02 알고리즘과 뉴스편집

송해엽 군산대학교 미디어문화학과 교수

1. 들어가며

알고리즘은 단순히 인간이 할 수 없는 방대한 일을 기계적으로 처리하는 도구에 그치지 않는다. 우리가 알고리즘에 대한 논의에서 주목하는 부분은 많은 양의 정보를 필터링하기 위해 기계적 방식을 활용한다는 점이다. 즉, 알고리즘은 정보 필터링(information filtering)을 위해 활용하는 기술의 하나로, 이용자가 관심을 가질만한 정보를 선별하여 추천하기 위해 사용하는 것이다. 예를 들면, 1억 개의 유튜브 영상 중 각자의 취향을 고려하여 10개를 뽑아서 1억 명에게 제시하는 것은 수백 명이 며칠 밤을 새워도 쉽지 않은 일이지만, 컴퓨터는 알고리즘에 따라 지치지 않고 작업을 수행한다. 단순히 기술적인 측면에서 알고리즘에 대해 논의하자면 다량의 정보를 처리하기 위한 기술이라는 관점에 머무르지만, 알고리즘에 관한 논의에서 가장 중요한 점은 규칙에 관한 부분이다. 알고리즘은 '컴퓨터가 문제를 해결하기 위해 수행하는 규칙'을 의미한다. 다수의 정보를 어떤 방식으로 처리할 것인지에 관해 컴퓨터는 알고리즘이 제시하는 규칙에 따른다.

규칙이 가지는 본질적 속성 중 하나는 개인의 행동을 특정한 방향으로 유도할 수 있다는 점이다. 자동차에 달린 내비게이션은 출발지에서 목적지까지 도달할 수 있는 경로를 제시한다. 개인은 목적지까지 도달하기 위한 탐색이 어떤 방식으로 이루어지는지 알기 어렵지만, 내비게이션이 안내하는 길에서 벗어난 경로로 운전하기는 쉽지 않은 일이다. 내비게이션이 말하는 최적의 경로가 과연 무엇인지는 오직 알고리즘을 만든 사람만이 알고 있다. 가장 빠른 시간인지, 가장 짧은 거리인지, 가장 경제적인 방식인지, 경로변경에 따른 효율이 고려되는지 이용자는 판단하기 어렵다. 일반 이용자가 복잡한 경로 추정 방식을 이해하는 것은 쉬운 일이 아니기 때문이다. 이용자는 알고리즘이 제시하는 결과가 옳은지 그른지 판단하기 어려움에도 알고리즘이 특정 규칙을 통해 제시하는 결정을 따르게 된다.

알고리즘이 이용자의 생각과 행동에 영향을 줄 수 있다는 사실은 다양한 연구 결과를 통해서도 확인된 바 있다. 페이스북과 美 코넬대학교의 연구진은 페이스북의 뉴스피드 알고리즘을 조작하여 689,003명의 이용자를 대상으로 감정이 전염될 수 있는지 실험했다. 실험 대상자의 뉴스피드에서 부정적인 내용을 줄였더니 상태 업데이트에서 부정적인 단어의 사용이 줄고 긍정적인 단어가 늘었으며, 긍정적인 내용을 줄였더니 긍정적인 단어는 줄고 부정적인 단어의 사용이 늘어났다(Kramer, Guillory, & Hancock, 2014). 이는 소셜미디어를 통해서도 이용자의 감정이 전염될 수 있으며, 플랫폼이 이용자의 감정 전염

(emotional contagion)을 조작할 수 있음을 실험으로 보여준 것이다. 검색엔진 조작효과(Search Engine Manipulation Effect, SEME)는 정치 투표와 관련한 실험에서 검색 순위 조작으로 부동층 20% 이상 투표에 영향을 줄 수 있다고 보았다(Epstein & Robertson, 2015). 특정 후보에게 우호적인 페이지를 상위에 노출하고 다른 후보에게 우호적인 페이지를 하위에 보여주는 것만으로도 투표에 영향을 미쳤으며, 실험에 참여한 대부분 사람은 검색 순위 조작을 인지하지 못하였다.

이와 같은 연구 결과는 알고리즘에 대한 담론이 규칙이라는 차원에서 논의될 필요가 있음을 시사한다. 알고리즘은 온라인 쇼핑몰에서 물건을 구매하고, 음원 사이트에서 취향에 맞는 곡을 청취하고, 소셜미디어를 통해 다양한 사람들과 관계 맺는 방식에 이르기까지 다양한 분야에 활용되고 있다. 알고리즘 활용은 이러한 분야를 넘어 포털의 온라인 뉴스 배열에도 적용되고 있다. '다음'은 오래전부터 알고리즘에 기반하여 개인에게 가장 적합한 맞춤형 기사를 제공한다고 이야기했고, '네이버'도 2019년부터 에어스(AiRS)라고 부르는 알고리즘 기반 자동 추천 기사를 전면 적용했다. 과거 종이신문 시절 어떤 기사가 신문의 1면에 나와야 하는지를 결정하는 것은 인간이었다. 충분한 경력을 쌓은 기자가 현장에서 보내는 기사를 취합하고 보도 방향을 결정했다. 뉴스는 온라인 쇼핑상품이나 음악과는 다른 특징을 가진 콘텐츠이다. 뉴스는 세상의 모습을 재현하여 사람들에게 전달하는 기능을 한다. 알고리즘에 기반하여 자동으

로 뉴스 기사를 배열하는 것은 마치 공정한 것처럼 보일 수 있으나 뉴스가 대중에게 가지는 영향력을 고려한다면 어떤 규칙에 기반을 두어 뉴스가 배열되어야 하는지 사회적 합의가 필요한 영역일 수 있다.

2. 포털의 뉴스배열 알고리즘 논란

포털의 뉴스 배열은 정치적 편향성과 관련하여 오랫동안 정치권의 공격을 받았다. 과거부터 뉴스편집에 대한 논란은 있었으나 네이버 스포츠가 한국프로축구연맹의 청탁을 받고 기사 배열을 변경한 사건이 2017년 10월 알려짐에 따라 논란은 더욱 거세졌다. '한국프로축구연맹, 누군가를 처벌할 자격이 있나'라는 2016년 10월 작성된 기사가 당시 네이버 스포츠 메인화면에 배치되었으나, 한국프로축구연맹 관계자는 K리그에 부정적인 기사를 보이지 않게 해달라는 청탁을 네이버 뉴스편집 담당자에게 전달했다. 네이버 담당자는 부하 직원에게 지시해 해당 기사를 메인에서 삭제했다는 것이다. 2017년 10월 말 국회 과학기술정보방송통신위원회에 출석한 이해진 네이버 창업자는 "알고리즘을 공개해서 어뷰징만 안 당한다면 외부에 (뉴스 배열 알고리즘을) 공개하고 장기적으로 감독을 받는 것이 중요하다고 생각한다"라고 말했다(김남희 외, 2017). 알고리즘을 공개하지 말아야 한다는 사업자의 논리 중 하나는 알고리즘 공개로 인해 악용될 소지가 있다는 것이다. 기사를 배열 상위에 노출하기 위해 알고리즘이 가진 규칙을 남용하여 혼란이 발생할 수 있다고 본다.

네이버는 이후 뉴스 배열의 공정성을 위해 모든 뉴스 배열을 알고리즘에 넘기겠다고 밝혔다(한지이, 2017). 네이버 모바일 뉴스 배열에서 사람이 담당하는 부분은 20% 정도였으나 인공지능과 외부 비중으로 모든 영역을 채우겠다고 결정한 것이다. 카카오도 속보성 뉴스만 편집에 관여하고 일상적 뉴스 추천과 배치는 알고리즘을 통해 이루어진다고 밝혔다. 하지만 네이버와 카카오의 주장처럼 포털뉴스 편집을 알고리즘이 담당한다고 하더라도 알고리즘 자체를 지속해서 평가하고 사회적 의견이 반영돼야 한다는 의견도 제기되었다. 따라서 2018년 1월 네이버는 기사 배열 공론화 포럼을 통해 외부 검증을 받기로 했다. 이어서 2018년 5월 만들어진 '네이버 뉴스 알고리즘 검토위원회'는 뉴스 서비스의 객관성과 투명성을 확보하고 향후 발전 방향을 논의하며 전반적인 뉴스 서비스 알고리즘에 대한 검토를 진행했다. 11인으로 구성된 전문가는 각 뉴스 서비스의 데이터 확보 과정, 실제 검색 결과를 결정하는 알고리즘과 주요 특징(feature)을 확인하고 이용자에게 서비스가 공개되는 과정과 업무 절차의 적절성을 검토했다고 밝혔다(서혜인, 2018).

지난 2019년 4월 네이버는 알고리즘에 따른 맞춤형 자동화와 언론사 편집판 적용으로 뉴스 서비스에서 네이버 내부 편집 영역을 제거했다. 공정성 논란에서 벗어나기 위한 네이버의 오랜 대응으로 포털뉴스와 관련한 알고리즘 논의가 수그러든 것처럼 보였다. 한동안 알고리즘 이슈가 가라앉은 것처럼 보였으나 2020년 9월 윤영찬 의원의 '카카오 들어오라 하세요' 문자 사건으로 인해 재확산되었다. 포털 사업자는 정치적 편향성 논란을 피하고자 뉴스 전 영역에 알고리즘을 적용한다고 발표했으나, 과연 알고리즘이 뉴스 배열을 온전히 담당한다고 할 수 있는지 다른 외부 영향에서 자유로운지에 대한 논란이 제기되었다. 신문법 제3조(신문 등의 자유와 책임)와 제4조(편집의 자유와 독립)는 신문 및 인터넷신문에 한해 언론의 자유와 독립을 보장하고 있다. 하지만 포털뉴스와 같은 인터넷뉴스서비스사업자의 경우 독립성 보장에 대해서 법적인 근거가 미비한 상태이다. 따라서 인터넷뉴스서비스사업자도 신문이나 인터넷신문처럼 자유 및 편집의 독립성 등을 보장할 수 있도록 규정하고, 이를 위반할 경우 2년 이하의 징역 또는 2천만 원 이하의 벌금에 처하도록 함으로써 언론의 투명성을 확보하고자 신문법 일부 개정안이 발의되기도 했다(의안번호 2103696. 2020. 9. 9.).

다양한 논의가 이루어졌으나 포털 알고리즘과 관련된 논란이 해결되었다고 보기는 어렵다. 포털 내부 관계자를 제외하면 알고리즘의 작동과정을 여전히 알기 어려우며, 네이버가 자체 운영한 알고리즘 검토위원회의 발표 내용 외에 아무런 검증 장치도 없다. 2020년 9월 윤영찬 위원은 이낙연 대표의 국회 연설 기사가 다음 메인화면에 빠지자 '카카오 들어오라 하세요'라는 문자를 보내는 것이 언론에 노출되어 논란을 일으켰다. 아직도 여전히 특정 정당 입장 혹은 특정 정치 성향의 뉴스 기사만 노출된다는 의구심은 여전하다. 최근 다음 뉴스에 조선일보 기사가 늘어났다는 이용자들의 의심은 조선일보 출신이 부사장이 되었다는 사실과 연결 지어 의혹으로 확대되기도 했다. 카카오는 "매체별로 송고되는 시간대가 다르므로 오전 시간대에 조선일보, 중앙일보, 동아일보가 많이 노출될 수 있고 심야 시간대에 한겨레, 경향신문 기사가 많이 올라오는 것처럼 보일 수 있다"라고 밝혔다. 이러한 의혹뿐만 아니라 알고리즘이 배열하는 뉴스 기사의 품질에 관한 의구심도 여전하다. 공들여 작성한 언론사의 기사 대신 해당 기사를 짜깁기한 기사가 상단에 노출되는 현상도 계속되고 있다.

네이버에서 운영한 '뉴스 알고리즘 검토위원회'가 발표한 6개월간 검토 결과는 네이버의 뉴스 자동화를 "공정성과 신뢰성 문제의 해결을 위한 효율적 대

안"이라고 평가했다(이정민, 2018). 뉴스검색 결과는 알고리즘을 통해 자동으로 배치되기 때문에 담당자의 개입은 원칙적으로 불가능한 구조이며, 뉴스검색 개발 및 운영개선에 관한 의사 결정이 모두 기록되고 있다고 밝혔다. 또한, 이용자 의견을 수집하고 반영해 검색 품질 만족도를 꾸준히 평가하고 있다고 보았다. 검토위원회는 에어스(AiRS) 뉴스 추천 서비스 영역에서도 네이버가 편집자의 개입 없이 자동으로 뉴스 이용자들의 피드백 데이터를 수집해 학습 데이터를 확보하고 있으며, 협력 필터(Collaborative Filtering)와 품질모델(Quality Model)을 결합해 이용자의 기사 선호도와 기사 품질을 고려한 개인화된 추천 점수를 계산하는 알고리즘을 활용하고 있음을 확인했다. 연예 및 스포츠 기사 추천 서비스에서는 수집되는 기사의 양과 형태, 뉴스 정보의 유효기간 등이 일반 분야 뉴스 기사와는 매우 다르다는 특징을 고려해 해당 영역의 특성을 반영한 기사 클릭 수치와 같은 이용자 정보를 바탕으로 비 맞춤형 뉴스 추천 서비스를 제공하고 있었다.

'뉴스 알고리즘 검토위원회'의 결과는 네이버가 뉴스검색 서비스 영역에서 명확한 지침을 기반으로 품질평가 방향을 정립하였으며, 앞으로도 계속 이러한 시스템을 정착시키는 것이 중요할 것이라고 보았다. 다만 연예 및 스포츠 기사의 경우 데이터 기반 기계 학습이 아닌 규칙 기반의 알고리즘으로 서비스가 운영되고 있는데, 시간이 지남에 따라 효용성에 변화가 있을 수 있으므로, 자질에 대한 꾸준한 검증이 필요할 것이라고 했다. 그 외에도 학습 데이터 구축 주기를 더욱 단축하고 다양한 연령층을 포괄하며, 알고리즘 개선을 위한 연구개

발에 힘쓸 것과 '좋은 기사'에 대한 공통 기준을 확립하고 기사 품질평가 방법을 개발해 주기적으로 실행하는 것을 요청하기도 했다. 실제 네이버는 협력 필터 기술로 비슷한 관심 분야를 가진 사람들이 본 콘텐츠를 통해 이용자들이 콘텐츠를 소비하는 방식을 분석한 뒤 나와 비슷한 관심사를 가진 사람들이 많이 본 콘텐츠를 먼저 선별해 보여준다. 이후 품질모델을 활용, 기사의 제목·본문·이미지·바이라인·작성 시간 등을 분석하여 정보량이 풍부한 뉴스를 판별한 뒤 조회 수, 체류 시간 등 다수의 이용자 소비 활동에 기반해 만족도가 높은 기사를 추천한다. 카카오는 전체 이용자 반응과 성·연령별 집단에 따른 반응, 노출된 기사의 클릭 여부, 클릭률, 열독률 등을 실시간으로 측정해 뉴스를 배열한다. 기사 제목이나 이미지, 작성시간, 카테고리별 추천 요인 등도 큐레이션에 반영하기 때문에 사용자마다 노출되는 기사가 다르다.

3. 중립성과 투명성 확보방안

네이버의 '뉴스 알고리즘 검토위원회'는 관리자의 개입이 원천적으로 불가능한 구조라고 결론을 내렸지만, 인공지능 알고리즘을 만드는 것은 결국 사람이 담당하기 때문에 누가 어떤 기준으로 인공지능 시스템을 만드는지에 대한 검증이 필요하다. 즉, 인공지능은 시스템을 설계하는 사람의 생각이 반영될 수밖에 없으므로 시스템이 차별하지 않는지, 정치적으로 중립적인지를 판단하기 위한 감사 시스템이 필요한 것이다. 이러한 중립성과 투명성을 확보하기 위해서 다양한 해결 방안이 제시되고 있다. 첫 번째는 포털뉴스 배열 중립성 확보를 위한 법제도 개선이다. 두 번째는 알고리즘을 투명하게 공개하는 것이다. 세 번째는 포털뉴스 배열 시 저널리즘 가치를 반영한 기사를 의무적으로 노출하는 것이다. 이와 같은 세 가지 방안이 중립성과 투명성을 확보하기 위한 대안으로 논의되고 있으나 각각의 해결 방안을 적용하는 것이 모든 문제를 해결하는 방안은 아니다. 각 방안을 적용하기 위해서는 고려해야 할 측면이 다수 존재한다.

첫째, 포털뉴스 배열의 중립성 확보를 위한 법제도 개선에 먼저 필요한 것은 중립성을 정의하는 것이다. 뉴스 배열의 중립성이라는 것은 개념적으로 정의하기 어려운 것이며, 개념적으로 정의되지 않은 것을 알고리즘을 통해 배열하는 것은 가능하지 않다. 알고리즘이 결과물을 보여주기 위해서는 기계가 이해할 수 있는 수준의 매우 구체적인 계량화가 필요하다. 예를 들면, 보수와 진보의 기사를 같은 숫자로 배열하는 것이 중립성인가? 보수와 진보라면 언론사는 조선일보와 한겨레인가? 동수로 배열하더라도 기사의 분량이 같아야 하는 것 아닐까? 중립성을 정의하기 위해서는 이러한 조건에 대한 질문이 꼬리를 물고 이

어질 수 있다. 그러므로 구체적인 중립성에 기준을 제시하지 못하면 기사 배열에서 중립성을 확보하라고 법안에 제시하기 어렵고, 중립성에 대한 위반 사안을 판단하는 것도 불가능하다. 실제 뉴스 배열에 대한 위반 사안은 배열을 보는 사람의 기준에 따라 맞춤으로 제시되기에 개인마다 다르고, 다르게 나타난 결과를 다시 재현하는 것도 불가능에 가까운 일이다. 이러한 이유로 법제도 개선을 통해 규제하는 것은 쉬운 일이 아니다.

둘째, 포털뉴스 배열 알고리즘이 가지는 다양한 기준과 가중치를 투명하게 공개하는 것도 다른 대안으로 제시된다. 하지만 알고리즘 공개가 강제되더라도 발생할 수 있는 가장 큰 문제는 검증 불가능성이다. 실제 알고리즘을 사실대로 공개한 것인지 일부만 공개한 것인지 외부에서는 판단할 수 없다. 알고리즘 대부분은 수학 공식이나 컴퓨터 코드로 이루어져 있으므로 공개하더라도 다수의 대중이 어떤 의미인지 파악하기 어렵고 실제 데이터를 활용한 결과물을 확인해야만 대략 어떤 방식으로 작동하는지를 파악할 수 있다. 즉, 법적 규제와 행정 절차를 통해 알고리즘 공개를 의무화할 수 있지만 공개한 내용이 사실인지 검증할 수 없다. 그리고 실제 검증 작업을 데이터와 함께 수행하더라도 구체적인 세부내용을 파악하기는 쉽지 않다. 구글의 경우 알고리즘에 관한 특허를 통해 일부 요인이 공개되어 있다. 언론사 기사 생산량, 기사의 길이, 보도 중요성, 속보성, 뉴스검색 이용 패턴, 수용자 수 및 트래픽, 언론사 뉴스룸 규모, 언론

사 지국 수, 실명 인용 보도의 수, 언론사 보도범위, 보도 기사의 글로벌 도달률, 글쓰기 스타일을 통해 뉴스의 순위를 매긴다. 하지만 구글도 모든 요인을 세부적인 수치와 함께 공개한 것이 아니라 특허를 등록하는 과정에서 주요 요인이 밝혀진 것에 불과하다. 해당 요인에서 어떤 기준을 좋은 것으로 평가하고 어느 지점부터 나쁜 것으로 보는지에 대한 내용은 알려지지 않았다.

셋째, 저널리즘 가치를 반영하는 기사를 노출하도록 강제하자는 주장도 있다. 이러한 논의에서 문제가 되는 부분은 저널리즘 가치를 반영하는 기사가 무엇인지 선정하는것과 관련한 문제이다. 기사에 대한 평가는 사람마다 다르므로 저널리즘 가치를 반영한 기사라는 것을 논의할 때 편향성에 대한 문제가 제기되게 된다. 한국언론진흥재단이 2016년부터 3년간 진행한 '뉴스 트러스트' 사업은 저널리즘 가치에 기반한 공적 배열 알고리즘을 개발하고자 시도하였다. 과연 실제 알고리즘을 정밀하게 구현할 수 있는지에 대해 의심의 시선이 있지만, 저널리즘 가치에 기반한 알고리즘을 만들기 위해 온라인 저널리즘에서 따라야 하는 규율이 무엇이고 좋은 기사가 무엇인지 논의를 통해 합의를 만들고자 하는 시도였다. 회의 결과를 담은 보고서를 살펴보면 각자의 시각이 달라 좋은 기사가 무엇인지 이야기하는 게 우리가 평소에 추상적으로 생각해왔던 것과 달리 쉽지 않다는 사실을 확인할 수 있다. 즉, 저널리즘 가치를 반영하는 기사 노출 공간을 만들었을 때 문제가 되는 부분은 다양한 이해집단 사이 합의가 어렵다는 것이다. 신문협회가 추진했던 온 뉴스(On news)는 2010년 12개 중앙일간지의 뉴스를 온라인으로 서비스했다. 하나의 애플리케이션에서 12

개 주요 신문의 뉴스 뿐만 아니라 실시간 속보까지 다양한 뉴스 콘텐츠를 이용할 수 있었다. 하지만 어떤 언론사를 먼저 노출하는지에 관한 문제, 언론사별 다른 결제 시스템, 언론사 묶음 서비스에 관한 갈등으로 인해 2013년 서비스가 종료되었다.

알고리즘 공정성 추구를 위해 논의되어 온 여러 가지 방법들이 가지는 나름의 제한점이 있었다. 가장 실현 가능한 방식으로 논의되는 것 중 하나는 알고리즘 자동배열과 관련하여 다양한 정보를 이용자에게 공지하는 것이다. 알고리즘이라는 용어가 미디어 분야에서도 널리 사용되면서 일반 대중들도 알고리즘이라는 용어에 점차 익숙해지고 있다. 오세욱과 송해엽의 연구보고서에 따르면 알고리즘 뉴스 배열을 인지한다고 응답한 포털뉴스 이용자는 69.1%였다. 생각보다 낮지 않은 수준으로 보이나 여전히 뉴스 배열을 알고리즘이 한다는 사실을 인지하지 못하는 이용자도 있다. 특히, 50대의 경우 알고리즘 뉴스 배열을 인지한다는 응답이 58.6%로 전체 평균보다 상대적으로 낮은 것을 확인할 수 있었다. 이러한 상황에서 알고리즘이 결정한 결과에 관해 최대한 자세한 설명을 하는 것이 알고리즘에 대한 이해를 높일 뿐만 아니라 뉴스 전반에 관한 신뢰를 높일 수 있는 수단이 될 수 있다.

미국 노스웨스턴대학 컴퓨테이셔널 저널리즘 랩(Computational Journalism Lab)에서 운영하는 알고리즘 팁스(Algorithm Tips)라는 사이트는 각종 알고리즘을 모아놓은 데이터베이스를 운영한다. 알고리즘이 적용되는 분야는 점차 확대하고 있으며 일상의 많은 부분이 인공지능 기반 기술 자동화에 의해 작동하고 있지만, 다수의 사람이 자동화된 결정의 결과물인지를 모르고 있는 경우가 많다. 따라서 기본적으로 이루어져야 하는 것은 어떤 부분이 자동화된 결정인지를 알려야 한다는 것이다. 알고리즘 팁스는 이를 위해 각종 알고리즘을 검색하고 결과물에 관한 내용을 알리는 방식으로 미국 연방정부에서 적용하고 있는 자동화된 의사 결정 내용을 정리하고 있다. 세금 납부, 복지, 보험, 채용, 보석금 산정에 이르기까지 시민에게 영향을 미칠 여지가 큰 부문에서 알고리즘으로 인한 자동화가 이루어지지만, 시민들은 이러한 문제가 왜 중요하고 삶의 어떤 부분과 관련되어 있는지 알지 못하고 있다. 알고리즘 팁스는 다양한 알고리즘이 삶의 어떤 부분과 관련되었는지를 유형화하고 시스템에 관한 다양한 정보를 제공함과 동시에 잠재적 위험성과 영향력을 다수에게 평가하도록 공개하고 있다. 이러한 방식은 알고리즘으로 인한 자동화의 문제를 직접 해결하지는 못하지만, 이용자가 스스로 결정할 수 있도록 정보를 공개하는

방식이라고 할 수 있다.

알고리즘 팁스에서 제공하는 사례 중 하나는 채용과 관련한 알고리즘이다. 자동화가 채용 절차의 효율성을 개선할 수 있으므로 기업이 인사부서를 계속 축소하고 있다. 이로 인해 인공지능 알고리즘이 가지는 매력은 점차 증가하고 있으나 자동화된 채용 절차는 기존 편견을 반영하고, 편견 있는 학습 데이터를 기반으로 지원자 명단을 형성하고, 채용 관행에서 차별을 영속화할 수 있음을 밝히고 있다. 고용 관행에서 차별을 다루는 많은 고용법이 있지만, 고용 심사 알고리즘에서 차별을 인식하고 완화하는 문제는 새로운 정책 문제를 제기한다. 편견을 반영한다는 증거가 점차 발견되고 있는 상황에서 원고의 증명책임에 대한 어려움이 있을 수 있으며, 고용주의 경우 결과가 충분히 정확하다면 합법적인 선 안에서 비즈니스 목표를 수행한 것으로 정당화할 수 있다. 이러한 사례는 채용과 관련된 부문이지만 뉴스와 관련해서 대중에게 알고리즘이 미칠 수 있는 영향력은 무엇이고 삶의 어느 부분과 관련되어 있는지 알리는 것이 필요할 수 있다는 점을 시사해준다.

4. 알고리즘, 사회적 합의가 이루어지는 공간

소셜미디어가 발달하고 개인끼리 연결될 수 있는 수단이 다양하지만, 여전히 뉴스는 정보를 다수의 사람에게 전달하며 사람들을 연결하고 공동체를 유지할 수 있도록 이바지하는 수단이다. 포털뉴스 이용 동기를 물어본 설문에서 가장 다수의 응답을 받은 것은 최신 뉴스를 빠르게 이용할 수 있기 때문이라는 것이었다. 국가와 공동체에 대한 소식을 빠르게 듣고 사람들과 함께 공유하기

를 원하는 것이다. 이러한 면에서 생각해볼 수 있는 부분이 지역 언론과 관련한 이야기이다. 지역 언론 홀대의 이야기는 지역과 관련된 사건이 발생했을 때 지역 언론의 기사가 상단에 노출되는 것이 아니라, 지역 언론을 베껴 쓴 중앙 언론의 기사가 먼저 노출되고 있다는 것이다. 지역에서 발생한 사건은 지역 언론과 중앙 언론 주재 기자가 경쟁적으로 보도하는 구조로 되어있다. 중요한 사안에 대한 추가 보도가 이루어지면 중앙 언론과 통신사에 기사량에서 밀리면서 지역 언론기사는 사람들에게 노출될 기회를 잃게 된다. 현재 포털 검색영역 알고리즘은 클러스터링 방식으로 하나의 주제에 관해 유사 기사를 묶어 노출하는 방식으로 운영하고 있다. 따라서 '1보'도 중요하지만, 추가 기사에 따라 기사가 뒤로 밀리며 노출되기 어려워지는 현상도 발생한다.

우리나라에서 서울을 중심으로 사고하는 것이 익숙한 일이지만 지역은 지역에서의 삶이 있고, 지역 언론은 지역 사람들의 삶에 관한 이야기를 전달하는 기능을 담당한다. 지역 언론은 중앙 언론의 정보량에 밀려 제대로 언급되지 않는 지역의 이야기를 다루고 사람들의 관심을 다룸으로써 지역의 공동체를 회복하는 기능을 한다. 기존 중앙 언론이 외면하는 소외계층과 소수 견해를 대변하는 기능이 있으며 지역주민들의 민원을 수렴하고 민·관을 연결하는 중계 역할도 있다. 따라서 지역 언론은 지역 운동, 생활서비스 제공, 공동체 의식 회복, 민원 중계, 대안 언론이라는 차원에서 주요 기능을 가진다. 지역신문발전지원특별법(지역신문법)이 2004년, 6년을 유효기간으로 하는 한시법으로 제정되었으나 현재까지 계속 이어지고 있는 이유도 지역 언론이 해당 지역에서 수행

하는 기능과 역할이 중요하다고 보았기 때문이다. 지역신문법 제1조는 "지역신문의 건전한 발전기반을 조성하여 여론의 다원화, 민주주의의 실현 및 지역사회의 균형 발전에 이바지함을 목적으로 한다"라고 밝히고 있다. 모두가 포털을 통해 뉴스를 소비하는 상황에서 포털뉴스 알고리즘의 지역 차별 문제도 해결해야 하는 과제 중 하나일 수 있다.

이를 해결하는 방법으로 GPS 데이터에 기반하여 메인화면을 각 지역에 맞춤으로 변경하는 방식이 논의되고 있다. 하지만 이러한 부분도 개인정보 동의에 관한 문제라든지 뉴스 이용자의 선택권 제약에 관한 논쟁을 피할수 없다. 물론, 각 지역에서 어떤 기사를 메인화면에 노출하는지는 기술적으로 가능해도 민간사업자의 선택 영역이지 입법으로 해결하기는 어려울 수 있다. 알고리즘 공정성과 비슷한 문제로 입법 추진 시 지역별 메인에 노출되는 언론사의 기준이라든지 언론사에서 노출되는 기사의 선정 문제와 같은 내용이 논의를 통해 규정되어야 한다. 모든 지역 언론사가 자사 기사를 먼저 상단에 노출하기를 원하기 때문에 포털 사업자 입장도 어려울 수 있다. 현실적으로 현재 네이버가 모바일에서 운영하는 '우리동네' 섹션이 대안은 될 수 있으나 포털 사업자의 사업적 판단이 필요한 영역이다. 그럼에도 언론이 가지는 사회적 역할을 고려하고 합의가 이루어진다면 적극적으로 알고리즘 영역에서 적용해야 하는 것이 무엇인지 생각해볼 수 있다는 점이다.

알고리즘에 대한 오해 중 하나는 알고리즘이 데이터 기반으로 학습한 결과를 그대로 보여주는 투명한 도구라는 생각이다. 이러한 생각은 알고리즘이 발생시키는 편향의 결과를 이용자 행동을 그대로 반영한 것일 뿐이라는 생각으로 이끈다. 일부는 사실이지만 일부는 사실이 아니다. 다수의 알고리즘은 실제로 데이터에 기반하여 가장 적합한 결과를 추정해내는 기술을 활용한다. 하지만 알고리즘이 그렇다고 하더라도 알고리즘이 발생시키는 편향에서 자유로운 것은 아니다. 최근 추세는 알고리즘으로 인해 발생하는 편향을 최소화하고 사회적으로 바람직한 결과를 제시하는 방안도 연구되고 있다. 예를 들면, 단어 빈도와 같은 말뭉치 규칙성이 성별 편향을 증폭시키는 부정적 결과를 낳을 수 있기에, 이러한 편향을 완화하기 위한 조정 작업이 필요하다는 사례도 있다(Wang, Lin, & Rajani, 2020). 알고리즘은 미디어 영역에서도 점차 강력한 영향력을 미치며 사람들의 일상을 변화시킨다. 단순히 이용자 행동의 반영이라고 넘길 것이 아니라 세계를 어떤 방식으로 드러내야 하는지 적극적인 고민과 사회적 합의가 필요한 공간이다.

참고 문헌

1) 김남희 · 김범수 · 심민관 (2017. 10. 31). ICT 기업인 총출동한 국감서 7시간 질문 공세…'이해진 청문회' 방불. 〈조선비즈〉. URL : https://biz.chosun.com/site/data/html_dir/2017/10/31/2017103100427.html
2) 서혜인 (2018. 11. 29). 네이버 뉴스 알고리즘 검토 위원회, 뉴스 서비스 검토 결과 발표. 〈플래텀〉. URL : https://platum.kr/archives/111887
3) 이정민 (2018. 11. 29). 네이버 뉴스 알고리즘 검토 위원회 "뉴스배치 관리자 개입 불가능 확인". 〈조선비즈〉. URL : https://biz.chosun.com/site/data/html_dir/2018/11/29/2018112901824.html
4) 한지이 (2017. 12. 7). 네이버 "뉴스 편집 손떼겠다"…인공지능이 전담. 〈연합뉴스〉 URL : https://www.yna.co.kr/view/MYH20171207019700038
5) Epstein, R., & Robertson, R. E. (2015). The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. 〈Proceedings of the National Academy of Sciences〉, 112(33). E4512–E4521.
6) Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive–scale emotional contagion through social networks. 〈Proceedings of the National Academy of Sciences〉, 111(24). 8788–8790.
7) Wang, T., Lin, V., & Rajani, N. (2020). Double hard–debias: Tailoring word embeddings for gender bias mitigation. 〈Salesforce Blog〉. URL : https://blog.einstein.ai/double–hard–debias–tailoring–word–embeddings–for–gender–bias–mitigation/