Check for updates

한국언론학보
2023년 67권 4호 238~271쪽
https://doi.org/10.20879/kjjcs.2023.67.4.007
http://www.comm.or.kr

팩트체킹 인공지능 기술과 사실성의 역학
현장 참여자의 심층 인터뷰 분석

박소영 조선대학교 신문방송학과 조교수
이정현 중앙대학교 인문콘텐츠연구소 HK연구교수

Artificial Intelligence Fact-Checking Technology and the Dynamics of Factuality*
An In-depth Interview Analysis of Field Participants

Soyoung Park (ORCID: https://orcid.org/0000-0002-9533-9191)**
(Assistant Professor, Chosun University)
Jeonghyun Lee (ORCID: https://orcid.org/0000-0003-4295-6783)***
(HK Research Professor, Humanities Research Institute, Chung-Ang University)

'Fact-check' has drawn attention as a countermeasure against false or misleading information that is produced in large quantities and rapidly distributed in the digital environment. In response, the number of fact-checking organizations as well as news outlets that try to emphasize fact-checking practices in the news production process has also increased in Korea. Since the late 2010s, attempts to automate fact-checking based on artificial intelligence (AI) technologies have developed with government-backed financial support. However, up to date, the social discourse on the necessity, limitations, prospects, and direction of AI-based fact-checking technologies has not been sufficiently developed. Facing this issue, this study aims to present the current status, challenges, and prospects of national AI fact-checking technology research and development through in-depth interviews with seven stakeholders who have been involved in two representative fact-check institutes. This study also conducted a phenomenological analysis of government policy documents and published research. The results of the study suggest that Korean AI fact-checking technology is currently stuck in an unfriendly research environment, including a lack of national research support and a lack of high-quality Korean

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

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

data, which has turned into a politicized debate. Participants also emphasized that social efforts are needed to find consensus among various stakeholders on the role and scope of AI in the fact-checking process. They also suggested that the factcheck technology development process should be accompanied by a mature social discourse given the subjectivity and politics inherent in fact-checking, a domain of social interpretation with embedded values and contextual information. Based on these findings, the study emphasizes the need to explore alternative governance systems to ensure and strengthen the independence and impartiality of fact-checking, and calls for self-reflection by all members of society to establish the fact-checking process as a virtuous cycle. This means that we all need to establish the correct perception of what constitutes a "fact," implement appropriate practices in the actual fact-checking process, and most importantly, respect and accept the fact-checking results presented without being bound by partisanship to further solidify the socio-cognitive foundation for a healthy social debate. All in all, our findings contribute to restructuring the sociotechnical discourse within the topography of various actors in the realm of the fact-checking ecosystem, thereby suggesting the future direction of AI fact-checking technologies in Korea. The significance of this study is that it documents the historical development process of trial and error of national AI-based fact-checking technology and raises the issue of overly politicized or blindly focused development of AI fact-checking research and regulation in Korea without mature social deliberation.

Keywords: Fact-Check, Artificial Intelligence, Automation, Journalism, Fake News

국문초록

디지털 환경 속에서 잘못되거나 왜곡된 정보들이 대량으로 생산되고 빠르게 유통, 소비되면서 '팩트체킹(fact-checking, 사실 확인)'이 대응책으로 주목 받고 있다. 국내에서도 팩트체크 전문기관이 등장했고 팩트체크 관행을 뉴스 생산 현장에서 강조하고자 하는 언론사 수도 증가해왔으며, 2010년대 후반부터는 정부의 지원에 힘입어 인공지능을 기반으로 팩트체킹을 자동화하려는 시도가 발전해 왔다. 하지만 정작 국내 인공지능 팩트체크 기술 개발의 필요성, 한계, 전망, 방향성 등에 대한 현장의 목소리는 사회적 담론으로 충분히 생산되지 못했다. 이 연구는 이에 문제 의식을 갖고 팩트체크 현장에의 참여 경험이 있는 이해관계자를 인터뷰함으로써 국내 인공지능 기반 자동화 팩트체크 기술의 현황, 과제 및 전망을 제시하고자 했다. 국내 팩트체크 전문기관에서 활동하고 있는 7인의 주요 이해관계자에 대해 심층 인터뷰를 진행했고, 한국형 인공지능 기반 자동화 팩트체킹 기술 개발과 관련된 정부기관의 기술개발 지원 현황과 학계 발표 연구 등을 현상 분석했다. 연구 결과, 현재 한국형 인공지능 팩트체크 기술은 국가적 차원의 연구지원이 소모적 정쟁 속에서 정치적으로 변질되고, 양질의 한글 데이터를 확보하기 어려운 비우호적인 연구 환경 속에서 연구가 답보 상태에 머물러 있음이 드러났다. 현장 참여자들은 팩트체킹 과정에서 인공지능의 역할과 범위에 대해 다양한 이해관계자 간 합의점을 찾기 위한 사회적 노력이 필요하다고 강조했다. 특히, 가치와 맥락 정보가 내재된 사회적 해석의 영역인 팩트체킹에 내재된 본질적인 주관성과 정치성을 감

안해 기술 개발 과정에 사회적 합의가 수반되어야 함을 강조하고 있다. 본 연구의 결과는 팩트체크 생태계 내에서 실제 활동 중인 참여자들의 시각을 사회적 담론으로 재구성하는 데 일조하고, 다양한 행위자의 지형도 안에서 한국형 인공지능 팩트체크 기술이 나아갈 방향성을 제시하고 있다. 본 연구의 의의는 국내 인공지능 기반 팩트체크 기술 개발 과정과 시행착오를 기술사의 일부로 기록하여 남기고, 사회적 숙의 과정을 거친 인공지능 개발 방향을 제안하는 데 있다.

핵심어 : 팩트체크, 인공지능, 자동화, 저널리즘, 가짜뉴스

1. 서론

디지털 환경에서 생산되는 지식과 정보의 양이 방대해짐에 따라 그 안에서 어떻게 사실과 진실을 분별하느냐가 중요한 시대적 과제가 되고 있다. 소셜 미디어의 부상과 함께 가속화된 '오정보(misinformation)' 및 '허위조작 정보(disinformation),'—이른바 '가짜뉴스(fake news)'와 프로파간다의 범람은 지난 2016년 당시 미국 대선 국면을 뒤흔들며 전세계 여러 국가에 경종을 울렸다. 비슷한 시기에 발생한 브렉시트(Brexit)나 미얀마 군부에 의한 로힝야인의 집단 학살 사태 등에서도 가짜뉴스는 부인하기 어려운 영향력을 행사해 왔다(York, 2022). 특히 코로나19 팬데믹으로 전세계가 진통을 앓던 과정에서 코로나 백신 접종이 에이즈나 치매를 유발한다는 등 코로나19 관련 허위조작정보는 확산속도와 파급력 때문에 정보(information)와 전염병(epidemic)을 합성한 인포데믹(infodemic), 즉 '정보 전염병'으로 일컬어지기도 했다(Zarocostas, 2020).

이처럼 잘못되거나 왜곡된 정보들이 디지털 미디어를 통해 빠르게 퍼져 나가는 현상이 만연한 작금의 상황은 객관적이고 균형 잡힌 정보를 통한 시민의 알 권리 충족을 기본적인 책무로 삼는 저널리즘 영역에 위기의식을 가중시키고 있다. 뿐만 아니라 미디어 융합 현상이 가속화되면서 신문과 방송의 뉴스로 대표되던 기존 저널리즘의 생산과 유통, 소비 방식에도 급격한 변화가 나타나고 있으며, 이는 오정보나 허위정보의 범람에도 영향을 미치는 구조적 환경으로 작용하고 있다. 주어진 정보에서 '참'을 가려내고자 하는 '팩트체킹(fact-checking, 사실 확인)'에 대한 국내외의 관심은 이처럼 증가하는 '가짜 뉴스' 사례에 대응해 사회적 경각심이 높아진 현실을 반영하는 것이다.

흔히 2016년 미국 대통령 선거 이후 페이스북(Facebook, 현 Meta)이나 트위터(Twitter) 등 소셜미디어에서 양산되는 오정보에 대한 대응책으로서 팩트체크 알고리즘을 홍보하면서 팩트체크라는 개념이 대중적으로 널리 알려졌다고 여겨지지만, 사실 팩트체크는 1990년대 미국의 선거에서 행해지던 언론의 흑색선전에 대한 사실 검증이 필요해짐에 따라 역사적으로 대두했고 이후 꾸준히 관련 조직과 체계를 구성하고 정비하는 과정에서 발전해왔다(최순욱·고문정, 2016). 미국의 팩트체크닷오알지(Factcheck.org), 폴리티팩트(PolitiFact), 더 팩트 체커(The Face Checker)뿐 아니라 국제 팩트체크 네트워크(International Fact-Checking Network: IFCN) 등이 조직되었으며, 유럽연합 디스인포랩(DisinfoLab)에서는 유럽 팩트체킹 기준 프로젝트(European Fact-checking Standards Network: EFCSN)를 통해 현재 정보 환경에서 준수되어야 할 팩트체킹에 대한 가이드라인을 수립하는 등 세계 각지에서 팩트체

크 전문 기관을 설립해 운영 중이며 이를 뉴스 생산 현장에 도입해 실천하고자 하는 팩트체킹 제휴 언론사의 수도 점차 늘어나고 있는 상황이다(Ye, 2023).

2010년대에는 디지털 미디어를 중심으로 미디어 이용자가 생산 및 축적하는 정보의 양 뿐 아니라 그 전파 속도 또한 기하급수적으로 증가하면서 보다 빠르고 효과적인 팩트체킹을 위한 혁신의 필요성이 대두되었다. 이에 디지털 기술과 인공지능을 기반으로 한 팩트체킹의 대량화, 실시간화, 자동화 시도들이 이루어지게 되는데, 미국에서는 텍사스대, 듀크대, 스탠포드대 연구진과 구글 리서치 팀이 공동으로 개발 중인 '클레임 버스터(ClaimBuster)', 구글의 '팩트 체크 도구(Fact Check Tools),' 인디애나 주립대 연구팀의 '지식 그래프(Knowledge Graph)'를 이용한 자동 팩트체킹 등이 대표적인 사례다(Posetti & Mattews, 2018). 국내에서도 팩트체크를 보조하는 인공지능 기술이 2017년 한국 과학기술 기획 평가원이 꼽은 10대 미래 유망기술 중 하나로 떠올랐으며, 한국언론진흥재단이 빅카인즈(bigkinds.or.kr)를 운영하고 서울대 언론정보연구소에서 인공지능 팩트체크 개발 프로젝트가 진행되는 등 팩트체크 자동화를 위한 기반 또한 점진적으로 조성되어 오고 있다.

가짜뉴스 문제에 대한 해법을 찾기 위한 이 같은 국내외의 노력에도 불구하고 가짜뉴스가 생산 및 배포되는 속도와 여기에 대응하는 팩트체크 기술 사이에는 여전히 큰 간극이 존재한다. 생성형 인공지능에까지 이른 현재의 디지털 기술과 인공지능은 누구나, 쉽게, 이전보다 훨씬 그럴듯한 가짜뉴스를 생성할 수 있게 한다. 디지털 인프라도 점점 확장되고 세밀해져 가짜뉴스를 유통하고 배포하는 과정에 정책적으로나 기술적으로 개입하고 통제하기도 버거워지고 있다. 하지만 팩트체킹의 대량화, 실시간화, 자동화를 위한 디지털 기술이나 인공지능 개발은 그 속도를 따라가지 못하는 실정이다. 국내 인공지능 팩트체크 기술 개발 역시 지금까지 답보 상태에 머물러 있으며, 세계 각지의 여러 기관과 연구자들에 비교하면 이에 대한 연구나 사회적 관심도 현저히 부족하다. 이 연구는 인공지능 팩트체크 기술이 확산되는 가짜뉴스에 대한 대응책으로서 대두되었음에도 불구하고 다른 인공지능 기술에 비해 가시적인 발전을 이루지 못하고 있다는 점에 문제의식을 갖고, 그 이유와 당면한 과제를 국내 팩트체킹 현장 참여자들의 목소리를 통해 살펴보고자 한다.

2. 문헌 검토

1) 팩트체킹 논의의 등장과 발전

팩트체킹의 개념과 중요성이 강조된 것은 사실 최근만의 일은 아니다. 뉴스의 재료가 되는 '사실'에 대한 확인은 뉴스의 기본 요소로 저널리즘의 발전 초기부터 이른바 금과옥조처럼 여겨지는 원칙으로 자리해 왔으며, 1890년대 말 황색저널리즘과 타블로이드지 등의 난립을 반면교사로 더욱 강조되었다(Posetti & Mattews, 2018). 저널리즘에서의 사실확인은 매우 고전적인 관행이나, '팩트체크'에 대한 논의가 재점화한 데는 정보 환경의 무질서와 인포데믹 현상에 대한 경각심이 전세계적으로 높아지게 된 최근의 상황과 밀접한 관련이 있다. 2016년 미국 대선 당시 러시아 정보기관이 도널드 트럼프 전 대통령에 우호적인 방향으로 인터넷 여론을 조작하여 선거에 영향을 미치려 한 시도가 미국 법무부 특별검사팀의 수사 보고서(일명 '뮬러 보고서(Mueller report)')에 의해 공개되며 인터넷 정보환경에 대한 전세계 여러 국가들의 믿음에 경종을 울렸다(Mueller, 2019). 이 사건을 시발점으로 인터넷 상의 정보가 특정한 조직과 이해관계에 의해 얼마나 쉽게 오염되고 왜곡되어 반대 진영을 공격하는 '무기'로 활용될 수 있는지(weaponization of information) 보여주는 사례가 곳곳에서 관찰되며 저널리즘과 민주주의에 대한 전례 없는 위협으로 떠오르게 되었다(Posetti & Mattews, 2018).

팩트체킹에 대한 논의는 이와 같은 정보 위기에서 유력한 대안으로 재부상했다. 온라인 공론장의 현재와 미래에 대한 위기의식을 공유하는 전세계 여러 뉴스기관, 저널리즘 연구 및 교육기관, 비영리 시민단체들이 팩트체킹의 중요성에 주목하기 시작하며 조직적인 행동과 실천을 촉구하기 시작했다. 대표적으로, 미국 저널리즘 연구기관인 '포인터(Poynter)'가 주창한 전 세계 팩트체크 기관 연합회인 국제 팩트체킹 네트워크(IFCN)를 발판으로 전 세계 백 여 개 이상의 기관들이 연합하여 팩트체킹의 성공적인 확산과 수행을 위한 원칙과 실천 강령을 수립 및 배포하고 있다. IFCN는 세계 각지의 다수 팩트체킹 기관들과의 논의를 통해 비당파성(상반되는 입장 어디에도 치우지지 않는 불편부당성)을 비롯해 공정성, 팩트체킹의 근거와 과정에 대한 투명한 공개 및 설명 원칙, 팩트체킹을 수행 및 후원하는 기관과 재정 운영에 대한 투명성, 사실성 판단에서의 오류 수정에 대한 개방성과 정직성 등을 주된 원칙으로 내세우고 있다(IFCN, 2016). 또한 비영리 독립단체인 '퍼스트 드래프트(First Draft)'는 2015년 9개 참여 기관의 주도로 출범한 이후 웹 상의 정보에 대한 팩트체킹을 위해 정확한 '출처'의 확인이라는 원칙으로 압축되는 실천적, 윤리적 가이드라인을 제공하고 있다. 이는 동영상과 사진 등 정보가 언제, 어디서, 어떤 맥락에서 게시되고 공유된 것인지, 그리고 정보의 공식적인 출처가 되는 저자와 기관을 신뢰할 수

있는지 판단하는 등의 과정을 포함한다(Dotto & Smith, 2019).

아울러, 이 같은 팩트체크 논의의 재점화 이후, 전문 저널리즘 영역에서도 확인해야 하는 '팩트'의 초점이 과거와 달라지고 있다(정은령, 2022). 과거 저널리즘에서는 취재원의 발언을 '팩트'의 핵심으로 보고 이를 자의적으로 바꾸거나 훼손하지 않고 그대로 옮기는 것을 강조했다면, 최근 팩트체킹의 핵심은 화자의 발언 내용에 관심을 두고 이것이 참인지 거짓인지 그 진실성 여부를 가리는 것에 치중한다는 것이다(정은령, 2018, 2019, 2022). 이러한 전환은 전세계적인 현상으로, 일례로 IFCN이나 듀크(Duke) 대학교 부속 리포터스 랩(Reporter's Lab)에서는 공적 인물이나 주요 단체로부터의 발언, 기타 공론화된 사회적 주장들이 정확한지 확인하고 이를 비당파적으로 보도하는 것을 팩트체킹의 주목적으로 규정하고 있다(IFCN, 2016). 팩트체킹의 대상이 되는 콘텐츠의 범위 또한 점차 확장되어 최근에는 '매니페스토(manifesto) 운동'으로 널리 알려져 온 정치인들의 정책공약 실천 여부에 대한 점검까지 팩트체킹의 대상으로 편입되고 있다(이나연, 2018).

팩트체킹은 어느덧 전세계적인 움직임으로 확장되는 추세를 보이고 있다. 북미권을 넘어, 유럽 대학 연구소(European University Institute: EUI) 산하의 유럽 디지털 미디어 옵저버토리(European Digital Media Observatory: EDMO)는 영국을 포함한 유럽 연합 국가에서 약 백여 개 기관이 팩트체킹 기관으로 활동하고 있음을 보여준다(EDMO, n.d.). 세계 유수의 언론기관들 또한 팩트체킹 부서를 신설하거나 웹사이트를 개설하는 등 사실확인을 통한 건강한 온라인 공론장을 형성하는 데 일익을 담당하고 있다. 국제적인 명성을 갖춘 워싱턴 포스트지의 팩트체커(Washington Post Fact Checker), 미국 공영 라디오 방송국(National Public Radio)의 팩트체크(NPR FactCheck), 로이터 팩트체크(Reuters Fact Check) 등이 대표적인 사례이다. 국내에서도 이러한 움직임에 적극적으로 동참하고 있다. 국내의 대표적인 팩트체킹 기관 및 실천의 사례는 서울대학교 언론정보연구소에서 설립한 SNU팩트체크센터를 주축으로 한 팩트체크(SNU팩트체크)와 한국 기자협회, 방송기자연합회, 한국PD연합회의 대표적인 3개 언론 현업단체 중심으로 출범했던 〈팩트체크넷〉, 그리고 각 언론사에서 자체적으로 수행하는 팩트체크를 들 수 있다. KBS, MBC, SBS 지상파 3사와 연합뉴스, JTBC 등 각 언론사에서 팩트체킹 담당 부서를 설치하여 뉴스 생산 과정에서 팩트체킹을 강조하고 있다. 이나연(2018)은 국내 주요 12개 언론사의 팩트체킹 기사를 분석하여 상당수의 기사에서 취재원 사용 등 기사 작성 과정에서의 투명성 원칙과 사안별 찬반 입장의 포함 등 불편부당성의 원칙을 준수하기 위한 노력이 나타나고 있음을 확인하였다.

2) '팩트' 체킹의 정치화

팩트체킹의 원칙과 대상을 규정함에 있어, 무엇을 '팩트'로 간주할 것이며 어떻게 '팩트'를 찾아야 하는지에 대한 논의는 가짜뉴스에 대한 개념적 층위에서 비롯한 '사실' 개념의 차이, '사실' 자체가 갖는 인식론적 불명확성, 사실확인에 개입하는 주체들 간의 역학관계와 같은 문제와 얽혀 팩트체크 자체를 정치화시켜 왔다.

먼저, '가짜뉴스'라는 용어 자체가 갖는 복잡다단한 층위가 존재하는데 이는 각 층위에 대응하는 '사실'이 무엇인지 판단하는 것을 더욱 복잡하게 한다. 앨리스 마윅과 레베카 르위스(Alice Marwick & Rebecca Lewis, 2017)는 가짜뉴스를 '온라인과 미디어 상에서 광범위하게 유포되는 오염된 정보'를 널리 아우르는 것으로 정의하며, 해당 용어가 도널드 트럼프 전 미국 대통령이 본인에게 비판적인 논조의 언론보도를 공격하기 위한 도구로 사용하는 과정에서 빠르게 쟁점화되고 정치화되었다고 진단하고 있다. 실제로 가짜 뉴스에 대한 대중적인 인식은 2016년 후반 미국 대선을 기점으로 급격히 확산되었는데, 클레어 와들과 호세인 데라크샨(Claire Wardle & Hossein Derakhshan, 2018)은 여기에 경각심을 표하며 '가짜뉴스'라는 용어 사용 자체를 문제시하기도 했다. 이는 '가짜뉴스'라는 용어가 광범위한 정보 오염 현상의 일부에 지나지 않는데다, 특히 특정 정치 세력이 스스로에게 부정적인 보도를 폄하하기 위해 사용하는 정쟁의 도구로 전락하기 쉽기 때문이다. 이에 대한 대안으로 와들과 데라크샨은 보다 엄밀하고 세분화된 개념체계를 '가짜뉴스'에 적용할 것을 제안했다. 즉, 정보의 허위(false) 여부와 해를 끼칠 의도(intent to harm)에 따라 오정보 또는 단순허위정보(misinformation), 허위조작정보(disinformation), 유해정보(mal-information)를 구분하자는 것이다.

여기서 오정보는 잘못된 정보이나 그 오류의 원인이 실수나 무지, 또는 부주의에 기인한 것들을 의미하고, 허위조작정보는 공중을 속이거나 기만하려는 명확한 의도를 가지고 사실을 왜곡하거나 날조한 정보들을 일컫는 것에 가깝다. 마지막으로 유해정보는 특정 집단이나 개인을 비방하고 모욕하는 것을 주된 목적으로 하는 정보들로, 차별이나 혐오표현이 여기에 해당한다. 앞서 언급한 개념들은 정보의 사실성을 강조하는 팩트체크와 유사한 맥락에서 논해지곤 한다. 모두 건강한 공론장과 언론 기능을 회복하고자 한다는 점에서 팩트체크와 인식론적으로 매우 근접하게 논의되는 것이다. 그러나, 본 연구의 화두인 팩트체크에서의 사실성은 의견과 사실의 분리, 취재원의 투명한 공개, 다양한 이견의 수용 등 보다 복합적인 검증 절차를 수반해 달성되는 것으로 정의된다. 즉, 특정 정보가 허위가 아니라는 이유만으로 그 정보가 '사실'로 판정되는 것은 아니라고 할 수 있다.

이처럼, 현재의 정보 오염 현상은 상당히 복잡다단한 개념적 층위에서 논해질 수 있기에 이

에 따라 각 현상에 대한 정확한 인식과 이해를 촉구하는 학계의 부단한 노력이 있어 왔다. 그러나 이러한 노력이 무색하게, 아직까지 정치권 그리고 일반 대중 사이에서는 '가짜뉴스'라는 용어가 남용되고 있는 실정이다. 트럼프 전 미국 대통령 사례가 시사하듯, 불리한 의제에 '가짜뉴스'라는 딱지를 붙여 대응하는 것은 쉽고 간편하면서도 상당히 효과적으로 대중에 소구할 수 있는 수단이다. 국내외에서 '가짜뉴스'가 정치 진영을 불문하고 여론전에서 유리한 위치를 점하기 위한 일종의 정치적 무기로 활용되는 현상이 심화됨에 따라 가짜뉴스 관련 의제는 정쟁의 도구나 정치적 수사의 하나로 전락하고 있다. 무엇보다도, 이로 인한 이슈의 '정치화(politicization)'현상은 사람들이 가짜 뉴스 관련 의제를 직관에 의해 당파적으로 해석하기 쉽게 할 수 있다는 점(Taber & Lodge, 2006)에서 건전한 팩트체크 환경의 성장에 위협이 될 수 있다. 실제로, 현재 가짜 뉴스를 둘러싼 논의는 지나치게 정치화되어 일종의 '이슈 블랙홀'이 되었으며, 그 과정에서 오정보, 허위조작정보, 유해정보, 사실 정보 등 각 세부 · 인접 영역과 관련된 고찰과 문제 해결을 위한 생산적인 논의의 발전을 저해하고 있다(Tong, Gill, Valenzuela, & Rojas, 2020).

둘째로, 어떤 사건에 대한 사실 확인 행위는 그 '사실성'의 해석에 관여하는 인간 의식의 불완전성으로 인해 본질적으로 편향성을 내포할 수밖에 없다는 점에서 정치화의 위험에 쉽게 노출될 수 있다. 과학적 명제와 달리, 사회적 사실에 대한 해석은 결국 시대와 공간, 기타 여러 상황적 맥락과 다양한 이해관계자들이 복잡하게 얽힌 총체에 대한 가치 판단과 다름없기에 누가 어떤 맥락에서 특정 사실을 바라보느냐에 따라 그 해석이 판이하게 달라질 수 있기 때문이다. 우신스키와 버틀러(Uscinski, 2015; Uscinski & Butler, 2013)는 이러한 관점에서 팩트체킹의 대표적인 한계점을 다섯 가지로 정리하고 있다. 그 첫째는 선택 효과(selection effect)이다. 사회적 자원은 한정적이기에 무수한 사회적 사실 모두의 사실성을 판단하는 것은 현실적으로 불가능하며, 효율적이지도 않기에 우선적으로 다뤄져야 할 문제를 '선택'하게 되는데 이 선택 과정 자체가 근본적으로 '정치적'일 수밖에 없다는 것이다. 예컨대 여러 나라에서 난민 등 이민자보다는 자국민의 문제를 우위에 두는 현실을 들 수 있다. 두 번째는 편집의 위험성이다. 다양한 '사실'들을 혼동되게 뒤섞거나 전체 맥락에서 떼어진 일부만 사실이라 주장할 가능성을 배제하기 어렵다. 세 번째는 성급한 인과관계 오류의 발생 가능성이다. 어떤 사회적 현상의 원인과 결과를 밝히기 위해서는 엄밀한 연구와 검증이 필요함에도, 단순한 선후 관계와 우연적인 상관관계 정도로 성급하게 그 인과를 단정해 버리는 위험성이 팩트체킹 과정에서 관행적으로 나타날 수 있음이 문제시되었다. 네 번째는 미래 예측에 대한 팩트체킹이 모순적 행태임을 지적하고 있다. 미래는 애초에 사실 '확인'이 불가능한 영역이기 때문이다. 마지막 다섯 번째는 증거 선택 기준의 모호함을 들고 있다. 특정 사안을 구성하는 수많은 정보 중 '사실성'을 판단하기 위해 선택되거나 걸러지는 정보

들이 있는데, 이 과정에서 어떤 정보가 어떤 이유로 취사 선택되는지 그 이유가 명확하지 않은 경우가 많다는 것이다. 종합하면, '사실성'의 판별은 근본적으로 매우 미묘하고 복잡한 과업이며, 그 검증 과정은 가치판단이 내재된 인간 인식의 주관성이 개입될 수밖에 없다.

마지막으로 근원적인 '사실' 자체의 인식적 불명확성에 덧붙여, 실천적 영역에서도 팩트체킹의 주체와 범위 등을 둘러싸고 있는 다양한 행위자들의 역학이 갈등을 일으킬 수 있다. 우선 팩트체킹 수행의 자격(즉 '누가' 팩트체킹을 해야 하는지)을 두고 문제가 생길 수 있다. 장기적인 훈련을 거쳐 저널리즘 등 '공식적인' 정보 생산과 검증에 종사해 온 기관이나 조직에 의해서 전문적이고 '객관적'인 팩트체킹이 이뤄질 수 있음을 강조하는 입장이 있는가 하면 (e.g., IFCN, SNU팩트체크센터), 독립적 단체나 기구부터 시민 개인까지 사회적인 이슈에 관심을 가진 누구나 팩트체킹에 참여하는 것이 중요하다는 시각도 있다(e.g., Reporter's Lab, 팩트체크넷). 관련해, 팩트체킹 주체의 정치적 중립성도 논란이 될 수 있는 요소이다. 예컨대 2023년 초 해산된 〈팩트체크넷〉의 사례를 들 수 있다. 팩트체크넷은 지난 2020년 말 문재인 정부 시기 방송기자연합회, 한국기자협회, 한국PD연합회, 사회적협동조합 빠띠가 공동으로 설립한 팩트체크 수행 기관으로, 언론인과 시민이 협업하는 시민참여형 팩트체크를 모토로 활동을 지속한 바 있다. 팩트체크넷은 정부의 허위 정보 대응을 위한 중점 추진 정책의 일환으로 2021년부터 방송통신위원회 산하 시청자미디어재단의 지원을 받기 시작했으나, 정치적으로 대립하는 진영의 인사들에 의해 팩트체크의 편파성 시비가 꾸준히 제기되어 왔다. 그 결과 관련사업 지원 예산은 순차적으로 대폭 삭감되었고 결국 올해 2023년 1월 말 해산을 결정하기에 이르렀다. 나아가, 검증해야 하는 '사실'들의 경합성 문제가 있다. 사회의 복잡성과 역동성으로 인해 필연적으로 수많은 정보와 쟁점들이 연속적으로 등장하며 사회적 관심을 놓고 경쟁하는 상황에서 어디에 우선순위를 두고 팩트체킹을 해야 할지에 대한 의견 차이다. 사실확인이 필요한 정보의 공적 가치를 염두에 두고 공공성이 큰 문제에 관심을 기울이는 것이 마땅하다는 시각이 있는가 하면, 온라인에서 유통되는 루머나 '낚시' 등 우리가 일상에서 마주하는 허위 의심정보 또한 검증의 우선순위에 오를 수 있다는 등, 근본적인 방향은 크게 다르지 않으나 미묘한 견해의 차이로 합의에 이르지 못한 지점들이 존재하고 있다.

3) 자동화된 팩트체킹

현재의 미디어 환경을 고려했을 때 우리가 직면한 팩트체킹의 어려움은 기하급수적으로 증가하는 절대적인 정보의 양과 전파 속도에 있다. 사실 확인을 위해 투입될 수 있는 인간의 노동력과 시간은 한정적인 데 반해, 검증이 필요한 정보는 온라인 공간에서 매순간 무수히 쏟아지고 순식간에

광범위하게 확산될 수 있기 때문이다. 클릭 한 번이면 자신과 연결된 수백, 수만의 이용자에게 (허위)정보를 공유할 수 있는 디지털 정보 환경에서 팩트체킹을 용이하게 함과 동시에 허위 정보 확산을 효과적으로 막기 위한 대안적 접근법을 촉구하는 사람들의 목소리는 그 어느 때보다도 높아졌다. 아울러, 작성자나 내용, 게시 날짜 등 정보 자체의 맥락을 보여주는 메타 정보부터 정보와 상호작용하는 다른 사회적 주체나 소셜 네트워크에서의 전파 및 확산 경로까지, 정보를 구성하는 요소들이 점차 복잡해지고 있다는 점도 팩트체크를 어렵게 하는 또다른 지점이다 (오미애·이정란·최호식·진재현·천미경, 2022). 인공지능 기반 자동화 팩트체킹(automated fact-checking)은 이와 같은 복잡한 미디어 환경 속에서 가장 유력한 대안 중 하나로 관심을 모으게 되었다.

자동화된 팩트체킹은 말 그대로 주장의 진실성(truthfulness, 혹은 veracity)을 관련 증거(evidence)를 바탕으로 확인하는 과업의 자동화를 이르는 것이다(Guo, Schlichtkrull, & Vlachos, 2022). 자동화 팩트체킹 관련 연구 및 개발은 컴퓨터 저널리즘과 인공지능학 분야의 접점에 위치하며, 특히 자연어 처리(NLP) 영역을 중심으로 발전해 왔다. 기계학습을 통해 대규모의 언어 데이터를 학습하여 모델을 개발하고, 이를 새로운 자료에 적용해 사실성을 예측하거나 분류 평가하는 것이다. 이 과정에서 그 성능을 산술적으로 측정하거나 다른 모델과 비교 평가할 수 있는 알고리즘의 정확도(accuracy)나 신뢰도를 함께 제시할 수 있는 표준화된 절차가 적용된다(오미애·이정란·최호식·진재현·천미경, 2022). 가장 널리 알려진 팩트체크 모델로 540만 개의 위키피디아 문서를 바탕으로 구축된 FEVER 모델이 있는데(Thorne, Vlachos, Christodoulopoulos, & Mittal, 2018), 그 기본적 작동 원리는 구축된 데이터셋을 기반으로 특정 주장에 관련되는 문서를 찾아내는 '문서 검색'의 과정, 문서 안에서 해당 주장의 증거가 될 수 있는 증거를 선별하는 '증거 선택'의 과정, 주장과 증거를 중심으로 참(true)과 거짓(false)을 판별(verdict)하는 '진실성 확인' 작업을 거치는 것이며, 모델의 성능 및 정확도는 그 지표가 되는 점수(FEVEROUS score)를 통해 산출된다.

공학적 관점에서 팩트체킹은 저널리즘 관점에서 정의하는 팩트체킹보다 좁은 의미를 지니고 있는데, 일례로 '사실확인(fact-checking)'과 사실의 '검증(verification)'의 구분된 용법을 들 수 있다. 여기서 검증은 출처와 날짜, 장소 등 '옳은' 사실들을 확인하는 상대적으로 제한적인 개념이라면 사실확인은 주장의 논리, 일관성과 맥락 등을 고려해 사실성을 판별하는 보다 넓은 범위를 아우르며, 이러한 시각에서 사실 검증은 사실확인 과정의 출발점으로 해석되기도 한다 (Thorne & Vlachos, 2018). 더불어, 저널리즘 영역에서 팩트체킹과 유사한 맥락에서 사용되고 있는 가짜 뉴스나 프로파간다의 탐지(detection) 또한 공학적 측면에서는 엄밀하게는 서로 다른 과업을 의미한다. 가짜뉴스 탐지는 '풍자(satire)'적 요소나 특정한 주장의 의도와 형식, 그

리고 여기에 노출된 수용자 등 저널리즘의 사실성 확인과 직접적인 관련이 없는 요소들 또한 탐지 대상의 범위에 포함하고 있어 팩트체킹과 구분되고 있다(Guo et al., 2022). 이처럼 일반적인 의미의 팩트체킹과 달리 팩트체킹 자동화 기술은 자동화되는 과업이 무엇이냐에 따라서 보다 좁은 의미에서 엄밀하게 정의될 수 있다.

구오 등(Guo et al., 2022)은 최근까지 팩트체킹 분야에서 이루어진 여러 연구들을 분석해 자동화된 팩트체킹에 수반되는 과업을 종합적으로 세 가지 단계로 정리했다. 그 첫 번째 단계는 우선 팩트체킹의 대상이 되는 주장(claim)을 탐지하는 것이다. 이 탐지 과정은 주장의 검증 가치(check-worthiness)에 기반하게 되는데, 한 마디로 특정한 주장이나 진술이 일반 공중이 그 사실 여부를 아는 데 관심을 가질 만할 것인지를 자동적으로 판단한다는 것이다. 가치 판단 과정에서 사회집단의 특성과 시간, 상황적 맥락 등에 따른 주관성이 개입하는 것을 방지하기 위해 개인적 경험이나 주관적 감상에 기반한 문장들을 자동적으로 걸러내는 방법 등이 제안되었다(Konstantinovskiy, Price, Babakr, & Zubiaga, 2021). 다음 단계는 증거 추출(evidence retrieval)이다. 탐지된 주장이나 진술을 뒷받침하거나 반박할 만한 관련 정보들을 수집해 내는 단계이다. 사전이나 뉴스 기사 등 어느 정도 신빙성과 신뢰도를 갖추고 있는 자료들을 바탕으로 증거를 추출하는 방법이 대표적이다. 마지막으로는, 주어진 주장에 대해 추출된 관련 증거들을 취합하여 사실성 여부를 판정(verdict prediction)하는 단계이다. 앞서 언급된 '참 또는 거짓'의 이분법적 분류부터 보다 세분화된 범주들을 바탕으로 주장이 (어느 정도까지) 사실인지를 나타내는 방식이 활용된다. 이 마지막 단계에서는 사람들이 판정 결과를 이해하고 납득할 수 있는 설명 근거를 함께 산출(justification production)하는 것이 점차 중요한 과업으로 강조되고 있다. 아무리 판정 결과가 그럴 듯해도 그러한 판정이 내려진 근거 없이는 그 정확성을 입증할 수 없다는 점에서, 판정에 대한 설명가능성(explainability) 원칙의 중요성이 부각되는 것이다.

지금까지 자연어 처리 기반 팩트체킹 자동화 연구는 지속적으로 발전해 왔다. 예를 들어, 인간 평가자가 매긴 점수나 순위, 기타 분류 처리를 거친 데이터를 학습해 검증 가치가 있는 진술들을 걸러내는 데 도움을 주는 언어 모델인 CLEF CheckThat! lab AI, ClaimBuster, ClaimRank, BERT, RoBERTa가 있고, 이전에 사실 검증을 거친 주장들을 탐지하는 Snopes, PolitiFact, Transformer-XH, Google의 Fact Check Explorer 등의 모델들도 개발되었다(Nakov et al., 2021). 문서나 문장 또는 벡터 수준의 유사성이나 순위를 기준으로 사실성 판단의 근거가 되는 증거를 추출하는 데 특화된 모델은 앞서 언급한 FEVER와 풀팩트의 알파를 꼽을 수 있다(Graves, 2018; Thorne et al., 2018). 훈련 데이터의 부족과 복합적 요소를 효과적으로 결합하기 위한 이론적 체계의 부재 등으로 인해 진척에 어려움을 겪고 있기는

하지만, 최근 딥페이크 등 시각적 허위정보가 확산하며 문자를 넘어 이미지, 소리, 영상, 그리고 연결망과 시간 등 맥락적 정보까지 포함하는 다중복합적(multimodal) 데이터를 활용해 문제를 해결하는 모델을 개발하려는 움직임도 나타나고 있다(Yao, Shah, Sun, Cho, & Huang, 2022).

이처럼, 인공지능 기반의 팩트체킹 자동화 연구는 정보의 양과 속도가 폭발적으로 증가하는 디지털 대전환 시대에 대량의 디지털 텍스트에 포함된 진술의 사실성을 신속하게 판별할 수 있는 방안을 제시하며 기존의 팩트체크 관행에 혁신의 가능성을 제시했다. 국내에서도 학계와 언론계, 시민사회 영역에서 허위의심 정보의 범람에 문제의식을 갖고 이에 대한 대응책의 하나로 인공지능 기반의 팩트체킹 자동화 기술을 제안하고 있으나, 그 기술의 구체적인 목표와 기능에 대해서는 충분히 공론화되지 않고 있다. 이처럼 성숙한 사회적 담론이 부재한 상황에서 최근 몇 년 사이 인공지능 기반 팩트체크 기술의 연구 개발이 급격히 증가하고 있다는 사실은 다양한 역할과 이해관계를 가진 국내 팩트체킹 행위자들의 담론 안에서 기술개발의 필요성과 역할에 대해 검토하는 연구가 필요함을 시사한다. 이에 본 연구에서는 국내 팩트체크 기관 이해관계자 심층 인터뷰 및 팩트체킹 관련 정책과 연구개발 관련 자료 등을 바탕으로 국내의 인공지능 팩트체킹 기술개발을 둘러싼 사회적 담론을 조망하여, 여기서 한국형 인공지능 팩트체킹 기술이 나아갈 방향성을 제시하고자 한다.

3. 연구설계

본 연구는 현재 다른 인공지능 기술에 비해 가시적인 성과를 거두지 못하고 있는 국내 인공지능 팩트체킹 기술을 둘러싼 사회적 담론의 부재에 문제의식을 갖고, 그 이유와 당면한 과제를 국내 팩트체킹 현장 참여자의 목소리 및 팩트체킹 관련 연구와 정책 및 보도자료들을 분석하여 살펴보고자 한다. 보다 구체적으로 이 연구가 현장 참여자 심층 인터뷰를 통해 탐구하고자 한 연구문제는 다음과 같다.

연구문제 1. 국내 팩트체킹 지형도 안에서 인공지능 기술의 도입, 개발, 발전은 어떻게 이루어져 왔는가?

연구문제 2. 인공지능 팩트체킹의 전망과 한계에 대한 내부 담론이 시사하는 기술의 방향성은 무엇인가?

위 연구문제에 대한 해답을 얻기 위해 이 연구는 가장 먼저 국내 팩트체크 전문기관의 주요 이해관계자들과 심층 인터뷰를 진행하고, 한국형 인공지능 기반 자동화 팩트체킹 기술 개발과 관련된 정부기관의 기술개발 지원 현황과 학계에 발표된 국내 연구 등을 2차 자료로 분석하였다. 이 연구는 국내 인공지능 기반 팩트체크 기술 개발 과정과 시행착오를 기술사의 일부로 기록하여 남기고, 그 사회적 맥락을 심층적으로 살피는 데 의의가 있다. 또한 선개발 후수습으로 일관했던 국내 인공지능 개발 사조에 문제를 제기하고 기술 개발의 필요성, 역할, 전망, 한계에 대한 사회적 담론을 구성해 다양한 행위자의 지형도 안에서 한국형 팩트체킹 기술이 나아갈 방향성을 제시하여 향후 연구 개발에 기여하고자 한다.

우선 심층 인터뷰는 국내 팩트체킹 수행 기관의 주요 이해관계자를 대상으로 했다. 심층 인터뷰 참여자를 정하는 데 국내 팩트체킹 수행 기관으로 한정하고 팩트체크를 실천하고 있는 개별 언론사는 인터뷰 대상에서 제외했다. 팩트체크를 '실천'하고 있는 개별 언론사보다는 다수의 언론단체와 유관기관들, 나아가 시민사회를 종합적인 관점에서 아우르며 공동의 '실천적 행동의 거점'으로 작용하고 있는 비영리 팩트체킹 기관에 집중하여 이들이 만들어 낸 국내 팩트체킹의 지형도를 읽어내고자 했다. 또한, 실제로 인공지능 기반 팩트체킹 기술 개발 관련 내용을 팩트체크의 한 방법으로 다루고 있는 기관에 집중하여 국내 팩트체킹 지형도 안에서 인공지능 기술의 도입, 개발, 발전의 과정을 읽고, 인공지능 팩트체킹 기술의 방향성에 대한 내부 담론을 이해하고자 했다.

인터뷰 시점 위 조건에 해당하는 두 개의 팩트체킹 기관을 선정하였는데, 첫째는 2017년 서울대학교 언론정보연구소에서 출범한 〈SNU팩트체크센터〉이고 둘째는 2020년 한국 기자협회, 방송기자연합회, 한국PD연합회의 대표적인 3개 언론 현업단체와 비영리 사회적협동조합 빠띠가 공동 출자하여 설립했던 〈팩트체크넷〉이다. 두 기관은 그 정체성 측면에서 각각 독특한 특징으로 구분된다. 우선 SNU팩트체크센터는 학교와 국내 34개 언론사들(2022.08.25 기준)과의 제휴를 기반으로 하는 협업 모델을 구축해 언론사들이 검증한 공적 관심사를 공중에 널리 알리는 것을 목표로 한다. 한편 앞서 언급한 3개의 대표 언론 현업단체 및 시민단체 중심의 팩트체크넷은 언론계와 각 분야의 전문가 뿐 아니라 시민들이 함께 협업하는 '크라우드소싱' 방식의 팩트체킹에 근간을 두었다. 전자가 전문직주의(professionalism)에 기반한 팩트체킹이라면 후자는 이른바 '풀뿌리(grassroots)' 팩트체크에 초점을 두고 있는 셈이다. 본 연구의 심층인터뷰는 〈SNU팩트체크센터〉와 〈팩트체크넷〉에 소속된 주요 이해관계자를 대상으로 하고 있는데, 이는 이처럼 독특하게 구분되는 각 기관의 특성이 국내 팩트체킹 지형을 보다 입체적으로 파악하는 데 일조하리라는 판단에 기인했다.

인터뷰 참여자의 구체적인 선정은 우선 각 기관의 설립자 및 대표를 포함했고, 이들이 각

기관 안에서 인공지능 팩트체크 기술을 개발 또는 활용하는 핵심 이해당사자를 추천해주는 스노우볼링 방식으로 확장했다. 대상 선정 과정에서 특히 고려한 부분은 특정 기관의 경험과 사례에 국한되지 않은 전반적인 국내 팩트체킹 지형을 평가할 만한 환경적 조건에서의 경험과 지식, 그리고 팩트체킹 도구나 기능의 일부로 인공지능 기반 기술을 개발 혹은 활용한 경험이 있는 이해당사자의 식견을 접하는 것에 있었다. 2022년 6월부터 7월에 걸친 약 두 달 여의 기간 동안 각 기관에서 총 일곱 명의 이해관계자를 연구 대상으로 섭외해 심층 인터뷰를 실시하였다. 섭외된 인터뷰 참여자들의 활동분야는 저널리즘(2인), 정책 및 언론학(1인), 공학 및 기술개발(2인), 팩트체크 수행(2인)으로 구분된다. 국내 팩트체킹 지형도 및 인공지능 기술의 도입과 발전 과정에 대해 더 이상 새로운 이야기가 나오지 않고 인터뷰에서 언급되는 사회적 맥락에 대한 2차 분석이 필요하다고 판단되는 시점에서 인터뷰를 멈추었으며, 인터뷰 과정 중에 언급된 정부 정책과 정부 지원사업 관련 자료를 2차 자료로 해석하는 과정에서 심층적인 이해를 위해 참여자 인터뷰가 더 필요한 부분을 보강하기 위해 인터뷰참여자 1인을 추가로 섭외하였다. 아래의 〈Table 1〉은 인터뷰 참여자들의 배경과 소속 기관에서의 구체적 역할을 보여주고 있다. 2022년 당시 코로나 상황을 감안한 비대면 화상 인터뷰를 진행했으며, 각 기관의 팩트체크 방법과 인공지능 팩트체크에 대한 전망과 한계 등을 묻는 열 가지의 공통 질문(〈Table 2〉 참고) 을 기반으로 맥락에 따라 자유롭게 논의를 조정하거나 확장하는 반구조화 인터뷰 방식을 채택해 평균적으로 약 1시간~1시간 30분 정도 인터뷰를 진행하고 이를 녹화하였다. 참여자는 공통적으로 국내 팩트체킹 지형도 안에서 자신의 역할과 각자의 시각에서 보는 현재 국내 자동화 팩트체킹 기술의 역할 및 영향에 대한 질문을 받았다. 공통질문은 '팩트체킹 전반'에 대한 질문과 '인공지능과 팩트체킹'에 대한 질문으로 구분되었는데 참여자의 전문 분야에 따라 한 영역에 대해서는 답변을 해주지 못하는 경우도 있었다. 이러한 경우 연구결과의 분석에 답변이 포함되지는 못했지만 그 자체로 시사하는 바가 있다고 판단하여 연구결과를 서술하는 데 반영하였다. 인터뷰 참여자는 질문 내용을 사전에 받았으며 연구의 취지에 대해 충분한 설명을 듣고 인터뷰 참여, 녹화 및 논문 직접 인용에 대해 동의하는 절차를 거쳤다. 인터뷰 참여자 보호를 위해 논문에서는 각각의 참여자의 직접적인 실명은 익명 처리하고 기호로 대체하여 기술하였다.

Table 1. Interview Participants

참여자	성별	소속	활동내용	기타 관련 경력
A	여성	〈팩트체크넷〉	시민팩트체커	팩트체크 교육 프로그램 수료 팩트체크 경연대회 입상 미디어 리터러시 교육자
B	남성	〈팩트체크넷〉	시민팩트체커	팩트체크 교육 프로그램 수료 팩트체크 경연대회 입상
C	여성	〈팩트체크넷〉	설립자/ 운영위원회 소속	팩트체크 활동가 전직 기자
D	남성	〈팩트체크넷〉	플랫폼 디자인 자문역	웹 데이터베이스 전문가
E	남성	〈SNU팩트체크센터〉	인공지능 기술개발자/ 운영위원회 소속	HCI & 데이터 사이언스 학자 인공지능 기술개발자
F	여성	〈SNU팩트체크센터〉	설립자/ 운영위원회 소속	미디어 정책학자
G	남성	〈SNU팩트체크센터〉	대표자/ 운영위원회 소속	언론학자 전직 기자

Table 2. Key Interview Questions

질문 유형	질문 내용
팩트체킹 전반	1. 팩트체킹 플랫폼, 기술, 조직 등과 관련하여 어떤 일을 하고 계시는지 본인 소개를 부탁드립니다. 2. 귀하의 기관에서 팩트체킹은 어떤 과정을 거쳐 이루어집니까? 각 단계에서 어떤 이해관계자나 매체 및 기술이 참여하게 되는지 함께 설명해주십시오. 3. 기억에 남는 팩트체킹 사례나 모범적인 팩트체킹 사례가 무엇입니까? 그 이유도 함께 공유해주십시오.
팩트체킹과 인공지능	4. 알고리즘 및 인공지능의 도입은 사회 모든 영역에서 이루어지고 있습니다. 저널리즘에서 알고리즘/인공지능을 무엇이며 어떻게 활용되고 있습니까? 5. 그렇다면 팩트체킹 영역에서의 알고리즘/인공지능은 무엇이며 어떻게 활용되고 있습니까? 현재 활용되고 있거나 개발중인 알고리즘/인공지능 기반 팩트체크 사례가 있으면 함께 설명해주십시오. 6. 인공지능 기술 개발 이후 팩트체킹 과정이나 팩트체커의 활동 등 팩트체크 지형 전반에 변화가 있었나요? 있다면 설명해주십시오. 7. 저널리즘과 팩트체킹 각각의 영역에서 알고리즘 및 인공지능의 역할이 무엇이라고 생각하십니까? 혹은, 어떤 역할을 해야 한다고 생각하십니까? 8. 알고리즘/인공지능이 팩트체크에 도입된다고 할때 다음의 영역에서 어떤 원칙이 적용되어야한다고 생각하는지 의견을 말씀해주십시오. 1) 인간과 인공지능의 역할 배분 및 기여 비율 2) 텍스트와 이미지 사실 검증 3) 알고리즘의 판단이 인간 팩트체커의 의견과 충돌할 경우 대응 4) 팩트체킹 기반 데이터 수집 및 트레이닝 5) 알고리즘 검증 및 정확도/신뢰도 개선 방법 9. 국내외 현재 출시가 되었거나 개발 중인 인공지능 기반 자동화 팩트체크 도입 및 활용을 위한 연구 및 개발 협업 등의 한계가 무엇이라고 보십니까? 10. 인공지능 팩트체크의 활용도가 높은 부분은 무엇이라고 생각하십니까? 어떤 점에서 인간 팩트체커와의 보완적/대체적 활용이 가능하며 향후 활용 및 도입 가능성을 어떻게 평가 혹은 예측하십니까?

나아가, 본 연구에서는 팩트체킹 분야에서 국내 인공지능 기술의 발전과 활용 과정을 보다 구체적으로 살피기 위해 주요 연구대상으로 삼은 팩트체킹 센터 관련 자료와 팩트체킹 또는 팩트체킹 기술에 관한 내용을 다룬 언론보도, 정부 정책과 각종 지원사업 관련 정보, 국내외 학술논문 및 연구보고서, 국내외 인공지능 기반 팩트체킹 기술의 개발 및 활용 사례들을 함께 분석하였다. 관련 자료를 통한 현상 분석은 심층 인터뷰 과정에서 참여자의 다각적인 의견을 도출할 수 있는 질문의 일부로 사용됨과 동시에 인터뷰 사후에 인터뷰의 내용을 보다 풍부하게 해석하기 위한 자료로 활용되었다.

4. 연구결과

1) 국내 팩트체크 지형도 내 인공지능 기술의 역사적 구성

팩트체크가 정보 위기에 대한 유력한 대안으로 떠오르는 국제적인 움직임 속에서 국내 유력 일간지나 지상파 방송과 같은 전통적 미디어뿐 아니라 군소 인터넷 언론사에서도 뉴스와 언론환경의 자성과 신뢰 회복을 위한 유력한 해법으로 팩트체킹 관행의 정착과 실천이 점차 강조되어 왔다.

듀크대학교의 리포터스 랩에서 구축한 전세계 주요 팩트체크 기관 데이터베이스에 따르면, 2020년 7월 기준 한국의 공식적인 팩트체크 활동 기구는 뉴스톱과 SNU팩트체크센터, 그리고 11개 언론사 부속기구(AFP, JTBC, KBS, MBN, SBS, the300, 시사위크, 연합뉴스, 오마이뉴스, 전북일보, 한국경제)를 포함하는 총 13개 기구로 집계된 바 있다. 그 후에도 2021년 팩트체크넷이 출범하고 SNU팩트체크센터의 제휴 언론사가 34개사로 확대되는 등의 사례는 국내 팩트체크 지형이 양적으로 꾸준한 성장을 거듭해 왔음을 보여준다.

이러한 한국의 팩트체킹 지형은 언론사 단독 모델, 둘째는 대학과 언론사의 협업 모델, 그리고 시민사회 중심의 '풀뿌리' 모델의 세 가지 유형으로 특징지을 수 있다. 우선 언론사 단독 모델은 팩트체크를 개별 언론사의 업무로 보는 시각을 견지하고 있다. 언론사들이 자체적으로 중점부서 혹은 팀을 개설 혹은 보강하여 뉴스 생산 과정에서 정보에 대한 판단과 검증, 즉 팩트체킹 기능을 강화하는 유형이며, 많은 언론사들이 공식적, 비공식적으로 활용하고 있는 방식이다. TV 뉴스나 신문사의 '팩트체크' 코너가 좋은 예로, 2017년 팩트체크 전문 언론사를 표방하며 출범한 '뉴스톱(NewsToF)'과 같은 언론사도 이 유형에 속한다고 볼 수 있다. 다음은 SNU팩트체크센터와 같은 대학과 언론사 협업 모델이다(최순욱, 윤석민, 2017). 학교 등 비영리 교육연구기관이 학문과 현장의 접경에서 팩트체크의 원칙과 실천 및 전달 방식에 대한 전문적 지식체계를 구

축하고, 그 취지에 동의하는 언론사들의 자발적인 참여와 실천을 독려하는 방식이다. 팩트체크를 위한 기반을 구축하고 팩트체킹 과업의 일관된 기준과 원칙을 제시하여 전문화와 표준화를 추구함과 동시에, 시민들에 사실이 검증된 질 높은 정보를 제공하는 것을 궁극적인 목적으로 한다. 마지막으로, 팩트체크넷의 예를 통해 제시되는 시민사회 중심의 풀뿌리 모델은 팩트체킹 과정에 언론사와 각 분야의 전문가뿐 아니라 시민 개개인이 참여하여 함께 협업하는 방식을 지향했으며, 여기서 시민들의 역할은 언론인과 전문가를 보조하는 것이 아닌 사실 검증의 주체로써 팩트체킹 과정을 주도하는 데 있었다. 시민단체는 시민들의 역량을 키우고 동기를 부여하기 위한 다양한 커리큘럼과 프로그램을 개발하고, 시민들이 보다 전문적인 도움을 구할 수 있는 창구가 되는 다양한 언론인과 전문가들과의 협력적 관계를 구축하는 데 중심적 역할을 하는 것이다.

이와 같은 국내 팩트체킹의 지형도 안에서 2020년을 전후하여 팩트체킹 인공지능 기술개발에 대한 사회적 관심이 커지고 있음을 확인할 수 있었다. SNU팩트체크센터의 2018년 〈팩트체크 컨퍼런스〉, 팩트체크넷의 2021년과 2022년 〈팩트체크 주간〉행사에서는 세계의 팩트체크 현황 및 세계 각지에서 인공지능기반 팩트체크 기술이 어떻게 발전하고 있는지 소개됐고, 스페인, 호주, 영국의 개발자들이 협력한 'ClaimHunter(클레임헌터)', 듀크대, 텍사스대, 구글 등이 협력 개발한 'Claimbuster(클레임버스터)'나 영국 최대 팩트체크 기관인 FullFact(풀팩트)에서 개발한 '알파(alpha)' 등이 대표적 서비스로 언급됐다.

국내의 인공지능 기술을 응용한 한국형 팩트체크 서비스의 본격적인 개발은 정부 주도의 지원 사업 운영을 통해 이루어졌다. 2019년부터 인터넷 환경에서의 허위정보 문제를 해결하기 위한 정책적 필요성이 대두되며, 방송통신위원회(이하 방통위) 및 산하 기관인 시청자미디어재단을 중심으로 '인터넷 환경의 신뢰도 기반 조성'사업을 기획하고 팩트체크 시스템 구축 및 고도화에 상당한 예산을 편성했다. 그 결과 2021년 SNU팩트체크센터의 이준환 교수를 중심으로 한 연구팀이 인공지능기반 한국형 자동화 팩트체크 모델 개발에 착수했다. 이준환 교수의 모델은 앞서 언급한 FEVER 데이터셋 모델의 원리를 차용하여 한국어 위키피디아와 인터넷 뉴스를 바탕으로 주장과 근거들을 구성해 8만여 건의 데이터셋을 구축하고, 인공신경망 기반의 머신러닝 모델을 학습시켰다 (〈Figure 1〉 참고).

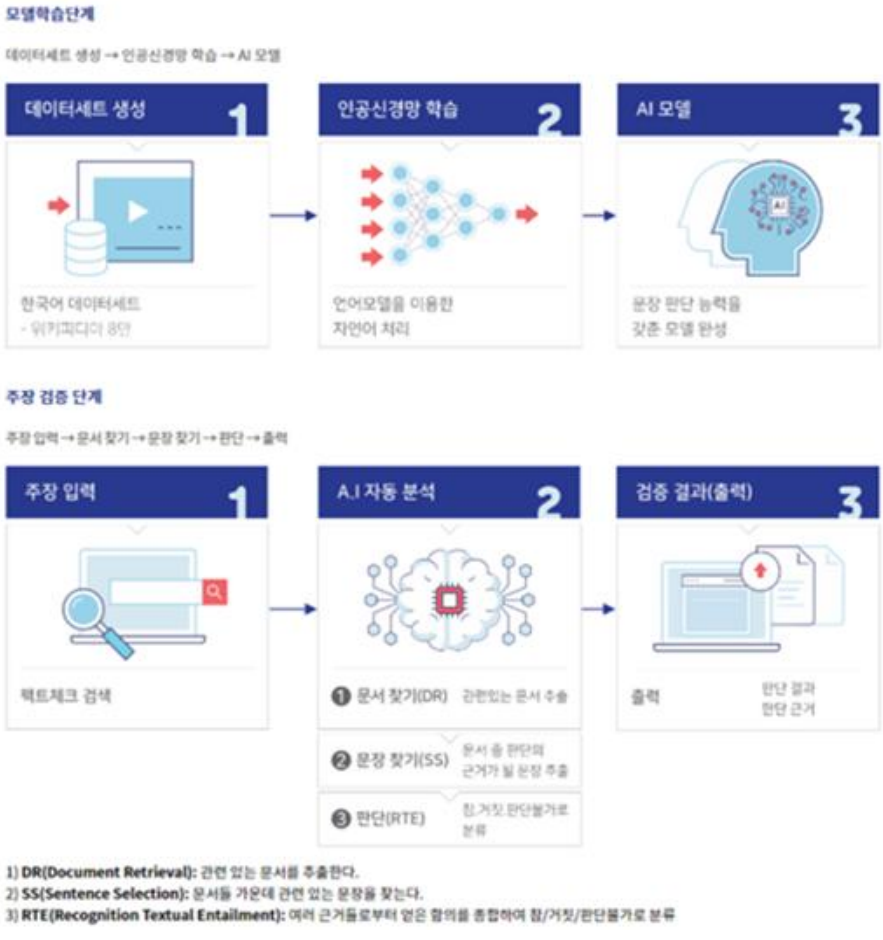

Figure 1. Korean AI Fact Checker: Working mechanisms. Source: Factchecknet. http://factchecker.or.kr/fc_autos

이 모델은 팩트체크넷의 AI 기반 자동화 팩트체크 서비스인 〈AINET〉에 탑재되어 시범 서비스 버전으로 운영되었으며, 모델 개선과 연구 확장을 위한 데이터셋 또한 함께 공개되었다. 〈Figure 1〉에 나타난 것처럼 입력한 주장에 대해 관련 문서와 문장을 찾아 추출된 증거와 함께 사실 여부를 판정하는 일련의 과정을 자동화한 서비스이다. 그러나 훈련 데이터 부족 및 데이터 저작권 문제로 인해 시범 서비스 단계에서 선보인 모델의 정확도를 개선하지 못하고 결국 2023년 초 서비스 자체가 중단되기에 이르렀다.

Table 3. National Research Trends on AI Factcheck

과업	인공지능 모델
주장 탐지	해당없음
증거 추출 및 주장 검증	NLP 기반 모델 • 뉴스 및 소셜 미디어 활용해 가짜뉴스 탐지(Hyun & Kim, 2018) • 요약기법(summarization techniques)을 활용한 가짜뉴스 탐지(Sim, Won, & Ahn, 2019) • 뉴스기사 맥락과 인간 검증 요소를 활용해 가짜뉴스 탐지(Youn & Ahn, 2018; Han & Kim, 2021; Lim, Kim, & Choi, 2021) • 뉴스 헤드라인과 세부 내용 활용한 가짜뉴스 탐지(Youn, 2018; Lee, Kim, Kim, Park, & Yang, 2019) • 한국형 AI 응용 팩트체크 모델 (Lee, 2021) 사회적 단서(social cues) 활용 모델 • 소셜네트워크 관계 데이터를 활용한 가짜뉴스 탐지(Jung & Ahn, 2022) • SNS이용자의 다양한 특성을 활용한 허위 주장 검증 (Jung & Lee, 2019) • 게시자 및 이해관계 집단의 협력자의 행동을 활용한 이상 정보 탐지(Lee, 2020)
다중감각(멀티모달리티) 활용	딥페이크 비디오 탐지(Lee & Moon, 2020)

이처럼 증가하는 팩트체킹 인공지능 기술개발의 필요성에 대한 인식과 달리 실제 국내에서 인공지능 팩트체크 기술의 발전 수준과 역할은 답보 상태에 머물러 왔다. 최근까지의 전반적인 국내 연구 동향을 살펴보면, 인공지능 기반 팩트체크 기술에 대한 연구는 지극히 제한적이다. 한국학술지인용색인(Korea Citation Index) 기준으로, 1)실제 팩트체킹 모델 개발과 관련이 있고 2)외국어가 아닌 한국어 기반 팩트체킹 과업을 다루고 있는 연구는 채 20건도 검색되지 않고 있는 실정이다. 〈Table 3〉은 국내에서 2018년부터 현재까지 진행되어 온 팩트체킹 기술 분야 연구들을 보여주는데, 국내 연구는 주로 텍스트와 소셜미디어 이용자 특성 및 네트워크 특성을 검증 요소로 활용하고 있으며, 엄밀하게 사실검증보다는 가짜뉴스 탐지에 보다 초점을 맞추고 있는 경향이 있다. 또한 검증할 만한 정보를 골라내고(claim detection), 증거를 추출한 후(evidence retrieval), 사실성을 판정하고(claim verification), 그 근거를 산출(justification production)하는 단계로 나뉘는 팩트체킹 자동화 기술의 과업 중 증거 추출 및 사실성 판정 단계에만 연구가 집중되어 있는 양상이다. 해외에서 논의되기 시작한 복합 요소(multimodality)를 활용한 사실 검증이나 "설명 가능한(explainable)" 모델 개발까지는 연구의 진척이 이루어지지 못했으며, 무엇보다도 현재까지 시도된 연구 중 실제 대중에 공개되어 활용되는 인공지능 응용 팩트체크 서비스는 거의 찾기 어려운 상황이다.

부족한 공적 지원과 연구자들의 저조한 관심 외에도, 한국형 팩트체크 자동화 연구가 직면

한 어려움은 연구에 우호적이지 않은 데이터 환경에서 찾을 수 있다. 자연어 처리 기반 팩트체크 모델의 경우 특정 주장의 검증을 위해 '사실성이 사전에 검증된(ground-truth)' 대용량의 데이터 코퍼스(예: 사전)로 사전 훈련된 언어 모델(pre-trained language model)을 구축하는 방식이 많이 활용되는데, 문제는 사실성이 검증되었으며, 연구에 활용하기 용이하게 정제된 대용량의 한국어 데이터가 부족하다는 데 있다. 백과사전 등 사전이 가장 좋은 데이터가 되겠지만 인공지능 기반 팩트체킹 기술 전문가인 참여자의 진술에 따르면 현재 국내 대부분의 사전 콘텐츠는 저작권법에 의해 보호되는 저작물로, 연구 목적의 자유로운 이용에 제한이 있으며 출판사 측에서도 연구 개발 참여는 환영하지만 연구 목적의 데이터 공유를 크게 반기지는 않았다고 밝히고 있다(Participant E). 이에 차선으로 고려된 웹 기반의 오픈 백과사전인 위키백과 등은 문장과 내용의 품질이 상대적으로 떨어져 이를 기반으로 훈련된 언어모델의 성능이 그리 뛰어나지 않았다고 평가되었다(Participant E).

> "인공지능을 활용한 팩트 체킹 같은 경우에는 저희가 지금 계속 데이터를 수집하고 분석하고 데이터를 수집해서 학습용 데이터를 만들고 이런 작업들을 거쳐왔는데 그 과정에서 이제 **저작권 문제라든가 데이터 학습용 데이터의 퀄리티 문제 이런 거에 따라서 결과에 많이 영향**을 미치는 거를 보아왔어요…(중략)…근데 이게 여러 가지 이유가 있을 수 있는데 아마 단순 사실 같은 것들보다 과학적인 사실이 한국어 위키피디아에 **좀 너무 없어서 문장이 많이 학습되지 않아서** 이럴 가능성도 있어요…(중략)…그 문장이 **두 문장 이상 정도 되면은 그런 문제들이 발생할 수가 있어서** 사실 저희가 지금은 아직 그 두 문장 이상의 주장을 판별해내지는 못하고 있어요 (Participant E)"

저작권법에서 자유롭거나 데이터 공유를 허락한 자료를 사용하는 방안도 있을 수 있다. 때문에 연합뉴스 신문기사, 통계청 데이터, 국회 보고서 등도 차선책으로 고려되었지만 근거 자료를 바탕으로 문장의 참, 거짓을 밝혀야하는 현재의 인공지능 팩트체크 시스템의 구조 상 근거 자료 자체의 팩트 체크가 선행되어야 하는 신문 기사나 시계열로 정보가 바뀔 수 있는 통계 데이터, 보고서의 경우는 사실상 활용하는 데 한계가 있음을 밝혔다(Participant E).

아울러, 정부 차원의 지원사업이 이루어질 때에는 인공지능을 통한 가짜뉴스 판별 및 자동화된 팩트체크라는 포괄적인 포부가 제시된 데 반해, 실제 이해관계자들은 팩트체크의 전 과정을 관리하고 판단하는 것은 어디까지나 인간의 영역이라는 입장을 고수하고 있었다. 인공지능이 도입이 되더라도 팩트체크에 필요한 여러 단계 중 일부분에 보조적인 기능이 가능할 뿐 '자동화된

팩트체크'나 인공지능만을 활용한 가짜뉴스 판별은 태생적인 한계가 있음을 기술개발자, 팩트체크 기관 설립자, 팩트체커 활동가 모두 공통적으로 지적했다.

이는 일차적으로, 팩트체킹 실천의 영역에서 적용 가능한 인공지능 기술의 한계 및 기술의 적용 범주에 대해 뚜렷한 구상이 제시되지 않은 채 연구 지원이 진행된 현실에 대한 인식을 반영한다고 볼 수 있다. 해외에서 개발된 모델을 국내 맥락에서 재현하는 데서 출발하고 있는 국내 기술은 연구 초기 단계에서부터 고전했다. 예컨대, 서울대학교에서 개발된 자연어 처리 기반 한국형 인공지능 팩트체킹 모델은 위키피디아를 데이터로 삼아 특정 주장에 대한 사실 여부와 근거 문장을 확인할 수 있는 서비스를 제공한다. 연구 초기 단계에서 공개된 시범 서비스는 두 문장 이상의 복잡한 문장을 검증하는 것이 불가능하고, 검증 정확도 이슈가 꾸준히 문제시되고 있어 실제 현장에 적용될 수 있는 단계는 아니었다. 시범서비스를 접한 인터뷰 참여자 역시 아직은 실제 팩트체킹 과정에 적용하기에는 어려움이 있다고 평가하고 있다(Participant A, Participant B). 그리고 2023년 현재는 해당 모델의 시범 서비스마저도 중단되어 있는 상황이다.

훈련 데이터가 갖는 한계에도 불구하고 국내 연구가 앞서 언급한 인공지능 팩트체크 기술 과업의 분류 중 증거 추출을 통한 참, 거짓 판별에 머무를 수 밖에 없었던 이유에 대해 인터뷰 참여자는 국내 미디어 환경에 대해서 언급했다. 아래 인용에서 볼 수 있듯이 인공지능 기술개발 영역의 한 참여자는 정보 확산 네트워크 특성을 활용한 가짜 뉴스 탐지에 인공지능 팩트체크가 기여할 수 있는 부분이 있을 것이나, 국내의 경우 메신저 내 정보 네트워크 양상을 파악할 수 없는 경우가 많기 때문에 현실성이 떨어진다고 밝혔다(Participant E). 이는 팩트체크 시스템 내에서 인공지능이 기술적으로 적용 가능한 부분과 실제 개발자가 접근할 수 있는 미디어 환경의 범위 사이의 괴리 때문에 인공지능 기술의 구현이 어려울 수 있는 영역이 있음을 보여준다.

> "네트워크를 분석을 하는 거는…메시지가 계속 네트워크를 타고 확산이 되잖아요. 불량 노드가 어딘지를 찾아내면…이 노드는 불량 노드…가짜 뉴스를 많이 보내는 그런 사용자다 〔라는 판단을 할 수 있게 되고〕 그럼 거기서부터 파생된 네트워크 그래프들은 불량일 가능성이 높다라고 판단을 하는 거죠…〔중략〕…예전에 이제 트위터가 굉장히 활발할 때는 이런 연구들이 좀 많이 수행이 됐었는데 트위터 같은 것들이 이제 많이 죽었잖아요. 그리고 페이스북이나 아니면 카카오톡 같은 경우에는 사용자의 네트워크를 우리가 볼 수가 없잖아요. 데이터를 수집을 할 수 없기 때문에 네트워크 연구를 하기가 굉장히 어려워요. 그래서 네트워크로 가짜 뉴스 확산을 보는 게 현실적으로 쉬운 일이 아닌 거죠(Participant E)."

2) 내부 담론을 통해 본 한국 인공지능 기반 팩트체크의 과제

참여자들이 공통적으로 지적하는 기술에 대한 우려는 팩트체킹에서 '필요한' 부문의 기술이 무엇인지에 대한 논의가 부재하다는 점이다. 현재 한국형 기술들이 지체되거나 대부분 시범서비스 단계에 있는 상황을 차치하더라도, 지금 이뤄지는 연구개발 과정들이 실제 팩트체킹을 실시하는 사람들이 정작 필요한 기술이 무엇인지에 대한 충분한 대화나 협의 없이 진행되고 있다는 것이다. 실제로, 현재 팩트체킹 과정에서는 문장의 참, 거짓을 자동으로 판정하는 인공지능 팩트체크 서비스를 활용하기 보다는, 일반적으로 구글 리버스(reverse) 이미지 검색이나 번역, 학술검색, 크라우드탱글(CrowdTangle)과 같은 소셜미디어 검색 도구 등 범용 디지털 도구들을 필요에 따라 복합적으로 활용하는 추세이다.

이런 상황에서 실제 팩트체킹을 하고 있는 인터뷰 참여자들은 주장의 참, 거짓 자체를 자동으로 판별해주는 인공지능 기술보다는 인간 팩트체커가 사실의 층위를 밝혀내는 과정에서 보조적인 역할을 하는 인공지능 기술개발 도입의 필요성을 제기했다. 일례로, 한 시민팩트체커는 해외 연구를 소재로 한 정보를 검증하는 과정에서의 어려움을 언급하고 있다. 전체 문서의 내용 중 사실성 검증에 관련이 있는 문장들로 범위가 좁혀져야 이를 번역해 검증하기 용이한데, 지금 국내에 이와 같은 기술은 존재하지 않기 때문이다.

> "(지금은 구글링 등 검색을 사용하는 것이 더 편한데)…보면 이 정도 부분일 거야 라고 그림으로 추측을 해서 〔자의적으로 판단할 수 밖에 없는 부분이 있지만 사실은〕 번역 시스템을 계속 돌려보고 조금 더 확인해 보고 이런 작업들이 필요한데 이런 부분을 좀 AI 팩트체크라든가 이런 AI검증 시스템들이…**좀 더 도와줄 수 있는 그러한 알고리즘을 좀 더 생성해 주면**…(Participant A)"

그 밖에도, 시민 팩트체커로 활동하고 있는 한 참여자는 기본적인 포털 검색을 통해 얻어지는 광범위한 정보의 사실 여부를 인공지능이 일차적으로 걸러내어 이를 팩트체크를 위해 선별된 근거자료로 활용할 수 있길 바랐고(Participant A), 전직 기자이자 기관 운영자인 한 참여자는 뉴스 기사 중 신빙성이 낮은 익명의 제보 인용, 따옴표로 엮은 기사, 기타 간접적 표현을 필터링 해준다면 사실이 아닐 가능성이 높은 정보를 일차적으로 걸러낼 수 있을 것이라고 밝혔다(Participant C). 또 다른 시민 팩트체커는 인공지능 팩트체크의 개발을 통해 기술과 사람의 팩트체크를 교차 검증하는 과정을 거쳐 사실 확인의 신빙성을 높일 수 있을 것 같다는 기대를 표했다(Participant B). 인공지능의 역할을 인간의 판단을 대신하기보다 인간의 판단을 용이하게 하

기 위한 도움을 제공하는 목적에 한정한다는 점에서 유사한 맥락에 있는 이러한 견해들은 정책 목표, 기술 개발 가능성, 팩트체크 현장 내 활용가능성 사이의 간극을 공통적으로 보여주고 있다.

나아가, 모든 참여자들 사이에서 가장 중요하게 인식되는 것은 본질적으로 팩트체크가 다루는 것이 외부 세계에 존재하는 객관적인 사실(Fact) 자체가 아니라 사회적 행위자들 사이에서 상호주관적, 사회적으로 구성되는 과정이라는 점에서 인공지능을 위시한 지능정보기술이 이 같은 복잡다단한 사실의 층위에 적절히 개입할 수 있을지에 대한 비판적인 시각이었다.

첫째로 무수한 사회적 정보의 사실성은 사회적 행위자들 각각이 인식하는 주관적 가치판단의 영향에서 자유로울 수 없기에, 사실 검증 과정에서 나타날 수 있는 잠재적인 편향이나 객관성의 문제를 기술이 효과적으로 처리하는 것이 가능한지, 나아가 그것이 과연 적절한지에 대한 의문이다.

> "사실 이제 인간도…판단을 할 때 잘못 판단하는 경우들이 생길 수 있기 때문에 서로 어떤 것도 맹신을 해서는 안 되는 부분이고…저희가 팩트체크에서 갖고 있는 굉장히 중요한 원칙 중 〔하나가〕 팩트체크는 100%의 확인이라는 게 없다라는 것이거든요. 그러니까 팩트는 끊임없이 찾아가는 과정인 것이지 이것 하나로 팩트〔가 완성이 되고〕 내가 하는 것이 완벽〔하다〕 라고 말하는 것이 아니기 때문에…팩트체크에서 수정의 원칙이 굉장히 중요하게 강조되고 있는 〔이유도 이 때문이구요〕 (Participant C)."

> "그 AI를 지원하고자 하는 욕망이라고 하는 건 사람이 했을 때의…선택 편향…같은 걸다 피해갈 수 있을 거라고 생각하는 그런 욕망이 있는 것 같아요…(중략)…우리가 다루고 있는 문제가 사회적인 문제잖아요…피지컬한 세상을 다루는 것이 아니기 때문에 그런 것들은 기본적으로 가치배제된 채로 볼 수 없는 문제들이 많이…(Participant G)"

이처럼 검증대에 오르는 '사실'들이 객관적 증명을 요하는 과학적 사실보다 주관적 가치 판단 과정에서 해석되는 사회적 사실에 가까워질 때, 팩트체크는 결국 사실성의 검증보다는 사실성에 다다르기 위한 '합의'의 도출에 비중을 두게 된다. 이러한 시각에서, 다음 인터뷰 발췌에서 조명되는 인간이 전담해 온 기존의 저널리즘 팩트체크 또한 다양한 관점에서의 판단을 통한 합의의 중요성을 강조하는데, 인공지능이 이 과정에 포함될 경우 발생할 수 있는 의사결정 체계의 혼란에 대한 우려 또한 엿볼 수 있다.

"…알고리즘이 하면 좀 더 중립적으로 하지 않겠느냐라고 하지만 제가 보기에는 알고리즘이라고 해서 〔편향성에 대한〕 질문을 그대로 피해갈 수 있을 건 아닐 것 같거든요…예를 들어 퀄리티 팩트 같은 경우에는 팩트체크 기사를 작성한 사람하고 그 기사를 데스킹 본 사람하고 판정을 일단 내려요. 판정을 내리면 전혀 그 판정 과정에 개입하지 않았던 다른 두 명의 에디터가 그거를 〔다시 보고〕…적어도 세 사람 중에서 두 사람이 동일한 판정을 내렸을 때에만 판정을 합의를 해요. 그니까 그런 종류의 판정을 합의하는 과정 같은 것들이 필요한데 그런 것이 인간 팩트 체크하고 알고리즘의 판단이 충돌할 경우라고 봤을 때에는 **아마도 인간과 인간이 합의하는 과정** 이런 것들이 다 들어가야 되겠죠. (Participant G)"

유사한 맥락에서, 한 참여자는 인공지능이 수행하는 팩트체크는 사실 기존의 팩트체크와는 맥락이나 범주가 다르며, 이를 명확히 구분해야 할 것을 강조한다. 참여자는 기술에 의한 팩트체크를 '팩트 베리피케이션(Fact verification)'이라는 용어로 구분했는데, 위의 인터뷰 발췌에서 지적한 '합의'를 기반으로 한 기존의 팩트체크와는 다른 '검증'의 범주로서 인공지능의 역할을 강조하는 것이다.

"저는 일부러 이제 팩트체킹 〔(fact-checking)〕이라기보다는 팩트 베리피케이션 〔(fact-verification)〕이라는 이야기를 많이 하고 있는데 그런 관점에서 한다면 실제로 이제 **사람이 하는 검색 행위 같은 것들을 자동화한다 라고** 보시면 될 거잖아요. 검색하고 판단하는 거를 자동화한다 (Participant E)"

그럼에도, 현재 국내 팩트체킹 영역에서 기술에 의한 팩트체킹과 인간에 의한 팩트체킹은 개념적으로 모호한 중첩성을 띄고 있다. 이는 '가짜뉴스'에 대응해야 하는 다양한 정치적 주체들의 주관적 인식과 이해 관계와도 얽히면서 인공지능 팩트체크에 대한 일관적인 정책 수립과 집행에 영향을 미치고 있다. 뿐만 아니라 팩트체크 의제 자체가 정치화됨에 따라 인간의 팩트체크 뿐 아니라 팩트체크를 위한 기술의 영역까지 정치적 해석에 좌우되는 양상이 나타나는 것이다.

한국의 인공지능 기반 팩트체킹 기술 발전을 둘러싼 학계, 언론, 시민사회, 정부 행위자를 아우르는 다양한 사회적 행위자들의 상호작용은 기술의 전반적인 발전 과정 속에서 유독 정부 행위자의 영향력이 두드러지게 나타난다는 특징이 있다. 우리 나라에서는 2019년부터 팩트체크에 대한 국가적 지원 및 인공지능 기술 도입의 필요성이 대두했다. 방송통신위원회, 방송통신심의위

원회, 정보통신정책연구원 등 국책연구기관의 지속적인 연구결과를 기반으로 방통위에 의해 '허위조작정보에 관한 전문가회의'가 조직되었으며, 학계·언론단체·관련전문가·시민단체 영역의 전문가 위원들이 문제해결을 위한 제언을 도출하고자 수 차례 회의를 거쳤다. 여기서 처음으로 AI를 통한 자동화된 팩트체크 시스템 개발 지원에 대한 사항이 정책과제로 언급되었다. 이후 사업 위탁기관인 시청자미디어재단을 주축으로 팩트체크 시스템 및 자동화 기술 개발, 팩트체크 시스템 구축 및 고도화에 각각 3.3억(2019년 편성, 총예산 6.1억), 14.95억(2020년 편성, 총예산 27.4억) 상당의 예산이 편성되는데, 이와 같은 국가주도적 지원의 본격화는 동 시기 거의 드러나지 않았던 민간 차원의 인공지능 팩트체크 개발 움직임과 확연히 대비된다.

그러나 이러한 정부 차원의 적극적인 지원은 오히려 개발의 안정적인 지속 노력에 걸림돌로 작용하게 된다. 이는 건전한 공적 담론 형성과 진작을 위해 부상했던 팩트체킹 관행이 과열된 정쟁 속에서 어느덧 또다른 정치적 무기로 변질되면서다. 지난 정권에서 중점적으로 추진된 팩트체크 지원 정책사업 전반에 대해 정권교체 이후 여권은 그 수혜 대상의 편향성과 편파성을 비판하며 예산 삭감을 주장해 왔고, 이에 따라 2021년과 2022년 팩트체크 정책 전반에 대한 전체 예산은 급격히 축소됐다. 그 중 팩트체킹 기술 지원 사업은 지원 대상에서 전면 배제되었으며, 팩트체크넷은 활동 예산 부족으로 인해 앞서 언급했듯 2023년 초 해산을 결의하기에 이르렀다. 팩트체킹 영역에서 이 같은 인공지능 발전의 '짧은 역사'는 언론의 중립성 문제와 민감하게 결부되기 마련인 팩트체킹 분야에서 균형을 유지한다는 것은 고도의 주의가 요구되는 문제이며(Participant F), 기술의 필요가 상업적 성격보다는 공적 기능을 수행하는 데 있을 경우 기술의 진보가 정치적 논쟁에 얼마나 취약해질 수 있는지 드러내고 있다(Participant E).

5. 논의 및 결론

이 연구는 국내 팩트체크 전문기관을 중심으로 언론계, 학계, 시민사회가 동참하고 있는 국내 팩트체킹의 전반적 지형에서 인공지능 기술의 역할과 범위에 대한 참여자 및 이해당사자의 인식을 알아보았다. 심층 인터뷰 결과에 대한 보다 입체적인 이해를 도모하기 위해 국가적 차원의 지원 및 육성 정책 추진 상황과 최근의 국제적 기술 발전 동향에 비춘 국내 개발 상황도 함께 살펴보았다. 심층 인터뷰와 현상 분석 결과를 토대로 국내 인공지능 기반 팩트체킹 기술의 역사적 구성과 이를 둘러싼 사회적 담론, 그리고 이를 바탕으로 한 앞으로의 과제와 전망에 대한 통찰을 정리했다.

참여자들을 대상으로 한 심층 인터뷰 결과, 인공지능 팩트체킹에 대한 정책적 지원과는 달리 실제 팩트체크 수행 과정에서 기술의 역할은 그리 두드러지지 않고 있음이 드러난다. 오히려 소모적인 정치적 논쟁 속에 몇 없는 팩트체크 기관 중 하나는 운영을 중단했고, 그 과정에서 그나마 유일하게 공개됐던 한국형 팩트체크 자동화 서비스는 현재 시범 단계에서 개발과 운영이 중단되기에 이르렀다. 정치화 논란으로 변질된 국가 차원의 연구지원과 이와 맞물린 저조한 연구관심으로 인해 현재 한국어 기반 팩트체크 서비스 기술의 발전은 답보 상태에 머물러 있다. 결과적으로, 가짜뉴스가 전세계적으로 문제시되기 시작한 2010년대 후반 이래 2018년부터 현재까지 인공지능 기반 정보 탐지와 사실성 검증 기술을 다루는 국내 연구는 상당히 지체되어 있다. 아울러, 저작권 등의 문제로 기계학습의 핵심적인 요소인 양질의 한글 데이터 공개가 제한적이며, 따라서 기술 발전을 뒷받침하기 위한 환경적 요건 또한 녹록지 않은 상황이다. 최근 OpenAI의 GPT를 시발점으로 생성형 인공지능(Generative AI)을 팩트체킹에 활용하거나 이 같은 거대언어 모델(large language model: LLM)을 기반 지식(Knowledge Base)이나 새로운 간접증거 등으로 팩트체킹에 활용하는 방식(Brown et al., 2020; Petroni et al., 2019) 등이 제안되고 있으나, 이 또한 거짓 주장을 사실인 양 전달하는 인공지능의 환각적 특성(hallucination)처럼 정보의 정확성 문제와 직결된 새로운 한계점들이 발견되며 난항을 겪고 있는 상황이다(Zuccon & Koopman, 2023). 인공지능 기술이 가짜뉴스에 대한 유일한 해결책은 아니지만 가짜뉴스를 생성하는 기술의 발전 속도에 비해 현저히 더딘 인공지능 기반 정보 탐지와 사실성 검증 기술은 언론계 전반과 팩트체크 제도가 당면한 과제를 여실히 드러낸다.

장기적인 관점에서 국내 팩트체킹 기술의 발전을 도모하기 위한 최우선의 과업은 팩트체킹 과정에서 인공지능의 역할과 범위에 대한 사회적 합의를 도출하는 데 있다. 연구 참여자들은 현재 연구개발 중인 기술이 실제 팩트체킹 과정에서 인간 팩트체커의 판단을 돕는 보조적 도구로서의 역할을 수행해야 한다는 점에 목소리를 같이하는 반면, 유용할 것으로 생각되는 인공지능 기술 적용 영역에 대해서는 각자 다른 구상을 가지고 있었다. 이는 성급한 기술 개발 이전에, 다양한 사회 구성원들의 필요와 요구를 수렴하고 접점을 찾기 위한 논의가 선행되어야 할 필요성을 부각시키고 있다. 아울러, 팩트체크의 본질적인 주관성과 정치화 현상은 사회적 합의의 중요성을 더한층 강조한다. 결국 가치나 맥락 정보가 내재된 사실성의 '사회적 해석'은 인간의 고유한 영역이며, 본 연구는 그 과정 안에서 기술이 개입할 수 있는 영역과 방식에 대한 충분한 사회적 합의가 선행되지 않는다면 팩트체크 기술 개발 과정이 어떻게 현장 적용 가능성과 사실성 경합의 정치화 속에서 결과 없는 시도로 남게 될 수 있는지 보여주었다.

이상의 문제에 대한 대안으로, 본 연구에서는 향후 팩트체킹 기술의 발전을 위해 개인적,

조직적, 사회적 측면에서 함께 고려해 볼 수 있는 몇 가지 중요한 지점들에 대한 제언을 제시하고자 한다. 먼저, 인공지능 연구의 토대가 되는 데이터 환경의 육성을 위한 사회 공동의 노력이 필요하다. 이는 사실 검증을 위한 데이터셋의 양과 다양성, 품질 등 데이터셋 자체에 기울이는 노력을 넘어 전사회적으로 과학기술 발전을 위한 데이터 개방과 공유의 보폭을 넓히는 연구 친화적 데이터 문화의 진작을 포괄한다. 한국형 팩트체킹 인공지능 모델의 성능 향상을 위해서는 방대한 양의 '사실 검증이 완료된' 한글 데이터셋을 통한 학습이 요구되기에, 사전이나 교과서, 사료, 검증된 뉴스기사 등 해당 조건에 부합하는 데이터들이 보다 자유롭게 공적으로 이용될 수 있는 방안을 고민해 보아야 할 것이다. 다음으로, 팩트체킹의 독립성과 공정성을 담보하고 강화하기 위한 거버넌스 차원의 담론이 지속되어야 한다. 언론의 공적 책무의 연장선상에서 유사한 사명을 띤 공적 역할을 수행하는 데 주안점을 두게 되는 팩트체킹 조직은 많은 경우 조직의 구성과 안정적인 운영을 위해 외부의 재원에 의존하게 되어 원치 않는 편향성 시비에 휘말릴 소지가 있다. 따라서 각 조직은 재원의 출처가 되는 개인 혹은 기관과의 객관적 거리 두기를 통해 조직 운영의 독립성을 확보함과 동시에 그 과정을 공중에 설득력 있게 전달하고 공정성에 대한 신뢰를 획득할 수 있는 구조적, 제도적, 담론적 장치가 무엇인지 끊임없이 고민하고 그 실행을 위해 지속적인 노력을 기울여야 한다. 마지막으로, 팩트체크가 사실성의 판정으로 완료되는 것이 아닌 끊임없는 합의의 과정 속에 위치한 과업이라는 점을 인지하고, 어떻게 이러한 팩트체크 과정을 사회적인 선순환 과정으로 정착시킬 수 있을지에 대해 사회 구성원 모두의 자성을 촉구하는 입장 또한 주목할 필요가 있다. 이는 '사실'이 무엇인지에 대해 우리 모두가 올바른 인식을 정립하고, 실제 팩트체킹 과정에서 적절한 실천 양식을 실행함과 동시에, 무엇보다도 소위 진영 논리에 얽매이지 않고 제시된 팩트체킹 결과에 대해 존중하고 수용하며, 여기서 파생되는 건강한 사회적 논의를 통해 올바른 사실성 판단을 위한 인식적 토대를 더욱 굳건히 하는 순환 과정을 포함한다. 한 인터뷰 참여자(Participant F)에 의해 강조된 이 마지막 제언은 인공지능 기반 팩트체킹 기술을 넘어 팩트체킹 영역 전반을 아우르는 건강한 사회적 공론장 형성을 위해 중요한 교훈이 될 수 있다.

본 연구를 통해 살펴본 국내 인공지능 기반 팩트체킹 기술은 전반적으로 가능성보다는 한계의 영역과 맞닿아 있다. 실제로 연구를 수행하는 과정에서 팩트체크와 인공지능이라는 교집합에 존재하는 기관도 〈SNU팩트체크센터〉와 〈팩트체크넷〉 두 곳에 불과하다는 것을 확인했으며 그 안에서 관련 전문가 혹은 활동가를 찾는 일 또한 쉽지 않았다. 이는 단순히 국내 연구자들의 관심과 데이터 부족 등 연구 환경의 문제가 아니라 복잡다단한 사회적 요인이 작용한 결과라고 볼 수 있다. 예를 들면, 정권의 교체 및 예산 삭감으로 인해 올해 초 그 길지 않은 여정을 끝내게

된 정부 출자 팩트체크 기관의 사례는 기술 뿐만 아니라 사회적, 정치적 요인이 한국의 팩트체크 환경이 급격히 위축되는 데 적지 않은 영향을 미치고 있음을 시사한다. 가짜뉴스와 팩트체크 관련 의제는 건전한 사회적 논의의 장을 형성하기보다는 정쟁의 소재가 되어 의제의 정치화를 부추기고 있다. 팩트체크 영역 자체에 대한 사회적 지지나 인정이 부족한 상황에서 팩트체킹 인공지능에 대한 예외적인 배려를 기대하기도 어렵다. 이 같은 사회적 분위기는 결국 팩트체크의 활성화와 관련 기술 개발을 통해 달성할 수 있는 언론 생태계의 재정비와 민주적 공론장 성장의 가능성이 후퇴하고 있는 작금의 상황을 여실히 보여주고 있다.

그럼에도 빠른 속도로 축적되는 방대한 정보를 효율적으로 처리하고, 감정이나 선입견 등 인간의 고질적인 편향에서 자유로울 것이라는 인공지능을 향한 일종의 '환상'을 좇는 근원적인 인간의 욕망은 인공지능에 의한 팩트체킹을 언제고 유력한 대안으로 다시 담론의 장에 소환해 낼 수 있다. (아직) 실현되지 않은 가능성의 연속선 상에서, 한국형 팩트체킹 인공지능에 대한 논의는 출발점에 서 있다.

References

Adair, B., & Stencel, M. (2016, June 22). *How we identify fact-checkers*. Duke Reporter's Lab. Retrieved 7/2/22 from https://reporterslab.org/how-we-identify-fact-checkers/

Alam, F., Cresci, S., Chakraborty, T., Silvestri, F., Dimitrov, D., Martino, G. D. S., ... & Nakov, P. (2022). A survey on multimodal disinformation detection. In *Proceedings of the 29th International Conference on Computational Linguistics*. doi:10.48550/arxiv.2103.12541.

Baik, J., Lee, S. E., Han, J., & Cha, M. (2021, October). *Objectivity in Korean news reporting: Machine learning-based verification of news headline accuracy*. Paper presented at the annual conference on Human and Cognitive Language Technology, Online. [백지수·이승언·한지영·차미영 (2021, 10월). <기계학습 기반 국내 뉴스 헤드라인의 정확성 검증 연구>. 제33회 한글 및 한국어 정보처리 학술대회. 온라인.]

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33*, 1877-1901.

Choi, S., & Koh, M. (2016). Media trend report. *SNU Institute of Communication Research Newsletter, 3*, 1-19. [최순욱·고문정 (2016). ICR Media Trend Report. <서울대학교 언론정보연구소 뉴스레터>, 통권 3호, 1-19.]

Choi, S., & Youn, S. (2017). The implications of collaborative fact-check service: Case of <SNU FactCheck>. *Journal of Cybercommunication Academic Society, 34*(2), 173-205. [최순욱·윤석민 (2017). 협업형 사실검증 서비스의 의의와 과제: <SNU 팩트체크>의 사례. <사이버커뮤니케이션학보>, 34권 2호, 173-205.]

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. <방송문화

연구>, 31권 1호, 47-101.]

Chong, E. (2022, May 10). *Fact-checking in the post-truth era*. OpenNet. Retrieved 5/7/22 from https://www.opennet.or.kr/20850 [정은령 (2022, 5, 10). 탈 진실시대의 팩트체크. <오픈넷>.]

Dotto, C., & Smith, R. (2019, October). *Newsgathering and monitoring on the social web.* First Draft News. Retrieved 5/12/22 from https://firstdraftnews.org/wp-content/uploads/2019/10/Newsgathering_and_Monitoring_Digital_AW3.pdf?x21167

EDMO. (n.d.). *Repository: Fact-Checking Initiatives in the EU (and in the UK)*. Retrieved 7/4/23 from https://edmo.eu/fact-checking-activities/

Factchecknet. (2021). *Rules and regulations for operation*. Retrieved 7/2/22 from https://factchecker.or.kr/terms/11 [팩트체크넷 (2021). 운영규정.]

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.

Han, Y., & Kim, G. (2021). A Study on automated fake news detection using verification articles. *KIPS Transactions on Software and Data Engineering, 10*(12), 569-578. [한윤진·김근형 (2021). 검증 자료를 활용한 가짜뉴스 탐지 자동화 연구. <정보처리학회논문지: 소프트웨어 및 데이터 공학>, 10권 12호, 569-578.]

Hyun, Y., & Kim, N. (2018). Text mining-based fake news detection using news and social media data. *The Journal of Society for e-Business Studies, 23*(4), 19-39. [현윤진·김남규 (2018). 뉴스와 소셜 데이터를 활용한 텍스트 기반 가짜 뉴스 탐지 방법론. <한국전자거래학회지>, 23권 4호, 19-39.]

IFCN. (2016). International fact-checking network fact-checkers' code of principles. Retrieved 7/2/22 from https://www.poynter.org/ifcn-fact-checkers-code-of-principles/

Jung, I., & Ahn, H. (2022). A study on the detection of fake news - The comparison of detection performance according to the use of social engagement networks. *Journal of Intelligence and Information Systems, 28*(1), 197-216. [정이태·안현철 (2022). 그래프 임베딩을 활용한 코로나 19 가짜뉴스 탐지 연구 - 사회적 참여 네트워크의 이용 여부에 따른 탐지 성능 비교. <지능정보연구>, 28권 1호, 197-216.]

Jung, S., & Lee, J. (2019). Possibility of automated fact checking through SNS user characteristics and confirmatory bias. *Korean Journal of Broadcasting & Telecommunications Research, 108*, 78-117. [정성욱·이준환 (2019). SNS 사용자 특성과 확증 편향을 통한 자동화된 팩트체킹의 가능성: 정치인

관련 트윗 데이터를 중심으로. <방송통신연구>, 통권 108호, 78-117.]

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.

Lee, D., Kim, Y., Kim, H., Park, S., & Yang, Y. (2019). Fake news detection using deep learning. *Journal of Information Processing Systems, 15*(5), 1119-1130.

Lee, D., & Moon, J. (2020). A method of detection of deepfake using bidirectional convolutional LSTM. *Journal of the Korea Institute of Information Security and Cryptology, 30*(6), 1053-1065. [이대현·문종섭 (2020). Bidirectional Convolutional LSTM을 이용한 Deepfake 탐지 방법. <정보보호학회논문지>, 30권 6호, 1053-1065.]

Lee, J. (2021). *AI factcheck*. Factchecknet. Retrieved 7/2/22 from https://ainet.factchecker.or.kr/ [이준환 (2021). AI 팩트체크. <팩트체크넷>.]

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.]

Lee, S. (2020). Characterization and detection of opinion manipulation on common interest groups in online communities. *Journal of Internet Computing and Services, 21*(6), 57-69. [이시형 (2020). 온라인 공간에서 관심집단 대상 비정상 정보의 특징 분석과 탐지. <인터넷정보학회논문지>, 21권 6호, 57-69.]

Lee, T. W., Yang, Y., Park, J. S., & Shon, J. G. (2021, November). *Fake news detection based on convolutional neural network and sentiment analysis*. Paper presented at the annual conference of Korea Information Processing Society, Yeosu. [이태원·양영욱·박지수·손진곤 (2021, 11월). <합성곱 신경망과 감성분석 기반의 가짜뉴스 탐지>. 한국정보처리학회 추계학술발표대회. 여수.]

Lim, D., Kim, G., & Choi, K. (2021). Development of a fake news detection model using text mining and deep learning algorithms. *Information Systems Review, 23*(4), 127-146. [임동훈·김건우·최근호 (2021). 텍스트 마이닝과 딥러닝 알고리즘을 이용한 가짜 뉴스 탐지 모델 개발. <경영정보학연구>, 23권 4호, 127-146.]

Marwick, A., & Lewis, R. (2017, May). *Media manipulation and disinformation online*. Data and Society. Retrieved 7/5/23 from https://datasociety.net/library/media-manipulation-and-disinfo-online/

Mueller, R. S. (2019, March). *Report on the investigation into Russian interference in the 2016 presidential*

election (Mueller report). Retrieved 5/12/22 from https://www.govinfo.gov/content/pkg/GPO-SCREPORT-MUELLER/pdf/GPO-SCREPORT-MUELLER.pdf

Nakov, P., Corney, D., Hasanain, M., Alam, F., Elsayed, T., Barrón-Cedeño, A., … & Da San Martino, G. (2021). Automated fact-checking for assisting human fact-checkers. doi:10.48550/arXiv.2103.07769.

Oh, M., Lee, J., Choi, H., Jin, J., & Chun, M. (2022). *A study on fake news detection through machine learning in the health and social welfare* (2022-48). Retrieved 7/7/23 from http://repository.kihasa.re.kr/handle/201002/42340 [오미애·이정란·최호식·진재현·천미경 (2022). 기계학습 기반 보건복지분야 가짜뉴스(fake news) 탐지 방법 연구. (한국보건사회연구원 연구보고서, 2022-48).]

Petroni, F., Rocktäschel, T., Lewis, P., Bakhtin, A., Wu, Y., Miller, A. H., & Riedel, S. (2019). Language models as knowledge bases? doi:10.48550/arXiv.1909.01066.

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 7/2/22 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. Retrieved 7/2/22 from https://reporterslab.org/fact-checking/

Shim, J. S., Won, H. R., & Ahn, H. (2019). A study on the effect of the document summarization technique on the fake news detection model. *Journal of Intelligence and Information Systems, 25*(3), 201-220.

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science, 50*(3), 755-769. doi:10.1111/j.1540-5907.2006.00214.x.

Thorne, J., & Vlachos, A. (2018). Automated fact checking: Task formulations, methods and future directions. doi:10.48550/arXiv.1806.07687.

Thorne, J., Vlachos, A., Christodoulopoulos, C., & Mittal, A. (2018). Fever: A large-scale dataset for fact extraction and verification. doi:10.48550/arXiv.1803.05355.

Tong, C., Gill, H., Li, J., Valenzuela, S., & Rojas, H. (2020). "Fake news is anything they say!"—Conceptualization and weaponization of fake news among the American public. *Mass Communication and Society, 23*(5), 755-778.

Uscinski, J. E. (2015). The epistemology of fact checking (is still naïve): Rejoinder to Amazeen. *Critical Review, 27*(2), 243-252.

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.

Yao, B. M., Shah, A., Sun, L., Cho, J. H., & Huang, L. (2022). End-to-end multimodal fact-checking and explanation generation: A challenging dataset and models. doi:10.48550/arXiv.2205.12487.

Ye, Q. (2023). Comparison of the transparency of fact-checking: A global perspective. *Journalism Practice*.

York, J. C. (2022). *Silicon values: The future of free speech under surveillance capitalism*. Verso Books.

Youn, T., & Ahn, H. (2018). Fake news detection for Korean news using text mining and machine learning techniques. *Journal of Information Technology Applications & Management, 25*(1), 19-32. [윤태욱·안현철 (2018). 텍스트 마이닝과 기계 학습을 이용한 국내 가짜뉴스 예측. <Journal of Information Technology Applications & Management>, 25권 1호, 19-32.]

Zarocostas, J. (2020). How to fight an infodemic. *The Lancet, 395*(10225), 676.

Zuccon, G., & Koopman, B. (2023). Dr ChatGPT, tell me what I want to hear: How prompt knowledge impacts health answer correctness. doi:10.48550/arXiv.2302.13793.

최초 투고일 2023년 02월 16일
게재 확정일 2023년 07월 13일
논문 수정일 2023년 07월 28일