한국언론학보
2024년 68권 4호 119~157쪽
https://doi.org/10.20879/kjjcs.2024.68.4.004
http://www.comm.or.kr

인공지능 팩트체크와 '사실성'의 기술사회적 정의

이정현 중앙대학교 인문콘텐츠연구소 HK연구교수
박소영 조선대학교 미디어커뮤니케이션학과 조교수

Artificial Intelligence Fact-checking Technology and the Sociotechnical Definition of 'Factuality'*

Jeonghyun Lee**
(Research Professor, Humanities Research Institute at Chung-Ang University)
Soyoung Park***
(Assistant Professor, Chosun University)

There has been a continuous movement to automate fact-checking through artificial intelligence (AI) technology as a countermeasure to the widespread production and rapid dissemination of misinformation, disinformation, and harmful information on the internet. However, this approach often appears to be a technology-centric solution with lacking two key perspectives. First, it proposes AI in the fact-checking process without sufficient consideration of the conditions and methods for AI implementation in the fact-checking process. Second, the current approach lacks an understanding of how factuality is negotiated and constructed within fact-checking process performed by AI. This paper addresses these issues by examining how AI, as a 'technology' of fact-checking, constructs 'facts' through domestic and international AI fact-checking technology cases. This paper explores how the 'factuality' constructed by AI fact-checking differs from the 'factuality' provided by traditional fact-checking journalism, which has historically shaped social reality by verifying and adding facts. To achieve this, the paper reviews AI fact-checking technologies presented by globally certified fact-checking organizations as of October 2023. The sources come from the International Fact-Checking Network (IFCN) and AI fact-checking technology cases published in an online database created by the Rand Institute in the U.S., aimed at fighting disinformation. Based on the objectives of each technology and how they are set up, this paper categorizes the AI fack-checking technology

* This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea(이 논문은 2017년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임) [NRF-2017S1A6A3A01078538].

** maryjlee1205@gmail.com, first author
*** sy.park@chosun.ac.kr, corresponding author

cases into: 1) detection of claims, 2) evidence extraction and verification of claim truthfulness, and 3) detection and control of information dissemination patterns. This paper then critically examines how each type of technology socially constructs 'factuality' compared to the ways in which the traditional journalism approach has defined 'factuality.' In sum, AI fact-checking involves the automation of the process through which external 'objective facts' and their requirements are defined in technically sensible ways, and the machine 'filters' and 'matches' information based on whether it meets these criteria. Ultimately, this process reconstructs the nature of facts within the 'requirements of facts' that the technology can construct. This study emphasizes the need to move beyond the current focus on whether AI technology can be adopted or implemented, and instead calls for a detailed examination of the specific conditions and methods under which AI is deployed within the context of fact-checking. This study reveals, through the analysis of case studies, that the definition of 'fake news' becomes highly ambiguous, depending on which aspects of the complex fact-checking process are automated by AI, and that the concept of 'factuality,' traditionally emphasized in journalism, can also be subject to political (re)definition.

Keywords: Fact-check, Artificial Intelligence, Factuality, Fake News, Journalism

국문초록

광범위하게 생산되고 빠르게 유통되는 인터넷 상의 오정보, 허위조작정보, 유해정보에 대한 대응책으로서 인공지능 기술을 통해 팩트체크를 자동화하려는 움직임이 지속되고 있다. 하지만 이는 팩트체크 과정 안에서 인공지능이 실행, 배치되는 조건이나 방식을 구체적으로 고찰한 후에 그 안에서 사실성이 협상되고 구성되는 방식을 이해하는 과정이 선행되지 않은 상태에서 기술 중심적인 해결책으로서 인공지능 기술이 제안되는 양상을 보인다. 이에 문제의식을 갖고 이 논문은 팩트체크 '기술'로서 인공지능이 '사실'을 구성하는 방식을 국내외 인공지능 팩트체크 기술사례를 통해 검토함으로써 인공지능 팩트체크가 구성하는 '사실성'이 사실을 더함으로써 사회적 실재를 구성하던 팩트체크 저널리즘과 어떻게 다른지 검토한다. 이를 위해 2023년 10월 기준 국제 팩트체킹 네트워크(International Fact-Checking Network, IFCN)에 의해 인증 받은 세계 팩트체크 인증기관들이 발표한 인공지능 팩트체크 기술과 세계적인 싱크탱크인 미국의 랜드 연구소(Rand Institute)에서 허위조작정보에 대한 대항(fighting disinformation)할 목적으로 구축한 온라인 데이터베이스에서 발표한 인공지능 팩트체크 기술 사례들을 검토하고, 각 기술 사례들의 목적과 '팩트 체크' 근거에 따라 1) 주장의 탐지, 2) 증거 추출과 주장의 진실성 검증, 3) 정보 확산 형태의 탐지와 통제로 유형화했다. 이어서 각 유형이 저널리즘이 구축해 온 '사실성'을 어떻게 기술사회적으로 구성하는지 비판적으로 검토했다. 요약하면, 인공지능 팩트체크는 외부세계에 존재하는 '객관적 사실'이나 그 요건을 사전적, 기술적으로 규정하고 해당 요건에 맞지 않거나 맞는 것을 걸러내고 대조하는 과정을 자동화하여 궁극적으로는 기술이 구성할 수 있는 '사실의 요건' 안에서 사실 자체를 재구성한다. 이 연구는 인공지능 기술의 도입 여부나 가능성 자체에 집중하고 있는 지금 상황에서 벗어나 팩트체크의 맥락 안에서 인공지능이 실행, 배치되는 조건과 방식을 구체적으로 살펴볼 필요가 있음을 강조하고. 팩트체크의 복잡한

과정 중에 인공지능에 의해 자동화되는 것이 무엇인가에 따라 '가짜뉴스'가 지칭하는 것은 매우 느슨하게 정의되며 저널리즘에서 강조해 온 사실성 역시 정치적으로 (재)정의될 수 있음을 기술 사례 분석을 통해 드러내고 있다.

핵심어 : 팩트체크, 인공지능, 사실성, 가짜뉴스, 저널리즘

1. 서론

소위 '가짜뉴스(fake news)'라고 불리는 것에 대한 사회적, 정치적 관심이 뜨겁다. 현 정부 출범 직후부터 가짜뉴스 규제는 정부 주요 과제로 강조되었고 방송통신위원회와 기획재정부는 '가짜뉴스 대응을 위한 팩트체크 사업' 예산을 증액해서 편성했다. 2023년부터 방송통신심의위원회가 착수한 '가짜뉴스 대응 방안 마련을 위한 연구', 문화체육관광부의 '가짜뉴스 신속 대응 자문단' 구성, 네이버의 뉴스 서비스 개선 및 가짜뉴스 대응 방안 마련을 위한 '뉴스혁신포럼' 출범 등 연일 '가짜뉴스'라 불리는 것들에 대응하기 위한 움직임이 눈에 띠게 늘어나고 있다.

'가짜뉴스'에 대응하는 과정에서 떠오른 또 하나의 키워드는 '팩트체크'다. 디지털 미디어 환경 안에서 광범위하게 생산되고 급속하게 퍼지고 있는 '가짜뉴스'에 대한 대응책으로서 팩트체킹(fact-checking)이 새로이 만들어진 것으로 오인하는 경우가 많지만 미디어를 통해 전달되는 정보의 사실관계를 파악하는 것은 저널리즘의 오랜 관행이자 언론사 본연의 기능이었다. 다만 디지털 환경 속에서 오정보, 허위조작정보, 유해정보 등 오염된 정보가 대량으로 생산되어 빠르고 광범위하게 유통, 소비되는 상황은 전통적 의미의 뉴스 및 보도에 대한 신뢰도와 연결되면서 저널리즘 영역에 위기의식을 가중시켰고 이는 저널리즘의 책임이자 실천으로서 강조되어 온 팩트체킹이 좀 더 조직적이고 제도적으로 구성될 수 있는 맥락과 당위성을 제공했다. 유럽 대학 연구소(European University Institute) 산하의 유럽 디지털 미디어 옵저버토리(European Digital Media Observatory)는 디지털 미디어 환경이 정착한 이후 영국 및 유럽 연합 국가 내에 약 백여 개의 팩트체킹 기관이 있는 것으로 보고하고 있으며(EDMO, n.d.), 미국의 워싱턴 포스트(Washington Post), 공영 라디오 방송국(National Public Radio), 영국의 로이터(Reuters)등 세계 유수 언론 기관 내에도 팩트체킹 부서가 신설되어 왔다. 국내에서는 서울대학교 언론정보연구소에서 2017년 SNU팩트체크센터를 설립했고, 한국 기자협회, 방송기자연합회, 한국PD연합회와 사회적협동조합 빠띠가 공동 출자하여 설립했던 〈팩트체크넷〉이 2021년 출범했으며, KBS, MBC, SBS, JTBC, 연합뉴스 등 각 언론사 및 방송사에서 팩트체킹 전담 부서를 별도로 설치하며 뉴스 생산 과정의 일부로 강조해 왔다.

이처럼 팩트체크는 정보의 사실관계를 파악하는 언론사의 오랜 관행에 바탕하여 디지털 미디어 환경에서 촉발된 정보위기에 대응하기 위한 저널리즘 영역 내부의 자성적인 움직임으로 구성되어 왔다. 하지만 팩트체크는 또 다른 한 편으로는 오정보, 허위조작정보, 유해정보, 정치적 저항담론 등이 광범위하게 생산되고 빠르게 유통, 소비되는 디지털 미디어 환경을 사회적으로 문제화하는 과정에서 이를 해결하기 위한 정부 주도 사업 및 정책 방향의 성과물로 두드러지기도

했다. 특히 팩트체크에 대한 정부주도 사업들은 '가짜뉴스'라는 사회적 문제를 효율적으로 해결하기 위한 기술을 개발하는 데 관심을 보여 왔는데 그 과정에서 인공지능 기술에 대한 관심이 두드러졌다. 대표적으로 2019년 방송통신위원회 및 산하 기관인 시청자미디어재단이 기획한 '인터넷 환경의 신뢰도 기반 조성 사업'은 팩트체크 시스템을 구축하고 고도화하기 위해 예산을 편성했다. 이러한 정부 주도의 지원 사업의 일환으로 SNU팩트체크센터의 이준환 교수 연구팀은 2021년부터 한국형 자동화 팩트체크 모델 개발에 착수했다. 이 모델은 한국어 위키피디아와 인터넷 뉴스를 바탕으로 주장과 근거 중심의 학습데이터를 8만여 건 구축하고, 학습 데이터에 기반하여 입력한 주장에 대해 참, 거짓을 판별하는 서비스이다. 개발된 모델은 시범 서비스의 형태로 인공지능 기반 자동화 팩트체크 서비스인 AINET에 2022년 말까지 한시적으로 공개되었으며, 연구 목적으로 공개된 피버(FEVER) 베이스라인 모델과 학습 데이터는 서울대, 숭실대 등의 후속 연구를 통해 지속적으로 개선되고 있는 상황이다. 현 정부 초기에 팩트체크 시스템 안정화 및 고도화를 위한 예산이 한 차례 삭감되었으나 2024년도 예산안에서 다시 일부 증액되면서 인공지능 팩트체크는 디지털 환경에서 가짜뉴스에 대응하기 위한 방안으로 정책적으로 꾸준히 강조되어 왔다.

하지만 저널리즘의 관행이면서 제도로서 구축되어 온 팩트체크에 인공지능 기술을 활용한다는 것은 말처럼 단순하지 않다. '가짜뉴스'가 전지구적으로 통용되기 시작한 직후에 눈에 띄게 늘어난 자동화된 사실확인 기술의 현황을 파악하고 한계를 지적한 연구(오세욱, 2017)나 국내에서 인공지능 팩트체킹 연구가 진행된 이후 현장참여자들의 심층인터뷰를 수행하여 한국형 인공지능 팩트체킹의 방향성을 탐색한 연구(박소영·이정현, 2023) 등이 지적하듯 인공지능과 팩트체킹의 결합은 그것이 기술적으로 구현 가능한 것인가의 관점에만 머물러 검토하기 보다는 이것이 팩트체크의 지형과 사실을 구성하는 역학을 어떻게 재구성할 것인지 탐구하는 과정이 수반되어야 한다. 이미 다양한 영역에서 경험했듯 사회에 이미 존재하던 제도, 구조, 체계에 인공지능 기술이 통합된다는 것은 사회적 구성물에 기술이 첨가되는 것에 그치는 것이 아니라 새로운 기술적 조건 위에 이들을 새롭게 배치되고 구성되는 결과를 낳는다. '인공지능 팩트체크' 역시 인공지능 기술이 저널리즘의 제도, 구조, 체계에 통합됨으로써 기존에 통용되던 개념이나 담론을 재구성할 가능성이 있음을 인식할 필요가 있다. 유사한 문제의식을 공유하면서 오세욱과 황구현(2018)은 자동 사실 확인 기술 사례들에서 '팩트'가 구성되는 형식적 요건을 탐색했으며, 박대민(2023)은 인공지능의 신뢰가능성이 언론이 구성해 온 신뢰가능성과 통약할 수 있는 가능성에 대해 검토한 바 있다.

이에 이 연구는 인공지능이 팩트체킹 과정에 활용되면서 팩트체크를 포함하는 저널리즘의

사실확인 관행이 실천적, 제도적으로 구성해 온 '사실'이라는 것이 어떻게 재구성되는지 이론적으로 고찰하고 이것이 국내의 '가짜뉴스' 대응책과 어떻게 기술사회적으로 결합할 수 있을지 검토할 필요성이 있음을 지적하고 있다. 즉, 저널리즘이 구성하고 추구하는 '사실'이라는 것은 팩트체크와 인공지능 기술 등 다양한 기술사회적 요소들의 역학에 따라 끊임없이 변화를 겪고 있으며, 인공지능 팩트체킹 기술의 작동 방식이 '사실'을 무엇으로 정의하며 '가짜뉴스'의 정치사회적 의미와 어떻게 경합하거나 협력하고 있는지 살펴보아야 한다. 이 같은 문제의식을 바탕으로 이 논문은 팩트체크 '기술'로서 인공지능이 '사실'을 구성하는 방식을 국내외 인공지능 팩트체크 기술사례를 통해 검토함으로써 인공지능 팩트체크가 구성하는 '사실성'이 기술사회적으로 어떻게 정의될 수 있으며 이것이 국내 '가짜뉴스' 담론에 어떤 영향을 미칠 수 있을지 비판적으로 탐구한다. 이를 위해 실제 팩트체크 과정에서 활용 가능한 인공지능 기술을 개발하거나 서비스하고 있는 기술 사례들을 조사하였는데 조사 대상은 2023년 10월 기준 국제 팩트체킹 네트워크(International Fact-Checking Network, IFCN)에 의해 인증 받은 세계 팩트체크 인증기관들이 발표한 인공지능 팩트체크 기술과 세계적인 싱크탱크인 미국의 랜드 연구소(Rand Institute)에서 허위조작정보에 대한 대항(fighting disinformation)할 목적으로 구축한 온라인 데이터베이스이다. 조사 대상에 포함된 국내 기술의 수는 상대적으로 매우 적은 편인데 이는 아직 국내에서는 대중 일반에게 공개되어 실제 서비스가 가능한 수준까지 진전을 이룬 인공지능 팩트체킹 기술 사례가 많지 않기 때문이다. 이후 각 기술 사례들을 목적과 '팩트 체크' 근거에 따라 유형화하고 이를 바탕으로 각 유형이 만들어내는 '사실'의 기술사회적 의미를 비판적으로 검토했다.

이 논문의 목적은 기술사례 연구를 통해 인공지능 시대 저널리즘의 사실성을 이론적으로 탐색하는 데 있다. 이를 위해 이 논문은 지식사회학의 관점을 바탕으로 사실이 사회적인 합의를 통해 상호주관적으로 구성된다고 보고, 언론을 지식과 사회적 실재를 구성하는 과정에서 개인의 사회적 실재를 구성하는 데 중요한 역할을 하는 타자로 개념화한다. 이를 바탕으로 인공지능 팩트체크가 어떻게 사회적 실재로서의 사실을 재구성하는지 이론적으로 탐구하고자 한다. 이는 '사실'의 새로운 맥락으로서 디지털 기술과 '사실'의 대척점에서 정의되는 '가짜뉴스'의 의미를 검토하는 것으로 시작한다. 이어서 탈진실 시대의 대표적인 사실확인 제도로서 팩트체크 저널리즘이 내용 검증과 검증절차의 투명성 보장을 통해 '사실'을 구성해 온 방식을 역사적, 이론적으로 살펴본다. 이후 국내외 인공지능 팩트체킹 기술의 주요 연구 사례가 어떤 방식으로 '사실'을 판정하거나 그 과정에 기여하는지 살펴보고 그 과정에서 '사실'이 기술사회적으로 (재)구성되는 방식을 탐구한다. 결론에서는 팩트체크 저널리즘이 구성하는 사실과 대비되는 인공지능 팩트체크가 재구성하는 사실이 '가짜뉴스'를 어떻게 정의할 수 있을지 논하며 이 같은 가짜뉴스의 의미화 방식에

따라 가짜뉴스에 대한 대응책이 어떻게 달라질 수 있을지 서술한다. 인공지능이라는 기술과 사회적 담론 사이의 공모 관계를 보다 가시적으로 드러내고 급변하는 저널리즘 생태계에서 '사실'이라는 것이 과연 무엇인지 비판적으로 검토함으로써 인공지능과 저널리즘이 공진화할 수 있는 가능성과 방향성을 탐구하는 데 이 연구의 의의가 있다.

2. 문헌검토: '사실'이 처한 두 가지 맥락

1) 디지털 기술과 '사실'

2000년대 중반 이후 디지털 기술이 급속도로 발전하고 플랫폼 기업이나 인공지능 기술이 정보를 생산 및 유통하는 전 과정에 지대한 영향을 미치게 되면서 '사실'을 둘러싸고 인식론적 전환이 일어나고 있다. 계몽주의를 바탕으로 한 서구 근대사회의 도래는 종교, 주술, 왕권 등의 통치성에서 '사실'을 발견하는 과학적 이성으로의 전환을 의미했다(Foucault, 2004/2012). 과학적 이성으로의 전환은 객관적이고 물리적인 사실을 발견하는 자연과학의 발전 양상으로 나타난 동시에 사회적인 영역 안에서도 사실이라는 것을 합의된 사회적 체계이자 제도로써 구축하는 과정으로 나타났다. 셸러(Max Scheler)나 만하임(Karl Manheim)은 지식사회학이라는 용어를 통해 과학적인 사실로 여겨지는 지식은 사회적으로 구성되며 이렇게 형성된 지식이 실재(reality)를 구성하고 이해하는 원료가 된다고 주장했다(윤태진, 2011). 즉, 지식은 객관적으로 존재하기보다는 사회적 요인들의 상호작용 속에서 생성되고 성장하여 구체적인 내용이 결정되며, 이를 바탕으로 개인이 인식한 실재는 제도화와 합법화의 과정을 거쳐 다시 지식으로 명명된다(Berger & Luckman, 1966). 지식사회학의 관점을 따르면 개인이 실재를 구성하는 방식은 외부 세계에 존재하는 객관적인 사실(fact)에 의해서가 아니라 상호주관적으로 형성한 사회에 대한 의미와 타인들이 구성한 의미 사이에 교류와 공유를 통해 구성되기에 실재는 사실의 총합이 아니며 오히려 사실은 사회적 실재로서 존재한다(윤태진, 2011).

이 과정에서 미디어는 현대사회에서 개인이 상호작용하는 가장 중요한 타인(significant other) 중 하나로 사회를 끊임없이 재현함으로써 개인이 현실을 이해하고 판단하여 지식을 형성하는 데 핵심적인 기준과 준거를 제공해 왔다(Silverstone, 1999). 그런데 2000년대 중반 이후 두드러진 미디어 생태계의 변화는 사회적 실재로서 존재하는 사실 자체에 대해 중요한 인식론적 전환을 가져왔다. 포털이나 사회관계망서비스 등 디지털 미디어 플랫폼이 신문이나 지상파 방송으로 대표되는 전통적인 미디어를 넘어 정보를 유통하는 핵심적인 채널로 부상하면서 미디어

이용자가 구성하는 사회적인 실재는 외부 세계에 존재하는 사실과 더욱 큰 괴리를 만들어냈다.

일례로 협업 필터링(collaborative filtering)을 바탕으로 만들어진 추천 알고리즘은 개인화된 추천 시스템이라는 명분 아래 플랫폼 기업에 의해 '유사한 집단'으로 분류된 이들의 의견만 듣게 되는 반향실 효과(echo chamber)나 유사한 내용의 정보에만 반복적으로 노출되면서 자신의 의견에 갇히게 되는 필터 버블(filter bubble)을 만들어냈다. 이후 딥러닝 기술의 발전은 정보의 분류뿐 아니라 생성에까지 가담했다. 딥페이크(Deep fakes)는 기존에 컴퓨터 그래픽을 통해 만들어지던 합성 이미지나 영상을 인공지능 기술을 통해 더욱 정교하게 만들어냈고, 적대적 생성 신경망(Generative Adversarial Network)이나 합성곱 신경망(Convolutional Neural Network) 등 생성형 인공지능은 이미지의 편집이나 합성뿐 아니라 생성까지도 가능하게 만들었다. 2023년 3월 공개된 GPT-4는 이미지 뿐 아니라 텍스트도 책, 기사, 논문 등의 형식에 맞추어 인간이 작성한 것과 구분이 어려운 수준까지 생성해 낸다. 이 같은 기술을 바탕으로 한 미디어와 상호작용하며 구성해내는 사회적 실재는 더 이상 외부세계에 존재하는 객관적인 사실을 향해 있지 않다.

디지털 및 인공지능 미디어 생태계와의 상호작용으로 구성되는 사회적 실재는 2016년에 《옥스퍼드 사전(Oxford Dictionary)》에 의해 올해의 단어로 선정된 '탈진실(post-truth)'로 대변된다. 탈진실은 더 이상 진실 혹은 사실이 무엇인지 확인하려 하지 않고 본래 자신이 믿던 것과 믿고 싶은 것만을 진실로 믿는 것으로, 사실이 더 이상 중요하지 않고 사실에 동의하지 않는 시대를 의미한다(정성욱, 2021). 이는 포털이나 사회관계망서비스 등 디지털 미디어 플랫폼을 통해 사회에 대한 상호주관적 의미가 구성되며 정보의 형태나 정보를 선택하고 생성하는 과정마저도 딥러닝 기술이나 알고리즘이 적극적으로 개입하는 시대에 '사실' 자체가 지속적으로 재맥락화되고 있음을 시사한다.

디지털 기술 및 인공지능 기술은 더욱 정교하면서도 적극적인 방식으로 사회적 실재로서 사실을 구성하는데 개입하기 시작했고 이는 사실이 구성되는 방식뿐 아니라 '사실' 자체에 대한 인식론적인 변화를 가져오고 있다. 뿐만 아니라 기존에 지식을 구성함으로써 사회적 실재로서 사실을 구성하는 데 결정적이고 독점적인 역할을 해 온 전문가 집단, 제도 및 미디어의 위치가 위협을 받기 시작했다. 전문 기관이나 전문가 뿐 아니라 일반 이용자까지 사회적 실재를 구성하는 과정에 적극적으로 참여할 수 있게 되었고, 그 과정에서 사실보다는 취향이나 의견을 추구하게 된다. 특히 디지털 및 인공지능 기술이 현대사회에서 개인이 사회적 실재로서 사실을 구성하는데 매우 중요한 타인으로 작동하면서 사실을 합의해 가는 과정이나 방식 자체에 대한 사회적인 재고가 이루어지고 있다.

2) 기호로서 가짜뉴스와 '사실'

디지털 및 인공지능 기술이 미디어와 결합하면서 만들어진 탈진실 시대의 대표적인 생산물 중 하나는 '가짜뉴스'다. 가짜뉴스와 디지털 기술의 연관성을 부인하는 학자는 없으며 가짜뉴스는 앞서 살펴 본 탈진실 시대를 야기한 핵심적인 요소 중 하나로서 서술되어 왔다(McIntyre, 2018; Peters, 2017). 하지만 기호로서 가짜뉴스는 '사실'의 대척점에서 유동적으로 정의되면서 사실을 구성하는 새로운 맥락으로 기능하고 있기에 가짜뉴스가 정의되는 방식을 사실을 구성하는 또 하나의 맥락으로 살펴 볼 필요가 있다.

이미 정치사회적으로 폭넓게 통용되고 있는 개념임에도 불구하고 '가짜뉴스'는 명확한 사전적 의미에 기반한 개념이라기보다는 경합 중인 담론이자 의미적 합의가 이루어지지 않은 기호에 가깝다. 가짜뉴스는 의미에 대한 정확한 합의를 거치기 전에 전지구적으로 통용되는 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 특별 검사팀은 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 입장에 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 포괄되어 왔다.

실제로 많은 언론인과 학자들이 가짜뉴스라는 용어 사용 자체에 대한 문제를 지적해 왔다. 밥 우드워드(Bob Woodward) 워싱턴포스트 기자는 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(Marwick & Lewis, 2017), 이를 허위(false) 여부와

해를 끼칠 의도(intent to harm)를 판단하여 오정보 또는 단순허위정보(misinformation), 허위조작정보(disinformation), 유해정보(mal-infromation)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 기호를 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(황용석, 2017; 황창근, 2017). 뉴스의 형식을 강조했던 것은 가짜뉴스를 새로운 사회현상으로 보고 이미 온라인상에 존재하던 허위사실, 단순 정보보고, 루머 등과 구분하기 위해서였다. 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 접근법은 사실상 텍스트로서 가짜뉴스가 어떤 '형태'를 갖는지에 따라 구분하는 것이었다.

한국언론진흥재단의 가짜뉴스에 대한 일반 국민 온라인 설문조사 역시 어떤 형태의 정보를 가짜뉴스로 인식하는지에 대한 질문이 담겨있다. 이에 대한 대중들의 대답을 살펴보면 정보와 텍스트의 형태면에서도 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). 즉, 가짜뉴스에 대한 담론이 팽배해짐에 따라 학술적으로 엄밀하게 구분되는 오정보, 허위조작정보, 유해정보 등이 서로 뒤섞여 가짜뉴스라는 하나의 기표에 수렴되는 양상을 보였다.

용어 자체가 가짜뉴스의 복잡다단한 생태계를 정확히 진단하지 못하게 하는 경향 때문에 국내에서도 2018년을 전후하여 가짜뉴스라는 용어를 폐기하거나(김민정, 2018) 해외 문헌처럼 가짜뉴스의 개념을 보다 세분화하여 이해하자는 주장(정은령, 2018)이 대두되었다. 하지만 여전히 대체 용어나 명확한 개념화에 대한 숙고 없이 가짜뉴스가 하나의 느슨한 담론으로 언론이나 정치캠페인을 통해 소비되고 있는 실정이다. 일례로 윤석열 대통령이 2023년 4.19 기념식에서 "가짜뉴스가 민주주의를 위협한다"고 발언한 다음 날 문화체육관광부는 〈'악성 정보 전염병' 가짜뉴스 퇴치 전면 강화〉라는 보도자료를 배포하여 국가 차원의 가짜뉴스 대응방안을 발표했다. 세부 내용은 언론진흥재단 내에 '가짜뉴스 신고상담센터'를 설치하고 가짜뉴스를 유형화하는 작업을 시작한다는 것이었다(문화체육관광부, 2023). 가짜뉴스에 대한 정부의 관심은 현 정부에서 처음 시작된 것은 아니다. 문재인 정부 때도 '가짜뉴스' 규제를 추진했다. 관련하여 2018년 10월 방송통신위원회를 중심으로 '허위조작정보 대응 범정부 종합대책'을 발표를 계획했으나 연기되었고, 이후 방송통신위원장이 교체되면서 별도의 '가짜뉴스' 규제책을 마련하지는 않았지만 '허위조작정보에 관한 전문가회의'를 이어 받고 전문가회의의 결과로 '허위조작정보 문제해결을 위한 제안'을 보고받아 채택했다. 당시 '허위조작정보'는 "허위사실임을 알면서, 정치적 경제적 이익 등을 얻을 목적으로, 정보 이용자들이 사실로 오인하도록 생산·유통된 모든 정보"로 언론사의 기사, 패러디, 풍자, 정치적 견해 등은 제외되었다(방송통신위원회, 2020). 반면 현 정부는 다시 '가짜뉴스'라는 기표를 사용하며 본격적인 '규제' 계획을 앞세웠다. 여기에는 언론사에 대한 규제 계획이 두드러졌는데 현재 방송통신위원장이 "가짜뉴스를 만든 언론사 원스트라이크 아웃"을 발표한 데서 이를 엿볼 수 있다. 또한 지금까지 가짜뉴스가 콘텐츠의 형태에 주목하여 정의되었던 것과 달리 '내용'에 집중할 가능성을 내포하는데, 문화체육관광부의 가짜뉴스 퇴치에 관련된 보도자료에서 "정부 정책 관련 가짜뉴스 사례"나 "정부 신뢰도를 떨어뜨리는 허위·왜곡 보도"에 대응하겠다고 밝힌 대목에서 '어떤 내용'을 중점적으로 살필 것인지에 대한 맥락을 엿볼 수 있다.

이처럼 국내에서 가짜뉴스는 온라인 환경에서 생산, 소비, 유통되는 다양한 층위의 오염된 정보를 정확히 구분하지 않고 때로는 서로 이질적인 사회현상을 지칭하는 공통 담론이자 기호로 구성되어 왔다. 이 과정에서 '가짜'는 여러 층위에서 '사실과 다른' 뉴스를 느슨하게 통칭했다. 이때 '가짜'와 대척점에 있는 '사실'은 객관적으로 존재하는 사실과 다른 오정보 뿐 아니라 구성원들에 의해 사회적 합의에 이르지 못한 사회적 실재 등을 광범위하게 포괄하고 있다. 앞선 절에서 살펴 본 디지털 기술이 사실성에 위기를 가져 온 기술적 조건이었다면 무엇을 가짜뉴스로 지칭하며 어떻게 제도화, 합법화할 것인지에 대한 담론이나 사회적 합의는 개인이 사회적 실재로서 사실을 구성하는데 매우 중요한 역할을 수행하고 있다.

3. 팩트체크와 저널리즘의 사실성 구성

앞서 살펴보았듯 디지털 기술과 기호로서 가짜뉴스는 우리가 '사실'을 사회적으로 이해하는 데 새로운 맥락을 제공함과 동시에 기존에 개인이 사회적 실재를 구성하기 위해 상호작용하던 전문가, 미디어 등의 집단이 사실을 '확인'하고 '검증'하는 역할과 방식을 재고하게 하는 결과를 낳았다. 특히 가짜뉴스로 통칭되는 오정보, 허위조작정보, 유해정보 등 온라인상에 급속도로 퍼지는 정보 오염 현상은 저널리즘에 전례 없는 위기를 초래했다. 언론은 사회를 관찰하고 이를 신뢰할 수 있는 정보의 형태, 즉 '사실'로 구성해내는 핵심기관이며, 현대사회에서 개인이 사회적 실재를 구성하고 조정하기 위해 상호작용하는 핵심 기관으로 기능해 왔다. 하지만 가짜'뉴스'라는 명칭에서도 드러나듯 디지털 미디어를 통해 생산 및 유포되는 다양한 형태의 오염된 정보 중 '뉴스'의 형태를 띠거나 뉴스 매체에 의해 오염된 정보가 생산되거나 유통되는 현상은 저널리즘이 디지털 기술과 가짜뉴스의 시대에 가져야 할 역할에 대해 재고할 필요성을 높였다.

국내에서 가짜뉴스를 조망하는 시각은 '가짜뉴스 대 전통 언론,' '거짓 대 진실'이라는 이분법적 구도로 진행되었다(정은령, 2019, 50쪽). 즉, 정보의 진위 여부만큼이나 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 짜깁기하거나 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 검증 없이 그대로 전달하는 '받아쓰기 저널리즘'(정준희, 2019), 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 연관하여 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. 이 같은 배경에서 국내 팩트체크는 언론 외부에서 생산되는 검증되지 않은 정보에 대한 대응책이자 반격의 성격으로 저널리즘 내부에서 시작되었다.

이 장에서는 이렇게 제도화된 팩트체크 저널리즘이 어떻게 '사실'을 구성하는지에 대해 서술한다. 즉, 팩트체크 저널리즘이 사회적 실재로서의 사실을 구성하는 과정에서 어떻게 '내용의 사실성'을 검증하고 '절차의 투명성'을 보장하는지 살펴보고자 한다.

1) 내용 검증을 통한 사실의 구성

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 & Mattews, 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 객

관주의 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 기자가 취재한 내용 및 발언의 정확성을 기자가 직접 평가한다는 점에서 객관주의 규범과 대조를 이룬다. 이것은 단순히 인용의 정확성을 확인하는 과정을 넘어서 보도하고자 하는 발언이 정확한지 기자들이 독자적인 조사를 거쳐 해석한 판정결과를 독자에게 제시하는 것이다(Amazeen, 2013). 언론이 "단순한 전달자가 아닌 책임 있는 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 언론은 '참' 혹은 '사실'이라는 것을 일차적으로는 '의견'과 구분하면서 이차적으로는 제도적 합의의 과정을 구축해 보장하여 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 실

명을 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). 취재기자, 팀장, 데스크, 편집기자, 편집부장, 편집국장 등 여러 단계를 거치는 게이트키핑 과정은 기사에 포함된 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 외부세계에 존재하는 객관적인 무언가를 찾는 것이 아니라 복합적인 검증의 원칙과 절차를 수반하여 사실 확인의 목적을 달성하여 왔음을 보여주며 팩트체크 저널리즘 역시 이와 유사한 과정을 따른다. 하지만 객관주의 저널리즘을 기반으로 한 전통 언론의 맥락에서 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 검증할 것인지부터 어떻게 해석하는가에 이르기까지 주관성과 당파성의 개입을 부정하기 어렵다. 우신스키와 버틀러(Uscinski & Butler, 2013)가 지적한 팩트체킹의 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 수밖에 없다. 또한, 사실성을 판단하기 위해 근거 자료를 선택하고 걸러내는 과정도 주관성과 당파성이 개입될 여지가 있기에 사실의 판정이라는 것도 주관적일 수밖에 없는 한계를 내포한다. 실제로 이런 이유로 동일한 검증대상에 대해서도 상이한 판정결과가 나올 수 있다(Graves, 2017).

2) 검증절차의 투명성

이처럼 내용을 검증함으로써 구성되는 사실성이 가질 수밖에 없는 필연적인 한계를 팩트체크는 절차의 투명성을 통해 보완한다. 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 국제 팩트체킹 네트워크(International Fact Checking Network: IFCN)는 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 세부항목[1]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 받았다. 이 같은 준칙을 통해 팩트체크가 강조하는 것은 객관성보다는 투명성이다. 예를 들어, 팩트체크닷오르그(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).

즉, 팩트체크가 추구해 온 사실성은 발견해야 할 사회적 '실재'가 아니라 정보의 사실성을 확보하는 과정을 투명하게 공개하는 데 있어 왔다. 앞서 언급했던 르포, 직접인용, 통계, 게이트키핑, 정보원 실명공개 등의 제도 역시 정보의 객관성보다는 투명성을 강조하고 있다. 이에 따라 IFCN의 인증을 받은 팩트체크 기관들은 팩트체크 검증대상 선정의 방법, 원칙, 판정기준을 명시적으로 밝히며, 검증에 동원된 정보원은 원자료에 접근할 수 있도록 하이퍼링크로 자료를 제공하고 있다. 검증결과를 판정한 이유를 제시하거나 오류를 수정한 경우를 공개함으로써 투명성을 통해 사실성을 확보해 온 것이다.

종합하면, 디지털 기술의 발전과 가짜뉴스라는 기호의 범람 속에서 사회적 실재로서 사실을 구성하고 있는 팩트체크 저널리즘은 르포, 현장인용, 정보원 검증, 게이트키핑 등 언론의 전

1) 5개의 기본 준칙에 따르는 세부항목은 1) 어느 한 쪽에 편증된 팩트체킹을 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 팩트체킹 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(IFCN, 2016).

통적인 사실성 제도를 복합적으로 거치며 기사를 통해 전달하려는 '사실'의 진위여부를 교차 및 반복 검증한다. 이 과정은 다시 말하면 정보원, 현장, 언론인, 데스크, 팩트체커, 이용자 등의 이해관계자들 각각의 '사실'의 조각들을 모으고, 쌓아가고, 맞추어 가며 상호주관적으로 사실을 구성하고 궁극적으로 언론을 통해 보도되는 사회적 실재로서 사실을 전달하는 것이다. 언론사 본연의 기능이자 언론의 전통적인 사실성 제도와 팩트체크가 구분되는 지점은 여기에서 더 나아가 '어떻게' 사회적 실재로서 사실이 구성되었으며 이를 '어떤 과정을' 통해 검증하여 사실이라고 부를 수 있는 것을 합의를 통해 구성해냈는지 투명하게 공개함으로써 사회적 실재의 구성에 일익을 담당하는 또 하나의 '중요한 타인'으로 자리매김하고 있다는 데 있다.

팩트체크 저널리즘의 '사실성'이 사실이 상호주관적 검증을 통해 뼈대를 갖추고 살을 입는 과정을 투명하게 공개함으로써 확립된다는 점은 후술할 인공지능 팩트체킹과 명확히 구분되는 지점이다. 다음 장에서는 인공지능 팩트체킹 기술사례 분석을 통해 팩트체킹의 과정에 인공지능을 활용함으로써 사실을 '모으고, 쌓고, 맞추어 나가는' 과정을 통한 사실의 구성이 어떻게 달라질 수 있는지 살펴보고자 한다.

4. 인공지능 팩트체킹과 기술사회적 '사실'

온라인 상 정보 오염 현상을 이르는 여러 하위 개념을 통칭하는 가짜뉴스가 범람하면서 팩트체크가 대안으로 떠오른 이유는 내용의 사실성 여부 판단을 통해 건강한 온라인 저널리즘 생태계에 기여하고자 하는 내부적인 성찰과 외부의 기대 때문이라고 할 수 있다. 하지만 정보의 진위여부 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 사실의 검증 과정에 개입함으로써 '사실'과 '검증 과정' 양쪽 모두를 기술사회적으로 새롭게 구성하고 있다.

인공지능 팩트체킹은 인공지능 기술을 팩트체크 과정에 활용하는 것으로 광범위하게 이해되고 있지만 이는 인공지능에 의해 자동화되는 것이 무엇인가를 기준으로 보다 엄밀하게 구분해 정의될 필요가 있다. 일례로 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 사실 여부의 검증 대상이 되는 특정한 '주장의 탐지'를 의미하며, 의도, 형식, 수용자 반응, 문장의 요소 등 정보의 형태 요소에서 파악한 단서를 토대로 특정한 주장이 사실 검증의 대상이 될 만한 가치(check-worthiness)를 지니는지 판단하는 것으로 사실 확인과는 구분된다(Guo, Schlichtkrull, & Vlachos, 2002). 현재로서는 팩트체크의 전 과정이 인공지능에게 맡겨진 사례는 없으며 여러 기술들이 여전히 개발 중에 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 일부만을 인공지능 기술을 통해 자동화하는 방식으로 이루어지고 있다.

이 때 인공지능 기술의 팩트체크 과정 중 '어느 단계'를 자동화하는지, '무엇을' 근거로 사실을 입증하는지에 따라 팩트체크 저널리즘이 역사적으로 구성해 온 사실성의 지형이 달라진다. 인공지능 기술이 도입되는 순간 사실 지형에 차이와 전환이 결과적으로 나타나며, 이렇게 인공지능이 사실성을 진단하는 대상이 사회적 실재라는 점에서 결과적으로는 저널리즘의 신뢰성에 결정적인 영향을 미친다. 때문에 사실 확인 과정에 인공지능 기술을 도입하는 것은 기술적으로 구현이 가능한가의 여부에만 의존하는 기술 결정론적 시각으로 운용할 수는 없다. 그보다는 인공지능 기술을 통해 '어느 단계'를 자동화하고 '무엇을' 근거로 사실을 입증할 수 있으며, 이 때 구성되는 사실이라는 것이 무엇인지에 대한 숙고를 바탕으로 이루어져야 할 과업이다.

이에 이 절에서는 2023년 10월 기준 국제 팩트체킹 네트워크(International Fact-Checking Network, IFCN)에 의해 인증 받은 세계 팩트체크 인증기관들이 발표한 인공지능 팩트체크 기술과 세계적인 싱크탱크인 미국의 랜드 연구소(Rand Institute)에서 허위조작정보에 대한 대항(fighting disinformation)할 목적으로 구축한 온라인 데이터베이스에 기록된 인공지능 팩트체킹 기술 사례들(Table 1 참고)을 이용 목적과 '팩트 체크' 근거에 따라 주장의 탐지, 증거 추출, 정보 확산의 형태 검증으로 유형화했다. 이를 바탕으로 앞서 서술한 내용 검증을 통한 사실성의 구성과 투명성 보장의 과정이 인공지능 팩트체크에서 어떻게 구현되는지 살펴보고, 그 과정에서 사실이 어떻게 재구성될 수 있는지 비판적으로 검토하고 있다.

Table 1. Cases of AI Fact-checking Technologies

IFCN 팩트체크 인증기관	랜드 연구소
클레임헌터(Claim Hunter) 클레임버스터(Claim Buster) 보토미터(Botometer) 혹시(Hoaxy)	
풀팩트 AI(Full Fact AI) 스쿼시(Squash) 레이더(Radar)	봇슬레이어(BotSlayer) 톱피버(Top Fiber) 헤밀턴 2.0(Hamilton 2.0) 이피코션트(Iffy Quotient) 트위터트레일(TwitterTrails)

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 그 사실 여부를 아는 것이 중요하거나 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 수밖에 없어 팩트체크 과정 자체의 편향성 논란을 불러온 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, Babakar, & Zubiaga, 2021).

주장 탐지 유형의 대표적인 기술 사례는 클레임 헌터(Claim Hunter) 이다(Beltran, Miguez, & Larraz, 2021). 2020년 스페인의 팩트체크 기관인 뉴트랄(Newtral)은 최고기술책임자인 루벤 미게즈(Ruben Miguez) 주도 하에 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(현. 엑스) 계정을 선별한 후, 클레임 헌터가 트위터 API를 활용해 선별된 계정에서 트윗을 수집한다. 수집된 트윗은 구글의 대규모 언어 모델인 BERT 기반 딥러닝 모델을 통해 자동으로 분류되는데, 사전에 전문 팩트체커들이 팩트체크 가치 정도를 판별해 가공(annotate)한 3만여 개의 트윗 데이터로 모델을 미세조정해 자동으로 긍정(팩트체크 가치가 있음) 또는 부정(팩트체크 할 만한 가치가 없음) 영역으로 분류하는 것이다. 긍정 영역으로 분류된 트윗은 인간 팩트체커의 최종 검토를 거쳐 기존의 데이터베이스에 포함되고, 이렇게 지속적으로 축적되는 데이터를 기반으로 팩트체크 가치 판별 모

델을 반복 학습시키며 보다 정확하게 새로운 주장의 검증 가치를 예측할 수 있게 된다.

유사하게 미국 텍사스 대학교 알링턴 캠퍼스 네이물 하싼(Naeemul Hassan) 교수의 연구팀은 2017년 주장 탐지에 특화된 클레임버스터(ClaimBuster)[2]라는 인공지능 도구를 오픈 소스로 공개했다. 클레임버스터는 소셜미디어 게시글, 정치인 연설로부터 수집된 정보 중 팩트체크할 만한 가치가 있는 주장들을 탐지하는데, 클레임 헌터와 유사하게 인간 팩트체커에 의해 검증 가치가 있는 주장들을 사전에 레이블링한 데이터 세트로 지도학습(supervised learning)한 모델을 활용한다. 모델이 팩트체크 가치가 있는 주장들이 가진 언어적 패턴과 구조를 학습하고, 이를 바탕으로 새로운 데이터에서 주장을 탐지하는 것이다. 탐지 결과 팩트체크가 필요한 트위터 메시지들은 리트윗되어 이용자들에게 알려진다.

마지막으로 영국 비영리 팩트체크 기관인 풀팩트에서 개발하여 유료로 제공하는 풀팩트 AI(Full Fact AI)[3]가 있다. 풀팩트는 2019년부터 아프리카의 아프리카체크(Africa Check) 및 아르헨티나의 체키아도(Chequeado)와 같은 팩트체크 기관이나 영국의 비영리기구인 오픈 데이터연구소(Open Data Institute)와 국제적 협력 관계를 유지하며 기술 개발에 힘쓰면서 인공지능 기반 팩트체킹 기술 발전의 선두에 서 있다. 풀팩트의 목표는 수많은 정보(주장)들 중 팩트체크해야 할 사안의 우선순위를 정하고, 이것이 이미 검증된 것이 아닌지 확인하고, 가능한 한 실시간으로 팩트체크가 가능하도록 전 과정을 자동화하는 데 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 '확인 가능한(checkable)' 부분으로 정의되며, 통계적 수치, 인과관계, 미래 예측 등이 주로 포함된다. 풀팩트AI 또한 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 팩트체크 된 문장을 걸러낸다.

이와 같이 주장을 탐지하는 팩트체크 자동화 기술은 두 가지 측면에서 살펴 볼 필요가 있다. 한편으로는 팩트체크가 인간과 기술이 협업하여 달성하는 과정이 될 수 있는 가능성을 제시한다는 측면이다. 전통적으로 팩트체크는 검증대상의 선정에서부터 검증과정, 검증결과 공개, 검증결과 수정에 이르는 절차를 전문 기자들이 수행하여 왔다. 이에 대한 대안으로 인공지능 기술과 분업을 통한 팩트체크의 방식이 대두된 것인데 이는 팩트체크를 수행할 인원이 제한적인 데

2) https://idir.uta.edu/claimbuster/

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

반해 검증해야 할 정보의 양은 점차 많아지는 상황에서 1차적으로 기계가 주목해야 할 정보를 걸러낸다는 측면에서 효율적인 분업 모델로 제시되었다.

반면 팩트체크가 사회적 사실을 구성하는 데 결정적인 역할을 하는 저널리즘 도구라는 측면에서 생각해보면 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 '무엇을' 팩트체크할 것인가는 선택의 문제라 정치적일 수밖에 없다는 비판을 받아왔는데 이 부분을 기계적으로 처리하여 팩트체크의 '객관성'을 높이려는 목적이다. 여기에는 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 기대와 맞닿아 있다. 하지만 이러한 관점은 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 확인 가능한' 주장들에 '사실'이 될 수 있는 기회를 부여할 가능성이 높다. 즉, 사실 확인이 필요한 주장이라는 것이 결국은 개별 사실의 '검증'이 가능한 주장들에 한정되는데, 앞서 언급했던 통계적 수치, 인과관계, 예측 등이 여기에 해당한다. 결국 인공지능 팩트체크 기술을 통해 탐지된 주장이 궁극적으로 검증해야 할 사실의 기반이 된다는 점에서 이 절에서 살펴 본 주장 탐지 기술은 마치 전통적인 언론환경에서의 '게이트키퍼'와 같이 사실성의 범위와 근본적인 방향성을 규정하는 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 적용해 왔는데, 기계학습을 통해 정해진 범위의 언어 데이터를 학습하여 모델을 개발하고, 이 모델로 새로운 정보나 기사의 사실성 정도를 예측하고 분류하는 것이다. 여기에는 탐지된 주장이나 진술을 뒷받침하거나 반박할 만한 정보들을 수집하는 단계가 수반되는데, 대체로 사전, 뉴스 기사, 국가 통계 등 권위와 신뢰도를 갖춘 출처에서 증거를 추출하지만 최근에는 팩트체크를 거친 진술을 데이터베이스로 구축한 아카이브에서 증거를 추출하는 경우도 있다. 어떤 주장의 진위를 판단하기 위해 추출한 증거를 주장과 '매칭(matching, 대조)'하여 참인지 거짓인지를 판정하는 것이다. 알고리즘의 작동 방식에 따라 참, 거짓의 이분법적 분류를 사용하거나 세분화된 범주를 바탕으로 계산을 수행하여 사실의 '정도'를 추정하는 경우도 있으나 주장과 증거를

대조하는 알고리즘의 작동 방식은 대체로 비슷한 맥락에 있다.

대부분의 경우 기술 자체의 작동 과정은 유사하다. 다만 '어떤 근거'를 바탕으로 사실 여부 및 정도를 판정하였는지가 관건인데 이는 판정 결과의 신뢰성이 기계적 의사결정 과정 자체보다는 판정 근거의 정확성에 좌우될 수 있는 상황을 드러낸다. 앞서 설명한 풀팩트 AI와 클레임버스터는 주장의 탐지에서 나아가 검증까지 자동화하였다. 각자의 방법으로 검증해야 할 주장을 탐지하고 나서 탐지된 주장의 검증까지 자동화하는 것인데 풀팩트AI의 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. 탐지된 주장에 대한 답변과 관련된 내용을 질문 응답 엔진인 울프람 알파와 구글에서 검색하는 것이다. 검색된 답변과 주장 사이 명확한 불일치가 있으면 판정이 유보되어 사용자에게 제시되고, 사실로 판정된 주장은 주장과 일치하는 문장과 그 주변 문장이 '맥락(context)'으로 묶여 사실을 뒷받침하는 '증거'로 사용자에게 보고된다.

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 스쿼시(Squash)[4]라는 플랫폼을 개발했다. 스쿼시는 정치 토론이나 연설 등 주요한 정치적인 이벤트에서 나오는 발언과 주장에 대해 실시간으로 자동화된 사실 검증을 제공하는 것을 목표로 만들어진 플랫폼이다. 스쿼시는 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 게시된다.

---

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 브리컬리(Dan Brickely)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.

사실검증 영역에서 가장 널리 알려진 모델 중 하나는 2018년 제임스 손(James Thorne, 현 카이스트 교수)을 필두로 영국의 쉐필드 대학교와 아마존의 영국 캠브리지 소재 연구소의 연구팀이 개발한 피버(FEVER, Fact Extraction and VERification)이다(Thorne et al., 2018). '사실 추출과 검증'이라는 이름이 가리키듯, 피버 모델은 위키피디아와 뉴스 웹사이트에서 수집한 자료 및 인간의 레이블링을 거친 데이터를 복합적으로 활용해 구축한 데이터셋으로 모델을 훈련시키고, 모델이 입력된 주장에 대한 증거를 추출하여 참/거짓을 판별하는 과정을 자동화한다. 여기서 핵심은, 피버 모델의 훈련을 위한 데이터셋이 사실 검증의 준거가 되는 '증거들의 집합'을 구성하는 것을 목표로 구축되었다는 것이다. 이를 위해 훈련된 인간 팩트체커들이 185,445 건의 샘플 문장(주장)을 다듬고, 각각에 대한 사실 여부를 판단한 후, 이를 뒷받침하는 증거까지 추출 및 교차 검증하여 데이터셋에 포함하는 레이블링 과정이 진행되었다. 실제 팩트체크가 수행될 때는 팩트체크하려는 문장을 입력하면, AI 모델이 지식 베이스가 되는 데이터셋에서 입력된 문장과의 관련성이 높은 문서와 문장을 선택 추출하며, 이것이 입력된 문장과 얼마나 일치하는지에 따라 지지(supported)와 반박(refuted) 여부를 판정한다. 만약 구축된 지식베이스에서 입력된 문장을 지지하거나 반박할 수 있는 내용을 찾지 못하는 경우 '충분한 정보 없음(not enough info)'이라는 판정을 내리게 된다. 피버 모델 개발을 위해 구축된 피버 데이터셋은 팩트체킹 자동화 연구 커뮤니티에서 가장 대표적으로 언급되는 벤치마크 데이터셋 중 하나이며, 이를 주제로 올해로 일곱 번째 국제 워크숍이 개최되는 등(http://fever.ai/) 상당한 후속 연구가 피버 모델을 채택해 이를 기준으로 새로운 모델의 팩트체크 수행 능력을 가늠하고 있다. 서울대 이준환 교수 연구팀이 개발한 국내 최초의 팩트체크 AI 서비스 또한 피버 모델을 벤치마킹 하였다. 이밖에 학술논문과 연구 보고서 등 학술 출판물을 토대로 싸이팩트(SciFact)와 코비드팩트(Covid-fact)는 코로나19 팬데믹 국면에서 과학적 사실 검증을 위해 구축되었으며(Saakyan, Chakrabarty, & Muresan, 2021; Wadden et al., 2020), 클라이메트피버(Climate-fever)는 기후와 환경 분야에 특화된 데이터셋으로(Diggelmann, Boyd-Graber, Bulian, Ciaramita, & Leippold, 2020) 이들이 사실 검증 데이터셋의 대표적인 예시이다. 폴리티팩트(PolitiFact)와 같은 기존의 팩트체크 웹사이트에서 사실 검증이 완료된 데이터들을 수집한 라이어(LIAR) 데이터셋 또한 대표 사례 중 하나이다(Wang, 2017).

주장의 탐지와 달리 추출된 증거를 기반으로 주장의 사실 여부를 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 구성하는 데 핵심적인 역할을 한다. 현재 인

공지능 팩트체크 기술이 판정 근거로 활용하고 있는 데이터셋은 위키피디아와 같은 웹 기반의 협업형 백과사전, 분야별 전문 사전, 뉴스 기사, 통계를 바탕으로 초기 모델을 훈련하며 이후에는 서술하였듯이 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 기술들은 판정의 근거를 클레임 리뷰와 같은 마크업 시스템이나 별도의 대시보드를 통해 공개하고 있다.

위와 같은 일련의 노력은 팩트체크 자동화 도구 및 기술의 투명성(transparency)에 대한 평가를 제고하는 요소가 된다는 점에서 더욱 중요시되는 측면이 있다. 팩트체크 기술 영역에서 투명성은 주로 분석기술의 공개와 판정의 설명가능성이라는 두 가지 맥락에 따라 논의된다. 전자는 모델이 어떻게 구성되었고 어떤 방식으로 작동하며 어떤 데이터를 활용했는지에 대한 투명한 공개, 후자는 모델이 판정한 '사실'에 대한 설명가능성(explainability)을 강조하는 것이다(Das, Liu, Kovatchev, & Lease, 2023). 여기서 설명가능성의 경우 특정 주장에 대한 참/거짓의 판정이 무엇을 참고로 했는지에 대해 근거 자료를 공개할 뿐 아니라 시스템이 그 판정에 대한 근거를 인간이 쉽게 이해할 수 있는 형태로 도출해야 할 것이 요구된다(Hartley, 2024; Kotonya & Toni, 2020). 앞서 서술한 IFCN의 팩트체크 준칙이 검증과정의 반복가능성을 합의에 다다르기 위한 사실성의 요건으로 고려했다면, 기술에 의해 자동화된 사실검증은 (공개된) 근거에 기반한 일관된 판단의 가능성을 사회적으로 합의할 수 있는 '사실'의 요건으로 강조하고 있는 셈이다.

지금까지 살펴본 사실검증을 위한 벤치마크 데이터셋 구축과 증거 추출을 위한 노력은 모두 팩트체크 절차의 투명성과 판정의 설명가능성을 높이기 위한 시도로써 그 의의를 평가할 수 있다. 그러나 현재 구축된 데이터셋과 기술 수준에 비추어 우리가 자동화 팩트체크의 판정을 온전한 사실로 받아들일 수 있는지에 대 해서는 아직 비판적인 평가가 잇따르고 있다(Rubin, 2022). 우선 모델을 학습시킬 만큼 대용량의 데이터이면서 '사실'을 검증할 수 있을 만큼 정확성이 있는 데이터가 충분하지 않다는 한계를 지적할 수 있다. 무엇보다도 대용량의 데이터를 구축하기 위해서는 사용자가 많거나 인구가 많은 언어로 적혀진 데이터일수록 유리한데 이는 사실상 영어를 제외한 언어나 영어로 번역될 수 없는 사실에 대한 데이터셋 확보가 한계에 부딪힐 수밖에 없음을 시사한다. 또한 기술의 작동 방식이 근거 자료와 주장을 대조하여 이루어진다는 점에

서 근거 자료는 사전에 사실성이 검증된 데이터이어야 하는데 그 범위가 매우 협소하다. 앞서 소개한 기술 사례들이 활용한 사전, 위키피디아, 통계자료, 뉴스기사 등이 대표적이나 백과사전의 경우는 대체로 저작권법에 의해 보호되는 저작물로 검증에 수반되는 전 자료의 공개가 권장되는 팩트체크 영역에서 적극적으로 활용되기에는 적합하지 않은 경우가 많다. 통계자료나 뉴스기사도 대안이 될 수 있지만 '사실'을 검증하는 데 있어 시간, 장소, 맥락에 따라 사회적 사실의 구성과 해석이 달라질 수 있는 뉴스기사나 시간의 흐름에 따라 지속적으로 정보가 변화하는 통계정보를 사용하는 것 역시 기술적인 한계가 따른다. 피버 모델이 사용한 위키피디아는 이러한 이유로 가장 현실적인 대안으로 주목받았는데 위키피디아 역시 자료가 실시간으로 수정 및 추가되며 협업형 모델의 특성 상 문장이 정제되어 있지 못하다는 점도 한계로 지적된다. 실제로 국내 유일의 인공지능 팩트체크 사례인 서울대 연구팀의 경우도 위와 같은 이유로 정제된 대용량의 한국어 데이터셋을 충분히 확보하는데 어려움을 겪은 바 있다.

이에 대한 대안이 복수의 전문 팩트체커들이 사전에 팩트체크를 수행한 데이터를 훈련 데이터셋으로 삼아 모델을 개발하는 방안이다. 팩트체커 1인에 의한 사실 판정이 아니라 다수의 팩트체커들 간의 합의가 팩트체크 과정 내에 수반될 수 있다는 점에서 '사실' 구성을 위한 인간과 기술 간의 합의 과정으로 해석할 수 있다. 하지만 모델 개발을 위한 데이터셋을 확보하는 과정이 사실을 확인하기 위해 풍부한 자료를 탐색하는 과정이라기보다는 인공지능 알고리즘에 최적화된 데이터로 그 범위를 1차적으로 좁혀 가기 위한 과정에 가깝다는 점에서 검증 '가능한' 사실의 의미적 구성은 결국 기술의 작동 과정에서 협소해지고 마는 한계가 있다. 또한 이 모든 과정이 결국은 알고리즘의 '판정'에 의거한다는 점에서 사실성에 다다르기 위한 '합의'를 도출하는 상호주관적 실재의 구성에 비중을 두는 팩트체크 저널리즘이 지향하는 사실성과 이질적인 맥락에 놓일 수 있다.

3) 정보 확산의 형태 감지

마지막 유형은 정보가 유통되는 네트워크의 구조와 정보 전파의 특징을 살펴 사실이 아닐 가능성이 있는 정보들을 파악하는 방법이다. 이들은 앞선 사례처럼 증거가 되는 정보 텍스트 위주로 '내용(content)'자체만을 판별하는 과정을 자동화하는 데 주력하지 않는다. 그보다는 특정 정보가 나타나고 퍼져 나가는 네트워크의 유형에 중점을 두고 관련 특징을 추출하여 사실성 여부를 판정하는 근거를 제공하기 위한 기술이다. 그래프 구조에 기반한 딥러닝 알고리즘의 일종인 그래픽 뉴럴 네트워크(Graphic Neural Network, GNN)에 기반한 이 접근법은 내용의 특성을 고려하면서도 정보가 공유되고 전파되는 사회적 맥락과 확산 패턴, 이용자 간 상호작용 등 추가적인 단서를 감지하여 사실에 반하는(counterfactual) 정보를 가려내는 데 일조한다(Phan,

Nguyen, & Hwang, 2023; Ünver, 2023). 이러한 기술 사례들은 대체로 트위터 등 초기 소셜미디어에 한정되어 이루어졌는데 허위 정보가 퍼져 나가는 흐름과 의심 계정을 탐지하여 정보의 흐름을 이용자에게 알리고 통제하는 기술사례들이 해당한다.

브라질의 대표적인 팩트 체크 기관인 아오스 파토스(Aos Fatos)에서는 2020년 왓츠앱, 유튜브, 트위터, 인스타그램, 페이스북 등 소셜미디어 상에서 퍼지는 허위정보의 흐름을 실시간으로 자동화 도구 레이더(Radar)[6]를 출시했다. 이 도구는 브라질의 정치 여론 뿐 아니라 코로나19 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 프로필, 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, 조회 수를 높이기 위해 공격적이거나 도발적인 단어를 선택하거나 자극적이거나 과장된 표현을 사용하는 경우, 과도한 대문자의 사용, 문법 오류가 있는 경우에 저품질 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 보토미터(Botometer)[7]가 있는데, 이는 특정 트위터 계정의 운영 주체가 인간이 아닌 봇일 가능성을 평가하는 도구이다. 보토미터는 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 혹시(Hoaxy)[8] 서비스가 출시되었다. 혹시는 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)[9]가 2018년 개발되었다. 비정상성 탐지 알고리즘을 활용해 해시태그, 링크, 계정

6) https://www.aosfatos.org/radar/#!/

7) https://botometer.osome.iu.edu/

8) https://hoaxy.osome.iu.edu/

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)[10] 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 마셜펀드(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)[11] 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)는 이피 코션트(Iffy Quotient)[12]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

9) https://osome.iu.edu/tools/botslayer
10) https://osome.iu.edu/tools/topfibers/
11) https://securingdemocracy.gmfus.org/hamilton-dashboard/
12) https://csmr.umich.edu/projects/iffy-quotient/

2014년 개발한 트위터트레일스(TwitterTrails)[13]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.

이 절에 제시된 기술들은 앞선 기술 유형과는 달리 정보의 '내용'을 판단하지 않는다. 그보다는 정보의 형태(구성 형식) 및 정보 확산 네트워크의 특성을 활용해 오정보, 허위조작정보, 유해정보 등을 감지하는 단서를 찾아내는 데 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 팩트체크 정의에 완전히 부합하는 것은 아니지만 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 기계적 대조에 의존하고 있기에 팩트체크의 합의 도출 과정과는 다른 방식으로 진행되는 측면이 있다. 이 때 정보 확산의 흐름을 파악하는 일은 정보의 유통 경로를 가시화한다는 점에서 '투명성을 통한 객관성을 확보'하는 팩트체크 저널리즘과 인식론적으로 더욱 가깝다고 볼 수 있다.

사실 온라인 상에 유통되는 수많은 정보들의 진위를 파악하기 위해 네트워크의 구조와 특성을 '가짜뉴스'로 통칭되고 있는 오염된 정보의 흐름을 파악하는 것이 기술적으로 가장 '적합한' 방법이라는 점에 대해서는 많은 기술자들이 동의하고 있다. 하지만 이는 현실적으로 어려운 방법론인데 앞서 설명한 트위터의 경우도 현재는 트위터가 API 제공을 중단함으로써 해당 서비스들

13) http://twittertrails.com/

은 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 현실적으로는 가장 어려운 유형에 속한다.

종합하면, 인공지능 팩트체킹 기술은 아직 초기 단계에 있으며 개선이 필요함에도 불구하고, 급변하는 디지털 정보환경에서 가중되는 사실성의 공백을 메우기 위한 발전을 지속하고 있다. 그 과정은 사실 여부를 가려낼 가치가 있는 주장의 탐지부터 근거 데이터셋에 기반해 내용의 사실성을 검증하고 정보가 유통되는 보다 넓은 미디어 환경을 통합적으로 고려하여 사실성을 검증할 수 있는 다양한 단서를 제공하는 방향까지 나아가고 있다. 그럼에도 인공지능 팩트체킹 기술의 도입에 앞서 보다 신중한 접근이 요청되는 이유는 인공지능 팩트체킹이 구성하고 있는 '사실성'에 내재된 한계 때문이다.

인공지능 팩트체킹은 언론의 사실성 제도를 통해 정보를 검증해가며 사실을 모으고, 쌓고, 맞추어 나가는 과정을 제도화하며 그 과정을 통해 상호주관적 실재를 구성해 주는 팩트체크 저널리즘과 달리 사실이 아닌 것을 기계에 의해 솎아내고(filtering), 사전에 합의 혹은 규정된 '사실'들의 집합에 검증의 대상이 되는 주장을 견주어 보는(matching) 방식으로 사실성을 구축한다. 즉, 팩트체크 저널리즘은 외부세계에 존재하는 객관적인 '사실'을 전제하지 않고 르포, 현장인용, 정보원 검증, 게이트키핑 등 언론의 전통적인 사실성 제도와 정보원, 현장, 언론인, 데스크, 팩트체커, 이용자 등 복합적 이해관계자들 각각의 상호주관적 검증을 거치는 과정 안에서 사회적 사실을 구성했다면, 인공지능 팩트체크는 외부세계에 존재하는 '객관적 사실' 및 그 요건을 사전적, 기술적으로 규정하고 해당 요건에 맞지 않거나 맞는 것을 걸러내고 견주어 보는 과정을 기계에 의해 자동화하면서 궁극적으로 사실 자체를 기술이 구성할 수 있는 '사실의 요건' 안에서 재구성한다. 이 과정 속에서 사실의 범위와 영역은 직관적이어지는 대신 상대적으로 경직되고, 정체되며, 협소해질 수밖에 없는 위험이 있다.

5. 결론

저널리즘에서 '사실(fact)'이라는 것은 객관적인 실재에 바탕하여 존재하기 보다는 사회적 사실을 구성하는 과정에서 다양한 사실성 관행을 통해 수립되어 왔다. 때문에 팩트체크 역시 총체적

으로 구성된 사실에 대해 복수의 검증자들이 합의하는 과정을 통해 사실을 (재)구성하고 (재)확인하는 과정을 거쳐 이루어져 왔다. 하지만 인공지능 팩트체크에 대한 사회적 접근법은 저널리즘이 구성하는 '사실'에 대한 명확한 숙고를 바탕으로 기술의 가용성을 탐색하는 것이 아니라 '가짜뉴스'라 불리는 것들을 보다 효율적이고 일괄적으로 퇴치하기 위한 방법론으로서 국가적으로 설계되고 장려되는 과정에서 '사실' 자체를 기존에 저널리즘이 추구해 온 것과 다른 방식으로 구성하고 있는 측면이 있다. 즉, 온라인 상에서 오정보, 허위조작정보, 유해정보가 생산되는 방식과 유통되는 과정에 대한 다각적인 이해나 저널리즘 영역 안에서 '사실'이 구성되는 과정과 조건에 대한 숙의 없이 '가짜뉴스'라는 기호를 다루는 방식의 일환으로 사실이 아닌 것을 기계적으로 제거해가는 과정을 자동화하고 있다.

앞선 절에서는 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 팩트체킹 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. 인공지능 팩트체크는 뉴스에서 주장을 탐지하고, 주장과 근거를 맞대어 견주어 보고, 정보 확산의 패턴을 감지함으로써 사실이 아닐 가능성이 높은 것을 제거하면서 '사실'의 범주를 제한하고 줄여 나가는 방식을 취한다. 즉, 인공지능 팩트체크의 사실성은 사실이 아닌 것을 걸러내고 남은 것을 합의하는 구조로 만들어진다. 이는 정보를 모으고, 쌓고, 맞추어 나감으로써 사실을 사회적으로 구성하는 팩트체크 저널리즘과 대조적이다.

'가짜뉴스'라는 기호에 대한 정책적 대응책으로 부상한 팩트체크 저널리즘이 인공지능을 활용하는 방식으로 확장, 혹은 부분적으로 전환되고 있다는 점에서 인공지능 팩트체크가 구성하는 사실성은 궁극적으로 걸러내어야 할 '가짜뉴스'가 무엇인지 기술사회적으로 정의하고 이에 어떻게 대응해야 하는지에 대한 새로운 담론의 방향을 제시한다. 주장의 탐지, 근거 추출과 주장의 검증, 정보 확산의 형태 감지로 유형화한 인공지능 팩트체크 기술이 '가짜뉴스'를 탐지하는 데 활용될 경우 그 기술적 양식에 따라 '가짜뉴스' 자체가 다르게 정의될 여지를 가지며 '가짜뉴스'에 대응하는 양상도 달라질 수 있기 때문이다.

우선 주장을 탐지하고 근거 데이터셋을 기반으로 주장을 검증하는 기술의 경우 인공지능 기술이 무수한 정보에서 검증할 만한 '주장'을 탐지하고 모델이 학습한 훈련데이터와 비교하면서 사실을 검증하여 개별 사실의 참, 거짓 여부나 정도를 밝힌다. 앞서 살펴본 것처럼, 인공지능 팩

트체크 기술이 우선순위를 부여하는 사실 확인이 필요한 주장은 '기술적으로 확인 가능한' 통계적 수치, 인과관계, 예측 등에 편중될 가능성이 높다. '가짜뉴스'의 생성 과정 자체에서 기술에 의한 선택 편향의 문제가 야기될 수 있는 것이다. 아울러, 현재 서비스가 진행 중인 대부분의 인공지능 팩트체크 기술은 근거 데이터를 구축하여 훈련한 모델이 사실의 여부나 정도를 판정하는 형태로 제공되는데, 이 경우 '가짜뉴스'는 내용을 검증할 근거 데이터가 없거나 근거 데이터와 맞지 않는 정보로 정의할 수 있으며 이를 제거함으로써 사실의 범주를 좁혀 '가짜뉴스'에 대응한다. 이 같은 접근법은 기본적으로 '가짜뉴스'로 통칭되는 인터넷 상의 오염된 정보의 '내용'을 확인 가능한 좁은 범위의 '사실' 내에서 대조해 사실 여부를 판정하는 형태를 취하고 있다. 이는 개별 요소의 사실성이 아니라 주장들이 맥락화하여 만들어내는 '총체적인 사실'을 '합의'의 과정을 통해 검증해간다는 팩트체크 저널리즘의 지향점과 극명하게 대비된다. 앞서 살펴보았듯이 언론은 과학적 명제나 객관적 실재를 탐구하는 것이 아니라 '실재의 사회적 구성'에 핵심적인 역할을 담당해 온 까닭에 '참' 혹은 '사실'이라는 것의 판정이나 주장은 제도적 합의의 과정을 통해 보장되어 왔다. 르포, 직접인용, 통계 등을 이용하여 사실성을 입증하거나 정보원의 실명을 공개하고 삼각확인이나 게이트키핑 등의 사실 검증 절차를 거치는 과정이 그러한 맥락에서 강조되어 왔다. 이 같은 제도들은 언론이 이미 존재하는 '사실'을 발견하는 것이 아니라 복합적인 검증 절차를 통해 사실을 구성하고 달성해왔음 보여준다. 이 과정이 본질적으로 정치적일 수밖에 없는 상황에서 사실의 판단을 인공지능 기술에 전가하는 상황은 무엇이 사실 판단의 근거 데이터가 되는지, 누가 그 데이터의 범위를 결정하느냐에 따라 여전히 정치적일 수 있음에도 불구하고 '객관적'으로 여겨진다. 검증 가능한 사실의 범위 역시 기술의 가용성에 따라 협소해지며 합의의 과정 없이 정보의 진위 여부를 판정해야 할 주체의 의도와 목적에 따라 특정한 사실(거짓)이 기계적으로 먼저 제거될 가능성을 낳는 것이다.

정보 확산 패턴을 감지함으로써 사실성의 '형태'를 추론하는 경우는 정보의 확산이 특정한 형태로 이루어지는 구조와 과정을 살펴봄으로써 '가짜뉴스'를 구성하는 행위자들의 연결망에 구조적으로 접근하는 경우이며, 이 때 가짜뉴스는 개별 정보의 진위여부보다는 정보의 생산 주체나 정보가 생산되고 유통되는 미디어 기술 환경에 대한 이해에 기반 한다고 볼 수 있다. 이 경우 '가짜뉴스'에 대한 대응책은 정보 확산이 연결고리를 끊거나, 혹은 보다 근본적으로 디지털 정보의 생산, 유통, 소비의 방식을 개선하기 위한 접근법을 강구할 수 있다. 즉, 가짜뉴스의 내용에 집중하여 정보를 제거하기 보다는 가짜뉴스의 유통에 가담하는 디지털 행위자들을 정렬하는 물질적 기능을 인공지능 팩트체크가 수행할 수 있다. 이는 정보 확산의 패턴을 감지하여 사실이 아닌 것을 제거하는 인공지능 팩트체크 기술이 더 나아가 이용자들이 (오)정보의 생산, 유통, 소비에 윤리적

으로 개입하여 사회적 실재로서의 사실을 더할 수 있는 여지를 열기도 한다. 즉, 인간과 기계의 이분법적 구분이 아니라 현재의 미디어 환경에서 '사실'을 구성하는 인간의 노동이 새로운 방식으로 협상되고 구성될 수 있는 여지를 열어 '가짜뉴스'라는 기호이자 현상에 대응할 수 있는 것이다.

저널리즘 영역 내부의 자성적인 움직임 외에도 정부, 뉴스 포털 기업, 기술자 등 다양한 이해관계자들이 복잡하게 연루되는 형태로 구성되고 있는 국내 팩트체크 시스템이 정착하는 과정은 '가짜뉴스'를 사실상 무엇이든 될 수 있는 기의 없는 기호로 만들었다. 특히, 인공지능 기술로 팩트체크를 자동화하려는 움직임은 협소하고 제한된 '실사'의 영역 내에서 사실이 아닐 가능성이 있거나 복합적인 이유로 사실이 아니게 만들어야 하는 주장을 거르고 제거하는 방식으로 팩트체크를 재구성했다. 결국 팩트체크의 복잡한 과정 중에 인공지능에 의해 자동화되는 것이 무엇인가에 따라 '가짜뉴스'가 지칭하는 것은 매우 느슨하게 정의되며 저널리즘에서 강조해 온 사실성 역시 정치적으로 (재)정의될 수 있는 여지가 있기에 팩트체크 전 과정에서 인공지능이 실행, 배치되는 조건과 방식을 보다 심층적으로 살펴 볼 필요가 있다. 살펴보았듯 아직은 진행 중이나 인공지능 기술이 팩트체크의 영역에 개입할 수 있는 여지는 다양하다. 저널리즘과 팩트체크는 개인이 사회적 실재를 구성하는 중요한 타자라는 점에서 인공지능 기술이 팩트체크 과정에 배치되는 조건이나 방식은 우리가 언론을 통해 사회를 이해하는 데 지대한 영향을 미친다. 인공지능 기술과 팩트체크의 결합은 단순히 '어떤 정보를 제거할 것인가'에서 설계되는 것이 아니라 인공지능 기술이 팩트체크의 맥락에 도입되는 과정에서 사실성이 어떻게 기술사회적으로 구성되는지를 구체적으로 바라보는 과정을 거쳐 시작될 필요가 있다.

팩트체크 환경은 생성형 인공지능의 등장으로 인해 또 다른 도전과 기회를 동시에 직면하고 있다. 생성형 인공지능의 확산은 저널리즘 뿐 아니라 콘텐츠 제작 전반의 방식을 근본적으로 변화시키고 있다. 딥페이크 등을 이용하여 만들어지거나 조작된 이미지의 품질은 사람이 만든 콘텐츠와 구별하기 어려울 정도로 점점 더 정교해지고 있으며, 챗지피티 등을 이용한 콘텐츠는 정보의 진위 여부와 관계없이 그럴듯한 어투로 사실처럼 느껴지는 정보를 무제한으로 생산해내고 있다. 지금까지 우리는 인공지능 기술을 통한 '자동화'의 측면에만 주목해 왔지만 인공지능은 이미 생성의 영역에 다다라 새로운 위기를 만들어 내고 있다.

때문에 최근의 국내 가짜뉴스 대응책은 생성형 인공지능을 향하고 있다. 특히 국내 뉴스 포털 기업들은 2024년 4월 제22대 국회의원 총선거를 앞두고 생성형 인공지능을 활용해 작성된 기사의 본문 상단과 하단에 관련 내용을 공지하며 인공지능과 로봇이 자동으로 작성한 기사에 대해 고지했다. 이는 생성형 인공지능을 활용하여 만들어내는 허위정보에 대한 국제적인 대응과 유사한 맥락을 공유하고 있는데, 일례로 오픈에이아이(OpenAI)사는 자사 이미지 생성 도구인 달

리(DALL-E)가 만든 이미지를 식별하기 위한 머신러닝 도구를 개발하고 있으며, 마이크로소프트(Microsoft)사에서도 동일한 역할을 하는 비디오 판별자(Video Authenticator)라는 도구를 공개했다. 구글(Google)역시 콘텐츠의 진위 여부를 식별하고 생성형 인공지능에 의한 인위적인 조작 여부를 가려내는 지그소우(Jigsaw)라는 별도의 조직 단위를 만들고 어셈블러(Assembler)나 스타일갠 탐지기(StyleGAN Detector)와 같은 도구를 발표했다. 뿐만 아니라, 네덜란드(Sensity AI), 영국(Logically AI), 캐나다(Originality AI) 등 세계 각국에서 적지 않은 기업가와 연구자들이 일반적으로 사람의 눈에는 보이지 않는 복잡한 패턴과 이상 징후를 분석하여 콘텐츠의 조작 여부를 판별하는 도구를 개발하기 위한 노력을 경주하고 있다. 이러한 현상이 시사하는 것은 팩트체크 분야 전체에 걸친 거대한 도전이다. 사실과 사실성의 구성에 있어 인공지능의 역할은 앞으로 더욱 확대될 것이며, 이러한 변화 한가운데서 '사실'과 '사실성'의 주체적인 구성 및 해석을 위한 인간 사회와 기술의 역할 및 양자 간의 균형에 대한 고찰이 그 어느 때보다 중요해질 것이다.

이 논문은 팩트체크 저널리즘과 대조하여 인공지능 기술 사례들이 팩트체크 과정 중 어느 단계를 자동화하며 무엇을 근거로 사실을 입증하는지를 살펴봄으로써 저널리즘이 역사적으로 구성해 온 사실성의 지형에 나타날 수 있는 변화를 비판적으로 검토해 보았다. 이는 상호주관적 사실을 더해 가는 합의의 방식으로 구성되어 온 사회적 실재가 사실이 아닌 것을 기계적으로 제거하고 남겨진 것들에 대한 동의를 이끌어내는 방식으로 사실이 기술사회적으로 구성되는 것으로 요약할 수 있다. 인공지능 기술이 어떤 방식으로 사회 안에 위치하게 될지는 언제나 기술과 사회적 담론 사이의 공모에 의해 정해진다. 특히 국내의 팩트체킹 지형은 팩트체크 전문기관, 언론사 내부 뿐 아니라 '가짜뉴스'를 정의하는 정부, 팩트체크 과정에 개입하는 기술의 특성, 상품으로서 뉴스를 관리하는 포털 기업 등 복잡한 이해관계자들에 의해 구성되고 있다. 이 논문은 저널리즘의 영역에서 전통적으로 팩트체크가 제도화되고 사실성이 정의되어 온 과정 위에서 인공지능 기술이 팩트체크 영역에 실행되고 배치되는 조건과 방식을 살펴봄으로써 기술사회적으로 진화하는 '사실성'의 한 층위를 이론적으로 탐색하고자 했다. 이는 앞으로 팩트체크나 '가짜뉴스'라는 기호를 다루는 방식, 인공지능 기술의 발전이 어떻게 그리고 어떤 방식으로 서로 얽혀나갈지 등 향후 지속적으로 추적하며 발전시켜나가야 할 주제이다. 또한 인공지능 팩트체크의 사실성을 이론적으로 검토한 이 논문을 바탕으로 '가짜뉴스'라는 담론을 구성하는 행위자를 보다 명확하게 분류하여 그들의 상호작용이 가짜뉴스라는 담론을 어떻게 기술사회적으로 구성하는지 살펴보는 작업 또한 이어져야 한다. 인공지능 팩트체크를 기술적으로 정교화시키려는 노력만큼이나 인공지능 기술이 저널리즘 생태계나 가치에 미치는 영향에 대한 탐구가 보다 다각적으로 이루어져야 할 것이다.

References

Amazeen, M. A. (2013). *A critical assessment of fact-checking in 2012*. New America Foundation.

Barnhurst, K. G. (2016). The problem of modern locations in US news. *International Journal of Media & Cultural Politics*, *12*(2), 151-169.

Beltrán, J., Míguez, R., & Larraz, I. (2021, April). *ClaimHunter: An unattended tool for automated claim detection on Twitter*. KnOD'21 Workshop.

Berger, P. L., & Luckmann, T. (1966). *The social construction of reality*. Open Road Media. 하홍규 (역) (2013). <실재의 사회적 구성>. 서울: 문학과지성사.

Choi, S., & Ko, M. (2016). ICR media trend report. *Seoul National University Media and Communication Resaerch Institute Newsletter, 3*, 1-19. [최순욱·고문정 (2016). 미디어 트렌드 리포트. <서울대학교 언론정보연구소 뉴스레터>, 3호, 1-19.]

Chong, E. (2018). The characteristics of Korea's fact check journalism: With a focus on fact check journalists' perception of "facts" and investigation of their fact verification processes. *Journal of Communication Research, 55*(4), 5-53. [정은령 (2018). 한국 팩트체크 저널리즘의 특징: 팩트체크 언론인들의 사실 인식과 사실 검증과정 탐색을 중심으로. <언론정보연구>, 55권 4호, 5-53.]

Chong, E. (2019). Fact check news and the recovery of credibility of South Korea's broadcast journalism: Focusing on the broadcast reporters' perceptions on formatting and news values of fact check news. *Studies of Broadcasting Culture, 31*(1), 47-101. [정은령 (2019). 팩트체크 뉴스와 한국 방송 저널리즘의 신뢰 회복: 방송 기자들의 팩트체크 뉴스 양식과 뉴스가치에 대한 인식을 중심으로. <방송문화연구>, 31권 1호, 47-101.]

Das, A., Liu, H., Kovatchev, V., & Lease, M. (2023). The state of human-centered NLP technology for fact-checking. *Information Processing & Management, 60*(2), 103219.

Diggelmann, T., Boyd-Graber, J., Bulian, J., Ciaramita, M., & Leippold, M. (2020). Climate-fever: A dataset for verification of real-world climate claims. *arXiv preprint arXiv:2012.00614*.

EDMO. (n.d.) *Repository: Fact-checking initiatives in the EU (and in the UK)*. Retrieved 8/16/24 from https://edmo.eu/resources/repositories/fact-checking-organisations-in-the-eu/

Foucault, M. (2004). *Naissance de la biopolitique: Cours au Collège deFrance, 1978-1979*. 오트르망 (역) (2012). <생명관리정치의 탄생: 콜레주드프랑스 강의 1978~79년>. 서울: 난장.

Graves, D. (2018). *Understanding the promise and limits of automated fact-checking*. 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.

Hartley, R. (2024). Efficacy analysis of online artificial intelligence fact-checking tools. *The International Review of Information Ethics, 33*(1).

Hwang, C. (2017). Legal measures to address fake news. *Media Arbitration*, *142*, 26-37. [황창근 (2017). 가짜뉴스에 대처하는 법적 방안. <언론중재>, 142호, 26-37.]

Hwang, Y. (2017). *'Fake news' when form and content are deliberately misleading: Issues in defining the concept of fake news*. Seoul, Korea: Korea Press Foundation. [황용석 (2017). <형식과 내용 의도적으로 속일 때 '가짜 뉴스': 가짜 뉴스 개념 정의의 문제>. 서울: 한국언론진흥재단.]

IFCN. (2016). International fact-checking network fact-checkers' code of principles. Retrieved 8/16/24 from https://www.poynter.org/ifcn-fact-checkers-code-of-principles/

Jung, C. (2019, September). Watergate investigative journalist calls for the term 'fake news' to be abolished. *Media Today*. [정철운 (2019, 9, 26). 워터게이트 특종기자, 가짜뉴스라는 말 폐기해야. <미디어 오늘>] Retrieved 8/16/24 from http://www.mediatoday.co.kr/news/articleView.html?idxno=202653

Jung, J. (2019). The dilemma of quotation journalism ② The banality of practices: The ordinary nature of evil. *Broadcasting Journalists, 47*, 12-14. [정준희 (2019). 따옴표 저널리즘의 딜레마2] 관행이란 이름의 범속함, 그 악의 평범성: 게으른 받아쓰기를 넘어 복화술 저널리즘으로. <방송기자>, 47권, 12-14.]

Jung, S. (2015). Empathy, compassion, and affect: The external extension of journalism analysis and criticism. *Communication Theories, 11*(4), 38-76. [정수영 (2015). 공감과 연민, 그리고 정동(affect): 저널리즘 분석과 비평의 외연 확장을 위한 시론. <커뮤니케이션 이론>, 11권 4호, 38-76.]

Jung, S. (2021). A study to increase the influence of fact check news: Focusing on the effect of fact check news on the audience. *Korean Journal of Broadcasting and Telecommunication Studies, 35*(1), 235-282. [정성욱 (2021). 팩트체크 뉴스의 영향력 확대를 위한 연구: 팩트체크 뉴스가 수용자에게 미치는 효과를 중심으로. <한국방송학보>, 35권 1호, 235-282.]

Kim, C. (2020). Ritual, defensive, intentional: A study on gatekeeping practices centered on fact-checking of major Korean newspaper editors. *Korean Journal of Journalism and Communication Studies, 64*(5), 5-45. [김창숙 (2020). 의례적, 방어적, 의도적: 한국 주요 신문 에디터의 사실 확인을 중심으로 한 게이트키핑 관행 연구. <한국언론학보>, 64권 5호, 5-45.]

Kim, M. (2018). From “fake news” to “disinformation” - Comparison of terminology and conceptual definitions in Korean legislative proposals to curb fake news and in the initiatives to curb fake news abroad, *Journal of Media and Defamation Law, 5*(2), 43-81. [김민정 (2018). 가짜뉴스에서 허위조작정보로: 가짜뉴스 규제 관련 국내 법안과 해외 대응책에 나타난 용어 및 개념정의 비교. <미디어와 인격권>, 5권 2호, 43-81]

Kim, S., & Kim, W. (2019). *Digital News Report 2019*. Seoul: Korea Press Foundation. [김선호·김위근 (2019). <디지털 뉴스 리포트 2019>. 서울: 한국언론진흥재단.]

Konstantinovskiy, L., Price, O., Babakar, M., & Zubiaga, A. (2021). Toward automated factchecking: Developing an annotation schema and benchmark for consistent automated claim detection. *Digital Threats: Research and Practice, 2*(2), 1-16.

Korean Communication Commissions. (2020, March 11). *Results of the 13th committee meeting in 2020*. Korean Communication Commissions press release. [방송통신위원회 (2020, 3, 11). <2020년 제 13차 위원회 결과>. 방송통신위원회 보도자료.]

Kotonya, N., & Toni, F. (2020). Explainable automated fact-checking: A survey. *arXiv preprint arXiv:2011.03870.*

Lee, H. (2020). The changing media ecosystem and media production studies. *Korean Journal of Communication & Information*, *101*, 81-112. [이희은 (2020). 미디어 생태계에서의 변화와 생산자 연구. < 한국언론정보학보>, 101호, 81-112.]

Lee, N. (2018). A content analysis of fact-checking of Korean news organizations in the 19th presidential election in South Korea: Based on principles of the international fact-checking network. *Journal of Communication Research, 55*(4), 99-138. [이나연 (2018). 한국 언론의 팩트체크: 19대 대통령선거에서의 후보자 검증 기사를 중심으로. <언론정보연구>, 55권 4호, 99-138.]

Marwick, A., & Lewis, R. (2017, May). *Media manipulation and disinformation online*. Data & Society. Retrieved 8/16/24 from https://datasociety.net/library/media-manipulation-and-disinfo-online/

McIntyre, L. (2018). *Post-Truth*. Cambridge, MA: MIT Press.

Ministry of Culture, Sports and Tourism. (2023, April 20). *‘Malicious information epidemic’: Strengthening measures to combat fake news*. Ministry of Culture, Sports and Tourism press release. [문화체육관광부 (2023, 4, 20). <‘악성정보전염병’ 가짜뉴스 퇴치 전면 강화>. 문화체육관광부 보도자료.]

Mueller, R. S. (2019, March). *Report on the investigation into Russian interference in the 2016 presidential election* (Mueller report). Retrieved 8/16/24 from https://www.govinfo.gov/content/pkg/GPO-

SCREPORT-MUELLER/pdf/GPO-SCREPORT-MUELLER.pdf

Oh, S. (2017). Current states and limitations of automated fact checking technology. *Journal of Cybercommunication Academic Society, 34*(3), 137-180. [오세욱 (2017). 자동화된 사실 확인(fact checking) 기술(technology)의 현황과 한계. <사이버커뮤니케이션학보> 34권 3호, 137-180.]

Oh, S., & Hwang, G. (2018). An exploratory study on the formal requirements of 'fact': Suggestion through analysis of SNU FactCheck information metadata. *Journal of Communication Research, 55*(4), 54-98. [오세욱·황구현 (2018). '팩트'의 형식적 구성 요건에 대한 탐색적 연구. <언론정보연구>, 55권 4호, 54-98.]

Oh, S., & Park, A. (2017). *Public perception of 'fake news'*. Seoul: Korea Press Foundation. [오세욱·박아란 (2017). <일반 국민들의 '가짜 뉴스'에 대한 인식>. 서울: 한국언론진흥재단.]

Park, D. (2015). A study of double validity claims in quotations: News source network analysis of news on the four major rivers project in the Dong-A Ilbo and the Hankyoreh. *Korean Journal of Journalism and Communication Studies, 59*(5), 121-151. [박대민 (2015). 사실기사의 직접인용에 대한 이중의 타당성 문제의 검토: 동아일보와 한겨레 신문의 4대강 추진 논란 기사에 대한 뉴스 정보원 연결망 및 인용문 분석. <한국언론학보>, 59권 5호, 121-151.]

Park, D. (2023). Journalism artificial intelligence based on trustworthy artificial intelligence: Toward a commensurability between media trust and trustworthiness of artificial intelligence system. *Media & Society, 31*(4), 5-47. [박대민 (2023). 신뢰할 수 있는 인공지능 기반의 저널리즘 인공지능: 언론 신뢰와 인공지능 신뢰성 간 통약가능성을 바탕으로. <언론과 사회>, 31권 4호, 5-47.]

Park, J. (2020). Actors and meaning network of 'fake news': Journalism, intormation technology, and cultural politics. *Journal of Cybercommunication Academic Society, 37*(4), 149-195. [박진우 (2020). '가짜뉴스'라는 기호를 다루는 사람들: 저널리즘, 정보기술, 그리고 대중들의 문화정치. <사이버커뮤니케이션학보>, 37권 4호, 149-195.]

Park, J., & Lee, W. (2008). Reporters and editors' attitudes toward the inverted pyramid and narrative writing style. *Korean Journal of Journalism & Communication Studies, 52*(6), 123-145. [박재영·이완수 (2008). 역피라미드 구조와 내러티브 스타일에 대한 기자와 에디터의 인식. <한국언론학보>, 52권 6호, 123-145.]

Park, S., & Lee, J. (2023). Artificial intelligence fact-checking technology and the dynamics of factuality: An in-depth interview analysis of field participants. *Korean Journal of Journalism & Communication Studies, 67*(4), 238-271. [박소영·이정현 (2023). 팩트체킹 인공지능 기술과 사실성의 역학: 현장 참

여자의 심층 인터뷰 분석. <한국언론학보>, 67권 4호, 238-271.]

Peters, M. A. (2017). Education in a post-truth world. *Educational Philosophyand Theory, 49*(6), 563-566.

Phan, H. T., Nguyen, N. T., & Hwang, D. (2023). Fake news detection: A survey of graph neural network methods. *Applied Soft Computing, 139*, 110235.

Posetti, J., & Matthews, A. (2018). *A short guide to the history of 'fake news' and disinformation.* International Center for Journalists. Retrieved 7/2/22 from https://www.icfj.org/news/short-guide-history-fake-news-and-disinformation-new-icfj-learning-module

Rubin, V. L. (2022). Artificially intelligent solutions: Detection, debunking, and fact-checking. In V. L. Rubin (Ed.), *Misinformation and disinformation: Detecting fakes with the eye and AI* (pp. 207-263). Cham, Switzerland: Springer International Publishing.

Saakyan, A., Chakrabarty, T., & Muresan, S. (2021). COVID-fact: Fact extraction and verification of real-world claims on COVID-19 pandemic. *arXiv preprint arXiv:2106.03794.*

Schudson, M. (2001). The objectivity norm in American journalism. *Journalism*, *2*(2), 149-170.

Shin, H., & Lee, Y. (2021). Journalists' awareness of misinformation issues: Focused on in-depth interviews. *Korean Journal of Journalism & Communication Studies, 65*(4), 239-272. [신혜선·이영주 (2021). 오보 문제에 대한 기자 인식: 심층 인터뷰를 중심으로. <한국언론학보>, 65권 4호, 239-272.]

Silverstone, R. (1999). What's new about new media? Introduction. *New Media & Society*, *1*(1), 10-12.

Thorne, J., & Vlachos, A. (2018). Automated fact checking: Task formulations, methods and future directions. *arXiv preprint arXiv:1806.07687.*

Throne, J., Vlachos, A., Christodoulopoulos, C., & Mittal, A. (2018). Fever: A large-scale dataset for fact extraction and verification. *arXiv preprint arXiv:1803.05355.*

United Nations. (2013). Information integrity on digital platforms. *Our common agenda policy brief 8*.

Ünver, A. (2023). *Emerging technologies and automated fact-checking: Tools, techniques and algorithms*. edam.

Uscinski, J. E., & Butler, R. W. (2013). The epistemology of fact checking. *Critical Review, 25*(2), 162-180.

van Dijk, T. A. (1988). *News as discourse*. Hillsdale, NJ: Lawrence Erlbaum.

Wadden, D., Lin, S., Lo, K., Wang, L. L., van Zuylen, M., Cohan, A., & Hajishirzi, H. (2020). Fact or fiction: Verifying scientific claims. *arXiv preprint arXiv:2004.14974.*

Wang, W. Y. (2017). "Liar, liar pants on fire": A new benchmark dataset for fake news detection. *arXiv preprint arXiv:1705.00648.*

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.

Yang, J. (2019). “News” and “fake news” as perceived by ordinary citizens. *Media Issue, 5*(1). Seoul: Korea Press Foundation. [양정애 (2019). 일반 시민들이 생각하는 ‘뉴스’와 ‘가짜뉴스’. <미디어이슈> 5권 1호. 서울: 한국언론진흥재단.]

Yoon, T. (2011). Affective participations and re-construction of reality: A study on Korean reality television shows. *Studies of Broadcasting Culture, 23*(2), 7-36. [윤태진 (2011). 정서적 참여와 실재(reality)의 재구성. <방송문화연구>, 23권 2호, 7-36.]

최초 투고일 2024년 02월 02일
게재 확정일 2024년 07월 29일
논문 수정일 2024년 08월 02일