정보통신정책연구 제 29 권 제 2 호 (2022. 6) pp.111-144
http://dx.doi.org/10.37793/ITPR.29.2.4

AI 추천 알고리즘 편향성과 규제에 관한 연구*

A Study on the Bias and Regulation of AI Algorithm

[1)]

김 승 현 (Seunghyun Kim)**
김 시 원 (Siwon Kim)***
안 정 민 (Jungmihn J. Ahn)****

국문초록

AI 기술이 비약적으로 발전하면서 다양한 플랫폼에서 이용자 맞춤형의 'AI 추천 알고리즘'이 활용되고 있다. AI 추천 알고리즘은 서비스 이용자의 취향에 맞추어 적절하게 콘텐츠를 추천하는 기술로써 방대한 정보의 늪에서 '선택'을 위해 필수적으로 사용할 수밖에 없는 기술이기도 하다. 플랫폼은 프로파일링을 토대로 한 이용자 취향을 그 선택의 기준으로 삼고 맞춤형의 정보를 추출하여 제공하다 보니 이용자는 자신과 유사한 경향의 사람들과만 소통하게 되어 점점 에코 챔버(echo chamber) 현상에 빠질 수 있다는 우려의 목소리도 높아지고 있다. 이러한 문제는 UGC와 결합하여 코로나 19과 같은 사회 공통적인 현상에 대해 잘못된 사실관계나 정보를 통해 여론 양극화 현상으로 이어지고 있다.

본 연구는 실험을 통해 유튜브의 AI 추천 알고리즘은 이용자가 허위거나 사실관계가 명확하지 않은 UGC 같은 영상을 시청하기 시작하면 계속해서 이런 UGC 위주로 추천함으로써 이용자가 공신력 있는 콘텐츠를 통해 정확한 정보를 받아들이기 어려운 상황에 빠져들 위험성에 노출되어 있다는 것을 확인하였다. 또한 AI의 부작용으로 인식되고 있는 편향성 발생 여부를 확인하기 위해 AI 추천 알고리즘 실태를 분석하고 이에 대한 해외 각국의 정책과 규제를 비교법적으로 탐색했다.

우리나라를 비롯한 여러 국가들의 알고리즘 관련 규제안과 가이드라인을 분석한 결과 '이용자 중심', 'AI 알고리즘의 투명한 공개' 등 상당한 부분에서 공통적인 내

※ 논문접수일: 2022. 5. 11, 수정일: 2022. 6. 27, 게재확정일: 2022. 6. 29
* 이 논문은 2021년도 한림대학교 교비연구비(HRF-202112-010)에 의하여 연구되었음
** 한림대학교 정보법과학전공 학생
*** 한림대학교 정보법과학전공 학생
**** 한림대학교 정보법과학 전공 교수

용을 찾아볼 수 있으며, 또한 비록 규범력에서 차이는 발견되고 있으나 규제안 제정에 이미 상호 영향을 미치고 있음을 확인하였다. 이를 기초로 본 연구는 AI 추천 알고리즘에 대한 우리나라의 정책과 규제 방향을 비교법적으로 분석하며 그 효과를 검토하고, AI 추천 알고리즘 편향성 문제를 해결하기 위한 대안으로 가이드라인의 구체화 방안과 이용자 선택권의 강화방안을 제시하였다.

주제어: AI 알고리즘, 편향성, 에코챔버, AI 알고리즘 규제

ABSTRACT

With the rapid development of artificial intelligence technology, user-customized AI recommendation algorithms are being used on various platforms. AI recommendation algorithms suggest content deemed of interest to users based on user history and are necessary, to a large extent, to allow 'selection' from a vast pool of information. As platforms offer recommendations based on user preference profiling, there are concerns that users increasingly interact only with like-minded individuals and thus fall into an echo chamber. Such a problem, combined with UGC, leads to a polarization of public opinion, caused by false facts and misinformation on common social issues such as COVID-19.

This study confirms, through experiments, that YouTube's AI recommendation algorithm makes it difficult for users to obtain accurate information through reliable content, once they begin watching UGC containing false or unconfirmed facts, as the algorithm continues to recommend UGC of similar nature. This study also examines AI recommendation algorithms to check for AI bias and compares relevant policies and regulations in different countries.

Our analysis of regulatory proposals and guidelines on algorithms shows considerable similarity between Korea and other countries, regarding 'user-centricity' and 'transparent disclosure' of AI algorithms. And, although differences have been observed in the scope of regulation, the examined proposals and guidelines are found to be drawing influence from one another. Based on such findings, this study provides a comparative analysis of South Korea's policies and regulatory directions for AI recommendation algorithms, considers their effects, and proposes such measures as establishing concrete guidelines and supporting user choices to solve the bias problem of AI recommendation algorithms.

Key words: AI Algorithm, Bias, Echo Chamber, Algorithm Regulations

Ⅰ. 서론

모든 일상은 디지털화되어 수많은 플랫폼과의 연결을 통해 일어나고 있다. 검색엔진, 이메일, 메신저, 쇼핑몰, 음식을 주문하는 배달앱부터 콘텐츠를 볼 수 있는 다양한 미디어 플랫폼에 이르기까지 우리는 방대한 양의 정보를 취향대로 고를 수 있는 플랫폼 전성시대에 살고 있다. 이용자들의 자발적인 활동을 통해 만들어지는 방대한 콘텐츠와 데이터는 플랫폼을 통하여 수집되고 더 가치 있는 정보로 정제되어 다시 사용자들에게 제공된다. 대부분의 플랫폼은 이용자의 성향을 사전에 파악하고 아이템을 추천하는 인공지능 알고리즘을 적극적으로 활용하고 있다.

AI 추천 알고리즘은 페이스북 친구들의 소식을 알려주고, 트위터 타임라인에 트윗들, 온라인 쇼핑몰의 추천 상품, 넷플릭스의 영화, 특정 검색어에 대한 검색결과, 구인구직 플랫폼의 구인 공고, 적절한 온라인 광고 등에서 이용자의 취향에 맞게 각종 콘텐츠를 추천한다(츠바이크, 2021, p.91). 서비스 이용자의 관심과 성향을 알아내고, 콘텐츠 이용 행태를 분석하여 이용자별 맞춤형 콘텐츠를 제공하는 AI 추천 알고리즘은 콘텐츠 서비스업자에게 이제 '선택이 아닌 필수'로 자리 잡았다.[1)]

그러나 이러한 AI 추천 알고리즘이 여러 플랫폼에서 반복적으로 사용되면서 이용자는 자신의 취향에 부합하는 소위 '맞춤형 정보'만을 제공받게 되고 결과적으로 자신과 유사한 생각을 하는 사람들과만 소통하게 되는 '에코챔버(echo chamber)'현상을 가져오게 된다. 사용 이력이 쌓일수록 AI는 이용자의 성향을 잘 알게 되고, 더욱 정밀하게 그에 부합하는 정보 위주로 필터링하여 제시하게 된다. 결과적으로 비록 AI 알고리즘이 목적한 바는 아니나 이용자에게 보이는 정보의 주제나 범위가 제한됨으로써 나와 다른 의견은 마치 존재하지 않는 것이 되어 버린다.

특히 AI 추천 알고리즘이 이용자에게 특정 정보를 제공하는 원칙이나 추천 학습 과정이 구체적으로 공개되지도 않고, 이용자는 AI 추천 알고리즘의 사용이나 거부에 대한 선택권이 거의 없기 때문에 이용자는 자신도 모르는 사이에 자신의 틀에 갇히게 된다. 이러한 현상을 보면 마치 수많은 정보의 홍수 속에서 편리함을 추구하기 위해 개발된 AI 추천 알고리즘이 독재국가에서 정권이 자신에게 유리한

1) 서봉원 (2016). 콘텐츠 추천 알고리즘의 진화.『방송트렌드 & 인사이트』, Vol.5, 19면.

정보만을 TV 등 미디어 매체를 통해 전달하던 구시대로 회귀하고 있는 듯하다. 오히려 지금은 정보를 왜곡하는 주체가 누군지도 모른다는 점에서 상황은 구시대보다 더 심각한 왜곡에 빠져있는지도 모른다.

최근에 겪은 코로나19는 AI 추천 알고리즘의 이러한 문제점을 단면적으로 보여준 사례였다. 코로나19의 유일한 대안으로 제시된 백신의 안전성 여부에 대한 사회적 관심이 증가하면서 백신 접종에 대한 많은 뉴스가 생산되었다. '백신 접종자에게 전자칩을 심는다', '백신을 맞으면 유전자가 변형된다' 등 근거가 명확하게 밝혀지지 않은 UGC(User Generated Contents)가 백신뿐만 아니라 정치, 경제 등 많은 사람의 삶에 관여하는 측면으로 유통되는 문제가 있었다(김춘식, 홍주현, 2020, pp. 403～440). 코로나19 외에도 가짜뉴스의 대부분이 정보 분별력이 약한 청소년이나 다양한 미디어 아웃렛을 활용하지 못하는 장년층의 올바른 정보공유에 위협이 되는 상황이며 이미 중요한 사회문제로 대두된 지 오래다.

알고리즘에 대한 편향성 중 특히 미디어 콘텐츠의 경우는 허위사실 또는 가짜뉴스가 사회적인 분열을 초래할 위험성이 있다고 많은 학자는 지적하고 있다(이상훈, 2020, 최승필 2020, 김유미, 2021, 이문한 2021). 학자들은 공통으로 신뢰할 수 없는 내용의 콘텐츠는 건전한 합의와 여론 형성을 막고 전반적인 사회 신뢰를 저하시킴으로써 종국적으로는 민주적 질서 훼손과 사회 공동체 형해화를 가져오기 때문에 규제가 필요하다고 주장한다. 그러나 미디어 콘텐츠 규제를 위해서는 '허위사실' 또는 '가짜뉴스'의 법적 개념이 우선으로 정립되어야 하나 현실적으로 쉬운 일은 아니다. 표현의 자유와 알권리가 가지는 헌법적 제한, 개념의 추상성, 허위와 진실의 구별 곤란성 등으로 인해 학설은 물론 대법원과 헌법재판소 역시 이에 대한 법적 정의를 내리지 못하는 상황이기 때문이다.[2] 이러한 개념 정립의 어려움과 블랙박스 같은 알고리즘 규제에 대한 정책이 갈피를 잡지 못하고 있는 사이에 '이용자 맞춤형'이라는 각종 AI 추천 알고리즘의 전방위적인 사용으로 우리의 다양한 정보공유 가능성은 더욱 위협받는 상황이 되었다.

이러한 왜곡 현상을 극복하기 위하여 사업자들도 자발적으로 그로 인한 문제점을 최소화하는 방안에 주력하고 있다. 구체적으로 보면 네이버는 2021년 11월 29일 서울대학교 인공지능 정책 이니셔티브(Seoul National University AI Policy Initiative, SAPI)와 'AI 리포트'를 발간했다.[3] 여기에서는 5가지의 'AI 윤리 준칙'

2) 현재 법적으로는 '일반적으로 바르지 못한 것 또는 참이 아닌 것', '진실에 부합하지 않는 사실' 등과 같이 극히 제한적인 기준만이 제시되고 있을 뿐이다.

에 대해서 설명한다. 먼저, 인간 중심의 가치를 최우선으로 삼고, 다양성을 존중하며 합리적인 설명과 사용자의 편리성을 고려하는 AI, 사람에게 유해한 영향을 끼치지 않는 AI 서비스 설계, AI의 개발과 이용 과정에서의 프라이버시 보호, 서비스 전 과정에서 정보 보안을 고려한 설계 적용을 선언하고 있다.[4)]

또한, IBM은 알고리즘의 편향을 막기 위해서 이미 2018년에 AI Fairness 360이라는 애플리케이션 프로그래밍 인터페이스(Application Programming Interface, API)를 개발했다. 이는 컴퓨터 또는 컴퓨터 프로그램 사이를 연결해주는 파이썬 패키지를 제공하여 알고리즘의 편향을 검출하고 이를 제거할 수 있는 알고리즘 적용이 가능한 서비스이다. IBM이 제공하는 오픈소스 툴키트는 편향을 완화할 수 있는 10가지 알고리즘을 포함하고 있다. 여기에는 알고리즘의 훈련 데이터, 학습 알고리즘 그리고 예측 등을 수정하는 알고리즘, 학습 데이터를 입력한 후 학습 전후에 편향성을 감지하고 완화할 수 있는 알고리즘 등이 있다.[5)]

하지만 인공지능이 발전함에 따라 알고리즘의 활용은 끊임없이 사회적 문제를 야기할 것이 예상된다는 점에서 업계의 자율규제만으로는 부족하며, 사회 전반의 공익을 위해 정부도 적극적으로 노력을 기울일 필요가 있다. 이하에서는 AI 추천 알고리즘이 우려하는 바와 같이 정보의 편향성을 가져오는지를 이용자의 관점에서 확인하고 이에 대한 우리나라와 해외 각국의 정책적 방향을 살펴보았다. AI 추천 알고리즘에 실제로 문제가 있다면 우리는 어떤 대응을 하고 있는지, 이러한 방법이 효과적인지 분석하고 해외 사례와 비교 분석함으로써 개선이 필요한 부분을 함께 제시하여 올바른 정책 방향을 모색하고자 한다.

3) 네이버 Agenda Research., & 서울대 인공지능정책 이니셔티브. (2021). NAVER-SAPI AI REPORT. Retrieved May 9th, 2022, https://www.navercorp.com/navercorp_/research/2021/20211129093002_2.pdf

4) 윤리 준칙의 실행을 위해 구체적으로 메일링 그룹 형태의 유연한 커뮤니케이션 채널을 마련해 프로젝트 진행 또는 서비스 개발 시 AI 윤리 준칙과 관련된 사안을 논의할 수 있는 커뮤니케이션 채널을 만들고 운영할 것과 사례 중심의 리포트 작성 및 공개, 운영 경과를 담은 리포트의 발간을 계획하고 있다고 밝혔다.

5) IBM Research Trusted AI. (n.d.). AI Faireness 360. Retrieved in May 9th. 2022, https://aif360.mybluemix.net/

Ⅱ. AI 추천 알고리즘 편향 실태

1. AI 추천 알고리즘의 원리와 정보편향 위험성

AI 추천 알고리즘은 여러 가지 항목 중 이용자가 선호할 만한 아이템을 추측해서 해당 이용자에게 적합한 것을 선택(information filtering)하여 제공하는 시스템을 일컫는다. 추천 알고리즘은 사용자의 과거 히스토리에 기반을 두어 사용자 프로파일을 생성하고, 이 프로파일 정보를 통해 다른 유사한 사용자들이 좋아한 영상, 혹은 사용자가 좋아한 아이템과 유사한 아이템을 추천한다.[6] 기술이 지속적으로 발전됨에 따라 최근에는 여러 가지 새로운 기법이 추가되기는 하지만 기본적인 추천 시스템은 협업 필터링(collaborative filtering)과 콘텐츠 기반 필터링 (content-based filtering)을 기반으로 한다.

협업 필터링이란 대규모의 기존 사용자 행동 정보를 분석하여 해당 사용자와 비슷한 성향의 사용자들이 좋아했던 항목을 추천하는 기술이다. 가장 일반적인 예는 온라인 쇼핑 사이트에서 흔히 볼 수 있는 '이 상품을 구매한 사용자가 구매한 상품들' 서비스이다. 협업 필터링은 충분한 이력이 쌓인 아이템에 대해 활용되는 한편, 소비 이력이 적거나 빠르게 아이템이 바뀔 때는 해당 항목 자체를 분석하여 추천을 구현하는 콘텐츠 기반 필터링을 사용한다. 음악을 추천하기 위해 음악 자체를 분석하여 유사한 음악을 추천하는 방식을 예로 생각할 수 있다.

이 외에도 최근에는 Bert, GPT3, T5 등 딥러닝(Deep Learning) 기술에 기반을 둔 새로운 방식의 알고리즘들이 여러 분야에서 놀라운 진전을 보이며 AI 추천 알고리즘은 각 기술의 장단점을 보완하기 위하여 여러 방식을 혼합하여 사용한다. 그러나 인간의 두뇌를 능가하는 인공지능의 비약적인 발전에 따라 더욱 정교한 AI 추천 알고리즘이 개발된다고 하더라도 AI 추천 알고리즘의 핵심은 정보를 필터링(filtering)해서 사용자에게 제시하는 것이다. 그리고 우리가 AI 추천 알고리즘에 대해 우려하는 문제는 이와 같은 알고리즘에 내재하는 원리에 의해 발생한다.

문제의 핵심은 AI 추천 알고리즘에 의존하는 순간 전체 정보를 볼 기회는 사라진다는 것에 있다. AI 추천 알고리즘은 많은 정보 중에서 이용자 개개인의 취향에 맞는 콘텐츠만을 노출함으로써 자신의 신념 또는 가치관에 부합하는 정보 속에

6) 정연오. et. al. (2013). 개인화된 전문가 그룹을 활용한 추천 시스템. 『한국지능시스템학회 논문지』, Vol. 23, No. 1, February 2013, 7-11

갇히는 필터 버블 현상을 일으킨다. 필터 버블은 주로 추천 알고리즘에 의해 생겨나는 에코챔버(echo chamber) 현상으로 우리의 선택과 관계없이 알고리즘에 의해 개인화되어가는 것을 의미한다.[7] AI 추천 알고리즘이 이용자에게 편의를 제공하면서 취향에 맞는 콘텐츠를 추천한다는 장점은 있지만, 이용자로서는 제공되는 정보를 수동적으로 받아들일 수밖에 없고 다른 의견에 대한 접근이 아예 배제되게 된다. 이와 같이 이용자의 신념 또는 가치관에 부합하는 정보만 지속적으로 노출되고 그 이외의 정보로부터는 배제되면 확증편향이 발생할 수밖에 없고, 결과적으로 사회의 극단화와 분열이 초래되기 때문에 미국의 유명한 헌법학자 캐스 선스타인(Cass R. Sunstein)은 에코 챔버를 민주주의의 가장 큰 위협으로 보기도 했다.[8]

2. 정보 편향성 분석 관련 선행 연구 분석

국내에서 AI 추천 알고리즘의 정보 편향성이 학계 및 업계에 문제 되기 시작한 것은 2019년 4월 포털 사이트인 네이버가 뉴스 서비스에 '에어스(AiRs, AI Recommender System)'라는 추천 뉴스 인공지능 편집을 도입하기로 하면서부터였다(강성원, 2019). 2017년에 네이버가 처음 선보인 '에어스'는 이용자의 콘텐츠 소비 패턴을 분석해 추천하는 일종의 AI 추천 알고리즘 시스템이다. 네이버는 2019년도에 '에어스' 도입 이전까지 편집자가 직접 메인 화면 뉴스를 배열하였으나 편집자가 미칠 수 있는 편향성 제거를 위해 AI 알고리즘을 도입하였으며 이후 그에 따라 뉴스 배열이 진행되고 있다.

그러나 AI 알고리즘에 의한 추천 정보의 위험성에 대한 인식이 사회 전반적으로 확산하기 시작한 것은 유튜브에 개인제작 영상(UGC)이 본격적으로 업로드되고 MZ 세대들이 글자로 된 기사보다는 동영상을 선호하기 시작하면서부터라고 할 수 있다. 유튜브는 이용자가 많은 OS 안드로이드 휴대폰의 기본 애플리케이션으로 제공되기 때문에 누구나 쉽고 편리하게 접근할 수 있어 이용률이 높은 미디어 제공 플랫폼이다. 콘텐츠 추천과 제공에 관해 고도화된 알고리즘으로 콘텐츠

7) Arguedes, A. et al. (2022). Echo chambers, filter bubbles, and polarisation: a literature review. Retrieved April 26, 2022, https://reutersinstitute.politics.ox.ac.uk/echo-chambers-filter-bubbles-and-polarisation-literature-review.

8) Sustein, C. (2018). The Echo Chamber Is the Enemy of Democracy. Retrieved April 26, 2022, https://www.bloombergquint.com/view/steve-bannon-the-new-yorker-and-free-speech

시장에서 압도적인 영향력을 가지고 있어 광고 시장에서도 독보적인 우위를 차지하고 있기도 하다.

한국언론진흥재단의 '2021년 소셜미디어 이용자 조사'에서 UGC가 주로 확산되는 소셜미디어가 무엇이라고 생각하느냐는 질문에 58.4%가 유튜브라고 답했으며, 이어 모바일 메신저(카카오톡)가 10.6%, 페이스북이 8.0%, 온라인 카페가 6.7%, 트위터가 5.0% 순으로 많이 확산된다고 응답했다(한국언론진흥재단, 2021, 결과표 pp.351).[9] 또한 정보통신정책연구원(Korea Information Society Development Institute, KISDI)에서 진행한「성별・연령대별 유튜브 및 넷플릭스 콘텐츠 이용행태 분석」을 보면 최근 1주간 연령대별 유튜브의 이용 비율은 20대 부문이 88.7%로 가장 높은 비율을 차지했고, 뒤이어 40대가 87.9%, 60대가 86.1%, 30대가 84.4%(김청희, 김남두, 2021, pp.3)로 여론 형성의 주요 계층인 청년층과 중장년층의 유튜브 이용 비중이 높다는 것을 볼 수 있다.

유튜브의 높은 영향력은 자연스럽게 유튜브 AI 추천 알고리즘에 대한 사회적 시선을 끌었고, AI 추천 알고리즘에 대한 각종 연구 결과와 인공지능학자 기욤 샬로(Guillaume Chaslot)의 유튜브 근무경험 등을 통해 추천 알고리즘이 사실상 사람들이 유튜브에서 더 많은 시간을 보내도록 하는 방향으로 설계되어 있음이 확인된 바 있다.[10] 오세욱(2019)은 '알고리즘으로 본 유튜브의 미디어 지향'에서 유튜브 추천 알고리즘과 이를 이용하는 사회의 문제를 지적하면서 유튜브 추천 알고리즘에 영향을 미치는 조회 수, 평균 시청 시간, 업로드 빈도 중에서는 시청 시간이 중요한 요인이며 그 외에도 영상의 길이가 길수록 영상이 추천되는 빈도가 높아진다는 연구 결과를 발표하였다.

한국언론진흥재단의『유튜브 추천 알고리즘과 저널리즘』에서는 다량의 데이터를 가지고 의문만 존재했던 알고리즘의 편향성에 대한 실험을 실제로 진행하여 유튜브에서 필터버블 현상이 일어나고 있는지를 확인하였다. 위 연구에서는 총 238,875개의 영상을 수집하여 유튜브 추천 알고리즘이 선호하는 영상의 특성을 분

9) 한국 갤럽이 2020년 9월부터 2021년 3월까지 진행한 '마켓70 2021'중 미디어, 콘텐츠, 소셜 네트워크 서비스 조사에서도 국내외 소셜 미디어 연간 이용률은 유튜브가 86%로 압도적인 우위를 차지했고, 뒤이어 네이버 밴드, 카카오 스토리 등의 순으로 확인되었다(갤럽리포트. 미디어, 콘텐츠, 소셜 네트워크 서비스 이용률.『갤럽리포트 마켓70』, 2021(2).).

10) Lewis, P. (2018). 'Fiction is outperforming reality': how YouTube's algorithm distorts truth. Retrieved April 26, 2022, https://www.theguardian.com/technology/2018/feb/02/how-youtubes-algorithm-distorts-truth.

석하였다. 이에 따르면 1) 전통적 언론사를 더 선호했으며; 2) 제목 안에 주요 키워드를 많이 포함하고 있거나 긴 제목을 선호하고; 3) 생중계 영상에 가중치를 두고 있으며; 4) 정치적 키워드에 대해 다른 장르의 영상을 추천하는 등 장르적 다양성을 보이고; 5) 특정 기간에 특정 이슈의 영상을 집중적으로 추천한다고 한다. 6) 시청 시간이 추천 알고리즘에 중요한 요인으로 작용하지만; 7) 개별 키워드의 이념적 성향에 따른 추천 결과에서는 유의미한 차이를 발견할 수 없었다고 하였다. 이러한 선행 실험은 다량의 데이터로 유튜브 추천 알고리즘의 특성을 살펴보고 추측에 근거했던 알고리즘의 편향성에 대한 통계적 분석을 진행하였다는 점에서 의미 있는 연구였다. 다만 이 연구 결과는 개인기록을 남기지 않은 비로그인 상태에서 알고리즘을 파악하는 방식으로 진행되었기 때문에 본 연구의 대상인 개인화에 의한 AI 추천 알고리즘 편향성에 대한 논리적 근거로 제시하기는 어려웠다. 이에 본 연구에서는 소량의 샘플이지만 일반 이용자 환경에서의 실험을 통하여 유튜브 AI 추천 알고리즘이 정보 편향성을 유발하는지 확인하고자 하였다.

3. 유튜브 AI 추천 알고리즘 정보 편향성 확인을 위한 탐색적 접근

본 연구에서는 앞선 선행 연구들을 토대로 유튜브의 AI 추천 알고리즘의 편향성을 확인하기 위하여 이용자의 관점에서 직접 사용해 분석해 보았다. 이를 위하여 정보편향성과 관련하여 최근 논란이 되는 UGC를 유튜브가 AI 추천 알고리즘을 통해 이용자의 성향에 따라 어떻게 추천하는가를 확인하고자 시도하였다.

(1) 테스트 환경 설정

앞선 선행 연구가 비로그인된 상태에서의 유튜브 AI 추천 알고리즘을 파악했다면 이 실험은 이용자의 연령층이 특정될 수 있도록 가상으로 계정을 만들고, 로그인 후 이용자 개인의 특성을 반영시킨 연구를 진행함으로써 선행 연구의 한계점을 보완하고자 하였다. 유튜브 영상의 경우 사용자가 안드로이드 휴대폰을 사용하기 위해 본인의 구글 계정으로 로그인/가입하게 되면 유튜브 영상을 단 몇 편만 시청하면 별도의 검색 없이 이용자가 시청한 영상과 연관된 추천, 맞춤 동영상 등을 제시하는 기능이 AI 추천 알고리즘을 기반으로 작동한다.

이러한 현상을 제거하기 위하여 본 연구에서는 기존 PC 사용자 정보에 의한 영

향이 없도록 미사용의 새 파이어폭스 브라우저를 설치하고, 기존에 사용하던 브라우저들의 쿠키 및 캐시를 모두 초기화하였다. 두 개의 실험 환경이 갖춰진 컴퓨터로 파이어폭스 브라우저를 사용해 두 개의 구글 계정(곽수철, 김용수)을 만들고, 실제 사용하는 계정처럼 보이기 위해 메일 주소를 영문 이름과 생년월일 순으로 생성하면서, 전화번호나 연락 가능한 메일 등 기존 사용자가 특정될 수 있는 가능성이 존재하는 정보는 제공하지 않는 환경을 구성하였다.

(2) 추천 경향 확인을 위한 테스트 방법

AI 추천 알고리즘을 통한 추천 경향을 확인하기 위하여 실험은 2개의 계정을 각각 언론사의 뉴스만 시청한 1번 계정과 UGC만 시청한 2번 계정으로 나누었다. 총 3가지의 키워드를 검색해서 키워드별로 4개의 영상을 선정해 시청했으며, 이 영상들을 토대로 유튜브 초기 화면으로 돌아가 계정 생성 당시 초기 추천 영상과 실험 시청 이후 추천 영상을 서로 비교해 보았다. 각 계정에서 시청할 영상의 키워드로는 사실의 진위는 정해져 있지만 이를 파악하기까지 상당한 시간이 소요되는 주제인 '김정은 사망설', '리설주 신변 이상', '박원순 사망'을 선택했다. 키워드별로 4개의 영상을 선정했으며, 상단 검색창에 키워드를 검색해 영상을 찾아 시청하는 방식으로 테스트를 진행하였다.

(3) 테스트 결과 및 분석

테스트에 대한 카이제곱 검정(chi-square) 결과 유튜브에서 계정에 따른 뉴스 및 UGC 시청에 따른 유튜브 알고리즘 추천과는 연관성이 있는 것으로 나타났다.

<표 1> 시청에 따른 추천 빈도 및 카이제곱 검정

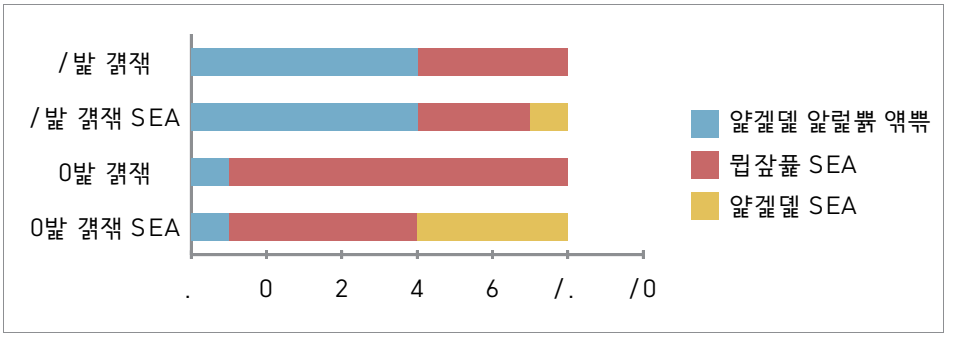
		유튜브 시청[11]		합계[12]
		뉴스 시청	UGC 시청	
추천	뉴스 추천	6	1	**8**
	UGC 추천	4	9	**12**
	합계	**10**	**10**	**20**

카이제곱($\chi 2$): 7.41, 자유도(df): 1, p-value:<0.01(6.63)

11) 시청한 24개의 영상 중 3개의 영상이 비공개 처리되어 현재는 열람이 불가하다. 실험과 실

1번 계정의 경우 언론사 영상과 UGC를 6:4 비율로 추천받았고, 추천받은 UGC 중 키워드와 관련 없는 콘텐츠는 한 개였다. 2번 계정의 경우 위 비율이 9:1로 나타났고, 키워드와 관련 없는 콘텐츠는 총 5개로 UGC 4개, 언론사 콘텐츠 1개가 추천되었다.

<그림 1> 밀접한 UGC와 연관된 UGC의 비율 시각화



우리가 진행한 테스트는 탐색적 접근방법에 불과하지만, 결과를 보면 유튜브의 AI 추천 알고리즘은 이용자가 허위거나 사실관계가 명확하지 않은 UGC 같은 영상을 시청하기 시작하면 계속해서 이런 UGC 위주로 추천함으로써 이용자가 공신력 있는 콘텐츠를 통해 정확한 정보를 받아들이기 어려운 상황에 빠져들 위험성에 쉽게 노출되게 된다. 이러한 결론은 앞서 언급한 한국언론진흥재단의『유튜브 추천 알고리즘과 저널리즘』에서 유튜브 추천 알고리즘이 전통적 언론사를 더 선호한다는 것과는 다른 결론으로, UGC 위주의 정보만을 시청하는 이용자에게 언론의 신뢰성이나 정보의 정확성에 대한 고려보다 단순히 이용자의 취향에 부합하는 흥미 위주의 콘텐츠를 우선하여 추천하고 있다는 합리적 의심을 가지게 한다.

험 결과 수집이 영상이 시청 가능했던 시기에 이뤄졌기에 실험 결과에 유의미한 결과를 가져다주지는 않는 요소로 판단된다.

12) 각 계정 별로 선정된 12개의 영상을 시청한 후, 유튜브 메인 화면에서 추천된 동일/관련 주제의 처음 10개 영상을 결과로 채택해 분석하였다.

Ⅲ. 알고리즘 편향성 규제를 위한 현 정책과 그 한계

1. 규제 필요성에 관한 논의 검토

앞에서 AI 추천 알고리즘은 이용자에 대해 축적한 과거의 데이터를 바탕으로 추천할 정보와 그렇지 않을 정보를 구분하고 이용자의 관심사에 따른 맞춤형 데이터를 추천함으로써 이용자는 필터링된 정보만이 주어지는 '필터버블' 현상이 발생하고 있음을 확인하였다. 이처럼 AI 추천 알고리즘은 사용자에게 편리함을 주는 대신 사용자의 '선택권'을 뺏어간다. 이런 현상은 이용자가 누려야 하는 정보 다양성을 제한하고 개인의 가치관 또는 취향과 유사한 정보 위주의 접근을 유도해 정보 편향성으로 인한 사회 양극화를 초래한다.[13)]

알고리즘 편향성과 차별 등에 관한 문제는 이미 많은 학자들이 제기하고 있다. 영남대학교 양종모(2017, pp.60～105) 교수는 인공지능의 의사결정이 인간의 결정과는 달리 객관적이고 편협하지 않을 것이라는 대다수의 추측과는 달리 인공지능 알고리즘에 의한 의사결정은 인간의 의사결정과 큰 차이 없이 오류 가능성뿐만 아니라 편향성과 차별적 결과를 보인다는 것을 시사했다. 편향성을 가진 인공지능 알고리즘에 의한 의사결정과 그로 인한 차별적 결과는 인공지능 알고리즘이 가져다줄 것이라고 기대했던 공정한 판단과는 달리 오히려 차별금지라는 우리 사회의 중요한 가치를 훼손할 우려가 더 크다고 하면서 알고리즘의 편향성에 관한 문제점을 지적했다.[14)]

이에 대한 원인으로 경남대학교 정원섭 교수의 연구를 눈여겨볼 필요가 있다. 정원섭 교수는 인공지능 알고리즘의 차별적인 의사결정 원인으로 먼저, 데이터 마이닝에서 목표 변수를 정의하는 과정 중 생기는 표본 편향과 배제 편향같이 특정 집단이 과잉/과소평가 되거나 아예 배제될 수 있다는 점; 둘째, 수집한 데이터를 데이터셋으로 재가공하는 레이블링 과정에서 부적절한 결과가 측정 편향과 회상 편향을 초래할 수 있으며; 셋째, 데이터의 특징 선택 단계 중 발생하는 비교 오류

13) 이미 유튜브에서 극우방송을 '틀튜브 "라고 부르고 주로 시청하는 이용자들을 "틀니충으로 부르고 있는 것이 그 대표적인 사례이다: 뉴시스, 홍준표 "틀튜브가 한국 보수 망쳐…안 보면 된다". Retrieved April 26, 2022, https://newsis.com/view/?id=NISX20211218_0001693028&cid.

14) 양종모 (2017). 인공지능 알고리즘의 편향성, 불투명성이 법적 의사결정에 미치는 영향 및 규율 방안.『법조』vol 66, no. 3, 통권 723호 60-105.

를 들었다.[15] 그리고 그 개선을 위해 알고리즘 설계 과정에서 불투명성을 해소하도록 강제하고, 특히 차별적 결과를 보이는 알고리즘의 경우는 작동 과정의 설명의무 부과 등으로 불투명성 문제를 해소하는 노력이 필요하다고 보았다.

사실 투명성이 담보되지 못한 AI 추천 알고리즘이 근거 없이 허위사실에 기반한 콘텐츠를 지속적으로 이용자에게 추천하다 보면 국민 대부분이 의식하지 못하는 사이에 AI 알고리즘이 나에게 추천해 주는 정보가 전부이며 진실한 것으로 받아들이게 될 위험이 있다. 이에 새로운 처벌규정을 신설하여 규제하는 방안도 조심스럽게 제기하는 견해도 있고(이문한, 2021), 자율규제와 미디어 리터러시와 같이 민관 협력적 수단과 유도 행정적 수단을 도입하자는 견해도 제기되어 있다(최승필, 2020). 이처럼 AI 추천 알고리즘에 대해서는 어느 정도의 강제적인 규제가 사용될 수밖에 없는 위험이 내재하여 있다는 것에는 논란의 여지가 없으나 그 정도와 방식을 정함에서는 많은 제약과 한계가 존재한다.

2. 형사처벌을 통한 규제

현행법하에서 사실에 근거하지 않는 정보를 생산하거나 유포하는 UGC 제작자에 대해서 형법상의 규제 또는 기타 비형사적인 규제가 적용될 수 있다. 허위사실에 기반한 UGC가 개인의 명예를 훼손하거나 사회 혼란을 일으키는 경우 형법 제307조 제2항에 따른 명예훼손죄 그 외에 업무방해죄, 위계에 의한 공무집행방해죄와 같이 적용할 수는 있는 형사처벌 규정이 있다. 그러나 허위적인 내용의 UGC의 제작이 우리 헌법 제21조 제4항에서 보호하는 언론·출판의 자유의 보호영역에 벗어나는지와 같은 위헌성 시비로 인하여 형사법상의 규정을 근거로 처벌하기에는 어려움이 있다. 실제로 현행법상 사회적 법익이나 국가적 법익 등 '공공의 이익'을 침해하는 허위사실 등은 선거의 경우에만 적용할 수 있는 공직선거법상 '허위사실공표'나 국가보안법상 '허위사실날조·유포' 정도의 규정만 있을 뿐이다.[16]

형사법적 규제의 특수성을 고려한다면 미디어 콘텐츠로 인해 발생하는 사실 왜곡이나 법익 침해 방지를 위한다는 목적으로 규제하기는 어려울 것으로 판단된다.[17] 정확하게는 AI 추천 알고리즘을 통한 정보 편향성은 아예 형사법적 처벌의

15) 정원섭 (2020). 인공지능 알고리즘의 편향성과 공정성. 『인간·환경·미래』, 2020년 가을 제25호, 55-73.

16) 공직선거법 제250조(허위사실공표죄); 국가보안법 제4조 및 제7조 참조

논의 대상이 되지 않는다고 할 수도 있다. 왜냐하면 형사법적인 처벌규정은 가짜뉴스 등 UGC 제작자나 이를 유포하는 자에게 적용된다. 이런 관점에서 보면 관련 영상을 자동추천하는 알고리즘의 경우 취향에 따른 이용자의 선택문제로 귀결될 가능성이 높다. 의무규정이 아닌 이상 반대되는 영상을 추천하지 않았다는 이유로 형사처벌 대상으로 삼기는 어렵기 때문이다.

3. 행정지도를 통한 규제

이처럼 규제의 필요성은 있지만 형사법적인 처벌은 어렵기 때문에 현재 유도적 행정규제 수단이 대안으로 제시되고 있다. AI 추천 알고리즘 관련 연구들은 공통으로 머신러닝을 통한 알고리즘은 그 작동 과정을 이용자에게 제대로 설명할 수 없고 감독할 수 없다는 불투명성을 원인의 하나로 지목하고 있다. 이에 대해 알고리즘 작동 방식에 대한 설명을 요구할 수 있는 권리를 보장하는 법안 등 제도적 수단의 도입이 고려되었으나 번번이 영업비밀 침해의 논란에 부딪히고 있다.[18)]

결국, 법적 강제수단은 가지지 못하지만, 일종의 정책적 방향을 제시하는 가이드라인이 업계의 자율규제와 병행되는 형식으로 정리되어가는 양상이다. 2021년 방송통신위원회에서는 「인공지능 기반 미디어(매체) 추천 서비스 이용자 보호 기본원칙」을 발표하였다. 여기에는 미디어 제공 플랫폼의 추천 알고리즘에 <3대 핵심 원칙>과 <5대 실행 원칙>을 추천 알고리즘 서비스 제공자가 실현해야 할 원칙으로 제시되어 있다. 3대 핵심 원칙은 2019년에 방송통신위원회가 발표한 '이용자 중심의 지능정보사회를 위한 원칙'에 기초한 특칙이며, ①투명성, ②공정성, ③책무성으로 이루어져 있다.

①투명성은 추천 알고리즘의 작동 과정과 작동에 필요한 정보에 대해 밝히고 설명하는 원칙이다. 이용자가 추천 알고리즘을 받는다는 사실을 인지할 수 있도록 서비스의 내용에 영향을 미치는 주된 요인과 효과를 설명하고, 추천 결과에 대해서 부정적인 요소가 발견되거나 이용자가 불만을 제기하는 경우 해당 결과에 도달하는 과정을 충실히 설명하도록 하는 내용을 담고 있다. ②공정성은 알고리즘의

17) 마찬가지로 정보통신망 이용촉진 및 정보보호 등에 관한 법률 제44조의7의 사람을 비방할 목적의 불법정보의 규제 역시 특별히 개인의 명예를 훼손하거나 비방하는 영상이 아니라면 적용하기 어려운 한계를 가진다.

18) 홍윤지 (2021). '알고리즘 설명요구권' 놓고 영업기밀 침해 논란도,. Retrieved April 26, 2022, https://www.lawtimes.co.kr/Legal-News/Legal-News-View?serial=171566.

편향성 문제에 대해 미디어의 다양성과 추천 알고리즘의 공정성 확보를 위해 서비스 제공자가 조치를 취해야 한다는 원칙이다. 추천 알고리즘 제공자가 추천 시스템의 편향성으로 인해 이용자의 권익 또는 미디어의 다양성이 훼손되지 않도록 콘텐츠 자동 배열의 기준 및 결과의 공정성 확보를 위한 조치를 강구할 것과 공정성 확보를 위한 조치로 이용자의 선택권 보장 및 불만 처리, 추천 서비스의 사전적, 사후적 평가, 지속적 시스템 개선 등 다양한 기술적, 관리적 조치를 들고 있다. ③책무성은 앞서 말한 투명성과 공정성을 추천 알고리즘 제공자가 지켜야 할 의무로 부과시키고, 문제가 발생할 경우 이를 해결해야 할 책임을 부여하는 것을 내용으로 한다. 서비스 제공자에게 투명성과 공정성을 제고하기 위해 기본원칙을 준수할 책무를 부여하고, 추천 서비스 운영 과정에서 기능적 오류, 오작동, 현행 법령 위반 등 부정적인 결과가 발생한 경우 이를 제거 및 시정할 책임과 그와 관련한 이용자의 불만 또는 분쟁을 해결하기 위해 노력해야 함을 정하고 있다.

한편, 5대 실행 원칙은 ①이용자를 위한 정보 공개, ②이용자의 선택권 보장, ③자율 검증 실행, ④불만 처리 및 ⑤분쟁 해결, 내부 규칙 제정을 그 내용으로 한다. 이용자를 위한 '정보 공개'와 '선택권 보장' 원칙은 이용자가 추천 서비스를 제공 받고 있다는 점과 추천 서비스가 이용자의 특정한 정보를 이용해 콘텐츠 자동 배열을 적용한다는 점을 공개해야 함을 밝히고 있으며, 이용자에게 해당 기능의 사용 여부를 정할 수 있는 권리를 부여한다는 내용이다. 자율 검증 실행 원칙을 통해 추천 서비스 제공자가 자율적으로 알고리즘 위험성을 상시 관리할 것과 분쟁 해결 원칙을 통해 이용자 권익 침해에 대한 빠른 해결을 추구하였다.

그 외에도 2021년 윤영찬 위원 등은 알고리즘 안정성의 문제점을 겨냥한 '알고리즘 및 인공지능에 관한 법률안'을 발의한 바 있다. 입법 취지와 내용을 살펴보면 알고리즘과 인공지능의 부정적 영향을 최소화하고 관련 산업을 육성하기 위해 '고위험인공지능'을 소개하며 국민의 생명, 신체의 안전 및 기본권의 보호에 중대한 영향을 미치는 아래 7개와 같은 인공지능을 고위험인공지능으로 특정하고 있다.

<표 2> 고위험인공지능에 해당하는 인공지능의 종류

호	내용
가.	인간의 생명과 관련된 인공지능
나.	생체인식과 관련된 인공지능
다.	교통, 수도, 가스, 난방, 전기 등 주요 사회기반시설의 관리·운용과 관련된 인공지능
라.	채용 등 인사 평가 또는 직무 배치의 결정에 이용되는 인공지능, 응급서비스, 대출 신용평가 등 필수 공공·민간 서비스 관련 인공지능
마.	응급서비스, 대출 신용평가 등 필수 공공·민간 서비스 관련 인공지능
바.	수사 및 기소 등 기본권을 침해할 수 있는 국가기관의 권한 행사에 이용되는 인공지능
사.	문서의 진위 확인, 위험평가 등 이민, 망명 및 출입국관리와 관련된 인공지능

출처: “알고리즘 및 인공지능에 관한 법률안” (5~6쪽을 재구성)

동 법안은 고위험인공지능을 개발 및 이용하는 과정에서 국민의 생명과 안전을 보호하고, 고위험인공지능과 알고리즘에 관한 기본원칙 및 정책 수립 등에 관한 사항을 심의 의결하기 위하여 국무총리 소속으로 고위험인공지능심의위원회를 두도록 하고 있다(안 제15조). 그 외에 고위험인공지능을 이용한 기술 또는 서비스에 대한 설명요구권, 이의제기권, 거부권 등 고위험인공지능 이용자를 보호하고(안 제19조), 이용자가 고위험인공지능의 기술 또는 서비스를 이용함에 있어 손해를 입으면 해당 고위험인공지능사업자에게 손해배상을 청구할 수 있도록 하면서(안 제20조), 분쟁조정을 위해 알고리즘 및 인공지능 분쟁조정위원회를 두도록 하는 등(안 제23조) 인공지능과 알고리즘으로 인한 피해 축소와 분쟁 해결과 같이 향후 발생 가능성이 높은 문제에 대응하려는 노력을 볼 수 있다.

최근에는 AI 추천 서비스를 직접적인 대상으로 하고 알고리즘 편향을 고려한 여러 법안이 논의되고 있다. 그 중 대표적인 법안으로는 전혜숙 의원이 대표 발의한 ‘온라인 플랫폼 이용자 보호에 관한 법률안’이 존재한다.[19] 이 발의안은 총 14가지의 주요 내용을 가지고 있는데 이 중에서 눈여겨 볼만한 항목은 ‘대규모 온라인 플랫폼 사업자는 콘텐츠 등의 노출 방식 및 노출 순서를 결정하는 기준을 공개하여야 한다’는 것이다. 이용자의 화면에 콘텐츠를 내보낼 때 어떠한 요소들을 고려하였는지, 그 방식이 무엇인지, 어떤 순서로 보여줄 것인지를 결정하는 기준을 명확하게 제공해야 한다는 것이다. 이와 같은 정보를 통해 이용자는 어떤 정보가 알고리즘 형성에 영향을 미쳤는지 알 수 있으며 알고리즘 편향을 인한 부작용을 감소

19) https://likms.assembly.go.kr/bill/billDetail.do?billId=PRC_F2L0A1F1D2T7J1W6Y5I1A0V1R6U9T2

시킬 수 있을 것으로 예상된다. 그러나 아직 기존 법률의 효과에 대한 분석이 부재한 상황에서 신규 규제를 도입하는 것은 바람직하지 않기 때문에 이에 대해서는 더 신중한 논의가 필요하다는 의견이 다수이어서 현재 이 안은 보류된 상태이다.

이상의 내용을 정리하자면 결국 현행법상의 한계로 인해 AI 추천 알고리즘 자체를 대상으로 하는 법적 규제는 실질적으로 불가능해 보인다. 만일 콘텐츠의 내용이 문제가 되는 경우라면 그 자체를 처벌 대상으로 하는 방법은 있으나 이러한 방식은 헌법상 표현의 자유를 지나치게 제한할 뿐 아니라 여전히 AI 알고리즘으로 인한 문제를 해결하지는 못한다. 나아가 AI 알고리즘을 규제하기 위한 새로운 법을 제정하는 것은 예측하기 어려운 과학기술의 발전과 혁신에 지대한 장애를 가져올 것이 우려된다. 아직 이러한 AI 알고리즘의 활용으로 어떠한 법률문제가 발생할지 정확하게 예상하기 어려운 상황인데 법익 침해의 우려만으로 법을 제정하는 우를 범해서는 안 될 것이기 때문이다.

결국, 혁신을 방해하지 않는 동시에 앞으로 기술이 나아가야 하는 방향을 제시할 수 있는 가이드라인이나 지침을 통해 AI 추천 알고리즘을 사용할 때 고려되어야 할 원칙과 기준을 제시하여 규제하는 방안이 가장 이상적이라 판단된다. 방송통신위원회에서 관련한 가이드라인이 제정된 바 있으나, 동 가이드라인은 이용자의 추천 서비스 이용 선택, 변경 기능과 자율 검증 실행 원칙에서 콘텐츠 유형, 소요 시간, 비용, 가용 기술 등을 고려해 합리적으로 실행 가능한 범위 내에서만 제공하도록 하고 있어 기업의 규모와 가용 기술에 따라서 선택적으로 제시하거나 제시하지 않을 수 있는 빈틈이 존재한다. 더구나 어떤 콘텐츠 유형에 따라 선택, 변경 기능을 제공하는지에 대해 구체적으로 밝히지 못하고 있어 명확성이 부족하다는 점도 한계로 작용하고 있다. AI 추천 알고리즘으로 인한 문제는 우리나라뿐만 아니라 전 세계가 당면한 과제이며 나라마다 다각적인 방면에서 해결책을 모색하고 있는 상황이므로, 이하에서는 해외의 AI 알고리즘 규제 현황을 살펴봄으로써 앞으로 우리나라는 어떤 방향으로 나아가야 하는지에 대한 시사점을 도출해보고자 한다.

Ⅳ. 해외규제 사례 및 시사점

1. 미국

AI 알고리즘 문제와 관련하여 미국은 초기에는 주로 AI 알고리즘을 활용한 기업의 부당거래를 연방거래위원회(Federal Trade Commission, FTC)의 조사대상으로 삼아왔다. 2018년 11월 FTC는 공청회를 열어 '알고리즘, AI, 또는 예측 분석에 대한 장단점, 기술발전 가능성 및 위험성'에 대하여 논하였다.[20] 2019년 6월에는 AI가 경쟁에 어떠한 영향을 미칠 수 있는가(Reduced Demand Uncertainty and the Sustainability of Collusion: How AI Could Affect Competition)에 대한 조사보고서를 발표하면서 불확실성을 제거하는 AI 기술이 어떤 방식으로 조정된 행동의 특성과 보급을 변화시키는지에 대해 다루기도 하였다(Wilson, 2019, pp.1～36).

2020년에는 그동안 FTC가 인공지능과 알고리즘에 관해 발표했던 모든 자료를 토대로 하여 "인공지능 및 알고리즘 사용(Using Artificial Intelligence and Algorithms)" 가이드라인을 발표한 바 있다(Smith, 2020). 이 가이드라인은 기업이 인공지능 시스템을 사용할 때 따를 다섯 가지 행동 모델을 권고하였는데 그 내용은 다음과 같다. 기업이 사용 혹은 사용할 인공지능 및 알고리즘이 1) 투명하고, 2) 사용자에게 설명할 수 있어야 하며, 3) 그 결과가 공정한 결정임을 입증할 수 있어야 하고, 4) 데이터와 모델이 실증적으로 타당하며, 5) 기타 발생할 수 있는 문제에 대해서 책임질 수 있어야 한다고 말하고 있다. 이러한 권고사항은 기업활동의 투명성과 공정성에 관한 것이라는 점에서 정보의 편향성을 직접적인 대상으로 삼고 있지는 않았다.

그러나 미국은 AI 및 알고리즘이 제공하는 자동 의사결정에 무의식적인 편견과 불공정한 결과가 존재한다는 것을 다수의 실례를 경험하면서 확인한 바 있다. 한 예로 2019년 페이스북은 광고주들이 성별, 인종, 종교 등에 따라 의도적으로 광고를 타기팅(targeting)할 수 있도록 허용한 적이 있다. 그러자 간호 또는 비서 업무에서의 구인 광고에서는 여성의 우선순위가 높았으나 청소부 및 택시 운전사에

20) Federal Trade Commision (2018). Hearing #7: The Competition and Consumer Protection Issues of Algorithms, Artificial Intelligence, and Predictive Analytics. Retrieved April 26, 2022, https://www.ftc.gov/news-events/events/2018/11/ftc-hearing-7-competition-consumer-protection-issues-algorithms-artificial-intelligence-predictive.

대한 구인 광고는 남성 특히 소수 민족 출신의 남성들에게 우선하여 표시되는 문제가 발생하게 되었고, 이후 페이스북은 인종, 성별, 종교를 타기팅(targeting)하는 것을 허용하지 않게 되었다(Martinez, 2019). 이런 경험들을 바탕으로 2021년 4월 “AI 활용에서 지향할 진실성, 공정성과 평등, 형평성(Aiming for truth, fairness, and equality in your company’s use of AI)”을 발표하였다(Jillson, 2021).

새로 발표한 가이드라인은 알고리즘 사용에 따른 편향성을 인지하고 이러한 문제점을 제거할 수 있는 AI 알고리즘의 활용 방향을 구체적으로 제시하고 있다는 점에서 의미가 있다. 새 지침은 총 7가지 사안에 대해서 1) 올바른 기준점에서 시작하여 불평등한 결과가 나올 가능성을 줄이고, 2) 알고리즘을 사용하기 전과 후를 주기적으로 비교하여 차별적인 결과가 나오지 않도록 해야 할 것, 3) 데이터와 소스코드를 외부에 개방하는 방식을 통해 투명성과 독립성을 유지하도록 해야 한다고 하고 있다. 또한, 4) 알고리즘이 편향되지 않고 언제나 공정한 결과를 낸다고 선언하는 것과 같이 알고리즘의 결과에 대해서 과장하지 않는 것에 대한 중요성도 강조하고 있다. 5) 데이터를 어떠한 방식으로 얻었는지, 어떻게 사용하는지에 대해 진실성을 유지할 것과 6) 공정한 알고리즘을 만들어 알고리즘 자체에 실보다 득이 크도록 할 것을 제안했다. 마지막으로 7) AI 개발자는 알고리즘의 성능에 대해서 책임을 져야 함을 규정하고 있다.

이 가이드라인은 FTC가 가지는 기업조사 권한과 결합해서 보면 단순한 권고 이상의 기능을 가지는 것으로 해석된다. 지난 수십 년 동안 FTC는 AI 개발자와 사용자에게 중요한 세 가지 법을 적용해왔다. 우선 연방거래위원회법 제5조(Section 5 of the FTC Act)는 불공정하거나 기만적인 관행을 금지하고 있는데 여기에는 인종적으로 편향된 알고리즘의 판매 또는 사용이 포함된다. 두 번째, 공정신용보고법(Fair Credit Reporting Act, FCRA)은 알고리즘이 사람들의 고용, 주택, 신용, 보험 또는 기타 혜택을 거부하기 위해 사용되는 특정 상황에 적용될 수 있다. 마지막으로, 신용기회평등법(Equal Credit Opportunity Act, ECOA)은 기업이 인종, 피부색, 종교, 출신국, 성별, 혼인 여부, 나이, 또는 공적 지원 여부 때문에 신용 차별을 초래하는 편향된 알고리즘을 사용하는 것을 불법으로 규정하고 있다. 이처럼 AI 알고리즘으로 인해 보호 계층의 신용에 대한 차별이 발생한다면 FTC법 및 신용기회평등법 위반으로 규제가 가능하다.

2021년 새로 발표된 가이드라인은 기존의 AI 관련 가이드라인과 달리 공정성을 유지하는 것과 차별 방지에 큰 비중을 두고 있다. 동 가이드라인은 FTC의 기업조

사라는 본연의 막강한 기능과 결합함으로써 권고적인 기능 이상의 효과를 가지게 되는데, 만일 알고리즘이 특정한 인종, 성별, 종교 등을 가진 이용자층을 타깃으로 하여 이용자들에게 상당한 피해를 줄 가능성이 있는 경우라면 언제든지 FTC의 개입이 가능하다는 점에서 큰 의미가 있다. 요약하면, 미국의 경우 AI 개발자에 대한 직접적인 규제보다는 개발된 AI가 이용자에게 어떠한 부정적인 영향을 끼칠 수 있는지에 대하여 FTC가 지속적으로 모니터링하고 있으며 이를 방지하기 위한 구체적인 가이드라인을 제시했다는 점에서 시사하는 바가 있다.

2. 유럽

2021년 4월 EU 집행위원회(European commission)는 인공지능으로 인해 이전까지 존재하지 않던 위험과 부정적 결과가 초래될 수 있다고 판단하고 신뢰할 수 있는 AI를 위한 법적 체계를 다루기 위해서 AI 규제법안을 발표하였다(European Commission, 2021).[21] 유럽 연합의 AI 규제법안의 경우 현재 신규 입법을 준비 중이며 발전하는 기술에 따라 계속된 수정이 필요하다고 보고 있기 때문에 4월에 제시한 법안은 초안에 불과할 뿐이다. 그러나 향후 입법이 된다면 EU 시장에 진출해 있는 우리나라 기업에도 영향을 미칠 수 있으며 AI와 관련한 법안을 준비하는 다른 국가들이 참고할만한 모델이 될 수 있기 때문에 주의 깊게 살펴볼 필요가 있다.

이 규제안의 목표는 다음과 같다. 1) EU 시장에 출시되어 사용되고 있는 AI 시스템의 안전과 기본권 및 EU 가치에 관해 현존하는 법을 존중하도록 보장한다. 2) AI에 대한 투자 및 혁신 지원을 위해 법적 확실성을 보장한다. 3) AI 시스템에 적용할 수 있는 기본권 및 안전 요건에 대한 기존 법률의 거버넌스를 강화하고 효과적으로 시행한다. 4) 합법적이고 안전하며 신뢰할 수 있는 AI 애플리케이션을 위한 단일 시장의 개발을 촉진하고 시장 분할을 방지한다.

EU 집행위원회는 위 네 가지 목표를 달성하기 위한 구체적인 규제방안을 8가지

21) “Regulation of the European Parliament and of Council: Laying down harmonised rules on artificial intelligence(Artificial Intelligence Act) and amending certain union legislative acts”: European Commission (2021). Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS. Retrieved April 26, 2022, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206.

로 구분하여 제시하였다. 첫째, AI 시스템의 서비스 및 사용에 적용되는 새로운 규칙의 주제와 적용 범위를 정의한다. 둘째, 금지된 AI 목록을 설정한다. 셋째, AI를 위험 수준에 따라 분류하고 이를 바탕으로 위험 관리를 하는 방안을 제시한다. 넷째, 특정 AI 시스템이 제기하는 조작의 특정 위험을 고려해야 하며 이때 시스템은 투명하게 공개되어야 한다. 다섯째, 혁신 친화적이고, 미래 보장적이며 탄력적인 법적 체계를 만드는 데 기여한다. 여섯째, 거버넌스 시스템 구축 및 EU 전체 데이터베이스 구축을 통해 모니터링 작업을 용이하게 하고, AI 관련 사고 및 오작동에 대한 사후 모니터링 및 의무를 규정한다. 일곱째, 비위험 AI 시스템 제공자가 고위험 AI 시스템 제공자에 대한 의무 요구사항을 자발적으로 적응하도록 유도하는 것을 목표로 하는 프레임워크를 만든다. 여덟 번째, 정보와 데이터의 기밀성 존중 및 위임과 실행 권한 행사 규칙을 규정한다.

본 연구 방향과 관련하여 이 중에서 주목해야 할 항목은 AI를 위험 수준에 따라 분류하는 점이다. 규제안에 따르면 위험 수준을 용인할 수 없는 위험 수준(unacceptable risk), 고위험 수준(high risk), 낮은 위험 수준(low or minimal)으로 구분하고 있다.

용인할 수 없는 위험 수준으로 분류되는 AI 시스템에는 기본권 또는 EU 공공의 이익을 침해하는 AI 시스템이 포함된다. 구체적으로는 알지 못하는 사이에 해를 끼치는 조작적 또는 착취적 시스템, 법 집행을 위해 공공장소에서 사용되는 실시간 원격 생체 인식 시스템, 그리고 사회적 행동이나 예측된 성격적 특징에 기초하여 개인의 신뢰도를 평가(social scoring)하는 AI 시스템이나 기술이 용인할 수 없는 위험 수준의 AI 시스템에 해당된다. 이러한 용인할 수 없는 위험 수준의 AI 시스템은 사용이 금지된다.

고위험 수준의 AI 시스템은 자연인(natural persons)의 건강과 안전 또는 기본권에 위험을 초래할 수 있는 AI 시스템을 말하며, 1) 제3자로부터 적합성 평가가 요구되거나 2) 생체인식 및 분류, 교육 및 직업훈련, 법 집행 등에서 사용되는 AI 시스템을 의미한다. 고위험 수준(high-risk)으로 분류된 AI 시스템은 특정 의무 요건 준주 및 사전 적합성 평가를 거친 후 사용이 허용된다. 충족시켜야 할 7가지 특정 의무 요건에는 1) 위험관리 시스템 구축(Risk management system), 2) 데이터 거버넌스 수행(Data and data governance) 3) 기술 문서화(Technical documentation), 4) 기록 보관(Record-keeping), 5) 이용자에게 투명성 및 정보 제공(Transparency and provision of information to users) 6) 사람에 의한 감독(Human oversight) 7) 정확성,

견고성, 그리고, 사이버보안(Accuracy, robustness and cybersecurity)이 해당된다. 한편, 낮은 위험 수준의 AI 시스템은 고위험 수준의 AI 시스템과 같이 요구사항을 충족할 필요는 없지만, 자발적인 행동 지침(code of conducts)이 권고사항으로 제안된다.

유럽연합위원회가 제시한 매우 구체적이고 상세한 법률안에 관하여 관련 업계에서는 AI 규제안의 마련에 대해서는 반기는 입장이지만 기본권 보호와 공공의 안전 사이의 균형에 대해서는 더 깊은 논의가 필요하다고 평가하고 있다(European Institute of Public Administration, 2021). EU의 AI 규제안에 대해서 디지털 유럽의 사무총장인 쎄실리아 보나펠드 데일(Cecilia Bonefeld-Dahl)은 고위험의 정의가 아직 불분명하기 때문에 이 부분을 명확하게 짚고 넘어가야 추후에 혼란으로 인한 문제가 발생하지 않을 것이라고 했으며 앞서 언급한 바와 같이 중소기업과 스타트업을 지원할 수 있는 방안을 마련해야 한다고 주장한 것이 그 대표적인 예이다(Bonefeld-Dahl, 2021).[22)]

유럽에서 발표한 규제안은 비록 업계가 전적으로 동의하지 않고 있긴 하지만 미국과 비교하면 국제적으로 적용이 가능한 공통 기준을 제시하고 있으며 AI 기술 활성화에 따른 지속 가능한 법적 체계와 그에 따른 문제점까지 고려하고 있다는 점에서 더 현실적인 방안이라고 평가할 수 있다.

3. 영국

2021년 11월 영국 정부는 AI 애플리케이션의 편견을 없애는 것을 목표로 하는 알고리즘 투명성 기준(Algorithmic Transparency Standard)을 발표했다(Central Digital and Data Office, 2021a). 이 기준은 영국 정부의 국무조정실 중앙 디지털 데이터오피스(Cabinet Office's Central Digital and Data Office, CDDO)와 데이터윤리혁신센터(Centre for Data Ethics and Innovation, CDEI)가 공동으로 제시한 것이다.[23)]

22) 또한, 광범위한 기술 기업을 대표하는 무역 기구인 컴퓨터 통신 산업 협회(Computer & Communications Industry Association, CCIA)는 AI 시스템을 위험 수준에 따라 분류하여 규제한다는 점에서는 반기고 있지만 불필요한 형식적 절차를 피하기 위해서는 규제가 더 명확해질 필요가 있다고 언급했다: Kahn, J. (2021). Europe proposes strict A.I. regulation likely to have an impact around the world. Retrieved April 26, 2022, https://fortune.com/2021/04/21/europe-artificial-intelligence-regulation-global-impact-google-facebook-ibm/.

23) 현재 발표한 기준은 최종적으로 발표된 기준이 아니며 시범 체계를 거쳐 취약점과 문제점

영국의 ‘알고리즘 투명성 기준’에 따르면 알고리즘 투명성이란 알고리즘 툴이 사용자의 의사결정을 어떻게 지원하는지를 투명하게 공개하는 것을 의미한다. 여기에는 알고리즘 툴과 의사결정 지원 과정에 대한 정보를 완전하고, 개방적이며, 이해하기 쉽고, 쉽게 접근할 수 있는 자유로운 형식으로 제공하는 것이 포함된다. 영국 정부는 공공 기관이 사용하는 알고리즘 툴과 왜 이러한 툴을 사용하는지에 대한 명확한 정보를 제공하는 데에 도움을 주고자 한다는 점을 이 기준제시의 기본 취지임을 명시하였다.

현재 영국 정부는 이 기준에 따라 공공 부문 기관들이 사용하고 있는 알고리즘 툴과 그 툴을 사용하고 있는 이유에 대해서 양식(template)을 마련해 두고 이를 작성하도록 하여 ‘알고리즘 투명성 기준’을 발전시키는 데에 필요한 정보를 수집하고 있다. 좀 더 구체적으로 살펴보면, 1) 기관에서 사용하고 있는 툴이 아래와 같은 정부 기준에 부합하는지를 판단한다(Central Digital and Data Office, 2021a). 정부 기준에 따르면 챗봇과 같이 사용자와 직접적으로 접촉하는지를 고려하고, 세부사항으로 딥러닝과 같은 머신러닝을 사용하는지, 개인 또는 인구에 법적, 경제적 또는 유사한 영향을 미칠 수 있는지, 인간의 의사결정을 대체하거나 결정에 도움을 주는지를 고려해야 한다. 2) 위 기준에 부합할 경우 이메일로 작성한 양식을 제출하는 방식으로 데이터를 수집하고 있다. 이렇게 제출된 양식은 알고리즘 툴에 대한 가장 기본적인 정보인 알고리즘 툴의 목적과 이유에 대한 정보를 제공하고, 알고리즘 툴의 소유자와 책임자가 누구인지, 그 용도는 무엇인지의 세부적인 요소에 대한 정보를 제공하게 된다. 그 밖에도 알고리즘을 개발하는 단계에서 사용한 데이터의 정보를 제공하고, 영향 평가 및 발생할 수 있는 위험에 대해서도 이 기준에서 언급하고 있다(Central Digital and Data Office, 2021b).

이처럼 정보 편향성을 극복하기 위한 구체적인 내용을 제시하고 있는 영국의 ‘알고리즘 투명성 기준’은 아직 초기 파일럿 단계로 권고수준에 머물러 있어 강제성을 가진 절차라기 보기 어렵다. 그러나 공공 부문의 기관으로 하여금 자체적인 참여를 유도하여 AI 시스템 관련 정보를 제공받고, 이를 통해 개선점을 찾아내 발전시키겠다는 점에서 완성된 형식의 규제안을 발표한 다른 나라들과 차별성을 가진다고 할 수 있겠다. 이와 같은 영국의 행보에 유럽중앙은행(European Central Bank, ECB)도 ‘전체(Union)’의 가치에 부합하는 신뢰할 수 있는 인공지능의 개발

을 파악해 개선할 예정이다.

과 마케팅, 이용을 위한 획일적인 법적 틀을 마련함으로써 내부 시장의 기능을 향상시킬 수 있다고 긍정적으로 보고 있다(ECB, 2021).

4. 해외 규제 사례의 시사점 분석

본 연구에서 진행한 실험을 포함한 다양한 연구결과는 유튜브 등의 다양한 미디어 콘텐츠 플랫폼 AI 알고리즘이 편향성의 문제를 보여주고 있다. 편향성이 발생하는 원인은 다양하다. 알고리즘은 인간이 설계한 프로그램이기 때문에 의도치 않은 개발자의 생각이 반영될 수 있고, 데이터 학습에 사용되는 데이터의 질적・양적 문제로 편향성이 발생할 수 있다. 또 당사자가 원하지 않더라도 제3자의 압력이나 외부 요인이 개입될 가능성도 있다. 이러한 우려에 기인해 각 국가에서는 알고리즘에 의한 편향을 예방하기 위한 노력을 하고 있으며 이에 따라 규제 및 가이드라인을 마련하고 있다.

국가는 사회의 여론 다양성을 보장하고, 생활의 모든 측면에서 발생하는 알고리즘으로 인한 침해로부터 국민을 보호하고 사회적 안정을 도모할 의무를 지고 있다. 따라서 국가는 AI 추천 알고리즘의 부작용에 대응하는 조치를 취하여야 한다. 이를 위하여 전 세계적으로 다양한 다국적 플랫폼의 AI 알고리즘이 국경을 넘어 사회 전역에서 이용되고 있고 이에 대한 통일적이고 일관된 규제 원칙이 필요하다는 점에서 적정 수준에서의 통일된 기준 적용 가능성과 구체적인 차이점을 살펴보는 것은 의미가 있다.

먼저 각국의 내용 유사성 측면에서 비교하면, 미국은 알고리즘 시스템의 활용에 있어 의도하지 않은 편견과 차별이 존재할 수 있다는 것을 확인하였다. 이러한 경험으로 “AI 활용에서 지향할 진실성, 공정성과 평등, 형평성”이라는 지침을 발표하여 이러한 문제점을 보완하기 위한 발판을 마련했다. 유럽은 특정 AI 시스템의 경우 조작의 위험성이 있으므로 시스템을 투명하게 공개하는 것을 우선으로 삼고 있는 것으로 보인다. 특히 영국의 ‘알고리즘 투명성 기준’에 따라 인간의 의사결정을 대체하거나 결정에 도움을 주는 AI 시스템을 사용하는 경우 정해진 양식(template)으로 알고리즘 툴에 대한 정보를 작성하도록 하여 사용자의 관점에서도 무분별한 AI 사용에 대한 책임을 고려하도록 하고 있다는 것은 주목할 만하다. 이러한 내용적 측면은 투명성, 공정성 및 책임성의 3대 기본원칙을 제시한 우리나라 방송통신위원회의 가이드라인에서도 찾아볼 수 있다. 이 점에서 전 세계적으로 AI

알고리즘으로 인한 정보편향 위험성을 해결하는 방안으로 투명성과 공정성을 제시하고 있음을 알 수 있다.

한편, 해외 규제에 대한 차이점과 관련하여 미국의 경우 AI 알고리즘에 대한 새 가이드라인은 기업보다는 이용자 편의에 더 중점을 두고 있다는 점이 다른 국가와 차이가 있다. 유럽의 경우 AI 알고리즘의 위험성을 수준별로 분류하여 AI 기술 발달에 따라 지속적으로 적용이 가능한 법체계 형성이라는 현실적인 방안을 제시하고 있다. 영국은 공공 기관의 자체적인 참여를 유도하여 AI 시스템 관련 정보를 제공받고 이를 통해 개선점을 찾은 후에 법체계를 만들겠다는 점에서 완성된 형식의 규제안을 발표한 다른 나라들과 차별성을 가진다고 할 수 있겠다.

구체적으로 비교해 보면 먼저, 미국의 경우 알고리즘의 결과에 대해서 과장하지 않는 것, 이용자들을 차별하거나 피해를 입힐 가능성이 있는 경우 이전에 존재하는 법안을 활용해 연방거래위원회인 FTC의 개입이 가능하다는 점 등은 단순한 권고적인 가이드라인의 수준을 넘어서 사실상 강제력을 가지는 조치로 평가된다.

또한, 유럽이 AI를 위험 수준별로 관리하는 방법은 2021년 윤영찬 의원 등이 발의한 '알고리즘 및 인공지능에 관한 법률안'에 상당한 영향을 미쳤던 것으로 보인다. 동 법률안을 살펴보면 앞에서 설명한 7가지의 항목에 해당하는 AI 시스템을 '고위험인공지능'이라고 특정하고 있으며 이 '고위험인공지능'으로 인한 피해를 예방하기 위해 고위험인공지능심의위원회를 만들어 국무총리 소속으로 둘 것과 고위험인공지능의 기술 및 서비스를 사용하다가 피해를 입을 경우 보상받을 수 있도록 하는 등 다양한 방안이 제시되어 있다. 이러한 방식이 유지된다면 우리도 유럽과 유사하게 AI 시스템을 위험 수준별로 분리하여 필요한 최소한의 규제체계를 마련할 수 있을 뿐만 아니라 다국적 기업에 대한 공조체계 형성에도 긍정적인 영향을 줄 수 있으리라 생각된다.

마지막으로 영국은 공공 부문 기관으로 하여금 현재 기관들이 사용하고 있는 알고리즘 툴과 사용 이유에 대해서 정해진 양식을 작성하여 제출하도록 하고 있음을 보았다. 모든 기관들이 제출해야 하는 것은 아니며 앞서 3장에서 언급한 정부 기준에 부합할 경우에만 양식을 작성하도록 권고하고 있다. 기관들의 자체적인 참여를 유도한다는 점에서 의미가 있으며 이렇게 수집한 정보를 활용해서 개선점을 찾고, 발전된 가이드라인 또는 규제안을 만들 것이라는 점에서 다른 나라와의 차별성을 가진다. 우리나라의 경우 영국과 같이 정해진 양식을 제출하는 등의 방안은 제시된 바 없으나 알고리즘의 투명성을 유지하고, 작동 과정과 방식을 일반

대중에게 공개할 필요가 있다는 것을 강조하는 점에서 영국의 기준과 목적은 유사하다고 볼 수 있다.

위에 제시한 해외 AI 관련 가이드라인 및 규제안들은 '이용자 중심' , 'AI 알고리즘의 투명한 공개' 등 상당한 부분에서 공통적인 내용을 찾아볼 수 있으며, 또한 비록 규범력에서 차이는 발견되고 있으나 규제안 제정에 이미 상호 영향을 미치고 있음을 볼 수 있다. 예를 들어 방송통신위원회에서 발표한 원칙들은 미국과 같이 이용자 중심적인 관점에서 알고리즘에 대한 규제를 고려한 것으로 평가받고 있는데, 이러한 공통적인 요인들은 앞으로 우리나라 가이드라인이 나아가야 할 방향을 제시하고자 할 때 중점적으로 고려되어야 할 요소로 판단된다.

Ⅴ. AI 추천 알고리즘에 대한 규제 방향

1. 「인공지능 기반 미디어(매체) 추천 서비스 이용자 보호 기본원칙」의 구체화

방송통신위원회는 2021년 6월 「인공지능 기반 미디어(매체) 추천 서비스 이용자 보호 기본원칙」을 발표하였다. 동 가이드라인은 <3대 핵심 원칙>과 <5대 실행 원칙>을 제시해 추천 알고리즘이 가지고 있는 편향성과 선택권 문제 등의 위험성으로부터 이용자 보호를 주목적으로 하고 있다. 방송통신위원회가 발표한 가이드라인은 동일하게 추천 알고리즘에 편향성이 존재할 수 있고 이에 따라 이용자 피해가 발생할 수 있다는 것을 전제로 하고 있으며 알고리즘으로 인한 여러 가지 부작용을 완화할 것으로 기대된다.

다만, 방송통신위원회의 「인공지능 기반 미디어(매체) 추천 서비스 이용자 보호 기본원칙」은 추천 알고리즘으로 인해 발생할 수 있는 위험성을 충분히 반영하지 못하고 있어 각국의 사례를 참고하여 불충분하게 제시되었던 부분을 최대한 구체적이고 명확한 체계를 제시해 줄 필요가 있다. 예를 들어, 가이드라인은 알고리즘 투명성과 이용자의 선택권에 기업의 규모와 기술력에 따라서 합리적인 선으로 따라야 한다고 하지만 스타트업과 같은 신생 기업의 경우 기술력이 완전하지 않아 자체적인 판단이 어려우며, 자율 검증 부서 등을 상설하기에 충분한 자본을 갖고 있지 않은 경우가 많기 때문에 합리적인 기준이 명확하게 제시되어야 한다.

가이드라인은 특히 미디어 제공 플랫폼의 AI 추천 알고리즘이 여론 양극화 현상을 야기할 수 있다는 문제에 대한 규제 필요성을 충분히 설명하지 못하고 있는데, 이렇게 여론 형성과 밀접한 관련이 있는 미디어 제공 플랫폼의 AI 추천 알고리즘이 가져오는 부작용에 대한 불충분한 설명은 오히려 AI 사업 운영에 정부가 과도한 개입을 시도하고 있다는 우려를 가져온다(AI 기반 추천 서비스 이용자 보호를 위한 기본원칙(안) 공개 토론회, 2021). 그 외에도 알고리즘의 사용으로 발생할 수 있는 사회적 위험성을 국민에게 알리는 다양한 방안을 마련하고, 추천 알고리즘을 프로그래밍하거나 데이터를 레이블링하는 전문가 교육을 통하여 편향이 배제된 공정한 데이터 가공을 위한 교육 등의 구체적인 정책도 함께 제시될 필요가 있다.

2. 이용자 선택권 강화

이용자에게는 이용자 자신이 사용하는 AI 추천 알고리즘으로부터 스스로를 보호할 수 있는 수단이 주어져야 한다. AI 추천 알고리즘을 사용하는 경우 이용자에게 자신의 취향으로 추정되는 맞춤형 추천 알고리즘을 사용할 것인지, 아니면 무수하고 다양한 정보를 보고 스스로 필터링할 것인지에 대한 선택권을 제공할 필요가 있다. 핸드폰의 예를 들어, 애플은 자사 계정(Apple ID)이 있는 사용자에게 자사 제품을 이용할 때 개인정보 수집을 통해서 맞춤형 광고를 제시할 수 있음을 알리고 해당 기능의 사용 여부를 사용자가 정할 수 있도록 한다. 애플은 이를 ‘Apple이 광고를 제공하는 방식 제어하기’라고 부르는데(Apple Support, n.d.), 이와 같이 AI 추천 알고리즘의 사용 여부를 이용자가 선택할 수 있는 권한으로 보장한다면 이용자는 자신이 목적하는 바에 따라 다양한 관점에서의 정보를 다양한 출처를 통해 제공받을 수 있으며, 이를 통해 블랙박스 같은 AI 알고리즘이 특정 콘텐츠만을 추천하는 것을 원천적으로 방지하고 사회 여론 양극화 현상을 완화할 수 있을 것이다. 이외에도 제공되거나 제시되지 않고 있는 정보의 종류와 범위를 이용자에게 고지할 수 있는 방법은 다양하다.

더 나아가 이용자 선택권 강화를 위해서는 「인공지능 기반 미디어(매체) 추천 서비스 이용자 보호 기본원칙」에서 제시하고 있는 이용자의 추천 알고리즘 사용 선택권을 합리적인 선 정도가 아니라 명확하게 원칙화하도록 할 필요성이 있다. 방송통신위원회의 「온라인 맞춤형 광고 개인정보보호 가이드라인」이나 개인정보

보호법도 이용자의 선택권 보호를 위해 구체적인 동의를 구하는 방식을 안내하고 있다. 이용자가 특정 기록을 추천 알고리즘의 목록에서 제외할 수 있거나 재가공되지 않도록 선택할 수 있는 방법과 같이 이용자에게도 추천 알고리즘을 제어할 수 있는 기능을 선택할 수 있도록 한다면 AI 알고리즘의 무분별한 콘텐츠 추천이나 이로 인해 제기되는 편향성 문제에 대한 우려를 완화시키고, AI 기술의 적용을 확대시킬 수 있는 계기로 삼을 수 있을 것이다.

Ⅵ. 결론

AI 추천 알고리즘은 유튜브 외에도 인스타그램, 페이스북, 트위터, 틱톡, 네이버, 다음(Daum) 등 대다수 이용자가 사용하는 다양한 플랫폼에 상시 작동하고 있다. 미디어 플랫폼을 포함한 서비스 제공자는 AI 추천 알고리즘을 기반으로 이용자의 취향에 부합할 만한 콘텐츠를 제공하지만 이용자는 그 알고리즘의 구동 방식이나 작동 과정은 알지 못한 채 주어진 결과가 정보의 전부라고 간주하게 만드는 경향이 있다. 다양한 연령층의 이용자를 가진 유튜브의 추천 알고리즘에 대한 본 연구의 실험은 알고리즘이 자체적으로 분석한 이용자의 '취향'에 적합하지 않은 정보는 아예 제공하지 않게 되어 이용자에게 노출되는 정보의 범위 자체를 축소시킬 수 있음을 보여주었다.

알고리즘 작동원리를 알 수 없는 이러한 실태는 AI 추천 알고리즘이 조작될 가능성과 정치적 목적을 위한 선거 개입이나 특정 여론 형성을 위해 악의적으로 사용된다 하더라도 그 직접적인 원인을 밝힐 수 없다는 문제점도 보여준다. AI 추천 알고리즘이 사회적 관계 형성이나 공동목표를 위한 협력에 큰 영향을 미치지 않는다면 AI 추천 알고리즘으로 인한 문제나 대안을 논의할 필요는 없다. 그러나 수많은 정보가 쌓여서 개인의 의견이 형성되고, 점점 더 온라인과 오프라인의 경계가 사라져가는 연결사회에서 특정 정보가 알고리즘으로 인해 이용자에게 아예 노출되지 않거나 편향적인 정보만 과도하게 노출되는 것은 사회적 문제일 수밖에 없다.

본 연구는 빠르게 변화하고 지속적인 혁신을 거듭하고 있는 인공지능 규제와 관련하여 새로운 법의 제・개정보다는 기술과 산업의 유연성과 다양성을 포섭할 수 있는 정부의 가이드라인을 활용하는 방안을 살펴보았으며, 법적 규제의 대안으

로 이용자의 선택권 강화, 현재 각국 인공지능 규제의 시사점을 포함한 다양한 AI 추천 알고리즘 규제 방안을 제시함으로써 바람직한 정책방향을 모색하였다. 인공지능이 인간과 사회에 주는 편익이 증가하고 있고 앞으로 인공지능을 활용한 추천 알고리즘의 편향성 문제와 이로 인한 혼란은 더욱 가중될 것으로 예측된다. 이에 따라 AI 추천 알고리즘에 대한 규제도 이러한 기술적 변화에 맞춰 지속적으로 개선해 나아갈 필요가 있다.

참고문헌

국문 학술지

갤럽리포트 (2021). 미디어, 콘텐츠, 소셜 네트워크 서비스 이용률.『갤럽리포트 마켓70』, 2021(2).

김유미 (2021). 소셜미디어의 가짜뉴스(Fake News)에 대한 제3자 효과 : 감염병 관련 허위정보를 중심으로.『한국방송학보』, 35(1).

김청희・김남두 (2021) [초점] 성별, 연령대별 유튜브 및 넷플릭스 콘텐츠 이용행태 분석.『KISDI Perspectives』, 2021 July No.3, 1-21.

김춘식・홍주현 (2020). 유튜브 공간에서 '가짜뉴스의 뉴스화': 관련 정치적 의혹 제기와 청와대 반응사례 연구.『정치・정보연구』, 23(2), 403-440.

서봉원 (2016). 콘텐츠 추천 알고리즘의 진화.『방송트렌드 & 인사이트』, 2016년 4_5월호(vol.05), 19.

양종모 (2017). 인공지능 알고리즘의 편향성, 불투명성이 법적 의사결정에 미치는 영향 및 규율 방안.『법조』, 66(3), 60-105.

오세욱 (2019). 알고리즘으로 본 유튜브의 미디어 지향,『관훈저널』, 2019년 봄호(통권 제150호).

이문한 (2021). 코로나19 관련 가짜뉴스 등 '공공의 이익'을 해하는 허위사실 표현에 대한 형사처벌과 헌법상 언론・출판의 자유.『법학평론』, 11.

이상훈 (2020). 가짜뉴스의 법적 규제에 대한 고찰.『법이론실무연구』, 8(1).

정연오 외 (2013). 개인화된 전문가 그룹을 활용한 추천 시스템.『한국지능시스템학회논문지』, 23(1), 7-11.

정원섭 (2020). 인공지능 알고리즘의 편향성과 공정성.『인간・환경・미래』, 25, 55-73.

최승필 (2020). 가짜뉴스에 대한 규제법적 검토 - 언론관련법 및 정보통신망법상 규제를 중심으로 -.『공법학연구』, 21(1).

한국언론진흥재단 (2021). 2021 소셜미디어 이용자 조사 결과표.『한국언론진흥재단』, 351.

URL 주소

강성원 (2019). AI 편집하는 네이버 뉴스 '감성 뉴스' 사라질라. Retrieved April 26, 2022, http://www.mediatoday.co.kr/news/articleView.html?idxno=147677.

네이버 Agenda Research., & 서울대 인공지능정책 이니셔티브. (2021). NAVER-SAPI AI REPORT. Retrieved May 9th, 2022, https://www.navercorp.com/navercorp_/research/2021/20211129093002_2.pdf

방송통신위원회 (2021).「인공지능 기반 미디어(매체) 추천 서비스 이용자 보호 기본원칙」. Retrieved April 26, 2022, https://www.korea.kr/news/pressReleaseView. do?newsId=156459220.

양소리 (2021). 홍준표 "틀튜브가 한국 보수 망쳐…안 보면 된다". Retrieved April 26, 2022, https://newsis.com/view/?id=NISX20211218_0001693028&cid.

윤영찬 외 (2021). [2113509] 알고리즘 및 인공지능에 관한 법률안. Retrieved April 26, 2022, http://likms.assembly.go.kr/bill/billDetail.do?billId=PRC_A2J1R1B1R1J0S1V6W3K9B0K6N6Q0Z9.

파이낸셜뉴스 (2021). "플랫폼 알고리즘은 영업비밀" 전자상거래법 손질 들어가나. Retrieved April 26, 2022, https://www.fnnews.com/news/202104061816369944.

홍윤지 (2021). '알고리즘 설명요구권' 놓고 영업기밀 침해 논란도,. Retrieved April 26, 2022, https://www.lawtimes.co.kr/Legal-News/Legal-News-View?serial=171566.

AI 기반 추천 서비스 이용자 보호를 위한 기본원칙(안) 공개 토론회, (2021). 방통위 추진 'AI 기반 추천서비스 기본원칙'에 업계・소비자 반응 엇갈려. Retrieved April 26, 2022, http://www.aitimes.com/news/articleView.html?idxno=138778.

Apple (n.d.). iPhone에서 Apple이 광고를 제공하는 방식 제어하기. Retrieved April 26, 2022, https://support.apple.com/ko-kr/guide/iphone/iphf60a6a256/ios.

Arguedes, A. et al. (2022). Echo chambers, filter bubbles, and polarisation: a literature review. Retrieved April 26, 2022, https://reutersinstitute.politics.ox.ac.uk/echo-chambers-filter-bubbles-and-polarisation-literature-review.

Bonefeld-Dahl, C. (2021). [BLOG] Seven questions to ask when looking at the EU's new AI regulation tomorrow. Retrieved April 26, 2022, https://www.digitaleurope.org/news/seven-questions-to-ask-when-looking-at-the-eus-new-ai-regulation-tomorrow/.

Central Digital and Data Office (2021a). Algorithmic transparency template. Retrieved April 26, 2022, https://www.gov.uk/government/collections/algorithmic-transparency-standard.

Central Digital and Data Office (2021b). Provide information on how you use

algorithmic tools to support decisions (pilot version). Retrieved April 26, 2022, https://www.gov.uk/guidance/provide-information-on-how-you-use-algorithmic-tools-to-support-decisions-pilot-version.

Connor, J., & Wilson, N. (2019). Reduced Demand Uncertainty and the Sustainability of Collusion: How AI Could Affect Competition. Retrieved April 26, 2022, https://www.sciencedirect.com/science/article/abs/pii/S0167624520301268.

EUROPEAN CENTRAL BANK (2021). Opinion of the European Central Bank of 29 December 2021 on a proposal for a regulation laying down harmonised rules on artificial intelligence. Retrieved April 26, 2022, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ%3AJOC_2022_115_R_0005&home=ecb.

European Commission (2021). Proposal for a regulation of the european parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts. Retrieved April 26, 2022, https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52021PC02 06

European Institute of Public Administration (2021), The Artificial Intelligence Act Proposal and its Implications for Member States. Retrieved April 26, 2022, https://www.eipa.eu/publications/briefing/the-artificial-intelligence-act-proposal-and-its-implications-for-member-states/.

Federal Trade Commision (2018). Hearing #7: The Competition and Consumer Protection Issues of Algorithms, Artificial Intelligence, and Predictive Analytics. Retrieved April 26, 2022, https://www.ftc.gov/news-events/events/2018/11/ftc-hearing-7-competition-consumer-protection-issues-algorithms-artificial-intelligence-predictive.

IBM Research Trusted AI. (n.d.). AI Faireness 360. Retrieved in May 9th. 2022, https://aif360.mybluemix.net/

Jilson, E. (2021). Aiming for truth, fairness, and equity in your company's use of AI. Retrieved April 26, 2022, https://www.ftc.gov/business-guidance/blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai.

Kahn, J. (2021). Europe proposes strict A.I. regulation likely to have an impact around the world. Retrieved April 26, 2022, https://fortune.com/2021/04/21/

europe-artificial-intelligence-regulation-global-impact-google-facebook-ibm/.

Lewis, P. (2018). ‘Fiction is outperforming reality’: how YouTube’s algorithm distorts truth. Retrieved April 26, 2022, https://www.theguardian.com/technology/2018/feb/02/how-youtubes-algorithm-distorts-truth.

Martínez, A. (2019). Are Facebook Ads Discriminatory? It’s Complicated. Retrieved April 26, 2022, https://www.wired.com/story/are-facebook-ads-discriminatory-its-complicated/.

Smith, A. (2020). Using Artificial Intelligence and Algorithms. Retrieved April 26, 2022, https://www.ftc.gov/business-guidance/blog/2020/04/using-artificial-intelligence-algorithms.

Sustein, C. (2018). The Echo Chamber Is the Enemy of Democracy. Retrieved April 26, 2022, https://www.bloombergquint.com/view/steve-bannon-the-new-yorker-and-free-speech