

인공지능 시대, 다변화되는 저널리즘의 역할 진단

뉴스통신진흥회 후원 세미나

일시 2023년 9월 22일(금) 14:00~17:00

장소 한국언론진흥재단 미디어교육원 디지털랩실

주최  **한국여성커뮤니케이션학회**
Korean Women's Association for Communication Studies

후원  **뉴스통신진흥회**
Korea News Agency Commission

한국여성커뮤니케이션학회 뉴스통신진흥회 후원

「인공지능 시대, 다변화되는 저널리즘의 역할 진단」 세미나

○ 일시 : 2023년 9월 22일(금) 14:00~17:00

○ 장소 : 한국언론진흥재단 미디어교육원 디지털랩실

○ 주최 :  한국여성커뮤니케이션학회
Korean Women's Association for Communication Studies

○ 후원 :  뉴스통신진흥회
Korea News Agency Commission

프로그램

시간	내용
14:00~14:10	개회사 : 홍지아 (한국여성커뮤니케이션학회 회장/경희대) 사회 : 이동후 (인천대)
1부 14:10~15:10	[발제1] 알고리즘 작동 플랫폼 시대, 뉴스 이용자의 저널리즘 이슈들 및 외국의 사례 - 이재원(이화여대) [발제2] 인공지능 팩트체크와 '사실성'의 기술사회학적 정의 - 이정현(중앙대) · 박소영(조선대) [발제3] 자동화 저널리즘과 기술-인간의 피드백 루프 : 뉴스 댓글창 진화 사례를 중심으로 - 홍남희(서울시립대)
15:10~15:30	휴 식
2부 15:30~17:00	[종합토론] - 김고은(한국기자협회) - 박서연(미디어오늘) - 양소은(서울대) - 이재영(연합뉴스) - 이설희(용인대) - 이희은(조선대)
17:00	폐 회

1

알고리즘 작동 플랫폼 시대, 뉴스 이용자의 저널리즘 이슈들 및 외국의 사례

이재원(이화여대)



알고리즘 작동 플랫폼 시대, 뉴스 이용자의 저널리즘 이슈들 및 외국의 사례¹⁾

이재원(이화여자대학교 커뮤니케이션·미디어연구소 연구위원)

1. 들어가며

전 세계적으로 주요 플랫폼의 영향력이 커지며 기존 언론사의 뉴스 이용률이 감소하고 있는 가운데 국내에서도 뉴스를 직접 생산하는 언론사가 아니라, 플랫폼을 통한 뉴스 이용이 증가하고 있다. 구체적으로 네이버나 다음 등 포털 사이트, 유튜브 페이스북 트위터 인스타그램 등 글로벌 플랫폼, 카카오톡 등 메신저를 통해 뉴스를 이용하는 추세가 늘어나고 있다. 이러한 플랫폼들은 알고리즘을 기반으로 이용자의 취향에 맞춰 서비스된다는 특징을 갖는다.

한국인의 디지털 뉴스 이용은 포털 사이트와 뉴스 수집 서비스를 중심으로 이뤄졌다(최진호·박영흠, 2022). <로이터 연구소 디지털 뉴스 리포트 2023>에 따르면 한국은 네이버나 다음과 같은 포털사이트 중심으로 뉴스를 소비해 온라인에서 뉴스를 보는 비중이 79%였고, 유튜브에서 뉴스를 시청하는 소비자가 늘어 소셜 미디어가 45%를 차지했다(Newman et.al., 2023).

특히 20대 이용자는 소셜 미디어와 뉴스레터, 모바일 알림 등의 방식을 다른 연령대보다 많이 이용하는 것으로 나타났다. 소셜 미디어에서 뉴스를 소비하는 20대 이용자는 23%로 12~16%인 다른 연령대보다 높았다. 뉴스레터나 휴대전화 알림으로 뉴스를 이용하는 20대는 15%로 역시 다른 연령대보다 비율이 앞섰다(최진호·박영흠, 2022). 이용률이 높지는 않지만, 젊은 세대의 뉴스 소비 통로가 포털 사이트에서 소셜 미디어로 이동하는 흐름을 예측할 수 있다.

이러한 뉴스 이용의 변화는 단지 뉴스 유통의 변화를 의미하지 않는다. 조사에 따르면 이용자는 플랫폼을 그 자체로 '언론'으로 생각한다. 이용자들은 포털 사이트(65.1%), 온라인 동영상 플랫폼(31%), SNS(17.8%)를 언론으로 인식하고 있다(한국언론진흥재단, 2021). 특히, 이용자들은 플랫폼을 언론으로 인식하는데 그치지 않고, 신뢰하고 있다는 점이 주목할 만하다. 신뢰도 높은 매체로 유튜브와 네이버를 꼽았고 전통 미디어의 신뢰도는 떨어졌다(한국언론진흥재단, 2021).

1) 본 연구는 현재 진행 중인 연구인 관계로 인용을 삼가주시기를 부탁드립니다.

이용자들은 뉴스가 지나치게 많다고 인식하며 이에 대해 피로감을 느끼고, 이에 따라 뉴스를 회피하는 움직임이 커졌다. 2022년 한국 이용자 67%는 뉴스를 회피(news avoidance)하고자 적극적으로 노력한다고 답했는데, 5년 전인 2017년 52%에 비해 15%p나 증가했다. 세계적으로도 뉴스를 선택적으로 피하는 비율(69%)이 크게 늘었다(최진호·박영흠, 2022). 한국은 대선을 치르는 과정에서 뉴스 소비가 활발해 다른 국가에 비해 뉴스를 회피하는 비율이 다소 낮았지만, 이용자들은 코로나19 관련 뉴스의 과잉으로 인한 피로감을 느끼는 것으로 나타났다.

특히 젊은층의 뉴스 무관심이 두드러져, 로이터저널리즘연구소 선임연구위원 닉 뉴먼은 이렇게 소위 ‘소셜 네이티브’ 세대가 뉴스를 보게 만드는 게 앞으로 과제가 될 것이라고 전망했다(최진호·박영흠, 2022). 젊은 세대가 뉴스를 회피하는 이유를 다음과 같이 분석했다. 첫째, 소셜 미디어를 통해 뉴스를 많이 이용하기 때문에, 뉴스가 어렵다고 느껴 피한다. 둘째, 전통 미디어에서 생산한 뉴스는 나이 많은 언론인들이 작성해 어휘가 어렵다. 셋째, 소셜 미디어에서 파편적으로 전달되는 뉴스를 접하다 보니 사안을 전반적으로 파악하기 힘들다. 즉, 소셜 미디어 중심으로 매개된 커뮤니케이션을 수행하는 게 자연스러운 이용자들에게 기존의 저널리즘 가치체계와 운영 방식을 고수하면 결국 뉴스를 회피하게 만든다는 의미라는 것을 알 수 있다.

알고리즘 저널리즘은 이용자 차원에서 다음과 같은 문제를 내포한다.

첫째, 뉴스의 탈맥락화와 뉴스의 질적 저하 현상이다. 전통 미디어에서는 하나의 이슈에 여러 기사를 묶어서 제공하기도 하고, 깊이 있는 심층 기사를 제공하기도 한다. 소셜 미디어에서는 이용자의 피드에 기사가 등장할 때, 기사들끼리 묶어서 전달되지 않는다. 이용자가 팔로우하는 수많은 계정과 더불어, 이용자의 로그 데이터를 기반으로 추천되는 광고 등의 피드와 함께 전달된다.

때문에, 소셜 미디어에서 개별 단위로 이용자에게 도달하는 기사는 탈맥락화된 다양한 종류의 정보들 사이에 섞여서 전달된다. 플랫폼 기업은 기존 언론사처럼 뉴스 서비스를 위해 만들어진 기업이 아니므로, 기사 이외의 각종 정보와 혼재되어 뉴스가 유통되는 것이다(이재원, 2021). 이러한 현상은 굳이 알 필요없는 정보 속에서 중요하고도 믿을 만한 뉴스가 누락될 우려가 있다. ‘가짜정보’, 질적으로 낮은 뉴스 등과 중요한 뉴스를 구분하기 어려운 상황이다.

둘째, 정보의 편향성이다. 소셜 미디어의 추천 서비스는 이용자에게는 더없이 편리한 서비스이지만, 정보 유통의 차원에서는 정보의 쓸림 현상을 야기한다. 정보 필터링 현상을 일컫는 ‘필터 버블(filter bubble)’(Pariser, 2011)은 이용자들이 다양한 정보를 접하지 못하게 만들 우려가 있다. 이용자는 자신의 취향, 개인정보 등을 분석해 개인화(personalization) 취향화(customization)된 콘텐츠를 소셜 미디어 기업으로부터 전달받게 되기 때문이다. 이로 인해 같은 정보만 지속적으로 소비하는 ‘반향실 효과(echo chamber effect)’(Jamieson & Cappella, 2008)가 나타나게 된다. 정보의 차이가 자신이 갖고 있는 신념을 강화하는 방식으로 작동하여, 양극화가 나타나 민주주의의 발전에 부정적인 영향을 줄 가능성이 있다.

셋째, 뉴스 과잉으로 인한 혼란이다. 지나치게 많은 기사, 정보, 가짜뉴스까지 쏟아져 나오면서 이용자

들은 뉴스 피로감(news fatigue)을 느끼며, 이는 뉴스 무관심으로 이어지고 있다.

본 연구에서는 알고리즘을 하나의 비인간 행위자로 보고, 알고리즘이 저널리즘 지형을 바꾸며 이용자가 저널리즘 차원에서 의미있는 뉴스를 접할 수 있는 방안에 대해 질문을 던질 예정이다. 저널리즘 지형에서 알고리즘을 감시하며, 이용자 및 언론인에게 알고리즘에 대한 교육을 마련하는 등의 활동을 벌이는 외국의 사례를 탐색하고자 한다. 이러한 작업을 통해 알고리즘 저널리즘 시대에 플랫폼과 언론사 뿐 아니라, 이용자가 함께 바람직한 저널리즘을 만들어갈 수 있는 논의의 장을 만들고자 한다.

2. 알고리즘 저널리즘 이슈들

1) 알고리즘 저널리즘 논의 - 생산자에서 생태계로

알고리즘(algorithm)은 문제 해결을 위한 '논리(logic)'와 문제 해결을 위한 전략인 '통제(control)'라는 두 요소로 구성된다(Kowalski, 1979). 2010년대 이후 알고리즘은 소셜 미디어가 갖고 있는 일종의 권력을 작동하게 만드는 기제로 널리 사용되고 있다. 밴딕(van Dijck, 2013)은 알고리즘은 기능을 계산하는 명령들이라고 규정하며, 소셜 미디어에서 데이터 계산을 가능하게 한다고 지적한다. 현재 알고리즘은 소셜 미디어 뿐 아니라 포털 사이트나 챗봇과 같은 광범위한 플랫폼과 소프트웨어에 두루 확대, 적용되고 있다.

알고리즘 저널리즘(algorithmic journalism)은 주로 컴퓨테이셔널 저널리즘(computational Journalism)의 범주에서 논의되어 왔다. 그동안 알고리즘 저널리즘은 “기사의 취재, 작성, 편집, 유통 등 저널리즘 과정 중 어떤 하나를 정해진 규칙에 따라 만들어진 알고리즘으로 대체 또는 보완하는 것”을 목표로 하는 저널리즘으로 주로 생산의 차원에서 논의되었다(오세욱, 2016, 53쪽). 때문에 알고리즘 저널리즘은 주로 자연어 생성(natural language generation, NLG)의 관점에서 인공지능이 뉴스의 전문적 생산이나 산업적 역할을 해 낼 수 있는지 논의하며 영역을 만들어왔다(Dörr, 2016).

오늘날 알고리즘 저널리즘은 소프트웨어로 기사를 작성하는 방식이나 로봇 저널리즘 등 생산에만 국한되어 논의할 수 없다. 이용자들이 기사를 전문적으로 생산하는 전통 저널리즘 주체가 아닌, 알고리즘을 기반으로 작동하는 소셜 미디어를 이미 언론으로 인식하고 있는 만큼 알고리즘 저널리즘 개념의 확장이 필요한 상황이다. 이용자들은 유튜브 페이스북 트위터 인스타그램 카카오톡 등 소셜 미디어를 언론으로 인식하고 있으며 이들 플랫폼은 알고리즘을 기반으로 작동하기 때문이다.

앤더슨(Anderson, 2013)은 컴퓨테이셔널 저널리즘을 사회학적으로 접근하며 빅 데이터, 알고리즘 분석과 저널리즘의 관계를 다각도로 조망했다. 앤더슨은 다음과 같은 차원에서 알고리즘 저널리즘을 논의해야 한다고 진단했다. 첫째, 정치와 공공정책(politics and public policy), 둘째, 경제학(economics), 셋째, 기관 및 분야(institutions and fields), 넷째, 조직 수준의 역학(organization-level dynamics),

다섯째, 문화사(cultural history). 여섯째, 기술과 뉴스(technology and the news) 등이다.

이러한 맥락에서 주로 뉴스의 생산 과정을 자동화하는 방식으로 사용했던 알고리즘 저널리즘(algorithmic journalism)(Anderson, 2012) 개념은 공익적 책무를 지녔다는 견해를 전제로 한다(Cohen, Hamilton, & Turner, 2011). 알고리즘 저널리즘은 “저널리즘의 책무성 기능을 보완하기 위한 알고리즘, 데이터, 사회과학적 지식의 조합(the combination of algorithms, data, and knowledge from the social sciences to supplement the accountability function of journalism)”(Hamilton&Turner, 2009)으로 보는 관점을 채택해야 좀 더 폭넓은 논의를 가능하게 한다.

이러한 시각을 바탕으로 본 논문에서는 ‘알고리즘 저널리즘’을 알고리즘을 기반으로 작동하는 소셜 미디어에서 일어나는 저널리즘 생산, 유통, 소비로 규정한다. 바람직한 알고리즘 저널리즘은 생산자와 이용자, 플랫폼이 함께 만들어 가는 생태계의 변화를 전제로 해야 한다는 관점이다.

2) 플랫폼별 이슈

(1) 포털 사이트 인공지능 알고리즘 편집 - 기사의 질적 저하

포털 사이트 네이버는 ‘뉴스콘텐츠’의 뉴스 배열을 ‘AI 추천인 시스템(AI Recommender System)’으로 운영하며 알고리즘으로 뉴스 지면을 편집하고 있다. 네이버의 ‘에어스’ 알고리즘은 다른 플랫폼처럼 이용자의 취향을 기반으로 하며, 유사한 취향의 이용자 로그 데이터를 활용한다. 여기에, 콘텐츠 소비 습성, 언론사들의 생산 방식 등을 반영하며 알고리즘을 수정해나간다.

네이버는 2017년 2월 에어스 서비스를 처음 도입할 당시, 공식 블로그 ‘네이버 다이어리’를 통해 두 가지 유형의 알고리즘을 적용한다고 밝혔다. 첫째, CF(Collaborative Filtering, 협력 필터링) 기술, 둘째, 인공신경망 기술인 RNN(Recurrent Neural Network)이다²⁾. 에어스 전면 도입 이후, 네이버 측은 알고리즘 원리를 좀 더 자세히 공개했다. 2020년 <관훈저널>에 ‘네이버 뉴스 알고리즘 이렇다’라는 제목으로 에어스 알고리즘은 첫째, 사용자 로그에 기반을 둔 협력 필터링, 둘째, 딥러닝 기반의 뉴스 기사 품질 모델, 셋째, 뉴스 기사 클러스터링이라고 공개했다(유봉석·최재호·최창렬, 2020)³⁾.

2) CF는 사용자별 콘텐츠 소비 패턴을 분석해, 유사한 이용자가 많이 본 콘텐츠를 추천해주는 방식이다. RNN은 이용자가 콘텐츠를 본 순서까지 고려해 스스로 학습하는 인공지능 딥러닝 방식으로, 더욱 정교하게 최신 콘텐츠를 보여주는 방식이다.

3) 먼저, 사용자 로그 기반의 추천이다. “추천시스템은 사용자가 검색어를 입력하지 않아도 사용자의 과거 이력과 성별, 연령, 위치 등의 컨텍스트(context) 정보를 기반으로 가장 적합한 아이템을 찾아주어야 한다”(50쪽)고 알고리즘 구성 정책을 설명했다. 전체 사용자의 70%인 로그인 사용자의 로그 기록을 바탕으로 로그 기반(log-based) 협력 필터링 알고리즘을 작동 중이다. 둘째, 다수 이용자의 소비 이력을 바탕으로 △기사 제목, △본문, △이미지, △비라인, △기사작성 시간 등 기사의 정보를 컴퓨터에 입력해 딥러닝 기술을 이용해 완성도 있는 기사인지 예측한다. 품질이 높은 기사를 누락시키지 않기 위한 알고리즘 설계이다. 셋째, 뉴스 클러스터링(clustering)은 “사용자 개인의 관심 여부와 상관없이 다수의 언론사에서 다루는 이슈에 대한 정보를 기반으로 뉴스를 추천하는 방식”(54쪽)이다. 이러한 방식은 “비슷한 시간대에 많은 언론사가 같은 주제로 생산하는 기사가 많다면, 대중적으로 관심이 많고 중요 이슈일 수 있다는 걸 전제로 한다”(54쪽)고 설명한다.

네이버 에어스는 여러 알고리즘을 결합해 통합적으로 뉴스 추천을 최종 결정하게 된다. 2021년 기준, 뉴스 추천 시 고려사항은 첫째, 실시간 사용자의 선호도 예측, 둘째, 자동화 방식으로 뉴스 품질 측정, 셋째, 시의적절한 주요 이슈 감지, 넷째, 확장성 있는 시스템 구조이다(네이버 서치엔테크 블로그, 2021). 특히 뉴스 품질 측정 시 고품질 기사는 △기사의 제목, 본문 등 콘텐츠 정보, △네이버 뉴스 이용자의 피드백 데이터(클릭수, 체류시간)을 함께 고려해 품질을 추론한다고 밝혔다. 고품질 기사란 “뉴스를 소비하는 사용자의 관점에서 많은 사용자들이 만족스럽게 읽은 기사”라고 규정했다. 이 때 이용자들의 만족을 많은 이들이 클릭했는가, 오래 동안 기사에 머물렀는가로 측정한다는 것이다.

이후 네이버는 2023년 7월 7일 ‘네이버 뉴스 AI알고리즘 안내’⁴⁾ 페이지를 공개했다. 네이버 측은 “생산자에게는 주요 기사의 편집권을, 사용자에게는 관심 있는 언론사의 선택권”을 제공한다”고 밝혔다. 첫째, AI알고리즘을 개인화 영역에서 Airs, 둘째, 기사 그룹핑은 클러스터링을 기반으로 한다고 설명했다(〈그림 1〉 참조).

〈그림 1〉 네이버 뉴스 AI알고리즘 개요(2023. 7. 7.)



구체적으로 Airs 추천에 영향을 끼치는 알고리즘 팩터(피쳐) 19가지를 공개했다(〈표 1〉 참조).

4) <https://media.naver.com/algorithm>

〈표 1〉 네이버 뉴스 알고리즘 팩터 (2023. 7. 7.)

추천모델	피처	설명
개인화 (CF모델)	NPMI(Normalized Point-wise Mutual Information) 점수	협업 필터(Collaborative Filtering) 알고리즘의 하나인 NPMI 모델을 활용합니다. 사용자가 '최근 클릭한 기사'들과 '함께 많이 클릭되는 기사'들을 후보로 추출하고 추천 점수를 부여합니다.
	NRMS (Neural news Recommendation with Multi-head Self-attention)	내용 기반 필터링(Content-Based Filtering) 알고리즘의 하나인 NRMS 모델을 활용합니다. 기사 내 단어의 중요도와 단어 사이의 관계를 분석하고, 주요 주제를 추론합니다. 이를 바탕으로 사용자가 읽은 기사에서 '사용자의 선호도를 나타내는 중요 뉴스'가 무엇인지 추론합니다.
개인화	섹션 선호도	사용자의 뉴스 섹션에 대한 선호도 : 사용자가 최근에 해당 섹션의 기사를 많이 클릭할수록 더 높은 값을 부여합니다.
	언론사 선호도	사용자의 언론사에 대한 선호도 : 사용자가 최근에 해당 언론사의 기사를 많이 클릭할수록 더 높은 값을 갖게 됩니다.
	제목 키워드 선호도	클릭한 기사의 제목에 존재하는 키워드에 대한 선호도
	엔터티(Entity) 키워드 선호도	클릭한 기사의 '제목+본문'에서 추출한 엔터티 키워드에 대한 선호도 : 엔터티 키워드가 포함된 기사를 자주 클릭할수록 선호도가 증가합니다.
	기자 구독 여부	기자 구독 여부 : 구독한 기자의 기사인 경우에 가산점을 부여합니다.
	언론사 구독 여부	언론사 구독 여부 : 구독한 언론사의 기사인 경우에 가산점을 부여합니다.
	숨김 언론사	사용자가 숨김 설정한 언론사의 기사는 후순위로 추천됩니다.
비개인화 (QE 모델)	기사 품질 점수	QE 모델을 통해, 자동으로 예측한 기사 품질 점수 : 기사 제목/본문/섹션/콘텐츠 타입(e.g. 글, 글+사진, 글+동영상)/이미지를 기준으로 기사 품질 점수 예측
비개인화 (SI 모델)	기사 최신 인기도 (Latest Popular) 점수	SI(User) 모델을 통해, 기사의 클릭수로 측정한 인기도 점수입니다.
	기사 클러스터 점수	SI(Press) 모델을 통해, 제목과 본문이 유사한 기사들을 묶고 '기사 클러스터 점수'를 부여합니다.
	기사 클러스터 최신성	최근에 생성된 기사 클러스터일수록 더 높은 값을 부여합니다.
	SI(Social Impact) 점수	SI(Press) 모델로 생성된 클러스터의 기사들이, SI(User) 모델에서 선정한 '최신인기(LP)' 기사에 해당하면 SI 점수 부여합니다.
비개인화	기사의 최신성	최근에 작성된 기사일수록 더 높은 값을 부여합니다.
	심층기사 여부	언론사에서 심층·기획 기사로 선정한 기사에 대하여 가산점을 부여합니다.
	시간당 기사 생산량	기사량이 적은 주말 또는 새벽 시간대에 클러스터 내 기사에 가산점을 부여합니다.
	지면 위치	온라인 신문 지면 위치를 뜻합니다. 앞 지면에 등장하는 기사일수록 가산점을 받습니다.
	노출 균형	과거 언론사별 노출 비중을 바탕으로 확률적 추정을 진행합니다. 이를 통해 언론사별 추천 기사량의 불균형을 완화합니다.

뉴스 알고리즘 팩터는 개인의 취향을 반영하는 개인화 팩터 9가지와 기사의 품질 중심으로 구성된 비 개인화 팩터 10가지로 구성되어 이들 팩터가 Airs에 쓰인다고 밝혔다.

둘째, 뉴스 클러스터링은 “같은 시간대 비슷한 내용의 기사 다수를 묶어낸다면, 해당 기사 그룹을 하나의 ‘이슈’로 볼 수 있다”는 관점으로 10분마다 기계적인 자동 분류를 한다고 밝혔다. “Airs 알고리즘이 사용자에게 클러스터를 기사처럼 추천하고, 또한 클러스터 안에 묶인 기사 중에서 어떤 기사를 더 선호할지 사용자별로 예측해 제공”한다고 설명했다.

포털 사이트에서 뉴스 편집을 전면 알고리즘에 맡긴 이후 언론사 기자들은 저널리즘 환경이 더 어려워졌다고 보고 있었다. 알고리즘이 표방하는 뉴스 가치를 정확히 알기 어렵다는 생각으로 알고리즘을 추적해 클릭을 위한 경쟁을 더욱 치열하게 해 나가며 질적 하락을 나타냈다.

네이버의 에어스 도입 이후 뉴스 생산자를 인터뷰한 결과 뉴스 생산 방식은 다음과 같이 변한 것으로 드러났다(이재원, 2020). 첫째, ‘가짜단독’의 증가이다. 뉴스 생산자들은 에어스 알고리즘이 ‘단독’에 가중치를 부여한다는 판단으로, 과거라면 ‘단독’이라고 표시하지 않을 기사에도 ‘단독’이라고 표기했다. 둘째, 소셜 미디어의 포스팅을 베끼어 쓰는 기사들이 늘어났다. 이는 SNS 내용을 그대로 쓴 기사가 ‘많이 본 뉴스’에 오르는 기사, 즉 보다 많은 독자가 클릭하는 기사의 유형이라는 것을 기자들이 학습했기 때문이다. 선행 자료가 있어야 작동할 수 있는 알고리즘의 특성을 파악한 조치다. 결국, 이러한 변화는 뉴스 생산자들이 알고리즘의 가중치를 부여받기 위한 ‘알고리즘 속이기’ 전략이며, 다른 기사 없이 홀로 독주하는 콜드 시스템(cold system)보다는 클러스터링을 형성하는 게 자사 기사 노출에 유리하다는 판단 아래 행해지고 있었다. 심지어 단독 기사를 발행할 때 타사 기자에게 정보를 제공하며 함께 기사를 작성하도록 유도했는데, 클러스터링을 만들기 위한 궁여지책이었다. 이러한 변화는 질(quality)적인 가치를 표방했던 저널리즘 차원에서는 양(quantity)이 확보되어야 하는 인공지능의 딥러닝 방식을 따라가느라 뉴스 가치에 혼선이 빚어지며 발생한 것이다.

네이버가 2023년 7월 알고리즘 팩터를 발표한 이후 뉴스 생산자 10명을 심층 인터뷰한 연구 결과, 알고리즘이 주도하는 생태계에서 뉴스의 내용은 클릭수가 많이 나올만한 내용인지 여부가 더 중요해져 질적인 하락이 심각해졌다. 이러한 변화는 첫째, 이슈 맞춤형 주제인지 중요해졌고, 둘째, 일반인의 기사가 증가했고, 셋째, 여론 중심으로 아이টে이 선택되는 경향을 보인다(이재원, 2023).

알고리즘이 저널리즘을 더욱 공정하게 만드는데 도움이 되는지에 관한 논의는 지속되고 있다. 네이버가 2019년 에어스를 전면 도입할 당시 언론은 네이버가 ‘드루킹 사건’에서 촉발된 댓글 논란 및 편집의 공정성 논란에서 자유로워지고자 이러한 선택했다고 분석하며, 알고리즘이 제대로 검토되었는지에 의문을 제기했다(이국현, 2019).

네이버는 뉴스 알고리즘을 공개적으로 검증해오고 있다. 2018년 뉴스 알고리즘을 검증하는 ‘알고리즘 검토위원회’를 구성해 전문가들이 직접 검토했고, 당시 필터버블 우려가 제기되었고, 3년만인 2021년 다시 위원회를 꾸리며 양질의 기사를 추천하는 방안을 검토했다(금준경, 2021). 두 번째 알고리즘 검토

위원회의 검토 결과 편향성은 없다고 판단했지만, 다음과 같은 네 가지 개선 사항을 권고했다. 첫째, 기사의 심층성 강화, 둘째, 언론사별 추천 기사 편차 가능성 개선, 셋째, 저널리즘 환경 변화를 반영한 품질 평가 가이드라인 재정립 및 검색 학습 데이터 확충, 넷째, 신규 알고리즘 반영 및 새로운 피쳐, 학습 데이터의 객관적 검증 등이다. 이후 네이버는 투명성을 높이고 심층 보도를 추천하는 개선 작업을 하고 있다. <뉴시스> 보도에 따르면 개선 전인 2022년 4월 19~27일과 개선 후인 4월 29일~5월 5일을 비교한 결과 심층·기획기사의 전체 추천 비중이 기존대비 685% 증가했다. 네이버는 검토위 권고사항을 반영해 기사의 심층성을 강화하고자 기사의 논증 구조를 분석해 다양한 논증 구조를 갖춘 기사를 심층 기사로 판단하기로 했다⁵⁾.

이처럼 알고리즘을 지속적으로 개선하고 있지만, 이용자의 로그 데이터를 기반으로 작동하는 알고리즘 체계에서 이용자의 클릭이 무엇을 의미하는지 확인할 필요성은 남아있다. 네이버 개발자들은 모든 클릭을 긍정적으로 전제했지만, 그렇지 않은 클릭도 존재하기에 이 문제를 해결해야 한다고 논의하기도 했다(Park, Lee, & Choi, 2017).

(2) 유튜브 알고리즘의 허위 정보 생산, 유통

한국은 다른 나라에 비해 한국에서는 소셜 미디어 중 유튜브(44%)에서 가장 많이 뉴스를 접한다는 특징을 갖는다(최진호·박영흠, 2022). 유튜브의 독특한 특성이 허위정보를 쉽게 만들고 퍼트릴 수 있다는 우려를 낳는다.

유튜브는 알고리즘 정책을 자세히 밝히지는 않지만, 유튜브 생산자들이 참고할 수 있게 안내를 하고 있다. 유튜브 측은 2021년 유튜브크리에이터아카데미에 유튜브의 알고리즘은 “시청자를 따라 가는 것”이라며 △시청자가 보는 것, △시청자가 보지 않는 것, △시청자의 동영상 시청 시간, △‘좋아요’ 및 ‘싫어요’, △‘관심 없음’ 의견 등을 살핀다고 밝혔다. 유튜브는 유튜브 알고리즘에 대해 “추천시스템은 실시간 피드백 루프를 통해 시청자 개인마다 다양한 관심사에 맞는 동영상을 찾는다”라고 설명했다.

피드백 루프(feedback loop)는 이용자 행동 데이터(평판, 반응, 측정 등)의 긍정적/부정적 피드백을 구분해 관리, 콘텐츠와 이용의 품질 유지하는 방식이다(최선영·고은지, 2018). 김인식·김자미(2021)는 “유튜브가 어떤 알고리즘을 사용하는지는 정확하게는 알려진 바는 없지만, 협업 필터링과 콘텐츠 기반 필터링을 활용하는 것으로 알려져 있다”(72쪽)고 설명했다. 협업 필터링은 이용자와 비슷한 성향의 이용자가 선호하는 콘텐츠를 추천하는 방식이며, 콘텐츠 기반 필터링은 이용자가 시청한 콘텐츠와 유사한 특성의 콘텐츠를 추천하는 방식이다.

뉴스 이용자에게 유튜브에서 뉴스를 소비할 때 우려되는 문제는 ‘가짜 뉴스’, 즉 허위정보이다. 한국언

5) https://m.blog.naver.com/naver_search/222875360677

론진흥재단 미디어연구센터가 온라인 설문조사를 벌인 결과 명백한 허위사실이라는 것을 알고도 해당 내용을 포함해 콘텐츠를 제작하거나 유포하는 행동에 대해 응답자의 87%가 ‘매우 심각한 문제임’이라고 답했다(양정애, 2021).

유튜브에서 허위 정보가 빠르게 확산되는 이유는 다음과 같다(이재원, 2021). 첫째, 크리에이터 중심의 콘텐츠 생산 방식이다. 메신저를 신뢰하면 그가 발행하는 내용을 모두 믿게 되어 사실 확인을 소홀히 할 우려가 있다. 둘째, 유튜브의 추천 알고리즘은 이용자가 오래 머물도록 하는 기준으로 운영되는 만큼, 자극적인 콘텐츠에 대한 검토가 이뤄지지 못한다. 유튜브 추천 알고리즘 개발팀에서 일한 전 구글 엔지니어 기욤 샬로(Guillaume Chaslot)는 “유튜브 알고리즘은 시청 시간에 최적화되어 있다”라며 “음모론이나 혐오 콘텐츠가 시청 시간을 늘린다면, 근거 없는 콘텐츠더라도 다른 콘텐츠보다 더 많은 추천될 것이고 이 같은 현상은 세계적으로 관찰되고 있다”라고 말했다(박세원, 2021). 셋째, 유튜브는 영상을 공유하기 쉬운 구조이기에 다른 플랫폼으로의 확산이 쉽다. 정정주·김민정·박한우(2019)는 여기에 온라인 공간에서의 집단적 유대감이 ‘사회적 동조’를 일으킨다면 더욱 위험할 수 있다고 경고했다. “정보전달과 소통 측면에서 허위정보 확산의 연쇄작용은 이러한 패턴을 끊을 수 있는 증거를 찾지 못하거나 설사 증거를 찾는다 하더라도 기존의 이용자는 증거에 주목하지 않은 채 불명확한 정보에 계속 의지할 수 있다는 것”(134쪽)이라고 우려를 나타냈다. 특정 이슈가 발생했을 때 몰려들어 허위정보를 양산하는 ‘사이버렉카’와 같은 현상이 나타나는 이유다.

특히 유튜브에서 허위 정보가 쉽게 전파되는 중요한 이유는 이용자들이 유튜브를 저널리즘의 공간으로 인식하고 있지 않다는 점이다. 시청률 조사기관인 닐슨은 ‘2019 뉴스 미디어 리포트-유튜브 저널리즘’에서 뉴스 미디어로써 유튜브의 역할에 주목하기 시작했다(정철운, 2019). 보고서에 따르면 유튜브에서 뉴스를 이용하는 이들은 △흥미성, △편리성, △다양성 등의 동기로 유튜브 뉴스를 이용했다. 문제는 전통 뉴스에서 중요하게 여기는 △신뢰성, △전문성은 유튜브 이용자에게는 주요 이용 동기가 되지 못한다는 것이다. 유튜브에서 뉴스 이용자가 기대하는 가치는 △재미(Fun), △유쾌한 장난(Frolic), △경박함(Frivolity)이었다. 유튜브를 찾는 이용자들은 기존 미디어의 보도와 동일한 뉴스를 기대하고 찾는다고 보기는 어렵다는 이야기다. 이러한 유튜브의 특성과 이용자가 기대하는 가치가 맞물려 허위 정보가 생산, 유통되기 쉬운 플랫폼이 되었다.

(3) 페이스북 알고리즘 사실보다 여론 중심의 뉴스 이용

전세계에서 가장 많은 이용자가 사용하는 플랫폼인 페이스북은 여전히 전세계 뉴스 이용자들이 가장 많이 사용하는 플랫폼이기도 하다. 그러나 한국에서는 유튜브 카카오톡에 이어 세 번째로 이용하는 뉴스 플랫폼이다. 페이스북은 2022년 7월 기준 국내 월간활성이용자(MAU) 1,109만 6,919명을 기록했는데, 2020년 5월(1,487만 910명) 이후 25% 넘게 감소한 수치이다(박수현, 2022).

페이스북은 이용자들이 자율적으로 게시물을 올리는대로 노출하는 방식을 5년간 이어왔으나, 반복적인 게시로 노출 순위가 올라갈 수 있다는 점에서 2009년부터 주목도가 높은 순서대로 게시물을 노출하도록 알고리즘에 개입하기 시작했다(공성운, 2022). 이후 노출 순서를 결정하는 알고리즘 정책은 지속적으로 바뀌었다. 페이스북 알고리즘을 규제하는 법안이 미국 의회에서 2022년 2월 발의되었다. 미국 공정거래위원회(FTC)가 어린이와 취약층을 보호하도록 유해 콘텐츠가 발견되면 알고리즘 조정을 요구할 수 있는 'SNS 네티지 액트' 법안이다.

페이스북에서 뉴스 소비는 사실이 아닌 의견 중심으로 이뤄진다는 문제가 제기된다. 개별 뉴스 단위가 아니라, 뉴스를 전달하는 페이스북의 이용자 중심으로 뉴스가 소비되고 유통되면서, 중요한 의제에 대한 감각을 페이스북 내 연결된 '친구'들 사이에서 키우게 된다. 예컨대, 페이스북 이용자들이 특정한 기사를 많이 공유하면, 해당 내용을 다른 소셜 미디어에서 검색해보고, 포털 사이트에서 확인하는 식이다. 전통 미디어처럼 혼자 뉴스를 소비하는 게 아니라, 소셜 미디어에서 함께 소비하는 뉴스의 '사회적 읽기'(social reading)가 새로운 현상이 된 것이다. 뉴스가치보다는 뉴스를 공유한 사람이 누구인지, 믿을만한지 여부가 뉴스를 읽을지 결정하는 요소로 작동한다. 전통 미디어에서 기사를 취할 때는 각 언론사의 성향이나 논조가 사회적으로 공유되어 있었지만, 독자는 자신이 원하는 성향의 신문이나 방송사를 택하여 기사를 소비했다. 하지만 미디어 단위가 아니라 비선형적으로 개별 뉴스가 친구에 의해 전달되는 페이스북에서는 '어떤 친구가 공유한 토픽인지'가, '어떤 미디어가 보도한 뉴스인가'보다 중요할 수밖에 없다(이재원, 2018).

이러한 현상은 두 가지 의미를 갖는다. 첫째, 소셜 미디어상 게이트키퍼의 중요성이 대두되었다는 점이다. 파파차리시(Papacharissi, 2010)는 소셜 미디어 이용자들이 게이트키퍼의 위치를 갖게 됨에 따라 '3단계 이론'이 가능하다고 보고, '2단계 이론'의 확장이라고 봤다. 정보 흐름에서 오피니언 리더(opinion leader)의 영향력을 인정한 '2단계 이론'이 소셜 미디어 시대에는 소셜 미디어의 오피니언 리더까지 포함되는 '3단계'로 정보가 확산하는 양상을 보인다. 실제로 페이스북 내에 '친구들'의 참여 정도에 따라 게이트키퍼(gatekeeping)가 단계적으로 이뤄지고 있었다. 페이스북에서 뉴스는 혼자 읽기가 다른 친구들과 함께 읽어나가는 구조인데, 여기에는 이용자들의 참여 정도에 따라 의제가 달라지고 있었다.

둘째, 사실 보다는 의견 중심으로 뉴스를 소비하는 양상이다. 기존의 미디어에서는 사건 사고를 중심으로 보도하는 '스트레이트 뉴스'와 배경을 설명하고 분석하는 '해설뉴스' '심층뉴스', 의견 중심의 '사설' '칼럼' 등이 구분이 되었다. 그러나 소셜 미디어에서는 이러한 기사들이 구분없이 공유되며 '친구'의 의견까지 덧붙여지면서 많은 '친구'들이 어떻게 해석하는지 중요하게 생각하게 된 것이다.

페이스북에서의 사회적 읽기는 언뜻 공론장에서의 소통처럼 보이기도 한다. 그러나, 이는 페이스북이라는 플랫폼에 한정된, '플랫폼상의 사회연결성(platformed socialty)'(van Dijck, 2013)에 기반을 두고 있을 뿐이다. 저널리즘의 이상이라고 볼 수 있는 '공론장에서의 자유로운 정보 교류와 의견 교류'가 페이스북에서 이뤄지고 있을까. '친구' 관계 중심으로 작동하는 플랫폼이기에 자유롭고 이성적인 논의보다

는 ‘친구’ 사이를 조율하는 방식으로 자신의 성향을 고수한다(이재원, 2019). 즉, 친구를 믿고 뉴스를 읽는 방식은 뉴스의 내용보다는 친구의 의견에 먼저 집중하게 만든다. 페이스북에서 어떤 뉴스가 생성되고 소멸하는 기제는 치열한 논쟁이 아니라, ‘유희의 문화’를 바탕으로 관계를 조율하는 방식으로 이뤄지는 것이다. 정치적으로 반대 성향의 ‘친구’가 보내는 뉴스에 자신의 의견을 피력하기 보다는 조용히 ‘친구’ 관계를 해지하거나 숨김처리하는 식이다.

3. 외국 알고리즘 저널리즘 사례

1) 이용자 보호 비영리 단체 운영

(1) 미국 -AI 및 디지털 정책 센터

AI 및 디지털 정책 센터(CENTER FOR AI AND DIGITAL POLICY·CAIDP)는 워싱턴 DC에 설립된 비영리 교육 기관으로, CAIDP는 인공 지능과 디지털 정책이 더 나은 사회, 더 공정하고 더 공정하고 더 책임 있는 사회를 촉진하는 것을 목표로 한다. CAIDP는 OECD AI 원칙 및 AI에 대한 보편적 지침을 포함해 AI 정책을 위해 글로벌 단위로 지도자와 지지자를 모으는 단체이다.

CAIDP는 AI 정책 분야에 대한 최초의 직업 윤리 강령을 발표했다. 목표는 AI 정책에 관한 조언을 제공하는 사람들이 높은 수준의 전문적인 행동을 하도록 하는 것이다. CAIDP 직업 윤리 선언문은 AI 시스템이 타인의 권리와 기회에 영향을 미치고 공공 안전에 위협을 초래할 수 있으며 이러한 편견이 계속되는 문제임을 인식하고 있다. 다른 의무 중에서도 성명서는 개인이 AI 시스템이 공정하고 책임 있고 투명하도록 보장할 것을 요구한다.

〈그림 2〉 AI 및 디지털 정책 센터 홈페이지 및 인증서



CAIDP는 2022년부터 열고 온라인 회의로 전세계 관계자들이 참여할 수 있는 인증 프로그램을 시작했다. CAIDP AI 정책 인증을 받으려면 주제는 AI 역사, AI 이슈 및 제도, AI 규제를 연구하고 AI 정책에 대한 직업 윤리 선언문에 서명해야 한다. 인증 프로그램을 이수하면 인증서를 받을 수 있다(〈그림 2〉 참조).

(2) 독일 - 알고리즘와치(ALGORITHMWATCH)

알고리즘와치는 자동화된 의사 결정(automated decision-making·ADM) 시스템과 그 시스템이 사회에 미치는 영향을 감시, 분석 및 분석하는 데 전념하는 비영리 연구 및 옹호 단체이다. EU 'AI법(AI ACT)' 관련해 독일 의회 디지털 위원회 청문회에 2022년 9월 26일 알고리즘와치의 성명을 전달했다. 알고리즘와치는 "제안서에 명시된 의무는 주로 AI 시스템 제공자를 대상으로 하지만 AI 시스템의 배포자(AI 법 용어로는 사용자에 대한 투명성과 책임 메커니즘을 강화해야 한다"고 전했다.

알고리즘와치는 홈페이지에 따르면 다음과 같은 방향성이 제시되어 있다.

a. 비전

일반적으로 기술, 특히 알고리즘 시스템이 개인을 보호하고 혜택을 주고 사회를 더 공정하고 민주적이며 포용적이고 지속 가능하게 만드는 데 사용되는 세상을 상상한다.

b. 미션

알고리즘 시스템의 사용이 소수가 아닌 다수에게 이익이 되도록 보장하기 위해 권위주의 대신 민주주의와 법치주의, 감시 대신 자유, 비인간화 대신 인권, 권력 비대칭 대신 자율성을 보호하는 알고리즘 시스템을 옹호한다. 편애와 불평등 대신 정의와 평등, 착취적인 삶의 방식 대신 지속가능한 삶을 지향한다.

이 미션을 달성하기 위해 알고리즘 의사 결정 프로세스의 사용과 인간과 사회에 미치는 영향을 분석하는 도구를 개발하고 사용한다. 일반 대중과 의사 결정자에게 그들의 위험, 갈등 및 가능한 피해를 지적하고 과학적으로 견고하고 윤리적으로 책임 있는 연구와 응용 프로그램을 구별하도록 돕는다.

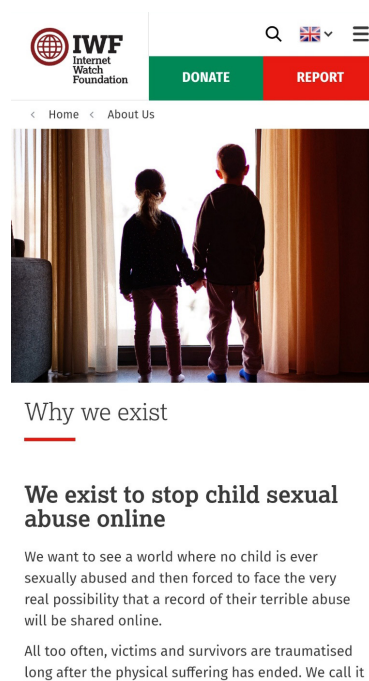
c. 가치

존경, 공정성, 연대, 무결성, 포용성, 다양성, 투명성, 책임 등 다른 사람에게 기대하는 것을 실천한다. 우리의 입장과 요구는 전문적인 분석과 높은 수준의 연구를 통해 얻은 증거를 기반으로 한다. 정치적으로 당파적이지 않다. 공익을 설득력 있고 존중하는 대표자가 되기 위해 노력한다.

(3) 영국 - 인터넷감시재단(Internet Watch Foundation)

영국의 인터넷감시재단은 1996년 온라인에서 어린이를 보호하기 위해 시작되었다. 어린이가 성적 학대를 당한 뒤 기록이 온라인에 확산되지 않기 위해 존재한다고 홈페이지에 밝히고 있다(〈그림 3〉 참조). 더불어 인터넷 산업을 보호하고 자율적으로 규제하기 위해 경찰, 영국 정부, 인터넷 사업자들의 노력으로 출범한 기구이다(KISO, 2015).

〈그림 3〉 영국 인터넷감시재단이 밝힌 재단의 목적



인터넷감시재단은 영국 정부가 2023년 가을 인공지능 규제에 관한 세계 첫 정상회의를 계획하자, AI 생성 아동 성 학대 이미지 문제도 논의해달라고 의견을 전했다(타이디, 2023).

2) 알고리즘 저널리즘 지원 활동

(1) 이용자의 데이터 기부 바탕의 알고리즘 저널리즘 연구

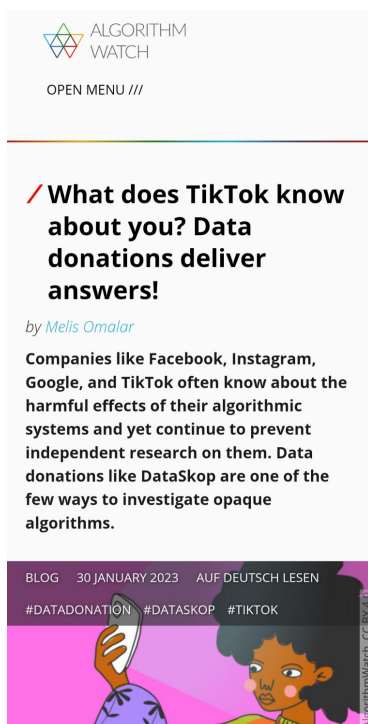
시민이 데이터를 자발적으로 기부해 알고리즘 연구 진행에 참여하는 협업형 프로그램 데이터 스콥(DATASKOP)을 운영 중이다. 독일 연방 정부의 재정 지원을 기반으로 알고리즘와치와 대학(ALGORITHMWATCH, EUROPA UNIVERSITÄT VIADRINA, MEDIALE PFADE, FH POTSDAM 및

UNIVERSITÄT PADERBORN)들이 연구에 참여한다.

DATASKOP의 데이터 기부 연구의 첫번째 결과로 5,000명 이상의 데이터 기증자 참여로 이뤄진 연구 결과가 나왔다. 2021년 가을, '선거 추천: 유튜브 알고리즘은 연방의회 선거에 대해 무엇을 보여줄까?'라는 주제로 6주간의 파일럿 프로젝트를 진행한 결과다. 데이터 기증자들이 유튜브에서 어떤 언론사 뉴스를 주로 시청하는지 확인한 결과, 독일의 미디어 대기업 AXEL SPRINGER 보수 언론사 DIE WELT가 유튜브 뉴스 섹션을 주로 차지해 유튜브에서 보수 언론이 더 자주 노출된다는 점을 알 수 있었다.

소셜미디어 알고리즘은 로그 데이터 기반으로 이용자의 이용 방식을 추적해 추천하지만, 각 이용자의 이용 방식은 개인정보여서 연구 데이터로 활용하기 어려웠던 점을 '기부'라는 방식으로 전환한 것이다.

〈그림 4〉 데이터스콕의 틱톡 대상 데이터 기부 안내



알고리즘와치는 2023년 1월, 두 번째 프로젝트로 틱톡 이용 방식을 분석하기로 하고 틱톡 이용자들에게 데이터 기부를 안내했다(〈그림 4〉 참조). 알고리즘와치는 “페이스북, 인스타그램, 구글, 틱톡과 같은 기업들은 알고리즘 시스템의 유해한 영향에 대해 알고 있으면서도 이에 대한 독립적인 연구를 계속 막고 있는 경우가 많다. 데이터스콕과 같은 데이터 기부는 불투명한 알고리즘을 조사할 수 있는 몇 안 되는 방법 중 하나다”고 주장했다⁶⁾.

6) <https://algorithmwatch.org/en/what-tiktok-knows-about-you-data-donations/>

알고리즘의 투명성을 높이기 위해서는 이용자에게 알고리즘이 어떻게 작동하는지 살펴볼 필요가 있다는 취지의 활동이다.

(2) 언론의 알고리즘 취재 지원

알고리즘와치는 언론 지원 프로젝트를 2022년 9월부터 시작했다. 44만 유로 예산 중 35만 4,000유로는 독일 우정 사업 본부(DEUTSCHE POSTCODE LOTTERIE)에서 후원한다. 이 프로그램은 유럽의 언론인 5명을 선발해, 급여와 교육을 지원해준다(〈그림 5〉 참조). 언론인 알고리즘 이해하고 저널리즘 깊이 수행하도록 교육하는데 주안점을 둔다. 이 펠로십의 목표는 다음과 같은 두 가지이다⁷⁾.

첫째, EU의 자동화 시스템에 대한 보도를 늘린다. 자동화된 결정 시스템에 대해 취재가 미치지 못하는 부분을 더 조사하고, 피해자들의 이야기를 언론인들이 듣는 것이 목표이다. 둘째, 이러한 이야기를 보도하면서 얻은 경험을 공유한다. 알고리즘의 책임성은 어렵고, 모든 언론인이 복잡하고 불투명한 자동화 시스템을 감당할 수 없기에 방법론과 팁을 공유한다.

〈그림 5〉 알고리즘 책무성 보도 지원 프로젝트 안내



7) <https://algorithmwatch.org/en/call-for-fellowships-accountability-reporting/>

이러한 프로그램은 언론인이 기존의 취재 관행에서 미처 다루지 못하는 알고리즘이 이용자에게 끼치는 영향을 취재할 수 있도록 지원을 해 준다는 의미가 있다. 또, 알고리즘의 투명성과 저널리즘의 책무성을 기업이나 언론 등 어떤 하나의 주체에 맡기는 것이 아니라, 함께 논의할 수 있는 공론장을 만들어준다는 의의가 있다. 언론인이 알고리즘을 제대로 이해해 양질의 기사를 만들어내고, 이용자에게 알고리즘 환경에서 고품질의 뉴스를 소비할 수 있는 기반을 마련해주겠다는 의미이다.

구글 역시 구글뉴스이니셔티브(Google News Initiative, GNI)를 운영하며 인공지능 기술을 접목한 취재 기법 개발을 돕거나 교육과 지원을 하고 있다. 구글뉴스이니셔티브는 2018년 시작해 연합뉴스 등 국내 매체가 혁신 프로그램 지원을 받았다. 구글 비즈니스 워크샵은 다양한 언론의 디지털 성장을 돕는 목표로 무료 온라인 교육 프로그램을 운영하고 있으며, 2023년 한국에서도 데이터 검색, 광고 수익화 등의 세미나를 개최했다.

4. 나가며

일상적으로 사용하는 플랫폼에서 인공지능 알고리즘의 역할이 확대되며, 알고리즘 저널리즘은 단지 인공지능 기술을 뉴스 생산 현장에 적용시키는 정도로 논의하기 어려운 상황이 되었다. 네이버와 같은 포털 사이트는 물론이고, 유튜브 페이스북 인스타그램 틱톡 등 알고리즘이 주도하는 추천시스템은 언뜻 이용자를 편안하게 맞춰주는 것처럼 보이지만, 저널리즘의 관점으로 보았을 때 과연 뉴스를 자동화된 의사 결정 시스템으로 추천 받는 게 보다 공정한지 질문이 제기된다.

중요하고 깊이있는 뉴스를 확인해야 하는 이용자로서는 도리어 정보에 소외되는 결과를 낳고 있다. 이용자는 관심사 기반의 뉴스를 반복적으로 접하며 편향적으로 정보를 소화할 우려가 생기고, 알고리즘 주도 환경에 휩쓸려 질적으로 저하된 기사를 쏟아내는 언론사 뉴스에서 중요한 뉴스를 찾지 못할 수도 있다. 심지어 '가짜 정보'와 진짜 정보를 구분하지 못하거나, 사실과 의견을 분리하지 못해 자칫 균형을 잃은 담론에만 귀를 기울일 수도 있다. 이런 혼란 때문에 젊은층은 뉴스를 회피하는 경향마저 발생했다.

이러한 현상은 국내의 문제만은 아니지만, 외국의 몇몇 사례들을 탐색하였다. 미국 영국 독일 등에서 AI관련 비영리단체들이 목소리를 내며 이용자 보호에 앞장서고 있었다. 정부나 기업 뿐 아니라, 알고리즘 저널리즘의 주체인 언론사와 이용자가 함께 참여해 바람직한 생태계를 조성하기 위한 프로그램들이 눈여겨볼만 했다. 구체적으로는 독일 알고리즘와치에서 이용자들이 자발적으로 데이터를 기부하는 시스템을 만들어 소셜미디어 로그 데이터를 분석할 수 있는 데이터스코프 프로그램을 운영하며 알고리즘의 투명성을 높이는 연구 환경 조성에 나섰다. 또한, 독일 우정사업본부의 지원으로 알고리즘 관련 보도를 지원하고 노하우를 공유하며 공론장을 만드는 언론인 지원 프로그램도 운영을 시작했다.

이러한 사례들을 바탕으로 이용자를 알고리즘 저널리즘 수행의 주체로 인식할 수 있는 방안이 필요하다. 그동안 알고리즘 정책의 과정에서 소외되어왔던 이용자의 의견을 선제적으로 수렴하는 방안이 필

요할 것으로 보인다. 플랫폼 기업들이 이용자의 로그 기록을 알고리즘을 작동시키는 자료로 활용해왔다. 알고리즘 생성 과정에서 이용자는 뉴스 이용 결과를 보여주는 데이터로만 기능할 뿐이다. 이제 알고리즘 저널리즘 시대에 걸맞는 뉴스 가치를 사회적으로 합의하는 과정에 이용자가 참여하도록 공청회를 개최하는 방안을 마련해 의견을 수렴할 수 있을 것이다. 예컨대, 현재 네이버의 뉴스 알고리즘 검토는 주로 학자들 중심으로 이뤄지나, 이 과정에도 시민과 언론인을 참여시키는 방안을 검토해볼만하다. 기술은 사회적으로 형성되는 동시에 사회를 형성해 간다는 시각(“SOCIALLY SHAPED AND SOCIALLY SHAPING”)이 필요하다(BUCKINGHAM, 2008, P.12).

이용자 대상의 알고리즘 저널리즘 리터러시 교육을 확대한다. 기존의 리터러시 교육 주체인 각 위원회, 재단, 신문사를 비롯해 알고리즘 기반의 플랫폼사들이 참여하는 형태로 교육의 주체를 출범시킨다. 이들이 뉴스 리터러시 교육과 캠페인을 운영해 알고리즘 저널리즘 생태계를 발전시키는데 함께 기여할 것을 제안한다.

퀄리티 저널리즘을 위한 언론사-이용자-플랫폼사-학계의 선순환 협업 시스템을 마련하는 방안을 검토해볼만하다. 현재 알고리즘 저널리즘 현장은 생산과 소비가 분리되기 어려운 구조이지만, 각 주체들은 개별적으로 존재하고 있는 상황이다. 플랫폼사의 후원으로 언론사의 퀄리티 저널리즘을 후원하고, 시민들은 데이터를 기부해 뉴스의 현황을 연구할 수 있도록 저널리즘의 한 영역으로 자리잡는 방안이다.

참고문헌

- (2022.9.15.). 뉴스 '투명성' 높이는 네이버...알고리즘 검토위 권고 개선 착수.<뉴시스>. URL: https://mobile.newsis.com/view.html?ar_id=NISX20220915_0002014185 (최종방문 2022. 12. 12.)
- 공성윤(2022.11.21.). 알고리즘으로 흥한 페이스북, 알고리즘으로 흔들리다[페이스북 분석]. <시사저널>. URL: <https://www.sisajournal.com/news/articleView.html?idxno=250789> (최종방문 2022. 12. 12.)
- 금준경(2021.3.17.). [단독] 네이버 알고리즘 검증 기구 꾸린다. <미디어오늘>. URL: <http://www.mediatoday.co.kr/news/articleView.html?idxno=212448> (최종방문 2022. 12. 12.)
- 김인식·김자미(2021). 유튜브 알고리즘과 확장편향. <한국컴퓨터교육학회 학술발표대회논문집>, 25권 1호(A), 71~74.
- 김영석·권상희·김관규·김도연·나은영(2017). <디지털시대의 미디어와 사회>. 나남.
- 박세원(2021.1.7.). “유튜브는 독 든 음식도 내주는 식당”...개발자의 고백 [이슈&탐사]. <국민일보>. URL: <http://news.kmib.co.kr/article/view.asp?arcid=0015402057&code=61151111> (최종방문 2021. 9. 1.)
- 박수현(2022.08.19. 08:28). 페이스북 국내 월 이용자 감소세... 2년간 25% 줄어. <조선비즈>. URL: <https://biz.chosun.com/it-science/ict/2022/08/19/663AIONOD5DYHLWDLNURK43FPM/> (최종방문 2022. 12. 12.)
- 박지호(2021). Week 31 - 이루다 같은 챗봇은 어떤 원리로 작동하는 걸까? URL: <https://jiho-ml.com/weekly-nlp-31/> (최종방문 2022. 12. 12.)
- 양선희(2020). 유튜브 저널리즘 시대, 전통적 저널리즘의 대응현황과 과제. <사회과학연구>, 31권 1호, 245~262.
- 양정애(2021.2.23.). 유튜브 이용자들의 ‘유튜버’에 대한 인식. <미디어이슈>. 7권 1호.
- 양형모·박주연(2020). 신문기자의 유튜브 1인 방송 콘텐츠 생산의 특징과 저널리즘 역할 인식 연구, <커뮤니케이션학 연구>, 28권 1호, 33~59.
- 오세욱(2016). 미디어로서의 봇(bot). <한국언론정보학보>, 79, 70~103
- 오세욱(2016). 저널리즘과 알고리즘의 융합에 대한 탐색적 연구. <사이버커뮤니케이션학보>, 통권 제33권 3호, 51~101.
- 유봉석·최재호·최창렬(2020). 네이버 뉴스 알고리즘 이렇다. <관훈저널>, 겨울호, 48~56.
- 이국현(2019.4.2.). 네이버, 뉴스 자체편집까지 ‘중단’...편향성 논란 중지부 찍나?.<뉴시스>. URL: https://newsis.com/view/?id=NISX20190402_0000607387 (최종방문 2021. 9. 1.)

- 이재원(2018). <소셜 미디어의 사회연결성과 이용자 참여 연구: 페이스북 토픽의 활성화 과정을 중심으로>. 이화여자대학교 언론홍보영상학과 박사학위 청구논문.
- 이재원(2019). 소셜 미디어의 토픽 활성화에 관한 연구 : 페이스북의 사회연결성을 중심으로. <한국방송학보>, 33(1), 121~160.
- 이재원(2020). 포털 사이트의 인공지능 뉴스 큐레이션 도입과 뉴스 생산 관행 변화에 관한 연구 : 네이버 연예뉴스를 중심으로, <방송통신연구>, 113호, 93~121.
- 이재원(2021). <알고리즘 저널리즘>. 커뮤니케이션북스.
- 이재원(2023). 인공지능 알고리즘 생태계에서 뉴스 생산자의 뉴스 가치 변화에 관한 연구, <언론과사회> 2023. 9. 15., 기획세미나 발표문.
- 이재원·양정은(2021). 연예 뉴스 생산자의 유튜브 저널리즘 인식에 관한 탐색적 연구 : 연예 매체 채널의 콘텐츠 전략을 중심으로, <정보사회와 미디어>, 22(3), 29~54.
- 정정주·김민정·박한우(2019). 유튜브 상의 허위정보 소비 실태 및 확산 메커니즘 생태계 연구 : 빅데이터 분석 및 모델링을 중심으로. <사회과학 담론과 정책>, 12(2), 105~138.
- 정철운(2019, 7, 28). '유튜브 저널리즘'의 시대가 오고 있다. <미디어오늘>. URL: <http://www.mediatoday.co.kr/news/articleView.html?idxno=201463> (최종방문 2021. 9. 1.)
- 최선영·고은지(2018). 넷플릭스 미디어 구조와 이용자 경험-행동경제학 관점에서 본 이용자와의 관계 맺기. <방송문화연구>, 30권 1호, 7~42.
- 최순욱·오세욱·이소은(2019). 딥페이크 영상 탐지 관련 기술 동향 연구. <한국정보과학학회 학술발표논문집>, 724~726.
- 최진호·박영흠(2022). <디지털 뉴스 리포트 2022 한국>. 한국언론진흥재단.
- 타이디, 조. (2023. 7. 18.). 인터넷감시재단, AI 정상회의서 '아동 성학대 문제' 논의해야, <BBC코리아>. <https://www.bbc.com/korean/articles/cjrlnj2vg48o>
- 한국언론진흥재단(2021). 인터넷 기반 플랫폼을 언론으로 인식하는 비율. URL: <http://hannun.or.kr/2021/3-5/> (최종방문 2022. 12. 10.)
- 한국언론진흥재단(2021). 전통매체 뉴스 이용률 및 인터넷 뉴스 이용률 추이 (2011~2020). URL: <http://hannun.or.kr/2021/3-2/> (최종방문 2022. 12. 10.)
- Anderson, C. W.(2013). Towards a sociology of computational and algorithmic journalism, *New Media & Society*, 15(7), 1005-1021.
- Barot, T. (2015). The botification of news. *NiemanLab*. URL: <http://www.niemanlab.org/2015/12/the-botification-of-news/>
- Burgess, J. & Green, J.(2018) *YOUTUBE-Digital Media and Society Series*, Polity

- Bruns, A.(2008). *Blogs, wikipedia, second Life, and beyond : From production to produsage*.
New York: Peter Lang Publishing.
- Cohen, S., Hamilton, J., & Turner, F.(2011). How computer scientists can empower
journalists, democracy's watchdogs, in the production of news in the public interest.
Communications of the ACM, 54(10), 66-71.
- Dörr(2016). Mapping the field of algorithmic journalism, *Digital Journalism*, 2016, Vol. 4, No. 6,
700-722.
- Hamilton, J. T., and Turner, F. (2009). *Accountability through Algorithm*. Center for Advanced
Study in the Behavioral Sciences Summer Workshop.
- KISO(사)한국인터넷자율정책기구, (2015. 3. 27.), 영국, 인터넷 감시재단(IWF, Internet Watch
Foundation), <KISO저널>, 제18호, <https://journal.kiso.or.kr/?p=6079>
- Kowalski, R. (1979). Algorithm = logic + control. *Communications of the ACM*, 22(7), 424-436.
- Manovich, L. (2013). *Software Takes Command*. New York & London & New Delhi & Sydney:
Bloomsbury.
- Park, P., Lee, J., & Choi, J.(2017). Deep Neural Networks for News Recommendations.
Proceedings of the 2017 ACM on Conference on Information and Knowledge
Management, 2255~2258.
- van Dijck, J. (2013). *The Culture of Connectivity: A Critical History of Social Media*. New York:
Oxford University Press.

인공지능 팩트체크와 '사실성'의 기술사회학적 정의

이정현(중앙대) · 박소영(조선대)



인공지능 팩트체크와 '사실성'의 기술사회학적 정의

이정현 (중앙대학교)

박소영 (조선대학교)

1. 서론

정보의 사실관계를 파악하는 것은 저널리즘의 오랜 관행이었다. 하지만 디지털 환경 속에서 오정보, 허위조작정보, 유해정보 등 오염된 정보가 대량으로 생산되어 빠르고 광범위하게 유통, 소비되는 상황은 '팩트체크(fact-checking)'을 저널리즘의 책임이자 실천으로 급부상하게 했고, 팩트체크의 대상으로서 '가짜뉴스(fake news)'는 오정보, 허위조작정보, 유해정보, 정치적 저항담론을 통칭하는 하나의 기호이자 담론으로 자리 잡았다. 동시에 팩트체크의 과정을 알고리즘이나 인공지능 기술을 통해 자동화하려는 시도가 나타났다. 국내에서도 2010년 후반부터 훈련 데이터를 기반으로 한 증거 추출을 통해 내용의 참, 거짓을 판별하는 인공지능 팩트체크 개발 연구가 정부 지원과 함께 진행되어왔다. 하지만 빠른 속도로 발전하는 알고리즘과 인공지능 기술이 점점 더 정교한 방법으로 가짜뉴스의 생산과 유통에 가담하고 있는 현실에 반해 인공지능 팩트체크 기술은 답보상태에 머물러 있다.

국내 인공지능 팩트체크 연구는 정부 주도의 지원 사업을 통해 설계되고 운영되었다. 국내에서도 온라인 상의 허위정보가 사회적, 정치적 위협으로 대두하면서 정부는 관련 사업에 상당한 예산을 편성하고 대책을 마련하기 위한 정책적인 움직임을 시작했다. 2019년에 방송통신위원회 및 산하 기관인 시청자미디어재단을 중심으로 '인터넷 환경의 신뢰도 기반 조성' 사업이 기획되었고, 이 사업의 일환으로 SNU 팩트체크센터의 이준환 교수 연구팀은 2021년부터 한국형 자동화 팩트체크 모델 개발에 착수했다. 이 모델은 한국어 위키피디아와 인터넷 뉴스를 바탕으로 주장과 근거 중심의 학습데이터를 8만여 건 구축하고, 입력한 주장에 대해 데이터를 바탕으로 참, 거짓을 판별하는 서비스이다. 개발된 모델은 데이터셋과 함께 인공지능 기반 자동화 팩트체크 서비스인 AINET에 탑재되어 시범 운영되었으나 모델의 정확도 및 범용성을 확보하는 데 한계를 해결하지 못하고 2023년 초 시범 서비스가 중단되었다.

국내의 인공지능 팩트체크 연구가 처했던 현실적인 어려움은 모델이 참, 거짓을 가리는 데 판단 근거가 될 양질의 한국어 훈련데이터를 확보하기가 어렵고 학습데이터로 활용하기에 마땅한 온라인 자료는 저작권 문제가 원만하게 해결되지 못했던 점이 꼽힌다(박소영·이정현, 2023). 하지만 국내의 인공지능

팩트체크 기술이 진전되지 못하고 있는 이유는 기술이나 연구 환경의 한계에만 있는 것이 아니라 인공지능 팩트체크에 대한 사회적 청사진이 정쟁의 도구로 기호화된 ‘가짜뉴스’의 담론 안에서 내용의 참, 거짓만을 가리는 기술적 도구로서 설계되었기 때문임을 무시할 수 없다. 즉, 온라인 상에서 오정보, 허위조작정보, 유해정보가 생산되는 방식과 유통되는 과정에 대한 다각적인 이해없이 ‘사실(fact)’에 반대되는 개념으로서 ‘가짜뉴스(fakenews)’를 정의하고 담론화해 온 현실 위에서 인공지능 팩트체크 기술도 설계되었다. 사실상 저널리즘에서 ‘사실’이라는 것은 객관적인 실재에 바탕하여 존재하기 보다는 사회적 사실을 구성하는 과정에서 다양한 사실성 관행을 통해 수립되어 왔다. 팩트체크 역시 총체적으로 구성된 사실에 대해 복수의 검증자들이 합의하는 과정을 통해 사실을 (재)구성하고 (재)확인하는 과정을 거쳐왔다. 하지만 인공지능 팩트체크의 접근법은 저널리즘이 구성하는 ‘사실’에 대한 명확한 숙고 없이 ‘가짜뉴스’ 퇴치를 위한 방법론으로서 국가적으로 설계되고 장려된 경향이 크다.

이 글은 팩트체크 ‘기술’로서 인공지능이 ‘사실’을 구성하는 방식을 국내외 기술사례를 통해 검토함으로써 인공지능 팩트체크가 구성하는 ‘사실성’이 기술사회학적으로 어떻게 정의될 수 있을지 탐구한다. 가장 먼저 인공지능 팩트체크의 맥락을 이해하기 위한 이론적 자원으로서 ‘가짜 뉴스’의 사회적 의미화 과정과 담론 구성 방식을 살펴보고, 가짜뉴스에 대한 대응책으로서 팩트체크가 어떻게 ‘사실성’의 역할을 다루어 왔는지 검토한다. 이어서 국내외 인공지능 팩트체크 기술의 주요 연구 사례가 어떤 방식으로 ‘사실’을 판정하거나 그 과정에 기여해왔는지 살펴보고 그 과정에서 ‘사실’이 기술사회적으로 구성되는 방식을 살펴본다. 마지막 장에서는 결론을 대신하여 ‘내용으로서 사실성’과 ‘구조로서 사실성’이 인공지능 팩트체크 기술 개발에 시사하는 바를 정리하고 각각의 정의가 ‘가짜뉴스’에 대한 사회적 담론과 공모할 가능성에 대해 비판적으로 서술한다. 이는 인공지능이라는 기술과 사회적 담론 사이의 공모 관계를 보다 가시적으로 드러내고 급변하는 저널리즘 생태계의 (비)가시적 구조를 보다 적극적으로 반영한 ‘팩트체크’의 가능성을 검토함으로써 인공지능과 저널리즘이 공진화할 수 있는 방안을 탐구하는 데 이 연구의 의의가 있다.

2. 문헌검토 : 인공지능 팩트체크의 맥락

1) 담론으로서 ‘가짜’뉴스

가짜뉴스는 보편적인 개념이라기 보다는 경합 중인 담론에 가깝다. 가짜뉴스는 의미에 대한 정확한 합의를 거치기 전에 전지구적인 용어가 되었다. 특히 가짜뉴스는 용어를 사용하는 주체에 따라 전혀 다른 맥락에서 사용되어 왔는데 이는 가짜뉴스가 정확히 무엇을 지칭하는 것인지에 대한 사회적인 혼란을 초래했다. 일례로 미국 법무부 특별검사팀은 2016년 미국 대선 당시 러시아 정보기관이 도널드 트럼프 전 대통령에게 우호적인 방향으로 인터넷 여론을 조작해 선거에 영향을 미치려 한 사건을 수사한 보고서를 공개했는데(Mueller, 2019), 이 때 여론 조작을 목적으로 생산된 러시아발 ‘가짜뉴스’를 지적하는 입장

에 대응하여 트럼프 전 대통령은 기성언론 중 자신에게 비판적인 논조의 언론보도를 '가짜뉴스'로 분류하고 명명했다. 이후 페이스북(Facebook, 현 Meta)나 트위터(Twitter) 등 소셜미디어에서는 '가짜뉴스'와의 전쟁을 선포하며 정보의 진위 여부를 기계적으로 가리는 자사의 팩트체크 알고리즘을 홍보했다. 여기서 '가짜뉴스'라는 동일한 기표를 사용하지만 각각이 정의하는 가짜뉴스의 요소는 상이했다. 사실과는 다른 오정보, 근거를 찾을 수 없는 허위정보, 의도적으로 조작된 정보뿐 아니라 특정인이나 집단에게 유해한 정보 등이 정확히 구분되지 않고 '가짜뉴스'라는 기표에 느슨하게 포괄되었다.

실제로 많은 언론인과 학자들이 가짜뉴스라는 용어 사용 자체에 대한 문제를 지적해 왔다. 밥 우드워드(Bob Woodward) 워싱턴포스트 기사는 가짜뉴스라는 용어가 트럼프 전 대통령이 언론에 대한 신뢰를 저해시키고자 정치적으로 악용하고 있음을 지적하며 용어의 폐기를 제안했다(정철운, 2019. 9. 26). 이는 '가짜뉴스'라는 용어가 특정 집단이나 정치 세력이 자신과 반대되거나 자신에게 부정적인 보도를 폄하하거나 공격하기 위해 정치적으로 사용하며 초래하는 사회적인 혼란과 용어의 부적절한 사용이 언론의 신뢰성에 가져오는 치명적인 위기에 대한 경각심을 표한 것이었다. 학계에서도 가짜뉴스 자체는 온라인과 미디어 상에 광범위하게 유포되는 오염된 정보를 통칭하는 개념으로 정의하고(Marwich & Lewis, 2017), 이를 허위(false) 여부와 해를 끼칠 의도(intent to harm)를 판단하여 오정보 또는 단순허위정보(misinformation), 허위조작정보(disinformation), 유해정보(mal-information)로 보다 엄밀하게 구분하여 사용할 것을 제안했다(Wardle & Derakhshan, 2018). 유사한 맥락에서 유럽연합 유럽진행위원회(European Commission) 고위전문가그룹(high-level group of experts)이 2018년 3월 발간한 보고서에서 '가짜뉴스'라는 용어를 '허위조작정보(disinformation)'로 대체하여 사용할 것을 제안했고, 영국 역시 디지털·문화·미디어·스포츠위원회(The Digital, Culture, Media and Sport Committee)가 2019년 최종 발표한 <허위조작정보와 '가짜뉴스'(Disinformation and 'fake news')> 보고서를 통해 가짜뉴스라는 용어의 폐기를 권고한 이후 허위조작정보를 정부 공식문서에서 사용하고 있다. 유엔(United Nations) 역시 가짜뉴스 대신 실수나 무지, 부주의에 기인한 오정보(misinformation), 공중을 속이거나 기만하려는 의도가 명확한 왜곡되고 날조된 허위조작정보(disinformation), 특정 집단이나 개인을 비방하고자 하는 목적을 가진 차별이나 혐오표현(hate speech)으로 용어를 구분하여 디지털 환경의 정보 오염 현상 전반에 대처하고 있다(United Nations, 2023).

국내의 상황은 다르다. 국내에서는 여전히 가짜뉴스라는 용어를 유연하게 사용하는 행태를 보이고 있는데 이는 결과적으로 '무엇이' 가짜뉴스이고 '누가' 가짜뉴스를 생산하는 주체인지에 대한 사회적 합의에 이르지 못한 채 범람하는 기표를 경험하게 했다. 국내에서 가짜뉴스는 미국 대선 직후인 2016년 후반 '페이크뉴스(fake news)'를 번역하여 소개되었다. 미국에서도 사회적인 용어로 대두했던 만큼 용어의 '수입' 직후에 무엇을 가짜뉴스로 규정할 것인지에 대한 논의가 이루어졌다. 초창기 가짜뉴스의 범주를 정의할 때 가장 중요한 점은 그 정보가 뉴스의 형식을 차용했는가 여부였다(황창근, 2017; 황용석·권오성, 2017). 뉴스의 형식을 강조했던 것은 가짜뉴스를 새로운 사회현상으로 보고 이미 온라인상에 존

재하던 허위사실, 단순 정보보고, 루머 등과 구분하기 위해서 였다. 동시에 언론/비언론이라는 정보 주체에 대한 구분도 강조되었는데(박진우, 2020), 이 역시 제도적 사실검증 과정을 거쳤다고 전제하는 언론의 보도와 대비되는 비언론 집단 발 거짓이나 왜곡된 사실을 가짜뉴스로 보는 경향을 보였다. 이 같은 구분은 사실상 텍스트로서 가짜뉴스의 형태에 집중한 개념화로 형식 안에 담긴 정보의 진위에 대한 접근법을 다루고 있지는 않았다.

한국언론진흥재단의 가짜뉴스에 대한 일반 국민 온라인 설문조사 역시 어떤 ‘형태’의 정보를 가짜뉴스로 인식하는지에 대한 질문이 담겨있다. 이에 대한 대중들의 대답은 정보와 텍스트의 형태 면에서도 가짜뉴스에 대한 이해가 얼마나 포괄적으로 이루어지는지를 보여준다. 한국언론진흥재단이 2017년 3월과 2019년 2월에 각각 실시한 가짜뉴스에 대한 일반 국민 온라인 설문조사에 따르면 2017년에는 언론사가 아니면서 언론보도인 것처럼 꾸민 정보만 가짜뉴스로 인식(오세욱·박아란, 2017)하던 데 반해 2019년 조사에서는 선정적 제목을 붙인 낚시성 기사, 클릭수를 높이기 위해 짜깁기 하거나 동일 내용을 반복 게재하는 기사, 한 쪽 입장만 혹은 전체 사건 중 일부분만 전달하는 편파적 기사 등 기존 언론사들의 왜곡, 과장보도 역시 가짜뉴스로 인식한다는 응답이 80%를 상회했다(양정애, 2019). 심지어 동일한 조사에 따르면 한국의 일반 시민들은 카카오톡과 같은 메신저 서비스를 통해 유포되는 일명 ‘찌라시’나 뉴스 기사 형식을 띤 조작된 콘텐츠, 언론사의 오보를 서로 다른 것으로 인식하지 않고 있었다(양정애, 2019).

용어 자체가 가짜뉴스의 생태계를 정확히 진단하지 못하게 하는 경향 때문에 국내에서도 2018년을 전후하여 가짜뉴스라는 용어를 폐기하거나(김민정, 2018) 해외 문헌처럼 가짜뉴스의 개념을 보다 세분화하여 이해하자는 주장(이향선, 2018; 정은령, 2018)이 대두되었다. 하지만 여전히 대체 용어나 명확한 개념화에 대한 숙고없이 가짜뉴스가 하나의 담론으로 소비되고 있는 실정이다. 일례로 윤석열 대통령이 올 해 4.19 기념식에서 “가짜뉴스가 민주주의를 위협한다”고 발언한 다음 날 문화체육관광부는 <악성 정보 전염병> 가짜뉴스 퇴치 전면 강화라는 보도자료를 배포하여 국가 차원의 가짜뉴스 대응방안을 발표했다. 세부 내용은 언론진흥재단 내에 ‘가짜뉴스 신고상담센터’를 설치하고 가짜뉴스를 유형화하는 작업을 시작한다는 것이다(문화체육관광부, 2023. 4. 20). 이는 가짜뉴스에 새로운 국면을 가져온다. 무엇을 가짜뉴스로 신고 받아 어떻게 상담할 것인지, 가짜뉴스 유형화의 근거는 무엇일지에 따라서 가짜뉴스라는 기표가 지칭하는 의미의 범위는 달라질 수 있다. 뿐만 아니라 이는 지금까지 가짜뉴스가 콘텐츠의 형태에 주목하여 정의되었던 것과 달리 ‘내용’에 집중할 가능성을 내포하는데, 동일한 보도자료에서 “정부 정책 관련 가짜뉴스 사례”나 “정부 신뢰도를 떨어뜨리는 허위·왜곡 보도”에 대응하겠다고 밝힌 대목에서 ‘무엇’을 가짜뉴스로 볼 것인지에 대한 맥락을 엿볼 수 있다.

이처럼 국내에서 가짜뉴스는 온라인 환경에서 생산, 소비, 유통되는 다양한 층위의 오염된 정보를 정확히 구분하지 않고 때로는 서로 이질적인 것을 지칭하면서 사회현상을 지칭하는 몇 개의 담론으로 구성되어 왔다. 이 때 ‘내용’에 주목한 가짜뉴스에 대한 접근법이 대두했는데 다음 장에서 살펴 볼 ‘팩트체크’가 대표적인 예다.

2) '팩트'체크와 사실성의 역할

뚜렷하게 정의되지는 못했지만 오정보, 허위조작정보, 유해정보 등 온라인 상에 급속도로 퍼지는 정보 오염 현상은 저널리즘에 위기를 초래했다. 국내에서 가짜뉴스를 조망하는 시각은 '가짜뉴스 대 전통 언론,' '거짓 대 진실'이라는 이분법적 구도로 진행되었다(정은령, 2019, 50쪽). 즉, 정보의 진위 여부 만큼이나 정보의 주체에 대한 관심도 높았는데 초기에는 비언론에서 뉴스인 것처럼 만든 정보만 가짜뉴스로 인식했던 데 반해(오세욱·박아란, 2017) 점차 선정적 제목을 붙인 낚시성 기사, 클릭수를 높이기 위해 짜깁기 하거나 동일 내용을 반복 게재하는 기사, 한 쪽 입장만 혹은 전체 사건 중 일부분만 전달하는 편파적 기사 등 기존 언론사들의 왜곡, 과장보도 역시 가짜뉴스로 인식하는 경향이 커지면서(양정애, 2019) 저널리즘 자체에 위기를 초래했다. 이는 보도자료나 정보원에 대한 검증없이 그대로 전달하는 '받아쓰기 저널리즘,' 인터넷 상 떠돌아다니는 정보를 거르지 않고 확인 없이 전하는 '카더라 저널리즘'(정수영, 2015), 세월호 오보 참사를 기점으로 정착한 '기레기'라는 경멸적인 호칭(정은령, 2019) 등 가짜뉴스라는 용어가 수입된 2016년을 전후하여 언론의 기사쓰기 행태에 축적된 불신의 영향도 컸다.

일부 언론은 가짜뉴스가 언론 자체에 대한 신뢰도에 영향을 미치는 현상에 경각심을 표하며 언론의 자성적 움직임을 촉구했고 팩트체크는 저널리즘의 신뢰회복을 위한 움직임 중 하나로 나타났다. 국내에서는 2017년이 "팩트체크 저널리즘의 원년"(김선호·김위근, 2017)으로 꼽히는데 2017년 19대 대통령선거를 앞두고 각 언론사들이 팩트체크를 시작하며 이전까지 오마이뉴스, JTBC, SBS 등 일부 언론사나 방송사의 내부 프로그램이나 코너로 존재하던 팩트체크가 언론사 전반에 급속히 확산된 영향이다. 2017년 대선 당시 온라인 포털 네이버(Naver)에서는 20여개 언론사가 참여하는 팩트체크 페이지를 제공했고, 여기서 각 언론사들은 대선 후보자 발언 공약 검증 등을 게시했다. 또한 언론사와 서울대 언론정보연구소가 협업하여 만든 팩트체크 모델인 'SNU팩트체크'도 2017년 출범했다. 당시는 세월호 참사 관련 오보와 권력편향보도, 국정농단 사건에 대한 방송사 간 보도 방식 차이 등에 대한 불만이 축적되면서 언론에 대한 신뢰도가 추락한 시점에서 2016년 후반 미국 대선과 관련하여 가짜뉴스라는 용어가 국내에 본격적으로 소개되기 시작한 시점과도 맞물렸다. 2017년 KBS와 MBC의 총파업 이후 양 사의 사장이 교체되었고 뉴스의 신뢰회복을 제1의 과제로 발표하면서 팩트체크 전담 조직이 신설되고, 2018년부터는 MBC, KBS, MBN, TV조선에서 순차적으로 메인뉴스 시간에 팩트체크 코너가 만들어졌다(이미나, 2018). 이 같은 배경에서 국내 팩트체크는 언론 외부에서 생산되는 검증되지 않은 정보에 대한 대응책이자 반격의 성격으로 저널리즘 내부에서 시작되었다.

팩트체크는 국내외에서 2016년에서 2017년 사이 빠르게 확산된 개념이지만 역사는 더 오래되었다. 기사에서 제공하는 정보의 '진위 여부'를 '확인'한다는 의미에서 팩트체크는 1890년대 말 황색저널리즘과 타블로이드지의 난립 등을 배경으로 중요성이 대두되어(Posetti & Matthews, 2018) 언론 내부에서 1988년 미국의 선거에서 행해지던 언론의 흑색선전에 대한 사실 검증의 필요성을 설파하기 시작하면서

꾸준히 관련 조직과 체계를 구성하고 정비해왔다(최순욱·고문정, 2016). 이는 기존의 저널리즘 보도에 대한 일종의 패러다임 전환이었다. 전통적으로 언론은 객관주의 저널리즘을 규범적으로 수용해 왔다(이나연, 2018). 이는 기자들이 사실과 가치를 분리하여 쓰도록 하는 것인데 정치적 논쟁을 다루는 경우 어느 쪽에도 치우치지 않고 논쟁의 주체들을 공정하게 재현하며, 논평이나 의견을 덧붙이지 않고 ‘뉴스’만 보도하도록 하는 것이다(Schudson, 2001). 따라서 진실(truth)이란 논쟁의 타당성 여부나 논쟁에 대한 언론인의 동의 여부와는 무관하게 “논쟁 자체를 그대로 정확하게 반영(mirroring)했는가에 제한”되는 것으로 인식되었다(Westerstahl, 1985; 정은령, 2018, 55쪽 재인용). 기사가 육하원칙, 역피라미드 구조를 활용하는 점도 간결하고 명확한 사실 전달을 위한 형식적인 원칙 준수의 일환이었다(박재영·이완수, 2008). 하지만 1988년 대선 당시 양 후보 진영이 보여준 정치 광고 등은 객관적으로 전달하기 어려울 정도로 사실성과 정확성이 결여된 흑색선전이었고, 이에 경각심을 느낀 기자들이 정치보도의 맥락에서 누군가 말한 사실을 전달하기만 하는 “방관자”에서 유권자에게 정치적 판단의 근거로 삼을 만한 것을 제시하는 “진실의 판정자”로의 전환을 촉구하면서 팩트체크가 기자 공동체 내부에서 저널리즘 혁신 운동으로서 제시되었다(정은령, 2019, 9쪽).

팩트체크는 기사가 취재한 내용 및 발언의 정확성을 기사가 직접 평가한다는 점에서 객관주의 규범과 대조를 이룬다. 이것은 단순히 인용의 정확성을 확인하는 과정을 넘어서 보도하고자 하는 발언이 정확하지 않거나 기자들이 독자적인 조사를 거쳐 해석한 판정결과를 독자에게 제시하는 것이다(Amazeen, 2013). 언론이 “단순한 전달자가 아닌 책임있는 해설자가 되기를 요구”하는 흐름의 일환으로 이어져 온 팩트체크의 핵심은 화자가 말한 내용을 그대로 옮겨 적는 데 있지 않고 인용하려는 말의 사실 여부를 가리는 데 있다(정은령, 2019, 9쪽). 사실을 전달하는 데 그치지 않고 그에 대한 해석을 제시하는 언론의 경향은 20세기 미국 뉴스의 성격과 양식을 특징짓는 핵심적인 요소이다. 팩트체크 저널리즘은 해석적 전환(interpretive turn)이라 불리는 새로운 저널리즘 맥락에서 뉴스의 수용자에게 사실에 대한 참 거짓의 판정을 제시하는 역할을 수행한다(Barnhurst, 2016).

언론인의 책임으로 부상하고 있는 팩트체크에 대한 공적인 취지는 대개 동의하지만, 팩트체크가 검증해야 할 ‘사실’이 무엇인지에 대해서는 다양하게 해석되고 있다. 즉, 확인해야 할 ‘사실’이라는 것이 무엇이며 무엇을 확인함으로써 사실을 검증할 수 있는지에 대한 논쟁이다. 과학적 명제나 객관적 실재를 탐구하는 것이 아니라 ‘실재의 사회적 구성’에 핵심적인 역할을 담당해 온 언론에게는 ‘참’ 혹은 ‘사실’이라는 것은 제도적 과정을 통해 보장되어 왔다. 예컨대, 말해진 사실의 진위를 밝히기 위해서 르포, 직접인용, 통계 등을 이용하여 사실성을 입증한다(van Dijk, 1988). 또한 개별 사실의 정확성을 확보하기 위해서 정보원을 활용하는 경우 정보원을 보호할 수 있는 범위 안에서 가급적 정보원의 실명을 공개하고 직접 인용하는 것을 원칙으로 하며, 이해관계자가 대립하는 경우 한 쪽으로 편향되지 않도록 균형 보도를 하고 다양한 정보원을 통해 삼각 확인 등의 검증 절차를 거치는 것을 원칙으로 한다(박대민, 2015). 취재 기자, 팀장, 데스크, 편집기자, 편집부장, 편집국장 등 여러 단계를 거치는 게이트키퍼 과정은 기사에 포

함된 사실을 확인하고 이것을 어떻게 연결하고 맥락화하여 총체적 사실로 만들지 확정함으로써 최종적으로 전달되어야 할 사실을 구성해 왔다(김창숙, 2020). 기사가 배포된 후에도 드물지만 정정보도나 반론보도 같은 언론사 내부의 자율 규제, 언론중재위원회나 심의위원회 등 언론사 외부 기관에 의한 강제 규제 등의 사후 조치도 언론의 사실성을 확보하는 기제로서 존재한다(신혜선·이영주, 2021).

이 같은 제도들은 언론이 이미 존재하는 '사실'을 찾는 것이 아니라 복합적인 검증 절차를 수반하여 달성하여 왔음을 보여준다. 객관주의 저널리즘을 기반으로 한 전통 언론의 맥락에서는 팩트체크 저널리즘이 '사실을 검증'한다는 것은 몇 가지 쟁점을 갖는다. 전통적으로 저널리즘이 주관적 의견을 개입하지 않고 객관성과 공정성을 유지하기 위해 양쪽의 입장을 모두 말하는 방식으로 사실성을 확보해왔다면, 팩트체크는 특정 입장의 근거를 확인하고 타당성을 가려 판정을 내리는 것이기 때문에 무엇을 검증하는가에서부터 어떻게 해석하는가에 이르기까지 주관성과 당파성의 개입을 부정하기 어렵다. 우신스키와 버틀러(Uscinski, 2015; Uscinski & Butler, 2013)가 지적한 팩트체크의 한계점에 따르면 사회적 사실을 모두 검증하는 것은 불가능하기에 우선적으로 다루어야 할 사안을 '선택'하게 되는데 이 때 누가 선택하고, 무엇을 선택하는지는 선택의 주체와 맥락에 따라 정치적일 수 밖에 없다. 또한, 사실성을 판단하기 위해 근거 자료를 선택하고 걸러내는 과정도 주관성과 당파성이 개입될 여지가 있기에 사실의 판정이라는 것도 주관적일 수 밖에 없는 한계를 내포한다. 실제로 이런 이유로 동일한 검증대상에 대해서도 상이한 판정결과가 나올 수 있다(Graves, 2017).

팩트체크는 내부적인 준칙과 규정을 통해 이 같은 논쟁에 대처하며 언론의 신뢰성을 회복하고 사실성을 구성하는 하나의 '제도'로 정착하고 있다. 대표적으로 전 세계 팩트체크 기관들의 자율적인 연대기구로 2014년 출범한 국제팩트체크네트워크(International Fact Checking Network: IFCN)은 각국 팩트체커들의 준칙(Code of principles)을 제정하며 팩트체크에서 정치적 불편부당성과 투명성의 준수를 강조했다. 준칙은 구체적으로 다음의 다섯 가지 영역으로 구성되었다: 1) 비정파성과 공정성(Nonpartisanship and Fairness), 2) 취재원 투명성(Transparency of Sources), 3) 재정과 조직에 관한 투명성(Transparency of Funding & Organization), 4) 검증방법의 투명성(Transparency of Methodology), 5) 공개적이고 정직한 수정(Open and Honest Corrections). 이를 바탕으로 IFCN은 각국 팩트체크 기관이 준칙을 준수하는 기관으로 인증을 원할 경우 5개 항목의 실행 여부를 세부항목¹⁾으로 심사해 자격을 부여한다. 2023년 5월까지 114개의 팩트체크 기관과 언론이 인증을 얻었고 한국에서는 2020년 1월 JTBC '뉴스룸'이, 2023년 팩트체크 전문 미디어 '뉴스톱(NewsToF)'이 인증을 받

1) 5개의 기본 준칙에 따르는 세부항목은 1) 어느 한 쪽에 편승된 팩트체크를 하지 않는다, 2) 증거가 결론을 이끌어내도록 한다, 3) 팩트체크하는 이슈들에 관해 어떤 정책적 입장도 취하지 않으며 옹호하지 않는다, 4) 팩트체크의 이용자들이 검증과정을 동일하게 반복할 수 있도록 정보원의 개인적인 안전이 침해당하지 않는 한 최대한 자세히 밝힌다, 5) 외부기관으로부터 재정지원을 받았을 때 지원자가 팩트체크 결과에 영향을 미치지 않는다는 것을 보장해야 한다, 6) 어떻게 검증대상을 선정, 조사하고 결과를 기술하고 편집하여 발행하는지 방법론을 이용자들에게 설명한다, 7) 오류가 있을 경우 공개적으로 투명하게 수정한다 등을 포함하고 있다(<https://www.ifcncodeofprinciples.poynter.org/>).

았다. 이 같은 준칙을 통해 팩트체크가 강조하는 것은 객관성보다는 투명성이다. 예를 들어, 팩트체크 닷오르그(Factcheck.org), 폴리티팩트(Polity Fact), 워싱턴포스트 팩트체커(Washington Post Face Checker) 등 미국의 주요 팩트체크 기관은 홈페이지에 검증대상 선정의 방법, 원칙, 판정기준을 명시적으로 밝히며, 검증에 동원된 정보원의 경우 하이퍼링크로 원자료에 접근할 수 있도록 하고 있다. 즉, 팩트체크 과정의 대상 선정, 검증 과정, 검증 결과에 따르는 편향성 및 객관성 논란에 대해 인용한 정보원을 공개하고, 검증결과를 판정한 이유를 제시하며, 오류수정을 공개함으로써 '투명성을 통한 사실성 확보'를 추구해왔다(정은령, 2019).

이처럼 정보의 사실 여부를 검증하고 검증 과정에 동원된 정보원과 판단 기준을 최대한 투명하게 공개함으로써 저널리즘의 사실성을 추구하고 있는 팩트체크는 가짜뉴스에 대한 '내용' 중심의 접근법에 가깝다. 온라인 상 정보 오염 현상을 이르는 여러 하위 개념을 통칭하는 가짜뉴스가 범람하면서 팩트체크가 대안으로 떠오른 이유도 이 같은 내용의 참, 거짓 판별을 통해 건강한 온라인 저널리즘 생태계에 기여하고자 하는 내부적인 성찰과 외부의 기대 때문이라고 할 수 있다. 하지만 정보의 진위 여부 전에 정보 자체의 양이 많아지는 현실에서 일부 팩트체크 기관이나 소수의 기자가 감당하기에는 팩트체크해야 할 대상이 지나치게 많아지면서 '효율적인' 팩트체킹을 위한 혁신의 필요성이 대두되었고, 인공지능을 활용한 팩트체킹은 인간 팩트체커의 한계를 극복할 수 있는 대안으로서 검토되거나 실행되고 있다. 하지만 인공지능을 도입한 사회의 모든 영역이 그러하듯 인공지능 기술의 도입의 과정과 결과는 예상만큼 단순하지 않다. 팩트체크의 경우 인공지능이라는 기술이 '사실'의 '검증' 과정에 개입함으로써 무엇이 사실이고 무엇이 검증 가능한 것인지에 대한 인식론적 전환을 가져오고 있다.

3. 인공지능 팩트체킹과 기술사회적 '사실'

인공지능 팩트체킹은 인공지능 기술을 팩트체크 과정에 활용하는 것으로 광범위하게 이해되고 있지만 이는 무엇이 인공지능에 의해 자동화 되는가를 기준으로 보다 엄밀하게 구분할 필요가 있다. 실제로 공학적 관점에서는 사실 확인(fact-check)과 사실의 검증(verification)을 구분하고 있는데 여기서 사실 확인은 주장의 논리, 일관성, 맥락 등 보다 총체적인 사실을 고려하여 사실성을 판별하는 것을 의미하며 검증은 출처, 날짜, 장소 등 확인 가능한 증거를 바탕으로 사실의 옳음을 확인하는 제한된 개념을 뜻한다(Thorne & Vlachos, 2018). 또한, 탐지(detection)는 특정한 주장의 의도, 형식, 수용자 반응, 문장의 요소 등 정보의 형태 요소를 파악하여 가짜뉴스나 프로파간다를 탐지하는 것으로 사실 확인과는 구분된다(Guo et al., 2002). 현재로서는 팩트체크의 전 과정을 인공지능에게 위임한 사례를 없으며 아직은 개발을 위한 연구 단계에 머물러 있다. 인공지능 팩트체크(artificial Intelligence Factcheck) 혹은 자동화된 팩트체킹(automated fact-checking)이라는 용어 자체가 주는 기대와는 달리 현재는 전 세계의 몇몇 팩트체크 전문 기관 및 유관 기관에서 팩트체크 과정 중 일부를 인공지능 기술을 통해 자동화하는 방식

으로 이루어지고 있다.

이 때 인공지능 기술의 팩트체크 과정 중 '어디를' 자동화하는지, '무엇을' 근거로 사실을 입증하는지에 따라 팩트체크 저널리즘이 역사적으로 구성해 온 사실성의 지형이 달라진다. 즉, 팩트체크 과정에 인공지능 기술을 도입하는 것을 단순히 기술 구현 가능성에만 의존하여 기술결정론적으로 운용하기에는 인공지능 기술 자체가 사실성의 지형에 차이, 전환, 결과를 만들어내는 힘을 무시할 수 없고 인공지능이 진단하는 대상이 사회적 실재라는 측면에서 결과적으로 저널리즘의 신뢰성에 결정적인 영향을 미친다.

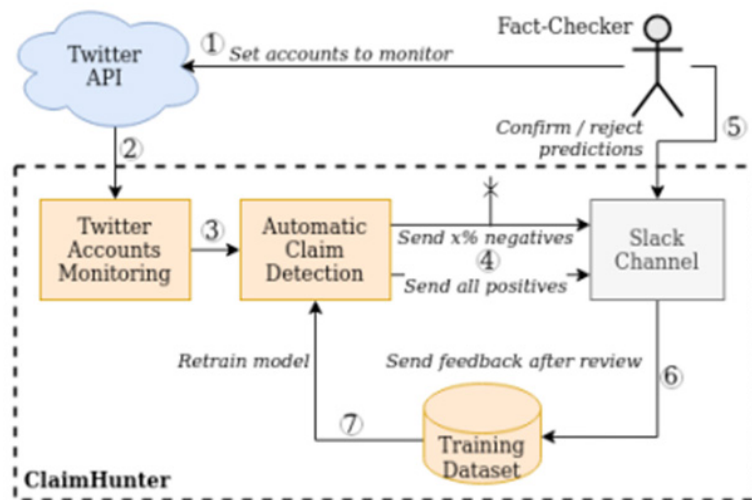
이에 이 논문은 실제 팩트체크 과정에 사용할 수 있는 인공지능 기술을 자체 개발하거나 서비스하고 있는 기술 사례들을 조사하고, 각 기술 사례들의 목적과 사실판단 근거에 따라 유형화하였다. 이를 바탕으로 각 유형이 만들어내는 '사실'의 기술사회학적 의미를 설명했다. 조사 대상은 2023년 8월 기준 국제 팩트체크 네트워크(IFCN)에 의해 인증 받은 세계 유수의 팩트체크 인증기관들과 세계적인 싱크탱크인 미국의 랜드 연구소(Rand Institute)에서 허위조작정보에 대한 대항(fighting disinformation)할 목적으로 구축한 온라인 데이터베이스이다. 국내의 기술은 조사 대상에 포함되지 않았는데 이는 아직 국내에서는 대중 일반에게 공개되어 실제 서비스가 가능한 수준까지 진전을 이룬 인공지능 팩트체크 기술 사례가 없기 때문이다. 앞서 언급한대로 2021년 서울대학교 언론정보학과 이준환 교수 연구팀에서 인공지능 팩트체크 툴을 개발하고 베타서비스 단계에서 대중 일반에 공개한 사례가 있으나 재정 부족과 데이터 부족 등의 문제로 본 서비스 수준으로 나가지 못하고 2023년 초 베타서비스 운영을 중단하였다. 하지만 국내 유일의 개발 사례라는 점에서 아래에 기술 사례를 유형화하는 과정에서 해당 유형에 함께 언급하였다.

1) 주장의 탐지

가장 먼저 무수한 사실 중 검증되어야 하는 '주장(claim)'을 탐지하는 기술 유형이 있다. 이론적으로 주장의 탐지는 주장의 검증 가치를 평가하는 과정을 자동화하는 것인데 일반 공중이 사실 여부를 아는 데 관심을 가질 만한 주장을 판단하여 무수한 정보 중 팩트체크 할 대상을 골라내는 것을 뜻한다. 앞서 서술했듯이 팩트체크에 있어 '무엇을' 검증할 것인가의 문제는 선택 과정에서 시대적, 상황적 맥락과 주관성이 개입될 수 밖에 없어 팩트체크 과정 자체의 편향성 논란을 불러 온 대목이다. 이 '선택'의 과정을 기계에게 맡겨 주관성이 배제된 '객관성'을 확보하고자 하는 것인데, 이를 위해 인공지능 기술의 주장 탐지에는 개인적 경험이나 주관적 감상에 기반한 문장들을 자연어 처리(Natural Language Processing, NLP) 기술을 통해 자동적으로 걸러내는 방법이 제안된다(Konstantinovskiy, Price, Babakr, & Zubiaga, 2021).

대표적인 기술 사례는 클레임 헌터(Claim Hunter)²⁾다. 2020년 스페인의 팩트체크 기관인 뉴트랄(Newtral)은 최고기술책임자인 루벤 미게즈(Rubén Míguez)의 지휘 아래 음성을 듣고 검증이 필요한 주장들을 탐지하는 클레임 헌터(Claim Hunter)라는 도구를 발표했다. 클레임 헌터는 훈련된 인간 팩트체커들이 공적 관련성을 기준으로 지속적인 관심을 기울여야 할 트위터 계정을 선별한 후, 클레임 헌터가 트위터 API를 활용해 선별된 계정에서 트윗을 수집한다. 수집된 트윗은 판별된 가치에 따라 긍정(팩트체크 가치가 있음) 또는 부정(팩트체크 할 만한 가치가 없음)영역으로 분류된다. 긍정 영역으로 분류된 트윗은 인간 팩트체커의 검토와 레이블링 과정을 거쳐 데이터베이스에 포함되고 모델은 이렇게 인간 팩트체커에 의해 정제된 데이터베이스로 학습한다. 학습된 모델은 구글의 BERT기반 딥러닝 모델을 통해 검증할 만한 가치가 있는 주장을 탐지한다. <그림 1>은 클레임 헌터의 작동 원리를 도식화한 것이다.

<그림 1> 클레임 헌터 작동 과정 및 원리



유사하게 미국 텍사스 대학교 알링턴 캠퍼스 네이물 하싼(Naeemul Hassan) 교수의 연구팀은 2017년 주장 탐지에 특화된 클레임버스터(ClaimBuster)라는 인공지능 도구를 오픈소스로 공개했다. 클레임버스터는 소셜미디어 게시물, 정치인 연설로부터 수집된 정보 중 팩트체크할 만한 가치가 있는 주장들을 탐지하는데, 클레임 헌터와 유사하게 검증 가치가 있는 주장들을 사전에 레이블링한 데이터 세트로 훈련한 모델이 지도학습(supervised learning)을 통해 팩트체크 할 주장을 탐지한다. 탐지 결과를 바탕으로 팩트체크가 필요한 트위터 메시지를 리트윗하여 이용자들에게 알리고 있다.

마지막으로 영국 비영리 팩트체크 기관인 풀팩트에서 개발하여 유료로 제공하는 풀팩트 AI (Full Fact

2) <https://ceur-ws.org/Vol-2877/paper3.pdf>

AI)³⁾가 있다. 풀팩트는 2019년부터 아프리카의 아프리카체크(Africa Check) 및 아르헨티나의 체키아도(Chequeado)와 같은 팩트체크 기관, 그리고 영국의 비영리기구인 오픈데이터연구소(Open Data Institute)와 국제적 협력 관계를 통해 기술 개발에 힘쓰며 인공지능 기반 팩트체크 기술 발전의 선두에 서 있다. 풀팩트의 목표는 수많은 정보들 중 팩트체크해야 할 사안의 우선순위를 정하고, 이것이 이미 검증된 것이 아닌지 확인하고, 가능한 한 실시간으로 팩트체크가 가능하도록 전 과정을 자동화하는데 있다. 풀팩트 AI는 TV 생방송, 온라인 뉴스 사이트, 소셜 미디어 페이지로부터 데이터를 수집해 문장 단위의 텍스트로 나누고, 여기서 '주장(claim)'을 추출한다. 여기서 주장은 모든 문장의 '확인가능한(checkable)' 부분으로 정의되며, 통계적 수치, 인과관계, 미래 예측 등이 주로 포함된다. 풀팩트AI는 구글에서 공개한 대규모 언어 모델인 BERT모델을 기반으로 자체 데이터 학습을 통해 모델을 미세 조정하여 추출된 주장들 중 보다 확인할 만한 가치가 있는 주장을 탐지할 수 있는 기술을 개발했다. 추가적으로 탐지한 주장을 자체로 확립한 팩트체크 데이터베이스 안에 있는 주장들과 대조하여 이미 팩트체크된 문장을 걸러낸다.

이와 같이 주장을 탐지하는 팩트체크 자동화 기술은 두 가지 측면에서 살펴 볼 필요가 있다. 첫째는 팩트체크 전 과정을 인간과 기술이 협업하여 달성할 수 있는 가능성을 제시한다는 측면이다. 팩트체크는 검증대상의 선정에서부터 검증과정, 검증결과 공개, 검증결과 수정에 이르는 절차를 전문 기자들이 수행하고 있다. 이에 대한 대안으로 인공지능 기술과 분업을 통한 팩트체크를 수행하는 것인데 이는 팩트체크를 수행 할 인원이 제한적인 데 반해 검증해야 할 정보의 양은 넘쳐나는 상황에서 1차적으로 기계에 의해 주목해야 할 정보를 걸러낸다는 측면에서 효율적인 분업 모델로 제시되었다. 하지만 팩트체크가 사회적 사실을 구성하는 데 결정적인 역할을 하는 저널리즘 도구라는 측면에서 생각해 보면 분업의 항목이 왜 하필 검증대상의 선정인지 생각해 볼 필요가 있다.

검증대상을 선정하는 과정을 인공지능 기술을 통해 수행하는 것은 팩트체크 과정에서 논쟁이 되어 온 '가치 판단'의 영역을 기술에 유보한 경우다. 앞서 서술했듯 팩트체크에서 '무엇을' 검증할 것인가는 선택의 문제라 정치적일 수 밖에 없다는 비판을 받아왔는데 이 부분을 기계적으로 처리하여 팩트체크의 '객관성'을 높이려는 목적이다. 이는 판단 주체에 따라 달라지는 선택의 과정을 기계가 하면 객관적일 것이라는 사회 전반의 믿음이 내재되어 있는데, 이는 인간 팩트체커에게 필연적으로 따르는 '선택 편향', 혹은 선택 편향의 논란을 인공지능 기술을 통해 피해갈 수 있을 것이라는 기대이다. 하지만 이는 결과적으로 '사실 확인이 필요한 주장'에 대한 새로운 맥락을 제시한다. 사실과 가치를 분리할 수 없다는 것을 전제로 한 팩트체크 저널리즘과 달리 주장을 탐지하는 인공지능 팩트체크 기술은 기술적으로 검증할 수 있는 사실만을 추출해 낸다. 주로 개별 사실의 '검증'이 가능한 주장들에 한정되는데 통계적 수치, 인과관계, 예측 위주로 포함된다. 사실상 위와 같은 과정을 통해 인공지능 팩트체크 기술이 추출한 주장을 검

3) <https://fullfact.org/about/ai/>

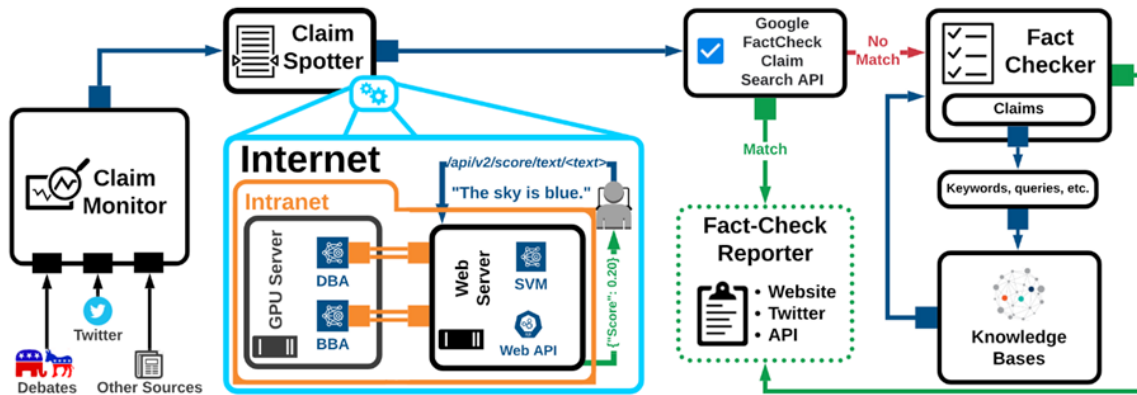
증하는 일도 인공지능 기술에 맡겨진다는 점에서 위와 같은 주장 탐지 기술이 사실 확인의 맥락 자체를 정하는 결정적인 역할을 하는 것이다.

2) 증거 추출과 주장의 진실성 검증

검증해야 할 주장의 탐지에서 나아가 주장 자체의 진실성(truthfulness)을 관련 증거를 바탕으로 확인하는 작업을 자동화한 경우가 있다. 이 경우 인공지능 기술 중에서도 자연어 처리 기술을 핵심적으로 이용하여 자동화되어 왔는데, 기계학습을 통해 정해진 범위의 언어 데이터를 학습하여 모델을 개발하고 이 모델이 새로운 정보나 기사의 사실성 정도를 예측해 분류하는 것이다. 이는 탐지된 주장이나 진술을 뒷받침하거나 반박할 만한 정보들을 수집하는 단계가 수반되는데 대체로 사전, 뉴스 기사, 국가 통계 등 신빙성과 신뢰도를 갖춘 자료에서 증거를 추출하거나 최근에는 팩트체크를 거친 진술을 데이터베이스로 구축한 아카이브를 바탕으로 증거를 추출하는 경우도 있다. 주장에서 추출한 증거들과 대조하여 사실성 여부를 판정하는데 참, 거짓의 이분법적 분류를 하는 경우도 있고 사실의 '정도'를 세분화된 범주를 바탕으로 계산하여 추론하는 경우도 있다.

이 유형의 경우 기술 자체의 작동 과정은 유사하다. 다만 '어떤 근거'를 바탕으로 사실성 여부 및 정도를 확인하였는지가 관건인데 이는 기계적 판정 과정 자체에 대한 의심보다는 판정 근거의 정확성이 결과의 신뢰도를 담보하는 상황을 드러낸다. 앞서 설명한 퍼팩트 AI와 클레임버스터는 주장의 탐지에서 나아가 검증까지 자동화하였다. 각자의 방법으로 검증해야 할 주장을 탐지하고 나서 탐지된 주장의 검증까지 자동화하는 것인데 퍼팩트AI의 경우 가려낸 주장에 나타난 주제, 추세, 값, 날짜, 위치 등을 식별하여 국가 통계 포털 등 공신력 있는 외부 사이트의 정보와 대조하는 방식으로 사실을 검증한다. 클레임버스터의 경우 구글 팩트체크 에이피아이(Google FactCheck Claim Search API)와 울프람 알파(Wolfram Alpha)와 같은 질문 응답 엔진을 이용한다. 먼저 질문 생성 도구가 생성하고 선택한 질문을 질문 응답 엔진인 울프람 알파로 보낸다. 울프람 알파에서 얻은 답변이 추출되고 그동안 동일한 질문이 구글에 HTTP의 형태로 전송되어 구글의 답변을 추출한다. 반환된 답변과 주장 사이 명확한 불일치가 있으면 판정이 유도되어 사용자에게 제시된다. 또 사실로 판정된 주장은 구글에 다시 검색 쿼리로 전송되고 상위 결과의 웹페이지를 다운로드하여 그 안에 주장과 일치하는 문장을 찾는다. 주장과 일치하는 문장과 그 주변 문장이 '컨텍스트'로 묶여 앞서 판정된 주장의 지지 혹은 반박 '증거'로 만들어져 사용자에게 보고된다(〈그림 2〉 참고).

〈그림 2〉 클레임 버스터 팩트체크 과정 도식화



이미 팩트체크 과정을 거쳐 '검증된 사실'의 데이터베이스와 대조하여 사실을 검증하는 방법도 있다. 미국 듀크대학교 빌 아데어(Bill Adair) 교수의 리포터스 랩(Reporter's lab)은 2021년 스퀘시(Squash)⁴⁾ 라는 플랫폼을 개발했다. 스퀘시는 정치 토론이나 연설 등 주요한 정치적인 이벤트에서 나오는 발언과 주장에 대해 실시간으로 자동화된 팩트체크 서비스를 제공하는 것을 목표로 만들어진 플랫폼이다. 스퀘시는 특정 이벤트에서 발화된 정치인의 발언을 문자로 변환하고, 앞서 언급한 클레임버스터를 활용해 검증가치가 떨어지는 주장들을 일차적으로 거른다. 이후 클레임리뷰(ClaimReview)⁵⁾와 같은 팩트체크 데이터베이스와 대조해 가장 주장과 관련성이 높은 정보를 추천한다. 이와 같은 자동화 과정을 통해 팩트체크가 완료된 정보는 훈련 받은 인간 편집자에 의해 최종적으로 선별되고 및 이용자가 볼 수 있도록 제시된다.

사실의 검증에 있어 가장 널리 알려진 모델은 2018년 제임스 손(James Thorne, 현 카이스트 교수)을 대표로 영국의 쉐필드 대학교와 아마존의 영국 캠브리지 소재 연구소의 연구팀이 개발한 피버(FEVER, Fact Extraction and VERification)⁶⁾다. '사실 추출과 검증'이라는 이름이 가리키듯, 피버 모델은 위키 피디아에서 수집한 자료, 자연어 처리, 인간의 레이블링을 거친 데이터를 복합적으로 활용해 만든 벤치마크 데이터셋으로 모델을 훈련시키고, 모델이 입력된 주장에 대한 증거를 추출하여 참/거짓을 판별하는 과정을 자동화한다. 훈련을 위한 데이터셋은 '어노테이터(annotator)'라고 불리는 훈련된 인간 팩트

4) <https://reporterslab.org/tech-and-check/>; <https://reporterslab.org/tag/squash/>

5) 클레임리뷰(ClaimReview, <https://www.claimreviewproject.com/>)는 스키마닷오알지(Schema.org)의 댄 브리켈리(Dan Brickley)와 구글의 저스틴 코슬린(Justin Kosslyn), 빙(Bing), 지그소(Jigsaw)가 듀크대학의 리포터스랩과 협력하여 만든 시스템으로 팩트체크 기사를 기술적으로 표준화하여 누구나 보편적으로 접근하고 활용하도록 한 마크업하여 제공한다. 이 마크업 시스템은 구글 팩트 체크 도구 페이지에 공개되어 있다. 클레임리뷰에 앞서 구글은 자사의 팩트체크 도구 페이지에 전세계 팩트체커들이 검증한 사실을 아카이빙하여 데이터베이스를 구축하고 API를 통해 이용자에게 공개하고 있다 (<https://toolbox.google.com/factcheck/explorer>). 최근에는 유사한 형식으로 팩트체크가 완료된 이미지, 영상, 오디오 데이터를 표준화 하고자 하는 미디어리뷰(MediaReview) 프로젝트 또한 듀크대학 리포터스랩에 의해 개발 중에 있다.

6) <https://arxiv.org/pdf/1803.05355.pdf>

체커들이 185,445 건의 샘플 문장(주장)을 다듬고 여기에 대해 사실 여부의 판단 및 증거를 추출해 레이블링하는 과정을 거쳐 구축하였으며, 여기서 사실 판단 및 추출된 증거는 다른 어노테이터들에 의해 교차 검증되었다. 실제 팩트체크가 수행될 때는 팩트체크하려는 문장을 입력하면, AI 모델이 지식 베이스가 되는 데이터셋에서 입력된 문장과 관련성이 높은 문서와 문장을 선택 추출하며, 이것이 입력된 문장과 얼마나 일치하는지에 따라 지지(supported)와 반박(refuted) 여부를 판정한다. 만약 구축된 지식 베이스에서 입력된 문장을 지지하거나 반박할 수 있는 내용을 찾지 못하는 경우 ‘충분한 정보 없음(not enough info)’이라는 판정을 내리게 된다. 피버 모델은 팩트체크 자동화를 위한 모델 뿐 아니라 팩트체크의 기반이 될 훈련 데이터셋 구축의 한 방식을 제안하고 있으며, 상당한 후속 연구가 피버 모델을 채택해 이를 기준으로 새로운 모델의 팩트체크 수행 능력을 기를 수 있다고 있다. 서울대 이준환 교수 연구팀이 개발한 국내 유일의 팩트체크 AI 서비스 또한 피버 모델을 채택하여 개발되었다.

주장의 탐지와 달리 주장을 검증하는 일은 그 결과가 이용자에게 공개된다는 측면에서 인공지능 팩트체크 기술의 가장 핵심적인 과정이다. 팩트체크 과정 중 시간이 가장 오래 걸리는 일을 기술적으로 ‘처리’하는 것인데 이 때 판정의 근거가 되는 데이터셋과 판정의 결과를 공개하는 방법이 ‘사실’을 구성하는데 핵심적인 역할을 한다. 현재 인공지능 팩트체크 기술이 판정근거로 활용하고 있는 데이터셋은 위키피디아와 같은 웹 기반의 협업형 백과사전, 분야 별 전문 사전, 뉴스 기사, 통계를 바탕으로 초기 모델을 훈련하며 이후에는 팩트체크가 완료된 정보를 데이터베이스로 축적하여 이를 훈련데이터로 활용한다. 결과적으로 인공지능 기술이 주장과 근거자료를 대조하여 사실을 판단한다는 측면에서 무엇이 훈련데이터로 사용되는지가 사실을 구성하는 데 결정적인 역할을 한다. IFCN이 권고하는 팩트체크 준칙은 이용자들이 검증과정을 동일하게 반복할 수 있도록 정보원을 최대한 자세히 밝히고, 검증대상의 선정, 조사, 결과 기술에 이르는 전 과정을 설명하도록 하고 있다. 때문에 위에서 서술한 주장 탐지 기술들은 판정의 근거를 클레임 리뷰와 같은 마크업 시스템이나 별도의 대시보드를 통해 공개하고 있다.

하지만 모델을 학습시킬 만큼 대용량의 데이터이면서 ‘사실’을 검증할 수 있을 만큼 정확성이 있는 데이터가 충분하지 않다는 것이 한계다. 우선 대용량의 데이터이기 위해서는 사용자가 많거나 인구가 많은 언어로 적혀진 데이터일수록 유리한데 이는 사실상 영어를 제외한 언어나 영어로 번역될 수 없는 사실에 대한 데이터셋 확보가 한계에 부딪힐 수 밖에 없음을 시사한다. 또한 기술의 작동 방식이 근거 자료와 주장을 대조하여 이루어진다는 점에서 근거 자료는 사전에 사실성이 검증된 데이터이어야 하는데 그 범위가 매우 협소하다. 앞서 소개한 기술 사례들이 활용한 사전, 위키피디아, 통계자료, 뉴스기사 등이 대표적이거나 백과사전의 경우는 대체로 저작권법에 의해 보호되는 저작물로 검증의 전 과정을 공개해야 하는 팩트체크가 적극적으로 활용하기에는 적합하지 않은 경우가 많다. 통계자료나 뉴스기사도 대안이 될 수 있지만 ‘사실’을 검증하는 데 있어 시간, 장소, 맥락에 따라 사회적 사실이 새롭게 구성될 수 있는 뉴스기사나 참여자의 범위나 단위를 통일하기 어렵고 시계열로 정보가 바뀔 수 있는 통계를 사용하는 것 역시 기술적인 한계가 따른다. 피버모델이 사용한 위키피디아는 이러한 이유로 가장 현실적인 대

안으로 주목받았는데 위키피디아 역시 자료가 실시간으로 수정 및 추가되며 협업형 모델의 특성 상 문장이 고르지 못하다는 점도 한계로 지적된다. 실제로 국내 유일의 인공지능 팩트체크 사례인 서울대 연구팀의 경우도 위와 같은 이유로 정제된 대용량의 한국어 데이터셋을 충분히 찾지 못해서 모델 개발이 진전을 이루지 못한 바 있다.

이에 대한 대안으로 인간 팩트체커가 사전에 팩트체크를 수행한 데이터에 대해 라벨링을 하고 이를 훈련 데이터셋으로 삼아 모델을 개발하는 방안이 시도되고 있다. 이는 팩트체크 과정 자체가 팩트체커 1인에 의한 사실 판정이 아니라 복수의 팩트체커들 간의 합의를 도출해내는 데 있다는 점에서 '사실' 구성을 위한 인간과 기술 간의 합의 과정으로 해석할 수 있다. 하지만 모델 개발을 위한 데이터셋을 확보하는 과정이 사실을 확인하는데 가장 적합한 자료를 탐색하는 과정이기 보다는 인공지능 알고리즘에 최적화된 데이터로 범위가 1차적으로 한정된다는 점에서 검증 '가능한' 사실이 기술적으로 제한되고 만다. 또한 이 모든 과정이 결국은 기계적인 '판정'에 이른다는 점에서 사실성에 다다르기 위한 '합의'를 도출하는데 비중을 두는 팩트체크 저널리즘이 도출한 사실과 상이한 맥락에 놓이게 된다.

3) 정보 흐름의 탐지와 통제

마지막 유형은 가짜뉴스의 '형태'를 기계를 통해 살펴 탐지하는 방법이다. 이들은 앞선 사례처럼 정보의 '내용'을 살피는 과정을 자동화하지 않는다. 그보다는 내용을 한정하고 관련 정보가 퍼져 나가는 형태를 추적하고 통제하는 과정을 자동화하는 기술이다. 지금까지는 대체로 트위터 등 소셜미디어에 한정되어 이루어졌는데 허위 정보가 퍼져 나가는 흐름과 의심 계정을 탐지하여 정보의 흐름을 이용자에게 알리고 통제하는 기술사례들이 해당한다.

브라질의 대표적인 팩트 체크 기관인 Aos Fatos(to the facts)에서는 2020년 왓츠앱, 유튜브, 트위터, 인스타그램, 페이스북 등 소셜미디어 상에서 퍼지는 허위정보의 흐름을 실시간으로 자동화 도구 레이더(Radar)⁷⁾를 출시했다. 이 도구는 브라질의 정치 여론 뿐 아니라 Covid19 팬데믹 시기 허위의심정보의 흐름을 효과적으로 모니터링하는데 일조했는데, 그 작동 과정은 다음과 같다. 먼저 검증할 주제를 선택하고, 주제와 관련된 단어를 포함한 게시물들을 API를 통해 수집한 후, 게시물의 저자, 이미지와 영상, 내용과 관련된 정보를 각각 추출한다. 레이더는 이렇게 정리된 데이터 중 허위 정보를 식별하기 위해 크게 두 가지 사항을 고려하는데, 첫째는 메시지가 허위정보 캠페인과 관련된 단어를 포함하는지, 둘째는 해당 메시지가 저품질의 콘텐츠가 가진 전형적 특징을 띄고 있는지 살펴보는 것이다. 메시지의 출처가 익명이거나 미심쩍어 신뢰할 수 없는 경우, 조회수를 높이기 위해 공격적이거나 도발적인 단어를 선택하거나 자극적이거나 과장된 표현을 사용하는 경우, 과도한 대문자의 사용, 문법 오류가 있는 경우에 저

7) <https://www.aosfatos.org/radar/#/>

품질 콘텐츠로 분류될 확률이 높다.

유사하게 미국 인디애나 대학교 소셜미디어 관측소(Observatory on Social Media) 연구센터에서는 소셜미디어상의 허위정보에 대응하기 위한 다양한 기술적 해법을 고안해내고 있다. 우선 2014년 공개한 머신 러닝 알고리즘 보토펜터(Botometer)⁸⁾가 있는데, 이는 특정 트위터 계정의 운영 주체가 인간이 아닌 봇일 가능성을 평가하는 도구이다. 보토펜터는 검증 대상이 되는 트위터 계정을 계정의 프로필, 친구, 소셜 네트워크 구조, 시간대별 활동 패턴, 언어, 정서 등의 특성을 기준으로 평가하고 있다. 2016년에는 트위터 상에서 이용자가 검색하는 주제와 관련된 정보가 퍼져 나가는 과정을 시간적 추세와 정보를 전파하는 네트워크 요소를 고려해 시각화하는 호시(Hoaxy)⁹⁾ 서비스가 출시되었다. 호시는 신뢰도가 낮은 출처에서 공유된 정보나 독립적인 팩트체크 기관에서 나온 정보를 시각적으로 추적할 수 있게 한다. 더불어 트위터에서 확산되는 정보의 조작 가능성을 추적하고 탐지하는 데 도움을 주는 애플리케이션인 봇슬레이어(BotSlayer)¹⁰⁾가 2018년 개발되었다. 비정상성 탐지 알고리즘을 활용해 해시태그, 링크, 계정 정보, 미디어 등의 전파와 확산 과정에서 비정상적인 움직임이 감지되는지 포착하는 방식으로, 이용자가 자신이 검색한 영역에서 봇 의심 활동이 나타나고 있는지 실시간으로 모니터링하는 데 도움을 준다. 가장 최근인 2022년에는 톱 파이버(Top Fibers)¹¹⁾ 대시보드가 개설되는데, 이는 트위터와 페이스북 등 소셜미디어에서 신뢰도가 낮은 정보를 가장 많이 퍼트리는 계정인 슈퍼전파자(superspreaders) 계정을 추적 및 보고하는 데 목적이 있다. 슈퍼전파자는 특정한 기간 동안 신뢰도가 낮은 정보들을 가장 많이 반복 공유한 계정으로 특정되며, 여기서 신뢰도가 낮은 정보의 기준은 이피 뉴스(Iffy.news)나 미디어 편향 팩트체크(Media Bias/Fact Check: MBFC)와 같은 독립적인 제3의 기관에서 신뢰할 수 없다고 평가한 출처가 해당 정보의 출처인 경우를 가리킨다.

이 밖에도 탐지해야 할 콘텐츠나 사이트를 한정하여 이들의 정보 생산 및 유통을 추적하는 기술사례도 있다. 북미와 유럽 대륙의 상호 협력과 이해를 추구하는 초당파적 미국 공공정책 싱크탱크인 마셜 펀드(German Marshall Fund of the United States)는 민주주의 수호 동맹(Alliance for Securing Democracy)의 구상 하에 2019년부터 해밀턴 2.0(Hamilton 2.0)¹²⁾ 프로젝트를 진행하고 있다. 해밀턴 2.0은 러시아, 중국, 이란 정부의 온라인 선전과 선동정보에 대한 실시간 정보를 제공하는 대시보드이다. 해당 도구는 명시된 국가의 정부와 모종의 관계가 있다고 추정되는 트위터나 유튜브 계정, 방송 채널, 뉴스 사이트, 외교부 성명서 등을 추적해 데이터를 수집하고 머신러닝과 자동번역 기법을 이용해 핵심 정보를 추출해 이를 대시보드에 공개한다.

또한 미시간 대학교의 소셜 미디어 책무 연구센터(Center for Social Media Responsibility: CSMR)

8) <https://botometer.osome.iu.edu/>

9) <https://hoaxy.osome.iu.edu/>

10) <https://osome.iu.edu/tools/botslayer>

11) <https://osome.iu.edu/tools/topfibers/>

12) <https://securingdemocracy.gmfus.org/hamilton-dashboard/>

는 이피 코션트(Iffy Quotient)¹³⁾라는 측정도구를 공개했다. 이는 의심스러운 뉴스 사이트를 특정하고 이들이 출처인 콘텐츠가 트위터나 페이스북에서 얼마나 높은 비율로 확산되었는지 측정함으로써 의심 사이트가 소셜 미디어 상에서 얼마나 많은 관심을 획득했는지 판단하는 것을 돕는다. 여기서 의심 사이트는 빈번하게 허위정보를 발행하며, 뉴스가드(NewsGuard) 나 팩트체크 전문기관 등에 의해서 신뢰도가 낮게 평가된 사이트를 의미한다. 유사한 서비스 개발 사례는 웰슬리 대학교의 소셜 인포매틱스 랩(Social Informatics Lab)에서 2014년 개발한 트위터트레일스(TwitterTrails)¹⁴⁾가 있다. 이는 트위터 상에 공유된 정보를 추적하고 정보의 신빙성을 판단하는 데 도움을 주는 도구인데 공유된 정보의 전파 형태를 파악하고 이용자들이 해당 정보의 신뢰도에 얼마나 회의적으로 반응하는지 분석한다. 정보 전파 양상에 더하여 이에 대한 이용자 전반의 반응을 살피는 것은 저널리스트 등이 사안을 조사하고 정보의 사실성을 판단하는 데 간접적인 도움을 줄 수 있다.

마지막 유형은 앞선 기술 유형과는 달리 정보의 '내용'을 판단하지 않는다. 그보다는 정보 확산 네트워크의 특성을 활용해 오정보, 허위조작정보, 유해정보 등의 '탐지'에 인공지능 기술을 활용한다. 이는 주로 트위터가 활발하게 사용되던 2018년을 전후하여 많이 연구되었는데 정보 확산 네트워크에서 잘못된 정보가 유통되고 소비되는 경로를 파악하여 통제하는 목적을 갖는다. 엄밀하게 말하면 사실의 진위여부를 판정하는 팩트체크의 범주에 포함할 수는 없지만 이용자의 입장에서 접한 정보의 출처, 신뢰도, 유통 경로를 파악하여 정보를 판단할 수 있게 한다는 측면에서 팩트체크의 맥락으로 이해할 수 있다. 이는 사실상 오정보, 허위조작정보, 유해정보에 대해 현재의 미디어 환경을 고려한 가장 포괄적인 이해를 제공한다. 현재의 미디어 환경에서는 정보 생산의 최초 단계에 있는 사람 뿐 아니라 정보의 유통 과정에 가담한 사람 및 매체도 생산의 맥락에 포함될 수 있다(이희은, 2020). 빠른 속도로 생성되고 확산되는 디지털 정보의 특성 상 내용의 참, 거짓보다는 정보가 생산되고 유통되는 관계망을 파악함으로써 정보의 신빙성을 판단할 수 있는 근거를 제공할 수 있다.

무엇보다 팩트체크가 다루어야 하는 대상이 과학적인 명제이기보다는 사회적으로 구성된 사실이라는 점에서 근거 자료와의 대조를 통해 사실을 판정하는 인공지능 팩트체크에 대해 회의적인 시각이 많다. 가치를 배제하기 어려운 사회적 사실에 있어 기계적으로 검증할 수 있는 사실이라는 것 자체가 굉장히 제한적인 범주에 속하며 그 방법론 역시 맥락을 고려하지 않은 대조에 의존하고 있기에 팩트체크의 합의 도출 과정과 상이하게 진행되는 측면이 있다. 이 때 정보 확산 네트워크를 파악하는 일은 정보의 유통 경로를 시각화하고 드러낸다는 점에서 '투명성을 통한 객관성을 확보'하는 팩트체크 저널리즘과 인식론적으로 더욱 가깝다고 볼 수 있다.

사실상 네트워크를 통해 '가짜뉴스'로 통칭되고 있는 오염된 정보의 흐름을 파악하는 것이 기술적으로 가장 '적합한' 방법이라는 점에 대해서는 많은 기술자들이 동의하고 있다. 하지만 이는 현실적으로 어려

13) <https://csmr.umich.edu/projects/iffy-quotient/>

14) <http://twittertrails.com/>

운 방법론인데 앞서 설명한 트위터의 경우도 현재는 트위터가 API제공을 중단함으로써 해당 서비스들은 2022년 중반부터 현재까지 일시적 운영 중지 또는 부분 운영 상태이다. 트위터가 예전만큼 정보 교류가 활발한 매체가 아니라는 점도 한계다. 특히 국내의 경우 카카오톡을 가장 많이 사용하는데 이 플랫폼의 경우는 사용자 네트워크가 일반인이나 연구자에게 공개되지 않는다. 데이터를 수집할 수 없기 때문에 정보 확산 네트워크를 파악하는 일이 불가능하여 관련 기술을 개발한다는 것이 현실적으로 가장 어려운 유형에 속한다.

5. 결론을 대신하여 : ‘가짜뉴스’의 기술사회학적 정의

이 글은 인공지능 기술이 팩트체크를 위한 도구로서 기능하며 ‘사실’을 검증, 판단, 구성하는 방식을 국내외 인공지능 팩트체크 기술사례를 통해 검토했다. 지금까지 영국, 아프리카, 아르헨티나, 미국, 브라질, 스페인, 기타 북미와 유럽 등지에서 인공지능 기반 팩트체크 기술을 자체 개발하거나 원활한 기술개발을 돕는 것을 목적으로 하는 상당수의 프로젝트가 진행되어 의미 있는 성과를 거둔 것을 알 수 있다. 인공지능 팩트체크 기술은 자체적으로 ‘사실’을 판정하거나 혹은 사실을 검증하는 과정에 기여함으로써 팩트체크 저널리즘이 추구한 것과는 상이한 맥락에서 기술사회적으로 사실을 구성해 왔다.

각각의 인공지능 팩트체크 기술이 검증하는 대상과 검증하는 방식은 두 가지 맥락에서 상이한 방식으로 사실성을 정의하며 이는 가짜뉴스를 각각 다른 방식으로 이해할 가능성을 보여준다. 첫 번째는 내용으로서 사실성이다. 이는 전통적으로 팩트체크 저널리즘이 추구해 온 사실 확인 모델을 기술적으로 구현 가능한 범위에서 재구성한 것으로 정보의 내용과 내용을 구성하는 개별 사실의 진위 여부를 판별한다. 즉, 인공지능 기술이 무수한 정보에서 검증할 만한 ‘주장’을 탐지하고 모델이 학습한 훈련데이터를 바탕으로 ‘대조’를 기반으로 한 사실 검증을 하여 개별 사실의 참, 거짓 여부나 정도를 밝힌다. 앞서 살펴본 것처럼 현재 서비스가 진행중인 대부분의 인공지능 팩트체크 기술은 주장을 탐지하고 근거 데이터를 구축하여 훈련한 모델이 사실의 여부나 정도를 판정하는 형태로 이루어진다. 현재 서비스가 중단되기는 했지만 국내에서 처음으로 개발된 인공지능 팩트체크 모델 역시 피버 모델을 차용해 위키피디아 문서를 훈련데이터로 삼고 정보의 참, 거짓을 판별하는 형태로 고안되었다. 하지만 이 같은 접근법은 기본적으로 ‘가짜뉴스’로 통칭되는 온라인 상의 오염된 정보의 ‘내용’을 기계적으로 검증해 사실 여부를 확인하는 형태를 취하고 있다. 이는 본래 팩트체크 저널리즘이 개별 요소의 사실성이 아니라 주장들이 맥락화하여 만들어내는 ‘총체적인 사실’을 ‘합의’의 과정을 통해 검증해간다는 점과 극명하게 대비된다. 인공지능 팩트체크 기술이 디지털 환경에서 오염된 정보가 광범위하게 생산되고 확산되는 환경에서 제한적인 팩트체크 인력을 대신하고 긴 검증과정을 자동화할 목적으로 만들어졌다는 점에서 이 같은 기술의 제한적인 방향성은 국내에서 ‘가짜뉴스’가 하나의 담론으로 구성된 방식과 공모할 가능성이 높다. 가짜뉴스는 정보의 목적과 범위에 따라 세분화될 수 있지만 그것이 사회적으로 통용되는 방식은 정보의 진위 여부

를 가려야 할 주체의 목적에 따라 취해야 할 진짜 혹은 사회적으로 배척해야 할 가짜의 이분법으로 분류되어 왔다. 이것이 '가짜뉴스'라는 용어 자체가 정치적인 수 있다는 비판을 받아 온 이유인데 이는 인공지능 팩트체크 기술을 설계하는 과정과 더욱 긴밀하게 공모하여 사실의 층위를 복잡하게 할 수 있다.

반면 '사실성의 구조'는 정보가 특정한 형식으로 생산, 유통, 소비되는 구조와 과정을 살펴봄으로써 사실성을 맥락적으로 이해하는 경우다. 이 경우 정보의 내용이나 생산 주체보다는 정보가 생산되고 유통되는 미디어 기술환경에 대한 이해에 가깝다. 이 때 사실성은 내용이나 개별 정보의 진위 여부에 의해 결정되는 것이 아니라 정보가 생산되어 퍼져 나가는 형태를 통해 무한히 유동적으로 구성된다. 과정을 드러냄으로써 사실성을 확보하는 이 같은 관점에서 가짜뉴스는 내용이 아니라 네트워크 안에서 가짜뉴스의 유통에 가담하는 디지털 기기들을 정렬하는 물질적 기능을 한다. 가짜뉴스라는 물질이 정렬한 디지털 기기들의 연결망을 살펴보는 관점은 생산, 유통, 소비가 더 이상 유의미하게 구분되지 않는 현대의 미디어 환경에서 가짜뉴스를 구조적으로 이해하고 이용자가 (오)정보의 생산, 유통, 소비에 다시 개입할 수 있는 여지를 열어준다. 이는 사회적으로 구성된 사실이 생산과 재생산을 거치며 퍼져 나가는 과정을 가시적으로 드러냄으로써 미디어 환경이 '가짜뉴스'의 경험과 그것이 파생하는 정동에 어떤 조건이 되는지 보여준다.

가짜뉴스와 인공지능이 동시에 사회적인 화두로 떠오른 현재의 상황은 답보 상태에 머무른 국내의 인공지능 팩트체크 기술을 또 다른 형태로 만들어낼 여지를 열고 있다. 단 몇 주의 간격을 두고 정부는 가짜뉴스에 대한 규제와 생성형 인공지능이 만들어내는 가짜뉴스 규제방안 마련에 대한 정부의 단호한 입장과 비전을 발표했다. 인공지능 기술이 어떤 방식으로 사회 안에 위치하게 될지는 언제나 기술과 사회적 담론 사이의 공모에 의해 정해진다. 감정, 선입견 등 사회적 선택 과정이 갖는 편향을 대체한 인공지능 기술을 기대하다 한 번의 '실패'를 겪은 한국형 인공지능 팩트체크 기술이 나아가야 할 방향에 대해 숙고가 필요한 시점이다.

참고문헌

- 김민정(2018). 가짜뉴스에서 허위조작정보로: 가짜뉴스 규제 관련 국내 법안과 해외 대응책에 나타난 용어 및 개념정의 비교. 미디어와 인격권. 5권 2호, 43~81.
- 김창숙(2020). 의례적, 방어적, 의도적 : 한국 주요 신문 에디터의 사실확인을 중심으로 한 게이트키퍼 관행 연구. 한국언론학보, 64(5), 5-45, 10.20879/kjcs.2020.64.5.001
- 박대민(2015). 사실기사의 직접인용에 대한 이중의 타당성 문제의 검토: 동아일보와 한겨레 신문의. 4대 강 추진 논란 기사에 대한 늦 정보원 연결망 및 인용문 분석. 한국언론학보, 59권 5호, 121~151.
- 박진우(2020). '가짜뉴스'라는 기호를 다루는 사람들: 저널리즘, 정보기술, 그리고 대중들의 문화정치. 사이버커뮤니케이션학보, 37권 4호. 149~195.
- 신혜선·이영주(2021). 오보 문제에 대한 기자 인식: 심층 인터뷰를 중심으로. 한국언론학보, 65권 4호, 239~272.
- 최순욱·고문정(2016). ICR Media Trend Report. 서울대학교 언론정보연구소 뉴스레터, 통권 3호, 1-19.
- 정은령(2018). 한국 팩트체크 저널리즘의 특징: 팩트체크 언론인들의 사실 인식과 사실 검증과정 탐색을 중심으로. 언론정보연구, 55(4), 5-53.
- _____(2019). 팩트체크 뉴스와 한국 방송 저널리즘의 신뢰 회복: 방송 기자들의 팩트체크 뉴스 양식과 뉴스가치에 대한 인식을 중심으로. 방송문화연구, 31(1), 47-101.
- 정준희(2019). [따옴표 저널리즘의 딜레마] 관행이란 이름의 범속함, 그 악의 평범성: 게으른 받아쓰기를 넘어 복화술 저널리즘으로. <방송기자>, 47권, 12-14.
- 양정애(2019). <일반 시민들이 생각하는 '뉴스'와 '가짜뉴스'>(미디어이슈 5권 1호). 서울: 한국언론진흥재단.
- 이나연(2018). 한국 언론의 팩트체크: 19 대 대통령선거에서의 후보자 검증 기사를 중심으로. 언론정보연구, 55(4), 99-138.
- Graves, D.(2018). Understanding the promise and limits of automated fact-checking. In Reuters Institute for the Study of Journalism. Reuters Institute for the Study of Journalism.
- Guo, Z., Schlichtkrull, M., & Vlachos, A.(2022). A survey on automated fact-checking. Transactions of the Association for Computational Linguistics, 10, 178-206.
- IFCN(2016). International Fact-Checking Network fact-checkers' code of principles. Accessed July 2nd, 2022 at: <https://www.poynter.org/ifcn-fact-checkers-code-of-principles/>

- Marwick, A. & Lewis, R.(May 2017). *Media Manipulation and Disinformation Online*. Data and Society. <https://datasociety.net/library/media-manipulation-and-disinfo-online/>
- Mueller, R. S.(March 2019). *Report on the Investigation into Russian Interference in the 2016 Presidential Election* (Mueller Report). Retrieved from: <https://www.govinfo.gov/content/pkg/GPO-SCREPORT-MUELLER/pdf/GPO-SCREPORT-MUELLER.pdf>
- Posetti, J., & Matthews, A.(2018). A short guide to the history of 'fake news' and disinformation: A new ICFJ learning module. International Center for Journalists, 7(2018), 2018-07. Retrieved July 2nd, 2022 from: <https://www.icfj.org/news/short-guide-history-fake-news-and-disinformation-new-icfj-learning-module>
- Reporter's Lab(2022). Global fact-checking map and database. Accessed July 2nd, 2022 at: <https://reporterslab.org/fact-checking/>
- Thorne, J., & Vlachos, A.(2018). Automated fact checking: Task formulations, methods and future directions. arXiv preprint arXiv:1806.07687.
- Uscinski, J. E., & Butler, R. W.(2013). The epistemology of fact checking. *Critical Review*, 25(2), 162-180.
- Wardle, C., & Derakhshan, H.(2018). Thinking about 'information disorder': formats of misinformation, disinformation, and mal-information. In C. Ireton & J. Posetti (Eds.), *Journalism, 'fake news' & disinformation* (pp. 43-54). Paris, France: UNESCO.

**자동화 저널리즘과
기술-인간의 피드백 루프
: 뉴스 댓글창 진화 사례를 중심으로**

홍남희(서울시립대)



한국여성커뮤니케이션학회-뉴스통신진흥회 후원 세미나 20230922

자동화 저널리즘과 기술-인간의 피드백 루프

: 뉴스 댓글창 진화 사례를 중심으로

서울시립대학교 도시인문학연구소 홍남희

AI와 자동화, 저널리즘

- 기술 변화는 저널리즘 영역에 직, 간접적인 변화를 야기
- 저널리즘 자체가 원래 기술 발전과 연관된 것

- 펜, 인쇄술(대량출판/팜플렛), 텔레비전, PC 및 모바일 기기, 디지털 기술

- 디지털 저널리즘의 출현과 여러 가지 변화들

- 새로 출현하는 관계와 기존 관계의 변화

예. 저널리스트와 수용자/저널리스트와 기술/저널리스트와 저널리스트/뉴스룸과 기술/수용자와 기술/수용자와 수용자 등 다양한 관계가 기술 매개로 전면화

- 기술의 진화와 자동화의 심화

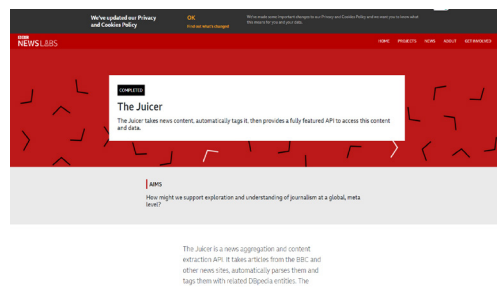
예. 플랫폼 중심의 뉴스 소비, 뉴스 생산-유통-소비의 즉각성, 생산/유통/소비 과정과 관련한 AI 도입, 기술 발전에 따른 AI 배치 및 작동 방식의 불투명화, 예측 불가능성 등

AI와 자동화, 저널리즘

- 새로운 기술 도입으로 무엇이 달라졌는가.
 - 저널리즘의 이상적 가치와 시민사회의 인간적 가치를 어떻게 확장/저해하는가
 - 한편으로는 저널리즘 과거에 대한 신화화 경계 필요
 - 폭로 저널리즘, 탐사 저널리즘은 과거에도 일부였음
 - 특정 시기, 예를 들어 아날로그 시대 저널리즘 이상화 경계
- (Michael Schudson, 이해림, 최순옥, 2016)

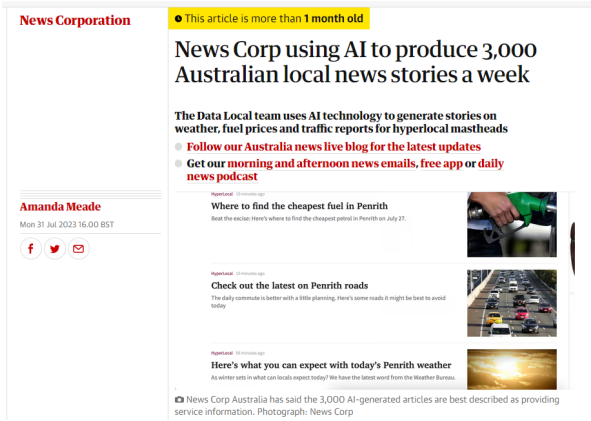
자동화 저널리즘의 사례

- 원래 저널리즘은 자동화를 다양한 방식으로 도입해 옴.
- LA 타임스 퀘이크봇 (지진 발생시 자동 기사 작성)
- 워싱턴포스트 헬리오그래프(워싱턴 D.C. 지역 고등학교 미식축구 경기 보도하는 자동화 소프트웨어)
- BBC The Juicer (뉴스 수집 및 콘텐츠 추출 API, BBC 및 기타 뉴스 사이트에서 기사를 가져와 자동 구문 분석 등)



최승영 (2023.8.1) AI 시대 저널리즘, 이용할 것인가 이용당할 것인가. 한국기자협회보.
<http://www.journalist.or.kr/news/article.html?no=54022>

생성 AI와 뉴스 자동 생산



Stories such as “Where to find the cheapest fuel in Penrith” are created using AI but overseen by journalists, according to a spokesperson from News Corp. There is no disclosure on the page that the reports are compiled using AI.

“I'd stress that all such information and decisions are overseen by working journalists from the Data Local team.”

The Guardian, 31 Jul 2023

생성 AI와 뉴스 저작권 문제

- 한국신문협회, <생성형 AI의 데이터 학습에 따른 뉴스 저작권 침해 방지를 위한 한국신문협회 의견> 문체부에 전달
 - 뉴스 콘텐츠가 AI 학습을 위해 이용될 경우 언론사 저작권 및 DB 제작자로서의 권리 침해한다고 주장
 - 언론사가 웹사이트 등을 통해 제공하는 뉴스 콘텐츠 제공 서비스는 '소재를 체계적으로 배열 또는 구성한 편집물로서 개별적으로 그 소재에 접근하거나 그 소재를 검색할 수 있도록 한 것' 이므로 저작권법상 데이터베이스에 해당하고 언론사는 데이터베이스 제작자
 - 데이터베이스제작자는 데이터베이스의 전체 또는 상당한 부분을 복제·배포·방송 또는 전송할 수 있는 권리를 갖는다

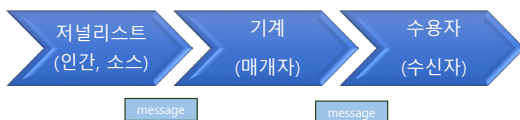
(박서연, 2023.7.26, 미디어오늘, 한국신문협회 “생성AI, 뉴스 데이터 학습 저작권 침해 방지해야”)

자동화 저널리즘의 정의 및 쟁점

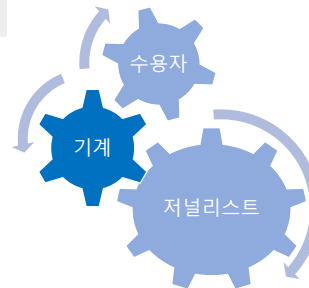
- 커뮤니케이션 과정과 노동을 자동화하는 기술 진화의 일부(Carlson, 2015).
- 기계적, 전자적 기기에 의한 뉴스 생산
- 맨 눈으로 볼 수 없는 무엇인가를 보게 하는 것
- 구조화된 데이터를 내러티브 뉴스 관점으로 자동 변환(Carlson 2015)
- 소프트웨어 생산 뉴스(Linden, 2017a)/컴퓨터이셔널 저널리즘/알고리즘 저널리즘 (Diakopoulos, 2014)/로봇 저널리즘(Clerwall, 2014).
- **주로 생산의 관점: 기계-저널리스트의 관계 구성의 관점에서 많이 논의됨**
- **그러나 기계-수용자 등 다양한 관계 부상**
- 자동화를 가능하게 하는 기술적 메커니즘은 인간이 커뮤니케이터이고 기계가 매개자라는 이론적 전제에 도전. 무엇이 커뮤니케이터를 구성하고 어떠한 사회적 관계가 인간-기계 사이에서의 교환 통해 구성되는지, 그 결과가 자아, 사회, 커뮤니케이션에 어떤 함의 갖는지 새로운 질문을 던짐 (Lewis et al., 2019)

자동화 저널리즘:인간-기계의 사회적 과정

알고리즘과 인공지능은 저널리즘 가치의 재평가라는 이미 지속되어 가던 과정의 일부이자 투명성, 데이터 보안, 책무성 같은 질문으로 저널리즘 영역을 확대(Lewis et al., 2019)



기존 저널리즘에 대한 이해:
소스로서의 인간-매개자로서의 기계-수신
자로서의 수용자



자동화 저널리즘:
저널리스트-기계-수용자의 상호유희 과정

자동화 저널리즘과 기계-저널리스트 관계

- 자동화 기술이 뉴스룸 소통 과정에 삽입됨으로써 기계-저널리스트 관계 중요해짐
- 기계 훈련 위해 저널리스트는 기계처럼 생각하도록 훈련
예. 기술의 어포던스에 적용하도록 훈련. 기사 작성 방식, 제목 작성 방식 등에 영향, 기술 매개한 수용자 반응 고려한 기사 취재 및 작성
- 기술에 의해 소진되었다고 느끼거나 전문성 타격 느낌

자동화 저널리즘과 기계-저널리스트 관계

- **자동화 불안(automation anxiety)의 사례**
 - 자동화 초기에는 기대에서 점차 직업, 지위, 전문성에 대한 위협, 공포(Linden, 2017, 72)
 - 디지털화에 대한 뉴스룸 내 세대 편차, 갈등/ 기술이 수익화로 이어지지 않는 상황/ 독자의 연성 뉴스 추구 성향에 따른 뉴스 생산 과정의 연성화 우려(홍남희, 이설희, 김수아, 2023).
 - 자동화 AI 시스템과 기사 편집 체계 사이의 불일치: 투명성, 책무성, 책임성 같은 저널리즘 가치 도달 어려울 수 있음(Lopez et al., 2022)
- **반복적 업무에서의 해방 도구로서 자동화**
 - 저널리즘 전문직주의 실천과 가치에 AI가 도움되도록 고안될 필요
 - 뉴스룸 재구조화, 콘텐츠 생산 다양화, 저널리스트들의 신기술 활용(Bossio & Nelson, 2021, 1377)
 - 기술을 뉴스룸 관행, 가치, 일상, 사회문화적 경험과 엮는 방식과 관점에서 바라볼 필요(Lopez et al., 2022).
 - 인간 저널리스트들의 창의성, 개성, 언어적 복잡성 등 전문성 강조
 - 자동화될 영역과 저널리스트 고유 영역 구분(Van Dalen, 2012, 651)

자동화 저널리즘과 기계-수용자 관계

- 디지털화 이후 뉴스는 저널리스트와 수용자, 기술 간 상호구성 관계로 진전
- 가장 큰 변화는 **수용자 간 상호소통**이 가능해진 것(댓글, 추천, 신고 등)
- 기술이 인간으로 자신을 제시할 경우 수용자 혼란(뉴스의 자동생산, 소셜봇 등)(Lewis et al., 2019, 11)
 - 수용자가 기계생산 콘텐츠 구분하는지 여부(독이성 신뢰성에서 차이 없지만 인간 소스에 미묘한 선호 표출)(Graefe et al., 2018)/저널리스트, 수용자 모두 소프트웨어에 더 긍정적(Jung et al., 2017)
 - 미디어 리터러시 논의, 가짜뉴스의 비판적 해독
- **댓글(수용자 반응)을 둘러싼 기계-인간, 인간-인간의 상호구성적 과정**
 - 댓글의 공적 역할에 대한 사회적 고민과 기술적 해법(AI 도입한 댓글 관리)
 - AI를 회피하는 수용자들의 실천

자동화 저널리즘과 인간-기계의 상호얽힘

- AI가 실행, 배치되는 조건, 방식, 인간 노동의 변화 등을 구체적으로 바라볼 필요성
- 인간과 기계를 대립시키고 경쟁시키기 보다 현장에서 인간-AI의 상호얽힘을 탐색하면서 인간됨(humanness)과 기계됨(machineness)이 협상, 구성되는 방식 탐색 필요(하대청, 2022)
- **뉴스 추천 알고리즘**
 - 필터 버블, 선별적 추천, 편향성과 유사성 기반한 편향적 뉴스 소비, 심층적 개인화
 - 플랫폼 기업의 알고리즘은 콘텐츠 내용 관계없이 이용자가 플랫폼에 오래 머물게 하는 것
- **댓글 차단, 삭제 위한 자동화 도입**
 - 악플의 부정적 영향에 대한 사회적 담론과 이에 대한 기술 기업의 해법. AI 개발을 통한 댓글 관리의 자동화
 - Data-hungry AI: 악플 학습 위한 수많은 데이터 학습/데이터 감지/삭제 및 차단

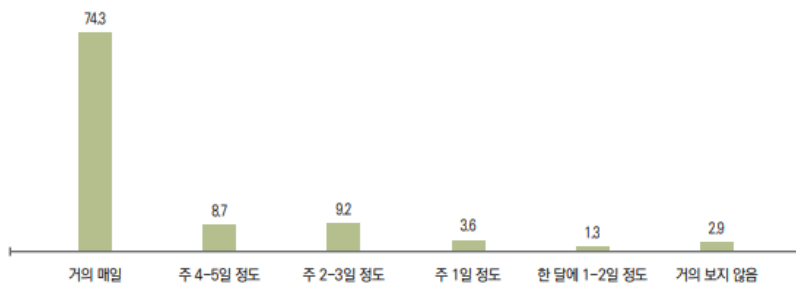
포털의 상품으로서 뉴스와 댓글

- 포털의 상품으로서 뉴스
- 포털의 상품으로서 댓글
 - 인터넷 규제 정책 대부분이 댓글 규제 정책
 - 인터넷 실명제(제한적 본인확인제), 임시조치(블로그 등의 내용 포함), 콘텐츠 관리
 - 자동 댓글 작성: 매크로 작업/소셜 미디어 봇 (computational propaganda, Bolsover & Howard, 2017)
 - 악플에 대한 포털 책임론
 - 실시간 검색어, 댓글 추천 및 공감, 베스트 댓글 등의 특정 행위 유발
 - 뉴스를 통해 이용자 포획

포털 뉴스 소비 빈도

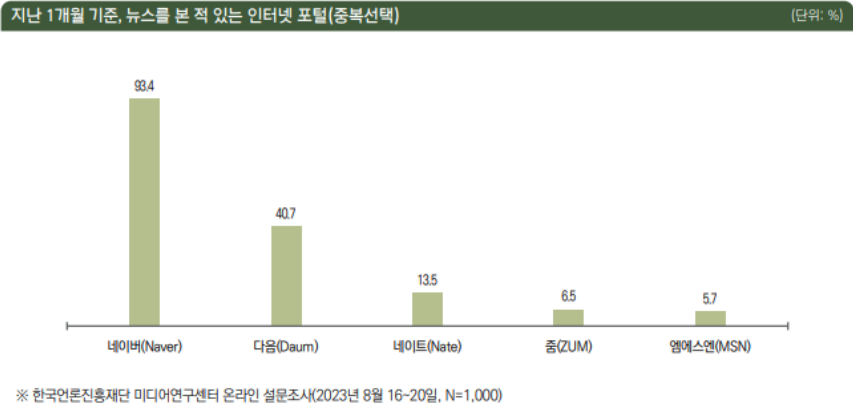
평소 인터넷 포털에서 뉴스 보는 빈도

(단위: %)



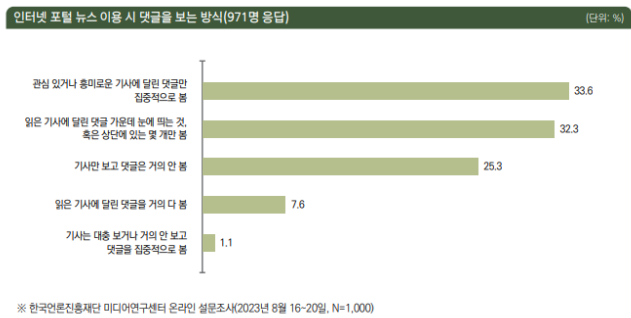
※ 한국언론진흥재단 미디어연구센터 온라인 설문조사(2023년 8월 16-20일, N=1,000)

뉴스 소비하는 주요 포털



뉴스 소비와 댓글

- 관심있거나 흥미로운 기사 댓글만 봄(33.6%)
- 댓글 중 눈에 띄거나 상단 댓글 몇 개만 봄(32.3%)



뉴스 소비와 댓글

• 댓글, 대댓글 작성 경험 여부

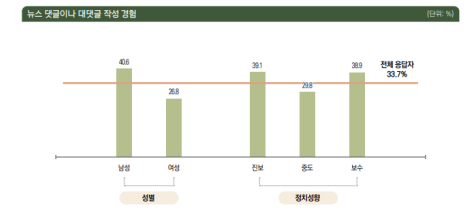
- 남성 > 여성
- 중도보다는 뚜렷한 정치성향 있는 경우
- 장애인 이동권 이슈 뉴스에 보수 성향 댓글 많았음(조원정, 이종혁, 2023)

• 뉴스 댓글에 공감 표시

- 10명 중 4명 정도
- 20대보다 40대, 50대가 더 많음
- 중도 < 진보(43.0%) < 보수(46.1%)

• 악플 신고 여부

- 신고하기 28.4%(20대가 가장 높음)
- 차단하기 22.7%(20대가 가장 높음)



포털의 댓글 관리 기술

- 설리, 구하라의 자살/정준영 사건/10.26 이태원 참사 등 이후 댓글창 문제 개선 위한 포털 책임, 의무론 목소리 커짐
- 해결책으로서 연예 뉴스면(이후 스포츠 뉴스면) 댓글 서비스 폐지(다음, 네이버)
- 연예뉴스 아닌 다른 면에 배치하여 댓글 유도함

네이버 연예뉴스는 댓글 공간이 건강하게 유지될 수 있도록 많은 노력을 기울여 왔습니다.

다각적인 노력에도 불구하고 인터넷 공간에서 연예인에 대한 인격 침해 댓글에 대한 우려가 커지고 있고, 현재의 기술적인 노력만으로는 이 문제를 해결하기에 부족한 점이 있었습니다.

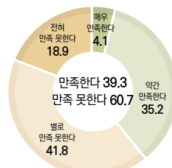
이에, 네이버 TV연예 서비스의 구조적인 개편이 완료될 때까지 연예뉴스 댓글을 잠정 폐지하기로 결정했습니다.

<https://help.naver.com/service/19620/bookmark/11471>

다음톡

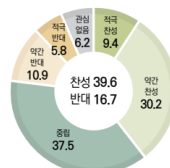
- 기사 작성 후 24시간 이내만 댓글 작성 및 노출이 가능
- 실시간 여론 알기 어렵고 이용자 의견 24시간 후 알수 없음(양정애, 2023.8.31)

포털 다음의 '타임톡'에 대한 만족도(이용 경험자 122명 응답) (단위: %)



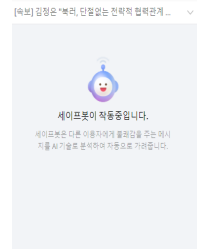
※ 한국언론진흥재단 미디어연구센터 온라인 설문조사 (2023년 8월 16-20일, N=1,000)

포털 다음이 도입한 '타임톡'에 대한 찬반 입장 (단위: %)



※ 한국언론진흥재단 미디어연구센터 온라인 설문조사 (2023년 8월 16-20일, N=1,000)

타임톡 Auto 순 23:57:49 발송



- 카카오톡처럼 변해서 불편해진 측면이 있다. 채팅 내용을 보려면 위로 스크롤을 올려야 하기 때문이다. 카톡처럼 바뀌니 발언이 가벼워졌다는 지적도 있다. 댓글은 영구박제되기 때문에 읽는 사람들이 힘들어했는데 그 부분은 개선됐다. 다만 본격적으로 사람이 개입해야 한다. AI(인공지능)가 관리하는 게 중립적이라고 하는데, 실제론 한계가 있다. 포털은 장소만 제공하니 언론사가 알아서 해야 한다는 입장이고 언론사는 포털에 달리는 댓글에 별도로 사람을 뽑아 관리할 생각까지는 없다. 해외 언론사 중에는 '커뮤니티 모데레이터'라며 댓글 관리만 하는 사람을 따로 뽑고 있다.”

(장슬기, 2023.8.26, 미디어오늘)

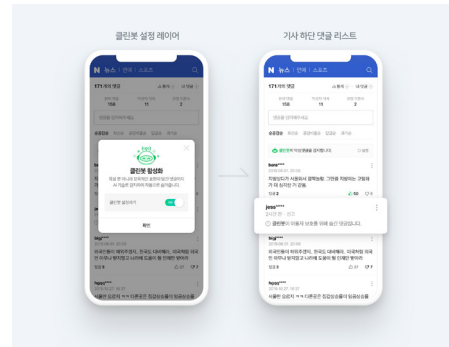
출처: <http://www.mediatoday.co.kr/news/articleView.html?idxno=311967>

네이버 클린봇

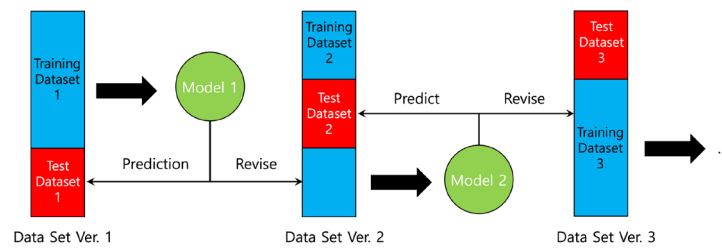
- 특정 표현을 자동탐지하여 숨기는 방식 (2020년부터 도입)
- 악성댓글 판단기준을 단어에서 문장 맥락으로 확대
- 비속어 없어도 문장 맥락 고려해 모욕적, 무례함 판단해 블라인드한다고 선전(정유림, 2020.6.14)
- 클린봇에 적용되는 악플 유형화는 사용자 반응 모니터링과 악플 관련 학술연구, 방송통신심의위원회 기준에 의해 이루어짐
- 뉴스, 연예, 스포츠 섹션 댓글 데이터 익명화해 샘플링하고 각 댓글 악플 여부 판별해 레이블링 함
- 정의된 유형 포함(2), 판단 애매함(1), 정의된 유형 표현 없음(0) 식으로 데이터셋 확보(35만건)

네이버 클린봇

- 욕설: 네이버 내부 욕설 데이터 포함된 표현
- 저속한 표현: 타인에게 불쾌감 주는 속되고 격이 낮은 표현
- 선정적인 표현: 성적으로 자극적인 표현
- 폭력적인 표현: 신체적 위협에 대한 표현
- 차별적인 표현: 지역/인종/국가/종교 기반 차별표현
- 비하적 표현: 상대방에게 모멸감, 수치감 주는 표현



- 레이블링은 사람에 의한 작업이기 때문에 얼마든지 오류가 발생할 수 있다. 학습 데이터에서 명백히 악플이 아님에도 불구하고 악플로, 그 반대로 잘못 레이블링한 사례를 찾을 수 있었다. 이러한 노이즈를 제거하기 위해 모델을 활용한 노이즈 정제 방법을 사용했다. 다음 절차를 1번부터 n번까지 반복하면 노이즈 정제 과정이 완료된다.



네이버 클린봇

“운영진은 클린봇 2.0을 준비하면서 출근부터 퇴근까지 악플만 모아놓은 엑셀 문치를 종일 들여다보곤 했다. 오랜 기간 악플에 집중적으로 노출된 셈이다. 황 매니저는 “클린봇 운영팀에 ‘악플과의 전쟁’은 괴로운 감정노동”이라며 “중간중간 바람을 쐬고 머리를 식힌다. 팀원들에게는 ‘귀가하면 좋은 생각을 하고 마음이 따뜻해지는 책을 보라’고 권한다”고 했다.

(박돈규, 2020.11.10. 조선일보).

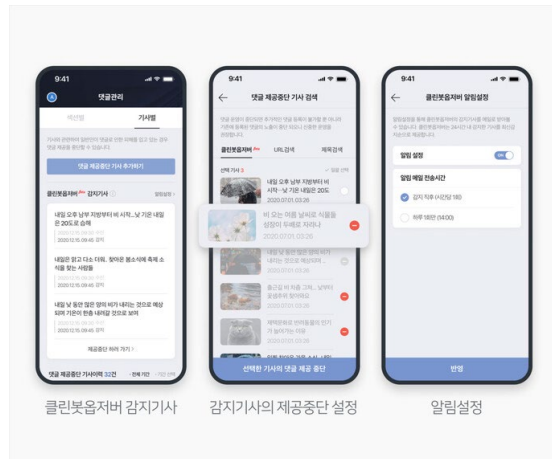
- 기술적 자동화 같지만 사람의 작업이 매 순간 투입되는 댓글 관리 작업
- 이미 포털 초창기부터 다양한 방식으로 댓글 관리 작업은 수행되어 왔으나
- 이용자가 남긴 댓글이 데이터가 되고 그것이 기계를 학습시켜 다시 댓글을 관리하는 시스템으로 작동하는 원리

기계학습 통한 자동화된 댓글 관리

- 단순욕설 및 라벨링에 해당하는 혐오댓글 탐지에 뛰어난 성능
- 스테레오타이핑, 분리하기 등을 악플로 판별하는 데는 부진함
- 혐오표현 탐지가 대상, 유형에 따라 편향된 방식으로 나타남(이신행, 2021).
- 여성, 이주민, 노인혐오성 댓글이 클린봇 작동에도 여과없이 게시(양혜승, 2022)
- 유튜브 댓글에서 욕설, 비속어보다 외모 비하, 조롱 목적 멸칭, 모욕적 상징을 함축한 고유명사 중심 악플 관련 언어적 변이(이신행, 2022).
- 기계학습 통한 악플 탐지 시스템의 한계 조명

네이버 클린봇 오퍼버

- 악성댓글 활동이 증가하는 기사 기계적 탐지해 언론사에 통지하는 서비스
- 2018년 10월부터 네이버는 언론사별 댓글제공방식을 선택할 수 있게 함
- 기사별 댓글 제공 중단 기능 추가: 2021년 8월부터 개별 기사 단위로 댓글 기능 제공 여부 정할 수 있게 함
- 클린봇과의 협업. 악플 급증 기사 알람을 받은 경우 기사 댓글창 운영 중단 등의 조치 취할 수 있음.
- 관련 패턴 파악해 언론사 설정에 따라 댓글 중단 선제적 관리 가능성
(임지선, 2023.7.14, 한겨레)



기계와 인간의 협업

- AI는 마치 인간인 것처럼 일하고 있지만, 그 뒤에서 인간은 마치 AI인 것처럼 일하고 있는 것이다. (하대청, 2018).
- AI가 인간을 대신하거나 도와주는 행위는 아주 낮은 가능성이 출현하더라도 미디어들이 앞다투어 보도하는 반면, 인간 노동자들이 AI를 도와주는 현실은 거의 재현되지 않는다는 점이다. 거의 매일 기업과 언론은 AI가 인간의 노동을 모방하거나 덜어주는 사례들을 소개하는 반면 AI를 여전히 훈련시키고 있는 인간의 숨은 노동들에 대한 관심을 두지 않는다. 인간을 모방하거나 능가하는 AI의 행위성은 놀라움과 두려움이 뒤섞인 채 주목하면서도 정작 이 AI의 능력 속에 여러 종류의 인간 행위성이 발휘되고 있다는 사실은 무시하고 있는 것이다 (하대청, 2018).

사회적 문제의 기술화

- 댓글 문제를 기술적으로 해결
- 사회적이고 정치적인 문제를 "기술적 해결책 필요한 기술적 문제로 보는 실리콘밸리에 만연한 사고방식"(Geiger, 2016, 791).
- 인간의 데이터 레이블링, 기술 기업 지속 성장 위한 기술개발 촉진으로 기술 기업의 사회적 역할 더욱 강화(Gillespie, 2020).
- 방대한 콘텐츠 양과 많은 이용자 수 등의 규모(scale) 문제가 AI 도입 정당화의 수사가 되고 있음
- 자동화는 여러 맥락에서 동일한 과정과 결과를 도출하는 절차화를 유도
- AI를 통한 데이터 트레이닝과 그 적용은 간단해 보이지만 인간-비인간 행위자들의 활동, 표현, 상호작용 토대로 하는 "다양한 사회공학적 스케일메이킹"(Gillespie, 2020).



댓글 관리의 자동화: 사회적 함의

- 기술적 자동화
 - 기술 중립성의 환상을 제공
 - 편향의 작동, 사람의 개입으로 인한 문제 예.악플에 성차별 제외/상업적 이윤추구
 - 기술-인간의 피드백 루프: 댓글의 데이터화, 기계학습, 악플 감지 인공지능 작동
- 악플 내용의 강도(intensity) 중심 규제 정착
 - 무례하고 저속한 표현 중심 규제
 - 신중의 표현, 유형화를 벗어난 표현, 텍스트가 아닌 표현 등을 제외시킴
- 포털의 상품으로서 뉴스와 댓글에 대한 사회적 규제가 부재
 - 저널리즘의 공적 가치에 대한 사회적 인식 부족
 - 댓글의 '정당정치적' 기능에 대해서만 규제 논의가 부상
 - 댓글 관리 또한 기술 기업의 기술 개발 논의로 수렴



THE END



감사합니다

A series of horizontal dotted lines for writing.

A series of horizontal dotted lines for writing, spanning the width of the page.