유료 구독 성장 전략의 전환: 번들링 모델의 부상	046
04	뉴스 유료화에서 발 뺀 언론사들	056
05	페이월 외 대안의 수익모델 찾아나선 언론사	063
06	AI, 뉴스 유료화에 어떤 영향 미칠까	070
07	나가며: 뉴스 유료화 성장과 쇠퇴 사이, 그래도...	077

이성규 블루닷에이아이 대표

이성규 대표는 미디어 기술, 비즈니스, 저널리즘에 오랜 시간 천착하며 '더코어(전 미디어고토사)'라는 미디어 분석 전문매체를 20여 년간 운영하고 있다. 언론사와 포털 플랫폼에서 기자 및 프로덕트 설계자로 활동하며 국내외 미디어 흐름을 다양하게 경험하고 관찰해 왔다. 미디어티에서 미디어 스타트업 엑셀러레이터로 활약하며 뉴닉, 어피티, 디에디트, 킵블 등을 발굴했다. 저서로는 《트위터, 140자의 매직》(2009), 《혁신 저널리즘》(2015), 《사라진 독자를 찾아서》(2018), 《AI, 빅테크, 저널리즘》(2024) 등이 있으며, 주간경향에 칼럼도 매월 기고하고 있다. 신뢰 있는 미디어 생태계를 복원하고자 블루닷에이아이를 창업했고, AI 검색 최적화 분석 플랫폼 '블루닷 인텔리전스' 개발에 매진하고 있다. (이메일 sungku@mediasphere.kr)

들어가며: 구독 피로 시대, 뉴스 유료화는 어디로

구독 피로(Subscription Fatigue)¹⁾라는 단어가 등장한 지 약 6년이 지났다. 반복되는 결제 대상물의 과잉으로 소비자들이 심리적으로 부담감을 느끼는 현상을 의미하는 이 용어는 구독 경제의 지속성에 대한 회의론을 상징한다(Nguyen, 2025). 이 용어는 넷플릭스와 같은 비디오 스트리밍 서비스의 고속 성장에 따른 후발 효과로 자주 언급됐다. 구독해야 할 옵션(넷플릭스, 아마존 프라임, 디즈니 등)은 빠르게 늘어나고, 경기 침체로 구독자들의 예산은 빠듯해지고, 인지 부담이 커지는 상황이 초래한 결과로 나타난 현상이었다. 하지만 최근 들어 이러한 구독 피로감이 '구독 비즈니스'로 급속히 전환하고 있는 뉴스 산업으로 전이되면서 유사한 회의론을 불러일으키고 있다.

뉴스 산업에 한정하면, 구독 회의론에 사실상 첫번째 못을 박은 언론사는 타임(Time)이다. 전 세계 대부분의 언론사들이 디지털 페이월(Digital Paywall)을 올리기 위해 혈안이 돼 있을 시점에, 타임은 돌연 페이월을 제거하는 정반대의 흐름에 올라탔다. 물론 1년 전인 2022년 쿼츠(Quartz)가 페이월을 없애며²⁾ 신호탄을 쏘아올리긴 했지만, 타임과는 규모와 파급 효과 측면에서 비교가 되지 않았다. 2023년 6월 1일, 타임은 웹사이트를 가로막았던 모든 유료 장벽을 완전히 걷어냈다. 하드 페이월에서 계량형 페이월로 돌아간 지 약 2년 만이었다. 2023년 타임 창간 100주년 기념 갈라에서, 제시카 시블리(Jessica Sibley)

1) <https://www.wired.com/story/netflix-subscription-fatigue-uk/>

2) <https://www.journalism.co.uk/news/quartz-pivots-from-paywall-to-registration-wall-making-the-bulk-of-online-content-free-to-read/s2/a919903/>

CEO는 다음과 같이 그 취지를 설명했다.³⁾

“타임의 사명은 세상을 움직이는 사람과 아이디어에 대한 신뢰할 수 있는 양질의 스토리텔링을 제공하고 지리적 위치나 사회경제적 지위에 관계없이 가능한 한 많은 사람이 정보에 접근할 수 있도록 하는 겁니다. 신뢰할 수 있는 정보에 대한 접근은 전 세계적인 필수 과제이며, 모든 인류가 그 정보를 이용할 수 있어야 한다고 믿습니다.”

당시 타임의 디지털 유료 구독자 수는 25만 명이었다.⁴⁾ 인쇄물 구독자가 130만 명인 점, 2011년 처음 디지털 유료화를 시도했다는 사실 등을 감안할 때 기대 이상의 디지털 유료 구독 성과를 달성했다고 평가하긴 어렵다. 그럼에도 25만 명의 디지털 구독 수익을 한번에 포기하기로 한 결정은 결코 쉽지 않았을 것이다. 사라진 유료 구독 수익만큼을 단 기간에 다른 수익원으로 채워넣지 않으면 재무적 상황은 악화 일로를 걸을 수밖에 없어서다. 이러한 위험을 감수하면서까지 그들은 유료 구독 철회로 방향을 선회한 것이다.

퀴즈에 이어 타임까지 유료 구독 대열에서 이탈하자 ‘유료화 포기’를 선언하는 언론사들이 잇달았다. 대표적으로 테크 전문 디지털 미디어 ‘테크크런치’(Techcrunch)를 들 수 있다. 테크크런치는 그동안 유료로 운영하던 테크크런치+를 2024년 2월 29일자로 폐쇄하고 모든 페이지를 제거하겠다고 발표했다.⁵⁾ 그리고 테크크런치+ 담당자 대부분을 해고했다. 서비스 폐쇄에 따른 비용 절감 조치의 일환이었다. 그리고 그동안 작성했던 모든 테크크런치+ 콘텐츠는 테크크런치 사이트에 통합해서 제공한다고 발표했다. 사실상 완전한 유료화 포기 선언인 셈이다.

2022년부터 확산하기 시작한 뉴스 유료화 회의론이 담론 수준을 넘어 현실화하면서 디지털 페이지월이라는 보편적인 뉴스 유료 구독모델은 이전과 다른 평가를 받기에 이르렀다. 더이상 생존을 위해 당연히 걸어가야 할 길로 인식되는 경향은 확연히 줄어들었다. 디지털 페이지월 이외에 다른 생존 경로가 있을 수 있다거나, 유료 구독은 상위 몇몇 언론사만이 선택할 수 있는 ‘그들만의 전유물’이라는 인식이 강해졌다. 그것의 개발, 운영, 성장 마케팅에 이르는 비용이 워낙 막대하기에 중소 규모 언론사들은 이러한 시스템을 지탱할 수가 없는 경우가 허다해졌다. 오히려 이전처럼 광고 수익과 오프라인 이벤트에 집중하면서 비용을 적절히 통제하는 모델에 대한 선호가 높아진 것이 사실이다.

3) <https://time.com/6274933/time-to-make-its-trusted-journalism-free-and-accessible-worldwide-on-time-com/>

4) <https://www.axios.com/2023/04/27/time-removes-digital-paywall>

5) <https://techcrunch.com/subscribe/>

그렇다고 디지털 뉴스 유료화를 시도하려는 움직임이 줄어들거나 멈춘 것은 아니다. 지금도 전 세계 대형 언론사를 비롯해 중소형 미디어들은 디지털 유료 장벽을 쌓으려는 시도를 멈추지 않고 있다. 무엇보다 챗GPT와 구글발 AI 검색의 보편화로 제로 클릭(Ze-ro-Click)⁶⁾ 현상이 두드러지면서 광고 수익이 하락할 것을 예상하고 페이지의 중요성에 공감하는 분위기도 역력하다. 자칫 광고 수익에만 과도하게 의존할 경우 조직의 미래를 장담할 수 없을 것이라는 불안감마저 팽배해진 상황이다. AI의 산업 내 침투로 뉴스 산업의 불확실성이 어느 때보다 증폭되는 시점에, 뉴스의 유료 구독을 둘러싼 상반된 흐름을 살펴보면서 각각의 조건에 최적화한 수익 모델을 찾아가는 건 불가피한 과정이 됐다.

6) <https://www.bain.com/insights/goodbye-clicks-hello-ai-zero-click-search-redefines-marketing/>

뉴스 유료화 회의론 속 언론의 세 가지 대응 모델

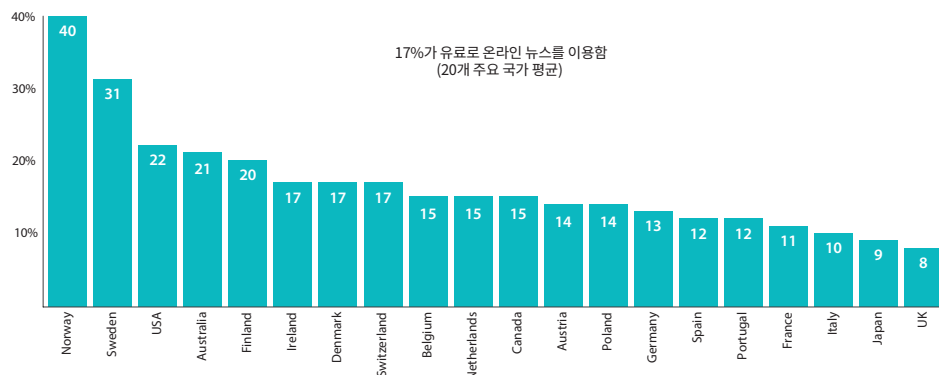
1) 뉴스 유료화의 확산 정체기와 승자독식

뉴스 유료화 회의론은 몇 가지 상징적인 통계를 통해 근거를 찾아볼 수 있다. 그 하나가 ‘유료 구독자 천장론’이다. 널 사람은 이제 다 돈을 내고 뉴스를 구독하고 있다는 주장이다. 뉴스에 대한 지불의사가 이제 성숙기에 접어들어 더이상 성장하지 않는다는 이 주장은 데이터에 의해 뒷받침된다. 대표적으로 로이터저널리즘연구소가 매년 발표하는 디지털 뉴스 리포트에서 이러한 흐름이 감지된다. 디지털 뉴스 리포트는 ‘지난해 온라인 뉴스에 비용을 지불한 비율’을 매년 측정한다. 2022~2025년까지 4년치를 비교해 보면, 전 세계 평균치는 17~18%로 정체돼 있다. 개별 국가마다 증감폭이 다르긴 하지만 평균 비율 자체는 4년 간 변화가 없었다. 영국과 같은 일부 국가에서는 비율이 2023년 9%에서 이듬해 8%로 줄어들기까지 했다. 어떤 지역은 늘어나고 어떤 지역은 줄어드는 현상이 동일한 평균치 하에서 반복되고 있다.

문제는 뉴스 유료 구독자 비율은 정체된 상태에서 일부 언론사만이 그 과실을 전유한다는 사실이다. 유료 구독자의 상위 언론사 쏠림 현상이 뚜렷하게 나타나는 것이다. 이른바 승자독식(Winner Takes All) 현상이 후발 사업자들의 의욕을 꺾는 징후가 포착됐다고 말할 수 있다.

“디지털 뉴스 보고서의 몇 년간 조사에서, 우리는 20개 시장 일부 국가의 설문 응답자들에게 어떤 온라인 뉴스 브랜드를 유료로 구독하고 있는지 물었다. 미국에서 뉴욕타임스가 지배적인 역할을 하고 있었고, 핀란드에서 헬싱킨사노마트(Helsingin Sanomat)가 강력한 구독자 기반을 갖추고 있는 등 여러 시장에서 승자독식 현상이 어떻게 전개되고 있

【그림 1】 지난 1년 동안 온라인 뉴스에 비용을 지불한 독자 비율(20개 주요 국가)



Q7a. 지난 1년 동안 온라인 뉴스 콘텐츠에 비용을 지불했거나 유료 온라인 뉴스 서비스를 이용한 적이 있습니까? (여기에는 디지털 구독, 디지털/인쇄물 결합 구독, 기부, 기사나 앱, 전자판에 대한 1회성 결제 등이 포함됩니다.) Base: 각 국가당 약 2,000명 대상

출처: 디지털 뉴스 리포트 2024

는지 확인할 수 있었다.”⁷⁾

승자독식 현상은 국가별 상위 1~2개 언론사를 제외할 경우 유료 구독자 확보에 높은 허들이 놓인다는 걸 의미한다. 투입되는 비용이나 감내해야 할 매출 감소세에 비해 성과가 뚜렷하지 않을 수 있다는 것으로, 이러한 경향이 누적되면 뉴스 유료화 회의론으로 곧장 번질 수밖에 없다. 특히 미국 시장에서 이러한 패턴은 두드러지게 나타나고 있다. 워싱턴포스트는 유료 구독자 수가 한때 기록했던 350만 명을 회복하지 못하고 있는 반면, 뉴욕타임스는 1,100만 명을 돌파하며 성장세를 이어가고 있다. 이 같은 대비는 전자의 정체와 후자의 확장이 뚜렷한 격차로 나타나고 있음을 보여준다. 뉴스 미디어 유료 구독에 두 개 이상 지출하지 않으려는 구독자의 일반적인 성향을 감안하면 어쩌면 자연스러운 결과라 할 수 있다. 노르웨이나 미국에서만 발생하는 특이한 현상이 아니라 거의 모든 국가에서 보편적으로 발생할 수밖에 없는 공통 현상인 것이다.

7) <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2024/how-much-do-people-pay-online-news-what-might-encourage-others-pay>

구독 플랫폼 개발 기업 주오라(Zuora)의 펠릭스 단작(Felix Danczak) 구독 마케팅 총괄에 따르면,⁸⁾ 언론사들이 디지털 페이지월을 포기하는 이유는 대략 다음 세 가지였다.

- 다른 수익화 수단에 집중하기 위해
- 유료 구독이 광고 감소분을 채워내지 못하는 경우
- 단기간의 비용 절감이 필요할 경우

특히 두 번째 이유가 유료화 포기의 가장 현실적인 동력으로 작동한다. 치열한 유료 구독자 유치 경쟁으로 인해 마케팅 비용과 인건비는 증가하기 마련이지만, 이를 통한 유료 구독 수익이 광고 수익 감소분에 미치지 못하면 뉴스 유료화를 포기하게 된다는 설명이다. 혹은 비용 절감이 급박하게 요구될 경우 유료 구독 활성화를 위해 채용한 인력을 내보내고 관련 비용을 줄이는 형태로 나타나게 된다. 결국 뉴스 유료화 시장의 정체 속에서 나타나는 승자독식 구조가 뉴스 유료화에 대한 회의론을 증폭시키고, 다른 전략적 대안을 찾는 움직임으로 현실화될 수밖에 없게 된 것이다.

2) 유료 구독 양극화와 대안 전략 찾기

뉴스 유료 구독에 대한 회의론은 크게 두세 가지 방향의 선택으로 분기가 됐다. 유료 구독의 성과를 경험한 상위 언론사들은 마케팅과 구독 전환 기술을 더 고도화하고, 관련 번들 및 협력 상품을 추가하고 있다. 반면 실적 하락을 경험한 언론사들은 유료화를 중도에 포기하고 광고 수익 활성화에 매진하거나 또 다른 수익모델에 집중하는 전략을 취하고 있다. 유료 구독 성과의 양극화가 빚어낸 2025년의 독특한 풍경인 셈이다.

디지털 뉴스 유료화 성과의 양극화에 대한 우려는 유료 구독 붐이 일기 시작한 초기부터 제기된 바 있다. 특히 해당 언론사의 규모, 콘텐츠의 독점성, 수익 채널 간의 수요 전환 효과에 따라 성과가 크게 엇갈릴 수 있다는 지적이 적지 않았다(Chung et al., 2019). 언론사의 규모에 따라 페이지월 마케팅 비용을 감당할 수 있는 수준이 달라지고, 지속적인 전환을 이끌어 낼 독점적이고 고유한 오리지널 콘텐츠를 생산할 역량이 갈려서다. 전반적인 트렌드와 분위기에 따라 디지털 페이지월을 도입하긴 했지만, 그 성과는 투자 여력에 따

8) <https://www.adweek.com/media/time-paywall-dropped-one-year/>

라 엇갈릴 수밖에 없었다는 것이다. 이러한 초기의 예상은 2022년 이후 들어맞았고, 규모가 모호하거나 투자 역량이 상대적으로 취약한 언론사들은 다른 전략을 검토하기에 이르렀다.

이러한 배경에 따라 나타난 세 가지 흐름은 다음과 같다.

1. 유료 구독 성공 언론사: 유료 구독 성장에 더 많은 자원을 투자하는 전략
2. 유료 구독 성과가 미미한 언론사:페이월을 내리고 광고 수익 강화에 나서는 전략
3. 환경 변화에 따라 대안적인 수익 모델을 검토하거나 비영리로 전환하는 전략

다음 장에서는 이를 구분해 각 사례를 구체적으로 분석해 보고자 한다.

유료 구독 성장 전략의 전환: 번들링 모델의 부상

유료 구독에 대한 피로도와 양극화, 그리고 성장 정체는 유료 구독을 핵심 수익으로 삼고 있는 언론사들에게도 전략적 변화를 요구했다. 뉴스라는 상품만으로는 이 위기 국면을 돌파할 수 없다는 인식이 강해졌다. 번들링 전략은 그래서 불가피했다. 유료 구독 시장의 양극화가 심화하는 가운데, 일정 수준 이상의 성과를 거둔 언론사일수록 번들링 전략에 대한 고민이 더욱 깊어졌다. 문제는 뉴스 외 번들에 결합시킬 다른 콘텐츠나 상품이 존재하느냐였다. 다른 상품군을 보유하고 있는 대형 미디어들은 개별 브랜드의 구독 전략에 집중하는 방식보다 이를 통합하는 번들 상품 개발에 더 많은 자원을 쏟아붓기 시작했다.

이 과정에서 주목할 점은 몇 가지 다른 갈래의 번들링 유형이 나타나기 시작한 것이다. 자사 내 다른 동질 콘텐츠와 번들링하는 방식, 자사 내 이질 콘텐츠와 번들링하는 방식, 타사 언론사 브랜드와 번들링하는 방식 등이 대표적이다. 이러한 번들 상품의 구성 유형은 에르브리히(Erbrich, 2024)의 구독 번들 구분법을 참고하되 일부는 더 세분화해 구분지었다.

[표 1] 번들 상품의 구성 유형

	동질(뉴스) 콘텐츠 번들	이질 콘텐츠 번들
자사 내 제품	헵스테드의 풀 티강(Full Tilgang)	뉴욕타임스의 번들
자사 외 제품	다인포메이션과 블룸버그	-

1) 유료화를 위한 번들의 확장: 자사 내 이질 콘텐츠 협업

(1) 뉴욕타임스의 사례

뉴욕타임스는 자사 내 이질 콘텐츠와 번들링을 시도한 모범 사례다. 사실 번들을 통한 유료 구독 강화 혹은 성장 모델은 뉴욕타임스를 빼놓고 상상하기 어렵다. 전 세계에서 가장 영리한 번들링 전략을 활용하고 있을 뿐 아니라 장기적인 전략 하에서 인수까지 시도한 대표적인 언론사이기 때문이다. 시도에 그치지 않고 그 효과를 증명하면서 ‘승자독식’의 구도를 굳히고 있는 곳이 바로 뉴욕타임스다.

뉴욕타임스가 번들을 통한 유료 구독 확장을 꾀한 것은 대략 2021년쯤이다. 그리고 이듬해부터 본격적인 번들 구독 상품 프로모션에 돌입했다. 사전 준비를 위해 뉴욕타임스 쿡킹(Cooking)을 독립적인 유료 구독 상품으로 만들었고, 게임 등을 추가했다. 오디오 앱 오듬(Audm) 인수(2020년)⁹⁾를 시작으로 쿡킹과의 연관성이 높았던 와이어커터(Wirecutter)를 매입해(2016년 10월)¹⁰⁾ 유료 상품으로 전환(2021년)했고,¹¹⁾ 워들(Wordle)이라는 게임도 추가로 구매(2022년)했다.¹²⁾ 스포츠 전문 미디어 디애슬래틱(The Athletic)을 인수한 것도 번들 전략을 본격화하던 이 즈음(2022년)이었다.¹³⁾ 2020~2022년까지 이어지는 일련의 버티컬 콘텐츠 서비스 매집 과정은 완성도 높은 번들 구독 상품을 출시하기 위한 담대한 전략의 일환이었다고 볼 수 있다. 2021년 3분기 컨퍼런스 콜¹⁴⁾에서 뉴욕타임스 CEO인 메리디스 코빗 레비엔(Meredith Kopit Levien)은 다음과 같이 기대감을 표시했다.

“당사의 핵심 제품인 뉴스의 성장과 함께, (요리, 게임 등) 독립형 제품들의 성공은 다중 제품 번들이 지닌 잠재력을 보여주기 시작했습니다. 교차 프로모션을 더 많이 시도하면서, 뉴스 독자들이 요리나 게임 콘텐츠에도 깊이 관여하는 모습을 점점 더 많이 확인하게 됐습니다. 시간이 지남에 따라 모든 제품을 아우르는 이러한 번들이 (독자들의 유료) 전환을 촉진하고, 가입자들이 더 오래 머무르며 활발히 서비스를 이용하게 할 중요한 동력이 될 것으로 기대하고 있습니다.”

9) <https://www.nytc.com/press/the-new-york-times-company-acquires-audm/>

10) <https://www.nytimes.com/2016/10/25/business/media/new-york-times-company-buys-the-wirecutter.html>

11) <https://www.niemanlab.org/2021/02/the-new-york-times-wants-to-test-wirecutter-as-a-subscription-product/>

12) <https://www.nytimes.com/2022/01/31/business/media/new-york-times-wordle.html>

13) <https://www.nytimes.com/2022/01/06/business/new-york-times-the-athletic.html>

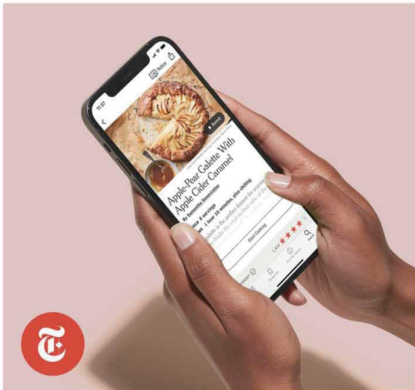
14) https://nytc-assets.nytimes.com/2021/11/Q3_2021-Earnings-Script.pdf

뉴욕타임스의 번들 모델은 ‘자사 내 이질 콘텐츠 통합 모델’의 모범 사례다. 뉴스라는 언론사 고유의 콘텐츠에 한정하지 않고 뉴스가 아닌 이질적인 비뉴스 콘텐츠와 묶어 판매함으로써 구독 상품의 가치를 끌어올리는 유형에 해당한다.

일반적으로 뉴스 그 자체는 시기에 따라서 구독 전환의 부침이 심한 콘텐츠 유형이다. 트럼프 범프(Trump bump)¹⁵⁾ 현상의 수혜를 입기도 하지만 반대의 경우도 적지 않았다. 뿐만 아니라 뉴스는 대규모 구독자 확장에도 여러 어려움과 허들이 존재하는 콘텐츠이기도 하다. 정치적 진영 논리에 휩싸이면 그것의 품질과 관련 없이 이탈이 크게 발생한다. 확장 범위의 제약도 발생한다. 이는 실제 뉴스만 구독하는 구독자 수 통계를 통해 입증된다. 2023년 1분기 358만 명이었던 뉴욕타임스의 뉴스 구독자는 2025년 1분기 179만 명으로 절반 이하로 떨어졌다. 전체 구독자에서 차지하는 비중은 이제 15%에 불과할 정도다. 뉴욕타임스는 폴리처상을 매년 수상할 정도로 품격 높은 저널리즘을 강점으로 삼고 있다. 하지만 이 저널리즘 하나만으로는 1,500만 명이라는 유료 구독자 수 목표치를 달성하기 어렵다는 걸 잘 알고 있었던 셈이다.

[그림 2] 뉴욕타임스 올 액세스(All Access) 번들 구성

Subscribers enjoy more with New York Times All Access.



Apple-Pear Galette With Apple Cider Caramel by Samantha Seneviratne

News
Engage with expert reporting, including culture coverage and analysis.

Games
Unwind with Spelling Bee, Wordle, The Crossword and more.

Cooking
Enjoy delicious recipes, advice and inspiration daily.

Audio
Enjoy full access to shows on our apps, Apple Podcasts and Spotify.

Wirecutter
Choose products confidently with reviews and real-world testing.

The Athletic
Follow in-depth, personalized coverage of your favorite sports.

▶ 출처: 뉴욕타임스

15) 도널드 트럼프 미국 대통령의 각종 논란과 발언이 뉴스에 대한 관심을 높이며, 뉴욕타임스 등 주요 언론사의 구독자 수가 급증한 현상을 뜻한다. 편집자 주

2025년 현재 뉴욕타임스의 번들 기반 구독 상품에는 이제 뉴스뿐 아니라 게임, 쿠킹, 오디오, 와이어커터, 디애슬래틱이 모두 포함돼 있다. 이에 맞춰 그들의 강조점도 ‘세계 최고 수준의 뉴스 보도’에 ‘선도적인 라이프스타일 제품을 갖춘 다양한 포트폴리오’가 추가됐다. 이제 뉴스와 저널리즘만을 강조하는 언론사에서 벗어나 라이프스타일 제품군의 가치를 뉴스와 동급으로 놓기 시작한 것이다. 뉴스와 동급이 된 라이프스타일 제품의 통합 번들은 구독 이탈률의 감소와 ARPU¹⁶⁾ 상승을 견인하는 결과로 이어지고 있다. 2022년 2분기 8.83달러였던 ARPU는 2025년 1분기 9.54달러까지 뛰어올랐다.¹⁷⁾ 번들 구독자 비중도 어느덧 50%에 육박했다. 번들의 힘이 그들의 목표인 1,500만 명 구독자 달성을 현실로 만들어주고 있는 것이다.

그렇다고 위기가 없었던 것은 아니다. 2025년 2월 뉴욕타임스 주가가 큰 폭으로 하락했는데, 이는 구독 성장세가 시장의 기대에 미치지 못해서다. 번들을 통한 구독 성장 전략을 공세적으로 펼쳐왔지만 잠재 고객 도달이 한계에 다다르면서 이러한 현상이 벌어진 것이다. 뉴욕타임스는 위기 국면을 탈피하기 위해 새로운 번들 전략, 타 언론사 구독 상품과의 교차 번들링을 시도했다. 특히 엘파이스(El País, 스페인), 폴리티켄(Politiken, 덴마크), 아이리시타임스(the Irish Times, 아일랜드), 코리에레델라세라(Corriere della Sera, 이탈리아) 등 유럽권 언론사와 교차 번들링 상품을 출시하며 도달 확장을 시도했고 결과는 꽤나 만족스럽게 나타나고 있다.¹⁸⁾

(2) 쉽스테드의 풀 틸강 번들 모델

2022년 노르웨이 미디어 그룹 쉽스테드(Schibsted)의 디지털 통합 구독 상품 ‘풀 틸강(Full Tilgang)’¹⁹⁾ 도입은 지난 10년간 디지털 구독 성장 둔화에 대한 직접적인 대응이었다. 이는 반전의 모멘텀과 경쟁 우위를 유지하기 위한 혁신의 시급한 필요성을 대변한다. 구독자에게 향상된 가치를 명확히 보여주고 이를 통해 유지율(retention)을 높이려는 전략이었다. 이 번들은 급변하는 미디어 소비 환경에서 장기적 생존 가능성을 보장하는 데 중요한 젊은 독자층에 소구하도록 설계됐다.

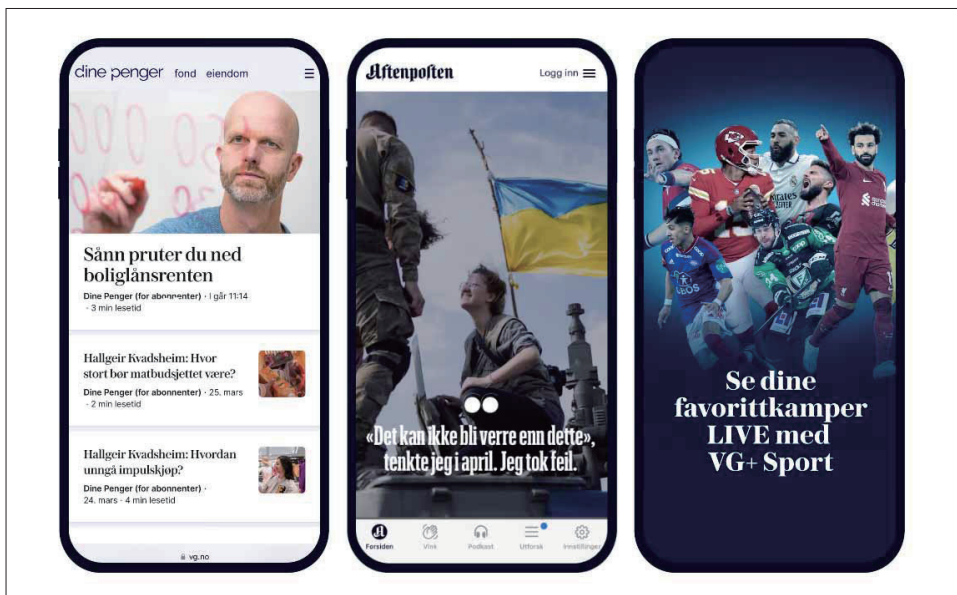
16) ARPU는 ‘Average Revenue Per User’의 약자로, 사용자당 평균 매출을 의미한다.

17) <https://nytco-assets.nytimes.com/2025/05/Q1-2025-Earnings-Presentation-Final-For-Distribution-beXb6ryl.pdf>

18) <https://www.niemanlab.org/2025/03/new-york-times-bundles-gives-european-publishers-a-subscription-boost/>

19) <https://fulltilgang.schibsted.no/>

[그림 3] 풀 틸강(Full Tilgang)의 번들 콘텐츠



출처: 쉽스테드

쉽스테드는 디지털 구독 성장률 감소세를 언급하며 풀 틸강이 이러한 도전에 대한 직접적인 해결책으로 출시됐다고 밝힌 바 있다.²⁰⁾ 쉽스테드와 같은 시장 선두주자에게도 개별 디지털 구독의 초기 성장세가 정체될 수 있음을 암시한다. 이들의 번들링 전략은 단순히 뉴스 제품 개선을 넘어 새로운 성장 동력을 확보하기 위한 근본적인 처방이라 할 수 있다. 콘텐츠 가치와 제공 폭을 크게 늘림으로써, ‘구독 피로’를 극복하고 독자들의 관심과 지출에서 더 큰 점유율을 확보하는 것을 목표로 했다. 이탈에 대한 방어적 조치이자 ARPU 및 유지율을 높이는 공격적인 전략을 모두 상징하는 것이다.

쉽스테드는 2022년 가을, 풀 틸강 번들 상품을 공식 출시하며 디지털 구독 전략의 이정표를 세웠다.²¹⁾ 다양한 신문, 광범위한 팟캐스트, 수많은 잡지에 대한 접근을 단일 구독으로 통합해 포괄적인 디지털 미디어 번들을 제공한다. 이 서비스는 노르웨이의 6개 주요 신문, 팟미(PodMe)와 같은 독점 팟캐스트(Krimpodden, AVHØRT, Papaya, Big 5 with

²⁰⁾ <https://www.twipemobile.com/the-power-of-bundling-how-publishers-can-offer-more-value-to-subscribers/>

²¹⁾ <https://www.m24.no/aftenposten-bergens-tidende-e24/schibsted-lanserer-alt-i-ett-abonnement-for-sine-nyhetsmedier/543821>

Nils and Harald, Ida med hjertet i hånden 등 인기 타이틀 포함), 그리고 44개에서 60개에 달하는 그룹 내 잡지를 포함하고 있다. 특히 노르웨이의 전국, 지역 및 금융 뉴스 브랜드들이 혼합되어 있다. 브이취(VG), 아프텐포스텐(Aftenposten), 스벤스카다그블라데트(Svenska Dagbladet), 아프톤블라데트(Aftonbladet), 옴니(Omni), 베르겐스티덴데(Bergens Tidende), 스타방에르아프텐블라드(Stavanger Aftenblad)와 같은 쉽스테드의 뉴스 브랜드 포트폴리오와 E24(노르웨이 최대 온라인 비즈니스 신문으로 'E24-podden' 및 'Stopp verden'과 같은 팟캐스트도 호스팅)는 번들의 가치 향상에 도움을 주고 있다.

번들은 명분상 '다양한 신문, 팟캐스트, 잡지'를 포함하며, 전국 구독자에게 "지역 및 현지 콘텐츠는 그들이 달리 소비하지 않았을 수도 있는 잠재적으로 유용한 추가 가치"라고 언급하고 있다. 이는 단순히 콘텐츠를 패키징하는 것을 넘어, 쉽스테드의 다양한 포트폴리오 내에서 교차 활용 및 콘텐츠 발견을 촉진하는 힘으로 작용했다. 독자가 어쩌면 개별적으로 찾아보지 않았을 수도 있는 제품을 탐색하도록 장려하는 방식이다. 대규모 미디어 그룹 내의 콘텐츠 사일로(content silo)²²⁾ 문제를 해결하고 모든 콘텐츠에 쉽게 접근하고 발견할 수 있도록 해주기도 한다. 풀 톨강은 콘텐츠 전반에 걸쳐 사용자 참여를 극대화하도록 설계된 일종의 큐레이션된 콘텐츠 생태계인 셈이다.

풀 톨강은 특히 젊은 구독자를 유치하고 장기적인 참여를 유도하기 위해 전략적 가격 모델도 활용하고 있다. 여느 번들 모델처럼 초기 할인(예: 첫 달 79 NOK)으로 제공되고, 프로모션 기간이 끝나면 정가(예: 30일당 349~379 NOK)로 전환된다. 이를 바탕으로 쉽스테드는 젊은 독자층을 성공적으로 유치해냈다. 특히 팟캐스트 플랫폼 '팟미'의 역할이 컸다. 쉽스테드는 2019년 첫 투자를 시작으로 2021년 팟미를 완전히 인수했다. 그리고 이듬해 풀 톨강 번들에 편입시켰다. 그들도 인정하다시피 팟미의 젊은 층 구독 전환 효과는 다른 뉴스 브랜드에 비해 상당히 높은 편이었다.²³⁾ 2023년 4분기 쉽스테드의 결산자료에는 "디지털 구독 서비스는 환율 변동 영향을 배제한 기준에서 12%의 견고한 두 자릿수 기본 매출 성장률을 기록했다. 이 성장률은 ARPU 개선과 매출량 증가가 결합된 결과로, 팟미와 풀 톨강 번들 서비스의 지속적인 성장도 주요 요인으로 작용했다"라고 적시돼 있다.²⁴⁾

22) 조직 내 각 부서나 브랜드가 콘텐츠를 독립적으로 관리하면서, 다른 부서나 플랫폼과 정보 공유와 연계가 원활하지 않게 되는 구조적 단절 현상을 의미한다. 이로 인해 사용자 경험이 단편화되거나, 중복 투자와 비효율이 발생하는 문제가 나타날 수 있다. 편집자 주

23) <https://www.kom24.no/kristin-ward-heimdal-podcast-podme/podme-sjefen-om-gigantunderskudd-okningen-i-kostnadene-er-derfor-som-forventet/630503>

24) <https://kommunikasjon.ntb.no/ir-files/17847482/2896/4038/Interim%20Report%20Q4%202023.pdf>

풀 티강은 현재 쉑스테드 디지털 구독 수익의 10% 이상을 차지하며 빠르게 핵심 수익 동력으로 자리매김했다. 뿐만 아니라 구독자의 60%가 번들 내에서 세 개 이상의 제품을 활용하고 있고, 이 구독자들은 유료 구독을 갱신할 가능성이 더 높다는 성과도 보고됐다.

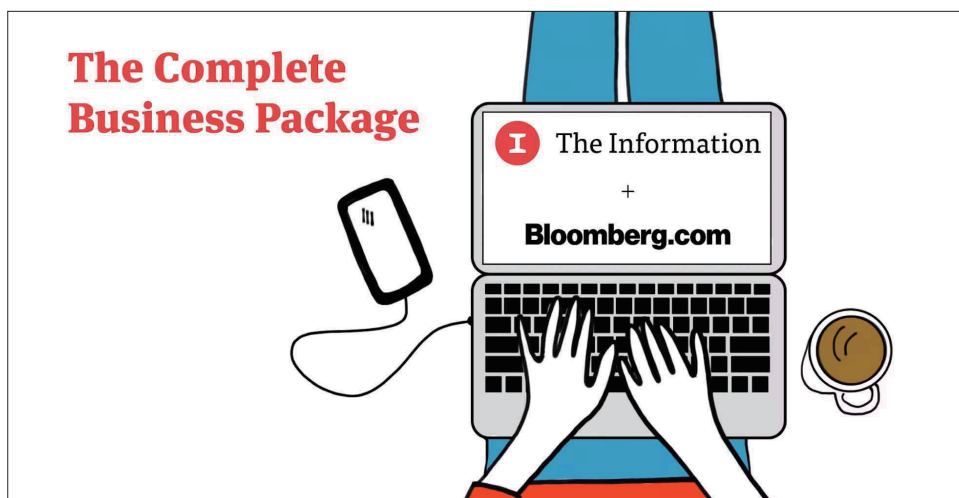
풀 티강의 성공은 단순한 구독자 유치를 넘어, 이들을 번들 사용자로 효과적으로 전환한 데서 비롯됐다. 쉑스테드 통합 번들링 전략은 보다 안정적이고 예측 가능한 수익 흐름으로 이어져 소비자 잉여를 증가시킬 뿐 아니라, 지불 의사와 독자 관심 측면에서 더 큰 점유율을 확보하고 이탈 위험을 줄일 수 있다는 명제를 확인시켜주고 있다.

2) 언론사 간 협업 모델: 자사 외 동질 콘텐츠 협업

(1) 디인포메이션과 블룸버그의 프리미엄 미디어 번들

디인포메이션(The Information)과 블룸버그(Bloomberg)의 번들 파트너십은 프리미엄 비즈니스 미디어 업계에서 협업을 통한 새로운 가능성을 제시한 사례다. 양사 간 번들링 논의는 2019년 9월 24일 엑시오스(Axios)의 보도로 처음 알려졌다. 이후 2020년 2월 19일, 공식적인 파트너십이 발표됐다.²⁵⁾ 제시카 레신(Jessica Lessin) 디인포메이션 편집장은 “뉴

[그림 4] 디인포메이션과 블룸버그의 더 컴플리트 비즈니스 패키지(The Complete Business Package)



출처: 디인포메이션

²⁵⁾ <https://www.theinformation.com/articles/a-bundle-with-bloomberg-media>

스를 통합하려는 플레이어는 애플, 페이스북, 구글뿐이었다”며, “언론사들이 협력해 양질의 저널리즘을 위한 독자층을 확대해야 한다”고 강조했다.

이 시기 유료 구독 환경은 경쟁이 치열해지고 있었으며, 언론사들끼리 공동 구독 번들을 판매하는 것이 새로운 트렌드가 될 수 있다는 분석이 제기됐다. 이러한 맥락 속에서 두 언론사 간의 파트너십은 상당한 주목받았다. 디인포메이션의 연간 구독료는 399달러, 블룸버그의 디지털 구독료는 415달러로 총 814달러였으나, 번들 가격은 499달러로 책정돼 구독자들은 315달러를 절약할 수 있었다. 번들 수익은 두 회사가 50대 50으로 분할하기로 합의했으며, 이는 한쪽이 더 많은 구독자를 유치하더라도 수익을 동등하게 나누는 구조였다. 월간 구독자들은 블룸버그를 포함시켜 업그레이드할 수 있었고, 연간 구독자들에게는 추후 해당 기능이 제공되는 방식이었다. 이 번들은 1년간 제공되는 한시적 구성으로, 블룸버그 구독 페이지에는 프로모션 가격이 1년 동안만 적용되며 이후 블룸버그 정기 구독 요금인 연 415달러가 청구된다고 명시돼 있다.

두 회사는 고급 비즈니스 구독자라는 타겟 고객층이 일치한다는 특징이 있었다. 디인포메이션과 블룸버그의 개별 구독료는 일반적인 소비자 구독료보다 높은 편이었다. 콤스코어(Comscore)에 따르면 2019년 12월 기준 디인포메이션은 10만 3,000명의 순 방문자를 기록한 반면, 블룸버그는 2,600만 명 이상의 순 방문자를 기록했다. 레신 디인포메이션 편집장은 “내가 번들 모델에 긍정적일 수 있는 또 하나의 이유는 뉴스 업계의 경쟁 방식 때문”이라며, “뉴스는 결코 ‘승자독식’이 아니다”라고 강조했다. 그는 “소프트웨어와 달리 뉴스 업계는 최대한 많은 매체가 최고의 작업을 할 때 최상의 상태가 된다”며, “서로의 성공을 응원할 수 있다는 점이 우리 사업의 매우 독특한 점”이라고 설명했다.

2019년 디지털 뉴스 리포트에 따르면 뉴스 구독료 지불을 선호하는 사람들조차 보통 하나의 구독만 선택하는 경향이 있다. 이러한 점을 고려할 때, 블룸버그와 디인포메이션의 전략은 동일한 독자가 기술 및 비즈니스 언론사 중 하나만 선택하도록 경쟁하기보다는 499달러를 양사가 나누는 것이 합리적이라는 판단에 따른 것이다. 파트너십에는 백엔드 시스템 통합이라는 기술적인 과제도 있었다. 디인포메이션을 통해 프로모션에 참여하는 사용자들은 블룸버그 사이트에서 구매를 완료해야 했다. 레신 편집장은 이 모델을 더 탐구하기 위해 ‘여러 언론사들’과 대화를 나누고 있다고 말했지만, 추가 번들에 대한 계획은 이후 진행되지 않았다.

이 번들링 파트너십은 구독 기반 미디어 업계에서 중요한 선례가 됐다. 빅테크 기업들처럼 언론사들 또한 접치는 보도 범위에 대한 유사한 파트너십으로 독자들의 지불 의

사를 높일 수 있다는 것을 보여줬다. 존 매카시(John McCarthy) 더드럼 온라인 선임 기자는 “비슷한 성향의 두 언론사가 교차 독자층을 확보하고 통합 패키지를 판매하는 것은 합리적”이라며, “이는 기술 플랫폼에 일방적으로 의존하는 기생적 협력 구조에서 벗어나려는 주체적 시도”라고 평가했다.²⁶⁾

결론적으로 디인포메이션과 블룸버그의 번들 파트너십은 프리미엄 비즈니스 미디어 분야에서 상호 보완적 협력의 성공적인 모델을 제시했다. 특히 기술 플랫폼에 대한 의존도를 줄이고 언론사들이 직접 구독자와의 관계를 강화하는 새로운 방향을 제시했다는 점에서 의미가 깊다. 이 사례는 구독 피로 시대에 언론사들이 경쟁이 아닌 협력을 통해 더 큰 가치를 창출할 수 있다는 사실을 입증했다.

(2) 워싱턴포스트와 파이낸셜타임스의 구독 협력 모델

2020년 9월, 미국 워싱턴포스트(The Washington Post)와 영국 파이낸셜타임스(Financial Times)는 공동 구독이라는 이례적인 파트너십을 발표했다.²⁷⁾ 이는 디인포메이션과 블룸버그 협업에 이은 대형 협업 프로젝트로, 세간의 관심을 모으기에 충분했다. 코로나19 팬데믹이 한창이던 당시, 두 매체의 협력은 위기 속에서 기회를 찾는 전략적 선택으로 풀이

[그림 5] 워싱턴포스트와 파이낸셜타임스의 공동 구독 파트너십 발표문



▶ 출처: 파이낸셜타임스

26) <https://medium.com/whats-new-in-publishing/bullish-about-the-bundle-the-information-and-bloomberg-trial-year-long-joint-subscription-548ec2d2a9e8>

27) <https://commercial.ft.com/news-insights/financial-times-washington-post-announce-joint-digital-offer-new-subscribers/>

됐다. 워싱턴포스트는 2020년 9월 14일, 팬데믹으로 인한 경제·사회 재건 노력을 조명하는 ‘더 로드 투 리커버리(The Road to Recovery)’ 섹션을 선보였다. 같은 날 발표된 파이낸셜타임스와의 파트너십은 워싱턴포스트의 광범위한 보도와 파이낸셜타임스의 국제 비즈니스 전문성을 결합해 시너지 효과를 창출하려는 목적이었다.

파이낸셜타임스는 워싱턴포스트 구독자에게 3개월 무료 구독 혜택을 제공하고, 워싱턴포스트는 파이낸셜타임스 구독자에게 90일 디지털 액세스를 제공하는 방식으로 진행됐다. 워싱턴포스트 구독료는 연간 39~49달러였으며, 파이낸셜타임스 구독 가격은 369.20달러였다. 프로모션 기간은 2020년 9월 14일부터 10월 8일까지 한 달도 채 되지 않았다.

파이낸셜타임스 컨설팅의 피트 두셋(Pete Doucette)은 워싱턴포스트의 글로벌 시장 확장 전략의 핵심 타겟으로 해외 거주 미국인과 국제 뉴스에 관심 있는 독자를 꼽았다. 파이낸셜타임스의 도달력을 빌려와 자사 구독자를 확장시키려는 전략인 셈이다. 워싱턴포스트의 CMO였던 미키 킹(Miki King)은 “파이낸셜타임스와의 협력은 성공적이었다”고 자평하며, 구독 사업이 연초 대비 40% 이상 성장, 특히 글로벌 구독 사업은 60% 이상 증가했다고 밝혔다. 다만 파이낸셜타임스와의 협업이 어느 정도 기여했는지는 별도로 밝히지 않았다.²⁸⁾

워싱턴포스트는 파이낸셜타임스와의 협력 성과를 높게 평가하며, 향후 캐나다 지역 매체들과의 협력도 계획하고 있다고 밝히기도 했다. 팬데믹으로 인한 구독 붐도 워싱턴포스트 성장에 기여했지만, 지역화 전략, 마케팅, 가격 정책, 번들링 등 워싱턴포스트의 꾸준한 노력이 결실을 맺었다는 분석이다.

워싱턴포스트와 파이낸셜타임스의 공동 구독은 한 달 남짓이라는 제한된 기간 동안 진행된 실험적인 파트너십이었지만, 워싱턴포스트의 글로벌 확장 전략에 중요한 이정표가 됐다는 평가를 받는다. 상호 보완적인 강점을 가진 언론사들이 협력을 통해 새로운 독자층을 확보하고 구독자 기반을 확장할 수 있음을 보여준 사례다.

28) <https://digiday.com/media/how-the-washington-post-is-expanding-its-global-subscriber-base/>

뉴스 유료화에서 발 뺀 언론사들

코로나 팬데믹 이후 뉴스 유료 구독을 둘러싸고 새로운 조어들이 등장하기 시작했다. 구독 정점(subscription peak),²⁹⁾ 구독 천장(subscription ceiling) 따위의 구독 회의론과 연결된 용어들이었다. 여기에 뉴스 산업 안에서 뉴스 회피(news avoidance)³⁰⁾라는 유행어도 빠르게 번져나갔다. 구독이 거의 모든 디지털 서비스의 기본 비즈니스 모델로 자리를 잡아가면서 나타난 후속 현상이다.

구독을 둘러싼 이러한 부정적 현상들은 뉴스 유료화 모델로 성장을 꾀하던 일부 언론사들의 고민거리로 자리 잡았다. 승자독식이 강화되는 마당에 구독 1등이 아니면 성과를 내기 어려운 비즈니스 구조를 깨뜨리기란 벅찰 수밖에 없었다. 기대한 만큼의 유료 구독 성장세를 만들어내려면 막대한 마케팅 비용이 지출될 수밖에 없는데, 이 또한 부담이었다. 때마침 코로나 팬데믹 이후 광고 시장이 회복세를 보이면서,³¹⁾ 광고 수익 창출에 대한 희망을 가질 수도 있게 됐다. 페이지를 올렸지만 구독자 증가세는 빠르지 않고, 트래픽은 줄어들어 광고 매출은 감소하자 언론사 경영진의 고민은 더욱 깊어지기에 이르렀다. 몇몇 언론사들을 중심으로 페이지 포기론이 힘을 얻게 된 대략적인 배경이다. 유료 구독으로 기대 매출을 달성하지 못할 거라면 전통적으로 강점을 지니고 있었던 광고 영역에 더 집중하자는 움직임이 2020년대 초부터 본격 등장하기 시작했다.

29) <https://www.theatlantic.com/health/archive/2022/03/why-subscriptions-are-hard-to-cancel/623885/>

30) <https://reutersinstitute.politics.ox.ac.uk/news/five-things-news-media-can-do-respond-consistent-news-avoidance>

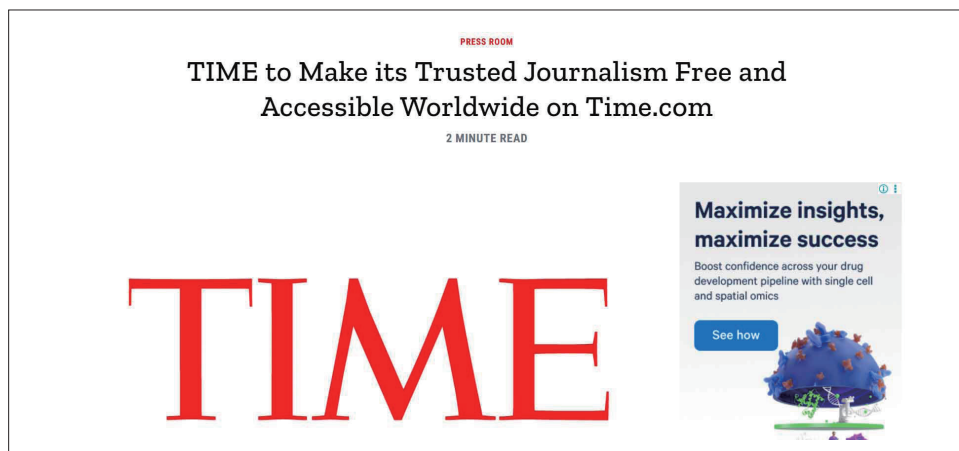
31) <https://www.statista.com/outlook/amo/advertising/worldwide>

1) 타임의 사례

타임은 2023년 6월 1일부터 웹사이트페이월을 완전히 없애 디지털 콘텐츠를 모든 독자가 무료로 이용할 수 있도록 하는 중요한 전략적 전환을 발표했다.³²⁾ 타임의 이러한 결정은 독자 수와 구독자 수가 모두 감소하는 어려운 시기에 이뤄졌다.페이월 제거의 표면적인 이유는 저널리즘을 모두에게 제공하기 위함이었지만, 업계 전문가들에 따르면 이는 지속적인 독자 개발 및 콘텐츠 전략 실패에 대한 전략적 대응이기도 했다.

타임은 2021년까지만 하더라도 2030년까지 1,000만 명 구독자 달성이라는 야심찬 목표를 내세우고 있었다.³³⁾ 이를 위한 구독 및 디지털 혁신을 위한 편집 리더십 구조 변경안까지 발표했다. 하지만 이런 기조는 제시카 시블리(Jessica Sibley)가 2022년 11월 CEO로 임명되면서 변화의 조짐을 보이기 시작했다. 그는 페이월 제거에 대한 입장을 설명하면서 “신뢰할 수 있는 정보에 대한 접근은 전 세계적인 필수 과제이며, 모든 인류가 그 정보를 이용할 수 있어야 한다”라고 밝혔다.³⁴⁾

[그림 6] 타임이 페이월을 내리고 무료로 기사를 제공한다는 안내문



▶ 출처: 타임

32) <https://time.com/6274933/time-to-make-its-trusted-journalism-free-and-accessible-worldwide-on-time-com/>

33) <https://time.com/6050783/time-announces-changes-to-editorial-leadership-structure/>

34) <https://pressgazette.co.uk/paywalls/peak-paywall-what-times-move-to-free-online-model-means-for-news-industry/>

타임의페이월 포기는 포화된 구독 시장에서 ‘도달 범위 대 수익’이라는 딜레마를 상징적으로 보여준다. 타임은 저널리즘을 모두에게 제공하려는 의도를 표명하며 페이월을 없앴다. 이러한 결정은 타임의 독자 수와 구독자 수가 감소하는 상황에서 이뤄졌으며, 이는 유료 구독자를 유치하거나 지키는 데 어려움을 겪고 있음을 시사한다. 뉴스 산업에서 최근 몇 년간 페이월 채택이 크게 증가했지만, 동시에 구독 증가율은 둔화되고 있었다. 이는 구독 시장이 성숙해져 잠재적으로 포화 상태에 도달했음을 나타낸다. 타임의 움직임은 저널리즘 가치로 포장됐지만, 실제로는 경쟁이 치열하고 성장세가 둔화되는 구독 시장에서 전략적 전환을 의미한다. 만약 한 매체가 유료 구독자를 유치하거나 유지하는 데 어려움을 겪는다면, 페이월의 기회비용(제한된 도달 범위, 광고 인벤토리 감소, 트래픽 저하, 신규 독자 유치 어려움 등)이 직접적인 구독 수익(특히 그 수익이 이미 감소하고 있다면)보다 커질 수 있다. 이러한 시나리오에서는 전략적 목표가 ‘개별 구독 수익 극대화’에서 ‘더 넓은 독자층 확보’와 ‘브랜드 영향력 재확립’으로 전환되며, 광고, 이벤트 또는 기타 다각화된 수익원을 통해 수익화될 수 있다. 이는 다양한 주제를 폭넓게 다루는 종합 매체 가운데, 유료 구독을 꾸준히 이끌 만큼 콘텐츠가 충분히 차별화되지 않았거나 ‘세계 최고 수준의 저널리즘’으로 평가받지 못하는 매체의 경우에는 구독자 증가가 한계에 이르는 ‘구독 정점(subscriber ceiling)’에 직면할 수 있음을 보여준다.

2) 퀴츠의 사례

비즈니스 중심의 뉴스 사이트인 퀴츠(Quartz)는 2022년 4월 계량형 페이월을 해제해 대부분의 콘텐츠를 무료로 제공했다. 페이월을 폐지한 대신, 자주 방문하는 독자에게는 매월 세 개의 기사를 열람한 뒤 이메일 등록을 요청하는 방식으로 운영 방식을 전환했다. 페이월을 로그인월로 변경한 사례라 할 수 있다. 퀴츠의 CEO이자 공동 창업자인 잭 시워드(Zach Seward)는 이번 결정이 “우리의 사명을 더욱 발전시키고 더 많은 사람들에게 우리의 저널리즘을 알리기 위한 것”이라고 강조했다.³⁵⁾ 그는 독자 설문조사 결과, 약 2만 5,000명의 유료 회원(연간 100달러를 지불) 중 대다수가 독점 콘텐츠 접근보다는 사이트의 사명과 저널리즘을 지원하기 위해 구독했다고 밝혔다. 이는 사명 중심의 멤버십 모델로의 전환을 시사한다.

35) <https://www.nytimes.com/2022/04/14/business/media/quartz-paywall.html>

페이월을 해제했음에도 불구하고, 퀴츠는 독자 기반 수익 모델을 완전히 포기하지 않았다. 그들은 ‘더포어캐스트(The Forecast)’, ‘위크엔드브리프(Weekend Brief)’와 같은 프리미엄 이메일 뉴스레터를 통해 멤버십을 계속 판매했고, 유료 구독자 전용 콘텐츠를 남겨뒀다.³⁶⁾ 또한, ‘퀴츠재팬(Quartz Japan)’과 같이 성공을 거둔 버티컬 구독 제품 개발을 계속 이어갔다. 하지만 잦은 소유권 손바뀜 속에서 이러한 정책도 일관성을 유지하지는 못했다.

퀴츠는 페이월을 포기하기 전 몇 년 동안 지속적인 비즈니스 문제에 직면해 있었으며, 2020년 초에는 광고 수익 감소로 인해 대규모 해고를 단행하기도 했다.³⁷⁾ 그해 말, 시워드와 직원들은 이전 소유주인 우자베이스(Uzabase)로부터 퀴츠를 인수해 독립적으로 운영했다. 이후 퀴츠는 지오미디어(G/O Media)로 매각됐고 2025년 4월에는 캐나다 소프트웨어 기업인 레드브릭스에 다시 넘어갔다.³⁸⁾ 재정적 어려움에 따른 소유권 불안정성이 페이월 접근 방식을 재고하는 결정에 영향을 미쳤을 가능성이 높다. 잭 시워드가 자신의 블로그에 남긴 글에 그 고충의 흔적들이 세세하게 기록돼 있다.³⁹⁾

퀴츠 또한 유료 구독 성장 전략의 실패에 따른 페이월 철회 사례로 해석할 수 있다. 퀴츠가 페이월을 내릴 당시 전체 매출액 중 유료 구독이 차지하는 비중은 크지 않았다. 당시 CEO였던 잭 시워드가 밝힌 바 있듯 광고가 여전히 회사 수익의 대부분을 차지하고 있어 2019년부터 시행해 온 페이월을 없앨 수 있었다.⁴⁰⁾ 이는 그들의 예상만큼 페이월을 통한 유료 구독 매출 증가가 이뤄지지 않았다는 메시지가기도 하다.

3) 테크크런치의 사례

테크크런치는 2024년 2월 29일 유료 구독 서비스 ‘테크크런치+(Techcrunch+)’를 공식 중단했다.⁴¹⁾ 동시에 모든 콘텐츠의 페이월을 제거했다. 무료 접근을 허용하는 전략적 방향 전환을 단행한 것이다. 이는 뉴스 미디어 업계에서 주목받은 대표적인 구독 서비스 역

36) <https://www.adweek.com/media/quartz-drops-its-website-paywall-in-an-unorthodox-about-face/>

37) <https://www.nytimes.com/2020/05/14/business/media/quartz-to-lay-off-80-employees.html>

38) <https://www.axios.com/2025/04/04/quartz-sold-redbrick-go-media>

39) <https://www.zachseward.com/what-was-quartz/>

40) <https://www.nytimes.com/2022/04/14/business/media/quartz-paywall.html>

41) <https://techcrunch.com/subscribe/>

전환 사례 중 하나다.

테크크런치+는 테크크런치가 운영한 프리미엄 구독 서비스였다. 월 15달러 또는 연 99달러의 구독료로 제공됐다. 서비스는 심층 분석 기사와 독점 콘텐츠를 포함했다. 스타트업 빌드업 가이드와 검증된 전문가 목록도 제공했다. 광고 제거 기능과 테크크런치 이벤트 할인 혜택, 빠른 읽기(Rapid Read) 기능까지 지원하는 종합적인 프리미엄 서비스였다. 테크크런치는 2019년 ‘엑스트라크런치(Extra Crunch)’라는 유료 구독 서비스를 시작한 뒤, 2021년 테크크런치+로 브랜드를 변경했다.⁴²⁾ 이러한 유료화 전략은 전통적인 광고 수익 모델의 한계를 극복하고, 보다 안정적인 수익원을 확보하기 위한 목적에서 비롯됐다. 특히 독자에게 심층적이고 전문적인 콘텐츠를 제공하려는 의도도 담겨 있었다.

테크크런치는 자사 웹사이트를 통해 테크크런치+ 서비스의 중단을 공식 발표했다. 회사는 2024년 초 해당 결정을 알렸으며, 즉시 신규 구독 접수를 중단했다. 2024년 2월 29일에는 서비스를 완전히 종료하고, 다음 날인 3월 1일부터 모든 콘텐츠를 무료로 전환했다. 서비스 종료와 함께 기존 테크크런치+ 콘텐츠는 메인 웹사이트에 통합됐으며, 빠른 읽기 기능도 전 사용자에게 무료로 제공됐다. 기존 구독자에게 제공했던 이벤트 할인 혜택 또한 2024년 2월 29일부로 종료됐다.

테크크런치는 서비스 중단의 공식적인 이유를 “전 세계 독자들에게 업계 최고 수준의 보도와 콘텐츠에 대한 접근성을 제공하기 위함”이라고 명확히 밝혔다. “스타트업 생태계에 제한 없는 접근을 제공”하는 것도 중요한 목적 중 하나였다. 이를 통해 테크크런치는 “재능 있는 기자들의 가치 있는 인사이트가 담긴 최신 뉴스에 무제한으로 접근할 수 있는 환경을 만들고자 했다”라고 설명했다.

페이월 포기 결정은 테크크런치의 비즈니스 모델에 근본적인 변화를 가져왔다. 수익 구조의 중심이 구독 수익에서 광고 및 이벤트 수익으로 이동했다. 더 큰 독자층을 기반으로 광고 가치도 증대됐다. 이벤트 참가자 확대를 통한 수익 다변화 역시 이뤄졌다. 콘텐츠 전략 측면에서도 변화가 있었는데, 모든 콘텐츠를 무료로 제공함으로써 접근성을 극대화하고 스타트업 생태계 전반에 대한 영향력을 키웠다. 전반적인 전략은 브랜드 가치와 신뢰도를 높이는 방향으로 재조정됐다.

테크크런치의 페이월 해제로 구독 모델의 한계와 지속가능성에 대한 재검토가 필요하다는 인식이 더욱 확산됐다. 표적 독자층의 특성에 따른 차별화된 수익 모델의 필요성

42) <https://techcrunch.com/2021/09/27/introducing-techcrunch-plus/>

도 제기됐다. 공공재적 성격의 저널리즘 정보에 대한 접근성 확보 중요성이 재조명을 받게 된 것이다. 테크크런치+의 중단은 단순한 서비스 종료가 아니다. 전략적 방향 전환의 결과로 평가된다. 단기적으로는 여러 성과를 거뒀다. 구독 수익을 포기함으로써 더 큰 시장에서 점유율을 확보했으며, 스타트업 생태계에서 필수적인 정보원으로서 입지를 더욱 강화했다. 이로 인해 광고주들에게도 한층 더 매력적인 플랫폼으로 자리잡게 됐다.

4) 와이어드의 사례

미국의 테크 전문 매거진 와이어드(Wired)는 앞선 사례와 달리 페이지를 부분적으로 해제한 특별한 사례다. 저널리즘의 명분과 실체가 일치하는 유형에 해당되기도 한다. 와이어드는 2025년 4월 언론자유재단(Freedom of the Press Foundation) 기고문을 통해 정보자유법(FOIA, Freedom of Information Act) 기반으로 작성된 기사에 대해서는 페이지를 적용하지 않겠다고 발표했다.⁴³⁾ 와이어드의 글로벌 편집 디렉터이자 언론자유재단 이사인 케이티 드러몬드(Katie Drummond)는 이 기고문에서 조건부 페이지 해제에 대해 다음과 같이 설명했다.

[그림 7] 와이어드 기사에 게재된 무료 공개 안내 문구

<i>WIRED has made this article free for all to read because it is based on reporting from Freedom of Information Act requests or public government documents. <u>Please consider subscribing to support our journalism.</u></i>	The Department of Justice has argued that extensive DNA collection activity at the border provides “an assessment of the danger” a migrant potentially “poses to the public” and will essentially help solve crimes that may be committed in the future. Experts say that the children’s raw genetic material will be stored indefinitely and worry that, without proper guardrails, the DNA dragnet could eventually be used for more extensive profiling.
--	--

▸ 와이어드는 저널리즘의 공공성과 투명성을 이유로 정보자유법(FOIA)에 기반한 정부 문서 보도를 전면 무료로 제공하고 있다. 그림은 해당 사실을 와이어드 기사에 안내문 형태로 명시한 모습. 출처: 와이어드

⁴³⁾ <https://freedom.press/issues/wired-is-dropping-paywalls-for-foia-based-reporting-others-should-follow/>

“트럼프 행정부가 출범한 첫 몇 달 동안, 와이어드 매거진과 같은 매체에서 발행하는 획기적인 탐사보도를 접할 수 없는 독자들에게는 이러한 경험이 특히 불만스러웠을 것이다. 다행히도 와이어드에는 해결책이 있다. 정보자유법을 통해 입수한 공공 기록에 주로 기반한 기사를페이월링하는 것을 중단하는 것이다.”

정보자유법 기반으로 작성된 기사는페이월 창 대신 “와이어드는 이 기사를 모든 사람이 읽을 수 있도록 무료로 제공했습니다. 이는 정보자유법 요청이나 공개된 정부 문서에 기반한 보도 내용이기 때문입니다”라는 문구가 나타난다. 비록 모든페이월을 제거하는 결정을 내린 것은 아니지만, 저널리즘 가치와 독자 기반 수익 창출 간의 균형이 가능한 모델을 제시했다는 점에 의의가 있다. 팬데믹 기간 동안 일부 공공적 기사에 대해페이월을 해제한 사례들은 많았지만, 특수한 상황을 제외하고페이월을 부분적으로 내린 경우는 와이어드가 처음이라고 할 수 있다. 포오폴미디어(404 Media)와 같은 소규모 뉴스 스타트업들도 이러한 움직임에 동참한 바 있지만, 와이어드라는 큰 규모의 언론사가 이러한 정책적 결정을 내린 것은 이례적이다.

와이어드는 2018년 처음으로페이월을 도입했다.⁴⁴⁾ 첫 구독자는 연간 12달러로 모든 디지털 콘텐츠에 접근할 수 있다. 2025년 2월 와이어드의 디지털 유료 구독자는 18만 명을 넘어섰다.⁴⁵⁾ 구독자 규모가 아직 제한적인 상황에서, 일부 기사에 한해페이월을 해제하기로 한 결정은 결코 쉽지 않았을 것으로 보인다.

44) <https://www.niemanlab.org/2018/02/learning-from-the-new-yorker-wireds-new-paywall-aims-to-build-a-more-stable-financial-future/>

45) <https://pressgazette.co.uk/paywalls/digital-subscribers-100k-club-ranking-worlds-biggest-paywalled-news-publishers-2025/>

페이월 외 대안의 수익모델 찾아나선 언론사

1) 가디언의 사례: 독자 기여와 페이월 병행 전략⁴⁶⁾

전 세계 언론사들이 페이월 모델 도입에 집중하는 사이 영국의 가디언은 완전히 다른 접근법을 선택했다. 가디언은 비영리 신탁 기구인 스콧 트러스트(Scott Trust)의 지원을 받는 독립적인 구조를 바탕으로 ‘저널리즘 독립성’을 강조하며 독자들의 신뢰를 구축했다. 이러한 독립성은 가디언이 미션 중심의 저널리즘을 추구할 수 있는 재정적 기반이 됐다. 독자들로 하여금 언론 자유와 민주주의 수호라는 가치에 공감하도록 만들기도 했다.

가디언의 수익 모델 전환은 미국 비영리 모델에서 영감을 얻어 시작됐다. 초기에는 소규모 구독 사업과 광고 모델을 통해 무료 오픈 웹 접근법을 유지했으나, 편집부와 커머셜팀 간 협력을 통해 독자들에게 직접적으로 지원을 요청하는 새로운 방식을 도입했다.

[표 2] 가디언의 연도별 수익 구조 변화

(단위: 파운드(£))

	2023/2024	2024/2025	변화율
총 수익	257.8M	275M	+6.7%
디지털 독자 수익	88.2M	100M	+20%
디지털 독자 수익 비중	34.2%	36.3%	

출처: 프레스가제트⁴⁷⁾

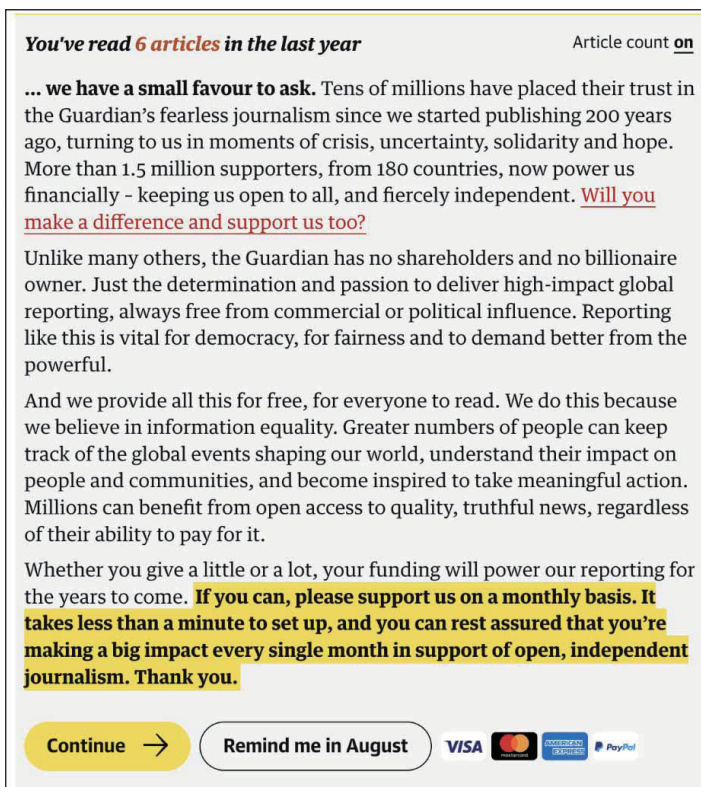
⁴⁶⁾ <https://www.inma.org/blogs/reader-revenue/post.cfm/the-guardian-bypasses-a-paywall-to-find-reader-support>

⁴⁷⁾ https://pressgazette.co.uk/media_business/guardian-grows-revenue-and-reduces-losses-for-year-to-march-2025/

가디언의 성공 핵심은 ‘에픽(EPIC)’이라고 불리는 독자 직접 지원 요청 시스템이다.⁴⁸⁾ 에픽은 웹사이트와 기사 하단에 게시되는 개인화 메시지로 가디언의 정체성과 독립적 저널리즘의 가치를 설명하며 독자의 재정적 지원을 요청한다. 이 시스템은 처음 도입 당시 많은 의구심을 불러일으켰지만, 예상을 뛰어넘는 성과를 보이며 완전히 새로운 수익원을 개척했다.

가디언은 효과적인 후원 유도 메시지를 에픽에 담기 위해 다양한 요소를 실험해왔다. 특히 후원 전환율이 높았던 에픽에는, 독자가 지금까지 읽은 기사 수를 안내하는 ‘읽기 횟수 표시’가 포함돼 있었다. 독자는 자신이 소비한 콘텐츠량을 확인함으로써, 가디언이 제 공한 저널리즘 가치와 공공성을 다시 인식하게 된다. 또한 모금 목표와 그 진행 상황을 시각적으로 보여줌으로써 독자에게 공동의 목적의식을 불러일으키고, 개인화된 메시지를

[그림 8] 가디언의 모든 기사 하단에 붙어서 나타나는 개인화 후원요청 ‘에픽 유닛’



48) <https://dicktofel.substack.com/p/the-amazing-story-of-the-guardians>

통해 독자 후원이 실제로 만들어내는 변화를 설명한다. 마지막으로 월간, 일회성, 연간 등 다양한 후원 옵션을 제공해 독자의 선택권을 보장한다. 특히 주목할 점은 ‘나중에 다시 알림(remind me)’ 버튼의 전략적 활용이다. 이 기능은 즉시 후원하지 않는 독자와도 이메일 관계를 형성할 수 있게 해주며, 향후 정기 후원으로 발전시킬 수 있는 중요한 접점 역할을 한다.

가디언은 사용자 행동과 동기를 분석해 독자를 세 가지 주요 그룹으로 분류하고 있다.⁴⁹⁾ 첫 번째는 미션 중심 서포터(Mission-driven supporters)로, 이들은 언론 자유와 민주주의 보호에서 가디언의 역할을 중시하며 주로 일회성이나 정기 기부를 선택하되 추가 혜택은 요구하지 않는다. 두 번째는 뉴스 애호가(News enthusiasts)로, 높은 참여도를 보이며 향상된 경험을 원하는 독자들로서 앱 접근과 광고 없는 읽기가 포함된 프리미엄 구독을 선호한다. 세 번째는 라이프스타일 지향 독자(Lifestyle-oriented readers)로, 요리 앱 ‘피스트(Feast)’ 같은 특정 콘텐츠에 관심을 보이며 미션보다는 실용적인 관심사 기반 콘텐츠에 더 끌리는 특성을 보인다. 이러한 세분화를 통해 가디언은 각 독자층의 니즈에 맞는 맞춤형 접근을 시도하고 있다.

가디언은 웹사이트와 모바일 앱에서 서로 다른 수익화 전략을 구사한다. 웹사이트에서는 완전히 개방된 접근을 유지하면서 3단계 지원 모델을 제공한다. 가장 낮은 단계는 월 4파운드의 정기 기부로 추가 혜택 없는 순수한 지원 형태이며, 중간 단계는 월 12파운드의 디지털 구독으로 광고 없는 읽기와 앱 접근을 포함한다. 최고 단계는 월 27파운드의 프리미엄 구독으로 인쇄판과 디지털 서비스를 번들로 제공한다. 반면 모바일 앱에서는 전통적인 계량형 페이지를 적용해 20개 무료 기사 이후 구독을 유도한다. 앱 구독료는 월 15달러 또는 연 110달러로 설정돼 있다. 가디언은 “앱은 프리미엄 포맷(Our app is a premium format)”이라고 명시하며, 웹에서의 오픈 저널리즘과 앱에서의 프리미엄 상품이 상호 배타적이지 않다고 설명한다.

이러한 이원화 전략은 가디언이 독자에게 보다 유연한 선택지를 제공하는 동시에, 충성도 높은 사용자로부터는 최대한의 수익을 이끌어낼 수 있도록 해준다. 일회성 기부든, 월간 지원이든, 앱 구독이든 모든 독자를 포용하는 이 접근법은 독자들과의 신뢰 관계를 구축하고 지속적인 참여를 유도하는 데 효과적이다.

49) <https://www.inma.org/blogs/reader-revenue/post.cfm/the-guardian-bypasses-a-paywall-to-find-reader-support>

가디언 후원-페이월 병행 모델의 핵심 성공 요인은 명확한 미션과 독립성에 대한 독자들의 신뢰다. “스콧 트러스트의 지원을 바탕으로 가디언을 영구적으로 지속시키고, 독립적이며 용기 있는 저널리즘과 고품질 탐사보도를 제공할 것이다”는 명확한 목적 의식은, 독자들로 하여금 단순한 뉴스 소비를 넘어 저널리즘 자체를 지원하고 있다는 공감대를 형성하게 한다. 가디언은 에픽의 언어, 길이, 디자인 등 모든 요소에 대해 성과와 시사점을 바탕으로 지속적인 테스트와 개선을 진행한다. 이러한 데이터 중심 접근법은 독자들의 반응을 실시간으로 반영해 지원 요청의 효과를 극대화하는 데 기여해왔다.

가디언의 최우선 과제는 현재 600만 명인 무료 이메일 구독자를 1,000만 명으로 확대하는 것이다. 이는 고품질 저널리즘을 기반으로 기부, 구독과 같은 고가치 행동을 촉진할 수 있는 직접 소통 채널을 강화하려는 전략의 일환이다. 이메일은 독자와의 직접적이고 지속적인 관계를 유지할 수 있는 가장 효과적인 채널로 평가된다.

가디언의 사례는 페이월이 뉴스 미디어 수익화의 유일한 해결책이 아닐 수 있음을 보여준다. 독자와의 신뢰 관계 구축, 명확한 미션 중심의 저널리즘, 그리고 다양한 지원 옵션 제공을 통해 지속 가능한 수익 모델을 구축할 수 있다는 걸 증명해 보이고 있다. 가디언의 ‘후원-페이월 병행 모델’은 전통적인 뉴스 미디어 수익화 전략에 대한 혁신적 대안을 제시하고 있다. ‘독자 수익 1억 파운드 달성’과 ‘130만 정기 구독자 확보’라는 성과는 미션 중심 저널리즘, 독자와의 신뢰 관계, 그리고 유연한 지원 옵션이 결합된 결과다.

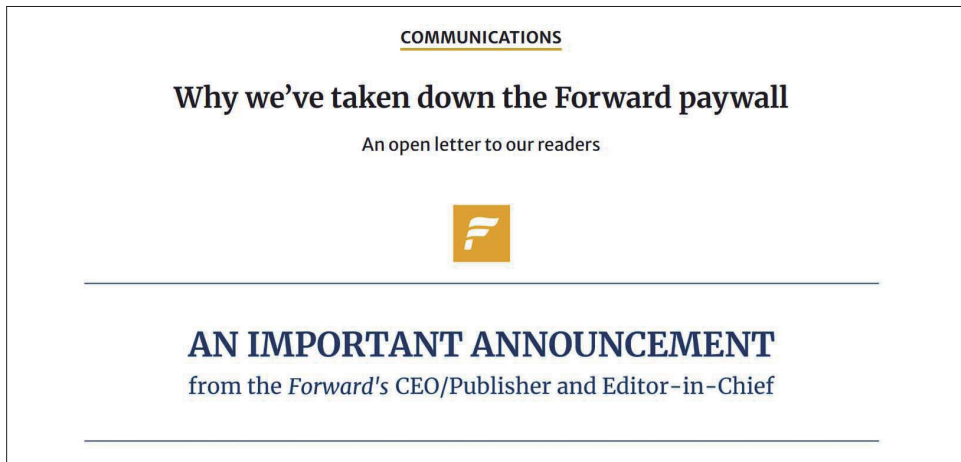
2) 더포워드의 사례: 페이월을 포기하고 멤버십으로

미국의 유대인 공동체 전문 비영리 미디어 더포워드(The Forward)는 2023년 12월 5일, 모든 콘텐츠에 대한 페이월을 제거하고 그들 역사상 처음으로 무료 기반의 독자 후원 멤버십 프로그램으로 전환했다.⁵⁰⁾ 재정상의 문제로 페이월을 내리고 광고 비즈니스로 넘어간 앞선 사례와 달리 멤버십 프로그램으로 재구성한 새로운 유형의 언론사라고 할 수 있다. 게다가 더포워드는 비영리 언론사라는 특별한 지위를 갖고 있기도 하다.

수십 년 동안 더포워드는 연간 약 500만 달러 상당의 손실을 기록했다. 운영을 유지하기 위해 그들의 다수 자산을 매각하는 데 의존해야 했다. 2002년에는 7,300만 달러 규모의 AM 주파수와 3,250만 달러 상당의 FM 주파수 슬롯을 매각했으며, 2010년에는 미

⁵⁰⁾ <https://forward.com/communications/571935/forward-paywall-membership-jewish-journalism/>

[그림 9]페이월을 내리고 멤버십으로 전환한다는 더포워드의 안내문



출처: 더포워드

드타운에 위치한 건물을 1,850만 달러에 처분했다.⁵¹⁾ 이러한 일련의 자산 매각 행보는 상당한 수익을 보충해줬지만, 장기적인 재정 건전성을 보장하지 못했다. 그저 인쇄 신문을 유지하는 데 활용할 수 있는 규모였다.

2008년 금융 위기는 한 달 만에 900만 달러의 급격한 손실을 초래했으며, 이 조직은 또한 마도프 폰지 사기⁵²⁾에 휘말려 35만 5,000달러를 손해보기도 했다. 2015년에는 잠재적인 명예훼손 소송과 같은 책임으로부터 자산을 보호하기 위해 대부분의 자산을 별도의 비영리 단체인 포워드펀드(Forward Fund)로 이전했다. 급기야 2019년 121년간 이어왔던 인쇄판 발행을 중단하고 편집자를 포함한 직원 20%(10명)를 해고하기까지 했다.

페이월 해제를 선언한 명분 중 하나는 “지역 사회가 번성하는 데 필수적인 고품질 정보를 제공하고 모든 사람이 무료로 접근할 수 있도록” 한다는 더포워드의 전통적인 사명이었다. 더포워드의 명시된 비전은 “공유된 역사에 기반을 두고 정의, 연민, 진실, 그리고 분열을 메우려는 공동의 책임에 전념하는 활기차고 연결된 글로벌 유대인 공동체”다. 그 사명은 “미국 저널리즘에서 가장 중요한 유대인 목소리”가 되어 독자들에게 영감을 주

51) <https://www.timesofisrael.com/the-forward-ran-low-on-cash-and-miracles-before-ending-121-years-in-print/>

52) 전직 나스닥 회장이자 투자운용사 대표였던 버나드 메이도프(Bernard Madoff)가 약 20년간 벌인 초대형 금융 사기로, 기존 투자자에게 돌아갈 수익금을 신규 투자자의 돈으로 지급하는 방식의 '폰지 사기(Ponzi scheme)'를 활용했다. 2008년 금융위기 속에 사기가 드러났고, 피해액은 약 650억 달러로 알려져 있다. 편집자 주

고 세대와 공동체의 다양한 부문 간의 대화를 활성화하는 것이다. 전쟁과 허위 정보, 반유대주의가 확산되는 혼란 속에서, 더포워드는페이월 해제가 필수적이라고 강조했다. 이는 단순한 공공 서비스 제공을 넘어, 더포워드의 유대인 관련 저널리즘이 갖는 고유한 가치가 더욱 중요해졌음을 시사한다. 특히 공동체 내에서 신뢰할 수 있는 정보에 대한 필요성이 커지는 시기에, 더포워드의 역할이 더욱 부각되면서 이러한 결정에 이른 것으로 보인다.

더포워드는 대안을 마련하기 위해 뉴스리베뉴허브(News Revenue Hub)와 파트너십을 맺고 광범위한 독자 분석 및 조사를 진행했다.⁵³⁾ 그 결과 대다수의 독자들은 유료 구독을 기부로 전환했을 때 더 많은 보도를 지원하기 위해 기부 금액을 더 늘릴 의향이 있는 것으로 나타났다. “독자들은 더포워드가페이월을 없애기를 원했고, 그 대가로 특별한 혜택을 원하거나 기대하지 않았으며, 단순히 더포워드의 저널리즘을 모두가 이용할 수 있도록 돕고 싶어 했다”는 것이다. 이러한 설문 조사 결과는페이월을 제거하는 것이 “기존 수익을 위태롭게 하지 않을 것”이라는 확신을 심어줬다.페이월을 제거하기로 한 결정은 오랜 재정적 어려움을 겪고 있는 언론사들에게 상당한 위협을 수반하게 된다. 더포워드는 이러한 위협을 완화하기 위해 엄밀한 심층 독자 분석, 타 뉴스룸과의 벤치마킹을 시도했다. 독자들의 기부 의사를 확인하고 잠재적 수익을 예측한 데이터 기반 접근 방식을 바탕으로 이사회 승인을 얻어 냈다.

페이월 제거를 위한 준비는 2023년 내내 진행됐으며, 전 부서에 걸쳐 상당한 규모의 작업이 동반됐다. 광범위한 데이터 마이그레이션, 웹사이트 업데이트 및 메시징 작업이 이뤄졌다. 가장 큰 운영 과제 중 하나는 세 개의 다른 공간에 저장돼 있던 후원자 데이터를 단일한 고객관계관리(CRM) 플랫폼으로 통합하는 것이었으며, 이는 뉴스리베뉴허브 기술 팀의 보조를 받았다. 후원자 데이터를 CRM으로 통합하는 작업은 기부자 기반 모델을 효과적으로 관리하는 데 반드시 필요하지만 겉으론 잘 드러나지 않는 기술적 인프라 과정이다.

더포워드의 과감하고 실험적인 전환 전략은 독자 수익의 상당한 증가로 이어졌다.페이월 제거 후 첫 3개월(2023년 12월 5일 ~ 2024년 3월 15일) 동안 더포워드는 5,000달러 미만의 기부금으로 약 58만 3,000달러를 확보했다. 이는 전년 동기 유료 구독 수익 대비

53) <https://americanpressinstitute.org/revenue-ama-how-the-forward-grows-audience-across-generations/>

37% 증가한 놀라운 수치였다.⁵⁴⁾ 그해 12월에는 전년 동기 대비 5,000달러 미만의 독자 수익이 103%나 증가했다. 더포워드는페이월이 후원 기반의 멤버십 모델로 전환되더라도 충분히 성공할 수 있다는 걸 보여준 사례에 해당한다.

⁵⁴⁾ <https://www.journalism.co.uk/news-commentary/the-forward-removed-its-paywall-and-generated-37-per-cent-more-revenue/s6/a1176966/>

AI, 뉴스 유료화에 어떤 영향 미칠까

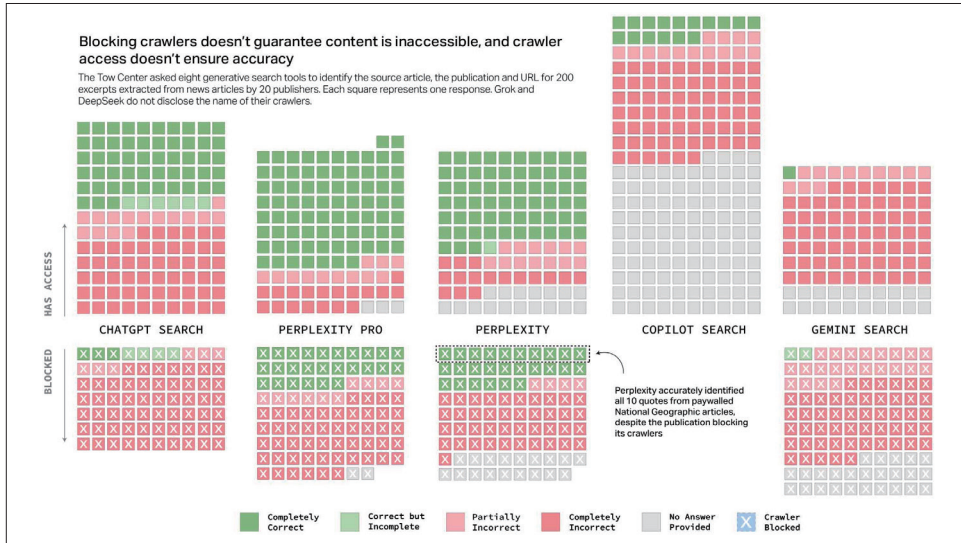
2025년 들어 뉴스 유료화에 영향을 미치는 새로운 변수가 등장했다. AI와 AI 검색이다. 현재 AI는 다양한 기술과 애플리케이션의 형태로 언론사들의 뉴스 유료화에 직접적인 영향을 미치고 있다. 유료 구독을 성장시키는 기술적 기반으로 작용하는 경우가 있는 반면, 디지털페이월 기술 자체를 무력화하는 도구로 활용되고도 있다. AI 플랫폼이 공개한 거대언어모델 API는 유료 구독을 신장시키기 위한 다양한 추천 기술의 장벽을 큰 폭으로 낮추는 데 도움을 줬다. 반면 챗GPT를 비롯해 구글의 AI 오버뷰·AI 검색 모드, 퍼플렉시티(Perplexity) 등 AI 기반 검색 기술은 뉴스가 소비되는 방식의 근간을 변화시키고 있다. 무엇보다 제로 클릭(Zero-Click)⁵⁵⁾과 답변 엔진으로 대변되는 이들 AI 검색의 정보 제시 구조는 검색-유입-전환이라는 뉴스 유료 구독의 전통적인 전환 경로 자체를 재설정할 것을 압박하고 있다. 뿐만 아니라 페이월에 가둬 둔 콘텐츠를 허락없이 긁어가 답변에 포함시키는 빈틈도 드러나고 있어 ‘뉴스 유료화가 이 환경에서 정상적으로 작동할 수 있는가’에 대한 회의감을 확산시키고 있다.

1) 페이월을 파괴하는 AI 봇: robots.txt 규칙 무시

AI 검색 플랫폼이 언론사가 뉴스 크롤링을 차단하기 위해 사용하는 표준 도구인 robots.txt 규칙을 무시하고 콘텐츠를 스크래핑하거나 검색하고 있다. 오픈AI(OpenAI)와 앤

⁵⁵⁾ 사용자가 검색 결과에서 링크를 클릭하지 않고도 검색 페이지 내에서 바로 정보를 얻는 행위를 말한다. AI 챗봇이나 요약 기능이 발달하면서 콘텐츠 소비가 출처 웹사이트 방문 없이 이뤄지는 경우가 늘고 있으며, 이는 언론사의 트래픽 감소와 수익 악화로 이어질 수 있다는 우려를 낳고 있다. 편집자 주

[그림 10] AI 검색의 크롤러가 부정확하게 답변했거나 차단했던 콘텐츠를 긁어가 답변에 활용한 사례를 시각화한 자료



출처: CJR⁵⁵⁾

트로픽(Anthropic)이 뉴스 미디어 사이트에서 robots.txt 지시를 우회하는 사례가 보고된 바 있으며, 퍼플렉시티의 크롤러도 포브스닷컴(Forbes.com)의 robots.txt를 무시하고 뉴스를 긁어가기까지 했다.⁵⁷⁾ 특히 퍼플렉시티의 무료 버전은 퍼플렉시티 크롤러를 허용하지도 않았고 이들과 공식 제휴 관계도 없었던 내셔널지오그래픽(National Geographic)의 유료 기사 10개를 모두 정확하게 식별해 답변에 포함시켰다.⁵⁸⁾ 이러한 비합리적 행위는 AI 검색 사업자와 언론사 간의 첨예한 갈등 이슈로 급부상하고 있다.

robots.txt 규약을 무시하는 행위는 AI 플랫폼 기업들의 '데이터 갈증'에서 비롯됐다. 가장 가치 있는 뉴스 콘텐츠(특종 보고, 포럼, 뉴스 아카이브 등)는 종종 페이지 뒤에 있거나 명시적인 비허용 규칙으로 차단돼 있다. 최근 들어 다양한 형태의 제휴로 언론사들의 독점 콘텐츠를 유료로 구매하는 경우가 늘어나고 있지만, 모든 AI 검색 플랫폼이 이러한 태도를 보이는 것은 아니다.

⁵⁶⁾ https://www.cjr.org/tow_center/we-compared-eight-ai-search-engines-theyre-all-bad-at-citing-news.php

⁵⁷⁾ <https://www.reuters.com/technology/artificial-intelligence/multiple-ai-companies-bypassing-web-standard-scrape-publisher-sites-licensing-2024-06-21/>

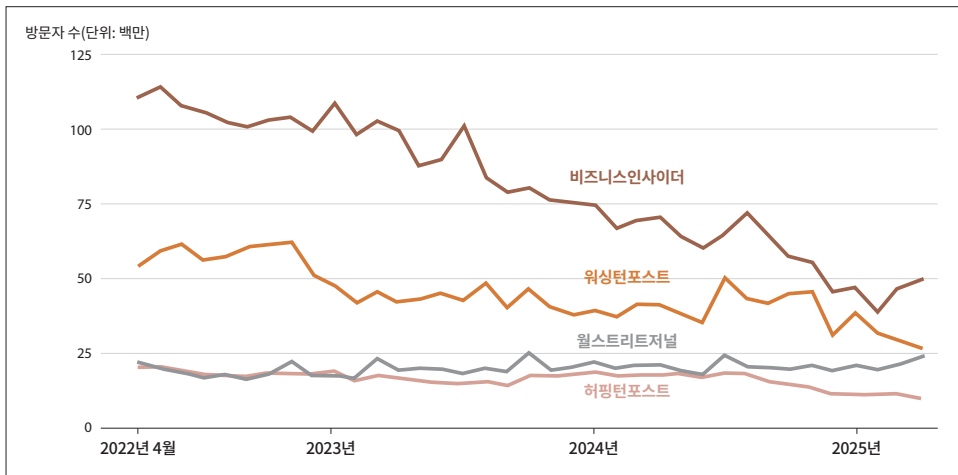
⁵⁸⁾ https://www.cjr.org/tow_center/we-compared-eight-ai-search-engines-theyre-all-bad-at-citing-news.php

언론사들은 robots.txt 파일과 법적 위협만으로는 무단 스크래핑을 막기에 불충분하다고 판단하고 있다. 이를 보완하기 위해 현재 서버 측 도구에 투자하고 있다. 일부 AI 봇은 수천 개의 IP 주소를 순환하며 인간의 브라우징 행동을 모방해 페이지를 우회하기도 한다. 이들은 페이지 구현의 약점(예: 보안되지 않은 API, 예측 가능한 접근 패턴, 불완전한 세션 유효성 검사)을 악용해 프리미엄 콘텐츠에 접근한다. 이러한 무단 접근은 언론사의 브랜드 인지도, 권위, 콘텐츠 수익화 능력을 약화시키고, 경쟁 우위 상실, 콘텐츠 오인, 서버 부하 증가 및 비용 상승으로 이어질 수도 있다. 언론 산업 단체 뉴스/미디어얼라이언스(News/Media Alliance)는 AI 기업들이 소속 언론사의 뉴스 콘텐츠를 보상 없이 활용하고 있다며 저작권 침해 소송을 제기하기도 했다. AI 검색 플랫폼의 무단 스크래핑은 뉴스 콘텐츠의 고유 가치를 떨어뜨리고, 페이지의 존재 이유를 약화시킬 수 있다는 점에서 ‘페이지를 파괴하는’ 현실적 위협 요인으로 작용할 수 있다.

2) AI 검색발 제로 클릭의 위험

AI 검색 크롤러가 표준 규약을 무시하는 것만큼이나 뉴스 유료화에 영향을 미치는 요소가 있다. 바로 AI 검색이 불러올 제로 클릭 경향이다. 다수의 언론사들은 구글 검색 등을 통해 독자를 유입한 뒤 페이지창에 부딪히게 함으로써 유료 구독 전환을 이끌어

[그림 11] AI 검색 확산 이후 주요 언론사의 월별 검색 유입 트래픽 변화



▶ 비교: 데스크톱 및 모바일 웹 검색 총합 기준

▶ 출처: 시밀러웹 데이터를 워싱턴포스트가 재구성함

낸다. 이미 여러 연구를 통해 밝혀진 사실이기도 하지만, 한 번의 클릭만으로는 결코 유료 구독 결제가 발생하지 않는다(Overby et al., 2024). 최소 10회 이상 페이지창과 만났을 때 독자들은 구독을 결심하게 된다. 문제는 구글 AI 오버뷰(AI Overviews)나 AI 모드(AI Mode)와 같은 AI 검색이 본격적으로 검색 결과창을 뒤덮게 된다면, 검색을 통한 유료 결제 유도는 사실상 불가능한 상황에 직면하게 된다는 것이다. 또다시 유료 구독 회의론이 번지게 되는 계기가 될 수도 있다.

[그림 11]에서 보듯, AI 검색이 초래한 검색 유입 트래픽 하락은 이미 현실이 됐다. 비즈니스인사이드나 월스트리트저널, 워싱턴포스트 모두 페이지를 통한 유료 구독을 시도하고 있는 언론사들이지만, 검색 유입 트래픽은 빠른 속도로 하락하고 있다. 비즈니스인사이드의 경우 3년 전에 비해 유기적 검색 유입 트래픽이 55%나 떨어졌다.⁵⁹⁾ 그나마 유입을 촉진하기 위해선 AI 검색 내 가시성을 높이는 전략이 제안되고 있다. 하지만 AI 검색에 가시성이 확보되더라도 클릭이 발생하는 확률은 매우 낮다는 통계도 나오고 있다.⁶⁰⁾ 페이스북과 같은 소셜 미디어 유입도 줄어든 마당에 독자와 직접 만날 수 있는 웹상의 접점이 점차 좁아지고 있는 것이다.

몇몇 언론사들은 이미 이에 대한 방비를 해왔다. 장기적으로 검색 유입이 줄어든 것을 대비해 자체 유입 창구를 늘려나간 것이다. 뉴스레터의 확대가 그런 차원에서 이뤄졌다. 자체 마케팅 퍼널(funnel)을 늘리거나 브랜딩 전략을 강화하는 노력도 경주하고 있다. 하지만 이는 높은 마케팅 비용을 감당할 수 있는 대형 언론사들 위주로만 시도되고 있다. 자본력이 취약한 소규모 페이지 언론사들은 독자 직접 유입을 확대하기 위해 다른 대안을 모색해야 하는 상황이다.

3) AI를 무기로 페이지를 성장시킨 사례

반면, AI는 페이지 전략의 효율성을 높이고 구독자 확보 및 유지에 기여하는 강력한 도구로도 활용되고 있다. AI는 단순히 효율성을 높이는 것을 넘어, 개인화된 경험을 제공하고 새로운 수익 모델을 창출함으로써 페이지의 가치를 증대시키는 데 핵심적인 역할을 하고 있다.

59) <https://www.wsj.com/tech/ai/google-ai-news-publishers-7e687141>

60) <https://www.emarketer.com/content/google-ai-overviews-decrease-ctrs-by-34-5-per-new-study>

1. 하이브리드페이월 최적화 및 효율성 증대

인도의 주요 디지털 미디어 기업인 '타임스인터넷(Times Internet)'은 '구독 의향' 모델과 같은 AI 기반 예측 알고리즘을 활용해 사용자가 유료 구독자로 전환될 가능성을 예측하고, 이에 따라페이월 배치를 최적화한다.⁶¹⁾ 이러한 전략적 AI 활용은페이월 효율성을 약 50% 향상시킨 것으로 보고됐다. 스페인의 디지털 뉴스 매체인 엘문도(El Mundo) 역시 AI 기반 동적페이월(dynamic paywall) 모델로 완전히 전환해, 편집팀이 수동으로 기사를 잠글 필요 없이 AI가 콘텐츠 접근을 자동화하도록 했다. 이 시스템은 사용자 행동, 광고 잠재력, 구독 전환 가능성을 실시간으로 분석해 각 방문자로부터의 수익을 최적화한다.

2. 푸시 알림 이탈 최소화과 트래픽 증가

AI는 독자 행동 데이터를 분석해 고도로 개인화된 푸시 알림 시스템을 구축하고, 이를 통해 독자 참여와 유지율을 향상시킨다. 미국 뉴스데이(Newsday)⁶²⁾의 경우, AI 기반 개인화 노력으로 일일 평균 알림 수를 175% 늘렸고, 그 결과 콘텐츠 트래픽이 38% 증가했다. 이러한 개인화된 알림은 오래된 콘텐츠를 재순환시켜 뉴스 기사의 소비 수명을 연장하는 데도 기여했다. 당연히페이월로의 유입도 증대시켰다. 워싱턴포스트는 AI가 생성한 질문을 함께 제공하는 '컨버세이션(Conversations)' 댓글 시스템을 도입해 독자의 댓글 참여를 유도했다. 또한 기사 상단에는 세 가지 핵심 요점을 요약한 AI 생성 문단을 배치해 독자의 이해도를 높였다.⁶³⁾

3. 광고 및 광고주와 독자의 정확한 매칭

광고의 효율을 높이고 수익을 확대하는 데 AI는 이제 필수적인 도구로 자리 잡아가고 있다. 뉴욕타임스와 월스트리트저널은 이 영역에 가장 돋보이는 결과를 입증하는 사례다.⁶⁴⁾ 뉴욕타임스의 경우 브랜드매치(Brandmatch)라는 광고 상품에 AI 기술을 접목해 광고 클릭률(CTR)을 32% 이상 향상시켰다. 광고주의 광고 캠페인에 가장 맥락이 적합한 내

61) <https://www.inma.org/blogs/conference/post.cfm/tech-powered-paywalls-increase-conversions-at-news-companies>

62) <https://www.newsday.com/>

63) <https://www.washingtonpost.com/pr/2024/12/16/washington-post-launches-new-commenting-experience-exclusively-subscribers/>

64) <https://www.inma.org/blogs/Generative-AI-Initiative/post.cfm/ai-helps-publishers-monetise-audience-improve-workflows>

부 뉴스를 AI 기반으로 분류하고, 관심이 높을 것으로 추정되는 독자를 찾아내는 데 AI를 활용하고 있다. 월스트리트저널 역시 AI를 활용해 기업의 브랜드 콘텐츠를 가장 관련성 높은 독자에게 노출하는 광고 모델을 운영하고 있다. 내부에선 이를 주제별 AI(Thematic AI)라고 부른다. 유사한 모델은 대만의 유나이티드데일리뉴스(United Daily News)에도 적용돼 있다.

4. 뉴스룸 워크플로우 개선과 기사 생산 효율화

한층 고도화한 에이전틱 AI(Agentic AI) 기술을 활용해 뉴스룸의 워크플로우를 개선한 사례도 소개되고 있다. 독일의 이펜디지털(Ippen digital)은 단일 거대언어모델을 활용하는 방식에서 벗어나 에이전트 기반 기술을 편집국 업무에 적용하고 있다. 엑소코텍스(exocortex)라는 이 시스템은 기자들의 업무를 잘게 쪼개 소규모 프롬프트로 나눈 뒤 개별 에이전트에 맡기고, 그 결과물을 다시 종합해 하나의 기사를 완성한다. 전형적인 에이전틱 기반 AI의 유형이다. 예를 들어 넷플릭스와 관련한 기사를 작성한다고 가정하면, 에디터 에이전트가 제목과 요약을 생성하고, 다른 에이전트가 관련 이미지를 생성해 이를 조합하는 형태다. 관련 기사를 찾아내는 에이전트도 별도로 구성돼 있다. 이펜디지털의 CTO인 마르쿠스 프란츠(Markus Franz)는 “소프트웨어와 기능을 확장해 더 나은 결과를 더 짧은 시간 내에 달성할 수 있다”⁶⁵⁾고 말했다.

위에 언급한 활용 사례들은 AI가 페이지 전략을 보완하고 수익을 강화하는 데 필수적인 도구라는 점을 보여준다. 더 높은 수준의 저널리즘을 생산하는 데도 보조적인 수단으로 활용할 가치가 있다는 사실도 증명해주고 있다. 그러나 AI는 효율성과 예측에 뛰어나지만, 저널리즘 가치 중심의 수익 모델을 창출하는 데는 한계가 있다. 자동화된 콘텐츠는 종종 인간 저널리즘이 제공하는 깊이와 미묘함, 독창성 등이 결여돼 있다. 더 많은 언론사들이 유사한 AI 도구를 공통으로 활용함에 따라 AI의 경쟁 우위는 곧 감소할 수 있다. 핵심은 페이지가 콘텐츠에 분명한 가치를 부여함으로써, 뉴스 미디어의 안정적인 수익원을 확보하고 독자와의 직접적인 관계를 구축하며 콘텐츠 품질 향상을 유도한다는 것이다. 이는 AI가 효율성을 높이는 보완적인 도구일 뿐, 수익 지속 가능성을 위한 근본적인 해결책은 여전히 고품질 저널리즘 기반의 강력한 페이지 전략에 있음을 시사한다.

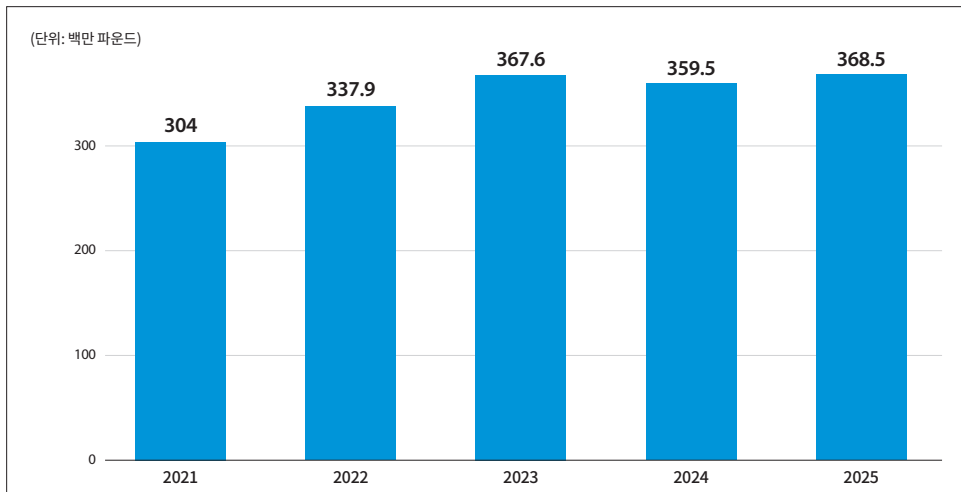
65) https://www.inma.org/modules/event/2025GenAIForMedia/replay/MarkusFranz_INMAMC25GENAI.pdf

AI 도구가 넘쳐나는 현 상황은 ‘무법천지’로 묘사될 정도로 표준이나 프로토콜이 부족하다. 이러한 환경에서 인간의 감독과 ‘인간 개입(human-in-the-loop)’ 정책의 중요성은 강조될 수밖에 없다. 특히 신뢰할 수 없는 AI 생성 콘텐츠의 급증은 오히려 독자들을 신뢰할 수 있는 뉴스 미디어로 다시 이끌 수 있다. AI 과잉의 정보 환경에서 신뢰와 품질은 핵심적인 차별화 요소가 된다. 뉴스 유료화가 작동할 수 있는 기반 환경이 지속될 수 있다는 걸 의미하기도 한다. AI 시대에 저널리즘은 단순히 정보를 전달하는 역할론을 넘어, 독점적이고 고품질이며 신뢰할 수 있는 콘텐츠를 통해 차별화된 가치를 제공함으로써 그 존재 이유를 재정의해야 할 필요가 있다.

나가며: 뉴스 유료화 성장과 쇠퇴 사이, 그래도...

영국의 이코노미스트는 2024년 한해 유료 구독 증가에 힘입어 최고의 매출을 기록했다.⁶⁶⁾ 프랑스의 르몽드도 디지털 유료 구독자의 꾸준한 증가세 등으로 인해 1978년 이후 최고의 발행 부수를 이뤄냈다.⁶⁷⁾ 뉴스 유료화를 둘러싼 여러 부정적 기운이 드리운 상황에서 일궈낸 성과다. 여전히 유료 구독은 뉴스 미디어가 선택해야 할 중요한 수익원 중 하

[그림 12] 이코노미스트의 매출액 추이



출처: 프레스 가젯

⁶⁶⁾ https://pressgazette.co.uk/media_business/economist-subscriptions-profit-revenue-2025/

⁶⁷⁾ https://www.lemonde.fr/en/media/article/2025/06/09/le-monde-group-s-2024-accounts-mark-a-landmark-year_6742164_22.html

나이며, 승자독식의 구조 속에서도 시도해야 할 핵심 수익 과제라 할 수 있다. 유료화 철회라는 반대의 흐름이 일부에서 관찰되고 있다는 사실을 부정하긴 어렵지만, 뚜렷한 대안 수익모델이 도출되지 않는 한, 유료화는 언론사들의 지속 생존을 위해 포기해야 할 대상은 아닌 것이다. 더포워드와 같이페이월을 포기했더라도 독자 기반 수익 모델 자체를 완전히 폐기하지 않은 사례도 주목해야 한다.

뉴스 산업의 유료화 전략에 대한 혼란스러운 입장은 디지털 시대가 초래한 여러 재정적 압력, 기술적 혁신, 변화하는 잠재 구독자 행동에 대한 복잡하고 지속적인 적응 과정을 잘 보여준다.페이월은 인쇄 광고 수익의 급격한 감소에 대한 필수적인 대응 전략으로 등장했지만, 그 구현은 성공과 실패가 엇갈렸다. 월스트리트저널과 같은 전문 콘텐츠 매체는 하드페이월로 성공을 거뒀지만, 일반 뉴스 매체는 초기에는 어려움을 겪었다. 뉴욕타임스가 개척한 계량형 및 동적페이월 모델과 같은 유연한 접근 방식은 독자 확보와 수익 창출 사이의 균형을 맞추는 데 더 효과적이라는 것도 확인됐다.

그러나페이월 도입은 저널리즘의 공공적 역할에 대한 역설적인 영향을 미치기도 했다. 재정적 생존을 위해페이월이 필요함에도 불구하고,페이월은 지역 뉴스 보도를 감소시키고, 미디어의 민주적 책임을 점진적으로 훼손할 위험이 있다. 이는 뉴스 조직이 상업적 생존 전략과 시민적 의무 사이에서 겪는 지속적인 긴장을 보여준다.

인공지능(AI)의 부상은 뉴스 수익화 환경에 새로운 차원의 파괴를 가져왔다. AI 기반 챗봇과 AI 검색은 뉴스 사이트로의 유입 트래픽을 줄이고 콘텐츠의 지식 자산을 도용해 '디지털 광고 붕괴 2.0'을 촉발할 수도 있다. 무단 스크래핑과 저작권 침해에 대한 우려가 커지면서, 언론사들은 AI 크롤러를 차단하고 톨비트(Tollbit)와 같은 혁신적인 솔루션을 통해 스크래핑되는 데이터를 수익화하는 등 대응책을 모색하고 있다. 동시에, AI 기업과의 라이선스 계약은 중요한 새로운 수익원으로 부상하고 있지만, 비공개적이고 파편화된 계약은 언론사들에게 힘의 불균형을 초래할 수 있다는 우려도 제기된다.

이러한 도전 속에서 일부 뉴스 조직은페이월 전략을 완화하고 적응하기 위한 다양한 방법을 모색하고 있다. AI 기반 동적페이월은 사용자 행동을 실시간으로 분석해 광고와 구독 수익을 모두 최적화함으로써 개인화된 접근 방식을 가능하게 한다. 번들링 전략, 특히 뉴욕타임스와 쉽스테드와 같은 사례는 구독자 참여를 크게 늘리고 이탈률을 줄이는 데 매우 효과적임이 입증됐다. 언론사 간 교차 번들 및 비뉴스 서비스와의 결합은 시장 도달 범위를 확장하고 잠재 구독자들에게 더 큰 가치를 제공하는 협업 모델의 중요성을 알려주고 있다. 끝으로, AI는 개인화된 콘텐츠 추천과 정밀한 광고 매칭을 가능하게 해 사

용자 경험을 향상시키고 새로운 수익원을 창출하는 데 핵심적인 역할을 한다. 그러나 AI 기반 수익화에서 개인 정보 보호, 투명성, 알고리즘 편향에 대한 윤리적 고려는 여전히 중요한 과제로 남아 있다.

궁극적으로 뉴스 수익화의 미래는 지속적인 적응, 독자와의 직접적인 관계 구축, 수익원 다각화, 그리고 AI와의 전략적 파트너십과 책임감 있는 활용에 달려 있다. 국내 언론사들의 경영진들은 “독자가 콘텐츠에 기꺼이 비용을 지불할 수 있도록 독자에게 실질적인 가치를 제공하는 모델을 찾는 것 외에는 다른 방법이 없다”라고 말한 덴마크 일간지 폴리티켄 디렉터의 코멘트를 두고두고 기억할 필요가 있다.⁶⁸⁾

참고 문헌

Chung, D. J., Kim, H., & Song, R. (2019). The comprehensive effects of a digital paywall sales strategy. Harvard Business School.

Erbrich, L. (2024). Digital News Bundles: Analyzing Consumers' Willingness to Pay for Cross-Publisher Journalistic Bundles. *Journalism and Media*, 5(3), 1279-1296.

Nguyen, D. (2025). Subscription Fatigue: Analyzing Financial Behavior in a Saturated Subscription Market.

Newman, N., Fletcher, R., Robertson, C. T., Ross Arguedas, A., & Nielsen, R. K. (2024). Reuters Institute digital news report 2024. Reuters Institute for the study of Journalism.

Overby, E. M., Pattabhiramaiah, A., & Kanuri, V. (2024). Please Subscribe to Continue Reading: An Empirical Analysis of Readers' Responses to Newspaper Paywalls. Available at SSRN 5146366.

⁶⁸⁾ <https://www.lanacion.com.ar/economia/negocios/troels-behrendt-jorgensen-el-competidor-de-hoy-puede-ser-el-socio-de-manana-nid08032025/>

3호

‘AI 네이티브 언론’ 전환 로드맵 A to Z

01	들어가며: ‘AI 네이티브 기업’의 등장	081
02	AI 네이티브 언론 구축 가이드	086
03	언론사 조직 차원의 AI 기술 도입 전략: 기술적 선택지	105
04	나가며: 일단 해 봅시다	117

저자 신현규 글리터컴퍼니 CEO

신현규 대표는 언론사에서 20년간 언론 산업 경험을 쌓은 기자 출신 사업가다. 현재 인공지능을 활용해 뉴스를 새로운 포맷으로 자동 가공하는 소프트웨어를 개발하고 있다. 스테이블 디퓨전(Stable Diffusion), 플럭스(Flux), 후안위안(Hunyuan) 등 이미지 생성모델을 1,000회 이상 파인튜닝 했으며, 젤마(Gemma), 라마(LLaMA), OSS-GPT 등 언어모델 역시 수십 회에 걸쳐 파인튜닝한 경험이 있다. RAG(Retrieval-Augmented Generation, 검색증강생성) 모델을 활용한 챗봇, 일정관리 앱 등의 개발 프로젝트를 수행하기도 했으며, 한국의 한 중견기업에서 인공지능 도입 컨설팅을 진행하고 있다. 런던정경대(LSE)에서 커뮤니케이션을 공부하면서 인공지능 기반 콘텐츠 생성 방법론을 연구했으며, 미국 스탠퍼드 대학교의 평생교육과정(Continuing Studies)에서 자연어 처리(NLP) 교육과정을 이수했다. 매일경제 재직 시절 뉴스레터 <미라클레터>를 창간했으며, 실리콘밸리 특파원을 역임했다. (이메일 haneul96@gmail.com)

감수자 이동욱 인프런 CTO

이동욱 인프런 최고기술책임자(CTO)는 소프트웨어 엔지니어링, 제품, 교육 비즈니스의 교차점에 오랫동안 경험을 쌓아온 기술 리더다. 개발자 학습 플랫폼 인프런에서 CTO로 일하며 AI 기반 다국어 자막·더빙, 미국·일본·베트남·중국 글로벌 확장, 강의 제작 효율화, 채용 서비스 ‘멜릿’ 등 핵심 프로젝트를 주도하고 있다. 국내외 테크 생태계의 흐름을 면밀히 살피며, 제품 설계와 데이터 기반 의사결정을 통해 교육 접근성을 높이는 일에 힘쓰고 있다. 저서로는 《스프링 부트와 AWS로 혼자 구현하는 웹 서비스》, 《개발자 원칙》(공저) 등이 있다. 인프런, AWS 서밋 서울(AWS Summit Seoul), 몽고DB 데이 서울(MongoDB Day Seoul) 등 여러 국내 기술 콘퍼런스에서 스타트업의 효율적인 기술 전략을 공유해왔다. 현장에서 쌓은 경험을 바탕으로 블로그(jojoldu.tistory.com)와 유튜브(https://www.youtube.com/@devbadak)를 통해 기술, 리더십, 조직 문화, 기업 성장 전략에 대한 콘텐츠를 꾸준히 발행하고 있다. (이메일 jojoldu@inflab.com)

들어가며: 'AI 네이티브 기업'의 등장

본질적인 관점으로 AI와 언론 산업의 관계를 보는 사람들은 아직 많지 않은 것 같다. 누구나 AI에 대해 말하고 있지만 말이다.

대다수는 AI를 사람을 보완하는 존재로 생각한다. 워싱턴포스트의 미디어 플랫폼인 아크XP(Arc XP)와 미디어 전문지 디지데이가 공동 발간한 <2025년 뉴스룸 내 AI 현황: 제작 공정 자동화를 넘어선 AI의 영향>¹⁾에 따르면 언론사의 97%가 AI 도입을 위한 예산을 늘리고 있다. 국제뉴스미디어협회(INMA)는 이 보고서를 소개하며 “인공지능이 기자들의 직관이나 사명, 독자와의 연결, 그리고 편집에 관한 판단을 재현할 수는 없다”고 선을 그었다. 결국 인공지능이 인간 기자의 존재를 완전히 대체할 수 없고, 따라서 기자나 언론사 운영자들이 그 기술을 활용해 신문, 방송, 유튜브 클립, 카드뉴스 등을 기존과 같이만 들어갈 것이란 전망이다. 하지만 과연 그럴까. 이 글은 그에 대한 파괴적 응답에서 시작한다.

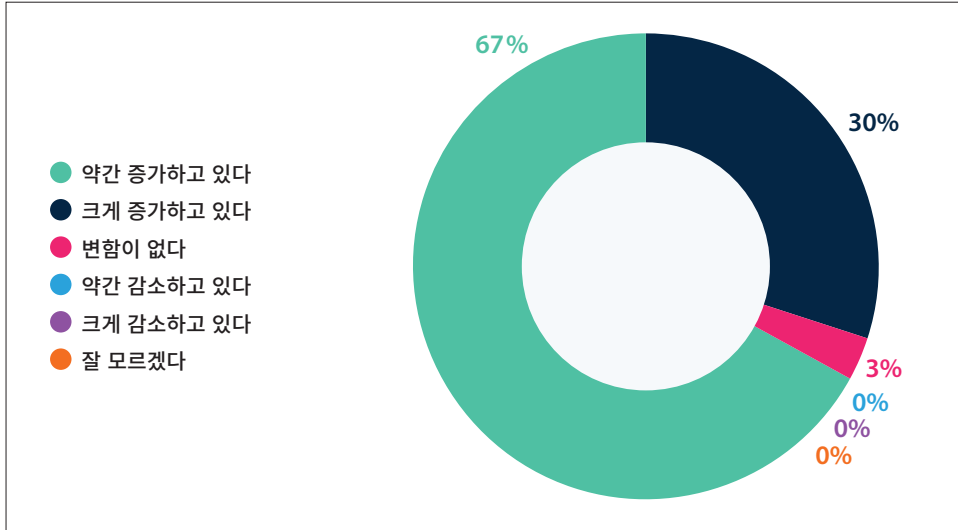
아니다^ㄱ

팔란티어(Palantir)를 보자. 이 회사의 일은 ‘온톨로지(ontology)’란 한 단어로 요약할 수 있다. 온톨로지는 데이터 조각들을 사실, 이벤트, 객체, 관계 등 AI 에이전트가 쉽게 이해할 수 있는 단위로 구분해 놓고, 이들의 연결성을 실시간으로 지도화(mapping)하는 것을 의미한다. 예를 들어 미 국방부가 인공위성 촬영 기록, 드론 촬영 영상, 적군의 통신 기록,

1) https://info.arcxp.com/hubfs/Downloadable%20files/Digiday_State_of_AI_in_the_Newsroom_Report.pdf

정찰 정보 등을 종합해 온톨로지로 해석할 수 있는 데이터베이스에 모아둔다고 상상해 보자. 팔란티어의 온톨로지는 적군의 움직임을 예측할 뿐만 아니라, 전쟁을 언제 어떻게 시작해야 하는지 판단하고 조언할 수 있을 것이다.

[그림 1] 2024년 대비 2025년 미국 뉴스룸 내 AI 도입 예산 변화



▶ 97%(약간 증가하고 있다 67%, 크게 증가하고 있다 30%)의 미국 언론사가 2024년 대비 2025년 뉴스룸 내 AI 도입 예산이 증가했다고 답했다.

* 출처: 아크XP, 디지데이

이를 뉴스룸에 투영해 보자. 언론사와 연계된 공공데이터 API(날씨뿐만이 아니라 정부 기관의 방대한 CCTV 정보, 위성 영상, 통신 기록, 수사 기록, 법원 기록 등을 제공하는 API), 언론사의 자체 취재 메모, 보도 사진 및 영상 아카이브, 독자 반응 데이터 등이 온톨로지로 엮이면, AI는 단순한 취재와 기사 작성을 넘어 ‘지금 취재해야 할 다음 사건은 무엇인가(agenda setting)’, ‘민주주의를 지키기 위해 지금 독자에게 제공해야 할 정보와 역사적 맥락은 무엇인가(fact finding and desking)’, ‘민주주의의 일원으로서 우리 언론이 지금 높여야 할 목소리는 무엇인가(opinion)’까지 제시할 수 있다. 그 판단의 근거를 온톨로지가 제시해 줄 수 있을 것이다.

이것이 AI 네이티브(AI Native) 기업의 모습이다. 언론 산업만 놓고 보면, 취재-편집-보도로 이어지는 언론사 업무를 하나의 워크플로우(workflow)로 간주하고 각 업무를 하나의 노드(node)로 인식한 뒤, 이들을 AI로 자동화하는 것이 ‘AI 네이티브 언론’의 모습이 될 것이다. 실시간 의사결정 엔진을 만드는 것이다.

팔란티어는 오래전 이미 이런 일을 시작했다. 흥미로운 사실은 2015년 오픈AI가 설립될 당시 함께 참여했던 샘 올트먼(Sam Altman) 오픈AI CEO와 일론 머스크(Elon Musk) 테슬라 CEO 모두 피터 틸(Peter Thiel) 팔란티어 회장과 매우 친밀한 관계라는 점이다. 일론 머스크와 피터 틸은 페이스북에서 함께 일하면서 충돌한 적이 있고, 샘 올트먼은 투자회사 와이콤비네이터(Y Combinator) 대표 시절 피터 틸을 자사 이벤트에 여러 차례 초청하며 친분을 보여주기도 했었다. 세 사람은 모두 각자의 영역에서 AI 네이티브 기업을 만들고 있다. 피터 틸은 국방 시스템의 의사결정 엔진을 만들었고, 샘 올트먼은 범용 의사결정 엔진을, 일론 머스크는 조금 더 크게 도시 전체의 교통 시스템을 자율주행이라는 자율적 판단이 가능한 에이전트로 최적화시키는 의사결정 엔진을 만들고 있다. 여기에 (욕심도 많아서) 테슬라 내부 공장에 들어가는 로봇을 활용해 제조업의 의사결정 엔진도 만들고 있다.

AI 네이티브 기업의 등장은 더 이상 먼 미래의 일이 아니다. 기술적으로 AI 에이전트(AI Agent) 기술이 발전에 발전을 거듭하고 있다. 각종 AI 활용 도구의 발달은 복잡한 AI 에이전트로 구성되는 워크플로우를 며칠 만에 설계하고 테스트까지 할 수 있게 하고 있다(물론 실제 운영에 적합한 수준의 안정성과 신뢰성을 확보하기 위한 심층 테스트와 개선 작업은 여전히 상당한 시간과 전문성이 요구된다). AI 덕분에 비개발자와 비전공자도 손쉽게 개발에 참여할 수 있게 됐다. 무엇보다 가장 무서운 것은 자본주의 사회를 움직이는 엔진, ‘돈’의 힘이다. 투자자는 이런 AI 네이티브 기업을 너무나 선호한다. 이전에는 1,000명이 1,000억 원을 버는 회사가 시장을 지배했는데, 이제는 100명이 1,000억 원을 버는 AI 네이티브 기업이 시장에서 빠르게 영향력을 확대하고 있다. 고용 감소는 사회 전체적으로 이익이 아닐 수 있지만, 개별 기업, 투자자, 기존 직원에는 고용 비용이 줄수록 이득이다. 회사가 벌어들이는 순이익이 투자자와 직원의 몫이기 때문이다. 우리가 잘 아는 이미지 생성 솔루션 회사인 미드저니(Midjourney)는 2023년 11명의 인력으로 연 매출 2억 달러(약 2,700억 원)를 달성했다. 러버블(Lovable)은 30명 미만의 인력으로 1,000만 달러(약 130억 원)를, 아라곤AI(Aragon.ai)는 단 11명이 500만 달러(약 65억 원)를 벌어들였다.

실리콘밸리에는 “우리는 이런 회사에 투자하겠다”라고 노골적으로 선언하며 시작하는 모험 자본, 즉 벤처투자자가 많다. 실리콘밸리에서 4년간 일했던 어떤 대형 투자기관의 한 실무자는 한국으로 돌아와 벤처투자 업계와 스타트업 시장을 둘러본 뒤 이렇게 말했다.

“미국에는 AI 네이티브 기업이 많고, 그런 회사에 투자하겠다는 모험 자본이 많은데, 한국에는 그런 스타트업도 많지 않고 그런 회사에 투자하겠다는 곳들도 많지 않은 것 같아요. 왜 그렇죠?”

여러 이유가 있을 것이다. 그중 하나가 바로 이러한 막연한 생각이다.

“AI는 결국 인간을 대체하지 못할 것이다.”

“인공지능이 기자들의 직관이나 사명, 독자와의 연결, 그리고 편집에 관한 판단을 재현할 수는 없다.”

많은 전문가와 유수의 기업조차 AI는 결코 인간을 대체하지 못할 것이라고 주장한다. 그들의 전망이 전적으로 틀렸다고 반박하려는 것은 아니다. 하지만 그러한 주장이 AI가 불러올 현실에 대한 각성을 흐리게 하고, 우리 마음속에 작은 안도감을 심어주는 데 그친다면, 그건 지적 성실성을 저버린 나태함으로 이어질 수 있다고 생각한다. 그리고 이러한 인식이 한국에 AI 네이티브 기업과 그에 투자하는 자본이 많지 않은 이유 중 하나일 수도 있다. AI는 더 이상 단순한 업무 보조 도구가 아니다. 운영의 중심이다. 사업의 존재 방식 자체가 AI를 전제로 설계되는 시대가 오고 있다. AI가 구동되지 않으면 회사는 생명을 잃게 될 것이다. 테슬라, 팔란티어, 오픈AI 같은 기업이 바로 그러한 변화를 보여준다.

언론사에도 같은 질문을 던질 수 있다. “우리는 AI를 기자의 손을 털어주는 도구로 쓸 것인가, 아니면 뉴스 생산·편집·유통·소비의 전 과정을 AI 중심으로 재설계할 것인가?” 결론은 나와 있다. 그저 AI를 기자의 손을 털어주는 정도로 이용하고 넘어간다면 마음은 매우 편안하다. 우리에게 아직 시간이 있고, 세상은 언론사에 몸담은 내가 은퇴할 때까지 평화롭다. 주변을 둘러봐도 한국 언론사 중 AI 도입에 두각을 나타내는 곳은 아직 없다. 오히려 튀는 언론사가 있다면 그들의 실패에 욕을 한 바가지 쏟을 준비가 돼 있다. 특정 언론사가 성공했다는 이야기가 들려오면 금방 따라 할 정도의 자세와 태도를 갖추는 정도로 대비하면 일단은 안심이다. 한국언론진흥재단의 《해외 미디어 동향》 정도로 해외 언론사들은 어떻게 AI 시대에 대비하는지 팔로우하면 더더욱 안심이다.

미안하지만 이 글은 그런 소극적 생각과 결을 달리한다. 정반대로 뉴스 생산·편집·유통·소비 전 과정을 AI 중심으로 재설계한다는 적극적 비전을 갖고 있는 분들에게 이 글이 도움이 되도록 부족하지만 최선을 다해 설계했다. 보다 더 정확하게 이 글은 AI 네이티브 언론을 만드는 방법에 대한 필자의 주관적 방향성을 소개하는 글이라 보는 것이 옳을 것 같다. 왜 이런 목적을 전제로 까느냐 하면, 한국 언론사가 여유를 즐기는 사이, 세상은 급격히 바뀌어 갈 것이기 때문이다. 아마존이 인수한 워싱턴포스트, 개발자들을 대거 채용한 뉴욕타임스 등 AI에 눈을 뜬 언론사들은 이미 많다. 이 글에 등장하는 미국의 가넷(Gannett)처럼 각자 당면한 문제를 제대로 극복할 방안으로 AI를 활용한다면, 우리나라에도 무서운 AI 시대에 대응할 새로운 언론 모델이 탄생할 수 있을 것이다. 그리고 보다 더

무서운 것은 AI 네이티브 인재들이 만드는 새로운 AI 네이티브 언론일 것이다. 애플의 창업자 스티브 잡스(Steve Jobs)의 말을 빌려본다. “지금 이 순간에도 애플을 무너뜨릴 젊은 천재 한 사람이 저기 저 불빛이 시들지 않는 도서관에서 연구하고 있을 것이다. 나는 경쟁자들이 두렵지 않다. 그들이 가장 두렵다.”

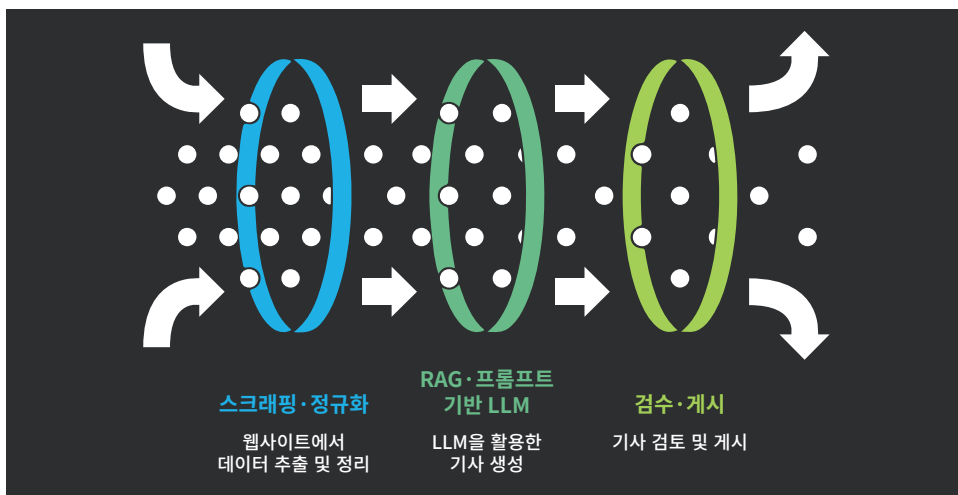
다행스러운 점은 AI 네이티브 언론의 비전이 거창한 구호가 아니라는 점이다. 팔란티어가 컨설팅을, 테슬라가 공장을 재정의한 것처럼, 뉴스룸을 하나의 지능형 운영체제로 전환하는 것일 뿐이다. 취재와 분석, 편집과 발행, 독자 관계와 수익화까지 모든 흐름이 데이터와 모델 위에서 하나의 피드백 루프로 연결된다면, 언론은 더 이상 ‘기사를 발행하는 조직’이 아니라 실시간으로 사회의 의제를 설계하고 조율하는 AI 네트워크로 진화할 것이다. 이 글을 그런 언론사를 만드는 하나의 사업계획서 초고 정도라고 보면 대략 맞지 않을까 한다.

AI 네이티브 언론 구축 가이드

1. AI 네이티브 뉴스 생성 시스템: 가넷의 사례

먼저 언론의 본질적 기능 중 하나인 ‘정보 전달’의 파이프라인을 AI 네이티브로 바꾸는 방법을 사례로 살펴보자. USA투데이를 비롯해 수백 개의 지역 매체를 보유한 미국 언론사 가넷은 각 지역 커뮤니티에 게시되는 공지사항을 스크래핑해 기사 초안을 자동으로 생성하고 있다. 예를 들어 미국 매사추세츠주에 위치한 가넷 소속 언론사 메트로웨스트 데일리(MetroWest Daily)는 보스턴 서쪽에 자리한 중소도시의 시청, 시의회, 구의회는 물론 지역 내 공공기관과 비영리기구가 발표하는 각종 자료와 공지사항, 게시판 글을 수집해

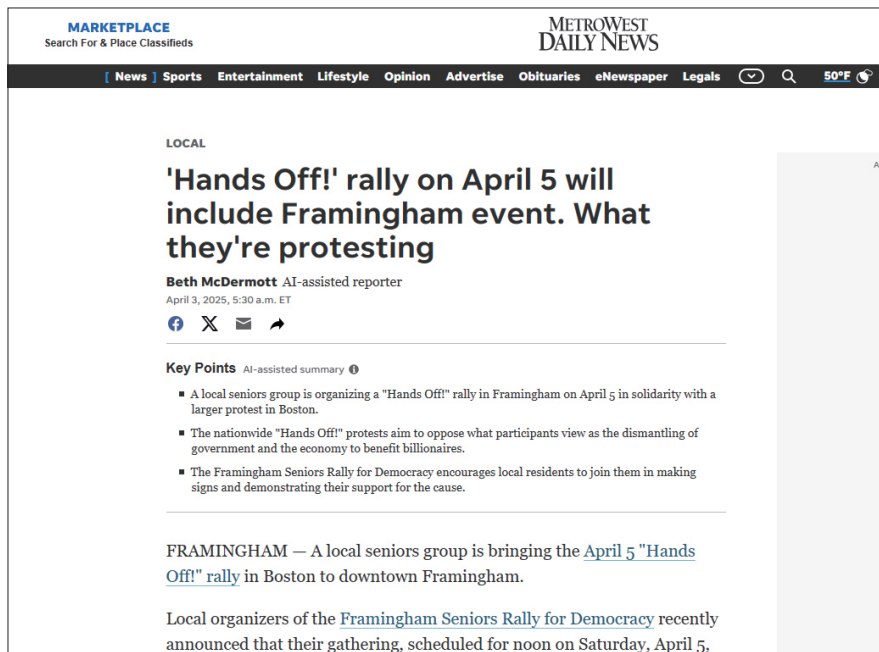
[그림 2] 기사 자동 생성 파이프라인



이를 자동화된 기사로 변환하고 있다. 간단하게 [그림 2]와 같은 워크플로우를 갖춰 두고 있음을 알 수 있다.

그렇다면 한국의 언론사들, 특히 지역에 기반을 둔 언론사들도 이와 유사한 기사 자동 생성 시스템을 도입하는 방안을 충분히 고려해 볼 수 있다. 예를 들어 가상의 지역 '신문읍' 이라는 곳이 있다고 해 보자. 이곳에서는 군이 편성한 주민복지 예산에 따라 매년 읍 단위로 주민들을 위한 작은 축제와 기념품 증정 행사가 열린다. 또한 읍내에 있는 '특종교회'에서는 어려운 처지에 있는 읍내 가정들을 돕기 위해 다채로운 이벤트들을 매달 진행한다. 읍내에 있는 전통 맛집 '편집국밥'에서는 싸고 신선한 재료가 들어왔을 때 지역 주민을 위한 스페셜 할인 이벤트를 진행한다. 그렇다면 이 신문읍에 위치한 '신문읍일보'는 신문읍의 축제 계획 발표 URL과 특종교회의 홈페이지 공지사항 URL, 그리고 편집국밥의 할인 이벤트가 올라오는 네이버 블로그 공지사항 URL 등을 지켜보고 있다가, 변동 사항이 생기면 대규모 언어모델(LLM)이 정리해서 기사화한 다음, 사람이 최종 편집·수정해서 발송하는 형태의 기사 작성 파이프라인을 갖출 수 있는 것이다.

[그림 3] 메트로웨스트데일리가 AI를 활용해 작성한 지역 기사



▶ 지역 시니어 단체가 보스턴에서 열리는 'Hands Off!' 집회에 연대하기 위해 2025년 4월 5일 프레임িং햄에서 집회를 연다는 소식을 담은 기사. 위 기사는 AI가 수집한 정보를 바탕으로 작성됐다.

*출처: 메트로웨스트데일리

사실 AI가 도입되지 않았던 과거에도, 정보 출처 목록만 잘 정비해 뒀다면 웹사이트 모니터링과 스크래핑·정규화 작업은 충분히 가능했을 것이다. 지금은 거의 쓰이지 않지만, RSS(새 글이 올라오면 자동으로 알려주는 웹 표준)가 이런 목적을 위해 설계된 시스템이기도 하다. 특히 지역 기반 언론사라면 정보 소스가 수십 개 이내일 가능성이 높으므로 더 더욱 어렵지 않을 수 있다. 먼저 목표가 되는 정보 소스(URL 등)를 정의하고, 해당 소스를 주기적으로 확인한 뒤, 새로운 정보가 있다면 해당 데이터들을 긁어 오는 스크래핑 코드를 만들어 둘 수 있다. 소스를 정기적으로 확인하려면 유닉스 기반의 예약 실행 도구인 크론(cron)을 활용하면 된다. 예를 들어 “매일 아침 7시에 군청 홈페이지를 확인하라”라는 식의 규칙을 크론에 등록해 두면, 사람이 일일이 접속하지 않아도 프로그램이 정해진 시각에 자동으로 실행된다. 또한 단순 HTML 코드만으로는 확인하기 어려운 동적 웹페이지(예: 버튼을 눌러야 게시물이 뜨는 경우)는 셀레니움(Selenium)이라는 도구로 해결할 수 있다. 셀레니움은 마치 사람이 브라우저를 직접 열고 클릭하듯이 가상으로 브라우저를 띄워 필요한 정보를 가져오는 방식이다. 이렇게 크론과 셀레니움 같은 자동화 도구를 조합하면, 과거에도 충분히 ‘자동 기사화의 전 단계’라 할 만한 정보 수집 시스템을 구축할 수 있었다.

하지만 이렇게 데이터를 긁어온다 하더라도, 과거였다면 이를 정리하는 데 상당한 시간이 걸렸을 것이다. 우선 의미를 다 파악하기도 어렵고, 업데이트가 될 때마다 시시각각 기사를 다르게 작성해야 하는 것도 불편한 점 중 하나였을 것이다. 단순히 생각해 보더라도 시청이나 시의회가 이미 홈페이지에 공시한 정보를 언론사가 다시 가공해 배포하는 것은 기자의 투입 시간에 비해 큰 부가가치를 생산한다고 보기 어렵다.

그러나 과거와 다른 점은 AI가 사전에 정리를 대신해 준다는 데 있다. 수만 문장에 달하는 시의회 회의록이나, 시청 홈페이지에 올라온 고발성 게시글처럼 수많은 데이터 속에 묻혀버렸을 정보들이 이제는 AI에 의해 정리되고 발견될 가능성이 훨씬 커진 것이다.

단, 이 경우 실무적으로는 한 가지 해결해야 할 일이 있다. 스크랩해 온 데이터가 대용량일 경우, LLM은 이를 처리하기 힘들다는 문제가 있다. 예를 들어 내가 사용하는 LLM이 오픈AI(OpenAI)나 클로드(Claude) 등과 같이 대용량 데이터 처리까지 지원하는 상용 모델의 API라면 문제가 없지만, 자체적으로 서버나 워크스테이션을 구축해 데이터를 처리하려는 경우라면 반드시 데이터 전처리 과정이 필수적이다. 이는 보통 자체적으로 LLM 모델을 설치했을 때, 실행 환경 컴퓨터의 GPU 메모리 용량이 제한돼 있기 때문이다. 예를 들어 구글이 만든 젤마2-9B(Gemma2-9B) 모델을 VRAM 16GB의 RTX 5080

GPU에서 실행한다고 가정해보자. 이 경우 한 번에 처리할 수 있는 텍스트 길이는 최대 약 1만 5,000 토큰에 불과하다.²⁾

1만 5,000 토큰이면 대략 200자 원고지 300~500매 정도 분량이기 때문에 큰 무리는 없을 수 있지만, 만일 시의회 회의록이나 시청 게시판의 대규모 게시글과 같이 방대한 텍스트라면 여러 방법을 고려해야 한다. 대표적인 방법론이 RAG(Retrieval-Augmented Generation, 검색증강생성)인데, 이는 앞서 스크래핑한 데이터를 AI가 한 번에 참조할 수 있는 적절한 크기로 분할(chunking)하고, 이를 AI가 이해할 수 있는 벡터 형태(vectorizing)로 정규화한 뒤, 필요할 때마다 꺼내 AI가 답변할 수 있도록 하는 작업이라고 말할 수 있다. 그리고 로컬 컴퓨터에서 실행되는 LLM이나 API 형태로 제공되는 LLM(챗GPT, 클로드 등)을 활용해 수집된 정보를 가공한 뒤 기사로 바꾸면 된다. 자세한 기술적 내용은 뒤에서 더 다루기로 하고, 단순하게 말하면 여기까지가 미국 언론사 가넷이 구축한 지역 기반 뉴스 자동 생성 시스템을 복제하는 방법이다.

Step by Step: 지역 기반 뉴스 자동 생성 시스템 구축하기

- 1단계: 시스템 환경 구성하기
 - 각 지역에서 나오는 뉴스의 소스 URL 확보(가장 중요!)
 - LLM API 확보
 - API 구매(오픈AI, 클로드 등)
 - 자체 구축
 - 리눅스가 돌아갈 수 있는 컴퓨터
 - 고성능 그래픽카드(VRAM 용량이 클수록 데이터 처리량 증가)
 - 오픈소스 LLM 모델(젬마, 덤시크, OSS-GPT 등)
- 2단계: 소스 모니터링 시스템 구축
 - 각 지역에서 나오는 뉴스 소스 URL이 업데이트되는지 모니터링(크론 프로세스 구축)
 - 뉴스 소스 URL 긁어오기(셀레니움 활용)
- 3단계: 뉴스 소스를 AI로 기사화하기
 - 수집한 데이터를 AI에게 요약·정리 시키기
 - 분량이 방대한 경우 RAG 등 방법으로 기사화하기

2) 젬마 모델이 파라미터 하나당 부동소수점 16자리(FP16)를 사용하는 방식을 부동소수점 8자리 이하를 버리고 8자리만 메모리에 올리는 양자화(quantization) 방식으로 바꾸고, 토큰을 읽고 기억하는 공간 역시 FP16 대신 FP8을 적용해 메모리 사용을 절반으로 줄였을 때 기준

2. AI 네이티브 뉴스 평가 시스템: 워싱턴포스트 ‘헬리오그래프’ 사례

별것 아닌 작업처럼 보일 수 있지만, 원자적(atomic) 작업이 중요한 이유는 다른 컴퓨터 소프트웨어와 마찬가지로 뛰어난 확장성(scalability) 때문이다. 예를 들어 만일 이 작업을 지역 언론사가 아니라 중앙 언론사가 시도해 본다면 어떻게 될까? 수천 개 이상의 데이터 소스를 갖고 실시간으로 URL 소스에서 데이터들을 스크랩해 온 다음, 새로운 것들이 올라올 때마다 기사 형태로 바꿔 줄 수 있다면? 언론사는 매 순간 전국 이슈를 모니터링할 수 있는 강력한 도구를 가질 수 있을 것이다. 물론 실제로는 이런 작업을 한다는 것이 매우 소모적이라고 느껴질 수 있다. 새로운 데이터 하나하나가 모두 가치 있는 정보를 담고 있지 않을 가능성이 크기 때문이다. 하지만 바로 그래서 일차적으로 AI가 뉴스 가치가 있는지 판단해서 언론사 내부의 데스크에 보고하는 작업이 필요하다. 언론사 내부에서 기자들이 자신의 정보 소스 중 뉴스 가치가 있는 것들을 골라서 ‘일보’를 보내는 것과 같은 과정을 자동화하는 것이다.

실제로 2016~2017년 무렵 국내에도 많이 소개됐던 워싱턴포스트의 AI 뉴스 모델 헬리오그래프(Heliograf)가 이 같은 역할에 초점을 맞춘 제품이라 할 수 있다. 헬리오그래프는 선거, 스포츠 경기, 재정 보고 같은 구조화된 데이터를 정해진 템플릿에 맞춰 자동으로 기사 형태로 바꾸고, 데이터 변화가 있을 때마다 기사 내용을 갱신하거나 이상치를 식별해 기자에게 경고를 보내는 시스템이다. 2016년 이후, 헬리오그래프가 수백 건의 지역 경기 결과 보도를 자동화함으로써 워싱턴포스트는 지역 커버리지를 훨씬 넓힐 수 있었다. 기자들은 반복적이고 단순한 보도보다는 인터뷰, 분석, 현장 취재 같은 고부가가치 작업에 집중할 여유를 얻었다. 로이터의 ‘뉴스 트레이서(News Tracer)’도 트위터에 올라오는 뉴스 소스를 교차 검증해서 비슷한 스크리닝 작업을 실시하는 제품이다.

이런 시스템이 가능한 이유는 앞서 말한 것처럼 자동화 가능한 정보 수집 루틴(스크래핑 및 모니터링), 그리고 뉴스 가치 판단을 거쳐 기사화를 누군가(또는 무언가)가 먼저 수행하는 구조 덕분이다. 중앙 언론사 차원에서 수천 개의 데이터 소스를 관리하고, 헬리오그래프처럼 알림·업데이트 기능을 더해 둔다면, 뉴스룸은 실시간 뉴스 흐름에 더 민첩하게 대응할 수 있을 것이다.

다만, 여기에는 까다로운 작업이 하나 필요하다. 바로 뉴스 가치에 대한 일차적 스크리닝 즉, 평가모델을 설계하는 일이다. 이는 언론사마다, 또는 설계자마다 서로 다른 가치와 기준이 개입될 수 있기 때문에 각자 모델이 달라질 수가 있다. 또한 이런 판단은 결국

[그림 4] 하버드대 니먼저널리즘랩 홈페이지에 실린 헬리오그래프 일러스트레이션



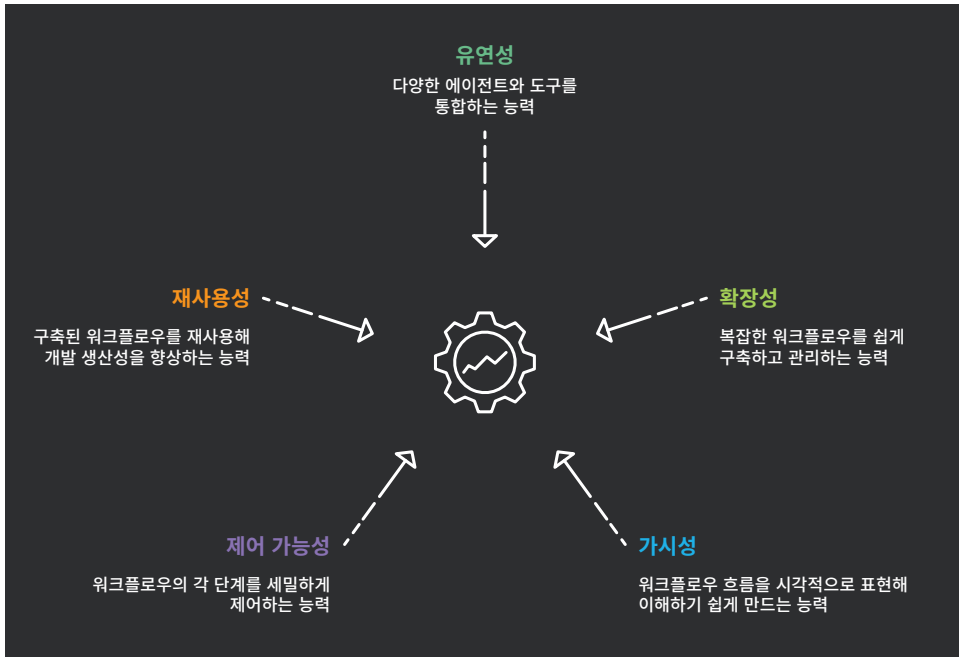
* 출처: 하버드대 니먼저널리즘랩

사람의 영역 아니냐는 질문도 나올 수 있다. 하지만 생각해 보면 일차적 스크리닝을 AI에 맡기는 것은 반드시 필요한 일일 수 있다. 첫째, 데이터의 소스가 많으면 많을수록 사람이 감당하기 어려운 수준의 기사를 AI가 생성할 수 있다. 둘째, 지금도 뉴스룸에서는 정보의 홍수 속 가치 있는 정보를 스크리닝하는 일상의 작업 때문에 정작 중요한 뉴스를 찾는 작업에 시간을 많이 쓰지 못하는 문제가 발생하고 있다.

더 이상 사람의 직관과 경험만으로 쏟아지는 정보를 감당하기 어렵다. 따라서 AI를 활용한 ‘뉴스 가치 판별 모델’이 필요하다. 이 모델은 단순히 특정 키워드의 빈도를 측정하거나, 소셜미디어에서 언급량을 감지하는 수준에 머물러서는 안 된다. 사건이 어떤 맥락 속에서 일어나고, 그 결과가 누구에게 어떤 영향을 주는지를 단계적으로 사고할 수 있어야 한다.

이를 구현하는 방법으로 LLM을 통해 인공지능의 업무 단계를 유연하게 설계할 수 있는 프레임워크인 랭그래프(LangGraph) 기반 노드 구조를 설계할 수 있다. 랭그래프는 AI 언어모델이 여러 단계를 거쳐 복잡한 업무를 수행할 수 있도록 하나의 워크플로우를 설계하고 운영할 수 있게 도와주는 파이썬 프레임워크다. 즉, 단순한 프롬프트 호출을 넘어서서 여러 단계의 의사결정과 도구 활용을 그래프 구조로 설계할 수 있도록 돕는다. 비슷한 프레임워크로 랭체인(LangChain)이 있는데, 랭체인이 직선적인 선형의 업무 구조를

[그림 5] 랭그래프의 핵심 특징



짜는데 특화돼 있다면 랭그래프는 피드백 루프를 만들 수도 있고 조건에 따른 업무 분기를 만들 수도 있는 등 훨씬 유연한 설계가 가능하다는 장점이 있다.

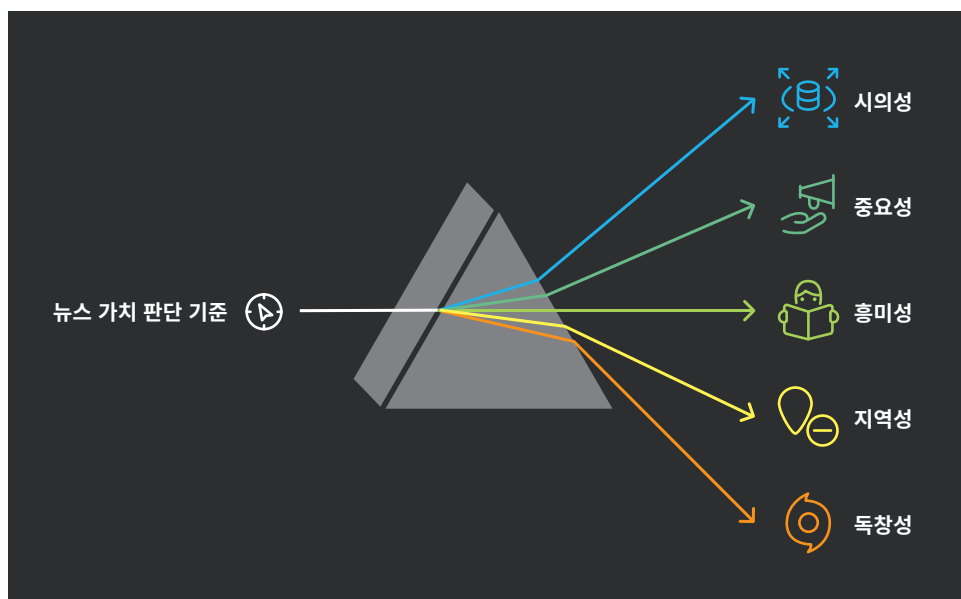
자동으로 생성된 기사 가치를 자동으로 판단하는 머신을 랭그래프로 설계하기 위해서는 AI 에이전트가 일을 하는 단위인 ‘노드’를 만들어야 한다. 그리고 그 노드를 연결하는 논리를 구성하면 그게 작업의 끝이다. 첫 번째 단계로 ‘사건 구조화 노드’를 만들 수 있다. 입력된 텍스트에서 주체(누가), 행위(무엇을), 영향(누구에게 어떤 결과를 미치는가)이라는 세 가지 요소를 뽑아내는 역할을 한다. 두 번째 단계는 ‘가치 평가 노드’다. 여기서는 전통적으로 언론학에서 정의해 온 뉴스 가치 기준들을 그대로 반영한다. 예컨대 시의성(얼마나 새로운가), 영향력(얼마나 많은 사람에게 영향을 주는가), 갈등성(대립이나 긴장이 포함되어 있는가), 지역성(독자와의 거리), 인간적 흥미(사람들이 공감할 만한 요소가 있는가) 같은 기준이다. 각 요소는 0~5점 등으로 점수화할 수 있고, 단순 합산보다는 가중치를 달리해 언론사별 성격을 반영할 수도 있다. 세 번째 단계는 ‘설명 생성 노드’다. 여기서 ‘생각의 고리(Chain of Thought, CoT)’ 방식을 써서, 모델이 한번에 결론을 내리는 대신, “왜 이 사건은 영향력이 높다고 볼 수 있는가?”, “왜 인간적 흥미는 낮다고 판단되는가?”와 같이 기준

별 근거를 단계별로 서술하게 만드는 것도 흥미로운 방법이 될 것 같다. 이렇게 하면 결괏값이 블랙박스로만 남는 것이 아니라, 기자가 그 과정을 검증할 수 있는 투명한 형태가 된다. 예를 들어 신문읍의 사례에서 ‘편집국밥 할인 이벤트’는 생활정보로 분류돼 시의성은 높지만 영향력은 낮다는 설명이 붙을 수 있고, ‘군청 축제 계획 발표’는 지역 주민 다수가 참여한다는 점에서 영향력과 지역성이 모두 높다고 평가될 수 있다.

마지막 단계는 ‘우선순위 종합 노드’다. 앞선 점수와 설명을 모아 최종적으로 보도 우선순위를 제시하는데, 단순히 숫자만 내는 것이 아니라 “이 사건은 지역사회 전체에 파급력이 크므로 헤드라인 가치가 있다”처럼 권고 형태로 나타난다. 기자는 이를 참고해 실제 기사 배치를 결정한다. 결국 이 시스템은 AI가 기자의 판단을 대체하는 것이 아니라, 방대한 정보 속에서 ‘무엇을 먼저 주목해야 하는가’를 도와주는 정교한 보조 장치가 된다. 과거에는 데스크의 경험적 판단에만 의존해야 했던 영역을, 이제는 AI의 단계적 추론과 여러 AI 노드들의 협업 프레임워크인 랭그래프 기반의 구조적 사고가 보완하는 것이다.

이렇게 어떤 뉴스거리에 대한 평가모델을 갖춰 놓게 된다면, 언론사 고객인 독자에 대한 ‘서비스’가 가능하다는 사실을 발견할 수 있다. 즉, 흥미성이 강한 뉴스에 관심을 보이는 ‘독자 A’에는 ‘흥미성’ 점수에서 높은 평가를 받았던 소스를 정제한 뉴스를 중점적으로 전달할 수 있고, 현재 내 홈페이지에 접속해 있는 ‘독자 B’에게는 시의성이 높게 평가

[그림 6] 뉴스 가치 판단기준 탐색 모델의 예시



받았던 소스를 기반으로 한 콘텐츠를 바로 보여줘서 홈페이지 체류시간을 높이는 등 전략을 취할 수 있는 것이다. 따라서 뉴스 소스에 대한 평가모델을 랭그래프 같은 도구로만 들어 두는 것은 언론사 구독자에게 ‘정보를 서비스하는 자동화 프로세스를 만든다’라는 관점에서 반드시 필요한 하나의 주춧돌 같은 것이라 할 수 있다.

Step by Step: 대규모 뉴스 수집 · 평가 시스템 구축하기

- 1단계: 데이터 수집 자동화
 - 수천 개의 뉴스 소스(URL, 트위터 피드, 보고서 등) 정의
 - 크론으로 주기적 실행, 셀레니움으로 동적 페이지 크롤링
 - 글 업데이트 · 변경 데이터 자동 확보
- 2단계: 뉴스 가치 평가모델 적용
 - 랭그래프 노드 설계
 - 사건 구조화(주체 · 행위 · 영향)
 - 가치 평가(의의성, 영향력, 갈등성, 지역성, 인간적 흥미 등)
 - 생각의 고리(CoT) 방식으로 “왜 중요한가” 설명 생성
 - 점수 · 설명 기반으로 우선순위 결정
 - 해당 점수는 구독자 맞춤형 뉴스 공급을 위해 활용
- 3단계: 기사화 및 배포
 - 템플릿+LLM 요약으로 초안 기사 생성(헬리오그래프와 동일 방식)
 - 데스크가 검토 후 발행
 - 독자별 맞춤 제공(예: 시의성 중시 독자 vs 흥미성 중시 독자)

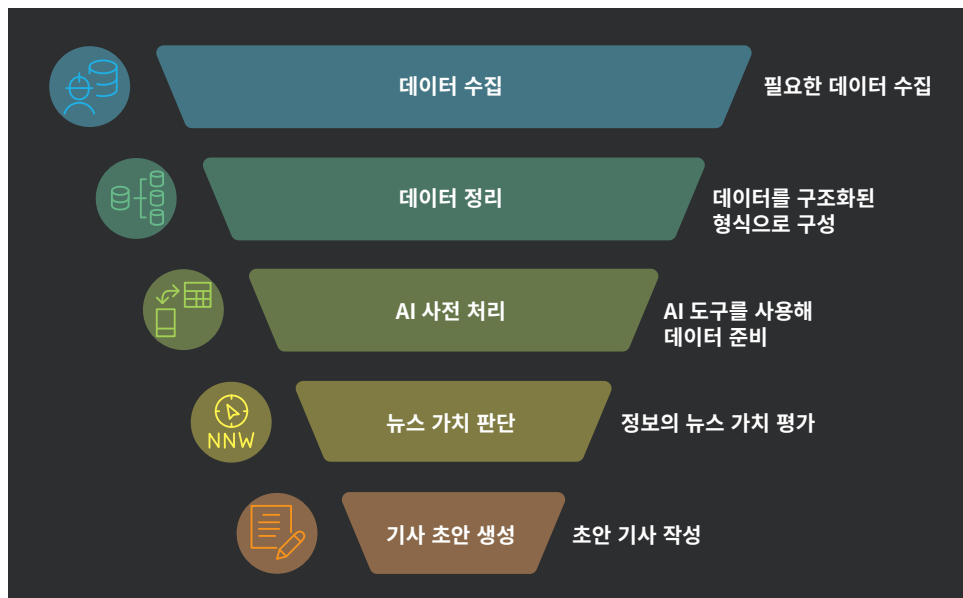
3. 왜 이런 시스템을 만들어야 하는가: 라즈 체티의 연구 사례

이런 아주 기초적인 자동 취재 및 뉴스 평가 시스템이 갖는 함의는 뭘까. 당연히 이런 시스템을 갖춘다고 해서 당장 돈이 되는 것은 아니다. 그러나 가넷 같은 곳들의 작업이 의미를 갖는 이유는, 이 방식이 확장(scale)될 가능성이 매우 크기 때문이다. 간단한 지역 기반 정보 취합 정리 서비스가 전국 단위로 확장되거나, 이 정보와 연결된 서비스 제공이 늘어날 때 굉장히 파괴력이 있는 서비스의 탄생이 가능할 수 있다. 예를 들어 현재도 언론사에는 보도자료를 공급받는 시스템(이메일, 메신저 등)이 존재한다. 정부 기관, 기업, 단체 등에서 언론사에 취재 요청 또는 기초자료 제공 목적으로 공급되는 보도자료들은 그 자체가 팩트의 집합이라 할 수 있다. 새로운 보도자료들이 RSS뿐만 아니라 이메일, 문자 메시지 등 다양한 비정형 데이터 형태로 제공되고, 이를 LLM이 읽어 들어서 뉴스 가치 판

단 모델에 적용하고, 다시 뉴스 가치 판단 모델을 통해 걸러낸 다음, 기사 초안을 생성하는 모델은 쉽게 생각해 볼 수 있는 뉴스 생성 자동화 시스템의 모습일 것이다. 보도자료 외에도 날씨 정보, 스포츠 경기 결과, 통계청 통계 등과 같이 API 형태로 제공되는 정보들은 이미 정형화돼 있기 때문에 뉴스 생성 자동화 시스템에 접목하기 딱 좋은 것들이다.

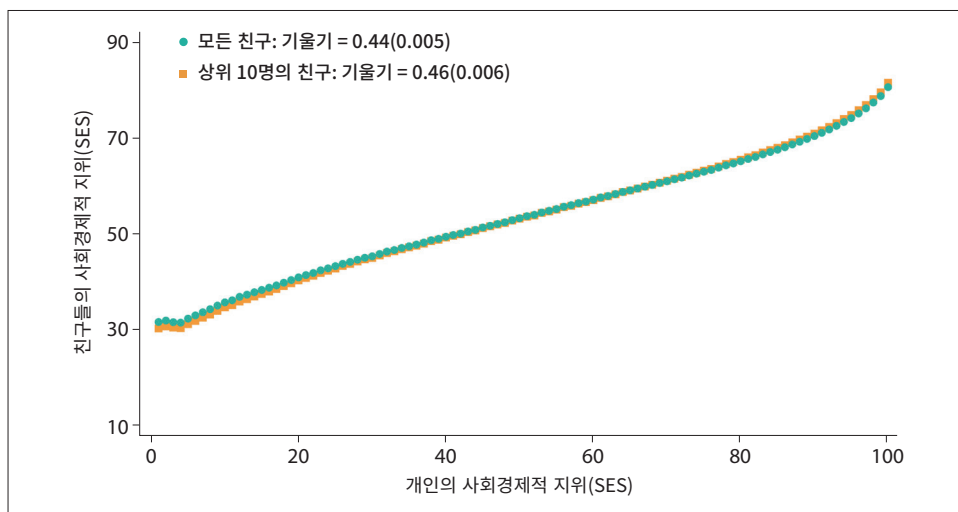
사회 전체 데이터를 전수조사 형태로 수집해 팩트 파인딩(fact finding)하는 시도는 이미 학계에서는 다수 진행되고 있다. 예를 들어 미국 하버드대의 경제학과 교수인 라즈 체티(Raj Chetty)는 페이스북이 갖고 있는 친구 관계 데이터에서 약 210억 건의 친구 연결 정보를 우편번호, 고등학교, 대학 단위로 분석해 ‘어릴 때 맺었던 친구 관계가 경제적 신분 상승에 큰 영향을 미친다’는 사실을 밝혀냈다.³⁾ 빈곤층에도 신분 상승의 기회를 제공하는 사회를 만들기 위해선 어린 시절부터 계층적으로 혼합된 환경을 설계해 줄 필요가 있다는 정책적 함의가 연구를 통해 도출된 것이다. 그것도 210억 건의 전수조사에 가까운 데이터 분석을 통해서.

[그림 7] AI 네이티브 기사 생산 프로세스



3) Chetty, R., Jackson, M. O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R. B., Gong, S., Gonzalez, F., Grondin, A., Jacob, M., Johnston, D., Koenen, M., Laguna-Muggenburg, E., Mudekereza, F., Rutter, T., Thor, N., Townsend, W., Zhang, R., Bailey, M., Barberá, P., Bhole, M., & Wernerfelt, N(2022), <Social capital I: Measurement and associations with economic mobility>, Nature, 608, pp.108-121, <https://doi.org/10.1038/s41586-022-04996-4>

[그림 8] 개인의 사회경제적 지위(SES)와 친구들의 사회경제적 지위(SES) 간 관계



▶ 논문 연구 결과의 핵심을 보여주는 그래프. 가로축이 자신의 경제적 지위를 말해주고, 세로축이 친구들의 평균적인 경제적 지위를 보여준다. 즉, 나의 경제적 지위와 어린 시절 친구들의 경제적 지위는 분명한 상관관계가 있는 것으로 나타난다.

* 출처: 라즈 체티 외(2022)

전 세계적으로 2010년대 후반 경제학계의 주목을 받았던 파리경제대 교수 토마 피케티(Thomas Piketty)는 국세청 납세 기록을 통해 수백 년에 걸친 부의 집중과 불평등의 현실을 전수조사로 입증했다. 학계에서 입증된 이런 연구들은 학자의 양심과 분석적 역량을 믿고 정부가 가지고 있는 데이터들을 있는 그대로 통째로 제공했기 때문에 확인될 수 있었던 결과다. 이처럼 국가 내부 시스템이 생산하는 구조적 데이터를 활용하는 접근은, 언론의 자동화된 팩트 파인딩과 심층 취재 역량을 실질적으로 강화할 수 있는 길을 제시한다. 비슷한 예로 오늘날에도 언론사들은 선거철 여론조사를 실시하고 판매하는 비즈니스를 하고 있기도 하다. 더 중요한 것은 공공데이터를 갖고 있는 공공기관(수사기관, 국세청, 통계청 등)이 언론사에 데이터를 공개하는 것이 유리한 구조, 즉 신뢰 관계다. 공공기관 입장에서는 통계를 알리는 것이 목적인데 신뢰하기 어려운 유튜브나 인플루언서에게 이 통계를 제공했을 때 자칫 악용될 위험을 걱정할 수 있다. 이러한 악용을 언론이 막아줄 수 있다는 사회적 신뢰가 있기 때문에, 언론은 꾸준히 데이터를 제공받을 수 있는 특별한 관계를 유지할 수 있는 것이다. 따라서 언론사들은 AI를 통해 더 많은 신뢰를 확보할 수 있다는 주장에 신빙성을 더해야 할 필요가 있다. 그런 주장이 받아들여져서 공공기관으로부터 더 많은 데이터를 받아 더욱 신뢰받는 보도를 할 수 있는 선순환을 만드는 것이 바람직한 방향일 수 있다.

4. 팩트체크 API 만들기

지금까지는 언론 본연의 업무를 잘하기 위한 AI 도입 방법에 대한 이야기였다. 그러나, 이제 언론을 비즈니스라고 생각하고 AI에 접근한다면 이야기는 조금 많이 달라진다. AI가 사람을 대체한다는 이야기가 아니라, “어떻게 언론이 AI를 이용해서 새로운 사업의 기회를 찾을 수 있는가”라는 질문이다. 이는 언론이 과거에는 없었던 새로운 가치를 다수 대중에게 어떻게 줄 것인지의 문제가 된다. 앞서 말했던 AI 제품들이 언론의 기존 업무를 효율화하는 관점이라면, 이제 ‘과거에 없던 새로운 혁신’을 AI로 어떻게 만들어 낼 것인지에 대한 이야기다. 특히 사업을 하는 사람으로서 늘 놓치지 말아야 하는, 그러나 기술이 발전하면 그 기술을 따라잡고 적용해 보느라 종종 놓치는, 정말 중요한 핵심은 바로 ‘고객’ 일 것이다. 즉, 언론 산업의 본질이라 할 수 있는 고객을 놓고 봤을 때 우리가 AI 네이티브 기업으로 변화하면서 고객에게 기존과 다른 가치를 줄 수 있는지 생각해 볼 필요가 있다.

예를 들어 고객들이 원하는 것은 이런 가치들일 것이다.

- 나 부자 되게 해 주는 정보가 어디 없나?
- 나만 바보 되지 않게 해 줘. 내 친구들이 아는 뉴스 나도 알게 해 줘.
- 세상 돌아가는 정치 이야기, 국제정세 이야기, 나도 잘 알게 해 줘.
- 어디 재밌는 일 없나?
- 누가 권력을 남용해서 나쁜 짓 하는 거 고발하고 잡아줬음 좋겠어!

AI를 이용해서 이 작업을 기존 서비스보다 두세 배 이상 잘할 수 있다면 사업적으로 매우 성공할 가능성이 높다고 할 수 있을 것 같다. 하지만 쉽지는 않은 일이다. 이미 영역마다 너무나 뛰어난 사업자가 존재하고 있기 때문이다.

- 나 부자 되게 해 주는 정보가 어디 없나? → 유튜브 강연
- 나만 바보 되지 않게 해 줘. 내 친구들이 아는 뉴스 나도 알게 해 줘. → 인스타그램 스레드 등
- 세상 돌아가는 정치 이야기, 국제정세 이야기, 나도 잘 알게 해 줘. → 인스타그램, 유튜브 등

- 어디 재밌는 일 없나? → 유튜브, 인스타그램
- 누가 권력을 남용해서 나쁜 짓 하는 거 고발하고 잡아줬음 좋겠어! → 유튜브의 각종 정치 채널들

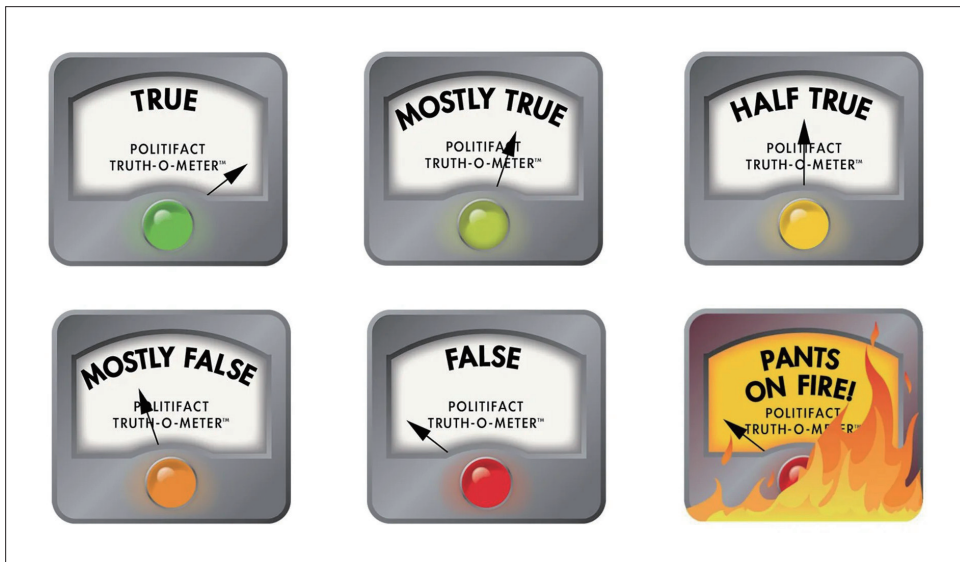
예를 들어 AI를 이용해 유튜브 강연보다 두세 배 더 확실하게 부자 되게 만들어 주는 정보를 전달해 줄 수 있다면, 또는 AI를 이용해 특정 고객에게 고객 주변의 또래보다 훨씬 더 정확하게 세상 돌아가는 이야기를 전달해 줄 수 있다면, 또는 AI를 이용해서 권력을 남용하는 공직자나 사회고위층의 비밀스러운 이야기를 더 정확하고 빠르게 폭로해 낼 수 있다면, 사실 이미 대중에게 공개돼 버려서 범용화된 AI 기술을 가지고 이런 뾰족한 가치를 제공하는 언론 모델을 만든다는 것은 쉬운 일이 아니다. 그러나 비즈니스로서 잘 될 것이라는 전제로 말하는 것이 아니라, 하나의 가능성 정도로 이야기해 본다면 다음과 같은 예시는 들어볼 수 있을 것 같다.

언론사가 AI를 통해 남들보다 더 잘할 수 있는 영역의 예시로서 ‘팩트체크’ 기능이 있을 수 있다. 인터넷 세상에 범람하는 수많은 정보와 음모론 중에서 ‘딱지신문’이 체크한 팩트 또는 픽션 딱지를 붙일 수 있는 기능을 블로그, 소셜미디어 등 여기저기에 유료로 제공하는 것이다. 생각해 보면 과거에도 제품 홍보나 기업 이미지 메이킹에 신문 기사를 활용한 사례가 많았다. 비슷한 효과를 줄 수 있도록 팩트가 검증된 경우 ‘딱지신문이 검증한 팩트’임을 붙여둘 수 있게 API를 붙여 판매하는 것이다. 또는 어떤 기업이 자사 제품의 우월성을 증명하기 위해 비용을 내고 ‘딱지신문의 팩트 챗봇’을 자사 홈페이지에 띄워 두면 접속자들이 “이 기업의 주장이 팩트 맞아요?”라고 ‘딱지신문’의 챗봇에 질문하면 사실관계를 답해주는 사업모델도 생각해 볼 수 있을 것이다. 게다가 AI가 콘텐츠를 창작하게 된 이 시대에는 할루시네이션(환각)이나 과거 데이터를 참조해 생성된 쓰레기 콘텐츠(garbage)를 방지하기 위해서라도 자동화된 팩트 응답 시스템에 대한 수요가 늘어날 수 있다. 요컨대, 사용자가 정말로 알고 싶어하는 특정한 팩트를 잘 체크해 줄 수 있다는 점에서 언론만큼 범용 서비스(예: 구글, 퍼플렉시티 등)에 비해 두세 배 이상 잘 만족시켜 줄 수 있는 기관도 없다고 필자는 생각한다. 문제는 이를 확장할 방법이 마땅치 않았다는 점, 그리고 구글이나 퍼플렉시티는 이를 확장할 방법을 잘 찾았다는 점인데, 그 부분에 있어서 AI가 어느 정도 해법을 제공해 줄 수 있지 않냐는 것이 필자의 가설이기도 하다(이견이 있을 수 있지만 일단 그렇게 가정하고 논의를 진행해 보자).

해외에선 이런 가설과 맞닿아 있는 유의미한 시도들이 다수 진행됐다. 미국의 폴리터

팩트(PolitiFact)는 정치인 발언이나 공공 정책에 대한 사실 검증을 꾸준히 축적해 온 대표적인 사례다. 이들은 ‘트루스-오-미터(Truth-O-Meter)’라는 직관적인 평정 체계를 통해 데이터를 공개하고 있으며, 일부 연구자나 개발자가 접근할 수 있는 비공식 API와 데이터셋을 제공해 왔다. 다만 공식적이고 상업화된 API 형태는 제한적이라는 한계가 있다. 미국의 스노프스(Snopes) 역시 방대한 팩트체크 데이터베이스를 운영하며 웹을 중심으로 신뢰 자산을 축적했지만, 데이터 활용의 외부 개방성은 크지 않았다. 한편 글로벌 차원에서 가장 체계적으로 API를 제공하는 곳은 구글 팩트체크 툴스(Google Fact Check Tools)다. 구글은 전 세계 언론사와 협력해 클레임리뷰(ClaimReview)라는 구조화 마크업을 기반으로 한 메타데이터를 수집하고, 이를 클레임 서치 API(Claim Search API)를 통해 검색할 수 있도록 지원한다. 이를 통해 특정 주장에 대한 다양한 출처의 검증 결과를 API 형태로 호출할 수 있으며, 언론사들은 구글 플랫폼에 자신의 팩트체크 결과를 구조화해 노출할 수 있다. 영국의 풀팩트(Full Fact)는 비영리 단체임에도 ‘풀팩트 AI(Full Fact AI)’라는 이름으로 자동화된 팩트체크 툴을 개발해 왔으며, 온라인상에서 반복적으로 등장하는 허위 주장을 자동 탐지하거나, 확산 경로를 추적하는 기능을 시험했다. 이들은 자동화와 인간 검증을 결

[그림 9] 폴리팩트의 트루스-오-미터(Truth-O-Meter)



▶ 폴리팩트는 트루스 오-미터라는 평정 체계를 운영한다. 팩트체크 결과를 진실(truth), 거의 진실(mostly truth), 반쯤 진실(half truth), 거의 거짓(mostly false), 거짓(false), 새발간 거짓말(pants on fire)로 구분해 제공하고 있다.

* 출처: 폴리팩트

[그림 10] 팩트체크 API의 잠재 고객들



합한 워크플로우를 강조하며 언론사와 협력해 왔지만, API 형태의 완전 개방보다는 특정 파트너와의 제한적 공유가 중심이었다. 이처럼 국가와 조직마다 차이는 있지만, 이미 여러 기관이 팩트체크를 구조화된 데이터, 나아가 API 서비스로 전환하기 위한 노력을 지속해 왔다는 점에서 시사하는 바가 크다.

그렇다면 팩트체크 API는 어떻게 만들 수 있을까? 아주 간단하게 말하면 이제까지 논의했던 작업을 하나씩 끼워 붙이는 것이 주된 작업이 될 것 같다. 즉, 가넷 모델에서 유추해 온 ‘자동 기사 생성 시스템’과 헬리오그래프 모델에서 유추해 온 ‘자동 기사 판단 시스템’을 대규모로 확장해 데이터셋(팩트의 집합)을 구축한 다음, 그를 기반으로 언론사가 자체적으로 보유한 데이터를 접목하는 것이다. 여기에 사용자들이 특별한 팩트의 검증을 요구하면, 다소 시간이 걸리더라도 사람이 개입해 응답하는 특별한 정보 확인 루트를 더하는 설계가 필요하다. 데이터셋이 갖춰졌다고 가정하면, 그 데이터를 사용자들의 요청과 어떻게 연결할지 설계하는 정도가 추가될 것이다.

이제 팩트체크 API를 만들기 위한 랭그래프 노드 설계를 다음과 같이 해 볼 수 있다.

팩트체크 API는 크게 입력 수집(Input), 자동 검증(Auto-check), 인간 검증(Human Review), 그리고 결과 배포(Output)라는 네 가지 축으로 나눌 수 있다. 단계마다 노드를 정의하고 상태 전환을 설정하면 언론사가 운영 가능한 하나의 완전한 팩트 검증 파이프라인이 완성된다.

첫째, 데이터 수집과 구조화 단계를 담당하는 노드가 필요하다. 이를 인풋 콜렉터 노드(InputCollectorNode)라고 이름 지어보자. 사용자가 팩트체크를 요청하는 방식은 다양하다. 기자가 직접 특정 정치인의 발언을 입력할 수도 있고, 일반 독자가 기업 보도자료의 문구를 붙여 넣을 수도 있으며, API를 호출하는 제휴사의 서비스에서 소셜미디어 게시물을 자동 전달할 수도 있다. 이 노드는 모든 입력을 동일한 구조의 데이터 객체로 표준화 시켜서 변환한다. 요컨대, 입력되는 값은 어지러운 사용자들의 다양한 요구 형식(정치인의 발언, 독자의 주장, 보도자료 문구, 심지어 유튜브 클립의 특정 부분)일텐데, 이 노드에서는 그걸 표준화된 형식으로 바꾸는 것이다. 특히 발언 내용, 발언자, 발언 시점, 출처 링크 같은 필수 정보가 담겨야 한다. 이 단계에서 중요한 것은 최대한 맥락 정보를 함께 수집하는 것이다. 그래야 이후 검증 단계에서 AI가 보다 정밀하게 교차 검색을 수행할 수 있다. InputCollectorNode는 결국 팩트체크의 출발점이자, 모든 흐름이 시작되는 기초자료를 제공하는 역할을 한다.

둘째, 자동 검증 단계다. 이 과정은 두 개의 핵심 노드로 나눌 수 있다. 하나는 오토 체크 노드(AutoCheckNode), 다른 하나는 플래깅 노드(FlaggingNode)다. AutoCheckNode는 첫 번째 단계인 InputCollectorNode를 통해 나온 결과물을 바탕으로 LLM을 호출해 1차 분류를 수행한다. 포인트는 단순히 모델의 직관적 답변을 받아들이는 것이 아니라, RAG 기법을 적용해야 한다는 데 있다. 예를 들어 “대한민국 남성의 80%가 탈모 유전자를 가지고 있다는 발언이 객관적 근거가 있는가?”라는 질문을 던지면, LLM이 자체적으로 답하는 것이 아니라 벡터 데이터베이스와 키워드 검색을 동시에 활용하여 통계청의 최신 데이터, 최신 의학 저널, 학술 데이터베이스 등 권위 있는 문서를 불러온 뒤, 그 근거를 바탕으로 판단을 내린다. AutoCheckNode은 사실(True)·거짓(False)·불확실(Uncertain) 같은 라벨과 함께 근거가 되는 데이터(근거 문장, 출처 링크, 인용된 통계 수치 등)를 포함한다. FlaggingNode는 이 AutoCheckResult를 검토해 불확실하거나 상충된 결과가 발견되면 ‘검증 필요’ 플래그를 붙인다. 예를 들어 LLM이 불일치 가능성을 감지했거나, 검색 결과가 오래된 문서에 의존하고 있으면 FlaggingNode는 이를 인간 검증 단계로 넘긴다. 반대로 명백히 사실로 판명된 경우에는 즉시 사실(True)·거짓(False) 레이블을 달아 API 응답으로

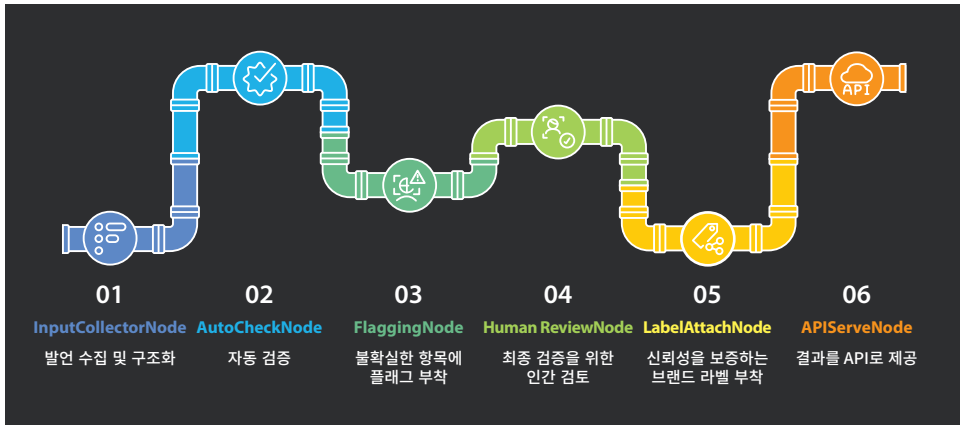
바로 보낼 수도 있다. 이처럼 자동 검증 단계는 대규모 요청을 빠르게 처리하는 효율성을 확보하면서도, 신뢰도가 떨어질 수 있는 항목은 반드시 다음 단계로 넘기는 안전장치를 마련하는 구조다.

셋째, 인간 검증과 품질 관리 단계다. 이 단계는 언론사의 강점이 드러나는 영역이다. 휴먼 리뷰 노드(HumanReviewNode)는 기자나 리서처가 AutoCheckResult와 FlaggingNode의 태스크를 입력으로 받아 최종 판정을 내리는 과정을 의미한다. 이때 기자는 LLM이 제시한 근거와 원문 데이터를 확인하고, 필요하다면 추가 자료를 직접 검색하거나 전문가 인터뷰를 통해 보완한다. HumanReviewNode는 결국 최종적인 사실(True)·거짓(False)·일부 사실(Partly True) 같은 레이블을 산출한다. 이를 파이널 버딕트(FinalVerdict)라는 데이터 객체로 정의할 수 있다. 이후 라벨 어태치 노드(LabelAttachNode)가 FinalVerdict를 받아서 브랜드 라벨을 자동 부착한다. 예컨대 ‘딱지신문이 확인한 팩트’라는 문구를 결과와 함께 API로 배포한다. 이 라벨은 단순한 결과 표시를 넘어, 언론사가 제공하는 신뢰의 보증서 역할을 한다. 사용자는 결과만 확인하는 것이 아니라, 해당 언론사의 권위와 책임까지 함께 느낄 수 있다. 이처럼 HumanReviewNode와 LabelAttachNode는 AI 혼자서 줄 수 없는 최종적 신뢰를 보장하는 장치다.

넷째, 결과 배포 단계다. 이를 담당하는 것이 API 서브 노드(APIServeNode)다. APIServeNode는 FactCheckLabel을 외부 서비스에서 호출할 수 있도록 REST API나 GraphQL API 형태로 제공한다. 언론사가 운영하는 자체 웹사이트뿐 아니라, 제휴 기업의 블로그, 전자상거래 플랫폼, 소셜미디어 위젯 등 다양한 곳에서 이 API를 불러올 수 있다. 예를 들어 소비자가 한 기업의 홈페이지에서 “이 제품의 성능이 경쟁사보다 우수하다는 주장이 사실입니까?”라고 묻는다면, 해당 홈페이지에 임베드된 챗봇이 APIServeNode를 통해 즉시 팩트체크 결과를 불러와 응답하는 식이다.

이러한 노드 설계는 단순히 기술적인 도식에 머무르지 않는다. 언론사가 직접 보유한 데이터와 브랜드를 자산화하고, 이를 자동화된 검증 서비스로 전환해 새로운 비즈니스 모델로 발전시킬 수 있다는 점이 중요하다. 자동 검증만으로는 신뢰성이 부족하고, 인간 검증만으로는 규모 확장이 어렵다. 하지만 두 가지를 랭그래프 기반의 노드 체계 안에서 결합하면, 효율성과 신뢰성을 동시에 확보할 수 있다. 또한 노드별로 상태를 정의하고 전환 경로를 분명히 했기 때문에, 오류가 발생했을 때 어느 단계에서 문제가 생겼는지 추적하기도 쉽다. 예컨대 AutoCheckNode가 특정 데이터베이스에 접근하지 못했을 경우, 상태가 실패로 기록되고, 곧바로 HumanReviewNode로 우회하도록 설계할 수 있다.

[그림 11] 팩트체크를 위한 랭그래프 노드 설계



정리하자면, 팩트체크 API를 위한 랭그래프 노드 설계는 [그림 11]과 같은 흐름을 가진다.

이런 식의 랭그래프 기반 설계는 단순한 AI 보조 기능을 넘어, 언론사가 자체적인 팩트 데이터 자산을 API로 전환해 수익화할 수 있게 한다. AI 콘텐츠가 범람하는 시대에 ‘자동화된 팩트 응답 시스템’은 필연적으로 수요가 증가할 수밖에 없다. 잘못된 데이터를 바탕으로 생성된 할루시네이션이나 가짜 뉴스성 콘텐츠를 걸러내기 위해서라도, 언론이 제공하는 검증된 라벨은 강력한 차별화 요소가 된다. 결국 팩트체크 API는 언론사에 있어 단순히 기술적 실험이 아니라, 독자와 사회 전체에 신뢰를 제공하는 새로운 사업모델이자 생존 전략이 될 수 있다.

5. 간단한 버전...

물론 제대로 한다면 이런 정밀한 과정을 거쳐야 하겠지만, 데이터셋 구축과 소프트웨어 개발에 큰 비용과 시간이 들어가는 프로젝트가 될 수 있을 것이다. 반면 아주 간단하게 서비스하고자 한다면 일차적인 팩트 검증 메커니즘은 퍼플렉시티처럼 최신 팩트 데이터를 이미 많이 갖고 있는 기업의 API를 사용하고, 그다음으로 고객이 원하는 특화된 팩트 데이터는 언론사 자체적인 데이터를 얹어서 제공하는 것이 빠르게 서비스를 할 수 있는 방법이 될 것이다. 스타트업 마인드로 이 팩트체크 API 설계에 접근한다면, 일단 퍼플렉시티 API를 얹은 챗봇을 운영하되 매우 정밀한 응답을 해야 하는 질문을 받을 경우 언

론사 내 기자에게 팩트체크를 요청하고 응답 받게 하는 서비스를 만들 수 있다. 이 서비스를 2주 내에 론칭하고 시장의 반응을 보면서 제품을 계속 업그레이드해 나가는 형태로 사업화할 수 있으리라 생각한다.

생각해 보면 바로 사용자들에게 팩트를 빠르게 정리해서 전달한다는 점에서 퍼플렉시티가 하고 있는 서비스가 우리가 설계하는 팩트체크 API와 유사한 성질을 갖고 있다고 할 수 있다. 다만 범용화된 서비스인 퍼플렉시티가 하지 못하는 수요자 특화 팩트체크 기능(위 랭그래프 설계에서 HumanReviewNode가 진행하는 업무)이 이 API에 붙어 있다는 것이 특징점이라 할 수 있다. 이렇게 설계한다면 수요자가 실제로 사용하는 만큼 과금을 받으면서, 해당 수입에서 퍼플렉시티 API 호출 비용을 차감한 만큼을 언론사가 가져가는 비즈니스 설계를 해 볼 수 있다. 팩트를 확인하고자 하는 사람들의 열망이 창출할 시장은 구글이 보여줬던 것처럼 어마어마하다. 그래서 퍼플렉시티 같은 기업들이 초기 설비비용과 운영비용 등과 같은 대규모 자금의 투자를 늘리며 언론사들의 팩트 데이터들까지 제휴와 수집을 통해 흡수하려고 하는 것이다. 따라서 시장은 분명히 존재한다고 할 수 있고, 각 사업 주체가 얼마만큼 더 고객을 만족시켜 줄 수 있느냐의 싸움이 되지 않을까 하는 생각이다.

언론사 조직 차원의 AI 기술 도입 전략: 기술적 선택지

거창한 제품 개발 외에도 언론사 내부 뉴스룸에서 벌어지는 수많은 업무를 AI로 바꿀 가능성은 매우 크다. 모델 컨텍스트 프로토콜(Model Context Protocol, MCP)를 활용해서 구글 독스(Google Docs)와 같은 무료 문서를 활용한 우리만의 CMS 만들기 프로젝트를 구성해 본다거나, MCP를 활용해 취재원과의 약속을 자동으로 스케줄링하는 AI 에이전트를 만들어 본다거나, 사진 또는 영상으로 촬영해 온 취재 결과물을 만화 또는 애니메이션 형태로 자동 전환해 자동으로 배포하는 시스템을 만들어 보는 등등... 여기 나열한 아이디어들은 모두 하나의 예시일 뿐이지만, 이들을 실제로 사용해 가치를 창출해 내려면 기술이 조직 내에 충분히 잘 스며들 필요가 있다. 이 때문에 언론사로서는 어느 정도까지 기술을 내재화할 것인지에 대한 고민이 커질 수밖에 없으리라 생각한다. 여기서는 언론사 조직 차원에서 AI 기술을 선택할 때 어떤 선택지들이 있는지, 그리고 가능하다면 어떤 기준으로 선택하는 것이 좋은지 제안하고자 한다.

1. LLM 직접 개발

언론사가 AI를 조직 차원에서 활용하려고 할 때, 가장 큰 선택지이자 동시에 가장 도전적인 전략은 대규모 언어모델(LLM)을 직접 개발하는 것이다. 이는 단순히 외부 API를 빌려 쓰는 수준을 넘어, 언론사가 독자적인 언어모델을 구축해 내부 데이터와 결합함으로써 뉴스룸에 특화된 AI 생태계를 가지겠다는 의미다. 장점은 명확하다. 첫째, 콘텐츠와 데이터에 대한 주권을 지킬 수 있다. 언론사가 축적한 기사 아카이브, 인터뷰 대본, 사진 캡션 같은 자료는 그 자체가 독보적인 학습 자원인데, 외부 API에 그대로 흘려보내면 사

실상 자산 유출이나 다름없다. 직접 모델을 만든다면 언론 고유의 문체는 물론, 엄격한 큐레이션과 검증을 거친 내부 데이터를 활용해 모델의 답변이 '사실성'을 지향하도록 훈련시키고, RAG와 같은 기술을 통해 최신 정보와의 연동을 강화할 수 있다. 둘째, 독립적인 기술 전략을 수립할 수 있다. 외부 API는 가격 변동이나 서비스 정책 변화에 따라 언제든지 리스크가 발생한다. 반면 자체 모델을 보유하면 비용 구조를 조정하거나 편집 워크플로우에 맞게 기능을 설계할 수 있다. 그리고 무엇보다, 언론사가 LLM을 자체 개발했다고 한다면 대외적인 이미지 제고 효과가 클 것이다. “와 ‘딱지신문’은 LLM을 자체 개발할 정도의 역량이 되는 곳이었어!”라는 평을 들을 수 있다는 말이다.

하지만 실제로 해야 할 일의 규모와 함께 들어가는 비용 등을 감안해 본다면 저런 장점들이 그렇게 대단하지는 않을 것이라는 사실을 대부분 알고 있을 것이다. 학습에 필요한 장비를 구축하는 것은 부차적인 일이다. 장비 차원에서만 보자면, 140GB 램을 장착한 최신 GPU인 H200 8장 정도를 하나의 클러스터로 구성해 3~5일 정도 풀가동할 수 있는 대여료와 전기료, 인터넷 요금 등이 있으면 누구나 자신의 목적에 맞게 LLM을 파인튜닝할 수 있다. 예를 들어 GPU 대여 전문 사이트인 바스트AI(vast.ai)에서는 H200 GPU 8대를 시간당 약 12.125달러(약 1만 5,000원)부터 빌려주고 있으니 3일 풀가동이라고 하면 약 873달러(약 121만 원), 5일 풀가동이라고 하면 1,455달러(약 202만 원) 정도에 나만의 LLM을 만들 수 있는 것이다. 참고로 대여라고 해서 물리적으로 GPU를 가져오는 것은 아니다. 이미 구성돼 있는 서버를 원격으로 접속해 사용할 수 있는 권한을 얻는 것이다. 더 싸게 빌리고 싶으면 한국에서는 광주광역시에 있는 인공지능산업융합사업단(AICA) 같은 곳에서도 정부 예산을 붙여서 저렴하게 대여 지원 사업을 하고 있고, 이미 구매해 둔 장비를 싸게 판매하거나 대여해 주는 국내의 대형 통신 회사도 많다. 요컨대, 장비는 큰 문제가 아니다.

정작 더 중요한 문제는 이 LLM을 어떤 목적으로 사용할 것인가, 그리고 어떤 데이터를 학습시킬 것인가 두 가지 질문이다. 특히 모델을 어디에 활용할 것인지 선명히 할 필요가 있다. 기사 초안 작성, 속보 요약, 헤드라인 생성, 아카이브 검색 보조 등 목표에 따라 필요한 데이터 종류와 모델 규모가 달라지기 때문이다. 뾰족하게 목표 설정만 할 수 있다면 데이터를 준비하는 데 필요한 시간과 비용, 그리고 노력을 획기적으로 줄일 수 있다. 예를 들어 헤드라인 생성만 놓고 본다면, 인풋(기사)과 아웃풋(헤드라인) 두 가지를 넣어서 모델 학습을 시킬 수 있다. 특정 스타일의 헤드라인만을 잘 생성하는 파인튜닝 모델을 만들고 싶다면 데이터셋 1만 5,000개 정도를 잘 준비하면 H200 GPU 8대 정도가 탑재된 환

[그림 12] 바스트AI에서 H200 GPU를 빌리기 위한 가격 일람표(2025년 9월 15일 기준)

m:34820	host:221867	Hungary, HU	Z13PN-D32 Series PCIe 5.0,16x 49.1 GB/s	↑601 Mbps ↓537 Mbps 899 ports	verified Max Duration 1 mon, 17d	\$12.125/hr
	8x H200					
vast.ai	428.2 TFLOPS Max CUDA: 12.8	140 GB 4032.1 GB/s 478.1 GB/s	XEON® PLATINU... 192.0/192 cpu 2064/2064 GB	SAMSUNG MZQL... 4243 MB/s 1039.9 GB 259.0 DLP/\$/hr	3140.4 DLPerf Reliability 99.89%	RENT
Type #19120413						
m:39263	datacenter:237688	Japan, JP	HPE Cray XD670 PCIe 5.0,16x 52.8 GB/s	↑4185 Mbps ↓2700 Mbps 12999 ports	verified Max Duration 15 days	\$23.533/hr
	8x H200					
vast.ai	428.2 TFLOPS Max CUDA: 12.6	140 GB 3119.8 GB/s 478.1 GB/s	XEON® PLATINU... 192.0/192 cpu 2064/2064 GB	SAMSUNG MZQL... 20193 MB/s 5871.9 GB 154.9 DLP/\$/hr	3646.1 DLPerf Reliability 99.38%	RENT
Type #2548523						
m:41307	host:288273	Michigan, US	0HCXFH PCIe 5.0,16x 47.3 GB/s	↑3636 Mbps ↓7917 Mbps 999 ports	verified Max Duration 5 mon.	\$17.886/hr
	8x H200					
vast.ai	428.2 TFLOPS Max CUDA: 12.2	140 GB 3120.0 GB/s 478.1 GB/s	XEON® PLATINU... 240.0/240 cpu 2064/2064 GB	Micron_7450_MT... 2334 MB/s 2395.0 GB 147.9 DLP/\$/hr	2646.4 DLPerf Reliability 99.19%	RENT
Type #2488322						
m:36199	datacenter:237688	Japan, JP	HPE Cray XD670 PCIe 5.0,16x 52.7 GB/s	↑1133 Mbps ↓951 Mbps 12999 ports	verified Max Duration 15 days	\$21.651/hr
	8x H200					
vast.ai	428.2 TFLOPS Max CUDA: 12.6	140 GB 3061.4 GB/s 478.1 GB/s	XEON® PLATINU... 192.0/192 cpu 2064/2064 GB	SAMSUNG MZQL... 20915 MB/s 25750.8 GB 167.3 DLP/\$/hr	3623.1 DLPerf Reliability 99.58%	RENT
Type #1988523						
m:42410	host:287287	France, FR	SB27C00631 PCIe 5.0,16x 50.4 GB/s	↑7741 Mbps ↓8200 Mbps 16383 ports	verified Max Duration 15 days	\$17.557/hr
	8x H200					
vast.ai	428.2 TFLOPS Max CUDA: 12.7	140 GB 4042.8 GB/s	XEON® PLATINU... 192.0/192 cpu 3096/3096 GB	SAMSUNG MZW... 33948 MB/s 10528.9 GB 92.8 DLP/\$/hr	1628.7 DLPerf Reliability 97.9%	RENT
Type #2588025						
m:32543	datacenter:135125	US	DGXH200 PCIe 5.0,16x 53.2 GB/s	↑1394 Mbps ↓6417 Mbps 1999 ports	verified Max Duration 2 mon, 8d	\$22.591/hr \$22.180/hr
	8x H200					
vast.ai	428.2 TFLOPS Max CUDA: 12.7	140 GB 4046.6 GB/s	Xeon® Platinum ... 224.0/224 cpu 2064/2064 GB	SAMSUNG MZW... 54684 MB/s 19128.3 GB 86.0 DLP/\$/hr	1908.0 DLPerf Reliability 99.933%	RENT
Type #2188025						
m:37254	datacenter:237688	Japan, JP	HPE Cray XD670 PCIe 5.0,16x 52.8 GB/s	↑1344 Mbps ↓2015 Mbps 12999 ports	verified Max Duration 15 days	\$21.652/hr
	8x H200					
vast.ai	428.2 TFLOPS Max CUDA: 12.6	140 GB 3114.5 GB/s	XEON® PLATINU... 192.0/192 cpu 2064/2064 GB	SAMSUNG MZQL... 16890 MB/s 6305.9 GB 86.3 DLP/\$/hr	1869.6 DLPerf Reliability 99.81%	RENT
Type #2399324						

* 출처: 바스트AI

경에서 1~3시간 안에 제작이 가능하다. 켄마, OSS-GPT 등과 같은 기본적으로 잘 훈련된 베이스 모델을 기반으로 해서 이 파인튜닝 모델을 접목하면 범용적인 서비스로도 만들 수 있다. 실제로 노르웨이의 미디어그룹 십스테드(Schibsted)는 자사 콘텐츠가 구글 등 검색엔진에서 최적으로 노출되도록 SEO 최적 헤드라인 생성 모델을 훈련한 적이 있다. 구체적인 모델 생성 비용과 데이터셋 규모 등은 공개되지 않았지만, 작업 결과 챗GPT 등에서 도출될 수 있는 결론에 비해 5배 정도 효과적인 헤드라인이 생성됐다는 보고가 있기도 했다. ‘무엇을 하고 싶다’가 잘 정의돼 있다면 시간과 비용이 수천분의 일로 줄어 들 수 있고, 현재 범용화되어 있는 LLM 서비스들을 능가하는 효과적인 제품이 나올 수도 있다.

자연스럽게 그다음으로 중요하고 어려운 단계로 넘어가게 된다. 바로 ‘데이터’다. 목적에 맞게 언론사가 내부적으로 갖고 있는 데이터를 가공·검증하고, 법적 윤리적으로 문제가 없는지 검토하는 작업은 여간 힘든 일이 아니다. 하나의 단일 단위로서의 작업은 어렵지 않은데, 그 수량이 수천수만 개를 넘어가기 때문에 힘든 것이다. 이는 다른 누구에게도 강요할 수 없고 언론사가 자체적으로 해야 한다. 바로 여기서 모델 퀄리티가 달라지기 때문이다. 특히 데이터 정제는 단순히 양만의 문제가 아니다. 언론 데이터는 고유한 특성을 가지고 있다. 같은 사건을 보도하더라도 언론사별로 서술 방식이 다르고, 심지어 한 언론사 내부에서도 기자마다 문체가 다르다. 또, 보도자료를 거의 그대로 옮긴 기사와 심층 취재를 거쳐 작성한 기사는 질적으로 차이가 크다. LLM 학습을 위해서는 이러한 데이터를 선별적으로 걸러내야 한다. 특정 기자의 문체를 반영하고 싶다면 해당 기자의 기사만 추출해야 하고, 특정 장르(경제 기사, 스포츠 기사 등)에 특화된 모델을 원한다면 그에 맞는 레이블링 과정이 필요하다. 결국 ‘데이터셋 설계 철학’이 모델 성능을 좌우하게 되는 셈이다.

문제는 여기서 끝나지 않는다. 법적·윤리적 검토도 반드시 필요하다. 기사에는 종종 실명, 주소, 민감한 개인정보가 포함된다. 이를 그대로 학습시킨 모델이 개인정보를 그대로 재현한다면 심각한 법적 문제로 이어질 수 있다. 따라서 개인정보 비식별화, PII 마스킹(개인식별정보를 가리는 기술), 날짜·숫자 표준화 같은 과정을 반드시 거쳐야 한다. 더 나아가 오보나 정정 기사까지 포함돼 있다면, 모델은 잘못된 사실을 ‘진실’로 학습할 위험이 있다. 이 때문에 데이터셋을 구축할 때는 단순한 크롤링을 넘어, 저널리즘적 검증 절차를 AI 데이터 파이프라인에 내재화해야 한다.

최근에는 이 과정을 완전히 수작업으로 하는 대신, 데이터 정제 자체를 또 다른 AI에 맡기는 방식이 늘어나고 있다. 예를 들어 앞서 보았던 기사 품질 평가모델을 돌려 사실성 점수, 출처 명확성, 편집 톤 등을 기준으로 데이터를 자동 분류하거나, 유사 기사 군집화를 통해 중복 데이터를 줄이는 식이다. 이렇게 일차적으로 AI가 데이터를 걸러주면 기자나 데이터 큐레이터는 최종 검수 역할에 집중할 수 있다. 즉, ‘AI를 위한 데이터’를 준비하는 단계에서도 이미 AI가 활용되는 것이다.

궁극적으로 언론사가 LLM을 직접 개발하려는 목적이 무엇이든, 가장 큰 병목은 장비가 아니라 데이터다. GPU는 돈만 있으면 빌리거나 살 수 있지만, ‘좋은 데이터’는 조직이 오랜 시간 축적해 온 저널리즘 자산 없이는 결코 만들어지지 않는다. 하지만 그 데이터 준비는 그야말로 ‘노가다’에 가까운 영역이다 보니, 먼저 LLM을 개발하려는 목적을 명

확하게 해 그 과정을 줄이는 노력을 하는 것이 효율적이다. 다시 말해 LLM을 직접 개발하려는 언론사라면, 모델 크기나 학습 시간보다 먼저 “우리가 가진 데이터는 어떤 특성이 있으며, 이 데이터를 어떻게 가공하면 경쟁력이 생길까?”라는 질문에 답해야 한다. 바로 그 지점이 언론사만이 할 수 있는 일이고, 동시에 언론사가 AI 시대에도 차별성을 유지할 수 있는 근거가 될 것이라고 본다. 만일 그러한 질문에 보족한 답이 없다면 필자는 범용화된 AI 제품들을 그대로 가져와 쓰는 것도 훌륭한 대안이 될 거라 생각한다.

2. RAG 도입 방안

LLM 직접 개발이 이상적이지만 현실적으로 쉽지 않은 길이라면, 언론사가 조직 차원에서 선택할 수 있는 또 다른 전략은 바로 RAG의 도입이다. RAG는 이미 학습된 대규모 언어모델에 새로운 데이터를 직접 학습시키는 대신, 모델이 필요할 때마다 외부 지식 베이스에서 정보를 검색·추출해 활용하는 방식이다. 쉽게 말해, 모델이 모든 것을 기억하고 있을 필요 없이 필요할 때마다 ‘자료를 찾아본 뒤 답을 만드는’ 접근이다. 언론사에는 방대한 기사 아카이브, 사진·영상 자료, 인터뷰 스크립트, 심지어 과거 정정 기사까지 포함된 데이터가 존재하는데, 이런 자료를 매번 모델에 재학습시키려면 시간과 비용이 천문학적으로 든다. 그러나 RAG를 적용하면 기존 LLM은 그대로 두고, 별도의 벡터 데이터베이스에 언론사 내부 자료를 인덱싱해 두면 된다. 모델은 독자의 질문이나 기자의 요청을 받았을 때, 아카이브에서 가장 관련성이 높은 문서를 검색해 가져와 답변의 근거로 삼는다. 이렇게 하면 모델이 ‘새로운 사실’을 매번 학습할 필요가 없고, 항상 최신 정보에 접근할 수 있다.

예를 들면 기자들이 기사를 입력하고 있는 와중에 해당 기사와 연관된 과거 팩트들이 아카이브에서 RAG를 통해 읽어져서 기사 형태로 기자에게 전송된다면, 해당 기자는 이른 시일 안에 데스크가 원하는 분량의 기사를 충분한 배경 데이터가 포함된 형태로 제공할 수 있게 된다. 이미 관련 서비스들이 많이 제공되고 있는데, 그 뒤에는 이런 RAG라고 하는 기술이 포함되어 있다고 보면 된다. 이미 앞에서 봤던 팩트체크 API와 같은 제품들을 설계하는데도 RAG는 기반 기술로 쓰이고 있다.

엄밀히 말하자면 RAG의 핵심 목적은 LLM의 지식 최신성 한계 보완, 환각 감소, 답변 근거 제시에 있다. LLM이 방대한 정보를 한꺼번에 GPU 메모리에 올려 처리하기 어려운 한계를 보완해, 필요한 정보를 외부 저장소에서 검색해 활용함으로써 컨텍스트 윈

도우(LLM이 한 번에 기억하고 처리할 수 있는 입력 텍스트의 길이 한계)를 확장하고 효율성을 높이는 부수적인 이점을 제공한다. 결국 RAG는 연산할 때 발생하는 GPU 메모리의 한계를 극복할 수 있게 해 주는 기술이다. 즉, 한꺼번에 GPU 메모리에 데이터들을 올릴 게 아니라, HDD나 SSD 같은 임시 메모리에 관련 데이터들을 저장시켜 놓고 필요할 때마다 쉽게 꺼내 쓸 수 있도록 해 두는 기술인 것이다. 예를 들어 이론적으로 전 세계의 모든 언어를 벡터화해 둘 수 있을 정도의 거대한 HBM 같은 것이 존재해서 거기에 모든 것이 저장돼 있고, 지속적으로 전기가 공급돼 해당 메모리가 살아있을 수 있다면, RAG 같은 기술 따위는 필요가 없는 것이다.

여기에서 RAG가 꼭 필요한 또 하나의 핵심적인 이유는 바로 LLM이 과거 데이터만 학습할 수 있다는 점 때문이다. 예를 들어 챗GPT나 제미니 등에 미국 대통령이 누구인지를 물어보면 웹 검색으로 들어가는 것을 볼 수 있는데, 이는 웹에서 최신 정보를 검색한 다음 RAG를 이용해 웹 문서를 참조해 대답해 주는 것이다. 왜 이런 일을 하느냐 하면, 두 서비스가 제공하는 모델들이 아직 도널드 트럼프 대통령이 취임하기 전의 데이터를 학습했을 경우, ‘미국 대통령은 조 바이든이다’라는 답을 해 줄 가능성이 크기 때문이다.

따라서 베이스가 되는 LLM 개발은 자체적으로 하지 않더라도, 언론사들이 RAG를 위한 시스템을 마련해 두는 것이 최신 데이터베이스를 기반으로 한 서비스를 내·외부적으로 하기 위해서라도 필요할 수 있다. 미국 언론사들 사이에서도 RAG는 가장 빠르게 도입되는 기술 중 하나다. 예컨대 미국 워싱턴포스트는 아카이브 검색 시스템을 개선해 기자들이 수십 년 치 기사를 몇 초 만에 조회할 수 있도록 실험 중이다. 또 뉴욕타임스는 외부 보도를 직접 학습시킨 모델보다는, 내부 데이터베이스를 안전하게 RAG로 연결하는 방식을 선호한다고 알려져 있다.

그렇다면 실제로 언론사가 RAG를 도입하기 위해서는 어떤 절차를 밟아야 할까. 일반적으로는 다섯 단계로 나누어 설명할 수 있다.

첫째, 데이터 수집과 정제 단계다. 언론사 내부에 있는 데이터들은 모두 바로 학습이나 검색에 활용될 수 있는 것이 아니다. 텍스트로 변환하고, 날짜·기자명·출처 같은 메타데이터를 정리해 뒤야만 검색 품질이 제대로 나온다. 데이터 준비 단계에서의 꼼꼼함이 곧 검색 정확도를 좌우한다.

둘째, 임베딩(embedding) 구축이다. 단순한 텍스트는 사람은 쉽게 이해할 수 있지만, AI 모델은 직접적으로 이해하기 어렵다. 따라서 문서를 의미 단위로 나눠 ‘벡터’라는 수

치로 변환해야 한다. 이를 통해 기사의 문맥과 의미가 수학적으로 표현되고, 나중에 유사한 질문이 들어왔을 때 관련성이 높은 자료를 찾아낼 수 있다. 예를 들어 “미국의 대통령은 도널드 트럼프다”라는 문장이 있다고 하자. 이 문장을 그대로 숫자로 바꿀 수는 없으니, 먼저 토큰화(tokenization) 과정을 거친다. 문장을 ‘미국’, ‘의’, ‘대통령’, ‘은’, ‘도널드 트럼프’, ‘다’와 같은 단어·어절 단위로 쪼개는 작업이다. 그 다음 이 토큰들을 어떤 모델(여기서는 BERT 모델)에 집어넣으면, BERT는 각각의 토큰을 고차원 공간의 좌표 값(벡터)으로 변환한다. 예를 들어 ‘대통령’이라는 단어는 768차원짜리 벡터로 바뀌는데, 이 벡터는 단순히 철자 정보를 담는 것이 아니라 주변 단어 맥락까지 반영한다. 그래서 같은 ‘대통령’이라도 ‘미국 대통령’ 문맥에 있는 경우와 ‘한국 대통령’ 문맥에 있는 경우 벡터 값이 달라진다. 이것이 바로 문맥적 의미(contextual meaning)를 반영하는 임베딩이다. 이렇게 얻어진 벡터들은 서로 간의 거리를 계산할 수 있다. 만약 ‘대통령 선거’라는 표현을 임베딩했을 때, 그 벡터는 ‘대선’이라는 표현의 벡터와 가까운 위치에 놓인다. 즉, 단어가 달라도 의미가 비슷하면 벡터 공간에서는 가깝게 모이게 되는 것이다. 이런 특성 덕분에 나중에 사용자가 “대선 관련 기사 찾아줘”라고 검색했을 때, 데이터베이스에서는 ‘대통령 선거’라는 표현이 들어 있는 기사도 자동으로 걸러낼 수 있게 된다. 언론사 입장에서는 이 과정을 거쳐야 기사 아카이브를 단순 키워드 검색이 아니라 의미 기반 검색으로 활용할 수 있다. 이처럼 사람이 이해할 수 있는 텍스트를 AI 모델이 의미를 파악할 수 있는 수치 벡터로 전환하는 과정이 임베딩이다.

실제 RAG 과정에서의 임베딩은 먼저 기사를 의미 단위로 잘게 쪼개는 것에서 시작한다. 한 편의 기사를 통째로 벡터화하면 질문과 직접 관련 없는 부분까지 검색 결과에 포함될 수 있기 때문에, 보통은 문단 단위나 일정 토큰 단위(예: 500~1,000 토큰)로 나눠 맥락을 보존하면서도 검색 효율성을 높인다. 이렇게 잘라낸 텍스트는 임베딩 모델에 입력되는데, 이 모델은 글을 고차원 벡터로 변환해 의미적 특징을 수치화한다. 덕분에 ‘대통령 선거’와 ‘대선’처럼 단어는 달라도 같은 의미를 가진 표현들이 벡터 공간에서는 서로 가깝게 위치하게 된다. 현재 널리 사용되는 임베딩 모델로는 오픈AI(OpenAI)의 텍스트 임베딩-3(text-embedding-3), 코히어(Cohere)의 임베딩 모델, 구글(Google)이 공개한 유니버설 센텐스 인코더(Universal Sentence Encoder), 그리고 연구자들이 자주 활용하는 센텐스-버트(Sentence-BERT) 계열 등이 있다. 이런 모델들을 활용하면 언론사가 보유한 기사와 문서들을 의미 단위로 정밀하게 매핑할 수 있고, 이후 검색 품질을 획기적으로 끌어올릴 수 있다.

셋째, 이렇게 벡터화된 데이터를 보관할 벡터 데이터베이스를 설계하는 단계가 필요하다. 대표적으로는 파이스(FAISS), 파인콘(Pinecone), 위비에이트(Weaviate) 같은 오픈소스·상용 솔루션이 있다. 언론사가 자체적으로 아카이브를 운영할 수 있고, 필요하면 보안 수준을 높여 내부 전용 검색망을 구축할 수도 있다. 결국 이 데이터베이스가 언론사 AI 시스템의 ‘지식 창고’가 되는 셈이다.

넷째, LLM 연동 단계다. 기자나 독자가 질문을 던졌을 때 먼저 벡터 데이터베이스에서 관련 문서가 검색되고, 그 결과가 LLM의 입력 프롬프트에 함께 전달된다. 오픈소스 모델(제마, OSS-GPT, 라마, 미스트랄, 덤시크, 큐웬 등)을 내부적으로 돌릴 수도 있고, 상용 API(오픈AI, 앤트로픽 등)를 활용할 수도 있다. 중요한 것은 모델이 자기 ‘기억’만으로 답을 내놓는 것이 아니라, 반드시 외부 근거 문서를 참고하도록 만드는 것이다.

다섯째, 마지막은 UX/UI 통합이다. 아무리 뛰어난 RAG 시스템이라도 기자나 독자가 쉽게 쓰지 못한다면 의미가 없다. 따라서 기자용 CMS 안에 ‘관련 아카이브 검색’ 버튼을 넣거나, 독자용 뉴스 앱에서 “지난주 기후 기사 모아줘” 같은 요청에 응답할 수 있도록 해야 한다. 이렇게 실제 현장에서 매끄럽게 활용될 수 있을 때 비로소 RAG의 효과가 드러난다.

이미 앞서도 몇 차례 살펴본 바가 있지만, 일반적으로 RAG는 그렇게 어려운 기술은 아니다. 중요한 것은 데이터를 수집하고 정제하는 것인데, 그 과정 또한 소규모 단위로 테스트해 볼 수 있기 때문에 수천수만 개의 데이터를 완벽하게 준비해야만 비로소 시작할 수 있는 LLM 학습 또는 파인튜닝에 비해서는 훨씬 수월하다고 할 수 있다. 또한 데이터를 재학습 시키지 않기 때문에 법적 리스크도 상대적으로 덜하다. 그러나 검색 품질을 높이기 위해 소프트웨어를 잘 만들기 위해 노력해야 할 것이고, 데이터베이스를 운영하고 관리하는 데 자원이 필요할 것으로 예상된다. 이 정도의 관리 역량을 갖춘다면 여러 가지 해 볼 수 있는 것들이 많으므로 언론사 차원에서 관심을 가질 필요성과 이유가 많다고 필자는 생각한다.

3. 오픈소스나 API냐

언론사가 AI 기술을 도입할 때 가장 실질적인 고민은 “오픈소스 모델을 직접 다룰 것인가, 아니면 상용 API를 활용할 것인가”라는 문제로 귀결된다. 두 접근법은 분명히 각자의 장단점이 있다.

오픈소스를 선택하면, 데이터 주권과 기술 자율성이 보장된다. 언론사가 축적한 아카이브를 외부 기업에 넘기지 않고 자체적으로 관리할 수 있으며, 필요하다면 언론 특화 모델로 마음껏 커스터마이징할 수 있다. 무엇보다 운영 과정에서 획득한 노하우와 새로운 사업 가능성을 발견할 수 있다는 점은 상용 API를 사용할 때 절대 가져올 수 없는 장점이다. 이는 단순히 비용 절감 차원을 넘어, 언론사가 스스로 기술 기업으로 성장할 수 있는 발판이 된다. 내부적으로는 기자와 개발자가 함께 문제를 해결하면서 조직의 디지털 역량이 높아지고, 외부적으로는 자체 기술을 바탕으로 다른 언론사나 기관과 협력하거나 새로운 수익 모델을 만들어 낼 기회가 열린다.

반대로 상용 API를 활용하면, 안정적이고 빠르게 서비스를 구축할 수 있다. 기자들이 내일 당장 쓸 수 있는 기사 요약기, 헤드라인 추천기, 아카이브 검색 보조 도구를 빠르게 만들 수 있는 것이다. 운영·보안·모델 업데이트는 모두 외부 업체가 맡아주기 때문에, 언론사는 기획과 데이터 품질 관리에만 집중하면 된다. 그러나 API 방식은 비용 변동성과 데이터 의존성이라는 리스크를 안고 있다. 가격이 오르거나, 정책이 바뀌거나, 특정 기능이 중단되면 언론사는 속수무책일 수 있다. 또한 내부 데이터가 외부 서버로 전송되는 과정에서 보안·저작권 문제가 발생할 가능성도 있다. 특히 대다수 상용화 서비스를 기자들이 쓸 때 민감한 내부 정보가 외부로 전달될 가능성이 늘 존재한다고 볼 수 있다.

하지만 현실적으로 더 관심이 있는 부분은 비용일 것이다. 결론은 분명히 나와 있긴 하지만(사용량이 많지 않다면 상용 API를 쓰는 것이 오픈소스 및 파인튜닝 모델을 자체적으로 운영하는 것보다 낫다) 둘의 비교방정식을 세워보는 것은 사용량에 따라 무엇을 선택하는 것이 나올지 스스로 판단해 보는 데 도움이 될 것 같다.

먼저 상용 API를 기준으로 가정해 보자. 현재 많이 쓰이는 오픈AI GPT-4o API의 텍스트 가격은 입력 100만 토큰당 2.50달러(약 3,500원), 출력 100만 토큰당 10달러(약 1만 4,000원)다. 입출력 비율을 5대 5로 가정하면, 총 1만 토큰(10k)을 처리하는 데 드는 비용은 약 0.0625달러(약 87원)가 된다. 기자 한 명이 하루에 30건 정도의 기사 요약, 제목 추천, 아카이브 검색을 AI에게 맡긴다고 하면, 기사 1건당 평균 2,000 토큰(입력+출력)을 소비한다고 산정해 볼 수 있다. 하루 사용량은 약 6만 토큰, 즉 0.4달러(약 557원) 수준이다. 기자 100명이 동시에 이 기능을 사용하면 월 약 1,200달러(약 167만 원) 수준이 된다. 따라서 내부적으로만 사용하기에는 상용화 서비스를 쓰는 것이 분명 이점이 있다. 하지만 실제 뉴스룸 단위에서의 요약·검색뿐 아니라 아까 언급한 팩트체크 API처럼 고객을 대상으로 서비스를 실시하면 토큰 사용량은 급세 기하급수적으로 늘어날 수 있다. 단순히 하

루 1만 토큰 정도를 사용하는 고객이 3만 명 정도 되는 대규모 서비스(매일 3억 토큰 사용)를 운영한다고 하면, GPT-4o API를 사용할 경우 하루 약 1,875달러(약 261만 원), 연간 약 68만 4,375달러(약 9억 5,000만 원)의 API 비용이 든다.

반대로 오픈소스 모델을 자체적으로 운영하는 경우를 보자. 우선 초기 투자비와 운영비를 나눠 생각할 필요가 있다. 초기에는 모델 학습·파인튜닝 비용이 발생한다. 예를 들어 기사 헤드라인 생성 전용 모델을 만들기 위해 H200 GPU 8장을 빌려 2~3일 정도 학습을 돌린다고 가정하면, 하루 약 230달러(약 32만 원), 총 700~1,000달러(약 97만~139만 원) 수준에서 원하는 파인튜닝 모델을 확보할 수 있다. 이는 일회성 비용이다. 또 내부 기사 및 외부 고객들에게서 오는 트래픽을 감당하기 위해 별도의 장비를 구매하기 위한 일회성 비용도 들어가는데, 하루 3억 토큰을 처리하는 트래픽을 감당하려면 최소 30장 정도의 RTX5090급 GPU가 필요하므로 해당 장비를 포함한 워크스테이션 구매 가격이 1억 원 이상 소요될 것이다(RTX 5090 하나당 2,400달러 가정). 여기에 해당 장비를 놓을 장소대여료, 전기료, 인터넷 요금(보통 이 세 가지를 한 번에 해결해 주는 서비스 제공업체들도 있다) 등이 연간 1억 원 이상 들어간다고 보면 된다. 설치에 필요한 인건비는 포함되지 않았다. 이를 감안하면 최초 설치 비용으로 일간 3억 토큰 정도를 처리하는 서비스를 제공하기 위해 언론사가 최초 연도에 투자해야 하는 비용은 약 1억 5,000만 원 정도가 될 것으로 추정된다.

[표] 상용 API와 오픈소스 모델 도입·운영 비용 비교

구분	상용 API (예: GPT-4o)	오픈소스 모델 (예: 젤마3)
도입 비용	초기 비용 없음	GPU 포함 워크스테이션 구매 (예: RTX5090 × 30대 ≈ 1억 원) 파인튜닝 모델 학습 비용 1개당 100만~200원 데이터 처리 비용 미포함
운영 비용 (내부 사용)	기사 100명 기준 월 1,200달러(약 167만 원) * 기사 요약·제목 추천 위주	전기·냉각·엔지니어 인건비 월 300~500만 원
운영 비용 (대규모 서비스)	하루 3억 토큰 처리 시 연간 약 9억 5,000만 원(API 사용량 기준)	같은 규모 처리 시 GPU 30대 클러스터 + 전기료 포함 연간 약 1억 원
장점	빠른 시작, 안정적 운영, 보안·업데이트 외주화	데이터 주권 확보, 커스터마이징 가능, 장기적으로 원가절감
단점	가격 변동·정책 변경 리스크, 데이터 외부 전송 부담	초기 투자 부담 운영·보안·기술 인력 필요

무수히 많은 가정이 포함돼 있어서 개별 상황에 따라 수치가 달라질 수 있는 비교이지만 대략적으로 이렇게 놓고 보면 일간 약 3,000만 토큰 정도의 처리 용량이 필요한 서비스를 언론사들이 제공할 수 있느냐가 하나의 기준점이 될 수 있을 것 같다. 그 이상의 서비스를 제공할 생각이라면 오픈소스 + 파인튜닝 모델을 자체적으로 운영하는 것이 나올 수 있고, 그렇지 않고 뉴스룸 내부에서의 사용이나 소규모 서비스 정도를 제공하고 싶다면 상용 API를 쓰는 것이 나올 수 있을 것이다.

강조하고 싶은 것은 오픈소스를 자체 운영하는 방식이 갖는 분명한 비용적 장점이 다. 단순히 생각해 봐도, 내부에서 모델을 운영한다면 비용 통제 및 효율화를 통해 원가를 절감할 여지가 많지만, 외부 모델을 사용할 경우 수동적으로 이미 결정된 가격에 따라야만 하기 때문이다. 게다가 이미 구매한 장비는 회사의 자산이 된다. GPU의 경우 중고 시장에 재판매하면 원가의 70% 이상 가치를 인정받는 경우가 흔하므로 회사 차원에서는 장비를 갖춰 둔다고 하여 크게 손해 볼 것은 없다. 2025년 9월 9일 미국의 언론 악시오스(Axios)는 이런 이유로 “AI 장비 투자는 마치 부동산 투자와 같다(What AI’s spending boom is really buying)”⁴⁾라는 내용의 기사를 내보내기도 했다. 예를 들어 언론사는 보다 효율적인 GPU가 나와서 지금 쓰고 있는 장비(예를 들면 RTX5090같은)보다 더 싼 장비로 일간 3억 토큰을 처리할 수 있다면 원가를 크게 줄일 수 있을 것이다.

물론 그 대가로 인프라 관리, 보안, 모델 업데이트, 운영 인력 확보 같은 부담이 언론사 내부에 쌓인다. 기술 인력이 충분하지 않거나, 조직 차원의 투자 의지가 약하다면 오히려 비효율적이 될 수 있다. 매년 엔비디아(NVIDIA) 등에서 새로운 장비가 발표되면 현 장비를 업그레이드해야 할 수요도 있을 수 있다. 아니, 그 전에 RTX5090과 같은 장비를 30대 정도 대량으로 구매하는 방법도 찾아봐야 할 수 있다. 물량 공급이 원활하지 않기 때문이다. 그러면 그때마다 장비 구매 비용이 늘어난다.

하지만 이런 고통을 해결하는 과정에서 조직 내 역량이 쌓일 수 있다. 필자가 강조하고 싶은 것이 바로 이 부분인데, 현재는 AI와 관련된 기술의 진보 속도가 매우 빠른 시점이므로 새로운 모델, 새로운 방법론, 새로운 장비 등이 등장하면 빠르게 적용해 보고 그를 통해 시장에서 필요로 하는 제품을 만들어 나가는 역량을 갖추 필요가 있다. 상용 API만 사용한다면 이런 역량이 절대 쌓일 수 없다. 예를 들어 ‘딱지신문’의 CEO가 이 글에 언급된 ‘팩트체크 API를 만들고 싶다’는 열망이 생긴다 하더라도 상용 API만 사용하고 있었

4) <https://www.axios.com/2025/09/09/data-centers-ai-spending-boom>

다면 그 열망을 실현할 방법을 찾기가 매우 힘들 것이다. 그러나 내부적으로 이미 오픈소스 모델들을 이용해 파인튜닝도 하고, 직접 고객에게 서비스할 수 있는 GPU 클러스터도 갖춰져 있다면 새로운 실험을 해 볼 수 있을 것이다.

요약하면, 소규모 매체나 파일럿 프로젝트 단계에서는 상용 API가 비용 효율성이 높고, 대규모 B2C 고객 서비스용 활용에서는 오픈소스 모델이 장기적으로 유리할 가능성이 크다. 또한 전략적 판단도 매우 중요하다. 장기적으로 AI의 활용 가능성이 크다고 보고 내부 역량을 쌓겠다는 판단이 있다면 기술을 내재화하는 것이 옳을 수 있다. 그러나 반대로 언론사의 본업이 AI를 활용해 무언가를 하는 것이 아니라, 진실을 취재하고 보도하는 것에 있다고 믿는다면 자체 모델과 서비스 인프라를 구축하는 것은 부수적인 일에 불과하다고 여길 수도 있다.

따라서 단순히 둘 중 하나를 택해야 한다기보다는, 혼합 전략이 현실적인 대안일 수도 있다. 초기에는 상용 API를 활용해 빠르게 내부 실험을 진행하고, 동시에 오픈소스 모델을 도입해 점진적으로 내재화 역량을 키워가는 방식이다. 예컨대 소규모로 GPU 클러스터를 조성해 뉴스룸 내부 일부 부서에서 사용하거나, 해당 부서가 고객 서비스를 하는데 활용하도록 하면서 가능성을 탐색하는 것이다.

나가며: 일단 해 봅시다

“인공지능이 기자들의 직관이나 사명, 독자와의 연결, 그리고 편집에 관한 판단을 재현할 수는 없다”(국제뉴스미디어협회)

필자는 이 말을 그다지 신뢰해서는 안 된다고 본다. 기자들의 직관은 이미 데이터에 의해 파괴되고 있다. 예를 들어 어떤 기자가 ‘부는 대물림 되는 것 같은데?’ 라는 직관을 갖고 있다면, 이미 AI는 자신이 파이프라인을 끼울 수 있는 데이터셋을 RAG를 통해 참조하고 자체적으로 리서치를 해 ‘부는 대물림 되는 것이 아니다. 어린 시절 같이 노는 친구그룹이 누구냐가 부에 가장 큰 영향을 끼친다(라즈 체티 외, 2022)’라는 인사이트를 도출해 낼 수 있을 것이다. 고객인 독자와의 연결도 그렇다. 이 글에서 본 것처럼 독자들이 원하는 팩트체크 서비스를 제공하는 일은 AI를 통해 충분히 자동화할 수 있다. 즉, 독자와의 연결이라는 가치를 AI가 제공해 줄 수 있는 기술적 가능성은 이미 나와 있다는 이야기다. 누가 이걸 멋지게 만들어서 기회를 먼저 가져갈까의 문제일 뿐이다. 결국 독자와의 연결 중 무의미한 부분은 AI에 의해 상당 부분 걸러지고, 유의미한 것들만 남게 될 가능성이 높다고 생각한다. 편집에 대한 판단 역시 전날 있었던 이야기를 ‘조간신문’이나 ‘9시 뉴스’와 같은 틀에 넣던 과거의 이야기일 수 있다. 세상이 시끌벅적해야 할 의제를 정하는 것은 이미 언론만의 역할이라고 보기에 힘들게 됐다. 영향력을 가진 ‘인플루언서’들은 이미 상당히 많이 탄생됐고 세상은 그렇게 변해버렸다. 이제 AI가 나의 성향과 니즈를 판단해서 나만을 위한 콘텐츠 추천을 순식간에 해 주는 시대다. 뉴스 역시 마이크로소프트의 퍼스널라이즈드 업데이트(Personalized Updates), 렛미노우뉴스(LetMeKnow.News), 스마트 뉴스(SmartNews) 등 사용자 관심 주제에 맞춰 특화된 추천을 해 주는 사례들이 수도 없이 많다.

거꾸로 필자는 저 문장을 완전히 뒤집어 생각해야 한다고 본다. “인공지능은 기자들의 직관과 사명을 수백 배 더 낮게 만들어 주고, 독자들과의 연결을 지금보다 수백 배 낮게 만들어 주고, 언론의 편집에 대한 판단에 수백 배 더 신뢰감을 줘서 영향력을 지금보다 훨씬 높일 수 있는 희망이자 돌파구”라고 말이다. 다소 과장된 표현이라 하더라도 그렇게 믿고 해 보는 자세가 가장 중요한 출발이라고 생각한다. 그렇지 않고 AI의 발전에 대해 “네가 아무리 발전해도 기자라는 직업과 언론이라는 산업은 영원할 거야”라고 말하는 것은 지적으로 성실한 자세도 아니고, AI에 대해서는 깊게 들어가지 못하는 스스로를 쉽게 놓아버리는 패배주의일 수도 있다고 본다. 과거 인터넷과 모바일 시기에 언론이 많이 겪었던 일들이기도 하다.

세상의 대다수 위대한 것들은 발명이라기 보다 발견이었다. 하다 보니 찾아진 거였다. 하버드대에서 친구들끼리 정보를 공유하던 사이트에서 시작한 페이스북은 타인을 염탐하고 싶은 인간의 욕구를 모바일이라고 하는 신기술로 절묘하게 발견해 냈다. 마크 저커버그가 그런 욕구를 처음부터 알고 사업을 시작한 것은 아니었다는 얘기다. 게임이 잘 돌아가는 반도체를 만들었을 뿐인 엔비디아는 이후 이 병렬로 연결한 반도체를 잘 통제할 수 있는 소프트웨어 CUDA를 설계했고, 그제 인공지능 학습에 엄청난 시간 단축을 가져다 준다는 사실이 스탠퍼드대의 앤드류 응(Andrew Ng) 교수 등에 의해 발견되면서, 오늘날 세계 반도체 시장에서 시가총액 1위 기업이 되었다. 젠슨 황(Jensen Huang) 엔비디아 CEO가 애초부터 인공지능을 염두에 두고 만든 제품도, 회사도 아니었다.

분명한 사실은 언론이 갖고 있는 본질적 가치는 너무나 확실하기 때문에 이를 AI에 점목시키려는 시도가 앞으로도 계속될 것이라는 점이다. 그 가치가 다른데 있지도 않다. 바로 국제미디어연합이 언급한 ‘기자의 사명(진실 파악)’ 그리고 ‘독자와의 연결(사실 보도)’ 그리고 ‘편집에 대한 판단(공동체에 필요한 아젠다 설정)’ 등이 바로 그들이다. 하나하나가 이 사회를 살아가는 개인들에게 너무나 소중한 가치들을 갖고 있기 때문에 지금보다 더 나은 방식(혁신)으로 이 가치들을 실현시킬 방법이 있다면 AI를 활용하고자 하는 니즈들은 언제 어디에서 출현하더라도 전혀 이상하지가 않다. 이런 상황에서 ‘인간의 가치는 영원할 것이다’라는 말만 믿고 아무것도 하지 않고 있는 것은 너무 안이하지 않을까.

따라서 지금 언론이 해야 할 일은 두려움 속에서 변화를 늦추는 것이 아니라, 작은 실험이라도 시작해 보는 것이다. 데이터 아카이브를 열어 RAG로 검색을 붙여 보든, 기사 개인이 기사 요약기를 활용해 보든, 혹은 소규모 팩트체크 API를 만들어 보든 간에 말이다. 일단 해 보는 것이 가장 좋은 출발점이고, 그 과정에서 고객이 원하는 진정한 수요가

발견되면 이를 빠르게 AI 등의 도구로 만족시키는 일련의 시행착오 과정이 언론 산업의 미래를 긍정적 방향으로 이끌 것이라고 믿는다.

4호

AI 시대, 저널리즘의 새 판을 짜다: DataJConf 2025 탐방기

01	들어가며: DataJConf 2025를 가다	121
02	생성형 AI가 조직을 뒤흔들다: 뉴스룸 ‘재건축’	125
03	뉴스룸을 바꾸는 작은 실험들	132
04	뉴스룸의 AI 리터러시	142
05	국경·언어·조직을 넘어: 데이터 기반 협업 저널리즘	146
06	AI 시대, 독자와 언론의 새로운 관계를 설계하라	153
07	나가며: DataJConf 2025가 남긴 것	162

들어가며: DataJConf 2025를 가다

그리스 수도 아테네에서 9월 8~10일 열린 ‘2025년 유럽 데이터 및 컴퓨테이셔널 저널리즘 콘퍼런스(DataJConf, The European Data & Computational Journalism Conference)의 화두는 단연 인공지능(AI)이었다. 근대 민주주의의 근간이 마련된 그리스에서, 민주주의를 구성하는 중요한 축인 언론의 미래에 대한 토론과 탐색이 이어졌다. AI의 등장에 따른 고민의 산물로 뉴스룸의 미래에 대해 어느 때보다 진지한 격론이 오간 것은 물론이다.

컴퓨터를 활용한 데이터 저널리즘에서 확장해 AI를 어떻게 뉴스룸에 적용할 것인지를 두고 언론학자들과 현업에서 뛰고 있는 저널리스트들이 머리를 맞댔다. 유럽에서 열린 저널리즘 콘퍼런스인 만큼 영국의 BBC, 파이낸셜타임스, 독일의 슈피겔, 스위스 타미디어(Tamedia) 등 유럽 언론사 종사자들이 주요 발표를 맡았다. 데이터 저널리즘이 발전한 미국에서도 블룸버그, 뉴욕타임스 등의 전·현직 기자뿐 아니라 주요 대학의 언론학자들이 발제를 맡아 현장 연구 사례를 소개했다.

DataJConf는 대학에서 시작된 아카데미한 포럼으로 저널리스트와 연구자가 만나 뉴스 기관과 사회 과학자의 경험을 공유하고, 대중의 이익에 도움이 되는 새로운 유형의 저널리즘을 함께 탐구하는 자리다.

첫 콘퍼런스는 지난 2008년 미국 조지아주 애틀란타의 조지아공과대에서 열렸다. 당시에는 인쇄술이 등장한 지 500여 년 만에 언론의 본질을 변화시킬 만한 기술로 컴퓨팅 기술 발전에 주목했다. 이후 컬럼비아대, 스탠퍼드대, 노스웨스턴대, 마이애미대 등에서 콘퍼런스가 열렸으며 2023년 스위스에서 10번째 행사가 개최됐다. DataJConf는 2017년부터 미국 콘퍼런스와 분리돼 아일랜드, 영국, 스페인, 스위스 등에서 열렸고, 이번 그리스 콘퍼런스는 유럽 DataJConf의 다섯 번째 행사다.

2022년 오픈AI의 챗GPT 출시 이후 DataJConf는 AI가 언론에 미치는 영향에 초점을 맞추고 다양한 강연과 워크숍, 패널 토론을 마련했다. 2023년 스위스에서 개최한 콘퍼런스에 이어 2025년 그리스 콘퍼런스까지 AI와 뉴스의 미래, 딥페이크(AI를 활용한 가짜 콘텐츠) 윤리, 알고리즘 책임성, 뉴스룸 자동화 등 AI를 중심에 둔 논의가 진행됐다.

특히 콘퍼런스 참가자들은 구글로 대변되는 인터넷 검색이나 소셜미디어의 등장보다 AI가 언론에 미치는 변화의 속도가 훨씬 빠르다는 데 공감했다. 인터넷 검색과 소셜미디어가 미디어에 활용되는 데 최소 1~2년 이상의 시간이 걸렸다면, AI는 단 몇 달 만에 언론계를 완전히 뒤흔들었다는 것이다.

콘퍼런스 첫날은 프레젠테이션과 비슷한 성격의 언론인 연합 건물에서 AI를 활용한 저널리즘 교육 워크숍이 열렸다. 독일 슈피겔, 스위스 타미디어 등 현직 저널리스트와 대학 연구진들이 최신의 트렌드를 소개하는 자리였다.

[그림 1] 그리스 아테네 언론인 연합 건물의 펜의 힘을 상징화한 구조물



데이터 시각화를 위한 바이브 코딩, 파이썬을 이용해 텍스트와 이미지에서 데이터를 추출하는 방법 등에 직접 실습해 볼 수 있는 워크숍도 마련됐다. 비정형 텍스트에서 AI를 이용해 저널리즘적 가치가 있는 데이터를 추출하는 방법, 이미지로 존재하는 텍스트를 내추럴 PDF(텍스트를 구조화해 기계가 읽고 분석할 수 있도록 만든 PDF)로 변환해 분석 가능한 데이터로 만드는 방법 등이 소개됐다.

또 ‘쇼 앤 텔(Show and Tell)’이란 제목으로 독일 슈피겔지의 데이터 저널리스트 마르셀 파울리(Marcel Pauly) 기자가 그래픽 및 기사 자동화 시스템에 대해 설명했다. 미국의 데이터 저널리즘 학자인 드루밀 메타(Dhrumil Mehta) 컬럼비아대 교수는 AI를 활용한 데이

터 저널리즘 교육 강연을 했다.

행사 둘째 날과 마지막 날인 9~10일에는 유서 깊은 아테네 학당과 나란히 자리한 아테네대로 장소를 옮겨 발표와 토론 등이 이뤄졌다. 컴퓨터이셔널 저널리즘(Computational Journalism) 분야의 세계적인 전문가인 닉 디아코폴로스(Nick Diakopoulos) 노스웨스턴대 교수의 ‘AI가 미래 정보 생태계에 미치는 영향’에 대한 기조연설로 본격적인 콘퍼런스가 시작됐다.

이어 BBC, 파이낸셜타임스, 오스트리아 APA 통신 등 언론과 아리스토텔레스대, 네덜란드 암스테르담자유대, 독일 루트비히-막시밀리안 뮌헨대(LMU) 등에서 AI 저널리즘에 대한 언론과 학계의 실황을 들려줬다. 이날 콘퍼런스의 마무리는 BBC의 기술 예측 책임자 로라 엘리스(Laura Ellis)가 맡아 ‘AI 시대에 우리는 무엇을 할 수 있으며, 해야 할까요?’란 질문을 던졌다.

콘퍼런스 마지막 날에는 AI를 활용한 훨씬 심층적인 사례 발표가 이어졌다. 행사의 좌장을 맡아 처음부터 마지막까지 이끈 바하레 헤라비(Bahareh Heravi) 영국 서리대 교수는 콘퍼런스가 ‘사람 중심의 이니셔티브’로 성장해 왔다고 밝혔다.

헤라비 교수는 “언론사는 AI 도입에 있어 ‘신중함’을 유지하면서도 기자들이 AI를 사용하도록 ‘열정’을 불어넣는 균형감을 갖춰야 한다”라고 강조했다. AI가 인턴기자과 같은 인간의 초급 일자리를 대체할 수 있다는 우려에 대해서도 “‘일자리 상실’과 언론인들이 자신을 보호하고 기술을 숙련하는 ‘배움’ 사이에서 균형을 찾아야 한다”라고 덧붙였다.

BBC에서 AI 리터러시 작업을 지원했던 헤라비 교수는 “BBC와 같은 대규모 조직에서 의무적인 교육 프로그램을 개발해 배포하는 데는 1년이 걸린다”라면서 “그 과정에서 프로그램이 이미 ‘구식’이 될 위험이 있다”라고 지적하며 지속 가능한 교육 방식의 필요성을 제기했다.

이번 콘퍼런스를 관통한 핵심 주제는 한마디로 ‘저널리즘과 기술이 만나 재정의되는 미래’였다. 닉 디아코폴로스 교수는 기조연설에서 “우리는 모두 매일 같이 미래를 고민하고 있다”라고 말하며, 언론과 기술의 교차점을 이해하려는 시도가 더 이상 선택이 아니라 필수임을 강조했다. 이번 행사를 공동 주최한 콘스탄티노스 무를라스(Constantinos Mourlas) 아테네대 교수 또한 “저널리즘과 기술이 만나는 지점을 함께 규명하고, 무엇을 배워야 하는지 논의하기 위한 자리”라고 이 행사를 소개했다.

생성형 AI라는 거대한 변화 앞에서 저널리즘이 어떤 원칙과 방향을 선택해야 하는지, 그리고 이 미래를 누가 어떻게 설계할 것인지에 대한 집단적인 사유의 장이 바로 이 콘퍼

런스였다.

특히 DataJConf는 학계와 현업 저널리스트가 파트너로 머리를 맞대는 드문 자리라는 점에서 더욱 의미가 컸다. 연구자는 기술의 가능성과 한계를 설명하고, 저널리스트는 실제 뉴스룸의 문제와 현실적 제약을 가져와 토론하며 함께 저널리즘의 다음 단계를 설계했다. DataJConf 2025는 기술의 속도를 좇는 자리가 아니라, 기술 시대의 저널리즘이 지켜야 할 기준과 미래의 방향성을 학계와 저널리스트가 함께 그리는 자리였다.

[그림 2] DataJConf에 참가한 참가자들의 모습



생성형 AI가 조직을 뒤흔들다: 뉴스룸 ‘재건축’

2024년 1월, 로이터저널리즘연구소의 닉 뉴먼(Nic Newman)은 〈2024 저널리즘, 미디어, 그리고 테크놀로지 트렌드 및 예측〉 보고서에서 “인공지능의 파괴적 힘이 올해 정보 생태계를 휩쓸 것”이라는 예언을 내놓는다.¹⁾ 약 2년이 지난 지금, 이 문장은 더 이상 과장이 아니게 됐다. 2022년 챗GPT의 등장 이후, AI는 일상의 도구를 넘어 뉴스룸 전략·조직 논의의 중심으로 이동했다. BBC는 “혁신과 성장을 가속하기 위해 AI를 배치해야 한다”는 데보라 터니스(Deborah Turness) 최고경영자의 메시지와 함께 2025년 ‘성장, 혁신 그리고 AI(Growth, Innovation, and AI)’라는 전담 조직을 신설했다. 각국 주요 매체도 가이드라인과 거버넌스를 급히 정비했다. 생성형 AI의 등장은 도구의 문제가 아니라 편집 원칙·조직 설계·업무 문화를 재구성하는 사건이 됐다.

DataJConf 콘퍼런스에서도 단연 화두는 생성형 AI였다. 행사 둘째 날 열린 ‘뉴스룸의 AI 혁신: 가능성에서 실현으로’ 세션에서는 생성형 AI 시대에 뉴스룸이 어떻게 적응하고 있는지에 대한 논의가 이뤄졌다. 바하레 헤라비 서리대 교수의 진행 아래 로라 엘리스 BBC 기술 예측 책임자, 올리버 호킨스(Oliver Hawkins) 파이낸셜타임스 데이터 사이언티스트, 티투스 플래트너(Titus Plattner) 타미디어 AI랩 선임 혁신프로젝트 매니저, 카렌 플리팅(Karyn Fleeting) 리치 AI 딜리버리 디렉터 등이 논의에 참여했다.

토론자들이 소속된 BBC, 파이낸셜타임스, 타미디어, 리치 등은 이미 수년 전부터 번역, 전사(Transcription), 추천, 광고 최적화 등 다양한 영역에서 AI와 머신러닝을 활용해 온 조직이다. 그럼에도 불구하고 토론자들은 챗GPT 공개가 분기점이 됐다고 기억했다. 이

1) <https://reutersinstitute.politics.ox.ac.uk/journalism-media-and-technology-trends-and-predictions-2024>

전까지는 AI가 일부 연구·개발 부서의 실험 대상, 혹은 광고·비즈니스 영역의 기술적 인프라에 머물렀다면, 생성형 AI의 확산은 '뉴스룸 전체가 AI에 어떻게 대응해야 하는가'를 고민하게 만드는 계기로 작용했다는 평가다.

로라 엘리스 BBC 기술 예측 책임자는 생성형 AI를 마주한 BBC의 초기 접근 방식은 매우 조심스럽고 폐쇄적이었다고 회고한다. BBC는 조직 차원에서 '데이터를 챗GPT에 넣지 말라', 'BBC 이메일 주소로 회원가입 하지 말라'라는 강력한 지침을 내리며 보수적인 태도를 견지했다. 이러한 보수적 접근은 보안 문제뿐 아니라, 기업 계정 사용 시 외부 서비스가 자동으로 생성하는 엔터프라이즈 캐시(enterprise cache)에 내부 정보가 기록될 위험을 차단하려는 조치였다. 기업 이메일로 챗GPT와 같은 외부 서비스에 로그인할 경우, 취재 정보나 내부 문서, 기획안 등 민감한 내용이 기업 단위의 캐시 또는 로그에 남을 수 있었다. 이에 BBC는 극히 제한된 테스트 환경에서 구체적으로 정의된 작업에만 AI 실험을 허용하는 방식을 택했다.

[그림 3] '뉴스룸의 AI 혁신: 가능성에서 실현으로' 세션에 참여한 패널들



▶ 왼쪽부터 로라 엘리스 BBC 기술 예측 책임자, 올리버 호킨스 파이낸셜타임스 데이터 사이언티스트, 티투스 플랫폼터 타미디어 시랩 선임 혁신프로젝트 매니저, 카렌 플리팅 리치 AI 딜리버리 디렉터

그러나 이러한 통제가 곧바로 기대한 효과를 내지는 못했다. 구성원들은 개인 계정을 통해 비공식적으로 실험을 이어갔고, 그 과정에서 자연발생적인 ‘핵심 사용자층’이 형성됐다. 금지 조치가 오히려 직원들의 호기심을 자극하며 BBC 특유의 ‘즐겁게 전복하는(subversive)’ 문화를 만들어내기도 했다. 공식·비공식 경로를 가리지 않는 자발적 실험을 통해 BBC 내부에 빠르게 전문성이 축적됐다.

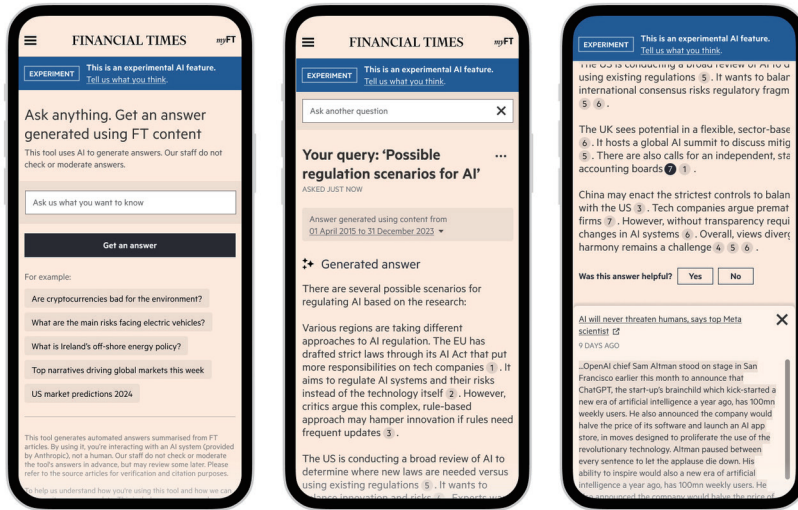
로라 엘리스 기술 예측 책임자는 조직이 매우 제한적이고 때로는 어수선했던 방식으로 대응했음에도, 이러한 보수적 접근이 결과적으로 BBC가 안정적인 실험 기반을 마련하는데 중요한 역할을 했다고 평가한다. 동시에 그는 “회사 내부의 누구도 지금 벌어지는 일들의 속도를 온전히 따라잡고 있지는 못한다”라며 “인사·재무·정책·전략·기술·콘텐츠 제작 등 조직 전 부문의 지속적인 논의가 필수적”이라고 강조했다. 생성형 AI 이후 내부 협업뿐 아니라 여러 미디어 조직 간 정보 공유와 가이드라인 논의가 활발해졌다는 점도 BBC의 중요한 변화로 꼽았다.

파이낸셜타임스의 사례는 생성형 AI가 기존의 ‘데이터팀의 실험’을 넘어 뉴스룸 전반의 아젠다로 확장되는 과정을 보여준다. 파이낸셜타임스의 데이터 사이언티스트인 올리버 호킨스는 “우리 팀은 이전에도 GPT-2, GPT-3 등 모델을 실험해 왔지만, 챗GPT가 공개된 순간 상황이 급변했다”라며 “그동안은 전문가나 괴짜가 하는 일로 보였던 AI 실험이, 하루아침에 모든 기자의 관심사가 됐다”라고 말했다.

파이낸셜타임스는 이후 AI 활용을 내부용(Internal)과 독자용(User-facing)으로 구분해 위험도가 낮은 영역부터 체계화해 나갔다. 모든 기자가 챗GPT나 클로드를 사용할 수 있도록 허용하되, 기사 작성과 같은 저널리즘 생산 과정에는 반드시 편집 정책을 준수하고 사례별 승인 절차를 거치도록 명확한 원칙을 세웠다. 동시에 파이낸셜타임스는 생산성을 높이기 위한 내부 도구를 개발했다. 대표적인 사례는 통신사 기사를 벡터화해 기사 관심사에 자동 매칭해 주는 시스템으로, 기자는 이 시스템 덕에 자신의 취재 분야에 맞춘 ‘개인화된 뉴스 피드’를 받을 수 있게 됐다. 프리미엄 독자를 위한 RAG(검색 증강 생성, Retrieval-Augmented Generation) 기반 아카이브 검색 서비스도 론칭했다.

생성형 AI를 안전하게 활용하기 위한 기술적 조치도 병행됐다. 파이낸셜타임스는 오픈AI와 협약을 맺어 회사 계정으로 챗GPT에 로그인하더라도 입력 데이터가 저장·학습되지 않는 환경을 마련했다. 그럼에도 탐사보도처럼 민감한 데이터를 다루는 영역에서는 외부 서버로 정보를 보내는 행위 자체가 위험하다고 판단해, 라마(Llama), 큐원(Qwen) 등 로컬 모델을 활용하는 실험도 진행하고 있다.

[그림 4] 파이낸셜타임스의 RAG 기반 아카이브 검색 서비스 애스크FT(AskFT)



스위스 언론사 타미디어는 생성형 AI를 효과적으로 도입하기 위해 조직 문화를 재편했다. 타미디어는 10명 규모의 AI 전담팀을 꾸려 다양한 도구를 직접 개발하고 있다. 티투스 플래트너 타미디어 AI랩 선임 혁신프로젝트 매니저는 “핵심은 기술 개발이 아니라 AI의 조직적 도입(AI adoption)과 AI 리터러시(AI literacy)”라고 강조하며, “기자들이 AI를 두려워하지 않으면서도 AI 결과물에 비판적 거리를 유지하도록 돕는 것”이 타미디어의 핵심 과제라고 설명했다.

타미디어는 비교적 일찍부터 AI를 도입해 왔음에도 AI의 활용에 대해서는 신중한 태도를 유지해 왔다. 이 과정에서 엇갈린 반응이 나왔다. 일부 직원들은 “이것도 금지, 저것도 금지한다”며 불만을 표하거나, “어차피 우리는 AI를 쓰지 않을 것”이라며 회의적인 태도를 보이기도 했다. 더 많은 정보와 실험 기회를 요구하는 이들도 있었다. 이러한 상반된 반응 속에서도 타미디어는 점진적으로 도구를 개선하고, 내부 실험을 확대해 나갔다. 150개의 내부 교육 과정을 제공하며 AI 리터러시를 높이기 위한 노력도 병행했다. 개발·사용 경험이 늘어나면서 조직의 AI 역량도 함께 성장했다. 티투스 플래트너 선임 매니저는 “이제 우리가 만든 도구들은 시장에 나온 서비스와도 견줄 정도”라고 말한다.

타미디어는 오픈AI와 같은 외부 서비스에 전적으로 의존하기보다, LLM(대규모 언어 모델)과 API를 결합한 자체 챗봇을 구축해 직원들이 안전하게 실험할 수 있는 환경을 마련했다. 민감한 데이터를 외부로 넘기지 않기 위해 특정 상황에서는 별도의 안전한 라이

선스를 확보해 사용하기도 했다.

한편 ‘인간이 개입하지 않은 AI 산출물이 외부에 그대로 노출되는 일은 절대 없다’라는 이른바 ‘휴먼 인 더 루프(Human in the loop)’란 타미디어의 원칙은 주목할 만하다. 생성형 AI가 높은 효율성을 제공하더라도, 사실 검증과 판단에 대한 책임을 AI에 전적으로 맡길 수는 없다는 판단에서 나온 원칙이다. 이제 ‘휴먼 인 더 루프(Human in the loop)’는 일부 AI 자동 발행 콘텐츠를 인간이 감독하는 ‘휴먼 온 더 루프(Human on the loop)’, 프라이버시 보호를 위해 인간 개입을 최소화하는 ‘휴먼 아웃 오브 더 루프(Human out of the loop)’의 원칙으로 발전해 가고 있다. 타미디어는 AI가 변환한 결과물을 다시 다른 AI로 교차 검증하거나, 기자가 작성한 기사 역시 AI 기반 팩트체크 도구로 재확인하는 방식도 실험하고 있다.

영국 최대의 상업 퍼블리셔인 리치는 120여 개의 브랜드와 4,000명의 직원이 분산된 대규모 조직이다. 이 때문에 AI 도입 과정에서도 ‘규모’와 ‘위험 관리’라는 두 가지 축이 핵심 변수로 작용했다. 카렌 플리팅 리치 AI 딜리버리 디렉터는 리치의 방향성을 “AI가 기자의 삶을 더 낫게 만드는가”란 기준으로 설명했다. AI 도입의 출발점을 혁신 그 자체가 아니라 현장의 업무 환경 개선에 두고 있다는 의미다.

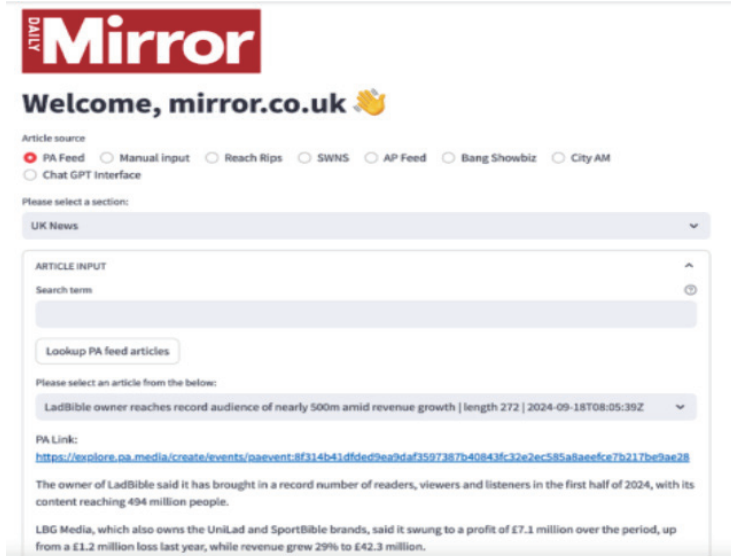
리치는 이미 오랫동안 다양한 부서에서 AI와 머신러닝을 폭넓게 활용해 왔다. 편집국은 소셜미디어 기반 속보 탐지 시스템인 ‘데이터 마이너(Data Miner)’를 활용해 왔고, 광고기술(AdTech) 영역에서는 머신러닝 기반 광고 플랫폼 ‘맨티스(Mantis)’를 운영해 왔다. 생성형 AI 확산 이후 리치가 직면한 가장 큰 과제는 기술 자체가 아니라 AI 리터러시와 조직적 도입이었다. 이를 위해 리치는 전 직원 의무 AI 교육, 정기적인 AI 뉴스레터 발행, 사내 활용 사례 공유 등 조직 전체를 아우르는 학습 구조를 마련했다.

위험 관리의 측면에서는 지적재산권(IP) 보호와 위험 완화를 위한 AI 거버넌스 체계를 구축했다. 새로운 AI 기술을 도입할 때는 법률, 사이버 보안, 데이터 보호 전문가가 참여하는 공식 거버넌스 그룹의 검토를 반드시 거쳐야 하며, 외부 AI 도구는 반드시 사전 승인을 받은 후에만 사용할 수 있다. 승인된 도구 목록은 사내 인트라넷을 통해 투명하게 공유된다.

내부 도구 개발도 적극적으로 진행되고 있다. 대표적인 사례는 기사 작성 도구인 구텀(Gutem)이다. 구텀은 다른 리치 계열 매체에서 생산된 기사를 각 매체의 스타일에 맞게 빠르게 고칠 수 있도록 지원한다. 또 앞서 타미디어에서도 강조된 ‘휴먼 인 더 루프’가 엄격히 적용되기 때문에 자동 생성된 콘텐츠가 편집자 검토 없이 외부로 게시되는 경우가

없도록 한다. 이 시스템을 기반으로 리치는 2023년 전체 콘텐츠의 25%를 AI가 일부 참여한 형태로 제작했으며, 이 콘텐츠들은 누적 18억 건의 조회수를 기록했다.

[그림 5] 리치의 AI 기반 기사 작성 도구 '구텀' 인터페이스 화면



▶ 타 매체의 기사를 불러와 데일리미러 스타일에 맞게 기사를 재작성하는 모습

생성형 AI의 등장은 단순한 기술 도입을 넘어 뉴스룸의 운영 방식과 조직 문화를 다시 설계하도록 만드는 강한 압력으로 작용하고 있다. BBC, 파이낸셜타임스, 타미디어, 리치 등 주요 언론사의 경험을 통해 확인되는 것은 공통적으로 세 가지다.

첫째, AI 도입의 출발점은 기술이 아니라 조직이라는 점이다. 각 조직은 보안·윤리·책임 구조를 재설계하고, 내부 합의를 형성하며, 때로는 ‘너무 보수적’이거나 ‘너무 개방적’인 접근 사이에서 균형을 찾아야 했다.

둘째, 작은 실험의 누적이 결국 조직 전체의 변화로 이어졌다는 점이다. BBC의 비공식 실험, 파이낸셜타임스의 생성형 AI 기반 검색 서비스, 리치의 구텀(Gutem) 등은 모두 거대한 시스템 변화가 아닌, 작고 구체적인 실험의 연속이었다.

셋째, 정확성과 팩트에 대한 책임 등 저널리즘의 핵심 가치는 AI 시대에 오히려 더 강화되어야 할 기준이라는 사실이다. 대부분의 언론사가 ‘휴먼 인 더 루프’를 전제로 기술을 적용하고 있었으며, 자동화가 확대될수록 편집 원칙과 위험 관리가 더욱 명확히 요구됐다.

결국 이들의 사례는 생성형 AI 이후의 뉴스룸은 기술이 아닌, 기술을 수용할 수 있는 조직과 문화, 그리고 책임 구조를 먼저 재건축한 곳부터 변화가 시작된다는 사실을 보여 준다. 앞으로의 뉴스룸 경쟁력은 ‘AI를 얼마나 활용하느냐’가 아니라 ‘AI를 어떤 철학과 구조 위에서 활용하느냐’에 의해 결정될 것이다.

뉴스룸을 바꾸는 작은 실험들

생성형 AI가 뉴스룸의 전략과 문화를 뒤흔들었다면, 그 변화를 실제로 완성한 것은 거창한 혁신이 아니라 작고 구체적인 실험들이었다.

파이낸셜타임스의 경우, 반복적으로 갱신되는 금융 시장 데이터에 주목한 것이 시작이었다. 하루에도 여러 차례 수치가 바뀌는 탓에 금융 시장 브리핑 기사는 업데이트 지연이 잦았고, 발행 시점에 이미 ‘구식’이 돼버리는 문제가 빈번하게 발생했다. 이에 파이낸셜타임스는 외부 업체와 협력해 자연어 생성(NLG, Natural Language Generation) 파이프라인을 시험했고, 기자가 업무 메신저인 슬랙(Slack) 안에서 금융 데이터만으로 자동으로 기사 초안을 불러오는 구조까지 만들었다.

하지만 이 파일럿이 보여준 핵심 문제는 AI 성능이 아니라 데이터 인프라의 취약성이었다. 파이낸셜타임스가 필요로 하는 실시간 시장 데이터를 완전히 제공하는 외부 API는 존재하지 않았고, 내부적으로도 데이터를 지속적으로 수집해 정제·적재할 체계가 없었다. 기자 경험(UX)도 문제였다. 도구가 슬랙 안에서만 작동하다 보니 기자들은 실제 기사 작성 화면과 슬랙을 계속 왔다 갔다 해야 했다. 기자들은 ‘업무 흐름이 끊긴다’라는 불평을 토로했다.

금융 기사 작성 파일럿 프로그램 개발에서 대두된 문제들은 이후 파이낸셜타임스의 AI 전략을 결정짓는 전환점이 됐다. 파이낸셜타임스는 ‘AI 기능을 덧붙이기 전에, 그 기능이 작동할 수 있는 데이터·워크플로 인프라를 먼저 재건해야 한다’라는 결론을 내렸다.

조나단 스트레이(Jonathan Stray) 콜롬비아대 저널리즘대학원 선임 연구원은 ‘탐사보도

를 위한 인공지능의 활용(Making Artificial Intelligence Work for Investigative Journalism)²⁾이라는 논문에서, 데이터 확보·정제·저장·검증 등 과정에 들어가는 시간과 비용이 지나치게 크기 때문에 AI나 머신러닝 기술을 저널리즘에 도입해도 경제성이 쉽게 확보되지 않는다고 지적한다. 그는 “단순히 모델을 돌리는 것보다, 데이터를 다룰 수 있는 인프라와 전처리 역량이 먼저 갖춰져야 한다”고 강조한다.

파이낸셜타임스가 맞닥뜨린 상황이 바로 이 문제였다. 파이낸셜타임스가 다루는 대부분의 데이터는 형식이 제각각이어서 추가적인 전처리 없이는 활용할 수 없었다. 보통 일반 기업의 과업은 보통 한 번 구축한 데이터 파이프라인을 오랫동안 반복 사용하므로 비용을 장기적으로 분산시킬 수 있다. 하지만 뉴스룸은 달랐다. 매번 기사 주제와 출처가 달라지기 때문에, 기사별로 새로운 데이터 파이프라인을 일일이 만들거나 바로잡아야 하는 구조적 제약이 있었다. 결국 파이낸셜타임스는 AI 자동화 시스템을 도입하면 할수록, ‘기사 하나를 만들 때마다 새로운 데이터 처리 비용이 반복해서 발생된다’라는 현실을 인정할 수밖에 없었다. 그리고 이 문제를 해결하려면 AI 도구를 도입하는 것보다 먼저 데이터 인프라를 재설계하는 일이 필요하다는 결론에 도달했다.

이 문제를 해결하기 위해 파이낸셜타임스는 내부 조직을 적극적으로 재정비했다. 고객 데이터, 콘텐츠 아카이브, 내부 API를 관리하는 데이터팀을 기반으로 삼고, 여기에 새롭게 ‘AI가속팀(AI Accelerator Team)’을 신설했다.

내부 진단 결과, 편집자와 기자가 AI 도구에 기대하는 기능이 다르다는 점도 명확히 드러났다. 편집자는 내부 데이터에 프롬프트를 적용해 사실 검증, 제목 제안, 소셜 게시물 생성 등 기존 워크플로를 효율화하는 기능을 원했다. 반면 기자들은 규제 공시, 기업 기록, 선거 자금, 정치 후원금 등 외부 데이터를 탐색하는 기능을 원했다. 즉 ‘건초더미에서 바늘 찾는’ 일을 도와줄 시스템을 요구했다.

이 요구를 반영해 파이낸셜타임스는 팀을 두 축으로 분리했다. 편집자용 도구는 내부 CMS인 스파크(Spark)에 적용했다. 예를 들어 이미지를 업로드하면 자동으로 대체 텍스트(alt text) 초안을 생성해 주는 기능이 스파크에 장착됐다.

기자들의 요구사항을 실현하는 일은 훨씬 복잡했다. 당시 파이낸셜타임스가 확보한 유일한 비정형 외부 데이터는 통신사 기사였고, 이를 벡터 임베딩해 개인화된 뉴스 피드

2) Stray, J. (2019), Making Artificial Intelligence Work for Investigative Journalism, Digital Journalism, pp.1076-1097, <https://doi.org/10.1080/21670811.2019.1630289>

를 만드는 실험이 첫 시도였다. 그러나 이 방식은 이미 공개된 기사만 재배열하는 방식이어서 특종 기사를 만들기에는 한계가 있었다. 이 문제를 해결하기 위해 파이낸셜타임스는 외부 데이터 소스를 지속적으로 확보하고 정제하는 전담 조직이 필요하다고 판단했다. 그 결과 AI가속팀은 상설 조직인 스토리파인딩팀(Story Finding Team)으로 전환됐다.

[그림 6] 영국 의원의 정치 후원 데이터 분석 화면

member_name	donor_status	name_of_donor	SUM of ft_amount
Siobhain McDonagh	individual	Waheed Alli	£1,200,000.00
	individual Total		£1,200,000.00
Siobhain McDonagh Total			£1,200,000.00
Jeremy Corbyn	company, registration 12956185 Total		£292,871.78
	company, registration Ltd company, registration 1		£2,916.67
	company, registration NI029999 Total		£319.98
Jeremy Corbyn Total			£296,108.43
Keir Starmer	individual	Lord Waheed Alli	£22,922.28
		Lord Waheed Alli	£16,200.00
	individual Total		£39,122.28
	company, reg The Football Association Premier League		£4,000.00
	Premier League		£3,000.00
	company, registration 02719699 Total		£7,000.00
	company, reg Crownhawk Properties (trading as ti		£4,500.00
	company, registration 03830322 Total		£4,500.00
	company, reg Tottenham Hotspur Ltd		£2,500.00
	company, registration 01706358 Total		£2,500.00
	company, reg Cain International UK Services Ltd		£2,400.00
	company, registration 10929931 Total		£2,400.00
	company, reg Crystal Palace Football Club		£2,142.00

▶ 파이낸셜타임스는 영국 의원의 비정형 정치 후원 데이터를 자동으로 추출·정제하는 데이터 파이프라인을 만들었다. 이를 통해 특정 의원과 개인·기업 간의 관계망을 한눈에 파악할 수 있다.

[그림 7] 영국 의원의 정치 후원 데이터 분석 결과로 생산된 파이낸셜타임스의 기사



▶ 파이낸셜타임스는 정치 후원 데이터 분석 결과를 활용해 노동당 의원이 제안한 임대료 개혁안에 숨은 잠재적 이해 상충, 유나이티(Unite)의 정치 후원 흐름 변화 등을 발굴해 보도했다.

스토리파인딩팀은 데이터 엔지니어·데이터 과학자로 구성돼 있다. 외부 데이터 수집·정제·저장·검색을 담당하는 지속적 데이터 인프라를 구축하는 것이 핵심 역할이다. 동시에 컴퓨테이션저널리즘팀(Computational Journalism Team)은 이 데이터 인프라를 실제 기사에 적용해 파이낸셜타임스만이 쓸 수 있는 차별화된 데이터 저널리즘을 구현한다. 두 팀의 결합은 다른 언론사가 쉽게 따라올 수 없는 데이터 기반 보도 역량을 만들어냈다. 작은 파일럿으로 시작된 금융 기사 자동화 실험은 파이낸셜타임스의 핵심 경쟁력을 강화하는 구조적 변화로 이어졌다.

데이터 인프라를 기반으로 파이낸셜타임스는 보다 정교한 AI 실험을 진행하고 있다. 규제 문서와 정치 후원 데이터를 RAG 기반으로 탐색하고, 미 도로교통안전국(NHTSA)에 접수된 수만 건의 소비자 불만 데이터를 LLM으로 분류한다. 솔라나(Solana) 등 블록체인 네트워크에서 암호화폐 거래 데이터를 직접 내려받아 분석하기도 하며, 주요 이벤트가 발생할 때 실시간으로 업데이트되는 라이브 블로그의 자동 게시 기능도 API 기반으로 구현했다.

물론 파이낸셜타임스는 생성형 AI를 활용하더라도, 모델이 만든 문장이 곧바로 독자에게 노출되는 것을 철저히 금지한다. 요약, 초안 작성, 데이터 분류 등 AI가 수행하는 모든 과정 뒤에는 반드시 인간 편집자의 검증이 따라붙는 ‘게이트키퍼 원칙’이 일관되게 유지된다. 파이낸셜타임스가 “독자가 궁금해하는 것은 우리가 AI를 썼는지가 아니라, 우리의 보도가 독점적인 통찰을 제공하는가다”라고 강조하는 이유도 바로 여기에 있다.

파이낸셜타임스처럼 대규모 언론사가 인프라를 재설계하며 AI를 끌어안는 방식이 있는가 하면, 훨씬 작은 조직이 ‘작은 문제’ 하나에서 출발해 의미 있는 변화를 끌어낸 사례도 있다. 규모나 자원은 다르지만, 핵심은 동일했다. 명확하게 정의된 문제를 중심으로 한정된 범위에서 AI를 실험하고, 점진적으로 확장해 나가는 접근이 효과적이라는 점이다. APA의 인포그래픽 대체 텍스트(alt text) 자동 생성 프로젝트는 이러한 전략이 소규모 뉴스룸에서도 어떻게 작동할 수 있는지를 보여주는 좋은 사례다.

APA는 오스트리아의 국가 통신사이지만, 전 세계에 수천 명 이상이 일하면서 글로벌 네트워크를 운영하는 AP나 로이터 등과 비교하면 직원 수는 500명 대로 규모가 크지 않다. APA는 오스트리아에서 ‘접근성법(Barrierefreiheitsgesetz)’이 2025년 시행되면서, 인포그래픽에 붙는 대체 텍스트를 단기간 내에 확보해야 하는 과제가 생겼다. 여기서 말하는 ‘접근성’은 장애 유무와 관계없이 모든 사용자가 동일하게 뉴스 콘텐츠를 이용할 수 있도록 보장하는 개념이다. 특히 시각장애인은 화면 읽기 프로그램을 통해 텍스트 정보를 듣

는 방식으로 콘텐츠에 접근하는데, 이때 이미지나 그래픽 요소는 그대로 두면 내용을 이해할 수 없다. 따라서 이미지·차트·인포그래픽 등 시각 요소에는 그 의미와 내용을 글로 풀어 설명한 대체 텍스트가 반드시 필요하다.

문제는 APA가 연간 2,000개가 넘는 인포그래픽을 제작하고 있다는 점이었다. 모든 인포그래픽에 대한 고품질 대체 텍스트를 사람이 직접 작성하려면 매년 약 7~8주의 작업 시간이 필요했다. 제한된 인력과 예산을 고려하면 현실적으로 감당하기 어려운 규모였다. 이 때문에 ‘인포그래픽 대체 텍스트 자동화’를 실현할 방법으로 AI 실험을 택했다.

APA는 우선 프로젝트 범위를 좁고 명확하게 설정하는 것부터 시작했다. 뉴스 제작 공정을 바꾸는 것이 아니라, 인포그래픽이라는 제한된 영역에서 AI가 실질적으로 어떻게 기여할 수 있는지 검증하는 데 집중했다. 이를 위해 데이터팁과 그래픽팁이 먼저 대체 텍스트가 갖춰야 할 최소 요건을 정교하게 정의했다. 모든 대체 텍스트 초안에는 그래픽 유형(막대그래프, 지도, 아이콘 기반 그래픽 등), 무엇을 설명하는지에 대한 핵심 문장, 데이터가 보여주는 주요 흐름, 출처 표기 등 네 요소가 반드시 포함되도록 프롬프트 구조를 고정했다.

이미지 파일의 형태도 고려해야 했다. APA는 PNG·JPEG 기반의 정적 인포그래픽을 많이 사용하기 때문에, 이 형식을 안정적으로 읽고 묘사할 수 있는 모델을 선택해야 했다. 내부 테스트 끝에 애저(Azure) 기반 GPT 모델이 선택됐다.

프로토타입이 완성된 뒤에는 실제 인포그래픽 수백 건을 대상으로 정확성, 명료성, 문체 조정, 사실 오류 발생 빈도 등을 평가하는 검증 과정이 진행됐다. 그 결과, 생성된 대체 텍스트의 약 3분의 2는 별도의 수정 없이 바로 사용할 수 있을 만큼 안정적인 품질을 보였다. 오류율도 APA가 진행해 온 다른 AI 프로젝트에 비해 현저히 낮았다.

프로젝트는 2025년 4월 1일에 시작해 6월 28일까지 시스템을 가동해야 하는 촉박한 일정 속에서 진행됐다. 법 시행일을 맞추기 위한 불가피한 일정이었다. 그럼에도 APA는 기한 안에 자동 생성 시스템을 구축해 인포그래픽 제작 파이프라인에 통합하는 데 성공했다. 현재 APA가 배포하는 모든 인포그래픽에는 자동 생성된 대체 텍스트가 기본적으로 포함돼 있으며, 최종 단계에서는 반드시 사람이 검수하는 ‘휴먼 인 컨트롤(human-in-control)’ 원칙이 적용된다.

이 과정에서 APA가 얻은 중요한 교훈은 두 가지였다. 첫째, 소규모 조직일수록 프로젝트 범위를 명확히 좁혀야 성공 가능성이 높다는 점이다. 실제로 APA 내부에서는 사진 서비스까지 자동화 범위를 확대하자는 논의가 있었지만, 인포그래픽부터 시작하는 것이 현실적이라는 판단이 프로젝트의 완주를 가능하게 했다.

둘째, 접근성 개선은 단순히 장애인을 위한 ‘부가 기능’이 아니라, 콘텐츠 품질을 근본적으로 끌어올리는 과정이라는 인식이 자리 잡았다. 실제 테스트 과정에서 인포그래픽의 정보 구조가 지나치게 복잡해 AI가 해당 인포그래픽을 제대로 설명하지 못하는 사례가 발견됐다. 이는 AI 실험이 기술 도입을 넘어, 인포그래픽의 설계 방식과 정보 전달 구조를 다시 검토하게 만드는 계기가 됐음을 보여준다.

APA의 사례는 작은 조직도 문제 범위를 정확히 정의하고 명확한 목표를 설정한다면, AI를 통해 뉴스 생산 과정의 일부를 효율화하고 조직의 역량을 강화할 수 있음을 입증한다. 대형 언론사든 소규모 뉴스룸이든 AI 도입의 핵심은 ‘거대한 혁신’이 아니라 ‘작게 시작해 빠르게 검증하는 실험 문화’라는 것이다.

한편 금융 정보에 특화된 블룸버그도 이미 10년 전부터 기사 자동 생성을 도입해, 그 경험을 기반으로 AI와 머신러닝을 폭넓게 활용하고 있다. 독자 수요가 많은 기업 실적 보도의 경우 2016년부터 기사 자동 생성을 도입했다. 현재는 AI를 더해 기업 실적 발표 자료에 전망, 해석 등과 같은 맥락까지 반영해 기사 완성도를 높였다. 이런 경우 공동 바이라인(double byline)을 써서 기사에 ‘자동화(automation)’가 사용됐다는 것을 알린다. 블룸버그는 실시간 금융 데이터와 국제 경매 보도를 넘어, 파생상품과 같은 복잡한 시장에서 나타나는 비정상적 움직임까지 포착하는 방향으로 기사 자동 생성의 범위를 확장하고 있다.

[그림 8] 블룸버그의 자동 생성된 기사 목록

Search News		Actions	Custom Searches	Translate	Key Themes	Page 1	News Feed
Bloomberg Auto Intelli...					Sources	All Dates	Time
Top Ranked News		More »		Top News		Background & Opinion	
1)	S&P 500 Index Falls 1.8%; Led by Communication Services Sector	BN	13:00				
2)	MSCI Nordic Index Falls 0.3%; H&M Drops	BN	12:15				
3)	Ukraine 1Q Tourism Revenue Falls 13%, Most Since March 2016	BN	08:29				
Time Ordered News							
4)	Square Social Media Volume Surges; Sentiment Somewhat Positive	BN	14:52				
5)	Liberty Gold Up 13%, Most in 11 Weeks as Canada S&P/TSX Declines	BN	14:44				
6)	Domtar Options Surge; 488,866-Share Block Trades	BN	14:40				
7)	Alphabet Shares Drop to Lowest in Five Weeks; Put Options Double	BN	14:36				
8)	Ranpak Holdings Class A Falls Most in 15 Weeks as S&P 500 Drops	BN	14:34				
9)	Brazil Curve Steady, CDS Spread Wider; Sovereign Update	BN	14:30				
10)	Alector Shares Down 12%, Most in More Than Three Months	BN	14:28				
11)	Office Depot Falls Most in Three Months; Put Options Quadruple	BN	14:25				
12)	Contango Oil & Gas Down 20%, Most in 15 Weeks; Options Quadruple	BN	14:24				
13)	Wynn Resorts Social Media Volume Surges; Sentiment Is Negative	BN	14:22				
14)	IAC/interactivecorp Options Surge; 350,000-Share Block Trades	BN	14:20				
15)	Riverview Bancorp Maintains Quarterly Dividend at 5 cents/Shr	BN	14:19				
16)	CyberOptics Shares Down 10%, Most in More Than 15 Weeks	BN	14:17				
17)	Pae Shares Down 9%, Most in 14 Weeks; Trading Volume Doubles	BN	14:16				
18)	Exxon Social Media Volume Quadruples; Sentiment Is Most Negative	BN	14:13				
19)	Pfizer Social Media Volume Quadruples	BN	14:11				

AI는 실시간 정보를 빠르게 확보하는 데 그치지 않고, 심층 분석을 통해 기사에서 부족한 맥락을 보완하고 서사를 풍부하게 만드는 역할까지 수행하고 있다. 이러한 기술은 반복적이고 시간이 많이 드는 작업을 자동으로 처리하는 데도 강점을 발휘한다. 예를 들어 코로나19 팬데믹과 같은 거대 이슈가 발생했을 때는 기사 자동 생산이 기자들의 데이터 정리 부담을 크게 줄여 효과적으로 활용됐다. 또한 시간적 여유가 없는 독자를 위해 핵심만 간추려 제공하는 요약 기능 역시 AI가 제공하는 중요한 기능이다.

주식 시장에서 상한가 또는 하한가를 기록한 종목을 감지하고, 해당 기업 관련 기사를 요약해 시장 움직임의 배경을 설명하는 것도 AI의 몫이다. 매일 아침 투자자들의 지침이 되는 모닝 뉴스레터 ‘데이브레이크(Daybreak)’ 역시 AI를 통해 여러 차례 자동 갱신돼 트레이더들이 밤사이 발생한 벌어진 상황을 빠르게 파악하고 하루를 준비할 수 있도록 돕는다. 기존에는 갱신 작업 때문에 데이브레이크 발행에 시간을 많이 잡아먹었지만, AI 덕에 이를 해결할 수 있었다.

블룸버그의 사례를 소개한 피터 스벤슨(Peter Svensson) 블룸버그 기자는 AI 활용의 목표를 “시장의 역동적인 움직임을 포착하고, 시장에 대한 통찰력을 독점적으로 발굴하며, 데이터 기반의 복잡한 계산 능력을 향상시키는 것”이라고 설명했다. 블룸버그는 이러한 관점에서 AI를 기사를 대체하는 기술이 아니라 금융 뉴스의 생산 방식을 정교하게 진화시키는 도구로 바라보고 있었다.

AI가 가져오는 생산성 향상에도 불구하고, 피터 스벤슨 기자는 ‘자동화의 확장과 편집 품질의 유지 사이의 균형’이 가장 중요한 과제라고 강조했다. 피터 스벤슨 기자는 그 균형을 맞추는 과정에서 마주한 현실적인 어려움 하나를 소개했다. 대규모 언어모델(LLM)이 생각보다 단순한 규칙도 제대로 지키지 못한다는 점이다.

블룸버그의 헤드라인은 반드시 64자 제한을 따라야 하지만, AI가 생성한 제목은 번번이 기준을 초과했다. 그는 “AI에게 64자를 정확히 지키게 만드는 일이 가장 어려운 과제 중 하나였다”라고 말하며 좌절감을 털어냈다.

이런 시행착오에도 블룸버그가 기사 품질을 유지할 수 있는 이유는 앞서 여러 매체들이 적용했던 ‘휴먼 인 더 루프’ 원칙을 엄격히 적용하기 때문이다. AI가 만든 초안은 반드시 편집자가 확인하고, 이 수정 내용은 다시 AI팀에 전달돼 모델의 동기화와 품질 개선에 활용된다. 자동화가 확대되는 환경에서도 최종 책임과 판단은 사람에게 있다는 원칙이 흐트러지지 않도록 운영하는 셈이다.

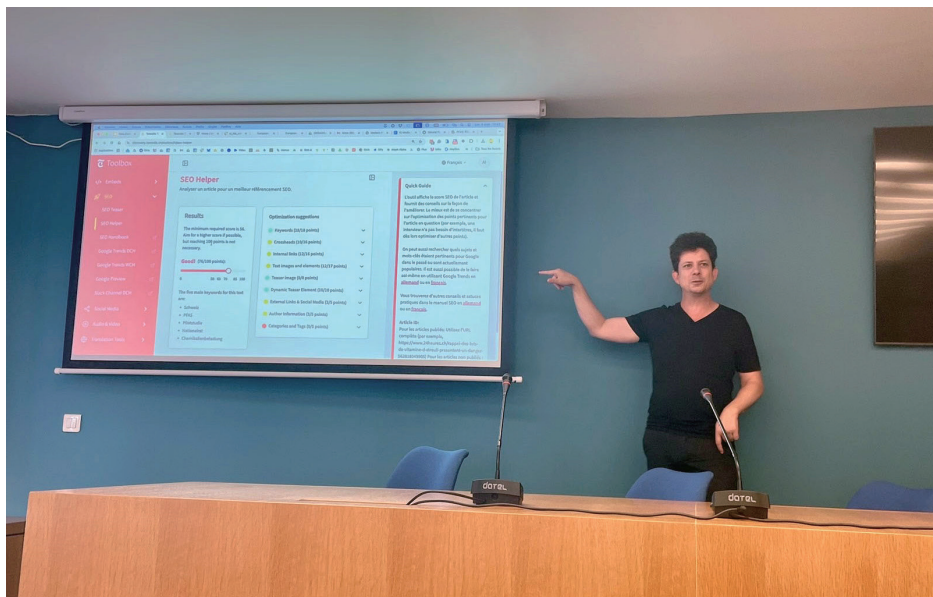
블룸버그의 사례는 AI 도입은 단순한 자동화의 문제가 아니라, 뉴스룸 고유의 리듬

과 전문성을 유지하기 위한 세밀한 조정 과정임을 보여준다. 기술이 강력해질수록 오히려 편집 기준과 검증 구조는 더 단단해져야 한다는 사실도 다시 확인된다. 그리고 이 지점은 또 다른 언론사가 직면한 고민과도 맞닿아 있다. 조직의 규모나 뉴스의 성격이 달라도, 생성형 AI를 어디까지 활용할 수 있으며 무엇을 인간의 판단으로 남겨야 하는지에 대한 문제는 누구에게나 공통의 과제로 다가오기 때문이다. 바로 다음 사례인 타미디어의 접근도 이러한 질문에서 출발한다.

타미디어는 2023년 11월 새로운 최고경영자(CEO)가 취임하면서 약 10명의 AI팀을 구성했다. 타미디어의 AI랩에 있는 티투스 플래트너 선임 매니저는 언론 환경에 AI를 도입하는 ‘문화 혁명’을 하는 중이라고 본인의 일을 소개했다. AI랩은 기자 교육을 위해 수백 개의 내부 강의 자료를 제공하고, 기자들이 반복적으로 활용하는 프롬프트(지시문)를 직접 제작해 검증한 뒤 필요할 경우 기사 작성 시스템(CMS)에까지 통합하는 일을 수행한다.

기자들은 CMS 내부에 마련된 일종의 ‘비밀 버튼’을 눌러 생성형 AI 클로드(Claude)로 문장을 보다 간결하게 다듬을 수 있다. 일부 기자는 이 기능을 3개월 동안 3,000회 이상 사용할 정도로 많이 사용하고 있다. 기사 길이를 사전에 정의된 형식인 4,150자 또는 5,800자로 자동 조정해 주는 도구도 개발 중이다.

[그림 9] DataJConf에서 강연하고 있는 티투스 플래트너 타미디어 AI랩 선임 혁신프로젝트 매니저



AI 도입으로 타미디어는 일부 업무에서 최소 30% 수준의 효율 향상을 확인했다. 특히 AI 프롬프트(지시문) 개발자를 포함한 프로그래밍 분야에서 효과가 뚜렷하게 나타났다. 다만 저널리즘 영역에서는 아직 가시적인 생산성 향상이 이뤄졌다고 보기는 어렵다는 평가다. 티투스 플래트너 선임 매니저는 “앞으로 일부 업무가 완전히 자동화되면서 생산성이 수십 퍼센트 더 높아질 것”이라고 전망하면서도, 조직 전체에 AI를 안정적으로 적용하는 과정에는 시간이 필요하다는 점을 강조했다.

타미디어가 AI 도입으로 가장 큰 혁신을 이룬 분야는 번역과 음성 전사였다. 스위스는 독일어와 프랑스어 등 공용어만 4개에 이르는 다언어 국가다. 음성으로 저장된 인터뷰를 텍스트로 옮기는 기술 자체는 새로운 것이 아니지만, 표준화된 문자 없이 구어로만 사용되는 스위스-독일어(Swiss German)의 경우에는 자동 전사가 사실상 불가능에 가까웠다. 티투스 플래트너 선임 매니저는 불과 2년 전까지만 해도 스위스-독일어를 표준 독일어로 전사하는 도구는 오류가 너무 많아 사용할 수가 없었다고 설명했다. 그런데 한 스위스 대학이 대규모 데이터셋을 공개하고, 얼마 지나지 않아 오픈AI가 이를 모델에 통합하면서 상황이 급변했다. 그는 “하루아침에 스위스-독일어 인터뷰를 고품질로 자동 전사할 수 있게 됐다”라며 “우리 녹음 자료 상당수가 이 언어로 돼 있다는 점을 고려하면 완전한 ‘게임 체인저’였다”라고 말했다.

타미디어는 앞 장에서도 언급했듯 AI 활용의 기본 원칙으로 ‘휴먼 인 더 루프’를 유지해 왔다. 그러나 최근에는 이 원칙을 한 단계 확장해, 필요할 때만 사람이 개입하는 ‘휴먼 온 더 루프(Human on the loop)’ 모델로 일부 업무를 전환하고 있다. ‘휴먼 인 더 루프’가 AI의 생성 과정마다 사람이 직접 판단에 참여하는 방식이라면, ‘휴먼 온 더 루프’는 시스템을 사람이 감시하되 정말 필요한 순간에만 개입하는 구조다.

타미디어는 리서치, 검색 결과 분류, 기사 재작성, 번역 등 다양한 공정에서 AI를 활용하고 있지만, 대부분의 기사 문장·표현은 여전히 사람의 검토를 거치는 체계를 유지하고 있다. 제목 생성이나 사진 설명 작성처럼 단순한 작업도 마지막 확인은 사람이 담당한다. 다만 매우 좁은 영역에서 결과의 신뢰도가 높은 경우에는 ‘휴먼 온 더 루프’로 전환하게 되는데, 이는 AI가 정보를 직접 게시하고 인간은 그 과정을 모니터링하는 방식이다.

스위스의 작은 지방 정부가 배포한 공식 보도자료를 예로 들면, AI가 이를 명확한 기사체로 자동 변환하고 문제가 없다고 판단되면 사람의 개입 없이 자동 게시한다. 이상 신호가 감지될 때는 기자에게 알림이 전달되고 사람이 최종 판단을 내린다. 이런 방식으로 타미디어는 하루 평균 80건의 지방 자치 관련 기사를 자동 생산하고 있다.

한편 타미디어는 이보다 더 제한된 영역에서 ‘휴먼 아웃 오브 더 루프(Human out of the loop)’ 모델도 적용한다. 이는 독자들이 직접 언론사 아카이브를 검색할 수 있도록 허용하는 경우 등에 적용된다. 언론사 기자들은 어떤 질문이 제기되는지조차 알 수 없으며, AI의 답변을 사전에 검토할 수도 없다.

티투스 플래트너 선임 매니저는 “AI 도입은 신중해야 하지만, 기다릴 여유는 없다”라고 강조했다. AI를 통해 뉴스 생산 방식을 바꿔내지 못하면 매체의 역할은 다른 주체가 대신할 것이라는 경고다. 그는 “AI에는 마법이 없다. 어떤 영역에서는 매우 잘 작동하고, 어떤 영역에서는 그렇지 않다”라며 AI의 한계를 투명하게 공유하고 지속적으로 실험하는 문화가 필요하다고 조언했다.

뉴스룸의 AI 리터러시

언론의 AI 리터러시는 회사 규모 또는 기자들의 세대에 따라 차이를 보였지만, 미국·유럽·아시아 등 지역 간 격차는 크지 않았다. 독일 루트비히-막시밀리안 뮌헨대(LMU)의 소르샤 세이버스(Sorsha Sabus) 연구원은 영국 기자 1,004명을 대상으로 실시한 저널리스트들의 AI 인식에 대한 조사 결과를 DataJConf에서 발표했다.

조사 결과에 따르면 AI에 대한 지식 자체보다, AI 도구를 실제 뉴스룸 업무에 적용해 본 경험이 AI를 기회로 인식하게 만드는 가장 강력한 요인으로 나타났다. 이론적 이해보다 실무 경험이 훨씬 큰 영향을 미친 것이다.

흥미로운 점은 기자 개인의 성향이나 개방성은 그리 중요한 변수가 아니라는 점이었다. AI에 대한 지식과 실무 경험이 있을수록 AI를 긍정적으로 바라보는 경향이 뚜렷했지만, 그렇다고 해서 AI 위협 인식이 자동으로 줄어들지는 않았다. 다만 AI를 실제 업무에 통합해 본 경험은 위협 인식을 완화하는 효과를 보였다.

또한 기자가 AI를 도입하는 과정에서는 ‘AI는 새로운 기회를 제공한다’는 인식이 결정적 요인으로 작용했으며, AI를 위협으로 보는 감정은 실제 도입을 가로막는 방해 요인이 되지 않았다. 조사에서는 세대 차이도 일부 발견돼, 더 젊은 기자들이 AI를 수용하는데 상대적으로 적극적인 것으로 분석됐다.

행사 셋째 날에는 ‘뉴스룸의 AI 리터러시’ 세션이 열렸다. 이 세션은 영국 서리대의 바하레 헤라비 교수가 진행을 맡았으며, AI 시대에 뉴스룸 구성원들이 어떤 역량을 갖춰야 하는지, 교육과 조직문화는 어떻게 달라져야 하는지를 중심으로 논의가 이뤄졌다. 패널로는 카렌 플리팅 리치 AI 딜리버리 디렉터, 로라 엘리스 BBC 기술 예측 책임자, 티투스 플래트너 타미디어 AI랩 선임 매니저, 캐서린 소티라코우(Catherine Sotirakou) IQ미디어

[표 1] 저널리스트들의 AI에 대한 인식

긍정적 인식	부정적 인식
- 저널리즘 업무의 효율성 증진 - 지루하고 반복적인 작업 자동화 및 감소 - 다국어 콘텐츠 제작 - 일상적인 작업 간소화 - 심층 보도를 위한 시간 확보 - 지루하고 반복적인 작업 자동화 및 감소	- 일자리 대체 - 창의성 상실 - 알고리즘 편향 - 투명성 부족 또는 불투명함 - 공정성 및 책임 등 윤리적 문제 - 콘텐츠가 정형화되고 비판적 시각이 부족해짐 - 직업의 의미 상실

* 출처: 독일 루트비히-막시밀리안 뮌헨대(LMU)

어 디렉터가 참여했다. 토론자들은 언론에서 이뤄지는 AI 교육이 단순한 기술 습득을 넘어, ‘AI를 책임 있게 활용하는 문화’와 ‘AI와 비판적으로 거리를 두고 판단하는 태도’를 길러내는 데 목적이 있다고 한목소리를 냈다.

로라 엘리스 BBC 기술 예측 책임자는 “AI 교육은 선택이 아니라 조직적 책임”이라는 원칙 아래, BBC가 가치 중심의 AI 리터러시 체계를 구축해 왔다고 설명했다. BBC는 모든 직원이 반드시 이수해야 하는 기초 교육 과정을 두고 있으며, 여기에는 ‘AI로 무엇을 할 수 있고, 무엇을 하지 말아야 하는가’에 대한 윤리·투명성 지침이 포함된다. 이어 실제 업무에서 코파일럿 도구를 어떻게 사용할지, 어떤 활용 방식이 안전한지 등 실무 중심의 단계별 교육이 뒤따른다.

BBC는 직원들이 실제 업무 맥락에서 AI를 경험할 수 있도록 돕는 일종의 슈퍼 사용자(super user)인 ‘공식 AI 리더’를 각 부문에 두고 있다. 이들은 현장에서 발생하는 질문과 문제에 즉시 대응할 뿐 아니라, 새로운 활용 아이디어를 검토하고 위험 요소를 점검하며 도구의 사용 범위를 설정하는 등 BBC의 AI 도입을 실질적으로 이끄는 핵심 허브 역할을 수행한다.

BBC의 AI 교육은 일방적이지 않다는 것이 특징인데, ‘블루룸(Blue Room)’이라는 참여형 공간이 대표적 사례다. 블루룸은 각 사업 부문의 의견을 모으고 다시 현장에 환류하는 피드백 허브로 기능한다. 이곳에서는 중국발 AI 모델 ‘딥시크(DeepSeek)’와 같은 신기술 논의부터, ‘AI가 특정 직무를 대체할 가능성’처럼 조직 전반에 영향을 미치는 주제까지 폭넓게 다뤄진다. 블루룸에서는 기술팀과 편집팀 간의 경계가 모호해지는 새로운 협력 방식이 나타나기도 했다.

BBC의 초기 AI 활용은 문서 정리나 메모 작성 등 단순한 생산성 향상 업무에 집중돼

있었다. 그러나 현재는 요약, 번역, 전사 등 편집 및 뉴스 제작 지원 작업으로 활용 범위가 확장되고 있다. 특히 인력 부족으로 처리하지 못했던 폴란드어 기사 번역을 AI로 가능하게 만든 사례처럼, AI가 기존에 수행하지 못하던 업무를 보완하는 효과도 나타나고 있다.

리치도 BBC와 마찬가지로 ‘조직 전체의 기본 역량으로서 AI 리터러시’를 구축하는데 방점을 두고 있었다. 약 4,000명의 직원을 보유한 리치는 120개 브랜드를 운영하는 대형 미디어 그룹으로, 특히 편집·제작 인력이 가장 큰 비중을 차지한다. 리치의 AI 리터러시 전략은 크게 ‘도입에 따른 위험을 통제하는 체계’와 ‘혁신을 빠르게 수용하는 실무 문화’의 두 축을 균형 있게 구축하는 것에 초점을 맞추고 있다.

리치는 최고경영자부터 인쇄소 직원까지 전 구성원에게 매년 의무적으로 AI 교육과 시험 통과를 요구한다. 이는 신규 입사자의 기초 교육에도 포함된 필수 절차로, 모든 직원이 AI의 기본 개념·작동 원리·한계·위험 요소·안전한 활용 방식 등 핵심 내용을 공통의 언어로 이해하도록 만들기 위함이다. 이를 통해 ‘AI를 어떻게 활용할 것인가’에 대한 조직 차원의 판단 기준을 통일하는 동시에, 특정 팀이나 개인의 역량 편차가 조직 전체의 AI 도입 속도를 저해하지 않도록 한다. 빠르게 변하는 기술 환경을 반영하기 위해 교육 프로그램은 주기적으로 업데이트된다.

각 팀에는 최소 한 명의 ‘AI 챔피언(AI Champion)’이 배치된다. 이들은 새로운 도구와 업데이트 소식을 팀원들에게 전달하고, 파일럿 실험을 지원하며, 실제 업무 적용 시 발생하는 문제를 조정하는 역할을 한다. 또한 리치는 전사 뉴스레터를 통해 내·외부의 AI 활용 사례와 기술 동향을 정기적으로 공유하는데, 열람률이 매우 높은 정도로 현장 관심도도 크다. 인트라넷에는 ‘Do/Don’t 가이드’와 사용을 승인한 AI 도구를 정리한 ‘AI 카탈로그’가 공개돼 있어 누구나 쉽게 확인할 수 있다.

리치는 AI 활용에 강력한 승인 체계도 운영한다. 직원이 새로운 AI 도구나 유스케이스를 제안할 경우 ‘AI를 사용하고 싶습니다(I Want to Use AI)’ 양식을 제출해야 하며, 법무·데이터 보호·보안 등 전문가로 구성된 거버넌스 그룹에서 이를 검토한다. 초기에는 편집자 승인 과정이 필요했지만, 활용 요청이 급증하면서 현재는 최고디지털책임자(CDO)가 중앙에서 일괄 승인하는 방식으로 전환됐다.

또한 리치는 직원들의 AI 활용 능력을 높이기 위해 게임화된 학습 방식도 도입했다. 예를 들어 특정 도구를 사용하려면 관련 교육을 먼저 이수해야 접근 권한이 열리는 방식이다. 이러한 구조는 단순히 교육 참여를 독려하는 것을 넘어, 안전하고 책임 있는 AI 사용을 조직문화로 정착시키는 장치로 기능하고 있다.

타미디어는 약 1,300명의 임직원 중 600명이 기자로 일하는 비교적 큰 규모의 언론사다. 티투스 플래트너 타미디어 AI랩 선임 매니저는 타미디어의 AI 리터러시 전략을 ‘조직 문화 전환’이라는 관점에서 설명했다. AI 교육이 단순한 기능 학습을 넘어, 뉴스 조직 전체가 새로운 업무처리 방식을 받아들이도록 만드는 문화적 작업이라는 것이다.

타미디어 역시 의무 교육과 AI 챔피언 제도를 운영하고 있다. 플래트너 선임 매니저는 “뉴스룸의 습관을 바꾸는 데 가장 큰 장애물은 시간”이라고 강조했다. 기자들은 늘 마감 압박에 시달리기 때문에 교육을 건너뛰 ‘그럴듯한 이유’가 항상 존재한다는 것이다. 이 때문에 경영진의 강한 의지가 무엇보다 중요하다고 했다. 타미디어는 2024년 1월 최고경영진과 2일간 집중 워크숍을 진행했고, 이 과정을 통해 ‘AI를 중심에 두고 회사를 재설계해야 한다’라는 공감대를 형성했다.

타미디어가 직원 교육에 공을 들이는 이유는 단순히 AI 오용을 방지하거나 품질 리스크를 통제하기 위해서만은 아니다. 플래트너 선임 매니저는 “언론 산업은 앞으로 인력이 줄어들 수밖에 없다”라고 지적하면서, 구성원들이 미래 노동시장에 맞는 역량을 갖추도록 돕는 것은 조직의 책임이라고 강조했다. 즉 AI 리터러시는 ‘생존을 위한 업스킬(up-skill)’이며, 회사가 직원에게 제공해야 할 최소한의 안전망이라는 말이다.

국경·언어·조직을 넘어: 데이터 기반 협업 저널리즘

2016년, 전 세계 언론사는 하나의 거대한 조세 은닉 지도를 마주하게 됐다. ‘파나마 페이퍼스(Panama Papers)’라는 이름으로 기록된 이 프로젝트는 80여 개국, 100여 개 언론사가 동시에 참여한 역사상 최대 규모의 국제 탐사 협업이었다. 한국에서는 뉴스타파가 일원으로 참여했다. 이 협업은 국경·언어·조직을 뛰어넘는 협력 저널리즘이 어떤 규모의 영향력을 만들어낼 수 있는지를 전 세계에 각인시켰다. 거대 권력에 맞서는 데 필요한 탐사 역량은 어느 한 언론사의 힘만으로는 충분하지 않으며, 함께할 때 언론의 존재 이유는 더 명확해진다는 사실을 이 프로젝트가 증명했다.

파나마 페이퍼스가 보여준 협업 저널리즘은 ‘언론 간 협력’이라는 1세대 모델이다. 지금 뉴스룸을 둘러싼 환경은 그때보다 훨씬 복잡하다. 오늘날의 문제는 더욱 구조적이고, 데이터는 기하급수적으로 방대해졌고, 기술은 언론의 업무 방식을 근본적으로 재편하고 있다. 이런 환경에서는 더 이상 언론만의 협력으로 저널리즘의 역할을 온전히 수행하기 어렵다. 시민 단체, 지역 커뮤니티 조직, 정책 연구기관, 연구자, 기술자, 법률가, 그리고 문제의 현장을 몸으로 겪는 시민들까지, 뉴스룸 바깥의 다양한 주체와 함께 움직이지 않으면 해결할 수 없는 문제들이 늘어나고 있다.

DataJConf가 주목한 것도 바로 이 지점이었다. 이 컨퍼런스에서 던진 또 다른 핵심 질문은 ‘언론이 혼자서는 다룰 수 없는 문제를, 각 사회 주체가 어떻게 협력과 거버넌스를 통해 함께 해결할 수 있는가’였다. 이 ‘협력 모델’을 선명하게 보여준 연사가 루시 ब्ल록(Lucy Block) 기후·커뮤니티 연구소(CCI, Climate & Community Institute) 데이터 커뮤니케이션 매니저였다. 그는 “시민사회 조직이 가진 가장 중요한 자산은 독점적 데이터, 문제에 대한 깊은 맥락적 이해, 그리고 현장과 연결된 네트워크”라며, “이는 뉴스룸 내부의 역량

만으로는 확보하기 어려운 자원”이라고 강조했다.

코로나19 초기, 루시 블록 매니저가 살던 뉴욕 퀸즈 엘머스트 지역은 미국 내에서도 가장 빠르고 심각하게 코로나19 피해가 집중된 곳이었다. 지역 병원에는 환자가 넘쳐났고, 사망자가 급증하며 지역 전체가 사실상 의료·주거·생계 위기 상태에 놓여 있었다. 루시 블록 매니저는 당시 주거 관련 시민단체에서 데이터 분석가로 일하고 있었다. 그는 팬데믹 초기에 드러난 이 지역의 극심한 피해가 단순한 ‘우연한 집중’이 아니라 기존의 구조적 불평등과 깊이 연결돼 있다는 점을 데이터로 확인하고자 했다.

루시 블록 매니저는 뉴욕시 보건국이 깃허브(GitHub)에 공개한 확진·사망 데이터를 과밀 가구 비율, 서비스 노동자 비율, 임대료 부담, 유색인종 비율 등 시민단체가 축적해 온 지표와 결합해 지역별 위험도를 시각화했다. 팬데믹의 충격이 특정 지역과 특정 집단에 불균형적으로 집중되고 있다는 사실이 이 분석을 통해 선명하게 드러났다. 감염 위험이 높았던 바로 그 지역이 생계와 주거 안전망까지 동시에 취약한 지역들이었다.

팬데믹이 장기화하면서 수만 명의 세입자가 일자리를 잃고 임대료를 감당하기 어려워지자, ‘감염 취약성’은 곧 ‘퇴거 취약성’으로 연결됐다. 뉴욕주가 한동안 주택 법원을 폐쇄해 퇴거 절차가 잠시 중단됐지만, 법원이 재개되면 어떤 지역부터 대규모 퇴거가 발생할지 예측할 방법이 없었다. 이런 문제를 해결하기 위해 루시 블록 매니저는 시민단체와 함께 뉴욕주 최초의 ‘퇴거 위기 모니터(eviction crisis monitor)’를 만들었다. 팬데믹 피해와 주거 취약성을 결합해, 퇴거 위험 지역을 미리 파악할 수 있도록 설계한 도구였다.

루시 블록 매니저와 동료들의 분석은 곧 뉴욕타임스의 스테파노스 첸(Stefanos Chen) 기자의 취재 출발점이 됐다. 루시 팀은 가공되지 않은 초기 단계의 데이터까지 적극적으로 공유했고, 기자는 이를 바탕으로 충분한 검증과 추가 취재를 진행할 시간을 확보할 수 있었다. 그 결과, ‘코로나19 고위험 지역의 세입자는 가장 낮은 위험 지역보다 퇴거당할 가능성이 네 배 높다’라는 분석이 인터뷰와 현장 사례와 결합돼 기사로 완성될 수 있었다.

이 협업을 계기로 루시 블록 매니저는 1년 뒤 스테파노스 첸 기자와 다시 한번 맨해튼 외곽 자영업자 부채 문제를 다루는 기사에서도 함께하게 됐다. 그는 원래 몇 달 뒤에 진행할 예정이던 임대료 데이터 업데이트 일정을 앞당겨 제공할 만큼 기자를 신뢰했고, 기자 역시 이 데이터를 토대로 지역 상권의 구조적 어려움을 설득력 있게 보도할 수 있었다.

루시 블록 매니저는 “협업은 반복될수록 더 빠르고 더 깊어진다”고 말한다. 한 번 신뢰가 형성되면 데이터팀은 더 정확한 분석을, 기자는 더 정교한 취재를 할 수 있게 되고, 그 결과는 더 나은 보도로 이어진다. 이 사례는 시민사회와 언론이 장기적으로 협업할 때

데이터 기반 보도의 질이 어떻게 자연스럽게 높아지는지를 보여준다.

[그림 10] 루시 블록 매니저와 스테파노스 첸 뉴욕타임스 기자의 협업 기사

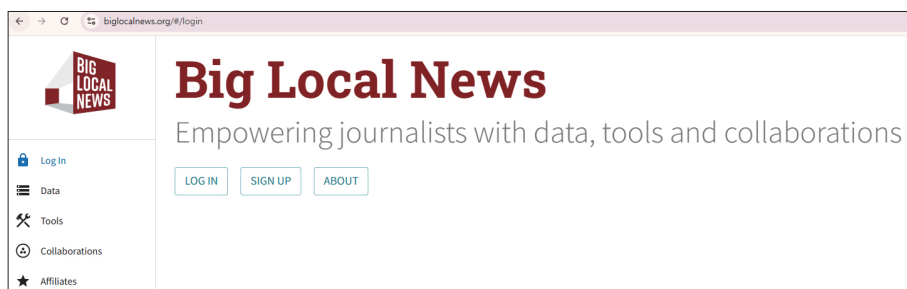


▶ 스테파노스 첸 뉴욕타임스 기자가 작성한 '코로나19 고위험 지역의 세입자는 가장 낮은 위험 지역보다 퇴거당할 가능성이 네 배 높다'라는 내용의 심층 기사

루시 블록 매니저의 발표가 시민사회와 뉴스룸이 함께 문제를 규정하고 해결하는 협업 모델을 보여줬다면, 셰릴 필립스(Cheryl Phillips) 스탠퍼드대 컴퓨테이션 정책연구소 교수의 발표는 그 협업 단위를 한층 확장해 '지역 뉴스 생태계를 위한 데이터 인프라 구축'이란 주제를 제시한다. 셰릴 필립스 교수는 기자 시절 수개월에서 1년 가까이 데이터를 수집하고 정제해 대형 기사를 내면서도, 그렇게 만들어진 데이터가 단 한 번의 보도 이후 다시 활용되지 못한 채 사라지는 현실을 반복적으로 목격했다고 회고했다. 이러한 문제의식이 그를 스탠퍼드로 이끌었고, 이후 지역 언론을 위한 데이터 공유·협업 모델을 고민하게 되는 출발점이 됐다.

필립스 교수가 구축한 '빅 로컬 뉴스(Big Local News)'는 바로 그 문제의식에서 출발한 데이터 공유 플랫폼이자 협업 인프라다. 특정 뉴스룸이 어렵게 만들어낸 데이터가 그 조직 내부에서 소모되고 마는 기존의 구조를 넘어, 지역 언론 전체가 함께 활용할 수 있도록 설계된 시스템이다.

[그림 11] 세릴 필립스 교수가 구축한 '빅 로컬 뉴스(Big Local News)' 플랫폼



필립스 교수가 강조한 핵심은 명확했다. 지역 언론이 겪는 위기의 본질은 ‘자원 부족’이며, 그 격차를 줄이는 가장 강력한 방법은 데이터를 공유하는 인프라를 구축하는 것이라는 점이다. 그는 “뉴스룸 하나가 만든 데이터는 그 뉴스룸에만 머무르지 않아야 한다”고 말하며, 경쟁적 기사 생산 체계 속에서도 협력할 수 있는 구조가 존재한다고 설명했다.

필립스 교수는 이를 위해 ‘연합(federated) 인프라’라는 개념을 제안한다. 이는 기자들이 서로의 데이터를 활용할 수 있는 발판을 제공하는 구조로, 데이터셋·분석 코드·보도 레시피·교육 자료 등 공익 자원을 공유하는 방식이다. 탐사보도 기술 장벽이 높아질수록, 개별 뉴스룸이 모든 역량을 갖추는 데는 한계가 있다. 그러나 연합 인프라 위에서는 자원이 부족한 소규모 지역 언론도 대형 언론사 수준의 데이터 기반 보도를 해낼 수 있다.

필립스 교수는 여러 사례를 통해 이 구조가 실제로 작동함을 보여줬다. 풀리처상을 받은 오피오이드 과다 복용 보도에서, 핵심적인 역할을 한 조직은 고작 네댓 명 규모의 작은 지역 뉴스룸이었다. 스탠퍼드 팀이 전국 데이터를 분석하고, 지역 기자들은 지역사회 네트워크와 현장 취재를 결합하며 입체적 기사를 완성했다. “작은 뉴스룸도 충분히 기여할 수 있다”는 그의 말은 선언이 아니라 증명된 사실이었다.

이러한 협업을 확장하기 위해, 필립스 교수는 제이 해밀턴(Jay Hamilton) 스탠퍼드대 교수와 함께 빅 로컬 뉴스 플랫폼 내에 ‘다트 매트릭스(DART Matrix)’를 개발했다. 다트 매트릭스는 데이터(Data), 알고리즘·도구(Algorithms and Tools), 보도 레시피(Reporting Recipes), 훈련·교육(Training)으로 구성된 지역 뉴스 지원 모델이다. 데이터 역량이 거의 없는 뉴스룸에는 ‘엑셀에서 정렬하는 법’부터 안내하는 레시피를 제공하고, 역량이 높은 뉴스룸에는 분석용 데이터 스트림만 제공하는 등 수준에 맞춘 지원이 가능하다. 이 모델을 적용한 학교 폐쇄, 경찰 불법행위 기록, 해고 통지 데이터 프로젝트 등은 소규모 언론사의 보도 역량을 실질적으로 끌어올린 사례다.

또한 그는 생성형 AI와 자동화 도구가 지역 저널리즘을 지원하는 새로운 기반이 될 수 있다고 전망했다. 필립스 교수팀이 빅 로컬 뉴스와 연계해 개발한 ‘아젠다 워치(Agenda Watch, 지역 정부 회의 안전 자동 크롤링 도구)’와 ‘데이터 토크(Data Talk, 데이터 기반 질의·팩트체크 도구)’는 그 대표적 사례다. 기자들은 복잡한 데이터 처리 과정을 직접 수행하지 않아도 되고, 인프라가 제공하는 데이터와 도구를 그대로 활용해 지역 보도에 집중할 수 있다.

그의 메시지는 단순했다. “지역 저널리즘을 되살리기 위해 필요한 것은 더 많은 기자가 아니라, 더 나은 인프라다.” 지역 데이터를 함께 구축하고, 함께 활용하고, 함께 확산시키는 시스템이야말로 지역 뉴스 생태계 회복의 실질적 해법이라는 것이다.

[그림 12] DataJConf에서 강연하고 있는 세릴 필립스 스탠퍼드대 교수



다음 연사인 안나 투리치(Anna Turic) 프리랜서 기자는 데이터 저널리즘이 국가·지역 단위를 넘어 ‘지정학적 내러티브’를 재구성하는 데 어떤 역할을 할 수 있는지를 보여주는 사례를 공유했다. 세르비아 출신 기자인 그는 중국 연구자, 데이터 분석 경험이 있는 동료 기자들과 함께 EU와 중국의 관계가 제3국에서 실제로 어떻게 작동하는지 추적하는 국경 간 협력 조사를 수행했다. 조사의 출발점은 발칸과 남캅카스 지역에서 반복되던 ‘중국의 영향력이 급증하고 있다’라는 정치적 내러티브였다. 팀은 공공조달 데이터, 국제개발은행(EBRD·EIB) 프로젝트 정보, 정부·중앙은행 통계 등 공식 데이터를 체계적으로 수

집·정제해 이 내러티브를 검증하기로 했다. 여러 국가의 불완전한 공개 시스템, 서로 다른 데이터 포맷, 투명성 격차 등으로 작업은 쉽지 않았지만, 국경 간 협업을 통해 데이터 기반의 분석 체계를 만들어냈다.

이 분석은 기존의 정치적 담론과는 전혀 다른 현실을 드러냈다. 발칸·남캅카스 국가들이 강조하던 ‘중국 의존 심화’ 서사와 달리, 실제 자금 흐름과 경제 지표를 전수 비교한 결과 이들 지역에서 EU가 압도적으로 재정적·경제적 영향력을 갖고 있는 것으로 확인됐다. 특히 EIB와 EBRD의 해외 프로젝트 데이터를 분석한 결과, EU가 제3국에 투자한 자금 가운데 약 13%가 중국 기업으로 유입되는 구조가 드러났다. 즉 국제 질서에서는 경쟁 구도를 유지하는 EU와 중국이, 제3국 개발사업에서는 공급망과 자금 구조 속에서 서로 얽혀가며 동시에 영향을 미치고 있었다. 무역·투자·대외부채 지표에서도 EU가 중국을 크게 앞서는 일관된 패턴이 나타나, 오랫동안 확산된 ‘중국 영향력 확대’라는 정치적 서사를 데이터로 다시 쓰는 근거가 됐다.

안나 투리치 팀의 작업은 단순한 ‘팩트체크’를 넘어, 데이터 접근성과 분석 역량이 국경 간 저널리즘 협업에서 얼마나 결정적인 역할을 하는지를 보여주는 사례이기도 했다. 세 나라 이상이 관여한 프로젝트였지만, 데이터 공개 수준과 포맷이 제각각이어서 팀은 초기에 ‘무엇을 기준으로 비교할 것인가’라는 질문부터 다시 설정해야 했다. 정부 부처에서는 ‘기밀’이라고 분류한 동일한 정보를 중앙은행에서는 제공하는 등 기관 간 불일치가 심했고, 공공조달 자료는 PDF로만 제공되거나 누락된 항목이 많아 수작업 정제가 필수적이었다. 안나 투리치 기자는 “팀 전체가 코드를 자유롭게 다룰 수 있었다면 훨씬 수월했을 것”이라고 말하며, 데이터 처리 역량의 불균형이 협업의 병목이 되는 순간들을 솔직하게 공유했다.

그럼에도 이 과정은 기자와 연구자가 국경과 전문 분야를 넘나들며 공동의 데이터 구조를 만드는 경험을 축적시켰고, 이는 오래 남을 협업 자산이 됐다. 안나 투리치 기자는 발표에서 “정치적 내러티브, 선전, 외교적 수사는 데이터 앞에서 비로소 실체가 드러났다”라며, 데이터 저널리즘이 국가 간 관계를 해석하는 새로운 기준선을 만들 수 있다고 강조했다. 이러한 접근은 단일 국가의 관점으로는 보이지 않는 지정학적 상호 의존 구조를 드러내고, 정책·외교 담론을 보다 현실에 근접한 근거 위에 올려놓는 데 기여한다는 점에서 큰 의미가 있다.

저널리즘이 단독으로 해결할 수 있는 문제의 범위는 갈수록 좁아지고 있으며, 데이터는 그 경계를 다시 설정하는 핵심 언어가 되고 있다. 팬데믹 당시 한 지역의 주거 위기부

터, 지역 뉴스 생태계의 구조적 붕괴, 국제 정치 담론의 왜곡에 이르기까지, 이들이 다룬 문제는 성격도 규모도 완전히 달랐지만 모두 협업 없이는 풀 수 없었다. 시민사회가 가진 현장 데이터, 대학과 연구기관이 구축한 인프라, 국경을 넘어 공유된 조사 구조가 결합될 때 오래된 문제는 비로소 새로운 설명을 얻었다.

DataJConf가 보여준 것은 바로 이 지점이었다. 저널리즘의 미래는 더 고립된 전문성이 아니라, 더 넓게 연결된 역량 위에서 가능하다는 사실이다. 데이터 기반 협업은 기자의 역할을 대체하는 것이 아니라, 오히려 더 넓은 맥락과 더 깊은 분석을 가능하게 하는 도구가 된다. 그리고 이러한 인프라와 네트워크는 특정 국가나 조직에 국한되지 않는다. 지역의 작은 뉴스룸도, 시민단체도, 국제 연구자도 함께 참여할 수 있는 개방적 구조가 마련될 때, 저널리즘은 다시 사회적 감시와 공공성이라는 본래의 임무를 수행할 수 있다. DataJ-Conf는 저널리즘이 더 이상 ‘혼자’가 아닌 시대의 시작을 보여주고 있었다.

AI 시대, 독자와 언론의 새로운 관계를 설계하라

닉 디아코폴로스 노스웨스턴대 교수는 ‘AI가 미래 정보 생태계에 어떤 영향을 미칠 것인가’를 탐색하기 위해 시나리오 기법을 활용했다. 그는 딥페이크 기술이 등장한 지 약 2년이 지난 2019년, 딥페이크가 선거 과정에 미칠 잠재적 영향을 분석하는 시나리오 8가지를 구성했다.

이 중 하나는, 존경받는 인물이 특정 개인의 이름을 직접 부르며 잘못된 투표소 정보를 전달하고 ‘부정선거가 있었다’고 주장하는 개인 맞춤형 딥페이크 영상이 유포되는 상황을 가정한 시나리오였다. 해당 시나리오는 2024년 미국 대선 때 뉴햄프셔에서 실제로 벌어진 사건과 상당히 비슷했다. 누군가 조 바이든(Joe Biden) 전 미국 대통령의 목소리를 딥페이크로 합성해 전화를 걸고, 유권자들에게 예비선거에서 투표하지 말라고 유도한 것이다.

그가 최근 진행한 AI 영향 예측 연구에서도 시나리오 기법이 활용됐다. 물론 시나리오 작성을 오직 전문가들에게 의존했기 때문에 생길 수 있는 ‘전문가 편향(expert bias)’의 한계가 있지만, 연구팀은 생성형 AI가 5년 뒤 뉴스 환경을 어떻게 바꿀지에 대한 시나리오를 100개 이상 도출했다. 이 분석에서 언론인과 AI의 관계는 대체로 긍정적으로 그려졌다. 반면 AI가 민주주의나 허위 정보에 미칠 영향은 전반적으로 부정적이거나 공·부정이 혼재됐다.

시나리오에서는 AI가 반복적인 작업을 자동화하고, 언론인이 절약된 시간을 더 중요한 저널리즘 활동에 재투자할 것이라는 예측이 매우 강력하게 나타났다. 저널리스트는 AI 시스템에 ‘확인된 사실(verified facts)’을 제공하고, 윤리적 감시자 역할을 하며, 편집 감독을 맡는 등 긍정적이고 필수적인 역할을 할 것이라고 봤다. 특히 언론인과 대중 간의 관

계는 100% 완벽하게 긍정적으로 나타났다. 시나리오에 참여한 전문가들은 언론인이 대중에게 신뢰를 제공하고 책임 있는 정보 전달자로서 긍정적 역할을 수행할 것이라고 확신하는 경향을 보였다.

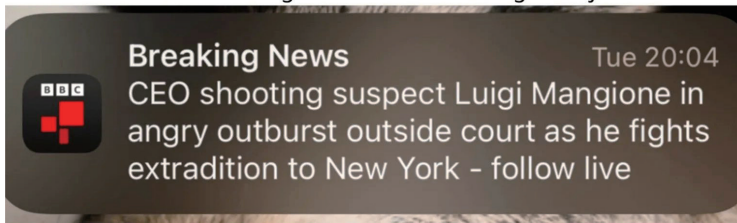
그러나 닉 디아코폴로스 교수는 저널리즘이 AI를 전적으로 신뢰해서는 안 된다고 못박았다. 번역이나 전사처럼 비교적 안정적으로 활용할 수 있는 영역이 존재하더라도, 최종 판단과 검증은 반드시 인간이 수행해야 한다는 점을 거듭 강조했다.

‘AI 쓰레기(AI slop)’가 저널리즘을 장악하는 최악의 시나리오가 제시되기도 했다. 닉 디아코폴로스 교수는 이미 프리랜서들의 가짜 필명으로 뉴스 매체에 ‘AI 쓰레기’가 게시되는 등 불량 정보가 유통되고 있다고 지적했다. 그는 “언론사가 오랫동안 쌓아온 신뢰와 신용도가 AI 쓰레기를 생산하는 시스템에 의해 급격하게 훼손되는 최악의 시나리오 중 일부가 시작되고 있다”라고 밝혔다.

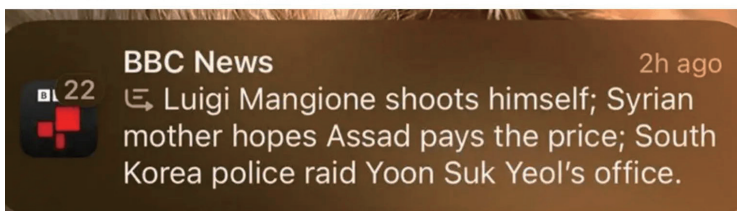
AI로 인해 언론사의 명성에 금이 간 사례는 쉽게 찾아볼 수 있다. 2024년 말, 애플의 AI 기능인 ‘애플 인텔리전스(Apple Intelligence)’가 BBC 뉴스 알림을 잘못 요약해 거짓 헤드라인을 생성하는 사건이 발생했다. 문제의 요약본에는 보험사 유나이티드헬스케어

[그림 13] BBC가 출처로 표기된 오정보를 전달한 ‘애플 인텔리전스’ 알림 화면

BBC News sends Breaking News alerts on a single subject



Apple sends an AI-generated summary of BBC News headlines - which can create inaccuracies



BBC

▪ BBC가 전송한 정상적인 속보 알림(위)과 애플 인텔리전스가 전송한 허위 속보 알림(아래). 애플 인텔리전스는 BBC의 속보 알림을 자동 요약하는 과정에서 루이지 맨지오니가 극단적 선택을 했다는 허위 정보를 만들어냈다.

CEO를 살해해 ‘영웅’ 취급을 받았던 루이지 맨지오니(Luigi Mangione)가 극단적 선택을 했다는 내용이 포함돼 있었는데, 이는 사실과 전혀 다른 정보였다. BBC는 즉시 애플에 항의했고, 국경없는기자회(RSF) 또한 “생성형 AI는 아직 대중에게 신뢰할 만한 정보를 제공하기에는 미성숙하다”라며 해당 기능의 중단을 촉구했다. 언론사 역시 ‘AI 쓰레기’로 인한 피해에서 자유롭지 않다는 사실을 드러낸 사례다.

한편 BBC는 간판 드라마 ‘닥터 후(Doctor Who)’ 홍보 뉴스레터 제작 과정에 AI를 사용했다가 시청자의 비난을 받았다. 아주 작은 문단에만 AI가 사용됐음에도, 독자들은 즉각적으로 어색하고 비인간적인 문체를 알아차렸다. 일부 시청자는 담당 프로듀서를 직접 찾아내 맹렬하게 비난했다. 시청자들은 “‘닥터 후’에서 AI를 봤는데, AI는 절대 좋은 방향으로 가지 않는다”라며 ‘불쾌한 골짜기’와 같은 근본적인 거부감을 드러냈다.

BBC는 ‘닥터 후 뉴스레터’ 사태에서 AI 사용의 투명성과 관련해 중요한 두 가지 교훈을 얻게 됐다.

첫째, 독자·시청자 집단의 특성에 따라 AI 활용의 민감도가 크게 달라질 수 있다는 점이다. 60년 넘게 방송된 공상과학 드라마 ‘닥터 후’의 팬층은 이른바 ‘후비안(Whovians)’으로 불릴 만큼 열정적이며 세계관·서사·세부 표현에 예민하게 반응하는 경향이 있다. BBC는 이러한 집단에는 아주 작은 AI 개입조차 강한 거부감을 촉발할 수 있음을 확인했다.

둘째, AI 활용 사실을 알릴 때는 ‘AI를 썼다’라는 선언만으로는 충분하지 않다는 점이다. 어떤 목적이었는지, 어느 문장에 어떤 방식으로 적용했는지, 인간 편집자의 역할은 무엇이었는지 등 사용 배경과 범위를 구체적으로 설명해야 한다는 것이다.

언론이 AI 시대에 직면한 또 다른 과제는 뉴스 회피 현상이다. 뉴스 회피는 단순한 관심 부족을 넘어, 사람들이 의도적으로 뉴스를 멀리하는 적극적 회피 행동을 의미한다. AI가 닥쳐오는 쓰나미라면 뉴스 회피는 이미 우리의 발을 적시고 있는 파도라고 할 수 있다.

니콜라오스 파나요투(Nikolaos Panagiotou) 아리스토텔레스대 교수는 “어떻게 하면 뉴스를 ‘섹시하게’ 만들 수 있을까요?”란 도발적 질문을 던졌다.

니콜라오스 파나요투 교수는 전 세계의 다양한 결과와 48개 시장에 걸쳐 직접 실시한 연구를 바탕으로, 현재 뉴스 회피 현상이 10명 중 4명꼴로 나타난다고 주장했다. 그리스는 미디어 신뢰도가 22%밖에 되지 않아 가장 낮은 국가 중 하나로 꼽히며, 뉴스 회피 비율은 유럽 평균보다 높은 10명 중 6명꼴에 이른다고 강조했다.

전국적 조사가 진행된 2025년 3~6월 사이, 그리스의 뉴스 회피 비율은 81%에 달해

[표 2] 뉴스 회피 연구 주요 결과

구분	내용	비고
전 세계적 뉴스 회피 현상	약 40%가 뉴스를 자주 또는 때때로 회피한다고 응답	2017년 29%에서 약 40%로 증가
뉴스 회피율이 높은 지역	불가리아, 튀르키예, 크로아티아, 그리스는 각각 60% 이상의 회피율 보여	그리스에서는 2025년 3~6월 사이 뉴스 회피율이 81%로 조사됨

출처: 니콜라스 파나이오투 아리스토텔레스대 교수

매우 높게 나타났다. 특히 과거 경제 위기 시기 미디어 산업이 거의 붕괴될 정도로 흔들렸던 영향이 여전히 깊게 남아 있음을 보여주는 결과다. 불가리아, 튀르키예, 크로아티아에서도 뉴스 회피 비율은 60% 이상으로 나타났다.

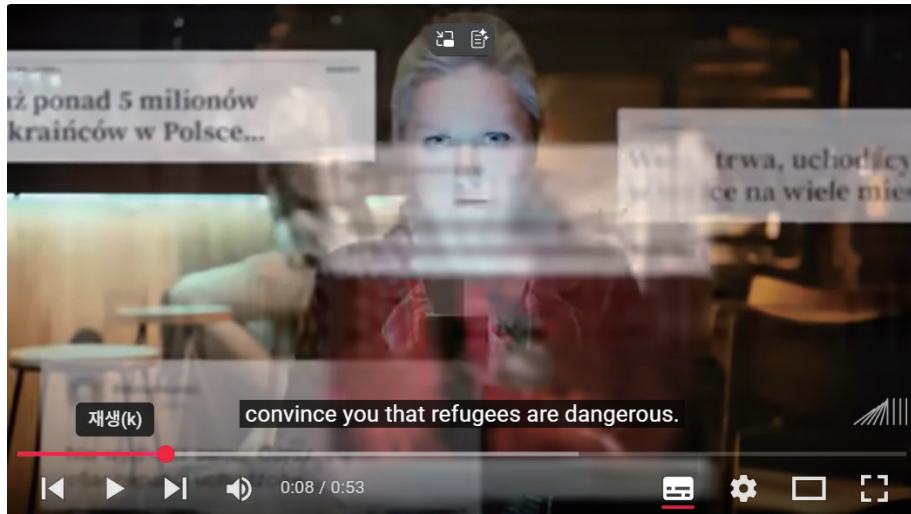
뉴스 회피 현상이 그리스에서 강하게 나타나는 이유는 시민들이 뉴스를 보면서 스트레스(61%), 분노(45%), 무력감(20%) 등 부정적인 감정을 경험했기 때문이다. 뉴스 회피를 하는 시민들의 52%는 '의도적으로 뉴스 소비 중단을 통해 휴식을 취했다'라고 밝혔으며, 87%는 이를 통해 '심리적 행복감을 개선했다'라고 답했다.

뉴스 회피는 미디어 산업을 위협한다. 10명 중 8명이 뉴스를 회피한다면, 다음 세대에서는 누가 뉴스를 볼까. 뉴스 회피 현상이 중요한 이유는 뉴스 회피자 중 과반(52%)이 의도적으로 뉴스를 멀리한다고 말했기 때문이다. '의도적으로' 뉴스를 회피했다는 답변은 그동안의 뉴스 제작 방식을 완전히 바꿔야 하고, 현재 저널리즘을 실행하는 방식을 전면적으로 포기해야 한다는 매우 구체적인 신호다.

니콜라오스 파나요투 교수는 저널리즘에서 AI가 위협이자 동시에 기회가 될 수 있다는 점을 인정하면서도, 지금 논의의 초점이 지나치게 AI에만 쏠려 있다고 지적했다. 그는 AI를 '방 안의 코끼리(the elephant in the room)'라고 비유하며, 그 코끼리를 논의하기 전에 더 크고 무거운 문제, 즉 이미 방 안에 자리 잡은 또 다른 코끼리라 할 수 있는 뉴스 회피와 미디어 신뢰의 붕괴를 먼저 다뤄야 한다고 강조했다.

뉴스 회피를 줄이기 위한 시도 중 하나로 구글이 유튜브를 통해 진행한 '사전 예방(prebunking)' 캠페인이 있다. 이 캠페인은 특정 국가에서 정보 조작 방식과 그 효과를 미리 알려 시민들의 인식을 높이도록 설계된 프로그램이었다. 잘못된 정보가 만들어내는 혼란과 피로감은 뉴스 회피의 중요한 원인 중 하나이므로, 이러한 접근은 독자가 뉴스를 피하기보다 정보를 스스로 식별하며 뉴스에 머물 수 있는 기반을 마련할 수 있다는 점에서 특히 유효하다.

[그림 14] 구글 ‘사전 예방’ 캠페인 영상의 한 장면



▶ 우크라이나 난민 관련 오정보의 확산 방식을 경고하는 내용이 포함돼 있다.

‘사전 예방’은 흔히 백신 접종에 비유된다. 사람들이 악의적 정보, 조작된 콘텐츠, 선전(propaganda)에 실제로 노출되기 전에, 그 전형적인 패턴과 기법을 미리 학습하게 해 스스로 감지하고 저항할 수 있는 심리적 면역력을 키우는 접근이다.

구글의 ‘사전 예방’ 캠페인은 잘못된 정보가 확산한 뒤에 이를 바로잡는 ‘사후 반박(debunking)’ 방식의 한계를 보완하기 위해 고안된 전략이다. 핵심 목표는 사람들이 온라인에서 접하게 될 오해의 소지가 있는 정보를 사전에 식별하고, 심리적으로 저항할 수 있도록 만드는 것이다. 거짓 정보로 인한 결과가 이미 발생하고 난 뒤에는 ‘사후약방문’이 될 가능성이 크기 때문이다.

캠페인 영상은 90초 내외의 짧은 유튜브 광고 형태로 제작돼, 일반 콘텐츠를 시청하기 전에 자연스럽게 노출된다. 흥미로운 점은 영상이 특정 주장이나 콘텐츠를 ‘가짜’라고 지적하지 않는다는 것이다. 대신 조작된 정보에서 반복적으로 등장하는 전형적 기법을 보여주며, 시청자가 실제 상황에서 이를 스스로 알아보도록 훈련한다.

구글이 강조하는 대표적 조작 기술은 크게 세 가지다. 첫째, 공포나 분노 같은 강한 감정을 자극해 이성적 판단을 흐리는 ‘감정 조작(Emotional manipulation)’, 둘째, 복잡한 사회 문제의 책임을 특정 개인이나 집단에 떠넘기는 ‘희생양 찾기(Scapegoating)’, 셋째, 마치 두 가지 선택지만 존재하는 것처럼 상황을 왜곡하는 ‘잘못된 양분법(False dichotomy)’ 등이 그것이다. 예를 들어 유럽의 복잡한 이주·난민 문제를 단순화해 ‘우크라이나 난민 때

문'이라고 책임을 전가하는 프레임은 전형적인 '희생양 찾기' 사례로 소개됐다.

이 캠페인은 2022년 우크라이나 전쟁 이후 폴란드, 슬로바키아, 체코에서 우크라이나 난민에 대한 오정보를 막기 위해 도입됐다. 2023년 인도네시아에서는 선거를 앞두고 감정 조작, 탈맥락화, 명예 훼손과 같은 주요 조작 전술에 대비하는 캠페인이 이뤄졌다. 이 캠페인은 실제 유튜브 환경에서 잘못된 정보를 식별하는 능력을 평균 5~10% 높이는 긍정적 결과를 보였다.

파나이오투 교수는 뉴스 회피 현상을 완화하기 위한 현실적 해법으로 학교·지역사회·언론사가 함께 참여하는 '뉴스 게임화(news gamification)' 접근을 제안했다. 그는 게임 요소를 활용하면 인지 부담을 줄이고 이해도를 높일 뿐 아니라, 뉴스에 대한 신뢰와 참여의 공간을 새롭게 만들 수 있다고 설명했다.

파나이오투 교수 연구팀은 그리스, 불가리아, 루마니아, 세르비아 등지에서 청소년을 대상으로 뉴스 회피를 줄이는 풀뿌리 캠페인도 직접 수행했다. 각 국가의 사회·문화적 맥락에 맞게 변형한 배드 뉴스(Bad News), 하모니 스퀘어(Harmony Square), 고 바이럴(Go Viral) 같은 게임을 학교 현장에서 활용했다. 게임의 목적은 잘못된 정보나 오해의 소지가 있는 뉴스를 포착하거나, 낚시성 기사를 피하는 등의 활동을 통해 학생들의 뉴스 소비 방식을 바꾸는 것이었다.

게임 기반 접근은 학생들의 뉴스 관련 토론을 활성화하고 뉴스를 더 올바르게 소비하는 습관을 형성하는 데 효과적이었다. 또 연구팀은 이러한 변화가 가정으로 확산하는 선순환도 확인할 수 있었다. 파나이오투 교수 연구팀은 이미 2019년 유럽의회 선거 당시 '아이들은 어떤 주제든 부모와 논의한다'라는 경향을 확인한 바 있다. 파나이오투 교수는 뉴스 회피 문제를 '아이들부터' 해결하는 전략이 가족 단위의 뉴스 참여를 확장하는 데 매우 효과적인 방법이라고 강조했다.

파나이오투 교수 연구팀이 시도한 또 다른 접근 방식은 지역 미디어를 사전에 지원하는 전략이었다. 지역 언론은 전국적인 미디어보다 훨씬 더 신뢰를 받고, 지역사회에 더 가깝다. 연구팀은 북마케도니아, 코소보, 불가리아, 그리스 등이 함께 개발한 게임 애플리케이션을 통해 젊은이들이 뉴스 과정의 일부가 되도록 초대했다. 관심 있는 이야기에 기여해 달라고 요청해 게임 애플리케이션이 일종의 뉴스룸이 되도록 시도했다.

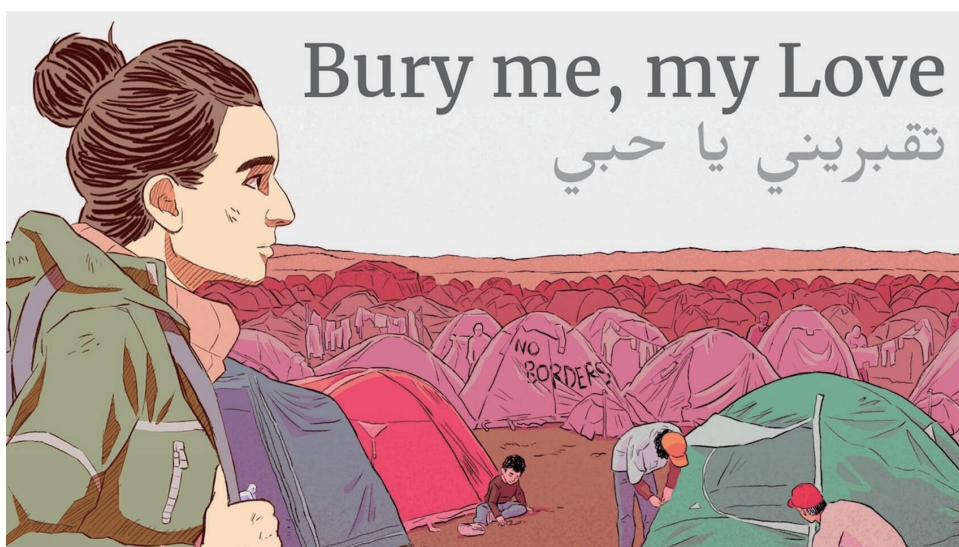
이러한 시도 끝에 참가자의 70%는 언론인이 어떻게 일하는지 더 잘 이해하게 됐으며, 62%는 뉴스를 공유하기 전에 출처를 확인하게 됐다. 또 참가자들은 게임화 과정을 통해 뉴스 소비를 일상생활의 일부로 삼게 됐다.

파나이오투 교수 연구팀의 게임화 시도가 오정보에 강한 시민을 길러내고 뉴스 소비 역량을 높이는 리터러시 중심의 접근이라면, 게임이라는 형식 자체가 언론과 독자의 관계를 회복하고 새로운 접점을 만드는 하나의 유효한 가능성임을 보여주는 사례들도 있다. 그중 하나가 유럽에서 가장 민감한 현안 가운데 하나인 난민 문제를 게임으로 풀어낸 ‘나를 묻어줘, 나의 사랑(Bury Me My Love)’이다.

이 게임은 2015년 내전 때문에 조국을 떠나야 했던 시리아 난민 다나(Dana)에 대한 프랑스 르몽드의 기사를 바탕으로 만들어졌다. 게임 주인공은 시리아 출신의 젊은 커플 누르(Nour)와 마즈드(Majd)다. 누르는 난민 지위를 얻기 위해 독일로 떠나고, 그의 파트너 마즈드는 어머니와 할아버지를 돌보기 위해 시리아에 남는다. 게임 플레이어는 마즈드의 위치에서 독일에 있는 남편을 만나러 갈지 결정하게 되는데, 50개의 서로 다른 장소와 19개의 가능한 결말이 있다. 이 게임은 난민 문제에 대한 유럽인들의 공감을 끌어냈다는 평을 받았다.

뉴스 게임화의 성공적 모델로는 뉴욕타임스의 낱말 퀴즈 ‘워드(Wordle)’이 꼽힌다. 워드는 2021년 구독자 100만 명을 달성했으며, 뉴욕타임스의 독자를 늘리는 데도 크게 기여했다. 영국 왕비 카멜라 해리스(Kamala Harris)는 영국 패션잡지 보그와의 인터뷰에서 잠이 오지 않을 때 손녀딸과 함께 매일 워드를 한다고 말했다. CNN 방송의 앵커 앤더슨

[그림 15] 게임 ‘나를 묻어줘, 나의 사랑’ 포스터



쿠퍼(Anderson Cooper)나 모니카 르윈스키(Monica Lewinsky)와 같은 영향력이 큰 유명인들도 워들에 중독됐다고 고백했다.

뉴욕타임스의 전략은 사용자가 게임하고 새로운 것을 배우는 ‘일일 습관’을 판매하는 것이었다. 워들뿐 아니라 파이낸셜타임스의 ‘우버 게임’은 깃 경제(gig economy, 필요할 때마다 계약·단건 형태로 노동을 제공하는 유연한 노동 방식)에서 생존하는 경험을 제공했고, 이코노미스트와 가디언은 뉴스 퀴즈를 통해 독자 참여를 유도했다. 이 외에도 워싱턴포스트, 뉴욕커 등 게임화 전략을 시도한 언론사는 많다.

게임화를 통한 수익은 저널리즘에 가치를 더하고 뉴스룸의 재정이 지속 가능하도록 도움을 준다. 또 젊은 층의 뉴스 회피 현상을 완화하고, 가정에서 뉴스 소비를 일상 생활화하는 데 긍정적 영향을 미친다.

물론 뉴스 게임화 전략에는 윤리적 과제도 뒤따른다. 난민 문제와 같은 복잡한 위기를 단순한 ‘놀이’로 축소하거나, 뉴스의 본질을 잃고 오락성만 강조하는 ‘경시(trivialization)’의 우려도 있다. 편향성, 투명성 부족 등으로 통합적인 시각이 아니라 부분적인 관점만 제시할 위험도 존재한다. 또 지나친 게임화에 독자들이 피로를 느낄 수도 있다. 그럼에도 게임화는 적절한 설계와 윤리적 기준을 갖춘다면, 뉴스 소비를 활성화하고 새로운 독자층을 유입할 수 있는 유효한 전략으로 평가된다.

AI 기술은 언론의 업무 방식뿐 아니라 독자와 언론의 관계를 재구성한다. 닉 디아코폴로스 교수 연구팀이 제안한 시나리오에서 볼 수 있듯, AI 시대에 언론인은 대중에게 신뢰를 제공하고 책임 있는 정보 전달자로서 핵심적 역할을 유지할 수 있다. 하지만 동시에 잘못된 AI 요약, 부정확한 자동 생성 문장처럼 작은 오류 하나가 언론의 신뢰를 무너뜨리고 독자에게 즉각적인 거부감을 불러일으킬 위험도 커지고 있다.

결국 AI 시대의 과제는 기술을 얼마나 빠르게 도입하느냐가 아니라, 이 기술을 사용하면서도 독자와의 신뢰를 어떻게 지켜낼 것인가라는 언론의 근본적 책무와 맞닿아 있다. AI의 도입이 열어주는 효율성보다 더 중요한 것은, 독자가 뉴스를 어떻게 경험하고 이해하며 다시 신뢰할 수 있게 만들 것인가 하는 질문이다.

이 지점에서 뉴스 회피 연구, 사전 예방 캠페인, 게임화 전략 등이 제안하는 접근은 하나의 공통된 메시지로 수렴한다. 독자는 단순히 정보를 소비하는 존재가 아니라, 뉴스 생태계의 심리적·사회적 맥락을 지닌 관계적 주체라는 점이다. 시민이 정보 조작에 넘어가지 않도록 면역력을 높이는 ‘사전 예방’ 캠페인이든, 뉴스 소비를 놀이적·참여적 경험으로 전환하는 게임화든, 이는 모두 결국 독자가 다시 뉴스에 머물고 관계를 복원하도록

돕는 장치들이다. 앞으로 언론이 AI 시대를 헤쳐 나가는 데 필요한 것은 더 정교한 기술이 아니라, 독자와의 신뢰, 참여, 이해를 다시 설계하는 능력이다. 기술은 그 목적을 위한 수단이며, 미래의 뉴스 생태계는 그 관계를 어떻게 회복하고 확장하느냐에 따라 달라질 것이다.

나가며: DataJConf 2025가 남긴 것

DataJConf 2025는 단순한 콘퍼런스를 넘어, 데이터와 AI가 저널리즘의 의미를 어떻게 다시 쓰고 있는지 집단적으로 점검하는 자리였다. 각 세션은 서로 다른 문제를 다뤘지만 논의는 크게 네 가지 흐름으로 수렴됐다.

첫째, 유럽에서 생성형 AI의 도입은 이미 기술을 넘어 조직 재설계의 문제로 이동했다. BBC·파이낸셜타임스·타미디어·리치 등은 모두 AI를 선택의 문제가 아니라 업무 방식과 책임 구조를 다시 짜야 하는 계기로 받아들이고 있었다. 초기에는 도구 실험이 중심이었지만, 실제 변화를 만든 것은 ‘AI가 안전하게 작동할 수 있는 조직을 어떻게 만들 것인가’라는 질문이었다. AI 리터러시 교육, 보안·윤리 기준 정비, 데이터 보호 체계 강화, 업무 플로우의 재정의가 함께 이뤄질 때야 비로소 기술은 조직 내부에서 효과를 발휘했다. 또 작은 실험의 축적이 대규모 전환을 가능하게 했다는 점에서, AI는 결국 기술보다 ‘문화와 구조’의 문제임을 확인하는 자리였다.

둘째, 데이터 인프라는 뉴스룸 경쟁력을 좌우하는 핵심 자산으로 부상했다. 파이낸셜타임스가 금융 기사 자동화보다 먼저 데이터 파이프라인을 재건한 이유, APA가 대체 텍스트 자동화를 위해 인포그래픽 구조 자체를 다시 검토해야 했던 이유는 분명하다. AI는 고품질 데이터 없이는 작동하지 않는다. 데이터는 수집·정제·저장·공유라는 고된 과정의 결과물이며, 이 인프라를 갖추지 않은 뉴스룸은 AI 실험조차 효율적으로 진행할 수 없다. 이번 DataJConf가 반복해서 던진 메시지는 결국 하나였다. “AI를 쓰고 싶다면 먼저 데이터를 다루는 집을 지어라.”

셋째, 협업은 더 이상 선택이 아니라 저널리즘의 기본 전제가 됐다. 루시 블록의 지역 기반 시민사회 협력 모델, 셰릴 필립스의 지역 데이터 공용 인프라, 안나 투리치의 국경

간 데이터 프로젝트는 모두 저널리즘이 기존의 범위를 넘어서는 새로운 해법을 필요로 한다는 점을 드러냈다. 팬데믹의 주거 위기부터 지역 뉴스 생태계의 붕괴, 지정학적 내러티브까지, 오늘의 문제는 어느 한 조직이나 한 국가의 관점만으로는 해결할 수 없다. 시민 단체는 현장의 데이터와 맥락을, 대학과 연구기관은 분석 인프라를, 기자는 취재와 검증을 제공할 때 협업은 비로소 힘을 발휘했다. DataJConf는 ‘함께해야만 보이는 진실’이 존재함을 실증적으로 보여줬다.

넷째, AI 시대의 신뢰·책임·독자 회복이라는 오래된 과제는 한층 더 무거워졌다. AI 쓰레기, AI 요약 시스템의 오류, 팬데믹 이후 확산된 뉴스 회피 문제까지, 기술 발전에 따른 부작용이 빠르게 커지고 있는 상황에 대응해 저널리즘의 신뢰 기준을 더욱 강화해야 한다는 논의가 이어졌다. 뉴스 회피가 그리스에서도 80%를 넘는 현실은 기술보다 ‘독자 경험’이 더 근본적 문제임을 보여주는 사례였다. 구글의 사전 예방 캠페인, 청소년 대상 뉴스 게임화 교육, 뉴스 게임화 시도 등은 기술적 대응을 넘어 독자가 뉴스를 이해하고 신뢰할 수 있는 환경을 만드는 것이 미래의 핵심 과제임을 시사했다.

DataJConf 2025가 남긴 가장 큰 메시지는 단순했다. 저널리즘은 기술이 아니라 사람과 구조 위에서 진화한다는 점이다. 생성형 AI, 자동화 도구, 데이터 기반 협업은 저널리즘의 역할을 대체하는 것이 아니라 확장하는 수단이다. 기술의 도입 여부가 아니라, 어떤 철학·규범·인프라 위에 기술을 얹느냐가 향후 뉴스 생태계를 결정하게 될 것이다. 다시 말해, 저널리즘의 미래는 더 고립된 전문성이 아니라 더 넓게 연결된 역량 위에서 열린다. DataJConf 2025는 이 연결 지도를 제시한 자리였다.

DataJConf 2025에서 확인된 흐름은 한국 언론에도 그대로 닿아 있다. 국내 주요 언론사는 다양한 형태로 생성형 AI를 실험하고 있다. 조선일보의 ‘조선 AI 기사 작성 어시스턴트’, 동아일보의 경제 관련 질의응답 서비스 ‘에스크비즈(AskBiz)’, 한국일보의 ‘하이(H.AI) 뉴스룸 도우미’ 등 여러 서비스가 이미 업무 현장에 스며들었다. 속보 경쟁과 인력 부족을 동시에 겪고 있는 한국 뉴스룸의 현실을 고려하면, 생산성을 높여주는 AI 도구의 도입은 자연스러운 선택이기도 하다.

그러나 DataJConf 2025가 보여준 가장 본질적인 메시지는 기술의 속도보다 ‘그것을 받아들이는 조직의 구조와 문화가 준비됐는가’가 AI 도입의 성패를 가른다는 점이었다. 생성형 AI는 ‘즉각적인 효율 향상의 도구’가 아니라, 데이터 인프라·업무 절차·책임 구조·교육 체계를 다시 설계해야 작동하는 복합적 시스템이라는 것이다. 이 기준을 한국 상황에 대입하면 선결과제가 더욱 선명해진다. 빠른 실험과 도입 흐름에 비해, 이를 뒷받

침할 조직적 기반은 아직 한국 언론에 충분히 갖춰졌다고 보기 어렵다. 기자별 AI 활용 수준의 격차, 취재·편집 프로세스와의 불완전한 결합, 데이터 정제 및 저장 체계의 부재, 내부 논의의 구조적 부재 등은 한국 언론이 지금부터 해결해야 할 공통의 문제로 보인다.

생성형 AI는 이미 뉴스 생산 기술의 일부가 되었지만, 이 기술이 저널리즘의 질을 끌어올릴지, 오히려 신뢰를 잠식할지는 조직의 선택에 달려 있다. AI 도입의 기준은 속도가 아니라 철학, 혁신의 크기가 아니라 책임, 성능이 아니라 독자의 신뢰가 돼야 한다. 생성형 AI는 기사를 대체하지 않으며, 대신 기자가 어디에 시간을 써야 하는지를 재배치하는 기술일 뿐이다. 반복적 작업을 줄여 더 중요한 취재에 집중하게 하는 도구가 되느냐, 혹은 콘텐츠의 신뢰도를 떨어뜨리는 위험 요소가 되느냐는 뉴스룸 설계에 달려 있다.

DataJConf 2025의 메시지는 한국 언론에도 동일하게 적용된다. 기술은 중심이 아니며, 중심에 있어야 하는 것은 인간의 판단과 저널리즘의 핵심 가치다.

2025
해의 미디어 동향
2025 WORLD MEDIA TRENDS



ISSN 2799-8304