

## 보고서: <언론사를 위한 오픈소스 LLM 활용 방안> 날카로운 질문은 인간의 영역 AI가 ‘동료 시스템’ 되려면

전창영 한국언론진흥재단 선임연구위원

생성형 인공지능과 대규모 언어 모델은 이미 뉴스룸의 일상 도구가 됐지만, 효율성과 신뢰성 문제는 여전히 남아 있다. 공개형 대규모 언어 모델과 뉴스 피드 기반 국제뉴스 자동 수집·분석 시스템을 중심으로, 언론사를 위한 LLM 시스템 개발 방향을 제안한 보고서 내용을 살펴본다. 편집자 주

### 언론사 뉴스룸 가상 시나리오

2027년, 지역 일간지 A사는 새벽 편집회의에서 전날 밤 국내외에서 발생한 사건·사고와 각 부처 보도자료, 지역 의회 일정, SNS에서 급상승한 지역 키워드를 한 화면에서 본다. ‘뉴스룸 LLM 시스템’이 밤새 RSS와 포털, 정부·지자체 공개자료, 자사 기사 아카이브를 긁어와 요약해 준 결과다.

편집국장은 화면에서 몇 개의 카드만 골라 ‘심층 취재 필요’, ‘속보 중심 처리’ 등의 라벨을 붙인다. 이후 취재기자들은 각자의 단말기에서 LLM과 대화를 주고받으며 관련 통계와 선행 보도, 핵심 쟁점을 정리해 취재에 들어간다. 독자는 아침 출근길에 이미, 자신의 관심사에 맞게 큐레이션 된 지역 뉴스 브리핑을 받아본다.

생성형 인공지능과 대규모 언어 모델(LLM)은 이미 많은 뉴스룸에서 ‘똑똑한 비서’로 쓰이고 있다. 그러나 국내 언론 산업이 직면한 수익 악화, 신뢰 위기, 인력 부족, 지역·중소 언론의 디지털 격차를 고려하면, AI를 단순한 도구가 아니라 ‘뉴스룸 인프라’로 설계하는 논의가 필요하다. 이 글에서는 한국언론진흥재단 연구보고서 <언론사를

위한 오픈소스 LLM 활용 방안>(전창영·이종혁·백강희)<sup>1)</sup> 연구 결과를 바탕으로, 언론사를 위한 LLM 시스템의 개발 방향을 제안한다.

## 뉴스룸에서 이미 일상화된 생성형 AI

연구에서 조사한 국내 언론인 103명 가운데 약 70%가 생성형 AI를 실제 업무에 활용하고 있다고 답했다. 이 중 30%는 매우 자주 사용하고 있었다. 인터넷 언론사와 전국 종합일간지, 경제·통신사에서 활용 비율이 특히 높았고, 지역 종합일간지는 50% 안팎에 그쳐 상대적으로 낮았다.

업무 단계별로 보면, 기자들은 취재 단계에서 △자료 수집·분석, △정보 검색, △전문용어·배경 지식 탐색, △녹취록 정리 등에 LLM을 가장 많이 활용했다. 기획 기사나 탐사보도에서 다뤄야 하는 방대한 보고서·통계 자료를 요약하고, 사건 일지나 시계열 흐름을 정리하는 데서 효용을 크게 느끼는 것으로 나타났다.

보도 단계에서는 △기사 초안 작성, △보도자료 재구성, △외신·제3외국어 번역, △그래프·인포그래픽 설명문 작성 등에서 활용이 두드러졌다. 특히 속보 경쟁이 치열한 통신사·방송사 기자와 저연차 기자일수록 “초안은 AI에게 맡기고, 문장 다듬기와 사실 확인에 집중한다”라는 응답이 많았다.

반면 배포·관리 단계에서의 활용은 아직 제한적이다. 독자 맞춤형 뉴스 추천, 요약 뉴스 서비스, 댓글·피드백 분석 등에서 가능성이 크다고 보지만, AI 요약 기사에 독자가 머무르면 정작 언론사 웹사이트로 유입이 줄어들 수 있다는 수익 구조상의 우려도 적지 않았다.

## 효율성 vs 신뢰성, 기자들이 느끼는 양가감정

연구진은 생성형 AI 활용 빈도와 숙련도가 높은 언론인을 대상으로 포커스 그룹 인터뷰(FGI)를 실시했다. 인터뷰 참여자 대부분은 생성형 AI가 반복적이거나 기계적인 업무를 크게 줄여준다는 점에서 높은 효용을 인정했다. 실제로 녹취록 정리, 방대한 통계자료의 전처리, 생소한 개념 설명, 수십 페이지 분량의 정책 문서 요약 등은 AI 도입 이전으로 돌아가기 어려울 정도로 이미 업무에 깊숙이 내재화됐다는 평가가 적지 않았다. AI는 기자의 시간을 절약하는 보조 도구를 넘어, 취재 준비와 정보 탐색을 가속화하는 기초 인프라로 자리 잡아가고 있는 셈이다.

그러나 AI가 생성한 텍스트를 온전히 ‘기사’로 인정할 수 있는지를 두고 견해차가 크게 갈렸다. 특히 정치·사회부처럼 현장성과 맥락 해석이 중요한 부서는 AI의 한계를 더욱 분명하게 지적했다. 기자들은 뉴스 가치 판단, 취재원 발언의 숨은 의도 읽어내기, 현장 취재를 통한 사실 확인은 여전히 인간의 영역이라고 강조했다. 한 정치부 기자는 사회·정치 기사를 “수많은 정보 중에서 가장 날카로운 지점을 찾아들이는 과정”이라고 표현하며, “LLM은 그 뽀족함, 즉 맥락 속에서 의미를 결정짓는 포인트를 스스로 찾아내지 못한다”라고 지적했다. AI가 제시하는 초안이나 정리는 충분히 유용하지만, 저널리즘의 본질적 판단을 대신하기는 어렵다는 인식이 강하게 드러난 것이다.

1) 보고서는 2026년 1월 한국언론진흥재단 홈페이지를 통해 공개될 예정이다. 보고서에는 국내 언론인들의 LLM 활용 능력 및 현황 조사 결과, RSS를 활용한 국제뉴스 자동 수집·분석 시스템 개발 결과 등이 포함된다.

이러한 기자들의 양가적 태도는 국내 뉴스 이용자들이 보이는 신뢰 구조와도 맞닿아 있다. 국내 뉴스 이용자들은 ‘언론 전반’에 대해서는 낮은 신뢰를 보이지만, ‘내가 선택해 읽는 뉴스’에 대해서는 상대적으로 높은 신뢰를 나타낸다.<sup>2)</sup> 이는 뉴스 소비가 점점 개인화·선택화되는 흐름 속에서, 이용자들이 스스로 필터링한 정보에 더 강하게 신뢰를 부여한다는 뜻이기도 하다. 그러나 이러한 환경일수록 AI가 개입한 기사에 대한 독자의 기대는 오히려 더 높아질 수 있다.

따라서 AI 관여도가 높아질수록 산출물에 대한 책임 소재는 어디에 있는지, 오류가 발생했을 때 어떤 기준으로 설명할 수 있는지, 정보의 출처와 생성 과정이 얼마나 투명하게 공개될 수 있는지에 대한 요구는 커질 수밖에 없다. 생성형 AI의 효율성이 아무리 높아지더라도, 뉴스룸이 이를 신뢰 가능한 도구로 정착시키기 위해서는 저널리즘 윤리와 투명성 확보를 중심에 둔 새로운 운영 기준이 함께 마련돼야 한다는 점이 분명히 드러난다.

## 왜 ‘오픈소스 LLM’인가

국내 언론사들은 이미 상용 AI 서비스(챗GPT, 클로드 등)를 널리 활용하고 있지만, 본격적인 ‘뉴스룸 LLM 시스템’으로 나아가기 위해서는 오픈소스 LLM을 기반으로 한 자체 시스템 구축이 필요하다. 그 이유를 세 가지로 정리해 봤다.

첫째, 비용 절감과 예측 가능성이다. 상용 모델 API를 대량 호출하는 방식은 단기 도입은 쉽지만, 트래픽이 늘어나면 비용 구조를 예측하기 어렵다. 반면 오픈소스 모델은 한 번 인프라를 마련하면 사용량 증가에 따른 추가 비용을 상대적으로 안정되게 관리할 수 있다.

둘째, 데이터 통제와 보안이다. 기사 원고, 미공개 취재 메모, 내부 편집 메모, 독자 정보 등은 언론사의 핵심 자산이다. 이를 외부 상용 모델에 그대로 보내는 것은 법적·윤리적 리스크가 크다. 로컬 혹은 전용 클라우드 환경에서 오픈소스 LLM을 운용하면, 학습·추론 과정에서 데이터 통제를 강화할 수 있다.

셋째, 뉴스룸 특화 커스터마이징이다. 언론사는 정치·경제·지역·국제 등 도메인별 용어와 기사 문법, 매체별 문체, 편집국별 데스크 기준이 뚜렷하다. 오픈소스 LLM에 자사 기사 아카이브와 스타일 가이드를 반영해 파인튜닝하거나, ‘검색 증강 생성(RAG·Retrieval-Augmented Generation)’을 통해 아카이브를 연결하면, 개별 언론사의 문체와 관점을 어느 정도 구현할 수 있다.

## 오픈소스 LLM의 언론사 활용 방안

연구 결과와 현장 인터뷰를 종합하면, LLM 기반 도구는 취재-작성-분석-배포에 이르는 뉴스 생산 전 과정에 걸쳐 다양한 형태로 적용될 수 있다. 중요한 점은 이러한 기능들이 단순히 업무 시간을 단축하는 수준을 넘어 뉴스룸의 운영 방식 자체를 재구성하는 전환점이 될 수 있다는 것이다. 반복 작업의 자동화는 기자의 인지적 부담을 덜어 더 깊이 있는 취재와 해석을 가능하게 하고, 아카이브·데이터 분석 기능은 기존에는 접근하기 어려웠던 장기적 흐름과 구조적 문제를 파악하는 데 도움을 준다. 또한 독자 상호작용 기능은

2) 이현우, 전창영, 김선호(2025), <디지털 뉴스 리포트 2025 한국>, 한국언론진흥재단

[표 1] 언론사 LLM 활용 시나리오

항목	활용 가능성
취재 보조 에이전트	<ul style="list-style-type: none"> <li>특정 정책·사건에 대해 핵심 쟁점을 요약하고, 관련 통계·선행 연구를 링크와 함께 정리</li> <li>취재 메모·녹취록을 입력하면, 기사 각 꼭지 후보와 추가 취재가 필요한 질문 리스트 제시</li> <li>생소한 전문용어·법률 개념·기술 개념을 기자 눈높이에 맞춰 설명</li> </ul>
기사 작성·편집 지원	<ul style="list-style-type: none"> <li>속보 상황에서 보도자료와 브리핑 전문을 바탕으로 1차 스트레이트 초안 생성</li> <li>자사 스타일 가이드에 맞춘 문장 다듬기, 중복 표현 제거, 리드 문장 변주 제안</li> <li>외신·제3외국어 자료의 초벌 번역을 수행하고, 인포그래픽에 들어갈 캡션·설명문 추천</li> </ul>
아카이브·데이터 저널리즘 도구	<ul style="list-style-type: none"> <li>‘2016년 이후 노동개혁 관련 사설·칼럼의 주요 논지’ 등과 같은 수년 치 아카이브 요약·분석</li> <li>의회 회의록·법안·예산안 등 방대한 공공문서에서 쟁점 문단을 뽑아주는 탐색 도구</li> <li>지역 의회·지방정부의 공개자료를 자동 수집해 ‘지역 현안 레이다’ 대시보드 구성</li> </ul>
독자 상호작용	<ul style="list-style-type: none"> <li>독자의 질문·검색 로그를 분석해, 자사 기사 중 적합한 콘텐츠를 추천하고 부족한 주제를 역으로 발굴</li> <li>뉴스레터·알림 서비스를 위한 맞춤형 요약·큐레이션 자동 생성</li> <li>허위 정보 의심 링크를 입력하면, 사실관계 검증에 필요한 자료와 체크 포인트를 정리해 주는 팩트체크 보조 도구</li> </ul>

독자의 질문과 관심사를 실시간으로 반영함으로써, 언론이 일방적 전달자에서 상호적 관계를 기반으로 한 정보 서비스 제공자로 전환하는 기반이 된다.

결국 LLM 시나리오는 단순한 기술 적용 목록이 아니라, 언론사가 자신의 정체성과 보도 철학을 유지하면서도 AI를 어떻게 통합할 것인지를 보여주는 설계도에 가깝다. 어떤 항목에 어느 정도로 AI를 개입시킬지, 자동화와 인간 판단 사이의 경계는 어디까지 허용할지, 그리고 독자에게 어떤 투명성 기준을 마련할지에 따라 뉴스룸의 AI 활용 가치는 달라질 것이다. LLM은 기사를 대체하는 존재가 아니라, 기자의 해석과 판단을 강화하는 ‘동료 시스템’에 가까워야 한다.

## 쏟아지는 국제뉴스에서 의미 있는 흐름을 포착하려면

이러한 문제의식은 추상적인 원칙에 머무르지 않고, 실제 뉴스 생산 환경에서 구현할 수 있는 시스템을 통해 구체화될 필요가 있다. 특히 설문과 FGI에서 많은 기자들은 생성형 AI와 LLM의 가능성에는 공감하면서도, 이를 실제 업무에 어떻게 적용해야 하는지에 대해서는 여전히 막연함을 느끼고 있다고 답했다. 본 연구는 이러한 현장의 요구를 출발점으로 삼아, 향후 LLM 활용을 염두에 둔 선행 단계로서 국제뉴스 영역의 자동화 가능성을 먼저 실험했다.

전 세계에서 동시에 쏟아지는 국제뉴스를 모두 따라잡는 일은 개별 언론사의 역량만으로는 불

[그림 1] RSS 기반 뉴스 자동 수집·분석 시스템 화면



가능에 가깝다. 본 연구에서 구축한 ‘RSS(Really Simple Syndication) 기반 국제뉴스 자동 수집·분석 시스템’<sup>3)</sup>은 바로 이 문제를 해결하기 위해 만들어졌다. 이 시스템은 북미·유럽·아시아 등 주요 해외 언론사의 공개 RSS 피드를 상시적으로 수집·정제·분석해, 국제 이슈의 발생과 확산, 변화 과정을 시간 흐름에 따라 파악할 수 있도록 설계된 지능형 플랫폼이다.

기존에 국내 언론사들이 국제뉴스의 시의성을 확보하기 위해 개별 모니터링이나 전문 뉴스레터, 외신 제공사 정보에 의존해 왔다면, 이 시스템은 그러한 과정을 자동화된 백엔드에서 일괄적으로 처리한다. 이를 통해 시차로 인한 정보 공백을 줄이고, 서로 다른 매체가 동일한 사안을 어떤 맥락과 관점에서 보도하고 있는지 비교·검토할 수 있는 환경을 제공한다. 수집된 기사들은 자동 생성된 주제·국가·지역 태그를 중심으로 탐색할 수 있으며, 특정 이슈를 둘러싼 보도량 변화와 국가별·매체별 경향 역시 함께 확인할 수 있다.

예를 들어 중동 분쟁, 글로벌 경기 전망, 반도체 공급망 이슈와 같은 주요 국제 현안이 어느 국가에서 먼저 보도되기 시작했고, 이후 어떤 지역과 매체를 중심으로 확산됐는지를 흐름으로 추적할 수 있다. 이는 국제부는 물론 경제부·산업부 기자에게도 강력한 도구다.

나아가 이 시스템은 단순한 모니터링을 넘어 뉴스룸 업무 흐름과의 연동 가능성도 염두에 두고 있다. 특정 국가나 테마, 기관과 관련된 보도량이 급증할 경우 이를 자동으로 감지해 알림을 제공하거나, 언론사 내부 CMS와 연동해 취재 메모나 브리핑 자료로 즉시 활용하는 방식으로 확장할 수 있다. 향후 LLM 기반 요약이나 이슈 브리핑 생성 기능이 결합될 경우, 수집·분석된 국제뉴스 데이

3) 이 시스템은 기술을 전혀 모르는 기자도 사용할 수 있도록 설계됐다. 컴퓨터와 인터넷만 있으면 된다. 시스템 주소는 다음과 같다. <https://jongh-hh-intnews-rss.hf.space/>

[그림 2] 주요 사건을 자동 정리·요약하고, 해당 사건 보도 기사 원문 링크를 제공하는 기사 분석 결과 화면



터를 바탕으로 주요 쟁점을 정리한 ‘국제 이슈 브리핑’을 자동으로 생성하는 것도 가능하다.

RSS 기반 국제뉴스 자동 수집·분석 시스템은 단순한 외신 모니터링 도구를 넘어, 국제 이슈의 맥락과 흐름을 구조적으로 파악할 수 있게 하는 뉴스룸 인프라의 한 형태다. 이를 통해 국제뉴스를 ‘따라가는 대상’이 아니라 ‘선제적으로 해석하는 정보 자원’으로 전환하는 데 기여할 수 있을 것이다.

### 기술 이전에 조직의 문제

다만 위의 RSS 기반 국제뉴스 자동 수집·분석 시스템은 LLM 활용을 전제로 한 자동화 기술이 뉴스룸 인프라로 작동할 수 있음을 보여주지만, 동시에 그러한 기술을 실제로 도입·운영하는 과정에서 어떤 조건과 기준이 요구되는지를 다시

물게 한다. 즉, 문제는 ‘무엇을 개발할 수 있는가’보다 ‘어떤 조건이 갖춰져야 안전하게 쓸 수 있는가’이다.

설문과 FGI에서 기자들이 꼽은 가장 큰 제약은 △AI 산출물의 정확성과 신뢰성 부족 △출처·학습데이터 불투명성 △저작권·법적 책임 문제 △기술 인력 부재와 인프라 비용 △조직 내부의 저항과 직무 불안이었다. 따라서 언론사를 위한 LLM 시스템 개발은, 최소한 [표 2]의 네 가지 항목을 동시에 고려해야 한다.

### 기자의 질문을 더 날카롭게 만드는 LLM

생성형 AI와 오픈소스 LLM은 분명 언론의 위기를 심화시킬 위험과 함께, 새로운 기회를 동시

[표 2] 언론사 시스템 개발 시 고려 사항

항목	내용
<b>품질·출처 검증</b>	<ul style="list-style-type: none"> <li>• LLM이 참조한 문서를 명시하고, 인용·요약 부분을 쉽게 대조할 수 있는 UI 설계</li> <li>• ‘AI 산출물 그대로 보도 금지, 반드시 기자 검증 후 사용’ 등을 명문화한 뉴스룸 준칙 제정</li> <li>• 허위·오류 사례를 수집해, 프롬프트와 데이터셋을 지속적으로 개선하는 내부 프로세스 구축</li> </ul>
<b>윤리·법·저작권 가이드라인</b>	<ul style="list-style-type: none"> <li>• 학습·추론 단계에서 활용할 수 있는 데이터 범위, 개인정보 처리 원칙, 제삼자 저작물 사용 기준 명확히 규정</li> <li>• 독자에게 ‘AI가 어느 단계에 어떻게 개입했는지’를 설명하는 투명성 원칙 정립</li> </ul>
<b>교육·역량 강화</b>	<ul style="list-style-type: none"> <li>• 프롬프트 설계, RAG·파인튜닝의 기본 개념, 모델의 한계와 편향 이해를 포함한 체계적인 교육</li> <li>• 연차·부서별로 다른 교육 커리큘럼: 저연차 기자에게는 기본 활용·프롬프트, 데스크와 간부에게는 전략·윤리·리스크 관리에 초점을 맞춘 교육</li> </ul>
<b>중소·지역 언론 자원 인프라</b>	<ul style="list-style-type: none"> <li>• 단일 언론사가 구축하기 어려운 GPU·스토리지 인프라를 공공 혹은 컨소시엄 형태로 공동 구축</li> <li>• 지역 언론을 위한 표준화된 오픈소스 LLM 아키텍처와 예제 코드, 매뉴얼 제공</li> </ul>

에 품고 있다. 이미 이용자들은 포털과 소셜미디어를 넘어 AI를 통해 뉴스를 접하고 있으며, 일부는 언론사를 거치지 않고 AI와 ‘정보 직거래’를 시도하고 있다. 언론사가 이 흐름을 방관한다면, 뉴스 생산과 유통의 주도권은 플랫폼과 AI 사업자에게 더 빠르게 넘어갈 것이다. 그러나 기술 자체가 저널리즘을 대체하는 것은 아니다. 포커스 그룹 인터뷰에서 한 기자가 말했다, “AI가 정교한 답을 내놓을 수는 있지만, 어떤 질문을 던질지는 결국 기자의 몫”이다. LLM 시스템이 기자의 취재·분석·해석 능력을 보완하고 확장하는 방향으로 설계될 때, AI는 언론의 경쟁자가 아니라 ‘동료 시스템’이 될 수 있다.

언론사를 위한 LLM 시스템 개발은, 기술 도입 여부를 넘어 ‘AI 시대에도 저널리즘의 공적 책임을 어떻게 구현할 것인가’에 대한 선택의 문제다. 지금 필요한 것은 각 언론사가 자신의 정체성과 독자, 조직 환경에 맞는 LLM 활용 전략을 주도적으로 설계하고, 이를 뒷받침할 공적·산업적 지원 체계를 함께 고민하는 일이다. 그 과정에서 오픈소스 LLM과 공공 인프라라는 한국 언론 생태계의 새로운 기반이 될 수 있을 것이다. 📌