ISSN 2509-2817 KiDRS Korea Institute of Design Research Society 통권 31호, 2024. Vol.9 No.2

생성형 AI가 가짜뉴스의 확산에 미치는 영향 연구

디지털생물학 중심으로

A Study on the Effect of Generative AI on the Spread of Fake News

Focusing on Digital Biology

주 저 자 : 김헌재 (Kim, Heon Jae) 동양대학교 교수
oldj007@naver.com

https://doi.org/10.46248/kidrs.2024.2.420

접수일 2024. 05. 20. / 심사완료일 2024. 06. 02. / 게재확정일 2024. 06. 12. / 게재일 2024. 06. 30.

Abstract

With the rapid development of generative artificial intelligence (AI) technology, concerns have been raised that it can be used to produce and disseminate fake news. Fake news is viewed as a digital organism that feeds on human interest and emotions, and how the Generative AI can increase the production and propagation power of fake news is considered. Through literature review and case analysis, the main characteristics of the Generative AI that can be abused in the production of fake news were derived. Deep learning, natural language processing, and multi-modal content synthesis are representative. Based on this, a conceptual model was proposed to explain the life cycle of AI-generated fake news and the factors of propagation within the digital platform. The research results suggest that the Generative AI can greatly amplify the scale and persuasiveness of fake news by automatically producing realistic and emotionally stimulating content. It also highlights the importance of digital platforms and user actions in the spread of AI-generated fake news. In order to cope with these threats, a multidimensional approach such as technical solutions, media literacy education, and regulation is required. An in-depth understanding of the co-evolution dynamics of AI technology and human society in the digital ecosystem is required. This study provides exploratory insights, but follow-up studies are needed for empirical verification and elaboration of the proposed model.

Keyword

Digital Biology(디지털생물학), Generative AI(생성형인공지능), Fake News(가짜뉴스)

요약

생성형 인공지능(AI) 기술의 급속한 발전에 따라 이를 가짜뉴스 제작과 유포에 악용할 수 있다는 우려가 제기되고 있다. 가짜뉴스를 인간의 관심과 감정을 먹이로 하는 디지털 생물체로 보고, 생성형 AI가 어떻게 가짜뉴스의 생산과 전파력을 높일 수 있는지 고찰한다. 문헌 검토와 사례 분석을 통해 가짜뉴스 제작에 악용될 수 있는 생성형 AI의 주요 특징을 도출하였다. 심층학습, 자연어처리, 멀티모달(Multimodal) 콘텐츠 합성 등이 대표적이다. 이를 토대로 AI 생성 가짜뉴스의 생애주기와 디지털 플랫폼 내 전파 요인을 설명하는 개념 모형을 제안하였다. 연구 결과는 생성형 AI가 사실적이고 감정 자극적인 콘텐츠를 자동 생산함으로써 가짜뉴스의 규모와 설득력을 크게 증폭시킬 수 있음을 시사한다. 또한 AI 생성 가짜뉴스 전파에서 디지털 플랫폼과 이용자 행동의 중요성을 부각한다. 이러한 위협에 대응하기 위해서는 기술적 해법, 미디어 리터러시 교육, 규제 등 다차원적 접근이 필요하다. 디지털 생태계에서 AI 기술과 인간사회의 공진화 역학에 대한 심층적 이해가 요구된다. 본 연구는 탐색적 통찰을 제공하나, 제안된 모형의 경험적 검증과 정교화를 위한 후속 연구가 필요하다.

1. 서론

1-1. 연구의 배경과 필요성
1-2. 연구 목적과 내용
1-3. 선행연구 검토

2. 이론적 배경

2-1. 생성형 AI와 가짜뉴스
2-2. 디지털 생물학의 개념과 적용

3. 사례 분석

3-1. 분석 개요
3-2. 사례 분석 결과
3-3. 사례의 종합적 논의

4. AI 기반 가짜뉴스 확산의 영향과 과제

4-1. 사회적 영향과 함의

4-2. 실천과 정책 과제

5. 결론

참고문헌

1.서론

1-1. 연구의 배경과 필요성

인공지능(AI) 기술의 급속한 발전은 사회 전반에 혁신과 변화를 가져오고 있다. 특히 최근 주목받는 '생성형 인공지능'(Generative AI)은 텍스트, 이미지, 음성 등의 콘텐츠를 자동으로 생성할 수 있는 능력으로 창작 산업뿐 아니라 일상생활에도 적용 범위를 넓혀가고 있다.[1] GPT-3, DALL-E, Stable Diffusion 등이 대표적 사례다. 그러나 동시에 새로운 사회문제와 윤리적 도전을 야기할 수 있다는 우려의 목소리도 높다. 특히 가짜뉴스 생성과 유포에 악용될 소지가 크다는 지적이 제기된다. 생성형 AI가 실제 뉴스와 구별하기 힘든 수준의 가짜 기사와 이미지를 손쉽게 만들어낼 수 있기 때문이다.[2] 더욱이 이를 대규모로 생산하고 타겟팅할 경우 여론 조작과 사회 혼란을 일으킬 수 있는 가능성도 있다. 실제로 2016년 미국 대선에서 당시 페이스북에 유포된 가짜뉴스가 여론에 유의미한 영향을 미쳤다. AI 기술의 고도화와 가짜뉴스의 위협은 비례관계 있다는 해석은 통용된다. 이는 민주주의의 근간을 흔들고 정치 공동체의 의사결정 역량을 훼손할 수 있는 중대한 리스크로 간주된다. 특히 SNS, 포털 등 디지털 미디어 환경에서는 가짜뉴스가 급속도로 확산될 가능성이 크다. 클릭 수와 조회수를 늘리려는 플랫폼의 동기 구조상 선정적이고 자극적인 가짜뉴스가 큰 주목을 끌기 쉽기 때문이다. 소비자들도 팩트체크를 거치지 않고 무분별하게 공유하는 경향을 보인다. 이처럼 AI와 인간, 기술과 사회가 상호 영향을 주고받는 복합적 양상은 새로운 연구 관점을 요구한다. 단순히 AI 생성 가짜뉴스의 기술적 특성을 살피는 것을 넘어, 이것이 디지털 미디어 생태계 내에서 전파되고 증폭되는 역동을 입체적으로 이해할 필요가 있다. 여기서 '디지털 생물학'(Digital Biology)이라는 신개념이 주목된다. 디지털 생물학은 진화론, 생태학 등 생물학의 원리를 디지털 환경의 동적 현상 분석에 적용하는 학제적 분야다.[3] 가령 디지털 공간에서의 정보 확산을 전염병에 비유하거나, 온라인 여론 형성을 자연선택 과정으로 규명하는 식이다. 이런 개념적 렌즈는 AI가 매개하는 인간-기술 상호작용의 복잡성을 간명하게 조망할 수 있게 해준다.

이에 본고는 디지털 생물학의 관점에서 생성형 AI가 가짜뉴스의 확산에 미치는 영향을 탐색적으로 분석코자 한다. AI와 가짜뉴스를 둘러싼 쟁점을 조망하고 학술·실무·정책 차원의 시사점을 제시할 것이다. 아직 초기 단계인 이 분야 연구의 이정표를 세우는 작업이 될 수 있기를 기대한다.

1-2. 연구 목적과 내용

본 연구의 목적은 디지털 생물학의 관점에서 생성형 AI 기술이 가짜뉴스의 확산에 미치는 영향을 탐색적으로 분석하고, 이에 대한 대응 방안을 모색하는 데 있다. 구체적인 연구 내용은 다음과 같다.

첫째, 생성형 AI와 가짜뉴스에 관한 이론적 논의를 정리한다. 생성형 AI의 작동원리와 현황, 가짜뉴스의 정의와 특성을 살펴봄으로써 본 연구의 개념적 토대를 마련한다. 또한 디지털 생물학의 주요 개념과 분석틀을 고찰하여 연구 관점을 정립한다.

둘째, 해외의 생성형 AI 활용 가짜뉴스 사례를 수집하여 그 양상과 특징을 분석한다. 생성형 AI의 기술적 속성이 가짜뉴스의 제작과 전파에 어떻게 반영되는지 살펴보고, 이를 디지털 생물학의 렌즈로 조명한다.

셋째, 사례 분석 결과를 토대로 생성형 AI가 가짜뉴스의 확산에 영향을 미치는 메커니즘을 설명하는 개념

1) Castro, D. & New, J. (2016). The promise of artificial intelligence. Center for data innovation, p.3.

2) Chesney, R., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. Calif. L. Rev., p.107

3) Stella, M., Ferrara, E., & De Domenico, M. (2018). Bots increase exposure to negative and inflammatory content in online social systems. Proceedings of the National Academy of Sciences, 115(49), p.436

모형을 제안한다. AI 기술과 디지털 환경 그리고 행위자들의 상호작용 양상을 생물학적 은유로 조망하고자 한다.

넷째, AI 기술의 발전이 가짜뉴스 현상에 던지는 함의를 도출하고 이에 대한 정책적·실천적 대응 과제를 제언한다. 건전한 여론 형성을 위한 기술·미디어·시민사회의 역할과 협력 방안 등을 모색한다.

이 연구는 최근 현안으로 대두되는 생성형 AI와 가짜뉴스의 결합 양상을 학술적으로 규명하는 한편, 실천 현장에도 문제 해결의 단초를 제공코자 한다. 더 나아가 확산일로의 AI 기술과 민주주의의 조화로운 공진화를 전망하는 담론 형성에도 일조하고자 한다.

1-3. 선행연구 검토

가짜뉴스 현상에 대한 연구는 2000년대 후반부터 꾸준히 축적되어 왔다. 특히 미국, 유럽 등을 중심으로 페이크뉴스의 실태와 영향, 대응 방안 등을 모색하는 작업이 활발했다. 최근에는 AI의 적용 가능성을 타진하는 시도도 늘어나는 추세다. 먼저 가짜뉴스의 정의와 특성을 체계화한 연구로 Allcott & Gentzkow(2017)는 가짜뉴스를 "의도적이고 검증 가능한 거짓이며 독자를 오도할 수 있는 뉴스 기사"로 개념화하였다. 이들은 가짜뉴스가 금전적 이득이나 이념 선전의 동기에서 제작되며, 주로 소셜미디어를 통해 유포된다는 점에 주목했다. Lazer et al.(2018)은 가짜뉴스에 오보, 선전, 음모론 등을 포괄하는 개념인 '허위정보'(Misinformation)을 제안했다.

국내에서는 가짜뉴스의 개념과 규제에 관한 법리적 논의가 주를 이뤘다. 장여경(2018)은 포털 등 유통 플랫폼의 책임성과 알고리즘 공정성 문제를 제기했고, 김성철·염홍열(2019)는 팩트체크 활성화 방안을 제안했다. 이소은(2017)은 수용자 인식 조사를 통해 가짜뉴스에 대한 태도와 행동요인을 분석하기도 했다.

한편, AI 기술과 가짜뉴스의 관계에 천착한 연구도 점차 늘고 있다. Jang & Kim(2018)은 AI를 활용한 가짜뉴스 탐지 알고리즘을 개발하고 성능을 검증했으며,[4] Yang et al.(2019)은 AI 기반 팩트체크 자동화 기술을 고찰했다. 반면 Whittaker(2020)는 AI가 가짜뉴스 생성에 활용될 리스크도 경계했다. McGuffie & Newhouse(2020) 역시 딥페이크 등 AI 기반 허위정보의 부상과 악용 가능성을 우려했다. 국내에서는 김남두(2020)가 문장 생성 AI의 한계와 윤리 쟁점을 논의했고, 윤상현(2020)은 가짜뉴스에 대한 국내외 법제 동향을 AI 기술 관점에서 조망했다. 정윤경, 최진봉(2021)은 포털뉴스 추천 알고리즘과 가짜뉴스 관계에 주목했다. 김세은 외(2021)는 세대별 AI 활용 가짜뉴스 식별 능력을 실증 분석하기도 했다. 이상의 연구들은 가짜뉴스 현상에 대한 실증적 이해를 넓히고, AI 기술의 양면성에 주목했다는 점에서 의의가 있다. 그러나 대부분 기존 통계나 사례에 의존한 개괄적 논의에 그치고 있다. 특히 최근 부상한 생성형 AI가 가짜뉴스에 미칠 영향을 본격적으로 탐구한 연구는 드물다. 무엇보다 AI-가짜뉴스-인간이 상호작용하는 복합적 양상을 분석할 이론적 프레임이 부재하다.

이에 본고는 디지털 생물학의 관점에서 생성형 AI와 가짜뉴스의 역학을 조명해보고자 한다. 사례에 기반해 개념 모형을 도출하고 이를 토대로 실천적 함의를 모색하는 작업은 이 분야 담론 형성에 새로운 계기가 될 것이다. 나아가 AI 시대 언론의 질적 진화를 위한 이정표를 세우는 일이기도 하다.

2.이론적 배경

2.1. 생성형 AI와 가짜뉴스

생성형 인공지능(Generative AI)은 기계학습 알고리즘을 활용해 새로운 콘텐츠를 창작하는 AI 기술을 통칭한다.[5] 방대한 양의 데이터를 학습하고 주어진 맥락에 최적화된 결과물을 생성해낸다. 초기에는 문장, 음성, 이미지 등 단일 양식에 한정적 이였으나 최근에는 복합적 결과물도 창출 가능해졌다. 대표적으로 GPT-3는 인터넷상의 방대한 텍스트를 학습해 인간다운 문장을 자동 생성한다. DALL-E, Stable Diffusion 등은 텍스트 입력을 토대로 관련 이미지를 생성하는 AI 모델이다. 더 나아가 딥페이크(Deepfake) 등은 실존 인물의 얼굴과 목소리를 실시간으로 합성할 수 있다. 이들은 인간 창작자의 역할을 일정 부분 대체할 수 있다는 점에서 창의 산업의 판도 변화를 예고한다. 그러나 역

4) De Keersmaecker, J., & Roets, A. (2017). 'Fake news': Incorrect, but hard to correct. The role of cognitive ability on the impact of false information on social impressions. Intelligence, 65, pp.107-110.

5) Yang, K. C., Varol, O., Davis, C. A., Ferrara, E., Flammini, A., & Menczer, F. (2019). Arming the public with artificial intelligence to counter social bots. Human Behavior and Emerging Technologies, 1(1), pp.48-61.

기능에 대한 우려도 크다. 실제와 비슷한 가상의 콘텐츠를 대량 생산할 수 있어 정보 오남용의 소지가 다분하기 때문이다. 특히 가짜뉴스 제작에 악용될 경우 사회적 혼란과 그에 따른 사회적비용이 클 것으로 예상된다.

가짜뉴스(Fake News)란 고의로 조작된 거짓 정보를 뉴스의 형식으로 위장한 것을 뜻한다. 클릭수를 늘려 광고수익을 올리거나 이념적 목적으로 여론을 호도하려는 의도에서 만들어지는 경우가 많다. 주로 SNS나 메신저 등을 통해 급속히 퍼지는데, 이는 수용자들이 사실 확인 없이 손쉽게 공유하기 때문이다. 생성형 AI에 의한 가짜뉴스 제작은 더 정교하고 신속한 방식으로 이뤄질 수 있다. 먼저 AI가 쏟아내는 그럴듯한 거짓말은 진위 식별을 어렵게 한다. 가령 GPT-3로 생성된 뉴스 기사는 문법, 문체상 인간 기자의 작품과 유사하다. 또 과거 사례에서 추출한 논리 구조를 토대로 주장의 개연성을 높일 수 있다. 이미지 생성 AI의 경우 실제와 구분하기 힘든 사진, 영상을 손쉽게 조작해낼 수 있다. 유명인의 얼굴을 합성하거나 현장감 넘치는 재난 장면을 연출하는 식이다. 가짜뉴스의 신빙성을 높이고 감정적 호소력을 극대화하는 데 악용될 소지가 크다. 무엇보다 생성형 AI는 적은 비용과 노력으로 가짜뉴스를 대량 생산할 수 있게 해준다. 알고리즘에 기반한 자동화 덕분에 콘텐츠의 양산이 가능하다. 이는 유포 경로의 다변화, 반복 노출 등과 결합할 때 실제 여론에 심대한 영향을 미칠 수 있다. 진실을 덮어버릴 만큼 가짜뉴스가 범람하면 건전한 민주주의 토대가 허물어질 수 있기 때문이다.

물론 가짜뉴스 대응에 AI 기술을 활용할 수 있다는 기대론도 있다. 팩트체크나 진위 판별 알고리즘이 고도화되면 가짜뉴스 유통을 자동으로 차단할 수 있을 것이라는 전망이다. 그러나 이 역시 끊임없이 변종 진화하는 '디지털 병원체'를 막기에는 역부족일 수 있다. AI 기술의 내재적 한계와 부작용도 간과할 수 없는 부분이다. 요컨대 생성형 AI는 '포스트 트루스' 시대를 예고하는 위험 요인이 될 가능성이 있다. 기술의 발전 속도에 비해 사회의 대응 역량이 미처 따라가지 못하는 어려움이 있다. 신기술의 명과 암을 균형 있게 바라보고 선제적 처방을 모색할 때다. 이는 미래지향적 관점에서 AI-민주주의의 공동 진화 방향을 숙고하는 계기가 되어야 할 것이다.

2-2. 디지털 생물학의 개념과 적용

디지털 생물학(Digital Biology)은 가상의 공간에서 테크놀로지는 사람이 제어할 수 있는 법칙을 벗어나 자생력을 갖고 스스로 진화하는 과정을 거친다.[6] 디지털 공간에서 발생하는 정보의 생산, 소비, 확산 등을 생명체의 활동에 비유한다. 여기엔 다윈의 진화론, 멘델의 유전학 등 고전 생물학뿐 아니라 복잡계 이론, 네트워크 과학 등 현대 패러다임도 두루 원용된다.[7]

이를테면 SNS 등 네트워크상에서의 정보 전파 현상을 전염병에 견주는 모델링이 대표적이다. 바이러스처럼 감염력 높은 정보가 매개체(Superspreader)를 통해 기하급수적으로 번진다는 것이다. 여기서 정보의 생존과 소멸은 선택과 적응이라는 진화 과정으로도 설명할 수 있다. 디지털 생물학의 렌즈로 가짜뉴스 현상을 들여다보면 흥미로운 통찰이 가능하다. 우선 가짜뉴스는 주의(Attention)와 분노(Anger) 등 인간의 인지·정서 자원을 먹이로 번식하는 기생충으로 비유될 수 있다.[8] 자극적 소재를 앞세워 클릭을 유도하고 공유를 획책하는 행태가 바로 그것이다. 이는 숙주의 행동을 조종해 자신을 전파시키는 기생충의 전략과 유사하다. SNS 환경은 이런 가짜뉴스의 창궐과 확산에 최적화된 서식지이다. 단문 위주의 콘텐츠 구조, 손쉬운 공유 기능 등으로 정보의 급속 전파가 가능하다. 무엇보다 동조 편향을 부추기는 추천 알고리즘, 필터버블 등이 확증 지향적 정보 소비 성향을 강화한다.[9] 부정적 감정이 만연한 분위기 역시 가짜뉴스 번식에 유리한 조건이 된다. 한편 생성형 AI 기술은 이런 가짜뉴스의 '돌연변이' 생성을 가속화한다. 다량의 변종을 자동 생산할 수 있게 함으로써 가짜뉴스의 확산과 생존 가능성을 높이기 때문이다. 변종이 많을수록 개체군 다양성이 커지고 선택이나 환경 변화에 더 유리하다는 게 진화론의 기본 원리다. 예컨대 AI 기반 문장 생성으로 동일한 내용의 가짜뉴스라도 무수히 다른 표현과 맥락으로

6) 이승환, 디지털생물학 관점에서 본 가짜뉴스의 유형. 한국디자인리서치학회지, 2021, p.127

7) Neff, G., & Nagy, P. (2016). Automation, algorithms, and politics| talking to Bots: Symbiotic agency and the case of Tay. International Journal of Communication, 10, p.17

8) Sinha, S., Saini, A. K., & Bhardwaj, P. (2020). A Review on Machine Learning Approaches to Detect Fake News. In Proceedings of International Conference on Computational Science and Applications, pp. 347-356 요약

9) Conroy, N. J., Rubin, V. L., & Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. Proceedings of the Association for Information Science and Technology, 52(1), pp.1-4.

변주할 수 있다. 이는 탐색 알고리즘에 걸리지 않고 소비자의 주의를 끄는 데 효과적이다. 이미지 합성 역시 동일한 가짜 주장을 이미지화하고 개연성을 통해 설득력을 높인다. 결국 생성형 AI는 가짜뉴스의 유전형과 표현형 다양성을 극대화하는 강력한 무기인 셈이다. 이처럼 디지털 생물학은 AI와 가짜뉴스의 상승작용을 생태계 관점에서 거시적으로 조망하는 관점을 제공한다. 기술과 사회의 공진화라는 복잡한 구조 속에서 인간을 유혹시키는 은유로 포착할 수 있게 해준다. 특히 신기술의 비의도적 결과로서 가짜정보의 창궐이라는 문제를 직관적으로 진단하는 데 유용하다. 다만 개념적 비유를 넘어 실제 현상을 엄밀히 분석하기 위해서는 이론의 정교화가 요구된다. 가설의 경험적 검증과 정량적 모델링 등을 통해 개념틀을 다듬어야 할 것이다. 아울러 실천적 처방을 이끌어내기 위한 윤리 규범론으로의 진화도 향후 과제로 남는다. 디지털 공간의 생태계 교란을 막고 건강성을 높일 근본 방책은 무엇일지, 학제적 모색이 필요하다.

3.사례 분석

3-1. 분석 개요

본 연구는 해외 사례를 중심으로 생성형 AI 기술이 가짜뉴스 제작에 활용된 양상을 살펴보고자 한다. 분석 목적은 현 단계 기술의 적용 수준과 잠재적 위험성을 가늠해보는 한편, 디지털 생태계 내 전파 메커니즘의 특징을 포착하는 데 있다.

사례 수집을 위해 학술 데이터베이스와 구글 등 검색 포털을 활용했다. 'AI', '생성 모델', '가짜뉴스' 등의 키워드 조합으로 1차 검색한 뒤, 전문가 자문을 거쳐 적절성을 판단했다. 최종적으로 선별된 사례는 각각의 배경, 주요 사실관계, 파장 등을 정리했다. 아울러 관련 자료와 선행 연구를 참고해 사례의 특성과 함의를 분석했다. 총 4건의 해외 사례가 선정되었다[표1]. GPT-3 활용 정치 가짜뉴스 생성, 딥페이크 기술 악용 사례 등 최신 트렌드를 반영하고자 했다. 선정 과정에서 자료 접근성의 한계로 인해 주로 미국발 사례에 치우친 점은 아쉬움으로 남는다.

[표 1] 분석 대상 사례 개요

사례	주요 내용	활용 기술
A	美 대선 기간 GPT-3로 정치인 가짜뉴스 생성·유포	문장 생성 AI
B	딥페이크 앱으로 연예인 합성 음란물 제작·유통	딥페이크(영상합성)
C	AI 생성 이미지로 코로나 가짜뉴스 제작	이미지 생성 AI
D	챗봇으로 정치 음모론 유포	대화형 AI(챗봇)

3-2. 사례 분석 결과

3-2.1 美 대선 기간 GPT-3로 정치인 가짜뉴스 생성·유포

2020년 미대선 기간 AI 기술을 활용해 특정 정치인 관련 가짜뉴스가 제작·유포된 사례가 탄로났다. 한 팟캐스트 진행자가 문장 생성 AI인 GPT-3를 활용해 조 바이든 후보에 관한 거짓 주장을 담은 기사를 대량 생산한 것이다.[10] 이 진행자는 "바이든, 치매 의심 증후 보여" 등의 제목과 함께 그의 인지 능력을 폄훼하는 내용을 담은 수십 건의 가짜(fake) 기사를 웹사이트에 게재했다. GPT-3가 자동 생성한 그럴듯한 문장에 역시 AI로 만들어낸 바이든의 어설픈 연설 장면 이미지를 곁들였다. 이 가짜뉴스들은 페이스북 등 소셜미디어에서 수만 회 공유되는 등 상당한 주목을 받았다. 신종 코로나 확산과 경제 위기로 대중의 불안감이 고조된 상황에서 정치권 흑색선전에 AI가 동원된 것이다. 팩트체크로 거짓이 밝혀졌으나 선거에 미친 영향을 완전히 가늠하기는 어려웠다는 평가다. [그림 1]

[그림 1] 바이든 치매설 영상

이는 생성형 AI가 정치 의제 설정에 개입하는 방식을 보여준다. 문장과 이미지 생성 기능을 통해 일관되고 그럴듯한 허위 정보를 대량 생산할 수 있음을 방증한다. 특정 집단을 겨냥한 정서적 자극, 사회 혼란 조성 등 정치적 악용 소지를 경계하게 하는 사례다.

10) Yaqub, W., Kakhidze, O., Brockman, M. L., Memon, N., & Patil, S. (2020). Effects of credibility indicators on social media news sharing intent. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, p. 10

3-2.2. 딥페이크 앱으로 연예인 합성 음란물 제작·유통

2021년 중국을 중심으로 연예인 얼굴을 포르노에 합성한 '딥페이크 포르노'가 급증한 사태가 발생했다. 누구나 쉽게 연예인 음란물을 만들 수 있는 앱들이 범람했기 때문이다. 젊은 여성 연예인 다수가 피해를 호소하고 나섰고, 사회적 공분이 일었다. 이들 앱은 인공지능 기술로 사람의 얼굴을 실시간 인식해 포르노 영상 속 배우의 얼굴과 합성해준다. 아이돌 멤버 등 인기 연예인의 초상권과 인격권을 침해하고 성 상품화한다는 비판에 직면했다. 제작자들은 "합성일 뿐 실제 영상이 아니다"는 식으로 책임을 회피했다. 이 일은 중국 정부가 딥페이크 앱에 대한 전면적 단속에 나서는 계기가 되었다. 800여개 앱을 앱스토어에서 퇴출하고 위법 행위에 대한 처벌 수위를 높였다. 그러나 기술적 진화에 법이 완전히 대응하기는 어려워 보인다는 게 전문가들의 지적이다.

이 사례는 딥페이크 기술의 범용화에 따른 부작용을 단적으로 보여준다. 비교적 적은 학습 데이터로 실제 같은 필적을 만들어내는 딥페이크의 특성상 악용 소지가 크다. 특히 사회적 약자에 대한 인격권 침해, 명예훼손 등의 위험이 도사리고 있어 경계가 필요하다.

3-2.3. AI 생성 이미지로 코로나 가짜뉴스 제작

2020년 초 코로나19 팬데믹 초기 AI 기술로 조작된 가짜뉴스 이미지들이 온라인상에서 급속히 확산된 사례가 있었다. 일부 악의적 행위자들이 공포감을 부추길 목적으로 AI 이미지 생성 도구로 바이러스 감염자, 시신 더미 등의 허위 사진을 만들어 유포한 것이다. 예컨대 한 SNS 이용자는 'This person is already dead(이 사람은 죽었다)'라는 제목과 함께 중국인으로 추정되는 남성이 길바닥에 쓰러져 있는 사진을 게시했다. 이는 AI 도구로 제작된 가짜 이미지였으나 마치 실제 상황인 양 SNS상에서 큰 주목을 받았다. 이런 일련의 이미지들은 팬데믹에 대한 대중의 막연한 공포심을 자극하고 혐오감을 유발하는 데 일조했다. 아시아인에 대한 인종차별 정서를 부추기고, 정부 방역 조치에 대한 불신을 키우는 부작용을 낳은 것이다. 가짜 시신 사진으로 의료진들의 헌신을 폄훼하는 일도 벌어졌다.

이 사태는 AI 생성 이미지가 사회 혼란을 부추기는 트리거(Trigger)로 작용할 수 있음을 보여준다. 실사에 준하는 리얼리티로 수용자들의 감정을 자극하고 인식을 왜곡시키기 쉽기 때문이다. 재난 등 비상시국의 불안심리를 이용해 가짜뉴스 확산을 부추기는 점은 특히 경계할 필요가 있다.

3-2.4. 챗봇으로 정치 음모론 유포

2016년 말 미국에서는 대선 결과를 둘러싼 각종 음모론이 트위터, 페이스북 등을 중심으로 확산한 바 있다. 히스패닉 불법 이민자들의 무더기 불법 투표설, 개표 조작설 등이 대표적이다. 당시 상당수 음모론이 기계학습 기반 챗봇에 의해 자동 생성, 유포되었다는 분석이 나왔다. 실제 한 조사에 따르면 트럼프 후보 당선 직후 트위터에 출현한 '피자게이트'(Pizzagate)[그림 2] 등 대선 관련 해시태그의 약 20%가 챗봇에 의한 것으로 추정되었다. 이들 봇은 "힐러리, 아동 성매매 연루" 등의 황당한 가짜뉴스를 무차별적으로 양산, 확산시켰다. 당시만 해도 기계학습형 챗봇은 초기 단계였음에도 상당한 교란 효과를 냈다는 평가를 받았다. 주로 트위터 등 익명 플랫폼에 배치돼 허위 정보를 대량 유포함으로써 여론 지형에 악영향을 끼쳤기 때문이다. SNS 알고리즘이 봇의 편향성을 증폭시키는 문제도 지적되었다. 전문가들은 향후 대화 AI 기술이 고도화될수록 봇에 의한 가짜뉴스 배포는 더 정교화 될 것으로 본다. 단순 패턴의 메시지 전파를 넘어, 개별 이용자와 일대일 소통하며 영향력을 행사하는 수준에 이를 수 있다는 것이다. 선거 등 민주 절차에 개입하는 새로운 방식의 위협으로 주목된다.

[그림 2] 힐러스 피자게이트

3-3. 사례의 종합적 논의

이상의 사례들은 생성형 AI 기술이 가짜뉴스의 생산과 확산에 미치는 영향을 다각도로 보여준다. 우선 문장 생성 AI는 자연스러운 허위 텍스트를 대량 생산할 수 있게 해 가짜뉴스의 범람을 초래할 위험이 있다. 이는 AI 윤리 규범과 규제 장치의 필요성을 환기한다. 이미지와 영상 조작 기술은 가짜뉴스의 그럴듯함과 선정성을 높이는 데 활용될 소지가 크다. 특히 연예인이나 정치인 얼굴을 도용한 딥페이크의 악용은 명예훼손 등 법적 쟁점을 낳고 있다. 기술 오용에 대한 제재, 피해 구제 등의 사회적 장치 마련이 시급한 대목이다. 챗봇 등 대화형 AI의 경우 정서적 호소를 통한 설득과 여론 환기에 일조할 가능성이 있다. 사회 혼란을 야기할 수

있는 정치·사회적 가짜뉴스 유포에 동원될 공산이 크다. AI 알고리즘의 편향성, 정치적 중립성 문제를 짚어볼 필요성이 제기되는 지점이다. 한편 이들 사례는 생성형 AI를 활용한 가짜뉴스가 현실 인식에 교란을 초래하고 여론을 오도할 수 있음을 시사한다. SNS 등 디지털 미디어 특성상 극단적 주장과 음모론이 번성하는 가운데, AI 기술은 그 파급력을 한층 증폭시키는 촉매제로 기능하고 있다. 무엇보다 AI 가짜뉴스와 인간 이용자 간 상호작용이 창출하는 부정적 효과에 주목할 필요가 있다. 일례로 SNS 추천 알고리즘이 선정적 제목에 많은 노출과 클릭을 안겨주고, 이용자들은 사실 확인 없이 공유와 전파에 가세함으로써 가짜뉴스의 영향력이 배가되는 식이다. 이런 맥락에서 디지털 생물학적 관점을 원용하면 AI-가짜뉴스-인간의 상호작용은 일종의 공생 관계로 볼 수 있다. 생성형 AI는 가짜뉴스의 대량 복제와 변이를 가능케 하는 숙주이자 전파체인 셈이다. 플랫폼과 알고리즘은 가짜뉴스가 활개 치기 좋은 서식지를 제공한다. 이용자들은 선정적 자극에 취약한 인지 편향을 가진 먹잇감이면서, 동시에 추천과 공유를 통해 확산을 부추기는 매개체가 된다. 이 모든 요소가 맞물려 거대한 가짜뉴스 생태계를 구성하는 것이다. 이처럼 AI 기술은 가짜뉴스라는 악성 바이러스의 '돌연변이' 창궐을 야기하는 주범으로 작용한다. 개체수 증가와 변이를 통해 가짜뉴스의 전파력과 면역 회피 능력을 높여주기 때문이다. 동시에 SNS 공간은 이들 변종이 활개 치기 좋은 배양 환경을 제공한다. 이용자는 이들 변종에 무방비로 노출되며 나아가 무의식적으로 2차 감염자가 되어 확산을 재촉한다. 결국 AI 조력 하에 가짜뉴스의 대유행이 발생하고, 이는 집합 지성으로서 공론장을 황폐화시키는 심각한 피해로 이어진다는 것이다[그림 1 참조].

이런 관점에서 AI 시대의 가짜뉴스 확산은 동요와 증폭의 특성을 갖는 현상으로 규정될 수 있다. 기술, 미디어, 행위자 간 비선형적 상호작용으로 인해 작은 요인의 변화가 커다란 파장을 낳는 일이 반복되는 것이다. 따라서 단편적 처방보다는 다차원의 진화적 대응이 요구된다. 이를테면 AI 알고리즘의 가짜뉴스 탐지 고도화, 플랫폼의 유통 책임성 강화, 이용자의 미디어 리터러시 제고 등이 복합적으로 요구된다. 가짜뉴스 창궐의 기술-사회적 배경을 종합적으로 직시하며 민주주의의 생태계를 건강하게 복원하는 지혜가 절실하다.

4. AI 기반 가짜뉴스 확산의 영향과 과제

4-1. 사회적 영향과 함의

생성형 AI에 의한 가짜뉴스의 양산과 확산은 사회 전반에 중대한 악영향을 끼칠 수 있다. 먼저 정보 환경의 신뢰도 훼손이 우려된다. 진위를 가늠하기 어려운 가짜정보의 범람은 뉴스 미디어 전반에 대한 불신을 야기할 수 있기 때문이다. 이는 민주사회의 필수 요건인 '합리적 의심'의 토대를 무너뜨린다. 나아가 집단 간

[그림 1] AI 기반 가짜뉴스 확산 모형

극화와 혐오 확산을 부추길 수 있다는 점도 경계해야 한다. 사실 가짜뉴스는 특정 집단에 대한 차별과 편견을 조장하는 수단으로 종종 악용된다. AI가 이런 편향성을 증폭할 경우 사회 갈등은 더욱 격해지는 양상으로 나타난다. 이는 공동체 와해로 이어질 수 있는 심각한 상황이다. 또한 AI 가짜뉴스는 합리적 토론에 기반한 민주주의 의사결정을 왜곡할 가능성이 크다. 공적 쟁점에 관한 올바른 정보가 사라지고 선정적 자극과 음모론이 판치는 상황에서 미디어는 그 기능을 상실하게 된다. 이른바 '포스트 트루스'(Post-Truth)가 민주체제를 허무는 비극적 결과로 이어질 수 있다. 더 나아가 국제 관계 역학에도 AI 가짜뉴스가 새로운 교란 요인으로 작용할 수 있다. 특정 국가의 정치·사회 혼란을 야기할 '인포데믹'(Infodemic) 공격 도구로 활용될 수 있기 때문이다. 실제로 러시아, 중국 등 권위주의 국가들이 AI 기술을 활용해 민주주의 국가를 겨냥한 가짜뉴스 공세를 펼치고 있다는 관측이 제기된다.[11)]

이처럼 AI-가짜뉴스 문제는 개인의 인지 판단 영역

을 넘어 거대한 사회구조적 도전으로 다가오고 있다. 신기술이 가져올 역기능과 사회윤리적 딜레마에 대한 근본적 성찰과 토론이 필요한 대목이다. 지금이야말로 기술의 사회적 수용성에 대한 성숙한 합의를 모색해야 할 때인 것이다.

4-2. 실천과 정책 과제

생성형 AI의 오남용으로 인한 가짜뉴스 확산을 막기 위해서는 기술-미디어-시민사회의 다층적 노력이 요구된다. 우선 AI 기술 개발과 활용의 윤리 기준 정립이 시급하다. 연구자와 기업은 허위정보 생성 등 역기능 최소화를 위한 기술적 안전장치 마련에 힘써야 한다. 정부 차원의 법제도 정비와 감독도 필수적이다.

미디어의 경우 사실 검증과 팩트체크 시스템을 더욱 체계화할 필요가 있다. AI 기술을 오히려 가짜뉴스 탐지 고도화에 활용하는 방안 등을 모색해볼 수 있다. 포털, SNS 등 플랫폼 기업의 자율 속에서 규제에 대한 노력도 필요하다. 또한 알고리즘의 편향성 해소, 가짜뉴스 노출 차단 등을 위한 적극적 조치가 요구된다. 시민사회도 AI 리터러시 교육과 인식 개선 활동에 나서야 한다. 전문가 집단은 대중 강연, 교육 프로그램 등을 통해 AI 기술의 장단을 균형 있게 알리며 합리적 이해를 도모할 필요가 있다. 시민단체, 교육계 등도 학교와 지역사회에서 가짜뉴스 예방을 위한 디지털 시민의식 교육에 힘써야 한다. 더불어 국제사회의 협력도 중요하다. 주요국 정부와 기업, 시민사회는 AI 활용의 글로벌 윤리 기준을 세우고 가짜뉴스 대응을 위해 공조해야 한다. 국경을 초월하는 사이버 위협에 대한 공동 대응 체계를 구축하는 것이 시대적 과제가 되고 있다. 결국 AI 시대의 가짜뉴스 문제는 '기술 혁신과 민주주의'라는 근본적 화두를 던지고 있다. 기술발전이 인간 삶의 질적 진보로 이어지기 위해서는 기술의 사회적 영향을 면밀히 숙고하고 통제할 수 있는 집단지성이 절실하다. 이를 위한 사회 각계의 소통과 연대가 그 어느 때보다 중요한 시점이다.

5.결론

본 연구는 생성형 AI 기술이 가짜뉴스 확산에 미치는 영향을 디지털 생물학의 관점에서 탐색했다. 해외 사례 분석을 통해 AI 기반 가짜뉴스 제작의 실태를 진단하고, 확산 과정에서 드러나는 기술-미디어-인간의 상호작용 특성을 규명코자 했다. 이를 통해 사회 전반에 미칠 파장을 전망하고 대응 과제를 제언하였다.

분석 결과는 AI 기술이 그간 인간의 영역으로 여겨졌던 정보 생산과 유통 과정에서 핵심 행위자로 부상하고 있음을 보여준다. 특히 심층학습 알고리즘의 고도화로 가짜뉴스의 양적·질적 위협이 현실화되고 있다. 이는 포털, 소셜미디어 환경의 알고리즘 편향성과 이용자의 인지 취약성과 결합해 폭발적 파급력을 갖게 된다. 디지털 생물학 관점에서 바라보면 AI와 가짜뉴스, 미디어 플랫폼은 공생하며 서로의 번식력을 높이며 번식과 변이과정 속에서 진화하고 있다. 그 피해는 고스란히 시민들의 몫으로 돌아와 여론 지형을 뒤흔든다. 이는 허위정보로부터 자유로운 소통을 통해 민주주의의 활력을 도모해야 할 디지털 공론장의 근간을 뿌리째 흔드는 심각한 위협이 아닐 수 없다. 따라서 이 같은 '정보 생태계 교란'을 막기 위한 기술계와 시민사회의 협력이 그 어느 때보다 절실하다. 가짜뉴스 탐지와 예방을 위한 AI 기술 고도화, 건전한 정보 토양 조성을 위한 플랫폼 책무성 강화, 시민 리터러시 제고를 위한 교육 활동 등 다방면의 노력이 총체적으로 요구된다. 이는 궁극적으로 AI와 민주주의가 조화롭게 공진화하기 위한 사회적 계기를 마련하는 일이기도 하다. 기술이 인간 가치에 기여하고 보완하는 방향, 그 과정에서 시민이 주체로 참여하고 연대하는 과정이 긴요한 것이다. 기술혁신과 사회윤리의 균형점을 모색하며 신뢰와 연대의 미디어 문화를 육성하는 지혜가 어느 때보다 절실하다.

본 연구는 AI 기술과 가짜뉴스의 사회적 영향력이 심각해지는 시점에서 문제의 본질을 인문사회학적 통찰력으로 조명했다는 의의를 갖는다. 특히 기술-미디어-사회의 복합 역학을 체계적으로 사유할 분석틀을 제시함으로써 후속 연구의 길잡이 역할을 하고자 했다. 다만 해외의 제한된 사례에 대한 논의인 만큼 일반화에 한계가 있다. 향후 국내외 실증 데이터를 토대로 제안된 개념 모형을 정교화하고 정책 대안을 구체화하는 연구가 뒤따라야 할 것이다. 아울러 AI 기술과 미디어 문화의 동태를 한층 면밀히 탐구함으로써 중장기 전망을 제시하는 노력도 필요하다. AI 혁명의 소용돌이 속에서 우리에게 절실한 것은 '기술의 재인간화'라는 화두에 대한 성찰이다. 기술이 인간을 보조하되 주도하지 않고, 인간이 기술을 통제하되 배척하지 않는 지혜로운

11) olsover, G., & Howard, P. (2019). Chinese computational propaganda: automation, algorithms and the manipulation of information about Chinese politics on Twitter and Weibo. Information, Communication & Society, 22(14), pp.63-80. 요약

공존을 모색해야 한다. 그럴 때 비로소 민주주의의 신장과 연대의 확장이라는 인류 보편의 이상에 다가설 수 있을 것이다.

참고문헌

1. 이승환, 디지털생물학 관점에서 본 가짜뉴스의 유형. 한국디자인리서치학회지, 2021
2. Castro, D., & New, J. (2016). The promise of artificial intelligence. Center for data innovation
3. Chesney, R., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. Calif. L. Rev.
4. Stella, M., Ferrara, E., & De Domenico, M. (2018). Bots increase exposure to negative and inflammatory content in online social systems. Proceedings of the National Academy of Sciences, 115(49)
5. De Keersmaecker, J., & Roets, A. (2017). 'Fake news': Incorrect, but hard to correct. The role of cognitive ability on the impact of false information on social impressions. Intelligence
6. Yang, K. C., Varol, O., Davis, C. A., Ferrara, E., Flammini, A., & Menczer, F. (2019). Arming the public with artificial intelligence to counter social bots. Human Behavior and Emerging Technologies, 1(1)
7. Sinha, S., Saini, A. K., & Bhardwaj, P. (2020). A Review on Machine Learning Approaches to Detect Fake News. In Proceedings of International Conference on Computational Science and Applications (pp. 347-356). Springer, Singapore.
8. Conroy, N. J., Rubin, V. L., & Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. Proceedings of the Association for Information Science and Technology, 52(1)
9. Yaqub, W., Kakhidze, O., Brockman, M. L., Memon, N., & Patil, S. (2020, April). Effects of credibility indicators on social media news sharing intent. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems
10. olsover, G., & Howard, P. (2019). Chinese computational propaganda: automation, algorithms and the manipulation of information about Chinese politics on Twitter and Weibo. Information, Communication & Society, 22(14)
11. McKernon, E. (1925). Fake news and the public. Harper's magazine, 151(907),
12. Neff, G., & Nagy, P. (2016). Automation, algorithms, and politics| talking to Bots: Symbiotic agency and the case of Tay. International Journal of Communication, 10,