DOI | https://doi.org/10.22837/pac.2023.9.2.37

인공지능과 허위조작정보에 관한 연구*

최 종 선**
한양대학교 공공정책대학원 겸임교수
방송통신심의위원회 기획조정실 법무팀 차장

국 문 초 록

본 논문에서는 인공지능 기술을 활용하여 생산·유통하는 허위조작정보와 관련된 법률적 쟁점 및 시사점을 다음과 같이 제언한다. 첫째, 인공지능 및 허위조작정보의 명확한 법적 개념을 마련하지 않고 규제 입법을 하는 경우 어떤 내용의 인공지능 또는 허위조작정보가 규제 대상이 되는지, 어느 범위까지 적법한지 등에 대한 예측가능성을 확보하기 어렵다. 따라서 인공지능 및 허위조작정보의 법적 개념, 보호 대상 등이 불명확하여 이를 규제하기 위한 법률 제정이 수월하지 않은 상황에서는 적용 가능한 현행 법률을 적극적으로 활용하여야 한다. 또한 인공지능 및 허위조작정보에 대한 사례를 분석한 결과를 축적하여 개념 정의, 보호 범위 등을 명확하게 마련할 수 있는 시점에 인공지능을 활용한 허위조작정보를 규제하는 법률을 제정하는 것이 타당하겠다. 둘째, 인공지능 알고리즘이 편향되는 경우에는 정확하지 않은 데이터를 기반으로 하거나 사실을 왜곡하는 정보 등으로 인하여 특정인·특정 집단·특정 의견 등을 무시하거나 차별할 우려가 있다. 특히 이용자가 이러한 알고리즘 편향성을 인식하

* 본 논문은 저자가 소속된 기관의 공식 의견이 아닌 개인적 견해임을 밝힙니다. 세심하고 정성스러운 심사를 해 주신 익명의 심사위원님들께 깊은 감사를 드립니다. Also I would like to thank Professor Jackie Harrison(Head of department) and Dr Irini Katsirea(Reader in International Media Law) for inviting me as Advanced Visiting Fellow at Sheffield University's department of Journalism in the United Kingdom.

** c0j0s0@hanyang.ac.kr, JO4CJ@sheffield.ac.uk, 법학박사

지 못하여 정확한 가치 판단의 기회를 상실할 수 있고, 인공지능이 허위조작정보 확산 수단으로 악용될 우려가 있다. 따라서 이용자들이 편향성 여부를 판단하기 어려운 알고리즘에 대하여 공인된 기관이 마련하여 제공한 알고리즘 검증프로그램을 통하여 언제든지 알고리즘을 투명하게 검증할 수 있도록 하고, 진실성 여부가 논란이 되는 정보에 대해서는 찬성과 반대 과정을 이용자들이 살펴볼 수 있도록 하여야 한다. 이와 함께 논란이 해소되지 않은 정보의 추천서비스를 자제하도록 하여 확증편향의 문제점도 함께 해소할 필요가 있다. 셋째, 허위조작정보를 규제하기 위해서는 해당 정보가 허위라는 사실을 규명하여야 한다. 사법부의 종국적 판단 이전에 행해진 행정부의 허위성 판단이 사법부의 종국적 판단과 상이한 경우에는 인터넷상 언론표현의 자유를 제한하거나 위축시킬 우려가 있다. 따라서 허위조작정보 규제 필요성에 대한 사회적 공감대를 형성하고, 유럽연합과 같이 일률적인 규제보다는 사회적 영향력이 큰 인터넷서비스제공자를 상대로 일정한 의무를 부여하는 등 선별적이고 순차적인 규제 사례를 참고할 필요가 있다. 넷째, 편향된 알고리즘으로 구성된 인공지능을 통하여 허위조작정보를 생산하고 유통함으로써 피해자가 발생한 경우, 고의 또는 과실로 인한 위법행위로 타인에게 손해를 가한 자에 대한 손해를 배상할 책임을 누구에게 물을 것인지 불분명하여 민사책임에 있어서의 행위책임 원칙을 훼손할 우려가 있다. 또한, 편향된 인공지능을 통하여 유통되는 허위조작정보를 예방하기 위하여 해당 정보 삭제 또는 차단 등의 조치를 요구하는 경우 누구를 대상으로 해야 하는지도 불분명하다. 따라서 가해자 특정, 인공지능이 악용되었다는 사실 및 특정 정보의 허위성 등을 피해자가 입증하기란 쉽지 않기 때문에 손해배상 청구를 위한 가해자의 고의 또는 과실 여부, 그로 인한 위법행위로 손해가 발생하였다는 인과관계 등의 입증책임을 가해자에게 전환하여 피해자 보호를 강화하여야 하며, 해당 정보 삭제 또는 접속차단 등의 시정요구 상대방을 명확하게 특정하여야 한다.

주제어: 인공지능, 허위조작정보, 알고리즘, 딥페이크, 여론조작

목 차

Ⅰ. 서론
Ⅱ. 인공지능과 허위조작정보
1. 정의
2. 인공지능 활용 및 허위조작정보 사례
3. 인공지능 및 허위조작정보 관련 법제도
4. 소결
Ⅲ. 법률적 쟁점 및 시사점 제언
1. 법률적 쟁점
2. 인공지능을 통한 허위조작정보 유통 방지를 위한 시사점 제언
Ⅳ. 맺은 말

Ⅰ. 서론

휴대폰 어플리케이션을 이용하여 주택 안의 전자 장비를 제어하거나, 공장 및 농장 등의 설비를 제어하는 스마트 컨트롤 시스템, 컴퓨터 시스템을 통한 자동차 자율주행, 지도나 지름길 안내(navigation), 텍스트나 음성 등을 활용하여 정보를 제공하는 챗봇(chatbot) 등의 서비스는 인간의 생활에 편의를 주는 긍정적인 역할을 한다. 최근에는 궁금한 정보에 대하여 질문을 하면 대화방식으로 수많은 정보를 빠르게 분석하여 필요한 정보를 제공해 주는 ChatGPT[1]가 많이 이용되고 있다. 한편, 특정한 목적을 달성하기 위하여 정보를 왜곡하거나 허위의 사실을 전파하는 데 고도의 인공지능 기술을 활용하는 경우에는 정보의 왜곡성 및 허위성 등을 판단하기 어렵고, 그로 인한 피해 확산 방지와 피해자 구제를 어렵게 하는 등의 부정적인 효과를 유발하기도 한다. 즉, 객관성·투명성·중립성 등을 갖춘 인공지능 기술 또는 진실한 사실 정보 그 자체는 문제가 없으나, 특정한 결과를 유도하거나 얻어내기 위하여 인간을 속이거나 오해를 유발하게 하려는 의도성을 지닌 인공지능 기술 또는 정보는 여

1) https://openai.com/blog/chatgpt (2023년 5월 4일 방문)

론을 호도하고, 국민의 알 권리, 개인의 명예훼손, 개인정보 침해, 사회적 차별 등 법적 권리를 제한[2]하여 국민의 올바른 판단을 어렵게 하여 민주주의의 핵심가치인 다양성 확보에 걸림돌이 될 수 있다.

진실과 거짓을 판별하는 것은 쉽지 않으며, 때로는 많은 시간이 지난 후 거짓이라고 믿었던 것이 진실로 판명되기도 하고, 그 반대의 경우도 발생할 수 있다. 특히 인공지능이라는 고도의 기술을 활용하여 특정한 목적을 위하여 의도적으로 일반인들이 진실과 거짓을 판별하기 어렵게 만드는 경우에는 선의의 피해자를 양산하고 정치적, 성별, 지역별, 국가별 대립 구도를 인위적으로 만드는 등 사회적 혼란을 유발할 수 있고, 인간으로서 스스로 올바르고 정확한 판단을 할 수 없는 환경을 제공하여 민의(民意)의 정확한 전달을 어렵게 한다. 또한, 특정 목적의 거짓된 정보 또는 유도성 정보 등으로 인한 피해자는 스스로 그 정보의 거짓 및 유도 의사 등을 입증한 후 손해배상 또는 해당 정보 삭제 등의 처리를 요구할 수 있으므로, 그 입증이 어렵거나 불가능한 기간 동안 피해구제를 받을 수 없게 되고, 이미 발생한 손해를 완전하게 구제할 수 없게 된다.

따라서 인공지능 기술을 활용하여 허위조작정보를 유통함으로써 발생되는 문제점을 예방하기 위해서는 인공지능 기술 또는 정보가 진실한 사실을 전달하는 것인지, 아니면 특정한 목적성·유도성을 지닌 허위의 사실을 전달하는지 여부를 이용자 스스로 판단할 수 있는 능력을 갖추게 하는 것이 필요하다. 또한 인공지능 및 허위조작정보에 대한 법률적 정의규정, 법적으로 보호받을 수 있는 보호 영역, 관리·감독 기구의 권한, 인공지능 기술을 이용하여 허위조작정보를 생산·유통 시킨 자에 대한 제재수단 등을 명확하게 규정하도록 법률적 토대를 마련하여야 한다.

허위조작정보에 대한 선행연구는 약 128편이 넘을 정도로 방대하다(DBpia 검색 기준). 특히, 신입 지원자들이 자신의 능력을 허위 혹은 과

---

[2] https://www.osce.org/files/f/documents/6/8/302796.pdf (2023년 5월 4일 방문)

장하여 채용담당자들의 검증방법에 대한 논문(구인혁, 2008)이 발표된 후, 2019년부터 2020년까지 “인터넷상 여론조작 실태(강지현, 2019), 가짜뉴스 규제법안(홍완식, 2019), 인터넷상 허위조작행위에 대한 규제방향(박상민, 2019), 가짜뉴스 용어 및 개념정의(김민정, 2019), 허위조작정보 규제에 대한 헌법적 문제점(지성우, 2019) 및 비판적 고찰(최종선, 2020), 가짜뉴스에 대한 규제법적 검토(최승필, 2020)” 등의 연구가 진행되었으며, 2021년에는 코로나19와 관련된 허위조작정보(정영주·홍중윤·박유진, 2021), 허위조작정보 규제의 헌법적 한계(이인호·이준형, 2021)에 관한 연구, 2022년에는 미국(고수윤, 2022), 독일(박용숙, 2022), 프랑스(최지선, 2022), (강명원·심우영, 2022)의 허위조작정보 규제에 대한 연구가 있었다. 선행연구 초창기에는 허위조작정보에 대한 규제를 중심으로 연구가 이루어지다가, 허위조작정보 규제의 헌법적 한계 등의 비판적 고찰 등에 대한 연구가 이루어져 허위조작정보로 인한 규제 중심에서 신중한 규제가 필요하다는 논의로 전환된 것이 특징이며, 최근에는 해외의 허위조작정보 입법사례를 중심으로 연구가 되는 점이 특징이다.

본 논문에서는 기존 선행논문에서 충분한 논의가 이루어지지 않은 인공지능과 허위조작정보의 개념, 인공지능과 허위조작정보의 상관관계 및 악용 사례, 인공지능 기술을 활용하여 생산·유통하는 허위조작정보와 관련된 국내외 법제도를 비교법적으로 살펴보고, 이와 관련된 법률적 쟁점을 살펴보고 관련 시사점을 제언하고자 한다.

Ⅱ. 인공지능과 허위조작정보

1. 정의

인공지능(Artificial Intelligence)이라는 용어는 1956년에 미국 컴퓨터

과학자인 존 매카시(John McCarthy) 교수가 개최한 다트머스 회의(Dartmouth Conference)에서 처음으로 사용되었다. 존 매카시는 인공지능에 대하여 "과학과 기술 영역에서 지능 컴퓨터 프로그램 등 지능형 기계를 만드는 것으로 인간의 지능을 이해하기 위하여 사용되는 컴퓨터와 관련이 있지만, 생물학적으로 식별가능한 방법으로 국한시킬 필요는 없다"고 설명하고 있다.[3] 또한 2017년 11월 벨기에 브뤼셀에서 개최된 유럽연합 가짜뉴스(fake new)와 관련된 회의에서, "뉴스라면 가짜가 아니고, 그것이 허위라면 뉴스가 될 수 없다"는 가짜뉴스(fake news)라는 용어의 모순성이 있으며, 모든 뉴스를 훼손하는 부정적 효과를 우려할 때, 가짜뉴스라는 용어를 사용하지 않는 것이 그 해결의 시작이 될 수 있다는 논의 후[4] 허위조작정보(disinformation)라는 용어를 사용하게 되었다.

인공지능 및 허위조작정보에 대한 법률적 정의는 명확하지 않지만, 표준국어대사전에 따르면 인간의 지능이 가지는 학습, 추리, 적응, 논증 따위의 기능을 갖춘 컴퓨터 시스템을 인공지능(artificial intelligence)으로 이해한다.[5] 한편, 유럽연합에 따르면, 경제적·정치적 등의 이득을 위하여 의도적으로 퍼뜨리는 거짓된 정보 또는 오해를 유발할 수 있는 정보를 허위조작정보(disinformaion)로 이해하고 있다.[6] 특히, 경제협력개발기구(OECD)는 경제적 이익을 위해 의도적으로 개인, 사회집단, 조직 또는 국가를 고의로 속이거나 조작 또는 유해를 가하는 정보를 허위조작정보(disinformation)로 이해하면서 목적성, 의도성 등 유무에 따라 오보정보(misinformation ; 거짓 또는 오해의 소지가 있는 정보가 자신도 모

3) http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html (2023년 5월 4일 방문)

4) https://www.unesco.org/en/articles/solving-fake-news-starts-avoiding-term (2023년 5월 8일 방문)

5) 국립국어원 표준어대사전, https://stdict.korean.go.kr/search/searchResult.do (2023년 5월 4일 방문)

6) https://digital-strategy.ec.europa.eu/en/policies/online-disinformation (2023년 5월 4일 방문)

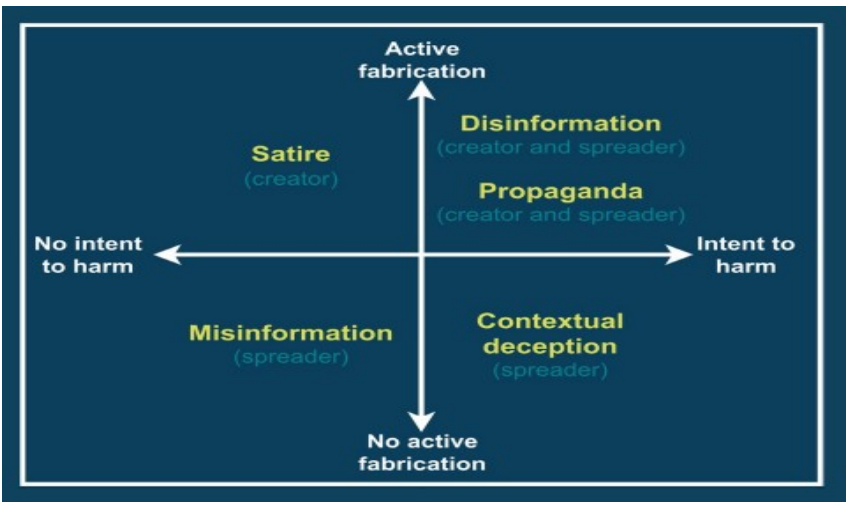
르게 공유되는 정보로 개인, 사회집단, 조직 또는 국가를 의도적으로 속이거나 조작 또는 해를 끼치려는 의도가 없는 것으로, 정보 유포자(spreader)가 초기 오보정보를 생산하거나 조작하지 않는 정보), 문맥상 기만정보(Contextual deception ; 해당 기사와 일치하지 않는 헤드라인 또는 자신의 이야기를 뒷받침하기 위하여 사실을 잘못 전달하는 등 허위조작정보와 달리 정보 자체는 사실이고 조작되지 않았지만 그 정보가 사용되는 방식이 정직하지 않고 사람을 교묘하게 조작하거나 해를 끼칠 의도를 가진 정보), 선전정보(Propaganda ; 정부, 민간기업, 비영리단체 및 개인이 집단적 사고방식, 가치, 견해 등을 채택하고 전파할 목적의 정보), 풍자정보(Satire ; 유머와 과장을 사용하여 사람이나 아이디어를 비판하는 언어, 영화 또는 기타 예술 작품으로 유머와 재치를 사용하여 사회의 이슈에 관심을 끌 수 있으나, 풍자정보가 공유되고 재공유됨에 따라 풍자의 본래 취지가 의도적으로 전파자에 의해 상실되는 경우 시청자들은 본래의 의미를 오해할 수 있다)와 구별하고 있다.[7] 유네스코 “저널리즘, 페이크 뉴스와 허위조작정보 : 저널리즘 교육과 훈련을 위한 핸드북”[8]은 특정인, 사회그룹, 조직 또는 국가에게 해를 가하기 위하여 의도적으로 생산한 거짓된 정보(disinformation), 해를 가할 의도로 생산되지 않은 거짓된 정보(misinformation), 특정인, 사회그룹, 조직 또는 국가에게 해를 가하는 데 사용되는 현실에 기초한 정보(mal-information)으로 구분하고 있다.

---

7) The Committee on Digital Economy Policy (CDEP), Disentangling untruths online : Creators, spreaders and how to stop them, OECD Going Digital Toolkit, 2022 ; https://goingdigital.oecd.org/data/notes/No23_ToolkitNote_UntruthsOnline.pdf (2023년 5월 8일 방문)

8) https://en.unesco.org/fightfakenews (2023년 5월 12일 방문)

〈표 1〉 인터넷상 거짓 정보 유형(A typology of untruths online)[9)]



인공지능 기술 및 구현방식에 대한 완벽한 이해가 부족하고, 정보의 허위성 및 게시자의 조작 의도라는 주관적 구성요건을 어디까지 허용할 것인가에 대한 사회적 합의가 부족하여 인공지능 및 허위조작정보의 범위 설정에 대한 이견(異見)이 법적 개념 정의의 어려움 중 하나라고 생각한다. 따라서 본 논문에서는 일관되고 완벽한 개념 정의는 아니지만 논의의 편의성을 위해서, 인공지능을 "특정한 결과를 도출하기 위해 알고리즘 설계·운영하는 등의 수단을 활용하는 자동화된 컴퓨터 시스템"으로 정의하고, 허위조작정보를 "방송·신문 등 기존 미디어 매체 또는 인터넷 기반의 새로운 유형의 미디어 매체뿐만 아니라 음성·이미지·영상·문자 등 정보 유형을 불문하고 허위의 사실을 진실인 것처럼 포장하는 내용의 정보[10)]"로 정의하기로 한다.

---

9) Molly Lesher, Leveraging AI, big data analytics and people to fight untruths online, 2022 ; https://oecd.ai/en/wonk/untruths-online (2023년 5월 8일 방문)

10) 최종선 (2020). 허위조작정보 규제 시도에 대한 비판적 고찰. 〈홍익법학〉. 제21권 제1호, 367.

2. 인공지능 활용 및 허위조작정보 사례

가. 인공지능

유럽연합은 인공지능에 관한 백서에서 인공지능이 우리 사회에 미칠 수 있는 영향력과 신뢰 구축의 필요성을 감안하여 인간의 존엄성과 개인정보 보호와 같은 기본권에 근본적인 가치를 두어야 함을 강조하면서, “인간중심과 감독(Human agency and oversight), 기술의 견고성과 안전성(Technical robustness and safety), 개인정보 보호와 데이터 관리(Privacy and data governance), 투명성(Transparency), 다양성·차별금지·공정성(Diversity, non-discrimination and fairness), 사회적·환경적 복지(Societal and environmental wellbeing), 책임성(Accountability)” 7가지를 인공지능 윤리지침 필수요소(Guidelines on trustworthy AI in April 2019)로 하고 있다.[11]

(1) 인공지능 시스템

인공지능 시스템이란 머신러닝, 로직(Logic) 및 지식 기반 접근법, 통계산출 등의 방식 중 한 가지 이상을 적용하여 개발된 것(ANNEX I)으로, 인간이 정한 목적을 위해 콘텐츠·예측·추천·환경에 영향을 주는 결정 등의 결과물을 만들어내는 소프트웨어를 의미한다.[12] 인공지능 시스템은 기계 또는 인간이 데이터를 입력하면 현실과 가상의 환경을 판별하고 자동화된 기계학습 분석을 통하여 원하는 결과물을 생성하는 구조

---

11) White Paper, On Artificial Intelligence - A European approach to excellence and trust, 2020 ; https://commission.europa.eu/system/files/2020-02/commission-white-paper-artificial-intelligence-feb2020_en.pdf (2023년 5월 9일 방문)

12) EU Artificial Intelligence Article 3 Definitions ; https://artificialintelligenceact.eu/the-act/ (2023년 5월 30일 방문)

로 시스템 자동화 정도에 따라 매우 다양하다.[13]

(2) 인공지능 활용

인공지능은 농업, 건강, 과학, 교육, 노동, 산업, 사회복지, 조세, 환경, 무역, 운송 등 다양한 분야에서 활용되고 있다.[14] 인공지능은 인간과 경제적 이익을 위한 전략적인 기술로, 의료분야에 있어서는 정밀한 의료진단과 효과적인 질병예방에 도움이 될 수 있으며, 농업 생산성 향상, 기후 변화 대응, 생산성 향상, 예측관리 기술을 통한 생산시스템 효율성 증진 등을 위해 활용되기도 한다.[15]

미디어 분야에서는 언론인들의 업무 효율성을 향상하고 이용자에게 보다 관련성 있는 콘텐츠를 제공하기 위하여 알고리즘(algorithms)과 자동화(automation) 기술을 사용하는 인공지능 저널리즘(AI Journalism), 데이터 또는 경험을 기반으로 자동적으로 학습하고 발전할 수 있는 것을 특징으로 하고 있는 인공지능 응용 프로그램을 뜻하는 기계학습(machine learning[16]) 등이 활용되고 있다.[17]

런던정치경제대학(London School of Economics and Political Science) 미디어 씽크탱크 폴리스(media think-tank Polis)가 인공 지능 및 관련 기술과 관련하여 32개국의 71개 뉴스 조직 116명의 기자들에게 설문 조사한 결과를 기반으로 작성된 Journalism AI 보고서[18]에 따르면, 응답자

13) OECD AI 정책연구소, https://oecd.ai/en/ai-principles (2023년 5월 8일 방문)

14) https://oecd.ai/en/policy-areas (2023년 5월 8일 방문)

15) https://ec.europa.eu/futurium/en/european-ai-alliance/have-your-say-artificial-intelligence-white-paper-european-approach-excellence.html (2023년 5월 9일 방문)

16) National Ai Initiative Act SEC. 5002. [15 U.S.C. 9401] DEFINITIONS. (11) MACHINE LEARNING ; https://www.ai.gov/wp-content/uploads/2023/04/National-Artificial-Intelligence-Initiative-Act-of-2020.pdf (2023년 5월 30일 방문)

17) https://www.eidosmedia.com/blog/tecnnology/AI-in-Journalism (2023년 5월 10일 방문)

18) https://blogs.lse.ac.uk/polis/2019/11/18/new-powers-new-responsibilities/ (2023년 5월 10일 방문)

절반 미만은 뉴스 수집(news gathering) 목적, 3분의 2는 뉴스 생산 목적, 절반 이상은 뉴스 배포를 위하여 인공지능을 이용한다고 답변하였다. 또한 '기자들의 업무 효율성 제고(68%), 이용자에게 보다 적절한 콘텐츠 제공(45%), 업무 효율성 향상(18%)'이 인공지능 사용 동기가 되었다고 하며, 대부분의 답변자는 언론기관이 윤리적이고 편집의 중립성을 보장할 수 있다면 인공지능 저널리즘은 긍정적인 역할을 할 수 있지만, 알고리즘이 편향되는 경우에는 데이터의 부정확성 또는 왜곡으로 특정 사회 집단이나 특정 견해를 가진 자에 대한 차별의 우려가 있고, 거짓된 정보 등을 판별할 수 없는 경우에는 인공지능이 거짓된 정보 확산에 악용될 수 있어 국민들의 편견 또는 갈등을 부추길 수 있는 문제점이 있다고 답변하였다.

나. 허위조작정보 사례

(1) 대선 후보자 비위 오인 사건

2016년 12월 미국 대선 과정에서 힐러리 클린턴 민주당 대선 후보가 위싱턴 인근 피자가게 코메트 핑퐁(Comet Ping Pong) 지하실에서 아동 성매매 조직을 운영한다는 내용이 인스타그램, 페이스북, 트위터 등 사이트를 통하여 확산되자, 이를 직접 확인하기 위해 피자가게를 직접 찾아가 총기를 난사한 사건이다.[19]

(2) 아동납치범 오인 사건

2017년 5월 인도 자르칸드 주(Jharkhand state)에서 아동 유괴 단체가 활동하고 있다는 잘못된 소문(rumours)을 진실로 믿은 사람들이 무고한

19) https://www.nytimes.com/2016/12/05/business/media/comet-ping-pong-pizza-shooting-fake-news-consequences.html (2023년 5월 12일 방문)

시민 6명을 아동납치범으로 오인하여 살해한 사건이 발생하였는데, 그 잘못된 소문이 언제, 어떻게 퍼지게 되었는지는 밝혀지지 않았지만[20], 서양의 국가와 달리 인도의 가짜 뉴스는 대부분 왓츠앱(WhatsApp)과 휴대전화 메시지를 통해 전파되는 점을 고려하면 인터넷 어플리케이션 등 휴대폰으로 접한 오인 정보가 급속도로 확산된 것으로 추정된다.[21]

(3) 유권자 오인 정보 유통 사건

공직선거법상 정보통신망을 이용한 선거운동이 허용[22]되면서(제82조의4) 후보자, 정당 등이 유튜브, 페이스북, 트위터, 인스타그램, 카카오톡 등 뉴미디어를 활용[23]한 선거운동(당선되거나 되게 하거나 되지 못하게 하기 위한 행위)이 활발해지고 있다. 특히, 국가정보원장이 산하 사이버팀 직원들과 공모하여 인터넷 게시글과 댓글 작성, 찬반클릭, 트윗과 리트윗 행위 등의 사이버 활동을 하면서 제18대 대통령선거에 개입한 행위[24]은 공무원의 지위를 활용하여 인터넷 선거운동을 함으로써 특정 후보자에게 유리한 유권자 오인 정보를 전달한 사건이다. 인터넷은 접근성, 전파성 및 비용 측면에서 선거운동비용을 획기적으로 낮출 수 있는 정치공간으로 평가[25]되기도 하지만, 정확하지 않은 정보 또는 딥페이

20) https://www.bbc.com/news/world-asia-india-39971412 (2023년 5월 12일 방문)

21) https://www.bbc.com/news/world-asia-india-40657074 (2023년 5월 12일 방문)

22) 공직선거및선거부정방지법[시행 1997. 11. 14.] [법률 제5412호, 1997. 11. 14., 일부개정] 제82조의3 (컴퓨터통신을 이용한 선거운동) ①선거운동을 할 수 있는 자는 선거운동기간중에 개인용컴퓨터를 이용하여 컴퓨터통신의 게시판·자료실등 정보저장장치에 선거운동을 위한 내용의 정보를 게시하여 선거구민이 열람하게 하거나 대제방·토론실 등에 참여하여 선거운동을 할 수 있다.

23) 인터넷상의 카페 활동과 구별되는 별도의 조직적인 활동으로서 공직선거법상 사조직을 갖춘 것으로 볼 수 있는지 여부는 해당 인터넷 카페의 개설 경위와 시기, 구성원 및 온라인과 오프라인상 활동 내용 등 제반 사정들을 종합하여 판단한다. ; 대법원 2013. 11. 14. 선고 2013도2190 판결

24) 대법원 2018. 4. 19. 선고 2017도14322 판결

25) 헌법재판소 2011. 12. 29. 선고 2007헌마1001 결정 등

크[26] 기술 등을 이용한 왜곡 정보 등이 인터넷에 확산되는 경우에는 유권자의 올바른 판단을 방해하기도 하는 양면성을 가지고 있다.

(가) 한국

1) 대선후보 아들 병역비리 의혹 제기 사건

이회창이 1997. 12. 18.자 제15회 대통령 선거에 출마할 신한국당 후보로 선출된 1997. 7.경부터 국민회의 의원들은 이회창 후보의 장남과 차남이 모두 병무청 징병검사에서 1급 현역판정을 받았다가 입영 시 정밀신체검사에서 체중미달로 5급 면제판정을 받고 귀향처리된 것에 비리가 있다고 언론보도를 통해 제기하였으나, 의혹에 대한 공식적인 조사나 수사가 이루어지지는 않은 채 대통령 선거가 실시되었고, 이회창은 약 2%의 지지율 차이로 낙선하였다.[27]

---

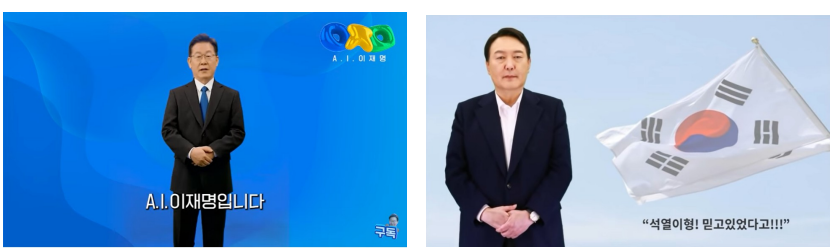
26) 딥페이크(deepfake)는 딥러닝(Deep Learning)과 가짜(Fake)의 합성어로 이미지 또는 영상속의 사람을 다른 사람으로 대체하는 등의 미디어 합성기술로 이해되고 있다. 딥페이크라는 용어는 2017년 레딧(Reddit) 이용자에 의해 처음으로 사용되었으며, 인터넷 뉴스와 집계 사이트에 얼굴 변환 기술을 사용한 포르노 영상을 공유하는 공간을 만들었다. ; https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained (2023년 5월 17일 방문)

27) 그 후 제기된 손해배상 청구사건에서 의혹 보도의 진실성 및 진실이라고 믿을 만한 상당한 이유가 없음을 밝히면서 “이 사건 보도들로써 원고(이회창 후보를 추천하고 선거운동을 했던 정당 ; 한나라당)에 대한 사회적 평가가 크게 훼손되었을 것이 명백하므로 피고들(의혹 제기자 및 이를 언론 보도한 자)은 그 행위로 인하여 원고가 입은 고통에 대하여 위자료를 지급할 의무가 있고, 피고(의혹 제기자)의 경우 자신의 발언 내용이 사회적으로 미친 파급효과가 클 것이라는 점을 예상하였거나 예상할 수 있었음에도 그 내용을 뒷받침할 객관적인 근거도 없이 무책임하게 병역면제비리 의혹과 관련한 제보를 일삼은 점에 비추어 배상할 금액을 1억원, 의혹을 보도한 자들에게는 각각 3천만 원, 2천만 원의 배상 금액을 정함이 상당하다”고 판단하였다. ; 대법원 2005. 4. 29. 선고 2005다13332 판결, 2002. 11. 27. 제16대 대통령 후보 이회창의 아들 이정연이 병역면제를 받았다는 것은 비리라는 취지로 이회창 후보를 반대하는 내용이 포함된 "179센티미터 45킬로그램 인간미이라"라는 제목의 책자를 출간하고, 2002. 12. 3. 한겨레신문에 179cm의 키에 49kg의 몸무게는 불가능하므로 이정연이 위 조건으로 병역면제를 받았다는 것은 비리라는 취지로 이회창 후보를 반대하는 내용이 포함된 광고물을 게재한 사실에 대하여, 적시한 사실이 진실이라는 점에 관한 입증이 부족하므로, 이것이 진실임을 전제로 피고인의 이 사건 책자의 발간과 광고물의 게재 행위에 관하여 위법성이 조각된다는 피고인의 주장은 더 나아가 살펴볼 필요 없이 이유

2) 인공지능 대선후보 선거운동

제20대 대통령 선거에서는 인공지능 기술을 활용한 AI 후보자가 등장하여 가상 후보자를 통한 선거운동이 등장하였다. 인터넷을 활용한 선거운동이 허용되므로 가상 후보자를 활용한 선거운동 자체가 불법은 아니지만(공직선거법 제82조의4제2항) 진실이 아닌 사실이나 유권자에게 오해를 불러올 수 있는 내용으로 선거운동을 하는 경우에는 유권자의 올바른 판단에 방해요소가 된다. 즉, AI 후보자는 실제 후보자의 외모와 목소리로 선거운동을 하고 있어 일반 국민들은 실제 후보자와 가상의 후보자를 구별하기 어렵고, 실제 후보자의 국정철학, 각종 정책기조들과 다른 내용이 전달되더라도 이를 그대로 받아들일 수 있는 우려가 있다.

[그림 1] AI 대통령 후보 이미지



(나) 미국

1) 대통령 선거 과정에서의 오인 정보

2016년 미국 대통령 선거 과정에서 프란치스코 교황이 트럼프 후보를 지지한다는 내용, 클린턴의 건강이 매우 좋지 않다는 내용, 클린턴이 IS를 포함한 이슬람 지하디스트에 무기 판매를 승인했다는 내용 등이 트위터 등 소셜미디어를 통하여 전파된 바 있다. 오하이오 주립 대학의 연구[28]에 의하면, 2012년 오바마 대통령 지지자 중 일부는 위와 같은 사

없다고 판단하였다. ; 대법원 2004. 6. 25. 선고 2004도2062 판결

28) Richard Gunther, Paul A. Beck, Erik C. Nisbet (2018). Fake News May Have Contributed to Trump's 2016 Victory, Ohio State University.

실이 아닌 정보를 진실이라고 믿고 2016년 미국 대통령 선거에서 오바마 대통령과 같은 당의 대통령 후보인 클린턴을 지지하지 않는 등 다수의 유권자들이 위와 같은 사실이 아닌 정보로 인하여 투표에 영향을 받은 것으로 밝혀졌다.

2) 트럼프 체포 이미지 유포 사건

2023년 3월 21일, 영국 온라인 탐사매체 벨링캣(Bellingcat) 설립자인 엘리엇 히긴스(Eliot Higgins)의 트위터에 도널드 트럼프 전 미국 대통령이 경찰에 의해 체포되어 연행되는 사진[29]이 게시된 후, 약 652만회의 조회수를 기록했다. 트럼프 전 대통령이 2016년 대선 직전 성비위 사건과 관련하여 뉴욕 맨하튼지검에 기소될 가능성이 있다는 언론보도 기사를 접한 엘리엇 히긴스는 이를 시각화하기 위해 인공지능 기술(AI art generator)을 이용하여 생성한 체포 이미지가 소셜미디어 이용자들에게 급속도로 확산되었고, 다수의 이용자들은 실제로 트럼프 전 대통령이 체포된 것으로 착각하는 등 논란이 확산되었다.[30]

3) 2020년 대통령 선거 개표 의혹 보도 사건

2020년 대통령 선거에서 바이든 후보가 선출되었음에도 불구하고, 트럼프 전 대통령은 투표 및 개표 집계기가 바이든 후보에게 유리하게 작동되도록 소프트웨어가 설치되었으며, 집계 기기에 결함이 있었다는 등의 주장을 하였고, 이러한 주장은 폭스뉴스 기사로 보도되면서 대통령 선거 결과를 의심하는 등 사회적 혼란을 야기하였다. 2021년 3월 투표 시스템 회사 도미니언은 이런 의혹 보도로 인하여 회사의 명예가 훼손되었다고 주장하며 폭스 뉴스 미디어를 상대로 16억 달러의 손해배상을

29) https://twitter.com/EliotHiggins/status/1637927681734987777 (2023년 5월 23일 방문)

30) https://www.washingtonpost.com/politics/2023/03/22/trump-arrest-deepfakes/ (2023년 5월 23일 방문)

청구하는 민사소송을 제기하였다. 도미니언 투표 시스템 회사(Dominion Voting System Corporation) 등의 원고와 폭스 뉴스 네트워크(Fox News Network) 및 폭스회사(Fox Corporation)의 피고 사이의 명예훼손 소송[31]에 대하여, 미국 델라웨어주 법원(Delaware Courts)이 배심원들과 재판을 시작하기 전인 2023년 4월 18일 양 당사자는 피고가 원고에게 약 7억 8,750만 달러(약 1조 391억원)를 배상하는 조건으로 합의를 하였다.[32][33]

(다) 유럽

1) 독일

2017년 실시된 연방의회 선거 이슈 중 하나가 난민 정책이었으며, 사회민주당(SPD)은 난민 가족입국 완화정책을, 기사당(CSU)은 난민 상한제를, 기민당(CDU)은 난민 가족입국 강화정책을 주장하였다. 기민당(CDU)과 사회민주당(SPD)의 양자토론 과정에서 방송 토론 패널이 사회민주당(SPD) 마르틴 슐츠(Martin Schulz)의 과거 난민에 대해 "난민들이 우리에게 가지고 오는 것은 금보다도 가치가 있는 것이다. 그것은 유럽 드림(dream)에 대한 확고한 믿음이다"라는 연설 내용을 "난민이 금보다 낫다"고 연설한 것으로 왜곡하여 질문하면서 보여준 영상 및 사진이 방송을 통해 전달[34]되면서 유권자들에게 '난민 수용을 무조건적으로 찬성한다'는 견해를 가진 것으로 오해하도록 만들었다.

또한 2017년 독일 공영방송 ZDF 기자들은 거짓된 정보가 많이 유통되고 있는 페이스북에 극우주의 성향의 게시글을 유통시키면 어떤 일이

31) US Dominion, Inc., et al. v. Fox News Network, LLC, et al.; https://courts.delaware.gov/opinions/ (2023년 5월 25일 방문)

32) https://www.foxnews.com/media/fox-news-media-dominion-voting-systems-reach-agreement-over-defamation-lawsuit (2023년 5월 25일 방문)

33) https://www.nytimes.com/video/business/100000008864999/dominion-fox-news.html?searchResultPosition=9 (2023년 5월 25일 방문)

34) https://m.blog.naver.com/PostView.naver?isHttpsRedirect=true&blogId=jongheesalon&logNo=221353750786 (2023년 5월 26일 방문)

발생하는지 살펴보기 위하여 가상의 계정(한스마이어)을 만들고 “난민들을 집으로 들이지 말라, 메르켈 총리의 난민 정책을 비판하면 자녀의 양육권을 박탈하기로 했다” 등의 극우주의 성향의 게시글을 유통시키자 극우 성향의 페이스북 이용자들이 친구 요청을 많이 하였고, 이러한 게시글에 대해 어떠한 반대 의견도 나타나지 않았으나, 실제 통계를 근거로 “난민 수가 증가하였지만 범죄는 크게 늘지 않았다”는 사실을 게시하자 무시하거나 비난하는 등 반응을 보이며 친구 관계를 끊는 것을 살펴볼 수 있었다.[35]

2) 영국

영국 디지털·문화·미디어·스포츠부(the Department of Digital, Culture, Media & Sport) 보고서에 의하면 허위조작정보는 분열·과장·왜곡을 위해 기술을 사용하는 새로운 형태의 전쟁이라고 평가한다. 러시아 투데이(Russia Today ; RT)와 스푸트니크(Sputnik)는 2016년 1월 1일부터 6월 23일까지 실시된 영국의 브렉시트 탈퇴 여부에 대한 국민투표와 관련된 261개의 미디어 사설을 발표하였는데, RT와 스푸트니크는 브렉시트 탈퇴 캠페인 기간 동안 투표 포기 정보보다 유럽연합을 반대하는 정보가 트위터에서의 적극적으로 이용된 것을 확인하였고, 스완지 및 버클리 대학교에서 수행된 연구에서도 브렉시트 탈퇴 캠페인 48시간 동안 러시아 계정에서 브렉시트 관련 트윗 156,252개의 게시물과 45,000개의 브렉시트 관련 메시지가 게시된 것을 확인한 바 있다.[36] 이 연구 결과는 영국 내 중요한 브렉시트 탈퇴 여부에 대한 국민투표에 러시아가 트위터를 통하여 허위조작정보 유통을 하면서 개입하였다는 정황을 보여주었다.

---

35) https://news.kbs.co.kr/news/view.do?ncd=3424373 (2023년 5월 27일 방문)

36) House of Commons Digital, Culture, Media and Sport Committee (2018). Disinformation and fake news_Interim Report, 43 ; https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/363/363.pdf (2023년 5월 28일 방문)

3. 인공지능 및 허위조작정보 관련 법제도

인공지능 기술 자체는 가치중립적이나 특정 목적을 위하여 의도적으로 악용하는 경우에는 현행 법률 위반 우려가 있게 된다. 문제는 인공지능 기술의 특수성으로 인하여 일반 이용자들이 그 악용 여부를 쉽게 판별할 수 없다는 것이다. 특히, 앞서 살펴보았듯이 민주주의의 중심축을 이루는 다양한 의견 개진 및 토론을 통한 사회적 공감대 형성을 하는 과정에서 인공지능을 이용한 허위조작정보의 유통은 이성적이고 정확한 판단을 함에 있어 방해요소가 되어 여론을 호도하는 교묘한 수단이 되고 있기 때문에 이러한 문제점을 예방하기 위한 법제도 마련이 필요하다. 다만, 인공지능 기술에 대한 규제 법률은 신기술 개발 창의성과 속도를 저하시키고, 인공지능을 활용한 허위조작정보 규제 법률은 민주주의 핵심 가치인 다양한 의견 수렴의 수단이 되는 표현의 자유라는 헌법적 가치를 제한하는 요소가 되기 때문에[37] 인공지능 악용 및 이를 이용한 허위조작정보 유통의 문제점 예방을 위한 규제 법률과 신기술 촉진 및 헌법상 기본권을 조화롭게 보호하기 위하여 어떠한 내용으로 균형점을 어디에 두느냐에 대한 구체적인 내용은 각국의 문화적·정치적 여건

37) 인공지능을 통한 허위조작정보를 규제하는 법률이 제 기능을 다하기 위해서는 명확한 규제 대상과 범위가 전제되어야 한다. 만약 이 전제가 충족되지 않는 상황에서 규제 법률이 시행된다면 규제자의 주관적 재량이 개입될 우려가 있고, 피규제자는 허용되지 않는 행위에 대한 예측이 어려워 어떤 표현을 함에 있어 위축될 수 있고(chilling effect) 이로 인하여 표현행위를 하기 전에 자기 검열을 할 우려가 크다. 이러한 위험성은 결국 헌법상 표현의 자유를 제약하게 된다. 또한 모든 허위조작정보가 헌법상 보호대상에서 제외되는 것이 아니라 불법성의 정도, 게시자의 의도 등을 종합적으로 고려하여 보호대상 여부를 판별해야 할 것이다. ; 김현귀 (2016). 표현의 자유와 혐오표현규제. 〈헌법이론과 실무〉, A-3, 70면(결론적으로 혐오표현도 일종의 표현행위인 이상 표현의 자유에 의한 헌법적 보호를 받을 수 있다는 데 의심의 여지가 없다. 아무리 그 내용이 혐오를 조장하거나 표현하는 것이라고 하더라도 그것이 사회적 공론에 아무런 기여가 없을 정도로 순수하게 혐오적인 표현이 아닌 이상에는 자신의 의사를 표현한 것이거나 공적인 관심사에 대한 정치적 표현 또는 학문적·예술적 표현인 경우까지 표현의 자유에 의한 보호를 받을 수 없다고 할 수는 없기 때문이다.)

에 따라 상이하게 나타난다. 따라서 인공지능 허위조작정보와 관련된 각국의 법제도를 살펴보고 바람직하고 필수적으로 필요한 내용이 무엇인지 살펴보도록 한다.

가. 인공지능 관련 법규범

(1) 국내

(가) 인공지능 관련 법안

국회의안정보시스템 제21대(2020-2024) 의안의 제안이유 및 주요내용에 '인공지능'이라는 용어로 검색(2023년 5월 29일 현재)을 하면, 총132건의 의안이 검색된다. 이 법안 중 인공지능 개발 및 활용을 촉진하거나, 인공지능으로 인간의 존엄성 침해, 차별·편견 등의 문제점들을 예방하기 위한 대표적인 법률안은 "[2110148] 인공지능교육진흥법안(안민석의원 등 10인), [2101823] 인공지능 연구개발 및 산업 진흥, 윤리적 책임 등에 관한 법률안(이상민의원 등 11인), [2103515] 인공지능산업 육성에 관한 법률안(양향자의원 등 23인), [2104772] 인공지능 기술 기본법안(민형배의원 등 10인), [2111261] 인공지능 육성 및 신뢰 기반 조성 등에 관한 법률안(정필모의원 등 23인), [2111573] 인공지능에 관한 법률안(이용빈의원등31인), [2113509] 알고리즘 및 인공지능에 관한 법률안(윤영찬의원 등 12인), [2115314] 한국인공지능·반도체공과대학교법안(안민석의원 등 13인), [2118726] 인공지능산업 육성 및 신뢰 확보에 관한 법률안(윤두현의원 등 12인), [2120353] 인공지능책임법안(황희의원 등 14인)" 10개 법안으로 이 중 7개 법안은 인공지능 기술 개발을 촉진하면서 인공지능 악용 등으로 발생할 수 있는 인간의 존엄성, 인권 침해 등의 문제점 예방도 함께 강조하고 있는 점이 특징이다.

(나) 인공지능 기반 미디어 추천 서비스 이용자 보호 기본원칙[38]

인공지능 기반 추천 서비스[39]의 편향적·차별적 정보 제공을 방지하고 이용자를 보호하기 위하여 방송통신위원회가 2021년 6월 30일 추천 서비스 제공자를 위한 자율적 실천규범을 제시하였다. 이 원칙은 인공지능 기반 추천 서비스 내용에 영향을 미치는 요인과 효과에 대한 정보 공개, 결과에 도달하는 과정에 대한 설명을 요하는 투명성, 편향성을 배제하는 공정성, 투명성과 공정성을 위한 사회적 책무성을 강조하고 있다.

(다) 인공지능 개발과 활용에 관한 인권 가이드라인[40]

국가인권위원회는 2022년 4월 11일 과학기술정보통신부장관, 개인정보보호위원회 위원장, 방송통신위원회 위원장, 공정거래위원회 위원장, 금융위원회 위원장에게 인공지능 관련 정책 수립이행 및 관련 법령 제·개정 시 인공지능 개발과 활용에 관한 인권 가이드라인에 기초할 것을 권고하였으며, 인공지능으로 인한 인권침해 및 차별이 발생하지 않도록 적극적인 관리·감독을 권고하였다. 가이드라인의 주요 내용은 “인간의 존엄성 및 개인의 자율성과 다양성을 보호, 인공지능 투명성 확보 및 설명 의무 부과, 최소한의 개인정보 활용, 특정 집단 또는 개인의 권리 등의 차별금지, 인공지능 위험성에 따른 규제 차별화” 등이다.

(2) 국외

(가) 미국

---

38) https://kcc.go.kr/user.do?mode=view&page=A05030000&dc=K00000200&boardId=1113&boardSeq=51454 (2023년 5월 29일)

39) 인공지능 알고리즘을 적용하여 완전히 또는 부분적으로 자동화된 콘텐츠 배열 시스템을 통하여 이용자에게 미디어 콘텐츠를 선별적으로 노출시키는 서비스를 뜻한다.

40) https://www.humanrights.go.kr/site/program/board/basicboard/view?boardtypeid=24&boardid=7607961&menuid=001004002001 (2023년 5월 29일 방문)

1) 트럼프 정부 인공지능 행정명령(American AI Initiative)[41]

2019년 2월 트럼프 대통령은 행정명령(제13859호)으로 미국 인공지능 이니셔티브를 발표면서 인공지능 연구 투자 확대, 인공지능 컴퓨팅 및 데이터 소스 활용, 인공지능 기술 표준 마련 등 5가지 목표를 핵심으로 한다고 밝혔다, 또한 인공지능 규제를 위한 가이드라인[42]을 발표하면서 경제와 국가 안보, 사생활 보호, 시민의 자유 및 표현의 자유·인권·법치주의·저작재산권 보호를 포함한 미국인의 가치를 함께 보호되어야 함을 명확히 하였다.

2) 2020 인공지능 이니셔티브법(The National AI Initiative Act of 2020)

2020 인공지능 이니셔티브법은 대통령으로 하여금 공공과 민간 영역에서 신뢰할 수 있는 인공지능 개발 및 이용 등의 목적을 위해 인공지능 이니셔티브를 설치하도록 하고 있으며, 법 시행일로부터 10년 후에는 그 효력을 다하는 일몰제로 규정되어 있다.[43]

(나) 유럽

1) 유럽연합(EU) : 인공지능의 조화로운 규제를 위한 제안(Proposal for a Regulation laying down harmonised rules on artificial intelligence_ Artificial Intelligence Act[44])

유럽연합 집행위원회는 2021년 4월 21일 인공지능의 조화로운 규칙 제정(인공지능법)을 제안하였다. 2017년 유럽 이사회는 높은 수준의 데이터 보호, 디지털 권리, 윤리 기준 등의 마련을 촉구한 바 있으며,

41) https://trumpwhitehouse.archives.gov/ai/executive-order-ai/ (2023년 5월 30일 방문)

42) Guidance for Regulation of Artificial Intelligence Applications ; https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf (2023년 5월 30일 방문)

43) National Ai Initiative Act SEC. 5101. [15 U.S.C. 9411] NATIONAL ARTIFICIAL INTELLIGENCE INITIATIVE. ; https://www.ai.gov/wp-content/uploads/2023/04/National-Artificial-Intelligence-Initiative-Act-of-2020.pdf (2023년 5월 30일 방문)

44) https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF (2023년 5월 30일 방문)

2019년에는 유럽 시민의 권리가 충분히 존중되어야 함을 강조(2019 Conclusions on th Coordinated Plan on the development and use of artificial intelligence Made in Europe)하기도 하였다. 또한 인공지능의 사회적·경제적 편익에도 불구하고 부정적인 영향 또는 새로운 위험을 유발할 가능성도 있기 때문에 2020년 2월 19일 인공지능의 우수성과 신뢰성을 향한 백서를 발표하면서 인공지능의 보급 확대와 함께 EU 가치, 기본권, 신뢰성, 인간의 존엄성 등의 위험에 대응할 필요성도 강조되었다. 특히 2020년 10월 21일에는 인공지능 시스템의 불투명성(opacity)·복잡성(complexity)·편향성(bias)·예측불가능성(unpredictability)·부분적 자동화(partially autonomous behaviour)에 있어 기본권 및 법률을 준수할 것을 촉구하기도 하였다.

2) 영국 : 인공지능 규제에 대한 혁신적 접근 방법에 관한 정책백서 (Policy paper, A pro-innovation approach to AI regulation)[45]

2023년 3월 29일 영국은 인공지능 규제에 대한 혁신적 접근 방법에 관한 정책백서를 발표하면서, 인공지능 기술로 인한 이점도 있지만 새로운 인공지능 기술로 인하여 새로운 위험 또는 인공지능 기술의 복잡성으로 인하여 국민들에게 불안감을 야기할 수 있고, 사생활 침해 및 인권 침해 우려가 있으며 이러한 위험·편견·차별에 대한 우려가 해소되지 않으면 인공지능에 대한 신뢰가 훼손될 수 있는 점 등의 문제점이 발생할 수 있기 때문에, 이러한 위험에 대응하고 대중의 신뢰를 구축하는 것이 인공지능 규제의 중요한 원동력이 된다고 밝히고 있다. 또한 경제 분야 인공지능 이용과 개발 책임을 알리고 안내하기 위해 "안전성·보안성 및 견고성(Safety, security and robustness), 적절한 투명성과 설명가능성(Appropriate transparency and explainability), 공정성(Fairness), 책임성과 거버넌스(Accountability and governance), 경쟁성과 보상(Contestability

45) https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper (2023년 5월 30일 방문)

and redress)"의 5대 원칙을 혁신프레임워크 토대라고 강조하고 있다.

3) 독일 : 연방정부 인공지능 전략(Federal Government Strategy on Artificial Intelligence)[46]

2018년 7월 18일 독일 연방정부는 인공지능 전략을 위한 주요 내용을 발표하면서, 인간의 존엄성·사생활보호 및 평등의 원칙 등의 유럽인의 기본적 가치를 존중·불법적인 차별금지·인공지능 개발에 있어서의 시민 참여·고도의 보안 수준 유지 등을 강조하였으며 인공지능 개발자와 이용자 사이에 윤리적·법적 한계를 알리고 높은 수준의 법적 확실성을 보장할 수 있는 규제 프레임을 마련하여야 한다고 밝힌 바 있다. 또한 2020년에는 인공지능의 이용으로 경제적·사회적·개인적 이익도 도모하지만 그에 따른 위험을 예방하기 위하여 인공지능의 투명성·추적가능성·적절한 통제수단·검증가능성·신뢰성·보안·견고성·지속가능성 등에 대한 필요성과 함께 알고리즘으로 인한 왜곡·차별·조작 또는 기타 오용에 대한 효과적인 보호 필요성을 강조하였다.[47]

나. 허위조작정보 관련 법규범

(1) 국내

국회의안정보시스템 제21대(2020-2024) 의안의 제안이유 및 주요내용에 '허위조작정보'이라는 용어로 검색(2023년 5월 29일 현재)을 하면, 총 19건의 의안이 검색되며, '가짜뉴스'라는 용어로 검색하면 총 17건의 의안이 검색된다. 허위조작정보와 가짜뉴스라는 용어가 지금까지도 혼용되

46) https://www.bmwk.de/Redaktion/EN/Downloads/E/key-points-for-federal-government-strategy-on-artificial-intelligence.pdf?__blob=publicationFile&v=1 (2023년 5월 31일)

47) https://www.ki-strategie-deutschland.de/files/downloads/Fortschreibung_KI-Strategie_engl.pdf (2023년 5월 31일 방문)

어 사용되고 있는 것은 양자의 명확한 법률적 정의가 마련되지 않은 점, 진실하지 않은 정보가 뉴스라는 미디어 매체를 통하여 미치는 영향력이 큰 점 등에서 기인되는 것으로 보인다. 허위조작정보 관련 법안 중에는 언론등의 고의 또는 중과실로 인한 허위·조작보도에 따라 재산상 손해 등이 있다고 판단되는 경우에 법원이 손해액의 5배를 넘지 않는 범위에서 손해배상액을 정할 수 있도록 하는 등 언론보도등으로 인한 피해구제의 실효성을 높이는 내용의 법안([2112222] 언론중재 및 피해구제 등에 관한 법률 일부개정법률안(대안)-문화체육관광위원장), 인공지능을 기반으로 영상과 음성을 조작해 가짜 영상·음성을 만들어 내는 딥페이크(Deepfake) 정보로 선거에 영향을 미치는 경우 형사처벌을 하는 법안([2114447] 공직선거법 일부개정법률안-민형배의원 등 12인), 정보통신망을 통해 타인의 명예훼손 또는 사생활을 침해하는 경우 징벌적 손해배상을 하도록 하는 법안([2110702] 언론중재 및 피해구제 등에 관한 법률 일부개정법률안-박정의원 등 10인), 미디어의 역기능 대응 등을 목적으로 하는 미디어교육을 강조하는 법안([2103202] 미디어교육 활성화 및 지원에 관한 법률안-정필모의원 등 12인) 등이 대표적이다.

(2) 국외

(가) 미국

미국 수정헌법 제1조는 “연방의회는 언론·출판의 자유 등 국민의 권리를 제한하는 법률을 제정할 수 없다”고 규정하면서 언론·출판의 자유를 두텁게 보호하도록 규정하고 있다.[48] 또한 통신품위법(Communication

48) https://www.senate.gov/civics/constitution_item/constitution.htm#amdt_1_(1791) (2023년 6월 3일 방문) ; Amendment I (1791) Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances.

Decency Act of 1996 ; CDA) 제230조[49]는 전기통신(telecommunication) 중 일반 통신사업자 규제(common carrier regulation)에 관한 조항으로 쌍방향 인터넷 서비스를 제공하는 통신사업자[50]가 “해당 정보가 헌법상 보호되는지 여부와 상관없이 음란 또는 성적 수치심을 야기하거나 폭력적 정보 등”을 선의로 차단 및 조치 등을 취하더라도 민사책임을 면책시켜 주는 근거 규정이다. 이처럼 인터넷상 표현의 자유를 강하게 보호하는 미국에서는 허위조작정보에 대한 문제점을 인식하면서도 이를 규제하기 위한 법률 제정에 대한 사회적 공감대 형성이 필요한 상황이다.

(나) 유럽

1) 유럽연합(EU) : 디지털서비스법(Digital Service Act[51])

불법 콘텐츠, 온라인상의 허위조작정보 또는 기타 사회적 위험에 대처하기 위하여 중개 서비스 제공자의 주의의무를 부과하는 등의 규정 필요성이 제기됨에 따라 제정된 디지털서비스법은 이용자 보호 원칙 등을 포함한 기본적 권리를 효과적으로 보호하고 안전성·예측가능성·신뢰성이 담보된 인터넷 환경을 조성하기 위하여 중개서비스 사업자의 적절한 기능 등에 대해 규율한다. 호스팅 사업자의 경우, 불법적인 행동 또는 불법 콘텐츠 유통 여부 등을 인지하지 못하는 경우 또는 불법 콘텐츠 유통 등을 인지한 경우 해당 콘텐츠를 삭제하거나 접근이 불가능하도록 조치 등을 한 경우에는 법적 책임을 면제하도록 규정(제6조)하고 있다. 또한 온라인 플랫폼 사업자를 포함한 호스팅 서비스 사업자의 경우, 플랫폼 서비스 제공자는 명백한 불법 콘텐츠를 빈번히 제공하는 이용자에게 상당한 기간을 정하여 사전 경고 후 그 이용을 중단할 수 있으며(제

49) https://uscode.house.gov/view.xhtml?req=granuleid:USC-prelim-title47-section230&num=0&edition=prelim (2023년 6월 3일 방문)

50) 면책 규정의 대상이 되는 쌍방향 서비스 제공자(interactive computer service)에는 유튜브, 구글, 페이스북, 트위터 등이 해당된다. ; https://www.supremecourt.gov/qp/21-01333qp.pdf (2023년 6월 3일 방문)

51) https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32022R2065 (2023년 6월 3일 방문)

23조), 월 평균 이용자 수가 4,500만명 이상인 대규모 온라인 플랫폼 서비스 사업자의 경우, 추천시스템(Recommender systems)에 적용되는 옵션 중 하나 이상을 제공해야 한다(제38조).

2) 영국 : 허위조작정보 대응을 위한 데이터 플랫폼
(Counter-Disinformation Data Platform ; CDDP[52])

CDDP는 잠재적 해를 유발하거나 공익을 위협하는 인터넷상 잘못되거나 오해의 소지가 있는 온라인 오보 및 해를 입히기 위해서 의도적으로 작성된 허위조작정보 등에 대한 정부 각 부처의 허위조작정보 대응 능력을 향상시키기 위해 개발된 플랫폼이다. 디지털·문화·미디어·스포츠부(Department of Digital, Culture, Media & Sport ; DCMS)에서 영국 개인정보보호법(UK General Data Protection Regulation ; GDPR)에 따른 개인정보에 대한 관리·감독 역할을 수행하며, 소셜미디어에 공개되어 있는 자료를 분석하여 인터넷상 허위조작정보 유해성에 대응하기 위한 증거기반의 접근법을 활용한다. 이러한 정보 수집의 근거 법률은 영국 GDPR(Article 6(1)(e)) 및 2018년 데이터보호법(Data Protection Act 2018[53] 8(d))으로, 공공기관 직무 수행 및 공익을 위해 데이터 처리가 필요하고, 특히 인터넷상 허위조작정보 처리를 담당하는 정부 부처로서의 기능을 수행하기 위해 이러한 데이터 처리의 필요성이 인정된다.

3) 독일 : 소셜 네트워크에서의 법 집행을 개선하기 위한 법률
(Gesetzes zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken ; NetzDG[54])

---

52) https://www.gov.uk/government/publications/counter-disinformation-data-platform-privacy-notice/counter-disinformation-data-platform-privacy-notice (2023년 6월 5일 방문)

53) https://www.legislation.gov.uk/ukpga/2018/12/section/8/enacted (2023년 6월 5일 방문)

54) https://www.gesetze-im-internet.de/netzdg/BJNR335210017.html (2023년 7월 10일 방문)

2017년 10월 1일 시행된 NetzDG은 콘텐츠를 다른 사용자와 공유하거나 공중이 접근할 수 있도록 하는 영리 목적으로 인터넷에서 플랫폼을 운영하는 텔레미디어 서비스 제공자(소셜 네트워크)에게 적용된다. 이 경우 텔레미디어 서비스 제공자가 책임지는 저널리즘 및 편집 디자인 등이 포함된 경우 이 법상 소셜 네트워크로 간주되지 않는다(제1조). 소셜 네트워크 서비스 제공 사업자에게 불법 콘텐츠에 대한 효과적이고 투명한 처리절차를 유지하도록 하는 의무를 부여하는 것으로, 민원이 접수된 후 24시간 이내에 해당 정보의 불법성 여부를 판단하고 불법성이 인정되는 경우 즉시 삭제 또는 접속을 차단하는 조치를 취하여야 한다(제3조). 불법 콘텐츠는 독일 형법 "국가·인종·종교 단체 등에 대한 증오를 선동하는 자, 단체나 개인을 모욕하거나 악의적으로 비방하거나 명예를 훼손하여 인간의 존엄성을 침해하는 등의 방법으로 공공의 평화를 훼손하는 행위(Section 130 Incitement of masses)", "공개적으로 모욕하는 행위(Section 185 Insult)", "타인을 비하하거나 여론에 부정적인 영향을 미치거나 신용에 해를 가하는 행위(Section 187 Defamation)", "성적 자기결정권, 신체적 존엄성 등에 반하는 불법행위를 하는 행위 등(Section 241 Threatening commission of serious criminal offence)" 등의 정보를 의미한다(제1조제3항[55]). 즉, NetzDG은 "공공연한 범죄선동, 국민선동, 악의적 비방, 명예훼손 등" 독일 형법에 따른 불법행위를 신속하게 규제[56]하기 위해 소셜 네트워크 서비스 제공 사업자에게 24시간 이내 처

---

55) (3) Rechtswidrige Inhalte sind Inhalte im Sinne des Absatzes 1, die den Tatbestand der § § 86, 86a, 89a, 91, 100a, 111, 126, 129 bis 129b, 130, 131, 140, 166, 184b, 185 bis 187, 189, 201a, 241 oder 269 des Strafgesetzbuchs erfüllen und nicht gerechtfertigt sind. ; German Criminal Code Section 1 Scope. Unlawful content shall be content within the meaning of subsection (1) which fulfils the requirements of the offences described in sections 86, 86a, 89a, 91, 100a, 111, 126, 129 to 129b, 130, 131, 140, 166, 184b in connection with 184d, 185 to 187, 241 or 269 of the Criminal Code and which is not justified.

56) 독일 형법상의 특정 범죄에 해당하는 표현물이 네트워크집행법에 적용을 받게 되는 것이다. ; 박용숙 (2022). 허위조작정보의 규제방안에 대한 시론적 고찰 – 독일의 네트워크집행법을 중심으로 –. 〈법이론실무연구〉, 제10권 제4호, 152.

리의무 등을 부여하고 있는 법률로, 표현 내용의 허위성을 직접 규제하는 것이 아니라 규제되는 불법정보에 허위조작정보가 포함될 수 있어[57] 허위조작정보 관련 법규범으로 소개하는 것이다.

4. 소결

우리나라의 경우 인공지능 기술 개발을 촉진하면서 인공지능 악용 등으로 발생할 수 있는 인간의 존엄성, 인권 침해 등의 문제점 예방 및 인공지능의 투명성, 공정성, 책무성도 함께 강조하고 있는 점이 특징이다. 미국의 경우에는 인공지능 활성화와 함께 사생활 보호, 시민의 자유 및 표현의 자유 등의 기본적 권리도 함께 보호해야 함을 강조하고 있으며, 유럽연합의 경우에도 인공지능의 부정적인 영향 또는 새로운 위험에 대한 대응을 위해 신뢰성, 인간의 존엄성 등 시민의 권리가 충분히 보호되어야 하며, 인공지능 시스템의 불투명성·복잡성·편향성·예측불가능성 등에 있어 기본권 및 법률을 준수해야 함을 강조하고 있다. 또한 영국 및 독일의 경우에는 정책백서 또는 인공지능 전략을 발표하면서 사생활 침해, 인권 침해 및 위험·편견·차별의 문제를 해소하여 인공지능에 대한 신뢰성 확보 및 인공지능의 투명성·추적가능성·적절한 통제수단·검증가능성·신뢰성·보안·견고성·지속가능성 등에 대한 필요성과 함께 알고리즘으로 인한 왜곡·차별·조작 또는 기타 오용에 대한 효과적인 보호 필요성을 강조하였다.

한편, 우리나라의 경우 허위조작정보 유통으로 인한 피해를 예방하기 위하여 다수의 법안이 상정되어 있으나, 허위조작정보의 명확한 법적 개념의 어려움, 피해자 구제를 위한 5배 손해배상제도 도입에 대한 언론표

57) 독일의 네트워크집행법의 대상이 되는 게시물은 모든‘가짜뉴스’가 아니라 ‘혐오표현’(hate speech)85)에 국한되기 때문이다. ; 지성우 (2019). 허위조작정보(소위 ‘가짜뉴스’) 규제에 대한 헌법적 문제점에 관한 연구. 〈공법연구〉, 제48집 제2호, 182.

현의 자유 침해 논란 등으로 허위조작정보 규제 입법에 대한 사회적 공감대 형성이 부족한 상황이다. 수정헌법 제1조에 따라 표현의 자유를 강하게 보호하는 미국의 경우에도 허위조작정보에 대한 문제점을 인식하면서도 이를 규제하기 위한 법률 제정에 대한 사회적 공감대 형성이 필요한 상황이며, 유럽연합의 경우에는 디지털서비스법을 제정하여 허위조작정보의 위험에 대처하기 위하여 호스팅 사업자에게 불법 콘텐츠 유통 여부 인지 시 해당 콘텐츠 삭제 또는 접근 조치 제한 의무와 함께 일정한 요건하에 법적 책임을 면하는 규정을 두고 있고, 온라인 플랫폼 사업자에게는 반복적으로 불법 콘텐츠를 제공하는 이용자에게 상당한 기간을 정하여 사전 경고 후 그 이용을 중단하도록 하고 있다. 독일의 경우에는 소셜네트워크 사업자에게 공공의 평화 훼손, 명예훼손, 모욕죄 등에 해당하는 정보에 대한 삭제 또는 접속차단 등의 의무를 부여하고 있다. 한편 영국의 경우에는 허위조작정보 등에 대한 정부 각 부처의 대응 능력 향상을 위해 소셜미디어에 공개되어 있는 자료를 분석하여 허위조작정보 유해성에 대응하기 위한 증거기반 접근법을 활용하는 데이터 플랫폼을 개발한 것이 특징이다.

Ⅲ. 법률적 쟁점 및 시사점 제언

1. 법률적 쟁점

가. 인공지능 및 허위조작정보 개념의 모호성

법제도를 통하여 특정 분야를 진흥시키거나 규제를 하기 위해서는 그 대상이 되는 개념을 명확하게 정의하여야 한다. 대상이 불분명하거나 특정하기 어려운 경우에는 법제도 내용에 따른 집행보다는 집행자의 재량

에 따라 각기 다른 결론이 도출될 수 있다. 인공지능 및 허위조작정보의 법적 개념은 살펴본 바와 같이 명확한 정의가 정립되지 않은 상황으로, 이를 규제하기 위한 법제도를 마련할 경우 피규제자 입장에서는 어떤 내용의 인공지능 또는 허위조작정보가 규제 대상이 되는지, 어느 범위까지 적법한지 등에 대한 예측가능성 확보를 어렵게 하고, 규제자 입장에서는 상황에 따라 또는 각자의 가치관 등 주관적 영향에 따라 규제 대상 및 내용 등이 상이할 수 있게 되는 등의 문제가 발생할 우려가 크다. 유럽연합이 인공지능의 영향력과 신뢰성의 확보가 필요하기 때문에 인간의 존엄성과 개인정보 보호와 같은 기본권 보호에 근본적인 가치를 두어야 한다고 주장하면서 "인간중심과 감독(Human agency and oversight), 기술의 견고성과 안전성(Technical robustness and safety), 개인정보 보호와 데이터 관리(Privacy and data governance), 투명성(Transparency), 다양성·차별금지·공정성(Diversity, non-discrimination and fairness), 사회적·환경적 복지(Societal and environmental wellbeing), 책임성(Accountability)" 7가지를 인공지능 윤리지침 필수요소를 강조하고 있지만 여전히 인공지능의 법적 개념, 규제의 대상 등의 불명확성으로 인한 피규제자 예측가능성 확보가 불충분하다. 허위조작정보의 경우에도 애초 뉴스 정보에서 거짓된 정보가 유통되면서 가짜뉴스라는 이름으로 논의되다가 거짓된 정보는 뉴스라는 것에 한정되지 않는다는 의견에 따라 의도적으로 거짓된 정보를 유통시키는 경우와 의도하지는 않았지만 잘못된 정보를 유통시키는 경우 등 목적성, 의도성 여부에 따라 오보정보(misinformation), 문맥상 기만정보(contextual deception) 등으로 세분화되는 등 법적으로 명확한 개념 정의가 충분하지 않은 상황이다.

나. 편향된 알고리즘을 통한 허위조작정보

특정 목적을 위하여 의도적으로 인공지능을 악용하더라도 이용자들은 그러한 목적성 및 의도성을 쉽게 알 수 없기 때문에 인공지능 알고리즘

설계자 및 작동 결과 등에 따라 인위적으로 가공된 정보를 접하게 되고, 이를 반복적으로 이용하는 경우에는 다양한 의견 접근 기회를 상실하고 편향된 사고를 하는 등 인공지능 알고리즘 편향성 문제가 발생할 수 있다. 인공지능을 작동시키는 구심점이 되는 알고리즘을 미디어 매체와 결합하여 악용하는 경우에는 미디어 수용자들로 하여금 매우 제한되고 수동적인 입장에서 목적성·의도성을 지닌 정보를 수용하게 되어 민주주의의 핵심가치인 다양성 확보에 큰 영향을 줄 수 있다. 즉, 인공지능 알고리즘이 편향되는 경우에는 정확하지 않은 데이터를 기반으로 하거나 사실을 왜곡하는 정보 등으로 인하여 특정인, 특정 집단, 특정 의견 등을 무시하거나 차별할 우려가 있으며, 이용자가 이러한 알고리즘 편향성을 인식하지 못하는 경우에는 정확한 가치 판단의 기회를 상실하고 허위조작정보 확산 수단으로 인공지능이 악용될 우려가 있다.

다. 진실과 거짓 판단의 어려움

허위조작정보의 법적 개념은 명확하지 않지만 진실에 반하는 허위의 사실 여부 판단이 전제되어야 인공지능을 이용한 허위조작정보 유통의 문제점을 예방할 수 있다. 어떠한 사실이 진실인지 허위인지 판단하는 문제는 간단하지 않으며, 특히 사회적으로 찬반이 팽배한 이슈에 대한 진실성 여부 판단은 쉽지 않다. 2023년 사이버범죄 트렌드[58]에 의하면, “인터넷을 이용한 허위조작정보 등으로 인한 피해사건이 증가하면서 사이버 명예훼손 및 모욕죄와 관련된 불법콘텐츠 범죄가 2021년 28,988건, 2022년 29,258건으로 증가한 것으로 평가”하고 있다. 사실 적시 명예훼손 범죄의 경우 진실한 사실로서 오로지 공공의 이익에 관한 경우에는 위법성을 조각하고 있는데, 대법원은 “진실한 사실이란 내용 전체의 취지를 살펴볼 때 중요한 부분이 객관적 사실과 합치되는 사실이라는 의

58) https://www.police.go.kr/user/bbs/BD_selectBbs.do?q_bbsCode=1001&q_bbscttSn=20230317103332394 (2023년 6월 7일 방문)

미로 세부에서 진실과 약간 차이가 나거나 다소 과장된 표현이 있더라도 무방하다"는 입장[59]이다. 이처럼 진실한 사실 여부는 고도의 법률적 지식을 가진 판사가 소송의 양당사자의 법정공방을 통하여 전체 문맥의 취지를 고려하여 객관적 사실과의 합치 여부를 판단하더라도 법적 쟁점이 된 진실성 여부는 재심절차 등을 통하여 번복될 여지가 있기 때문에, 법원 판결을 통하여 소송 당사자 간의 진실 여부에 대한 소송 결과가 나오더라도 사법부의 판결을 존중하는 측면에서 판결에 승복을 하는 것과 법적 분쟁의 대상이 되었던 사실에 대한 진실성 여부에 대한 판단결과가 언제나 진실하느냐 하는 문제는 구별하여야 한다. 즉, 허위조작정보 규제에 앞서 진실과 거짓 여부의 판단이 전제가 되어야 하는데, 허위조작정보의 진실성 여부 판단의 주체를 법률을 집행하는 행정부로 할 경우 사법부의 종국적인 판단에 앞서 허위조작정보 여부를 판단하고 인터넷상 유통 여부를 규제하게 되는 결과를 초래[60]하므로 인터넷상 언론표현의 자유를 제한하거나 위축시킬 우려가 있다.

라. 법적 책임 주체

편향된 알고리즘으로 구성된 인공지능을 통하여 허위조작정보를 생산하고 유통함으로써 피해자가 발생한 경우, 그 피해 배상의 책임을 편향된 알고리즘 설계자에게 물을 것인지, 아니면 편향된 인공지능을 이용하여 허위조작정보를 생산하고 유통한 자에게 물을 것인지 즉, 고의 또는 과실로 인한 위법행위로 타인에게 손해를 가한 자에 대한 손해를 배상할 책임을 누구에게 물을 것인지 불분명하여 민사 책임에 있어서의 행위책임 원칙을 훼손할 우려가 있다. 또한 편향된 인공지능을 통하여 유

59) 대법원 2023. 2. 2. 선고 2022도13425 판결

60) 인터넷 내용규제에 있어 자율규제가 정부의 법적 규제보다 나은 대안으로 간주되고 있다는 견해도 있다. ; 김민정 (2020). 유럽연합의 온라인 허위조작정보(disinformation) 자율규제 사례 고찰. 〈언론과 법〉, 제19권 제1호, 196.

통되는 허위조작정보를 예방하기 위하여 해당 정보 삭제 또는 차단 등의 조치를 요구하는 경우 누구를 대상으로 해야 하는지도 불분명한 문제점이 있다.

2. 인공지능을 통한 허위조작정보 유통 방지를 위한 시사점 제언

가. 인공지능 및 허위조작정보 악용 사례 분석을 통한 단계적 접근

인공지능 기술을 악용하는 사례는 대표적으로 딥페이크(deepfake)를 활용하여 특정 사람의 모습, 목소리 등을 조작하는 것으로 실제 당사자의 초상권, 명예훼손 등의 법익을 침해하는 동시에 이용자로 하여금 사실관계를 혼동케 하는 것이다. 허위조작정보 악용의 대표적 사례는 인공지능 기술을 활용하여 생성한 허위조작정보를 유튜브·페이스북·트위터·인스타그램·카카오톡 등 뉴미디어를 통하여 유통시킴으로써 선거과정에서 상대후보를 비방하거나, 선거 이슈에 대한 입장을 왜곡하는 등의 방법으로 유권자의 그릇된 판단을 유발하여 선거 결과에 영향을 주는 것이다. 인공지능의 악용사례 및 허위조작정보 사례는 우리가 예상하지 못하는 영역에서 얼마든지 발생할 수 있기 때문에 인공지능 및 허위조작정보 법적 정의가 명확하지 않은 상황에서 전체를 아우르는 규제 법률을 마련하기 보다는[61] 현재 문제점으로 발생하는 사례 위주로 분석하여 공통의 분모가 무엇인지 살펴보는 것이 중요하다. 또한 이미 발생한 인공지능을 악용한 허위조작정보는 형법의 허위사실적시 명예훼손, 모욕죄, 신용훼손, 사기죄 등, 공직선거법상 정보통신망을 이용한 선거운동 위반죄, 성폭력범죄의 처벌 등에 관한 특례법상 허위영상물 등의 반포

61) 기존의 법률과 사회관계망 서비스의 자율규제만으로는 가짜뉴스에 대응하기 어렵다는 결론이 나오면, 그때 입법에 착수해도 늦지 않을 것이다. ; 안수길 (2019). 가짜뉴스에 맞서는 독일 사회관계망법집행법의 내용과 쟁점. 〈법학논총〉, 제36권 제1호, 134-135.

죄, 민법상 손해배상책임 등 현행 법률 위반에 따른 법적 책임을 묻는 것이 타당하겠다. 즉, 인공지능 및 허위조작정보의 법적 개념, 보호 대상 등이 불명확하여 이를 규제하기 위한 법률 제정이 수월하지 않은 상황에서는 현행 법률을 적용할 수 있는 경우 이를 적극적으로 활용하고, 인공지능 및 허위조작정보에 대한 사례 분석 결과가 축적되어 개념 정의, 보호 범위 등을 명확하게 마련할 수 있는 시점에 인공지능을 활용한 허위조작정보를 규제하는 법률을 제정하는 것이 타당하겠다.

〈표 2〉 인공지능을 악용한 허위조작정보 : 현행법상 적용 가능한 법률

구분	현행 법률	비고
형법	제307조(명예훼손) ②공연히 허위의 사실을 적시하여 사람의 명예를 훼손한 자는 5년 이하의 징역, 10년 이하의 자격정지 또는 1천만원 이하의 벌금에 처한다. 제308조(사자의 명예훼손) 공연히 허위의 사실을 적시하여 사자의 명예를 훼손한 자는 2년 이하의 징역이나 금고 또는 500만원 이하의 벌금에 처한다. 제309조(출판물 등에 의한 명예훼손) ①사람을 비방할 목적으로 신문, 잡지 또는 라디오 기타 출판물에 의하여 제307조제1항의 죄를 범한 자는 3년 이하의 징역이나 금고 또는 700만원 이하의 벌금에 처한다. 제313조(신용훼손) 허위의 사실을 유포하거나 기타 위계로써 사람의 신용을 훼손한 자는 5년 이하의 징역 또는 1천500만원 이하의 벌금에 처한다. 제347조(사기) ①사람을 기망하여 재물의 교부를 받거나 재산상의 이익을 취득한 자는 10년 이하의 징역 또는 2천만원 이하의 벌금에 처한다.	허위사실 적시 명예훼손 등
정보통신망법	제70조(벌칙) ②사람을 비방할 목적으로 정보통신망을 통하여 공공연하게 거짓의 사실을 드러내어 다른 사람의 명예를 훼손한 자는 7년 이하의 징역, 10년 이하의 자격정지 또는 5천만원 이하의 벌금에 처한다.	비방 목적 거짓 사실로 인한 명예훼손
공직선거법	제82조의4(정보통신망을 이용한 선거운동) ②누구든지「정보통신망 이용촉진 및 정보보호 등에 관한 법률」제2조제1항제1호에 따른 정보통신망(이하 “정보통신망”이라 한다)을 이용	허위 사실 유포 금지

구분	현행 법률	비고
	하여 후보자(후보자가 되려는 사람을 포함한다. 이하 이 조에서 같다), 그의 배우자 또는 직계존·비속이나 형제자매에 관하여 허위의 사실을 유포하여서는 아니되며, 공연히 사실을 적시하여 이들을 비방하여서는 아니된다. 다만, 진실한 사실로서 공공의 이익에 관한 때에는 그러하지 아니하다.	
성폭력처벌법	제14조의2(허위영상물 등의 반포등) ①반포등을 할 목적으로 사람의 얼굴·신체 또는 음성을 대상으로 한 촬영물·영상물 또는 음성물(이하 이 조에서 “영상물등”이라 한다)을 영상물 등의 대상자의 의사에 반하여 성적 욕망 또는 수치심을 유발할 수 있는 형태로 편집·합성 또는 가공(이하 이 조에서 “편집등”이라 한다)한 자는 5년 이하의 징역 또는 5천만원 이하의 벌금에 처한다.	허위영상 유포 금지
민법	제750조(불법행위의 내용) 고의 또는 과실로 인한 위법행위로 타인에게 손해를 가한 자는 그 손해를 배상할 책임이 있다.	손해배상

나. 신뢰성 확보 및 진실성 판단 등을 위한 기초정보 제공

인공지능을 악용한 허위조작정보 유통을 예방하기 위해서는 인공지능을 악용했다는 사실, 유통되는 정보가 진실하지 않다는 점 등을 이용자, 관리·감독 주체들이 판단할 수 있다. 그러나 특정한 목적을 가지고 의도적으로 인공지능을 악용하여 진실하지 않은 정보를 생산·유통하는 자들이 스스로 인공지능을 악용했다는 사실, 유통되는 정보가 진실하지 않다는 것을 밝히는 것을 기대하기 어렵다. 따라서 알고리즘 편향성 여부, 진실성 논란의 정보 등을 이용자 스스로 판단할 수 있는 기초정보를 제공하는 것이 중요하다. 알고리즘의 경우 이용자들이 쉽게 그 편향성 여부를 판단하기 어렵기 때문에 투명하고 언제든지 검증할 수 있는 알고리즘 검증프로그램을 공인된 기관이 마련하여 제공할 필요가 있으며, 진실성 여부가 논란이 되는 정보에 대해서는 찬성과 반대 과정을 이용자들이 살펴볼 수 있도록 하고 이러한 정보에 대한 추천서비스를 자제하도록 하여 확증편향의 문제점도 함께 해소할 필요가 있다. 유럽연합이 인공지능의 영향력과 신뢰성 확보를 위해 인간의 존엄성 보호에 큰 가

치를 두고 “인간중심과 감독, 투명성, 책임성, 기술의 견고성 및 안정성” 등 7가지의 인공지능 윤리지침을 강조하는 사례와 허위조작정보 진실성 여부 판단을 사법부의 종국적인 판단 이전에 행정부가 하는 경우 초래될 우려가 있는 인터넷상 언론표현의 자유 제한 등의 문제점을 고려하여 규제 법률 마련에 신중한 입장을 취하는 미국의 사례도 그 시사점이 크겠다.

다. 사회적 공감대 형성 및 인터넷서비스제공자에 따른 선별적·순차적 규제

인공지능 활성화 방안과 함께 인공지능 알고리즘 편향성 등으로 인한 문제점을 예방하기 위하여 우리나라는 인공지능의 투명성, 공정성, 책무성, 인간의 존엄성 보호, 인공지능 투명성 확보 및 설명 의무 부과, 최소한의 개인정보 활용, 특정 집단 또는 개인의 권리 등의 차별금지, 인공지능 위험성에 따른 규제 차별화” 등을 강조하고 있다. 해외의 사례를 비추어보면 인공지능을 악용하는 문제점을 예방하고 인공지능을 활성화하기 위한 조화로운 규제 방안을 마련하기 위하여 가이드라인 발표, 법률 제정을 위한 제안, 규제 법률 마련 등의 과정을 통하여 사회적 공감대 형성을 위한 시민 참여를 유도하고 있으며, 인공지능 설명가능성, 적절한 통제수단, 검증가능성, 지속가능성, 윤리적·법적 한계 명시 등 규제 프레임 마련에 필요한 구체적이고 실효성 있는 방안을 점진적으로 제시하고 있다. 허위조작정보의 경우 우리나라는 법원으로 하여금 언론 등의 고의 또는 중과실로 인한 허위조작보도에 따라 발생한 재산상 손해에 대한 손해배상액을 최대 5배까지 부과할 수 있도록 하는 법안이 발의되었으나 찬반 의견이 팽배하여 진척을 이루지 못하는 상황이다. 미국의 경우 인터넷상 표현의 자유를 강하게 보호하기 위하여 허위조작정보 규제에 소극적인 경우도 있으나, 독일과 같이 소셜 네트워크 서비스 제공사업자에게 독일 형법에 위배되는 불법정보에 대한 24시간 이내 삭제

또는 접속차단 등 조치를 취하도록 하는 법률을 제정한 경우도 있다. 유럽연합의 경우에는 허위조작정보 위험에 대처하기 위하여 호스팅 사업자, 온라인 플랫폼 사업자 등 인터넷서비스제공자 또는 월 평균 이용자 수가 4,500만명 이상인 대규모 온라인 플랫폼 서비스 사업자 등으로 구분하여 차등적인 의무를 부여하는 것이 특징이다. 특히 영국의 경우 잠재적 해를 유발하거나 공익을 위협하는 인터넷상 잘못되거나 오해의 소지가 있는 온라인 오보 및 해를 입히기 위해서 의도적으로 작성된 허위조작정보 등에 대한 정부 각 부처의 허위조작정보 대응 능력을 향상시키기 위해 개발한 허위조작정보 대응을 위한 데이터 플랫폼(CDDP)을 마련한 것이 특징이다.

인공지능을 악용한 허위조작정보 규제를 위한 사회적 공감대 형성이 필요한 상황에서 이를 규제하기 위한 법률을 제정하기보다는 이용자·인터넷서비스제공자 등 이해당사자들의 의견을 충분히 수렴하고, 허위조작정보 유통으로 사회적 영향력이 큰 인터넷서비스제공자 등을 상대로 선별적이고 순차적인 규제 법제도를 도입할 필요가 있다. 즉, 인공지능을 이용한 허위조작정보로 인한 피해를 최소화하기 위해서는 허위조작정보 대응을 위한 데이터 플랫폼을 마련하여 국가적 차원에서 대응 능력을 높이는 영국의 사례를 참고할 필요가 있으며, 사회적 공감대 형성이 필요한 상황에서 일률적인 규제보다는 사회적 영향력이 큰 인터넷서비스제공자를 상대로 일정한 의무를 부여하는 등 선별적이고 순차적인 규제 제도를 도입한 유럽연합의 사례도 시사점이 크다고 하겠다.

라. 입증책임의 전환 등 피해자 보호

인공지능을 악용한 허위조작정보로 인한 피해자가 손해배상을 청구하기 위해서는 가해자의 고의 또는 과실로 인한 위법행위로 인하여 손해가 발생하였다는 점을 입증하여야 하나(민법 제750조), 가해자 특정, 인공지능이 악용되었다는 사실 및 특정 정보의 허위성 등을 피해자가 입

증하기란 쉽지 않기 때문에 손해배상 청구를 위한 가해자의 고의 또는 과실 여부, 그로 인한 위법행위로 손해가 발생하였다는 인과관계 등의 입증책임을 가해자에게 전환하여 피해자 보호를 강화할 필요가 있다. 다만, 허위성 여부의 입증을 가해자에게 전환하더라도 그 입증이 어려운 점은 동일하기 때문에 피해자라고 주장하는 자에게는 약한 정도의 소명의무를 부여하고, 인공지능을 악용하지 않았다는 사실 및 특정 정보가 진실하다는 사실에 대해 가해자의 일반적 입증의무를 부여한 후, 사법부의 판단에 따라 손해배상책임 여부를 결정해야 할 것이다.

Ⅳ. 맺은 말

정보의 바다라고 불리는 인터넷은 좋은 정보만 이용할 수 있는 이상적인 환경이 구축되지 않는 한 나쁜 정보, 정확하지 않은 정보 등이 함께 유통되기 때문에 이용자들이 좋은 정보를 선별하여 이용할 수 있는 능력을 갖추고 정보제공자는 좋은 정보만 제공할 수 있는 인터넷 환경이 구축되도록 노력할 필요가 있다. 특히, 특정 목적을 위해서 의도적으로 허위정보를 유통시켜 여론을 호도하여 민주주의의 다양성이라는 핵심가치를 훼손하는 행위에 대한 규제 필요성은 점점 높아지고 있다. 그러나 인공지능, 허위조작정보의 개념 정의의 어려움으로 관련 규제 입법은 쉽지 않은 상황이다. 특히, 인공지능을 악용하여 생산한 허위조작정보를 유통하여 유발되는 민주주의의 위협, 선의의 피해자 양산 등의 문제점을 해결해야 한다는 당위성은 있지만, 효과적인 규제 입법 마련이 어려운 상황이다.

본 논문에서는 인공지능 기술을 활용하여 생산·유통하는 허위조작정보와 관련된 법률적 쟁점 및 시사점을 다음과 같이 제언한다. 첫째, 인공지능 및 허위조작정보의 명확한 법적 개념을 마련하지 않고 규제 입법을 하는 경우 어떤 내용의 인공지능 또는 허위조작정보가 규제 대상

이 되는지, 어느 범위까지 적법한지 등에 대한 예측가능성을 확보하기 어렵다. 따라서 인공지능 및 허위조작정보의 법적 개념, 보호 대상 등이 불명확하여 이를 규제하기 위한 법률 제정이 수월하지 않은 상황에서는 적용 가능한 현행 법률을 적극적으로 활용[62]하여야 한다. 또한 인공지능 및 허위조작정보에 대한 사례를 분석한 결과를 축적하여 개념 정의, 보호 범위 등을 명확하게 마련할 수 있는 시점에 인공지능을 활용한 허위조작정보를 규제하는 법률을 제정하는 것이 타당하겠다. 둘째, 인공지능 알고리즘이 편향되는 경우에는 정확하지 않은 데이터를 기반으로 하거나 사실을 왜곡하는 정보 등으로 인하여 특정인·특정 집단·특정 의견 등을 무시하거나 차별할 우려가 있다. 특히 이용자가 이러한 알고리즘 편향성을 인식하지 못하여 정확한 가치 판단의 기회를 상실할 수 있고, 인공지능이 허위조작정보 확산 수단으로 악용될 우려가 있다. 따라서 이용자들이 편향성 여부를 판단하기 어려운 알고리즘에 대하여 공인된 기관이 마련하여 제공한 알고리즘 검증프로그램을 통하여 언제든지 알고리즘을 투명하게 검증할 수 있도록 하고, 진실성 여부가 논란이 되는 정보에 대해서는 찬성과 반대 과정을 이용자들이 살펴볼 수 있도록 하여야 한다. 이와 함께 논란이 해소되지 않은 정보의 추천서비스를 자제하도록 하여 확증편향의 문제점도 함께 해소할 필요가 있다. 유럽연합이 인공지능의 영향력과 신뢰성 확보를 위해 인간의 존엄성 보호에 큰 가치를 두고 “인간중심과 감독, 투명성, 책임성, 기술의 견고성 및 안정성” 등 7가지의 인공지능 윤리지침을 강조하는 사례와 허위조작정보 진실성 여부 판단을 사법부의 종국적인 판단 이전에 행정부가 하는 경우 초래될 우려가 있는 인터넷상 언론표현의 자유 제한 등의 문제점을 고려하여 규제 법률 마련에 신중한 입장을 취하는 미국의 사례도 그 시사점이 크겠다. 셋째, 허위조작정보를 규제하기 위해서는 해당 정보가 허위라는 사실을 규명하여야 한다. 사법부의 종국적 판단 이전에 행해진 행정부의

62) 박용숙, 위의 논문, 161.

허위성 판단이 사법부의 종국적 판단과 상이한 경우에는 인터넷상 언론 표현의 자유를 제한하거나 위축시킬 우려가 있다. 따라서 허위조작정보 규제 필요성에 대한 사회적 공감대를 형성하고, 유럽연합과 같이 일률적인 규제보다는 사회적 영향력이 큰 인터넷서비스제공자를 상대로 일정한 의무를 부여하는 등 선별적이고 순차적인 규제 사례를 참고할 필요가 있다. 넷째, 편향된 알고리즘으로 구성된 인공지능을 통하여 허위조작정보를 생산하고 유통함으로써 피해자가 발생한 경우, 고의 또는 과실로 인한 위법행위로 타인에게 손해를 가한 자에 대한 손해를 배상할 책임을 누구에게 물을 것인지 불분명하여 민사책임에 있어서의 행위책임 원칙을 훼손할 우려가 있다. 또한, 편향된 인공지능을 통하여 유통되는 허위조작정보를 예방하기 위하여 해당 정보 삭제 또는 차단 등의 조치를 요구하는 경우 누구를 대상으로 해야 하는지도 불분명하다. 따라서 가해자 특정, 인공지능이 악용되었다는 사실 및 특정 정보의 허위성 등을 피해자가 입증하기란 쉽지 않기 때문에 손해배상 청구를 위한 가해자의 고의 또는 과실 여부, 그로 인한 위법행위로 손해가 발생하였다는 인과관계 등의 입증책임을 가해자에게 전환하여 피해자 보호를 강화하여야 하며, 해당 정보 삭제 또는 접속차단 등의 시정요구 상대방을 명확하게 특정하여야 한다.

참 고 문 헌

강지현 (2019). 인터넷상 여론조작의 실태와 규제에 대한 형사법적 고찰. 〈비교형사법연구〉, 제21권 제1호, 287-315.

고수윤 (2022). 허위조작정보의 규제에 대한 소고 - 미국의 정보통신서비스 제공자 규제를 참고하여 -, 〈법이론실무연구〉, 제10권 제4호, 73-110.

구인혁 (2008). 신입사원의 허위정보 제공 성향의 상관 요인과 검증방법 연구. 〈인적자원관리연구〉, 제15권 제1호, 37-51.

김민정 (2019). 가짜뉴스(fake news)에서 허위조작정보(disinformation)로 - 가짜뉴스 규제 관련 국내 법안과 해외 대응책에 나타난 용어 및 개념정의 비교 -. 〈미디어와 인격권〉, 제5권 제2호, 43-81.

_____ (2020). 유럽연합의 온라인 허위조작정보(disinformation) 자율규제 사례 고찰. 〈언론과 법〉, 제19권 제1호, 193-231.

김현귀 (2016). 표현의 자유와 혐오표현규제. 〈헌법이론과 실무〉, A-3, 1-120.

박상민 (2019). 인터넷상 허위조작행위에 대한 합리적 규제방향 - 관련 현행법 및 입법안 검토를 중심으로 -. 〈법학연구〉, 제19권 제3호, 1-24.

박용숙 (2022). 허위조작정보의 규제방안에 대한 시론적 고찰 - 독일의 네트워크집행법을 중심으로 -. 〈법이론실무연구〉, 제10권 제4호, 137-168.

지성우 (2019). 허위조작정보(소위 '가짜뉴스') 규제에 대한 헌법적 문제점에 관한 연구. 〈공법연구〉, 제48집 제2호, 157-192.

안수길 (2019). 가짜뉴스에 맞서는 독일 사회관계망법집행법의 내용과 쟁점. 〈법학논총〉, 제36권 제1호, 121-147.

이인호·이준형 (2021). 허위조작정보 규제의 헌법적 한계 — 언론중재법 개정안(대안) 및 5·18민주화운동특별법을 중심으로 —. 〈언론과법〉, 제20권 제3호, 223-263.

최지선 (2022). 프랑스의 허위조작정보 대응에서의 국가안보 관점 - 〈2018 정보조작 근절에 관한 법률〉을 중심으로 -. 〈언론과법〉, 제21권 제2호, 85-128.

최종선 (2020). 허위조작정보 규제 시도에 대한 비판적 고찰. 〈홍익법학〉.

제21권 제1호, 364-391.
홍완식 (2019). 가짜뉴스 규제 법안에 대한 입법평론. 〈법과 정책〉, 제25권 제1호, 329-357.

House of Commons Digital, Culture, Media and Sport Committee (2018). Disinformation and fake news_Interim Report.
Molly Lesher (2022). Leveraging AI, big data analytics and people to fight untruths online.
Richard Gunther, Paul A. Beck, Erik C. Nisbet (2018). Fake News May Have Contributed to Trump's 2016 Victory, Ohio State University.
The Committee on Digital Economy Policy (CDEP) (2022). Disentangling untruths online : Creators, spreaders and how to stop them, OECD Going Digital Toolkit.
White Paper (2020). On Artificial Intelligence - A European approach to excellence and trust.

헌법재판소 2011. 12. 29. 선고 2007헌마1001 결정 등
대법원 2004. 6. 25. 선고 2004도2062 판결
대법원 2005. 4. 29. 선고 2005다13332 판결
대법원 2013. 11. 14. 선고 2013도2190 판결
대법원 2018. 4. 19. 선고 2017도14322 판결
대법원 2023. 2. 2. 선고 2022도13425 판결

공직선거 및 선거부정방지법 (1997. 11. 14.)
민법 (2023. 6. 28.)
성폭력범죄의 처벌 등에 관한 특례법 (2022. 7. 1.)
정보통신망 이용촉진 및 정보보호 등에 관한 법률 (2023. 7. 4.)
형법 (2021. 12. 9.)

https://artificialintelligenceact.eu/the-act/
https://blogs.lse.ac.uk/polis/2019/11/18/new-powers-new-responsibilities/
https://commission.europa.eu/system/files/2020-02/commission-white-paper-artificial-intelligence-feb2020_en.pdf

https://courts.delaware.gov/opinions/
https://digital-strategy.ec.europa.eu/en/policies/online-disinformation
https://ec.europa.eu/futurium/en/european-ai-alliance/have-your-say-artificial-intelligence-white-paper-european-approach-excellence.html
https://en.unesco.org/fightfakenews
https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32022R2065
https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF
https://goingdigital.oecd.org/data/notes/No23_ToolkitNote_UntruthsOnline.pdf
http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html
https://kcc.go.kr/user.do?mode=view&page=A05030000&dc=K00000200&boardId=1113&boardSeq=51454
https://m.blog.naver.com/PostView.naver?isHttpsRedirect=true&blogId=jongheesalon&logNo=221353750786
https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained
https://news.kbs.co.kr/news/view.do?ncd=3424373
https://openai.com/blog/chatgpt
https://oecd.ai/en/ai-principles
https://oecd.ai/en/policy-areas
https://oecd.ai/en/wonk/untruths-online
https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/363/363.pdf
https://stdict.korean.go.kr/search/searchResult.do
https://trumpwhitehouse.archives.gov/ai/executive-order-ai/
https://twitter.com/EliotHiggins/status/1637927681734987777
https://uscode.house.gov/view.xhtml?req=granuleid:USC-prelim-title47-section230&num=0&edition=prelim
https://www.ai.gov/wp-content/uploads/2023/04/National-Artificial-Intelligence-Initiative-Act-of-2020.pdf
https://www.bbc.com/news/world-asia-india-39971412
https://www.bbc.com/news/world-asia-india-40657074

https://www.bmwk.de/Redaktion/EN/Downloads/E/key-points-for-federal-government-strategy-on-artificial-intelligence.pdf?__blob=publicationFile&v=1
https://www.eidosmedia.com/blog/tecnnology/AI-in-Journalism
https://www.foxnews.com/media/fox-news-media-dominion-voting-systems-reach-agreement-over-defamation-lawsuit
https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper
https://www.gov.uk/government/publications/counter-disinformation-data-platform-privacy-notice/counter-disinformation-data-platform-privacy-notice
https://www.humanrights.go.kr/site/program/board/basicboard/view?boardtypeid=24&boardid=7607961&menuid=001004002001
https://www.ki-strategie-deutschland.de/files/downloads/Fortschreibung_KI-Strategie_engl.pdf
https://www.legislation.gov.uk/ukpga/2018/12/section/8/enacted
https://www.nytimes.com/2016/12/05/business/media/comet-ping-pong-pizza-shooting-fake-news-consequences.html
https://www.nytimes.com/video/business/100000008864999/dominion-fox-news.html?searchResultPosition=9
https://www.osce.org/files/f/documents/6/8/302796.pdf
https://www.police.go.kr/user/bbs/BD_selectBbs.do?q_bbsCode=1001&q_bbscttSn=20230317103332394
https://www.senate.gov/civics/constitution_item/constitution.htm#amdt_1_(1791)
https://www.supremecourt.gov/qp/21-01333qp.pdf
https://www.unesco.org/en/articles/solving-fake-news-starts-avoiding-term
https://www.washingtonpost.com/politics/2023/03/22/trump-arrest-deepfakes/
https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf

ABSTRACT

A Study on Artificial Intelligence and Disinformation*

CHOI, JONGSUN
Adjunct Professor of Public Policy Graduate School at Hanyang University and
Deputy General Manager of the Legal Team of the Planning and Coordination
Office of the Korea Communications Standards Commission

This paper discusses legal issues and implications concerning the production and dissemination of disinformation using artificial intelligence (AI) technology. First, defining clear legal concepts of AI and disinformation poses challenges, leading to uncertainty in regulating and determining their legality and scope. Therefore, when legal concepts and protections of AI and disinformation are unclear and enacting laws to regulate them is challenging, existing laws should be actively used. Second, biased AI algorithms may lead to the risk of neglecting or discriminating against specific people, groups, and opinions based on inaccurate data or distorted facts. In addition, if users are unaware of such algorithm biases, there is a risk of abusing AI to spread false and manipulative information. Therefore, preparing and providing an algorithm verification program that allows users to verify biases at any time and offers transparency in the approval and opposition process is essential to mitigate this. Third, regulating disinformation necessitates

* I would like to thank Prof. Jackie Harrison (Head of the Department) and Dr. Irini Katsirea (Reader in International Media Law) for inviting me as an Advanced Visiting Fellow at the Department of Journalism Studies at Sheffield University in the United Kingdom.

confirming its falsity. However, premature making false judgments before the judiciary's final judgment may impinge on the freedom of speech on the Internet. Therefore, it is necessary to refer to the introduction of selective and sequential regulatory systems, such as forming a social consensus on the need to regulate disinformation and imposing certain obligations on Internet service providers with significant social influence rather than uniform regulations like those in the European Union. Fourth, the responsibility for damages to others resulting from the intentional or negligent production and dissemination of disinformation through biased algorithms remains unclear. Therefore, shifting the burden of proving the perpetrator's intention or negligence can strengthen victim protection.

Keywords: artificial intelligence, disinformation, algorithms, deepfakes, public opinion manipulation

[논문투고일 2023. 6. 14. 논문수정일 2023. 7. 17. 게재확정일 2023. 7. 24.]