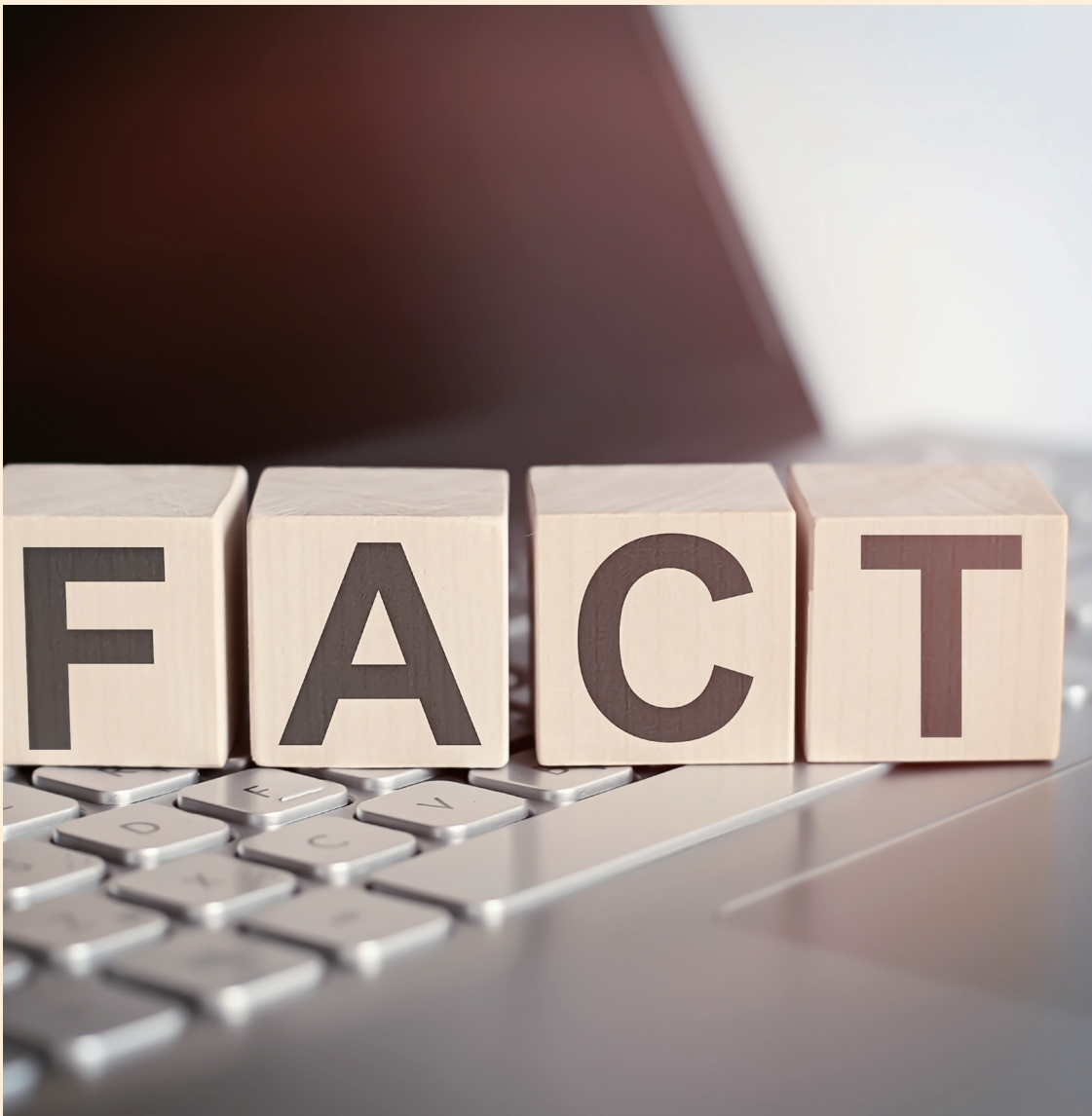


02

플랫폼 팩트체크 현황과 발전을 위한 제언

김경희 한림대 미디어스쿨 교수



I. 팩트체크의 개념과 중요성

미국 언론에서 100여 년에 걸쳐 정립된 <저널리즘 기본 원칙>을 저술한 코바치와 로젠스탈이 제시한 첫 번째 원칙은 '진실'이다(Kovach&Rosentiel, 2003). 가장 중요한 원칙이지만, 진실을 밝히는 일은 매우 힘든 일이다. 진실을 밝히기 위해서는 과정이 필요한데 그 과정의 핵심은 사실 검증이다. 사실을 검증해 보도하는 일은 기자라면 반드시 해야 하는 일이지만, 정치인이나 유튜버의 말을 확인 없이 그대로 받아쓰기 하는 언론들이 대세를 이루면서 팩트체크 저널리즘이 새로운 뉴스 유형으로 자리 잡고 있다.

팩트체크는 말 그대로 사실을 확인하는 과정으로, 1920~30년대 미국 뉴스잡지 업계에서 처음 등장한 것으로 추정된다(정은령, 2022). 초기 팩트체크는 기자들의 원고를 한 단어씩 검토하며 정확성을 확인하는 작업이었지만, 1990년대 이후에는 단순한 받아쓰기가 아니라 발언 내용 자체의 진실성을 검증하는 방식으로 변화했다. 현대적 팩트체크는 20세기 초 미국 언론계에서 확립된 객관주의 보도와는 차이가 있다(정은령, 2023). 객관주의 보도는 사실과 가치를 분리할 수 있다는 원칙 아래, 기자들이 정치적 논쟁을 공정하게 재현하고 논평 없이 뉴스만 전달해야 한다는 규범을 따른다. 반면, 팩트체크 저널리즘은 정치인의 발언을 그대로 전달하는 것이 아니라, 해당 주장의 정확성을 직접 평가한다는 점에서 차별화된다. 단순히 발언을 올바르게 인용했는지가 아니라, 그 내용이 사실인지 검증하는 과정에 초점을 맞춘다. 즉, 현대적 팩트체크는 독립적인 조사와 분석을 통해 발언의 진위를 판별하고, 그 결과를 독자들에게 제시하는 평가적 과정을 포함한다. 이를 통해 단순한 정보 전달을 넘어, 공적 담론의 정확성을 유지하고 허위정보(이 글에서는 허위조작정보+오정보를 허위정보라 표기함) 확산을 방지하는 적극적인 언론 활동으로 자리 잡고 있다.

팩트체크의 중요성이 더 부각된 것은 소위 '가짜뉴스'라고 불리는 허위조작정보가 사회를 혼란스럽게 만들고 있기 때문이다. 허위조작정보(Disinformation)는 상대를 속이기 위해 의도적으로 퍼뜨리는 거짓 정보를 의미한다(Bittman, 1985). 그냥 잘못된 정보(Misinformation, 오정보)와는 다르게, 허위조작정보는 '나쁜 의도'를 가지고 계획적으로 만들어지고 유포되는 정보라는 점에서 오정보와는 차이가 있다. 오정보는 한마디로 사실이 아니거나 오해를 불러일으키는 정보를 의미한다(Fetzer, 2004). 오정보도 사람들을 난처한 상황에 놓이게 하기 때문에 팩트체크가 필요하다. 그러나 최근에 더 빠르고 정확한 팩트체크가 필요한 이유는 개인이나 사회

전체에 심각한 피해를 줄 수 있는 허위조작정보가 확산되고 있기 때문이다. 더구나 아주 교묘하게 사실과 거짓을 섞어서 사람들을 속이는 허위조작정보가 늘어나고 있어 그 위험성이 더 커지고 있다.

그럼에도 불구하고 팩트체커들의 현황은 녹록지가 않다. 미국에 본부를 둔 비영리 언론 교육 기관인 포인터 연구소(The Poynter Institute for Media Studies)는 국제 팩트체크 네트워크(International Fact-Checking Network, 이후 IFCN으로 표기)를 운영하면서 전 세계 팩트체크 단체들의 협력을 촉진하고 기준을 정하는 역할을 하고 있다. IFCN의 보고서(2023)에 따르면, 137개 팩트체커 조직(69개국)이 참여한 설문에서 팩트체커들은 자금 조달과 지속 가능성을 주요 도전과제로 꼽았다. 우리나라에서도 한국 최초 팩트체크 플랫폼인 SNU팩트체크센터가 네이버의 자금지원 중단으로 출범 7년 만인 지난해 8월 무기한 활동 중단에 들어갔다. SNU팩트체크센터는 30여개 언론사들과 협업한 비영리 팩트체크 플랫폼으로 언론의 팩트체크 보도를 비교할 수 있고 근거자료를 명시하도록 하는 등 자체적으로 요구 기준을 마련해 한국의 팩트체크 생태계를 만들고 확장하는 데 기여해왔다(박재령, 2024, 8, 19). 허위조작정보로 인해 사회적 양극화가 더욱 심화되는 지금, 팩트체크의 중요성에 대한 사회적 인식이 확산되고, 이를 지속 가능하게 하기 위한 시민들의 자발적인 참여와 제도적 지원이 그 어느 때보다 필요한 시점이다.

II. 디지털 환경에서 팩트체크의 도전과 과제

1. 에코챔버 및 확증편향

에코챔버(Echo Chamber)는 비슷한 의견을 가진 사람들끼리 소통하며 자신의 신념을 강화하는 환경을 의미한다(Sunstein, 2001). 이러한 환경에서는 반대의견이 차단되거나 무시되어, 특정 신념이 증폭되는 현상이 나타난다. 확증편향(Confirmation Bias)은 자신의 기존 신념과 일치하는 정보는 받아들이고, 그렇지 않은 정보는 무시하거나 거부하는 경향을 의미하는데, 이는 인간의 인지적 편향 중 하나로, 객관적인 판단을 방해하고 특정 신념을 강화하는 역할을 한다(Nickerson, 1998). 에코챔버와 확증편향은 상호 작용하며 소셜 미디어 환경에서 더욱 심화되는 경향이 있다. 소셜 미디어의 알고리즘은 이용자의 기존 관심사와 성향에 맞는 정보만을 선별적으로 제공하여 에코챔버를 형성하고, 이는 이용자들이 자신이 믿고 싶은

정보만을 수용하도록 유도하면서 확증편향을 강화시킨다. 이러한 정보 환경에서는 객관적 사실보다 특정한 관점을 강화하는 정보가 더욱 확산되기 쉬우며, 정치적 양극화를 심화시키고 허위조작정보의 확산을 가속화하는 결과를 초래할 수 있다. 결국, 이러한 구조적 문제로 인해 팩트체크의 효과가 제한될 뿐만 아니라, 객관적인 검증 결과조차 받아들이지 않으려는 경향이 강해져 팩트체크의 어려움이 더욱 커지고 있다.

김새론 사건을 통해 에코챔버와 확증편향이 대중의 인식에 어떤 영향을 미쳤는지 확인할 수 있다. 2022년 음주운전 사고 이후, 비난 여론이 형성되면서 그에 대한 부정적 프레임이 강화되었고, 생활고 주장과 카페 아르바이트 논란, 홀덤펍 방문 등이 잇따라 보도되며 '생활고를 가장한다'는 인식이 확산되었다. 이러한 정보들이 필터링 없이 반복적으로 소비되면서 대중은 기존의 부정적 인식을 강화하는 방향으로 해석하는 확증편향을 보였다. 2025년 그가 사망한 이후에야 과거 논란이 재조명되면서 뒤늦게 알려진 피해 보상과 사과 행보로 인해 일부 동정 여론이 형성되었다(박서연·금준경, 2025, 2, 18). 김새론 사건은 디지털 미디어 환경에서 확증편향과 에코챔버 효과가 어떻게 특정한 서사를 강화하고, 팩트체킹을 무력화하는지 보여주는 대표적 사례다. 카페 아르바이트 사진이 조작이라는 주장은 검증 없이 빠르게 확산되었고, 반박 증거가 존재했음에도 불구하고 기존의 부정적 서사를 강화하는 방향으로 여론이 흘렀다. 이는 팩트체킹이 단순히 사실 여부를 검증하는 것이 아니라, 사람들의 신념과 감정이 결합된 정보 환경에서 진실을 전달하는 어려운 과정임을 시사한다. 팩트체킹이 허위정보가 자리 잡기 전에 적극적으로 개입하고, 건강한 정보 순환 구조를 만드는 과정임을 인식할 필요가 있다.

2. AI 기반 허위정보와 딥페이크

2023년 오픈AI의 DALL-E 3와 메타의 이미지 생성 모델 출시 이후 텍스트, 이미지, 동영상을 아우르는 복합적 허위정보 생성이 가능해졌다. 실제로 구글과 듀크대 연구팀이 팩트체크 사이트 및 미디어 단체와 공동으로 집필한 논문에 따르면, AI가 생성한 가짜 이미지가 2023년 초 이후 크게 늘어났다고 한다(김태중, 2024, 5, 31). 일례로 미국 대선 기간, 한 팟캐스트 진행자가 GPT-3를 활용해 조 바이든 후보의 인지 능력을 폄훼하는 가



짜뉴스를 대량 생산했다(김헌재, 2024). ‘바이든, 치매 의심 증후 보여’ 등의 제목과 함께 AI가 생성한 어설픈 연설 장면 이미지도 첨부해 웹사이트에 게시했고, 이 가짜뉴스들은 페이스북 등 소셜 미디어에서 수만 회 공유되며 확산됐다. 문제는 이런 가짜뉴스가 팩트체크 뉴스보다 더 많이 공유되고 더 많은 참여를 유도해내고 있어 그 영향력이 막강하다는 점이다. 버즈피드의 분석에 따르면, 2016년 전 세계적으로 가장 유행했던 가짜뉴스 50건 중 23건이 미국 정치와 관련된 내용이었는데, 이 가짜뉴스들은 총 110만 건의 페이스북 참여(공유, 좋아요, 댓글 등)를 유도한 것으로 분석됐다(Silverman, 2016, 12, 30). 또 미국 대선 기간 동안 일부 가짜뉴스는 주요 기성 언론사의 뉴스보다 더 많이 공유되기도 했다(Silverman, 2016, 11, 17).

딥페이크(Deepfake)도 허위정보를 생산하는 데 악용되고 있어 팩트체커들에게는 도전적 과제이다. 딥페이크는 ‘딥러닝(Deep Learning)’과 ‘가짜(Fake)’의 합성어로, 인공지능을 활용하여 사람의 이미지나 영상을 합성하거나 변형함으로써 실제와 구별하기 어려운 콘텐츠를 생성하는 기술(Kietzmann, 2020)을 의미한다. 이러한 기술은 긍정적인 활용 가능성에도 불구하고, 허위정보 확산, 명예훼손, 프라이버시 침해 등과 같은 윤리적 문제를 야기하고 있다. 여성가족부 산하 기관인 한국여성인권진흥원의 디지털성범죄피해자지원센터에 따르면, 딥페이크 영상물 피해 지원 건수가

2018년 177건에서 2019년 417건, 2020년 상반기 222건으로 증가세를 보이고 있으며, 2020년 6월부터 2023년 8월까지 방송통신심의위원회에 심의 요청된 딥페이크 기술을 활용한 성적 허위영상물이 9,006건에 달하는 것으로 나타났다(신성원, 2023).

딥페이크 기술은 정치적 선전 도구로도 적극 활용되고 있다. 특히 선거기간 동안 유권자들의 인식을 조작하려는 시도가 증가하면서, 딥페이크 기술이 민주주의에 미치는 영향이 더욱 커지고 있다. 영국 디지털 신원 확인 보안 업체에 따르면 2024년 선거가 예정되어 있는 국가에서 딥페이크 사건이 전년 대비 눈에 띄게 증가한 것으로 나타났다(김희정, 2024). 인도(280%), 미국(303%), 남아프리카공화국(500%), 멕시코(500%), 몰도바(900%), 인도네시아(1550%), 한국(1625%)이 크게 증가했고, 불가리아(3000%), 포르투갈(1700%), 벨기에(800%), 스페인(191%), 독일(142%), 프랑스(97%) 등 많은 국가에서 딥페이크 사건이 전년 대비 크게 증가한 것으로 분석됐다.

특히 딥페이크 영상이나 음성은 육안이나 단순한 기술적 분석만으로는 진위를 가리기 어렵기 때문에, 기존의 팩트체크 방식은 한계에 부딪치고 있다. 이러한 문제를 해결하기 위해 팩트체커들은 AI 기반 딥페이크 탐지 기술 도입, 출처 검증을 위한 메타데이터 분석, 전문가 네트워크와 협력을 통한 크로스체크 등 새로운 방식의 검증 기법을 개발하고 있다. 하지만 딥페이크 기술 또한 빠르게 발전하고 있어, 팩트체커들은 계속해서 진화하는 조작 기술에 대응할 수 있는 정교한 검증 시스템을 구축해야 하는 도전에 직면해 있다. 결국 생성형 AI와 딥페이크 시대의 팩트체크는 단순한 정보 검증을 넘어, 첨단 기술과 협력하여 진실과 허구를 가려내는 지속적인 싸움이 될 것으로 보인다.

3. 팩트체크의 시간적 지체 문제

소셜 미디어를 통한 허위정보의 확산 속도는 매우 빠른 반면, 이를 검증하는 팩트체크 과정에는 상당한 시간이 소요된다. 기존 연구(Vosoughi et al., 2018; Wu et al., 2019)에 따르면, 하나의 주장에 대한 검증에는 평균적으로 24~72시간이 걸리는데, 이 기간 동안 허위정보는 수백만 명의 이용자에게 도달할 수 있다고 한다. 팩트체크가 지연되는 주요 원인 중 하나는 인간 팩트체커들이 신뢰할 수 있는 여러 출처를 교차 검증해야 하기 때

문이다. 이를 해결하기 위해 자동화된 팩트체크 기술이 개발되고 있지만, 이러한 기술의 신뢰성과 정확성은 여전히 중요한 도전과제로 남아 있다. 결과적으로, 팩트체크가 이루어지기 전 허위정보가 대중의 인식에 영향을 미칠 수 있으며, 이로 인해 팩트체크의 교정 효과가 감소하는 문제가 발생한다(Lee, Xiong, Seo, & Lee, 2023).

포터와 우드(Porter & Wood, 2021)의 실험 결과에 따르면, 팩트체크는 허위정보에 대한 신념을 유의미하게 감소시키는 것으로 나타났다. 그러나 팩트체크가 지연될 경우 특히 허위정보에 노출된 후 2주 이상이 지난 뒤에 이루어진 팩트체크는 허위정보 교정 효과가 63% 감소하는 것으로 분석됐다. 이는 허위정보가 일단 사람들에게 내면화되면, 시간이 지나면서 교정이 훨씬 더 어려워진다는 것을 시사한다.

김형지 등(2020)은 연구를 통해 가짜뉴스는 소셜 미디어, 포털, 메신저 등을 통해 빠르게 전파되는 반면, 팩트체크 뉴스는 비교적 신뢰도가 높은 뉴스 포털이나 방송 매체를 통해 뒤늦게 전달되는 경향을 보여주었다. 이로 인해 허위정보가 사람들의 인식에 먼저 각인된 후 교정되는 시간적 지연이 발생하고 있고, 이는 가짜뉴스의 영향력을 더욱 강화하는 요인으로 작용하고 있었다. 이러한 문제는 플랫폼의 알고리즘 구조와도 밀접한 관련이 있다고 연구자들은 지적한다. 팩트체크 뉴스가 효과적으로 작용하려면 빠른 대응과 높은 가시성이 필수적이다. 따라서 플랫폼 차원의 알고리즘 개선이 필요하며, 팩트체크 정보가 신속하게 유통될 수 있는 구조 마련이 필요하다. 즉, 알고리즘을 통해 검증된 정보를 보다 가시적으로 제공하고, 허위정보보다 먼저 노출될 수 있도록 팩트체크 기사의 노출 우선순위를 상향 조정해야 하며, 팩트체크 속도를 높이기 위해 AI 기반의 자동 검증 기술을 활용하고, 팩트체크 결과가 빠르게 공유될 수 있도록 해야 하며, 허위정보가 포함된 콘텐츠에 팩트체크된 정보를 자동 연결해 이용자가 진실 여부를 즉시 확인할 수 있도록 해야 한다고 제안한다. 허위정보의 영향력을 줄이기 위해서는 팩트체크를 넘어, 플랫폼과 이용자가 협력하는 정보 환경을 구축하는 것이 필수적이다. 허위정보 확산 속도를 줄이고 팩트체크가 더 효과적으로 작용하기 위해서는, 플랫폼이 단순히 정보 전달자가 아닌 책임 있는 정보 필터링 역할을 수행해야 할 것이다.

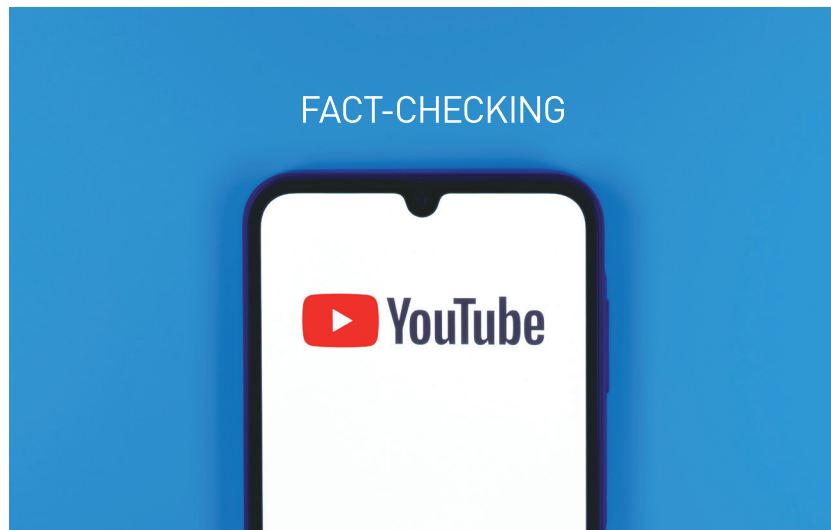
III. 주요 플랫폼의 팩트체크 정책 과 운영 현황

1. 메타(페이스북, 인스타그램)의 팩트체크 정책과 운영 현황

메타는 2016년부터 허위정보 확산을 방지하기 위해 제3자 팩트체크 프로그램을 운영해왔다(Meta, 2018, 6, 14). 이 프로그램은 IFCN의 인증을 받은 독립적인 팩트체크 기관과 협력하여 메타 플랫폼 내 콘텐츠의 정확성을 검토하는 것이 핵심이다. 전 세계 60여 개 언어로 콘텐츠를 검토하는 80여 개의 팩트체크 기관과 협력하여, 허위로 판명된 콘텐츠의 노출을 줄이고 경고 라벨을 보여주었다. 팩트체크 파트너가 허위로 판정한 콘텐츠는 뉴스피드 하단으로 이동시켜 노출을 억제하고, 이러한 콘텐츠를 반복적으로 공유하는 페이지나 도메인에 대해서는 수익화 및 광고 권한을 제한하는 정책을 펼쳤다.

그러나 2025년 1월 메타는 표현의 자유 확대와 콘텐츠 관리 오류 감소를 이유로 미국에서 제3자 팩트체크 프로그램을 종료하고, 커뮤니티 노트(Community Notes) 모델로 전환하겠다고 발표했다(Meta, 2025, 2, 20). 이에 따라 2025년 2월 20일 페이스북, 인스타그램, 스레드에서 커뮤니티 노트 기능이 도입되어 승인된 기여자들이 거짓이거나 오해를 불러일으킬 수 있는 콘텐츠를 발견하면 이에 대한 추가적인 맥락을 제공하는 노트를 첨부할 수 있게 됐다. 메타는 이 기능을 통해 플랫폼이 직접 콘텐츠를 검열하기보다는 이용자들이 자율적으로 정보의 신뢰성을 평가하는 구조를 만들겠다는 목표를 내세웠다. 커뮤니티 노트는 이용자 참여 기반의 팩트체크 시스템으로, 이용자가 특정 게시물에 노트를 작성하고, 다른 이용자들이 해당 노트의 유용성을 평가하는 방식으로 운영된다. 다양한 관점을 가진 이용자들로부터 유용하다고 평가된 노트만 게시물에 공개되며, 이를 통해 허위정보에 대한 보완 정보를 제공할 수 있다.

그러나 커뮤니티 노트 모델에는 몇 가지 한계점이 존재한다. 최소 6개월 이상 활동한 사용자, 활성화된 전화번호를 등록한 사용자, 플랫폼 규정을 위반한 적이 없는 사용자면 승인된 기여자가 될 수 있어 누구나 커뮤니티 노트를 작성할 수 있는 것이기 때문에 일반 이용자가 작성하는 노트의 신뢰성과 정확성에 대한 우려가 있다. 또 유효한 팩트체크도 특정 그룹이 다수 의견을 형성하면 공개되지 않을 수 있다. 가령, 어떤 노트가 실제로 정확한 정보를 제공하고 있더라도, 투표한 기여자들의 성향이 다양하지 않다면 '유용한 정보'로 평가되지 않아 일반 사용자들이 볼 수 없다. 2023년 10월,



비영리단체 ‘디지털증오대응센터(Center for Countering Digital Hate)’의 보고서에 따르면, 미국 선거와 관련된 허위정보 게시물 283건 중 74%에 해당하는 209건의 게시물에서 정확한 정보를 제공하는 커뮤니티 노트가 표시되지 않았다(CBS News, 2025, 1, 10). 이는 커뮤니티 노트 시스템이 허위정보를 교정하는 역할을 수행하는 데 한계가 있음을 보여준다. 특히 전문 팩트체커의 부재로 인해 복잡한 이슈나 전문 지식이 필요한 콘텐츠에 대한 검증이 부족할 수 있으며, 이는 오히려 허위정보가 정당성을 얻는 결과로 이어질 가능성도 있다. 커뮤니티 노트 기능이 효과적으로 작동하려면, 노트 작성자의 신뢰도를 평가하는 강화된 검증 시스템과 악용을 방지하기 위한 추가적인 조치가 필요하다는 지적이 나오고 있다.

2. 유튜브의 팩트체크 정책과 운영 현황

유튜브는 플랫폼 내 허위정보 확산을 방지하고 이용자들에게 신뢰할 수 있는 정보를 제공하기 위해 제3자 팩트체크 프로그램을 도입하고 있다(Ma & Feldman, 2022, 11, 29). 이는 구글 뉴스 이니셔티브(Google News Initiative)의 일환으로, 전 세계의 독립적인 팩트체크 기관과 협력하여 콘텐츠의 정확성을 검증하고 있다.

또 유튜브는 2023년 8월 의료 오정보 확산을 막기 위해 새로운 정책

을 발표하고, 신뢰할 수 있는 건강 정보를 제공하는 채널을 우선적으로 노출하는 시스템을 도입했다(University of Nebraska Medical Center, 2023, 8, 15). 유튜브는 미국 의학한림원(NAM), 세계보건기구(WHO), 의학전문학회협의회(CMSS) 등 국제적으로 권위 있는 기관의 원칙을 바탕으로 공신력 있는 출처를 선정하고, 기준을 충족하지 않는 콘텐츠에 대해 검토하는 방식으로 의료 오정보를 식별하고 있다(Graham, 2024, 1, 30). 유튜브 헬스 기능이 적용되면 ‘대한민국의 면허를 소지한 의사’ 등 공신력 있는 출처의 동영상임을 확인할 수 있는 라벨(보건 정보 출처에 대한 정보 패널)이 표시되고 건강 관련 주제를 검색했을 때 먼저 추천되는 ‘보건 콘텐츠 섹션’에 포함된다(최현석, 2024, 2, 1).

이와 함께 유튜브는 X(옛 트위터)와 유사한 형태로 커뮤니티 노트 기능을 도입했다(Salinas, 2024, 6, 19). 커뮤니티 노트는 선별된 일부 기여자들이 동영상에 대한 추가 정보를 제공하고, 서로 다른 성향의 사용자들이 동일하게 ‘유용하다’고 평가한 경우, 해당 노트가 공개될 가능성이 높아진다.

그럼에도 불구하고 유튜브 콘텐츠에서 팩트체크가 제대로 이루어지고 있다고 신뢰하는 사람들은 많지 않을 것이다. 유튜브가 허위정보를 차단하고 신뢰할 수 있는 정보를 제공하려면, 좀 더 적극적인 대응이 필요하다. 팩트체크 기관과의 협력 강화와 알고리즘의 투명성을 확대하고, 이용자들의 미디어 리터러시 향상을 위한 노력도 기울여야 한다. 또 팩트체크 시스템과 커뮤니티 노트 기능이 실제 운영 과정에서 얼마나 공정성과 신뢰성을 유지할 수 있을지에 대한 지속적인 감시와 개선도 이루어져야 할 것이다.

3. 네이버와 카카오 팩트체크 정책과 운영 현황

네이버에서는 2017년부터 SNU팩트체크센터가 여러 언론사들과 협력하여 팩트체크 기사 비교 서비스를 제공함으로써 허위정보 확산을 방지해왔다. 그러나 2023년 8월 네이버는 SNU팩트체크센터에 대한 재정 지원을 중단했고 네이버 뉴스의 ‘팩트체크 코너’도 개편했다. 2018년부터 뉴스 홈에 ‘팩트체크’ 메뉴를 운영해왔는데, 2023년 9월 26일부터 뉴스 서비스 개편을 단행하며, 기존 SNU팩트체크센터의 기사 제공 방식을 폐지하고 개별 언론사들이 자체적으로 생성한 팩트체크 기사를 모아볼 수 있는 방식으로 변경했다(김소현, 2023, 9, 26). 팩트체크 페이지에는 언론사에서 공적 관

심사를 자체적으로 검증한 결과를 바탕으로 작성한 기사가 최신순으로 배열된다. 네이버는 이에 대해 각 언론사에서 생성한 팩트체크 기사를 한 곳에서 볼 수 있도록 개편했다고 밝혔지만, 기존 SNU팩트체크센터가 유지했던 엄격한 검증 절차와 표준이 사라지면서 팩트체크 기사의 신뢰도가 낮아질 가능성이 크다는 우려가 제기되고 있다.

또 네이버는 정정보도 모음 페이지의 접근성을 강화하고, 자살 관련 기사 댓글을 자동으로 가려주는 기능을 추가했다. 고침기사, 언론중재법에 따른 정정·반론·추후보도, 인터넷선거보도심의위원회로부터 불공정 선거보도로 경고·주의를 받은 기사를 모아서 제공해 이용자들이 더 쉽게 정정보도 기사를 확인하고 정확한 정보를 얻을 수 있도록 해당 페이지를 강화했다는 설명이다(김소현, 2023, 9, 26). 특히 모바일 메인 MY뉴스판, 언론사 편집판 등에서도 정정보도 모음을 확인할 수 있는 배너를 제공하고, 뉴스 서비스 상단 메뉴를 통해 관련 페이지로 쉽게 이동할 수 있도록 했다. 이러한 네이버의 팩트체크 개편은 정보의 접근성을 높이는 장점이 있을 수 있지만, 독립적인 팩트체크 검증 시스템을 폐지함으로써 허위정보 방지 측면에서는 오히려 퇴보하는 길을 선택한 것으로 평가된다.

카카오는 허위정보 및 딥페이크 콘텐츠 확산 방지를 위해 기술적 조치, AI 기반 기사 표기, 캠페인 진행, 신고 시스템 운영 및 정부 기관과의 협력 등 다각적인 접근 방식을 도입하고 있다고 밝혔다(Kakao, 2024, 3, 13). 먼저 기술적 조치로, AI로 생성된 이미지에 비가시성 워터마크를 도입하여 딥페이크 콘텐츠를 식별할 수 있도록 하고 있다. 이 워터마크를 통해 이미지의 가시적 훼손 없이 AI 생성 여부를 확인할 수 있게 하여 허위정보 유통을 방지한다는 것이다. 두 번째로, 언론사가 AI를 활용해 작성한 기사에는 '자동 생성 알고리즘에 의해 작성된 기사'라는 문구를 표기하도록 했다. 이를 통해 이용자들이 AI 생성 콘텐츠를 명확히 인지할 수 있도록 하여 허위정보 확산을 차단한다는 설명이다.

세 번째로는 다음 카페, 티스토리, 카카오톡스토리, 브런치스토리 등 다양한 플랫폼을 활용하여 건전한 선거 문화 정착 및 딥페이크 근절을 위한 캠페인을 진행함으로써 이용자들에게 허위정보의 위험성을 알리고, 올바른 정보 소비 문화를 확산시키는 데 노력하고 있다고 밝혔다. 네 번째로는 악의적인 딥페이크와 선거 관련 허위 콘텐츠에 대한 신고 시스템을 구축하여, 이용자가 이러한 콘텐츠를 발견하면 즉시 신고할 수 있도록 허위정보 신고

시스템을 운영한다. 이를 통해 허위정보 확산을 신속하게 차단하고 플랫폼 내 정보의 신뢰성을 높이는 환경을 조성한다는 것이다.

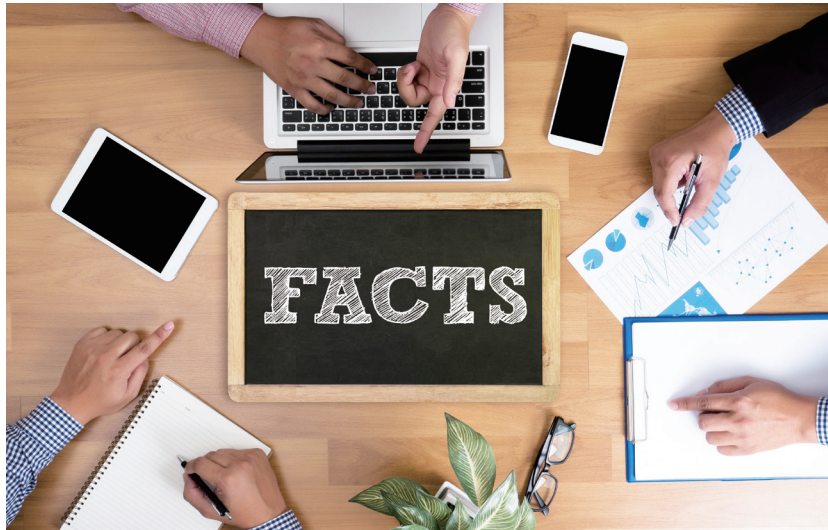
한편, 2017년 대선 정국에서는 대선 특집 페이지(Kakao, 2017, 4, 21)에 ‘가짜 뉴스 바로 알기’ 코너를 마련해 가짜 뉴스 관련 기사, 사례, 판별법, 신고 방법 등을 안내했고, ‘언론사별 팩트체크’ 코너를 통해 언론사가 대선 관련 주요 이슈, 주장, 공약 등에 대해 사실 관계를 분석한 기사를 모아서 제공하기도 했다.

이러한 네이버와 카카오의 팩트체크 현황은 정보 접근성을 높이고 기술적 대응을 강화하는 측면에서 의미가 있지만, 독립적인 검증 체계의 약화와 팩트체크 신뢰성 저하라는 문제를 초래할 우려가 있다. 특히, 네이버의 SNU팩트체크센터 지원 중단은 엄격한 검증 절차와 표준이 사라지는 결과를 낳았고, 카카오의 허위정보 대응 시스템 또한 AI 기반 자동 탐지 기술과 팩트체커와의 협력 강화가 필요하다는 지적이 나온다. 따라서 단순한 정보 제공 방식의 변화가 아닌, 신뢰성과 객관성을 확보할 수 있는 지속적인 보완책 마련이 필요하다.

IV. 플랫폼 팩트체크 발전을 위한 제언

허위정보 확산이 빠르게 이루어지는 디지털 환경에서 플랫폼들의 팩트체크 시스템은 더욱 정교하고 신뢰할 수 있어야 한다. 특히 국내 플랫폼들은 글로벌 기준에 맞춘 기술적·제도적 보완이 필요한 상황이다. 이에 따라 국내 플랫폼을 중심으로 플랫폼 팩트체크의 발전 방향을 몇 가지 제안하고자 한다.

먼저, 독립 팩트체커(IFCN 소속 기관)와의 협력 강화가 필요하다. 현재 유튜브는 IFCN 인증 팩트체크 기관과 협력하여 허위정보를 검증하는 시스템을 운영하고 있다. 그러나 네이버와 카카오 등 국내 플랫폼은 자체적인 검증 시스템을 운영하거나 법적 기준을 중심으로 허위정보를 판별하고 있어 신뢰성이 떨어지거나 대응이 미흡할 수 있다. 특히 네이버는 SNU팩트체크센터와 연계하여 운영하던 팩트체크 시스템을 2023년 중단하며, 개별 언론사들의 자율 검증 체계로 전환했는데, 팩트체크의 신뢰성과 공정성을 유지하기 위해서는 플랫폼이 독립적인 검증기관과 협력하는 것이 바람직하다. IFCN은 전 세계 팩트체커들이 지켜야 할 규범인 △비정파성과 공정성 △취재원 투명성 △재정과 조직의 투명성 △검증방법의 투명성 △공개적이고 정직한 수정 등의 5개의 준칙과 이 준칙의 세부사항을 준수하는 팩트



체크 기관을 인증기관으로 인정하고 있다(정은령, 2023). 네이버와 카카오는 독립적인 팩트체크 기관(IFCN 소속 조직)과 협력하여 검증 시스템의 신뢰성을 높여야 한다.

둘째, AI와 인간 팩트체커의 병행 운영을 제안한다. AI 기술이 발전하면서 허위정보 탐지 기술이 고도화되고 있지만, AI만으로 팩트체크를 전적으로 수행하는 것은 한계가 있다. 페이스북은 SimSearchNet++를 이용해 이미지 기반 허위정보를 탐지하지만(IEEE Spectrum, 2021, 8, 5), 맥락을 이해하지 못하는 경우가 많다. 따라서 AI 기반 자동 탐지 시스템과 인간 팩트체커의 협업이 필요하다. 1차적으로 AI가 허위정보 가능성이 높은 콘텐츠를 선별하고, 2차적으로 IFCN 소속 팩트체크 기관이나 전문가들이 검증하는 이중 시스템을 구축해야 한다.

셋째, 이용자 신고 시스템의 개선 및 처리 속도 향상이 필요하다. 틱톡의 커뮤니티 가이드라인 집행보고서(TikTok, 2021)에 따르면, 2021년 2분기 동안 전 세계적으로 81,518,334개의 동영상상을 커뮤니티 가이드라인 또는 서비스 약관 위반으로 삭제했으며, 이 중 93.0%는 게시된 지 24시간 이내에, 94.1%는 이용자 신고가 있기 전에 삭제했다고 밝히고 있다. 네이버와 카카오는 선거법 위반 댓글만 즉각 삭제하고, 기타 허위정보는 수동 검토를 거쳐 대응하는 방식이라 대응 속도가 느리다. 악의적 신고나 법적 문제, 검증의 어려움 등 다양한 문제가 있겠지만, 신고 후 검증까지의 프로세스를

‘AI+인간 팩트체커’ 협업을 통해 지금보다는 신속하게 처리하고, 신고 결과를 이용자에게 투명하게 공개할 수 있는 방안을 찾기 위해 노력해야 할 것이다.

넷째, 글로벌 협력 확대 및 데이터 공유 체계 구축이 필요하다. 자발적인 참여는 아니지만 틱톡 등 글로벌 플랫폼은 EU DSA(Digital Services Act) 정책에 따라 허위정보에 대한 실천 강령에 참여하고 있다(Korea-EU Research Centre, 2023, 1, 15). 허위정보 확산 문제는 국내에 국한된 것이 아닌 만큼, 국내 플랫폼들도 글로벌 팩트체킹 네트워크와 협력하여 국제 표준에 맞춘 대응 체계를 구축하고, 정보 공유 네트워크에 적극 참여할 필요가 있다.

허위정보 확산은 플랫폼의 노력만으로 해결될 수 없다. 특히 정치적·사회적 영향을 고려할 때, 정부 차원의 제도적 지원과 협력이 필수적이다. 플랫폼들이 독립적이고 효과적인 팩트체크 시스템을 구축할 수 있도록 정부가 보완적인 역할을 수행해야 한다. 먼저, 정치적 압박으로부터 팩트체크 기관을 보호해야 한다. 팩트체크 기관이 특정 정치 세력이나 정부의 영향력을 받게 되면, 검증 자체가 왜곡될 위험이 있다. 팩트체크 기관이 정부나 기업의 영향력에서 자유롭게 활동할 수 있도록 법적·제도적 보호 장치를 마련해야 한다. 또 정부는 플랫폼이 독립적인 팩트체크 기관과 협력할 수 있도록 정책적 유인책을 마련해야 한다. 플랫폼이 IFCN 인증 팩트체크 기관과 협력하도록 인센티브를 제공하고, 팩트체크 데이터를 공유할 수 있는 공공 데이터 허브 구축을 지원하여, 플랫폼 간 협력이 가능하도록 해야 한다. 마지막으로 허위정보 문제를 근본적으로 해결하기 위해서는 이용자들이 허위정보를 스스로 판별할 수 있는 능력을 키우는 것이 중요하다. 정부는 공교육과 성인 교육 과정에서 미디어 리터러시 교육을 강화하고, 시민들이 허위정보를 판별할 수 있도록 지원해야 한다. 허위정보 대응은 팩트체커와 팩트체크 기관만의 문제가 아니라, 정부, 플랫폼, 시민사회가 함께 해결해야 할 사회적 과제이다. 지속 가능한 팩트체크 환경을 조성하기 위해서는 각 주체들의 협력과 제도적 뒷받침이 필수적이다. 더 나아가, 허위정보 문제는 단순히 개별적인 대응을 넘어, 교육, 미디어 리터러시 강화, 정보 유통 구조의 개선 등 장기적인 사회구조적 차원에서 접근해야 하며, 이를 위한 지속적인 연구와 정책적 노력이 병행되어야 한다. 🍵

참고 문헌

- 김소현 (2023, 9, 26). 팩트체크 페이지 리뉴얼·이용자보호 강화...네이버, 26일 뉴스서비스 개편. <ITBizNews>. URL: <https://www.itbiznews.com/news/articleView.html?idxno=112167>
- 김태종 (2024, 5, 31). "지난해 초 이후 Si가 만들어낸 이미지 이용한 가짜정보 급증". <연합뉴스>. URL: <https://www.yna.co.kr/view/AKR2024053001220>
- 김형지, 정은령, 김은미, 양소은, 이재우, 강민지 (2020). 가짜뉴스와 팩트체크 뉴스 노출 집단의 미디어 이용과 뉴스 인식, 그리고 리터러시 관계. <한국언론정보학보>, 제101호, pp232-267.참조
- 김현재 (2024). 생성형 AI가 가짜뉴스의 확산에 미치는 영향 연구. <한국디자인리서치>, 9(2), 제9권 제2호, pp420-429.참조
- 김희정 (2024). 딥페이크 기술을 이용한 신종범죄에 대한 법적책적 시사점: 외국의 법적책 대응을 중심으로. <서강법률논총>, 제13권 제3호, pp97-128.참조
- 박서연·금준경 (2025, 2, 18). 25세 김새론 배우의 죽음, 언론의 자성 그 이상이 필요하다. <미디어오늘>. URL: <https://www.mediatoday.co.kr/news/articleView.html?idxno=324411>
- 박재령 (2024, 8, 19). '정치권 외압' 시달리던 SNU팩트체크 결국 활동 무기한 중단. <미디어오늘>. URL: <https://www.mediatoday.co.kr/news/articleView.html?idxno=320301>
- 신성원 (2023). 딥페이크 기술을 활용한 디지털 성범죄에 관한 연구. <한국치안행정논집>, 제20권 제4호, pp147-168.참조
- 정은령 (2022). 팩트체크란 무엇인가 <올바른 저널리즘 실천을 위한 언론인 안내서: 팩트체크편>(김경희, 정은령, 김수진, 최재원, 송영훈). 한국언론진흥재단
- 정은령 (2023). 팩트체크 글로벌 동향과 SNU팩트체크. <서울대학교 아시아센터 아시아브리프>, 제3권 제44호, pp1-31.참조
- 최현석 (2024, 2, 1). 유튜브, 공신력 있는 의료정보 국내 유통 확대한다. <연합뉴스> URL: <https://www.yna.co.kr/view/AKR20240131155200017>
- Bill Kovach & Tom Rosenstiel (2001). *The Elements of Journalism* (1st ed.). 이재경 (역) (2003). <저널리즘의 기본요소>. 서울: 한국언론진흥재단.
- Bittman, L. (1985). *The KGB and Soviet Disinformation: An Insider's View*, Virginia: Pergamon-Brassey's International Defense Publishers.
- CBS News (2025, 1, 10). What is Community Notes? Twitter/X, Facebook and Instagram's fact-checking tool explained. (CBS News). Retrieved from <https://www.cbsnews.com/news/what-is-community-notes-twitter-x-facebook-instagram>
- Graham, G. (2024, 1, 30). 의료 전문가들이 유튜브를 활용할 수 있는 새로운 방법. <Youtube Official Blog> URL: https://blog.youtube/intl/ko-kr/news-and-events/2024_01_health-corpus-expansion-korea
- Fetzer, J. H. (2004). Information: Does it Have To Be True?. *Minds and Machines*, 14(2), 223-229.
- IEEE Spectrum (2021, 8, 5). How Facebook's AI tools tackle misinformation. <IEEE Spectrum>. Retrieved from <https://spectrum.ieee.org/how-facebooks-ai-tools-tackle-misinformation>
- International Fact-Checking Network (2023). State of fact-checkers 2023. Poynter.
- Kakao (2017, 4, 21). 카카오, 다음 대선 특집 페이지 2차 개편. URL: <https://www.kakaocorp.com/page/detail/8288>
- Kakao (2024, 3, 13). 딥페이크 허위 조작 정보 근절을 위한 카카오의 노력. URL: <https://www.kakaocorp.com/page/detail/10943>
- Kietzmann, J., Lee, L., McCarthy, I., & Kietzmann, T. (2020). Deepfakes: Trick or treat?. *Business Horizons*, 63(2), 135-146.
- Korea-EU Research Centre (2023, 1, 15). EU-틱톡, 데이터 보안·허위 정보·EU 디지털 규제 논의 (1.10). Korea-EU Research Centre. URL: <https://k-erc.eu/2023/02/europe-trends/12868/>
- Lee, S., Xiong, A., Seo, H., & Lee, D. (2023). "Fact-checking" fact checkers: A data-driven approach. *Harvard Kennedy School Misinformation Review*, 4(5). <https://doi.org/10.37016/mr-2020-126>

- Ma, O. & Feldman, B. (2022, 11, 29). How Google and YouTube are investing in fact-checking. Google. Retrieved from <https://blog.google/outreach-initiatives/google-news-initiative/how-google-and-youtube-are-investing-in-fact-checking/>
- Meta (2018, 6, 14). Hard Questions: How Is Facebook's Fact-Checking Program Working?. Meta. Retrieved from <https://about.fb.com/news/2018/06/hard-questions-fact-checking/>
- Meta (2025, 2, 20). Community Notes: A New Way to Add Context to Posts. Meta. Retrieved from <https://transparency.meta.com/features/community-notes>
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175–220.
- Porter, E., & Wood, T. J. (2021). The global effectiveness of fact-checking: Evidence from simultaneous experiments in Argentina, Nigeria, South Africa, and the United Kingdom. *PNAS*, 118(37), e2104235118. <https://doi.org/10.1073/pnas.2104235118>
- Salinas G. (2024, 6, 19). YouTube adds Community Notes feature. 〈Creator Handbook〉. Retrieved from <https://www.creatorhandbook.net/youtube-adds-community-notes-feature/>
- Silverman, C. (2016, 11, 17). This Analysis Shows How Viral Fake Election News Stories Outperformed Real News On Facebook. 〈BuzzFeed News〉. Retrieved from https://www.buzzfeed.com/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook?utm_term=.qdjwmRJR4#.qjBoL7mkK
- Silverman, C. (2016, 12, 30). Here are 50 of the biggest fake news hits on facebook from 2016. 〈BuzzFeed News〉. Retrieved from https://www.buzzfeed.com/craigsilverman/top-fake-news-of-2016?utm_term=.tcRMlxvQ#.gIMVQkmb0
- Sunstein, C. R. (2001). *Republic.com*. Princeton University Press.
- TikTok (2021). 커뮤니티 가이드라인 집행 보고서 2021년 2분기. URL: <https://www.tiktok.com/transparency/ko-kr/community-guidelines-enforcement-2021-2>
- University of Nebraska Medical Center (2023, 8, 15). YouTube announces new policies to target medical misinformation. Retrieved from <https://www.unmc.edu/healthsecurity/transmission/2023/08/15/youtube-announces-new-policies-to-target-medical-misinformation/>
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151.
- Wu, L., Morstatter, F., Carley, K. M., & Liu, H. (2019). Misinformation in social media. *ACM SIGKDD Explorations Newsletter*, 21(2), 80–90.