



미디어 알고리즘과

민주주의

알고리즘 매개성과 정치적 양분사회 : 기술과 사회심리의 상호작용효과

황용석 (교수, 건국대학교 미디어커뮤니케이션학과)



웹에서 PDF 바로 보기

요약

디지털기술과 민주주의는 밀접한 관계를 가져왔다. 기술의 변화가 시민의 정보습득, 참여방식에 영향을 미치기 때문이다. 인공지능시대에 알고리즘으로 대변되는 새로운 소프트웨어 기술은 정치 및 정보매개자로서 사회의 접착제 역할을 수행하고 있다. 알고리즘은 우리의 일상 현실을 재구성하고 정보생산 및 소비양식을 변화시키고 있다.

알고리즘은 단순한 매개자가 아니라 사회구조와 인간의 인식 등에 광범위한 영향을 미친다. 알고리즘은 입력데이터와 모델 등 다양한 요인에 의해 편향을 가질 수 밖에 없다. 알고리즘 편향의 결과는 다양하게 나타난다. 우선, 차별의 문제를 들 수 있다. 알고리즘의 차별은 의도적인 차별적 처우와 결과에 의해 발생하는 차별적 효과로 구분해서 설명할 수 있다. 전자는 설계상의 문제이고, 후자는 학습 데이터에 의한 문제이다. 또한 알고리즘은 사람들의 인지나 판단과정에도 영향을 미친다. 알고리즘이 제시하는 정보의 배열순서만 바꾸어도 정치후보자에 대한 태도가 변화하고, 제공하는 소셜미디어 상의 감정전이가 일어난다. 알고리즘은 전달하는 내용과 관계없이 그 자체로 고유한 효과를 만든다.

알고리즘은 오늘날 우리가 겪고 있는 정치적 갈등, 특히 좌우 이념과 의견대립으로 나타나는 사회의 양극화와 분리할 수 없다. 필터버블과, 에코챔버와 같은 용어로 대변되는 온라인상의 소통단절과 분극화 현상은 알고리즘 효과와 연관이 있다. 필터버블은 이용자의 개인맞춤 알고리즘에 의해 생기는 정보편식현상을 의미하며, 이용자가 제한된 주제의 정보에 갇혀 정보의 다양성과 여론지각의 균형성이 떨어진 상태를 말한다. 필터버블은 같은 정보를 공유하는 유사상종의 정보 및 의견집단을 형성시켜, 사회적 분극화를 부추기는 원인의 하나이다. 유사한 집단 내에서 일어나는 상호작용은 또 다른 심리적 기제인 동조화 효과를 양산하고 그 결과 집단극화가 일어난다.

에코챔버효과는 밀폐된 반향실에서 자신과 같은 목소리가 메아리치고 증폭되는 현상이다. 디지털공간에서 유사한 사람들끼리만 소통하게 되어 점차적으로 편향된 사고가 강화되는 효과를 뜻한다. 네트워크 관점에서 보면 이런 현상을 호모필리 또는 유유상종 효과라고 부를 수 있다. 필터버블과 에코챔버는 사람들의 확증편향을 강화시켜서 허위조작정보(가짜뉴스)를 생산하고 유통시키는 원인이기도 하다.

알고리즘에 의해 매개되는 편향된 여론지각 환경과 닫힌 커뮤니케이션 구조는 우리 사회를 양분시키고 각각의 집단이 극단의 방향으로 움직이게 한다. 언론과 같은 전통적 정치매개집단의 신뢰와 영향력이 낮은 가운데, 민주주의의 건강성을 확보하기 위한 노력이 필요하다.

알고리즘의 영향으로부터 자율적인 판단능력을 기르기 위해서는 정확한 정보를 판단할 수 있는 판단기준의 제시, 다양한 이견의 교차성, 다른 생각을

인정하는 관용, 대화와 표현능력이 필요하다. 기술을 매개로 연결되는 관계 속에서 시민주의 능력(civic competence) 점점 중요해 지고 있다. 이에 따라 디지털리터시에 많은 관심이 필요하다.

또한 알고리즘이 작동되는 방식을 어떻게 설명하는가에 따라 알고리즘의 결과물을 받아들이는 비판적 능력이 달라진다. 따라서 알고리즘의 설명책임을 어떻게 실행할 것인지를 디지털플랫폼 사업자들이 고민해야 한다.

1. 정치커뮤니케이션을 매개하는 새로운 기술

디지털은 민주주의를 촉진시키는가?

이 같은 질문은 인터넷의 초창기부터 많은 사회과학자들의 궁금증이었다. 오늘날 우리가 겪고 있는 정치적 갈등, 특히 좌우 이념 및 의견대립과 정치적 양분사회 현상은 디지털 민주주의에 대한 기대를 허물고 있다. 그런 가운데 인공지능 알고리즘으로 무장한 디지털플랫폼은 정치사회적 극단화와 양분화의 한 원인으로 지목되고 있다. 필터버블, 에코챔버와 같은 용어로 대변되는 온라인상의 소통단절과 분극화 현상이 그 대표적인 예이다. 이 글에서는 알고리즘 매개 시대에서 우리가 당면한 정치적 소통의 문제를 새로운 기술의 매개성과 사회심리적 요인을 중심으로 되새겨 보고자 한다. 이를 위해 디지털기술이 전통적인 정치매개집단을 대체하는 현상과 그 대체물로서 알고리즘의 고유한 정치적 효과, 그리고 이 기술을 매개로 나타나는 분극화되고 양분화된 사회현상을 소개하고자 한다.

1) 전통적 정치매개집단을 대체하는 비인간행위자

월드와이드웹의 상용화와 대중화가 일어난 1990년대 후반을 기점으로 많은 학자들은 전자민주주의에 대한 낙관론을 전개했고, 새로운 디지털 세대에 환호했다. “디지털 네이티브”(Prensky, 2001) 또는 “넷-세대”(Tapscott, 1998)라 불리는 새로운 세대가 기술역량을 갖추고 보다 능동적인 정치참여문화를 만들 것으로 인식했었다. 하워드 레인골드(Rheingold, 1991)는 디지털 미디어가 단순히 정보의 전달 매체가 아니라, 적은 비용으로 보통의 사람들이 정치 공동체를 구성할 수 있게 하는 공공의 장이 될 것이라고 기대했다.

디지털기술이 민주주의에 낙관적 미래를 가져다 줄 것이라는 이 같은 믿음은 시민참여에 들어가는 거래비용을 감소시켜서 참여를 촉진할 것이라는 직접민주주의의 이상에 근거해 있었다. 실제로 대의제 민주주의에서 정치커뮤니케이션의 핵심 연결자 역할을 했던 정치매개집단(언론, 정당, 시민단체 등)의 영향력이 급격히 약화되고 있다. 반면에 소셜미디어와 같은 디지털플랫폼을 통해 개인간 연결이 중요한 소통통로가 되고 있다. 전통적인 정치매개집단을 대신해서 디지털 공간에서 개인을 연결짓는 새로운 정치매개기술이 등장한 것이다.

새로운 커뮤니케이션 환경에서 개인들은 제도화된 미디어가 아닌 개인미디어를 통해 다양한 방식으로 소통하고 콘텐츠를 창출하고 있다. 수동적인 수용자(audience)가 정보생산자(prosumer)로 전환되면서 사적인 영역의 콘텐츠들이 공적인 영역으로 급속히 유통되기 시작했다.

이러한 현상들은 기존의 사회여론과정을 급격하게 변화시켰다. 정치과정에서 제 주체들간의 커뮤니케이션 네트워크를 다양하게

형성되면서, 인터넷 이용자들이 의제를 생산하고 유포하는 내발적 여론확산 공간이 확대되었다. 확산되는 정치 이슈나 의제는 전적으로 선호적 연결(preferential attachment)을 통해 형성되고 널리 전파되었다. 알고리즘과 같은 새로운 기술이 전통 언론을 대체하면서, 네트워크화된 여론공간을 창출한 것이다.

알고리즘으로 무장한 유통 디지털플랫폼(Digital Platform) 또는 디지털정보매개자(Digital Intermediaries)의 영향력은 여러 조사에서 확인된다. 영국 옥스퍼드 대학 부설 로이터 저널리즘 연구소가 발표한 『디지털 뉴스 리포트(Digital News Report) 2019』에 따르면, 뉴스 이용을 위한 온라인 채널을 묻는 질문에서 한국 조사응답자의 48%가 검색, 27%가 뉴스를 모아 보여주는 뉴스 수집 서비스인 애그리게이터(aggregators), 9%가 소셜 미디어 등 총 84%가 디지털 플랫폼을 통해 주로 뉴스를 접한다고 답했다(Reuter Institute, 2019). 디지털플랫폼에서 높은 비율로 뉴스가 소비되는 것은 전 세계적인 현상이지만, 한국이 조사대상 38개국 평균인 55%보다 30%포인트나 높아서 그 의존도가 매우 컸다.

2) 개인의 부상과 접착제로서 알고리즘

매스미디어로 대변되는 대중사회에서 대중은 원자화되고 고립된 무기력한 존재였다. 그러나 디지털 네트워크 사회에서 개인은 소프트웨어에 의해 상호연결된 자아로 존재한다. 관심을 공유하는 개인들로 구성된 연결망 사회 속에서 알고리즘과 같은 소프트웨어는 접착제와 같은 역할을 한다. 특히, 소셜미디어는 이용자의 선호정보와 행동정보를 바탕으로 분석한 알고리즘을 통해 관계를 확장시킨다. 개인의 자아가 투영된 커뮤니티가 형성됨으로써, 자아의 표출과 자의식의 강화, 선호의 강화가 동시에 일어난다.

이 같이 알고리즘에 의해 매개되는 소셜미디어 환경은 어떤 특징을 갖는가? 먼저 미디어 메시지의 개인화와 선별적 소비가 가속화되고 사회적 규범압력이 약화되었다. 다른 한편으로 개인성이 부각되면서 정치나 사회적 사안에 대해 분화가 촉진되고 대중매체나 정당, 대의제 기구와 같은 정치 매개집단의 규범적 영향력이 급격하게 약화되었다.

또한 네트워크 공간에서 개인은 단순한 수동적 소비자가 아니라 정보를 주체적으로 생산하고 유통할 수 있는 존재로 바뀌었다. 미디어시스템은 더 이상 사회 규범의 일방향적 생산자로 자리매김하기 어려워졌다. 이처럼 사회적 규범의 압력으로부터 상대적으로 자유로워진 디지털 시대의 개인들은 욕구 기반의 콘텐츠 소비를 가속화시키고 있다. 콘텐츠 장르가 다양화되고 탈사회규범적 콘텐츠가 생산 및 소비되는 한편, 마이너리티

문화도 부상하고 있다. 이 같은 현상만 보면, 개인에게 보다 폭 넓은 선택성이 부여된 것은 맞지만, 그것은 어디까지나 제한된 선택행위이다. 알고리즘에 의해 인도된 선택지 내에서 일어난 결정물이기 때문이다. 따라서 네트워크 공간의 자유로운 개인의 상은 실제와 다른 허상일수도 있다,

셋째, 개인간의 대화가 정치커뮤니케이션의 중심에 자리잡기 시작했다. 벤클러(Benkler, 2006)는 이런 현상을 상업저널리즘 또는 매스미디어로 대변되던 일방향적인 '시장 모델'(market model)을 대체하는 '사회적 대화 모델'(social conversational model)로의 변화라고 칭했다. 벤클러는 언론인과 정치인, 전문가로 구성된 엘리트 정치매개집단에 의해 형성된 사회적 의제가 이용자에게 전이되는 방식에서, 누구나 정보의 발화자가 될 수 있는 '네트워크 공론장'으로 전환되었다고 보았다. 이 같은 공론장은 개인들간 공통된 관심으로 연계된 동등집단(peer group)으로 상호협력과 공조를 일으킨다. 탈중심적 구조를 갖는 네트워크 공론장은 구성원들간에 평등한 커뮤니케이션을 가능하게 한다는 점에서 긍정적이지만, 다른 외부효과 즉, 정보의 필터링이나 의견간 상호작용에 의한 극단화 현상이 나타난다.

전자민주주의에 대한 낙관적 시각은 기술을 중립적으로 바라보고, 이성적이고 합리적인 인간행동과 공동체의 자기 정화기능, 그리고 상호작용이 이상적 담화로 이어질 것이라는 순진한 믿음에 기초해 있었다.

그러나 웹이 발명된지 30년이 더 지난 현재, 우리는 민주주의에 대한 도전과 시련에 직면해 있다. 개인미디어를 통한 허위조작정보(가짜뉴스)의 확산, 디지털 사회갈등과 대립 등 과거 매스미디어 시대에서 경험하지 못한 새로운 부작용이 등장한 것이다. 개인 자율성의 증대는 주류미디어의 권력을 약화시키고 탈중심적 구조를 창출했지만, 다른 한편으로 허위기만정보가 걸러지지 않고 급속하게 유통되는 구조를 양산했다. 디지털기술을 통해 개인주의적 문화가 조장되고, 검증되지 않은 정보가 유통되고, 공동체의 연대가 약화되는 부작용이 확인되고 있는 것이다.

이런 현상을 이해하는 데 있어서, 우리는 기술이 매개하는 방식과 기술에 의해 만들어진 소통구조, 그리고 일어나는 커뮤니케이션 양식에 주목할 필요가 있다. 이 글에서는 전통적 정치매개집단을 대체하고 있는 비인간행위자인 알고리즘의 영향에 초점을 맞추고자 한다. 네트워크 공간에서 알고리즘이 기능하는 바와 인간행동에 미치는 사회심리적 특성과 현상의 상호작용을 다루고자 한다.

1) 현실의 재구성자로서 알고리즘

알고리즘(algorithms)은 문제를 해결하는 데 필요한 단계의 순서를 명시하는 구체적인 계산법(MacCormick, 2011)으로 일반적으로는 어떤 과업을 수행하거나 문제를 해결하는 일련의 절차를 의미하기도 한다. 어떠한 목적을 달성하기 위해 미리 짜여진 수학적 모형이라 할 수 있다. 최근 들어 정보통신기술에 인공지능 기술이 빠르게 접목되는 소위 '지능정보혁명'이 본격화되면서 딥 러닝(deep learning)과 같은 인공지능 알고리즘이 4차 산업혁명과 미래산업의 핵심 키워드로 급부상하고 있다(황용석, 2017).

알고리즘은 다양한 응용서비스 영역에 침습되어 있는 수학적 모형이자 구현체로서 그 구현방식이 우리의 경험 영역에서 확인되지는 않는다. 그러나 그 결과물은 우리의 일상 현실을 재구성하는 영향력이 큰 기술이다. 알고리즘은 세계를 바라보는 우리의 인식을 바꾸고, 상품이나 정보를 선택하는 행동에 영향을 미친다. 또한 사람들이 여론의 흐름을 인식하고 집단행동을 하는 데 필요한 준거 정보를 제공한다. 경제적으로는 인터넷상의 다양한 상품시장에서 시장 참여자들의 경쟁행위에도 영향을 미친다.

이처럼 알고리즘은 사회 전 분야에 영향을 미치고 있다고 해도 과언이 아니다. 디지털기술에 대한 의존도가 점점 높아질수록 인간은 보다 복잡한 정보환경에 놓이고 되고, 그 결과 인지적으로 정보를 처리하는 데 어려움을 겪게 된다. 알고리즘은 인간의 의사결정을 보조하는 것을 넘어서서 인간의 일상을 재구성한다.

오늘날 우리가 경험하고 있는 초연결 사회(Hyper-Connected Society)는 사람과 사람, 사람과 사물, 사물과 사물이 지능적으로 연결되는 네트워크 기반 사회로서, 상시연결과 접근이 가능하며 끊임없이 데이터가 생성되고 수집되는 구조를 갖는다(Fredette, et al., 2012). 네트워크가 복잡할수록 알고리즘에 대한 의존도는 커질 수 밖에 없다. 그런 면에서 우리는 정보서비스의 주체로서 스스로 선택하는 권리를 행사하는 것처럼 보이지만, 보이지 않는 기술에 의해 선택당하고 있다 해도 과언이 아니다.

유형	사례	
검색 애플리케이션 (search applications)	general search engines	e.g.,Google search,Bing,Baidu
	special search engines	e.g.,genealogy:Mocavo,pictures:Shutterstock,social media:Social Mention
	meta search engines	e.g.,Dogpile,Info.com
	semantic search engines	e.g.,Yummly
	questions &answers services	e.g.,Ask.com
콘텐츠 집적 애플리케이션 (aggregation applications)	news aggregators	e.g.,Google News,nachrichten.de
관찰/감시 애플리케이션 (observation/surveillance applications)	government/intelligence surveillance	e.g.,Raytheon RIOT
	monitoring of corporate/private ICT infrastructures and usage	e.g.,Spector,Spytec,Splunk
	detection of illegal content	e.g.,PhotoDNA for child pornography
진단/예측 애플리케이션 (prognosis/forecast applications)	predictive policing	e.g.,PredPol
	predictive profiling predicting developments	e.g.,success of music (scoreAhit,Music Xray)
	diffusion of diseases	Google Flu Trends
필터링 애플리케이션 (filtering applications)	spam filter	e.g.,Norton
	child protection filter	e.g.,Net Nanny
recommendation applications	recommender systems	e.g.,for music (Spotify),films (Netflix)
스코어링 애플리케이션 (scoring applications)	reputation systems:music,film,etc.	e.g.,ebay's buyer/seller reviews
	news scoring	e.g.,reddit,Digg
	credit scoring	e.g.,Creditkarma
	social scoring	e.g.,Klout
콘텐츠 제작 애플리케이션 (content production applications)	algorithmic journalism	e.g.,Quill;Quakebot
배치 애플리케이션 (allocation applications)	computational advertising	e.g.,Google AdSense,Yahoo!Bing Network
	algorithmic trading	e.g.,Quantopian

[표] 알고리즘의 기능에 따른 유형화
출처: Latzer et al.2015.

2) 알고리즘, 데이터 권력과 편향의 우려

알고리즘에 대한 비판주의적 시각은 파스 쿠 알레(Pasquale, 2015)의 저서에서 확인된다. 그의 설명은 지식의 통제를 다루는 푸코주의(Foucauldian) 관점과 맞닿아 있다. 이 입장은 알고리즘과 그것이 구현되는 디지털플랫폼을 권력체계로 바라본다. 푸코주의에서 보면, 소셜미디어와 같은 디지털플랫폼에서 알고리즘은 사람들의 대화를 조절하는 매커니즘으로 이해된다. 디지털플랫폼은 광범위하게 데이터를 수집하고 데이터에 대한 통제권을 행사한다. 알고리즘은 디지털플랫폼이 데이터를 통제하는 수단이자 그 결과물이다. 이용자들이 디지털플랫폼에서 경험하는 환경은 알고리즘에 의해 조율된 환경이자 이용자들간의 관계이다.(Gozdecka, Ercan 및 Kmak, 2014).

디지털플랫폼은 광범위한 이용자 데이터를 수집함으로써 이용자를 '수량화 된 자아들(quantified selves)'로 만들고 이용자들의 일상적인 선택행동에 영향을 미친다. 데이터를 둘러싼 힘의 불균형은 알고리즘이 '블랙 박스 (black box)'로 은유화되는 데서 찾을 수 있다(Timmermans, 2019). 디지털플랫폼을 이용하는 이용자들은 자신의 데이터가 어떻게 수집되는지 그리고 자신에게 제시되는 추천 서비스가 어떤 원리로 제공되는지 알 수 없다. 온라인 애플리케이션 프로그램을 사용하는 동안 사람들은 자신의 정보가 어디로 전송되는지, 정보가 어떻게 사용되는지 그리고 그러한 행동의 결과를 알 수 없다.

코멘과 라이저슨(Cormen & Leiserson, 2009)은 알고리즘은 이용자의 모든 행동데이터를 추적하면서 입력과 산출을 반복하는 동태적인 구조를 갖는다고 말했다. 알고리즘의 각 항목은 조작 가능한 조건에 놓여 있으며, 어떻게 데이터에 속성을 부여하는가에 따라 다른 결과를 산출할 수 있다.

알고리즘의 영향력은 사회적 차원에서 차별의 문제로 대두되었다. 알고리즘 차별은 두 가지로 구분할 수 있다. 먼저, '차별적 처우(disparate treatment)'이다. 이는 알고리즘 설계에 '의도적'으로 차별적 요소를 포함시켜 생긴 결과를 말한다(Zafar, Valera, Rodriguez, & Gummadi, 2017). 알고리즘 설계 과정에서 특정 집단의 속성에 차별적인 값을 부여하는 것을 말한다. 미국에서는 알고리즘 기반 보험·금융 서비스 등에서 사회적 약자(protected class)에게 차등적인 서비스를 제공하는 것을 법으로 금지하고 있으며(Steel & Angwin, 2010), 마찬가지로 유럽에서도 보험사가 특정 집단에 더 불리한 보험 조건을 제시하는 것을 불법으로 판시한 사례가 있다(European Court of Justice, 2011). 의도적으로 약자에 대한 차별적 모델링을 하는 것은 사회적 차별을 양산한다.

차별의 또 다른 차원은 알고리즘의 '차별적 효과(disparate

impact)'이다. 차별적 효과는 알고리즘의 수학적 모델이 '비록 정확하고 근거가 충분하고, 투명한 방식으로 결정되었다 하더라도' '결과적으로' 특정 속성의 집단(성, 연령, 지역 등)에게 차별적 처우가 일어나는 것을 말한다(Feldman et al., 2015). 그 원인은 학습데이터의 편향 등 다양하다. 황용석과 김기태는 추천 알고리즘의 편향과 차별 문제를 파악하기 위한 방법론적 논의를 정리하면서 다양한 편향의 원인을 제시한 바 있다(황용석, 김기태, 2019).

인공지능 알고리즘이 중립적이고 공정할 것이라는 일반인들의 믿음과 다르게 최근 알고리즘이 차별적이고 편향된 결과를 도출한다는 부정적인 사례가 보고되고 있다. 한국에서도 음원사이트의 순위 조작, 영화 추천 사이트의 평점조작이 사회적 이슈로 등장해 알고리즘 편향 문제가 부각된 바 있다.

알고리즘 편향과 관련해서는 다양한 증거들이 제시되고 있다. 미국 백악관이 2016년 5월에 발표한 '빅데이터: 알고리즘 시스템, 기회와 시민권' 보고서는 알고리즘이 편향된 결과를 내는 데이터 선택상의 요인 4가지를 밝혔다(Whitehouse, 2016). 첫째, 데이터 자체를 잘못 고른 것. 둘째, 불완전·부정확하고 시기에 안 맞는 데이터를 사용한 것. 셋째, 편향적인 선택. 넷째, 역사적인 편향성 등이다. 이 보고서가 의미하는 바는 편향적인 결과는 일부러 의도하지 않더라도 데이터 그 자체의 문제점 때문에 공정하지 않고 편향된 결과가 나올 수 있다는 것이다.

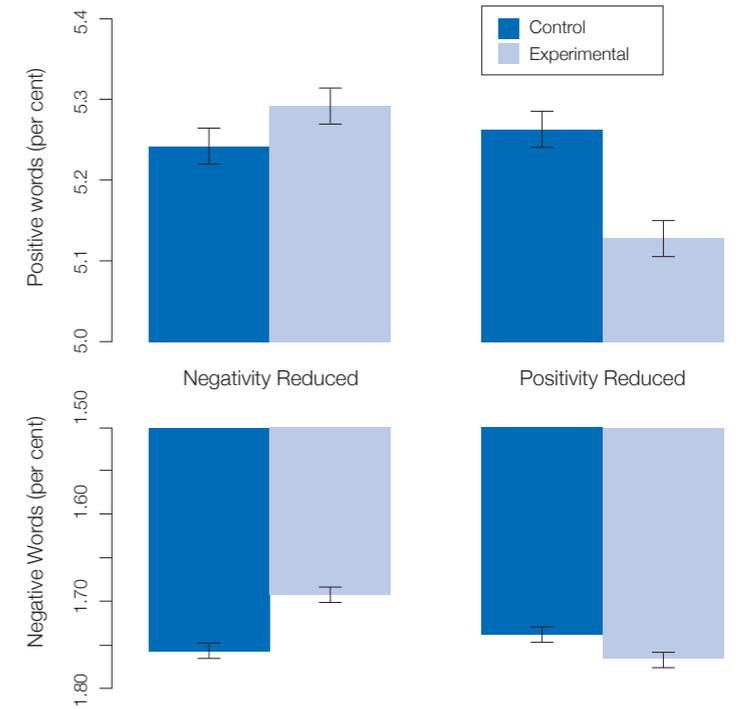
미국 비영리 인터넷 언론 '프로퍼블리카(Propublica)는 알고리즘이 가져올 수 있는 '기계의 편향(Machine Bias)'이라는 제호의 기사에서 데이터의 객관성 속에 숨겨진 문제점을 경험적으로 제시했다. 프로퍼블리카는 컴파스(Compas)라고 불리는 미국 법원의 재범 위험 평가 알고리즘을 문제 삼는다. 컴파스의 위험 평가 알고리즘에 따르면, 미래의 범죄자로 낙인 찍힐 위험은 흑인 피고인이 백인 피고인에 비해 두 배에 달하고 백인 피고인들은 흑인 피고인들에 비해 종종 낮은 위험을 가진 것으로 잘못 판정되고 있다는 것이다. 최근 데이터 사이언티스트 아베 공(Abe Gong)은 컴파스의 알고리즘을 통계적으로 재검토한 후 컴파스에서 나타나는 이러한 인종적 편향은 데이터 자체의 부정확성이나 컴파스 알고리즘의 연산공식상의 오류에서 기인하기보다는 단지 외부 현실에 존재하는 편견 혹은 편향(인종 편향)을 데이터와 알고리즘이 그대로 반영한 결과일 뿐이라고 지적한다. 이와 유사하게 2015년 카네기 멜론 대학 연구 결과 구글의 온라인 광고시스템은 여성보다 남성에게 높은 임금의 직업광고를 추천한다는 점이 드러나 논란이 되기도 했다(황용석, 정재관, 김기태, 2018). 이와 같은 사례는 표본통계와 달리 알고리즘에 사용되는 빅데이터는 모수를 그대로 반영하기 때문에 알고리즘이 데이터에 의해

학습된 현실세계의 편견을 재강화하는 측면을 보여주고 있다. 구글의 광고가 남성에게 더 높은 임금의 구인광고를 내보낸 것은 구글이 데이터를 조작했기 때문이 아니라 기존의 축적된 데이터를 모아서 추천하다보니 현실을 왜곡할 수 있다는 것이다. 이처럼 알고리즘 편향의 원인을 찾아내기 어렵다는 점은 현재 우리가 당면한 난제이다.

3) 알고리즘 처지에 따른 인지적·판단적 효과

알고리즘 조작에 의한 사람들의 인지·선택·판단상의 편향 현상은 사회과학계에서 여러 번 실험으로 검증한 바 있다. 구안과 컷렐(Guan & Cutrell, 2007)은 시선 추적(eye tracking) 방법을 사용하여 검색 알고리즘에 의해 결정되는 검색 순위가 사람들의 검색에 어떤 영향을 미치는지를 검증하였다. 이용자들은 두 번째에 나타난 검색 결과가 보다 관련성이 있더라도, 첫 번째 검색 결과를 클릭하는 경향이 보였다. 레즈닉과 알버트(Resnick & Albert, 2014)도 역시 시선 추적 방법을 사용하여 상업적 페이지 내에서 웹 배너 광고의 위치가 배너광고를 무시하려는 이용자의 경향성(Banner blindness)에 영향을 미침을 보여주었다. 알고리즘은 여러 측면에서 현실세계에 큰 영향을 끼칠 수 있다. 특히 풍부한 정보에 접근 가능하도록 도와주는 검색 알고리즘이 일련의 조작에 의해 편향을 드러낸다면 그 영향력은 더 크다 할 수 있다.

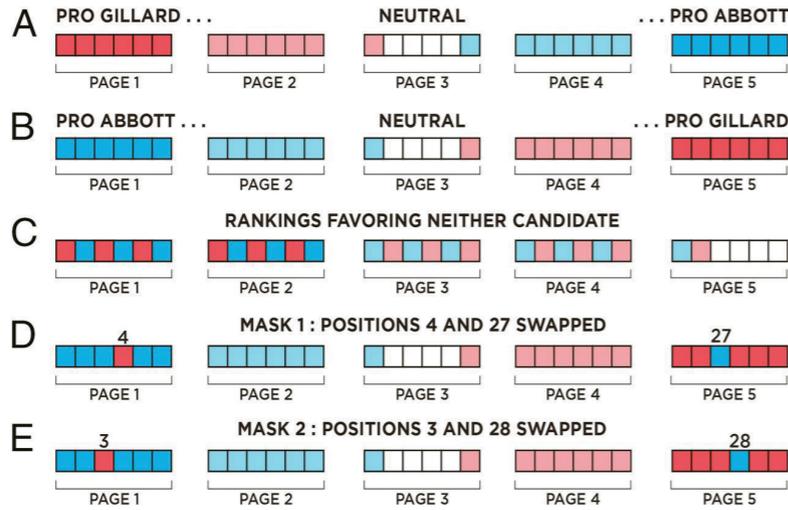
알고리즘을 의도적으로 조작했을 때의 영향력을 추적한 연구도 있다. 대표적인 사례는 페이스북(Facebook) 코어데이터과학팀 소속 연구원 아담 크레이머가 실시한 감정전이(emotional contagion) 실험이다.(Kramer, Guillory & Hancock, 2014) 감정전이란 다른 사람의 감정이 내게 옮겨오는 현상을 일컫는다. 친구가 울면 내가 슬퍼지는 것은 감정전이 때문이다. 페이스북은 이런 현상이 SNS에서도 나타나는지 알아보기 위해서 실험군과 비교군으로 나누고 실험군 68만9003명의 뉴스피드 알고리즘을 조작했다. 그 결과 페이스북에서도 감정 전이 현상이 나타났다. 긍정적인 게시물이 줄어들면 이용자는 긍정적인 표현을 줄이고 부정적인 게시물을 더 많이 올렸다. 반대로 뉴스피드에 나타나는 부정적인 게시물이 줄어들면 이용자는 긍정적인 게시물을 더 많이 올렸다. 즉 뉴스 추천 알고리즘 조작으로 인하여 이용자의 뉴스 콘텐츠 선택·이용 상의 감정전이라는 일종의 편향이 발생한 것이다.



[그림] 페이스북의 감정전이 연구결과
출처: 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*, 201320040.

알고리즘이 투표행동에 영향을 미칠 수 있다는 연구결과를 보여준 앱스타인과 로봇손(Epstein & Robertsson, 2015)의 연구를 주목할 필요가 있다. 이들은 광범위한 온라인 실험을 통해 검색결과의 순서효과를 입증했다. 내용의 변화 없이 단순히 검색결과의 순서배열만으로도 후보자에 대한 태도가 변화한다는 것을 입증했다. 미국과 인도에서 실시한 이들의 연구에 따르면 부동산에게 특정정당에 우호적인 검색결과를 보여주도록 프로그래밍한 검색엔진을 만들어서, 피실험자들에게 선거 정보를 찾도록 했다. 그 결과 동일한 검색결과이지만 검색결과 순서만 변경되었음에도 20%의 지지후보가 옮겨진 것으로 확인되었다.

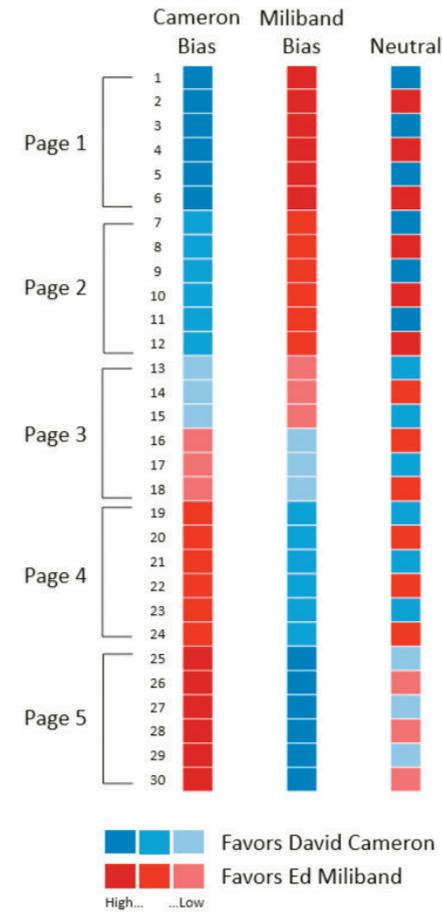
이후 앱스타인과 동료들(Epstein, Robertson, Lazer, & Wilson, 2017)은 2015년의 연구(Epstein, & Roberson, 2015)를 발전시켜 검색 결과가 편향되어 있음을 고지하였을 때 조작된 검색결과가 선호도에 미치는 영향에 차이가 있는가를 실험을 통해 검증하였다.



[그림] Epstein 연구의 2015년 1차 실험설계
 출처: 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

그들의 후속연구에서는 2015년 실시했던 기존의 실험 설계를 유지하는 한편 검색엔진에서 제공된 검색 결과가 특정 후보자에게 편향되었음을 알리는 메시지 창을 검색 결과 상단에 노출할 때(실험1)와 메시지 창의 노출과 더불어 각각의 검색 결과 항목이 어느 후보자에게 유리한 내용을 담고 있는지를 표시할 때(실험2)로 구분했다.

실험결과 조작된 검색 결과가 피실험자에게 미치는 영향은 노출된 경고 메시지가 강력할수록 줄어들었다. 반면 아무런 메시지가 표지되지 않은 실험그룹에 비해 편향 메시지가 노출된 그룹(그룹1: Low Alert 그룹)이 검색 후 후보자의 호감도 차이가 줄어들었다. 강력한 편향 경고 메시지에 노출된 그룹(그룹2: High Alert 그룹)의 경우, 한 쪽으로 편향된 결과가 제공된 검색결과를 읽은 후 오히려 초기 호감도가 검색 결과의 반대로 이동하는 것을 발견하였다. Epstein과 동료들은 이 연구에서 이용자들이 편향을 인지할 경우 이를 검색 결과에 반영하여 인지된 내용을 조절하려는 경향이 있음을 지적하며 검색 결과에 대한 공정하고 이해가능한 설명은 이용자로 하여금 검색 결과의 공정성과 신뢰성을 판단하는 데 기여할 수 있음을 제안하였다. 한편, 연구자들은 이 같은 연구결과가 검색의 편향성을 경고하는 가상의 조작방식으로 검증된 것임을 지적하며 실생활에서의 검색 엔진은 이와 같이 통제된 환경에서 결과가 주어질 수 없음을 지적하였다(황용석, 정재관, 김기태, 이현주, 2018).



(a) Search rankings by bias group assignment.

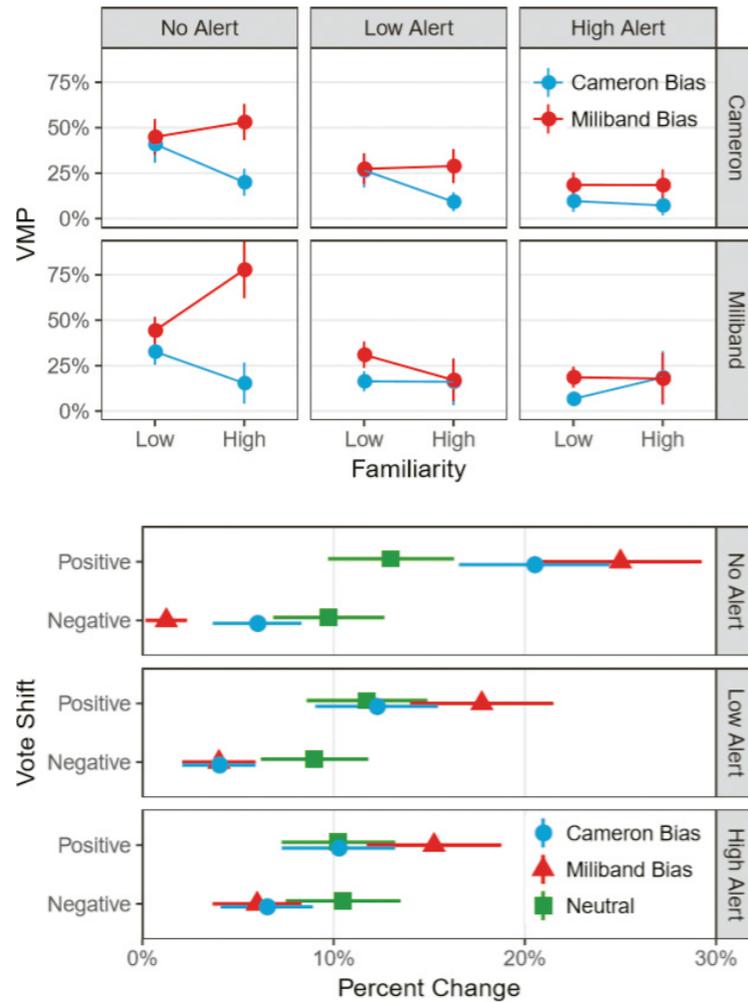


(b) Low alert search condition.



(c) High alert search condition.

[그림] Epstein 연구의 2017년 2차 실험설계
 출처: Epstein, R., Robertson, R. E., Lazer, D., & Wilson, C. (2017). Suppressing the search engine manipulation effect (SEME). Proceedings of the ACM: Human-Computer Interaction, 1, 42.



[그림] Epstein 연구의 2017년 2차 실험결과
출처: Epstein, R., Robertson, R. E., Lazer, D., & Wilson, C. (2017). Suppressing the search engine manipulation effect (SEME). Proceedings of the ACM: Human-Computer Interaction, 1, 42.

3. 알고리즘과 사회심리의 상호작용은 정치적 양분사회를 부추기나?

1) 필터버블과 집단극화

한국사회는 심각한 의견 대립을 겪고 있다. 의견의 분포가 양분되는 것을 넘어서서 상호 적대적인 태도와 극단주의적 행동이 다양한 정치사회적 사안에서 나타나고 있다. 뿐만 아니라 가짜뉴스로 통칭되는 허위조작정보는 온라인공간의 정보질서를 훼손하면서 사회갈등을 부추기고 있다. 이런 현상을 설명하는 데는, 정치현안의 특징과 사회문화적 변수를 고려할 수 있을 것이다. 그러나 오늘날 대부분의 사회적 담론이 온라인 공간에서

펼쳐지고 있다는 점을 고려할 때, 그 공간의 기술 매개적 특성과 참여자의 사회심리학적 특성을 동시에 고려할 필요가 있다.

그 중심 용어는 필터버블(filter bubble)과 에코챔버(eco chamber)이다. 이들 현상은 다양한 선택지에 놓인 개인의 심리적 요인과 알고리즘의 매개성이 상호작용한 결과물이다.

언론과 같은 전통적인 정치매개집단은 사회의 중요한 의제를 정리해서 전달하는 역할을 수행한다. 언론이 중요하게 생각하는 의제는 다양한 주제를 포괄하게 되며, 이런 주제들은 대부분 정치나 사회문제 등 경성 의제에 가깝고, 다양한 관점을 포괄하는 경우가 많다. 그러나 개인의 선택권이 주되게 작동한다면, 사회적 의제의 중요성 보다는 개인의 선호가 보다 강하게 작용한다. 이른바 자신이 갖고 있던 신념이나 선유경향이 정보를 선택하는 데 더 크게 영향을 미친다. 자신의 선호가 정보를 선택하는 데 영향을 미쳐 원하는 정보만 이용하는 것을 '선택적 노출'이라고 한다. 선택적 노출을 통해 개인이 갖는 편향이 온라인에 투영되고 알고리즘에 의해 추천 및 강화되는 현상은 자신과 유사한 정보에 갇히게 되는 필터버블 현상으로 이어진다.

필터버블은 일라이 파리저(Eli Pariser, 2011)가 자신의 저서에서 소개한 개념으로 일종의 알고리즘 효과이다. 우리가 정보를 습득하는 과정에는 개인의 경험과 태도, 동기에 의해 만들어진 선유경향에 영향을 받는다. 정보를 편향적으로 소비하는 이유는 자신에게 익숙하거나, 동기를 충족시키고, 자신과 유사한 의견을 확인함으로써 심리적 만족감과 안정감을 얻을 수 있기 때문이다.

우호적이고 선호적 정보만을 추구하는 행동은 알고리즘의 학습체계에도 영향을 미친다. 대부분의 추천알고리즘들은 행동의 유사성을 기반으로 확률적으로 다시 일어날 행동을 추출해 제안한다. 즉, 온라인 정보추구 행동의 동질성을 강화시키는 방향으로 유사한 정보를 계속해서 추천한다. 이런 추천방식은 자기 생각과 다른 이질적인 의견은 배제해서 우호적인 정보나 여론에 갇히는 결과를 부른다.

결과적으로 필터버블은 개인맞춤 알고리즘에 의해 생기는 정보편식현상을 의미하며, 이용자가 제한된 주제의 정보에 갇혀 정보의 다양성과 여론지각의 균형성이 떨어진 상태를 말한다.(김선호, 2018)

정치커뮤니케이션 관점에서 필터버블은 같은 정보를 공유하는 유유상종의 정보 및 의견집단이 형성되는 결과를 낳는다. 상호교류가 없이 동질성이 높은 의견집단은 사회적 분극화를 부추지게 된다(Pariser,2011). 유사한 집단 내에서 일어나는 상호작용은 또 다른 심리적 기제인 동조화 효과(confirmative effect)를 양산하고 더 나아가 집단극화(group polarization)로 이어진다. 사회심리학에서 집단극화는 집단의 의사결정이

구성원들의 평균적인 의견보다 상호작용 이후 더 극단적 방향으로 이동하는 것을 말한다. 정낙원(2014)은 디지털미디어가 집단극화를 유발하는 과정을 세 가지로 설명했다. 첫째, 온라인 환경이 기존의 태도를 강화시키는 선택적 커뮤니케이션(예를 들어, 선택적 노출)에 유리한 속성을 제공하기 때문이다. 둘째, 온라인 활동 중 필연적으로 발생할 수밖에 없는 이질적 커뮤니케이션도 이해와 관용의 폭을 넓히기보다 선유경향을 강화하는 편협한 인지정교화를 유발할 가능성이 높기 때문이다. 마지막으로, 익명성으로 인해 이용자들이 탈개인화(deindividuation)되고, 이는 탈규범화(denormalization) 및 집단정체성(group identity) 강화로 이어져 집단 극화 현상이 나타난다는 것이다. 집단극화는 극단화와 구별된다. 집단극화는 토론이나 정보교류를 할수록, 점점 더 한 쪽방향(예를 들어, 보수 또는 진보)으로 집단의 평균값이 이동하는 현상을 말한다. 온라인 공간에서의 상호작용은 집단극화를 강화하는 심리적 조건을 제공하는데, 여기에 개인화 추천 알고리즘이 개입하면 집단극화에 더 유리한 의견집단이 형성된다.¹

소셜미디어의 맞춤형 알고리즘은 이러한 집단 구성원간의 의견 동조화를 더욱 촉진시켜 극단적인 정치적 대립각을 더욱 노골화시킨다는 견해도 있다.

2) 확장편향을 부추기는 에코챔버효과

앞서 언급한 대로, 필터버블은 개인화 알고리즘의 결과물로서, 검색엔진이나 소셜미디어가 선별적으로 정보를 제공하거나 선별적으로 관계를 추천해서 생기는 현상이다. 사용자의 동의에 의해서라기 보다는 알고리즘이 구성한 환경에 놓이는 것을 말한다. 그 결과 이용자들이 자신의 관점과 멀리 있는 정보가 필터링 되어 이념적 거품에 가두어지는 것이다. 이견이 줄어들고 대신 같은 목소리가 반복됨으로써 개인의견이 강화되는 증폭효과가 발생한다. 네트워크 관점에서 보면 필터버블은 호모필리(homophily) 또는 유유상종 효과를 부른다. 호모필리는 말 그대로 사회적 지위나 성향이 비슷한 사람들이 상호작용한다는 의미이다. 호모필리는 사회인구학적으로 유사한(성별, 경제력, 학력 등) '사회적 호모필리'와 신념이나 가치 중심의 '가치적 호모필리'로 구분할 수 있다(정낙원, 2014). 일반적으로 디지털공간에서는 개인의 사회인구학적 속성이 확인되기 어렵다. 온라인 환경에서 나타나는 탈개인화 특성

1 집단극화를 설명이론은 설득주장이론(persuasive arguments theory), 사회비교이론(social comparison theory), 자기범주화이론(self-categorization), 선택적 노출과 지각(selective exposure & Attention), 명성 효과의 사회정체감 모델(Social Identity model of Deindividuation Effects: SIDE) 등이 있다.

때문이다. 이로 인해 개인의 관심분야나 정치적 신념을 중심으로 한 호모필리가 형성되는 경향이 있다.

호모필리에서 같은 목소리가 울리는 것을 반향실 효과 또는 에코챔버(echo chamber)효과라고 한다(Sustein,2009). 이것은 밀폐된 반향실에서 자신과 같은 목소리가 메아리치고 증폭되는 현상으로 점차적으로 편향된 사고를 강화시킨다.

“에코챔버효과”는 “필터버블”과 편향으로 이어지는 방식이 다르다. 필터버블은 알고리즘의 추천효과에 초점이 맞춰져 있다면, 에코챔버는 이용자의 심리적 특성 즉, 자발적인 선택에 더 주목한 개념이다(김선호,2018). 두 개념이 유사하지만, 편향을 만들어내는 메커니즘을 설명하는 방식에서 차이가 있다.

에코챔버 개념이 소개된 이후 다양한 문헌에서 그 효과를 강조하고 있지만, 실제 여러 경험연구에서 효과가 상반되게 나타나고 있다. 소셜미디어에서 친구로부터의 관계를 맺는데 승인을 받거나 관계의 순위를 보여주는 알고리즘이 정보소비에 어떤 영향을 미치는 가를 탐구한 앤드류 구스 등(Guess et al., 2018)의 연구에서는 에코챔버효과가 잘 나타나지 않았다. 그러나 콜레오닐(Colleoni, Rozza, & Arvidsson, 2014)의 트위터 연구에서는 유사한 정치성향간 관계강화나 토론증진 효과가 확인되었다.

가렛(Garrett, 2009)은 인터넷과 온라인 뉴스를 통해 어떠한 정치적 정보에 자신을 노출시킬 것인가에 대한 더 많은 통제력을 지니게 되지만, 사람들이 완전히 다른 생각들로부터 자신들을 고립시키는 것은 아니어서 인터넷으로 에코챔버가 형성한다는 근거가 없다고 주장했다. 소셜미디어 자체가 이견을 노출시키는 효과가 있다는 반대되는 주장도 있다. 박시 등(Bakshy et al., 2015)은 페이스북 로그데이터를 분석한 결과 페이스북 이용자의 친구 중 평균적으로 20%가 개인과 반대되는 이데올로기를 가지고 있었고, 소셜 미디어에서 생각보다도 반대의 담론에 많이 노출된다고 밝혔다(Bakshy et al, 2015).

에코챔버의 효과를 발견한 연구도 다수이다. 브라이트(Bright, 2016)는 23개국 90개의 정당에 대한 토론네트워크를 고려하여 트위터 데이터를 분석한 결과, 이념적인 측면에서 서로 떨어져 있는 정당 집단들끼리는 덜 상호작용하고 있었으며, 이데올로기적으로 극단에 있는 정당이나 개인이 에코챔버를 형성할 가능성이 특히 높은 것으로 나타났다.

필터버블과 에코챔버의 메커니즘이 온라인 상에서의 집단극화와 분극화를 강화하기는 하지만 전반적인 효과는 제한적이라는 주장도 있다. 소셜네트워크와 검색엔진이 사람들 간의 이데올로기적 거리를 확장시키기는 하지만 동시에 정치적으로 다른 의견에 대한 개개인들의 노출 역시 강화한다는 설명이 있다(Flaxman, Goel and Rao, 2016).

3) 에코챔버와 허위조작정보(가짜뉴스) 확산에 대한 연구

에코챔버현상은 속해 있는 집단의 동질성만으로 효과가 나타나는 것은 아니다. 의견 동조자가 얼마나 많은지에 따라 확증편향의 강도가 달라진다. 또한 가짜뉴스와 같은 허위조작정보를 진실된 것으로 여기거나 신념강화의 도구로 여겨 확산시키는 행동은 동기의 강도 즉, '동기화 추론(motivated reasoning)' 정도에 따라 그 결과가 다르다. 쿤다(Kunda, 1990)는 사람들은 자신이 원하는 결론에 도달하려는 방향성 동기와 객관적이고 합리적인 결론에 다다르려는 정확성 동기 중 방향성 동기가 강할수록 확증편향이 작동하며, 동시에 자신과 다른 의견에 대한 불인정편향(disconfirmation bias)이 강하게 작용한다고 밝혔다.

최근에는 가짜뉴스로 불리는 허위조작정보의 확산과 에코챔버효과를 연결짓는 경험적 연구들이 발표되고 있다. 코다 등(Cota, Pastor-Satorras, & Starnini, 2019)은 정치커뮤니케이션 네트워크에서 허위조작정보가 퍼지는 현상과 에코챔버의 관계를 분석했다. 이 글에서는 이 연구를 보다 구체적으로 소개해 이해를 돕고자 한다.

그동안 에코챔버 효과가 허위조작정보를 유통시키고(Del Vicario, A., Zollo, , Scala, & Quattrocchi, 2016), 의견양극화를 불러오며(Wojcieszak, 2010), 오프라인의 커뮤니케이션에도 부정적인 영향을 미친다는 연구(Dubois & Blank, 2018) 등 다수의 연구들이 발표되었다. 허위조작정보와 관련해서도 조사결과들은 다소 상이하게 나타났다. 그 이유는 분석하는 대상이 연구논문마다 다르고 데이터처리방식과 조사모형이 크게 차이가 나서일 수 있다.

여기서 소개하려는 코다 등(Cota, Pastor-Satorras, & Starnini, 2019)의 연구는 트위터 이용자들의 정치적 성향을 계량화(quantifying)하고 이를 통해 에코챔버 효과를 검증하고자 한 흥미있는 연구이다. 이들은 2016년 브라질 전 대통령 딜마 루세프(Dilma Rousseff)의 탄핵에 대한 찬반여론을 분석했다. 2016년 3월 5일~12월 31일까지 트위터 API에서 탄핵과 관련한 323개의 키워드로 검색된 게시물을 수집한 후, 이 중 리트윗(Retweet:RT) 데이터는 콘텐츠 생산정도가 약하고 의견전파의 경로가 명확치 않아 배제하고 트윗(Tweet) 메시지만 분석에 사용했다. 그

이유는 로봇에 의한 기계적인 트윗을 제거하기 위해 팔로잉-팔로워 관계가 약한 교류(소셜봇의 경우 일방향적이므로)를 제거하고 상호교류가 강한 관계(SCC: Strongly Connected components)만을 분석에 사용했다. 이 과정을 통해 분석에 사용된 표본은 총 31,412명(탄핵찬성:13,925명, 반대:16,257명, out-degree:26.5, 인터랙션: 1,552,389회)이다.

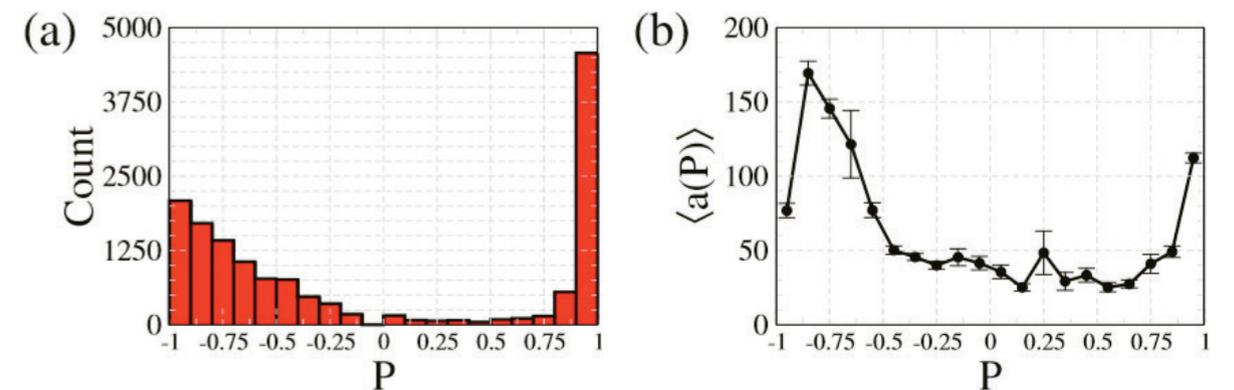
가. 의견양극화

분석과정을 보면, 먼저 의견양극화가 확인되었다. 이를 위해 각각의 트윗메시지의 헤시테그를 기준으로 -1:반대, 0:중립, 1:찬성으로 코딩한 후, 한 이용자의 모든 트윗메시지에 대한 평균치를 산출하여 그 이용자의 정치적 성향을 -1~1사이의 값으로 산정했다.

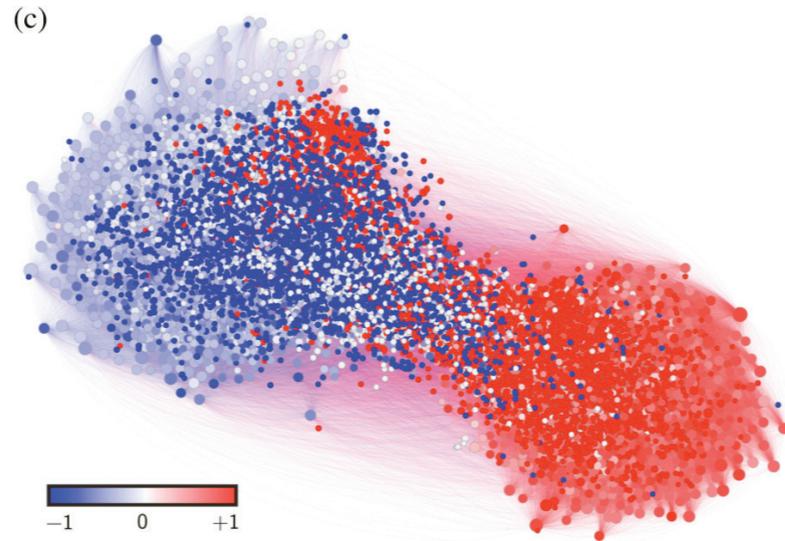
예를 들어, 라는 이용자가 ai만큼의 트윗을 작성하면 그 평균치로 정치적 성향을 계산했다.

$$P_i \equiv \frac{\sum_{t=1}^{a_i} S_t}{a_i},$$

분석결과 대부분의 트윗은 매우 반대(1) 혹은 매우 찬성(-1) 사이의 양 극단에 위치해 있다. 그리고 반대의견은 점진적으로 분포되는 반면 반대의견은 극단적 수치(1)에 편향되어 있는 것을 알 수 있다. ([그림] 참조). [그림]에서 보듯이 찬반의견과 게시글 간의 관계를 살펴보면 활동량이 높은 이용자일수록(트윗 게시물이 많을수록) 찬반여론의 극단에 위치함을 알 수 있다.



Tweet의 교류(팔로워와 팔로잉을 통한)를 살펴보면 같은 정치성향을 가진 집단간에만 활발한 교류가 일어나는 것을 볼 수 있다. 흰 색으로 표시된 정치적 중립집단만이 두 집단과 활발히 교류하지만 정치성향이 다른 두 집단은 거의 교류가 일어나지 않음을 알 수 있으며 탄핵에 반대하는 집단(빨간색)의 고립이 더 극명히 드러난다.



특정 변인으로 커뮤니티를 분류해주는 루바인 모듈러리티(Louvain modularity) 기법을 적용하여 커뮤니티를 분류해보면 세 번째로 강한 반대의견과 두 번째로 강한 찬성의견에 대부분의 이용자가 집중되어 있는 것을 알 수 있다(9,937명, 10,502명).

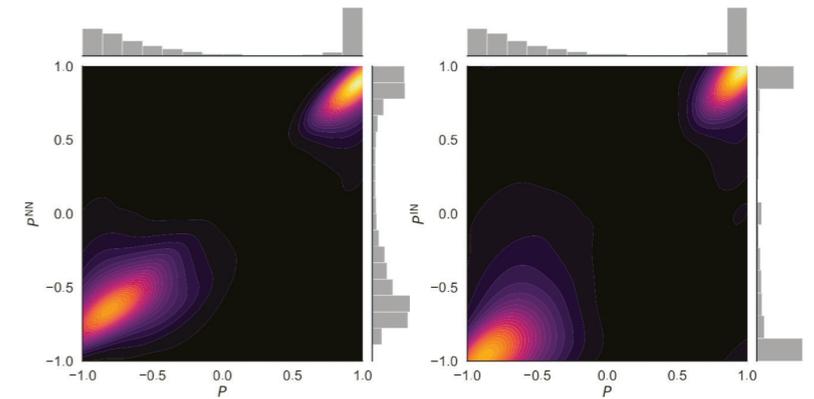
나. 에코챔버의 위상적(Topological) 증거

코다 등(Cota, Pastor-Satorras, & Starnini, 2019)은 이 연구에서 특정 이용자의 정치적 성향(P)과 가장 상호작용이 활발한 이용자의 정치적 성향(PNN) 및 받은 트윗메시지의 성향(PIN) 간의 상관관계를 살펴보았다. 아래 그림에서 보듯이, 이용자의 정치적 성향은 가장 가까운 트위터친구의 정치적 성향과 상관관계가 있으며 작성한 트윗의 정치성향 또한 받은 트윗메시지의 성향과 통계적으로 유의미한 상관관계를 보이고 있다.

Pearson correlation: P와 PNN $r=0.89$, $p<.001$

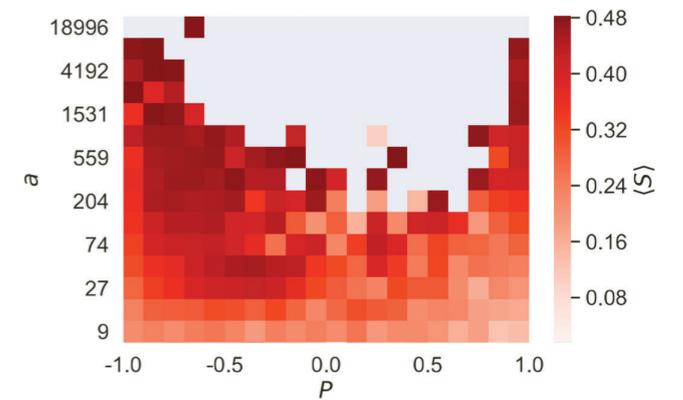
P와 PIN $r=0.80$, $p<.001$

한가지 흥미로운 점은 정치성향이 강할수록 상관관계가 크며(같은 성향끼리 고립되어 있으며), P가 0보다 클 경우(탄핵에 반대하는 사람일 경우) 그 고립정도가 더 큰 것을 알 수 있다.

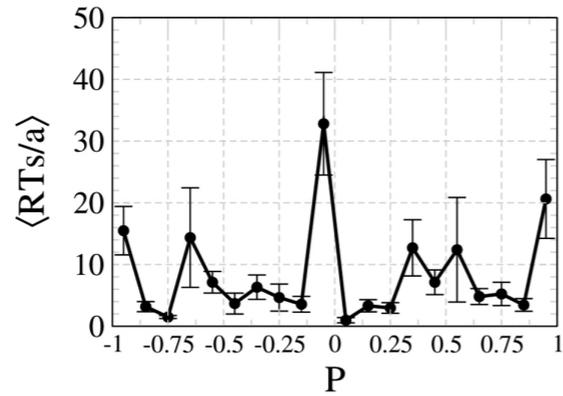


다. 정치적 성향이 정보보급에 미치는 영향

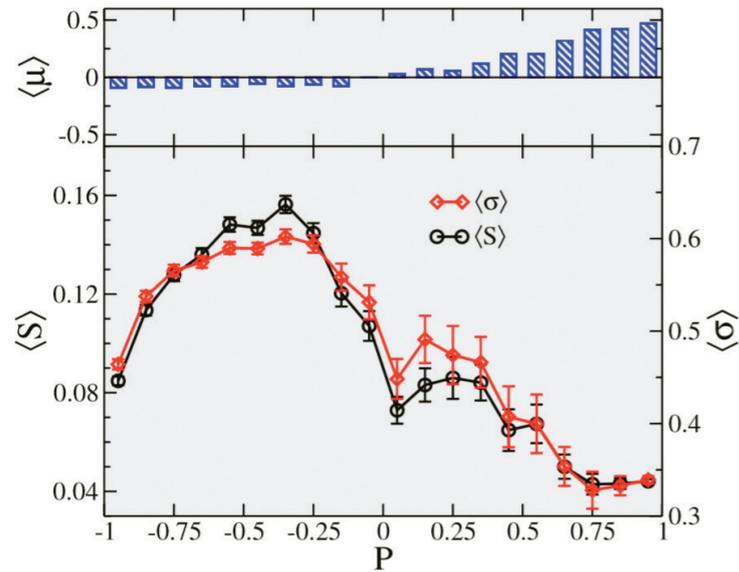
아래의 표를 보면 정치성향이 강한 트윗 이용자일수록 더 많은 리트윗을 받고 있는 것처럼 보인다.



그러나 그들의 트윗수를 리트윗수로 나누어 살펴보면(아래표 참조) 정치적 성향이 강한 이용자의 상대적 리트윗수(정보파급력)이 적은 것을 알 수 있다.



이들 연구자들은 이용자의 활동량(트윗수)를 통제한 후 그들의 정치적 성향이 정보파급력(spreadability)에 미치는 영향을 살펴보았는데, 중도-약한 찬성의 의견이 가장 넓게 전파되는 것을 알 수 있으며 강한 찬성과 강한 반대의 경우 전파력이 감소함을 발견했다,



위의 그래프에서 (메시지의 다양성)을 가진 경우 그 파급력(S) 또한 크게 분포함(그래프 상에서 위 아래로 뻗은 막대그래프 길이를 참조해서 해석) 알 수 있으나 의견의 극단에 있는 그룹의 경우 그 영향력이 거의 없음을 확인할 수 있다. 에코챔버에 갇힌 사람들은 다른 의견을 가진 사람들의 태도에 큰 영향을 못 미친다는 것이 확인되었다.

코다 등(Cota, Pastor-Satorras, & Starnini, 2019)의 연구는 최근에 발표된 논문으로 이전 논문들과 다르게 데이터의 전처리 및 검증방법 측면에서 훌륭한 연구설계를 보여주었다. 물론, 이 논문으로 디지털플랫폼에서 알고리즘에 의해 매개되는 에코챔버현상을 일반화하기는 어렵다. 그러나 그동안 다수의 사회심리적 이론과 알고리즘의 논리에 기반한 주장들이 현실과 동떨어져 있지 않다는 경험적 증거를 제공하고 있다.

4. 알고리즘과 민주주의를 위한 제언

우리는 알고리즘이 매개하는 초연결 사회에 진입해 있다. 이제 연결은 사람과 사람, 사람과 정보, 더 나아가 사람과 사물 및 사물간 연결이 가속화되는 밀도 높은 네트워크 사회로 진입했다. 연결의 복잡성이 커질수록 알고리즘의 필요성은 커지고 그 영향력 역시 확대될 것이다.

앞서 다룬 내용처럼, 알고리즘은 우리의 생활세계에 깊숙이 파고들고 있지만 우리는 그 존재감을 인식하지 못채 여러 가지 판단사항을 위임하거나 알고리즘이 추천하는 것을 자연스럽게 받아들이고 있다. 정치커뮤니케이션 현상을 이해함에 있어서도, 자신의 선택동기에 충실하고 학습된 알고리즘의 추천 메시지에 갇혀 편향된 여론지각과 정보습득을 하기도 한다.

이러한 편향된 지각과 닫힌 커뮤니케이션 구조는 우리 사회를 양분시키는 한 원인이다. 언론과 같은 전통적 정치매개집단의 신뢰와 영향력이 낮아질수록 이런 현상은 더욱 가속화된다. 우리에게 사실을 확인시켜주고 이질적인 정보를 교차시켜주는 언론의 신뢰 하락은 네트워크 사회에 큰 구멍을 만들었다.

이런 맥락에서 민주주의의 건강성을 확보하기 위한 노력을 정리해 보면, 다음과 같다.

첫째, 디지털시대 기술을 매개로 연결되는 관계 속에서 시민주의 능력(civic competence)의 중요성은 더욱 커지고 있다. 시민의 능력을 함양할 수 있는 디지털리터시는 유용한 정책대안 중 하나이다. 디지털리터시는 시민의 참여능력을 강조하는 디지털시민성교육(digital citizenship education)으로 그 개념이 확장되고 있다(황용석, 2018). 디지털리터시는 주어진 텍스트를 비판적으로 읽고 이해하는 능력, 그리고 자신의 사고를 표현하는 능력 뿐 아니라, 사회를 구성하고 있는 다양한 구성원들의 여러 의견과 그 가치를 식별하는 동시에 존중하며, 사회를 구성하는 시민으로서 소통하고 관계할 수 있는 능력(communicative competency), 즉, 문화적인 능력 (cultural competency)이자 시민적인 능력 (civic competency)을 포함하는 개념으로 이해되어야 한다.

즉, 이 개념은 단순히 도구적이고 기술적 능력의 함양에 국한하지 않고, 디지털 공간에 존재하는 시민들간의 소통능력의 증진을 강조한다.

특히, 디지털리터러시에서 빠질수 없는 중요한 하위 개념은 비판적 사고이다. 미래학자 니콜라스 카가 이야기한 '사유의 무능'을 겪는 무비판적 집단주의를 경계할 필요가 있다.

니콜라스 카(Nicholas Carr)

"당신이 볼 수 있는 것은 기기에 사로잡혀 있는 사고다. 온라인상에서 우리는 종종 우리 주변에서 일어나는 일을 망각한다. 기기를 통해 전달되는 상징과 자극의 홍수를 처리하면서 실제 세상의 모습은 점차 흐릿해지고 있다." (니콜라스 카, 최지향 옮김, 『생각하지 않는 사람들: 인터넷이 우리의 뇌 구조를 바꾸고 있다』, pp.176-177.)

둘째, 알고리즘은 다양한 형태로 우리의 판단에 영향을 미친다. 우리가 알고리즘을 완전하게 통제할 수 없다면, 알고리즘의 작동방식을 설명하는 것만으로도 일정정도 효과를 얻을 수 있다.

황용석, 정재관, 김기태, 이현주의 연구(2018)²에서도 알고리즘의 설명책임과 책무성이 알고리즘 효과를 약화시키는데 유의미하다는 것을 경험적으로 증명했다. 이들은 가상의 검색엔진 사용에 대한 실험연구를 통해 알고리즘 책무 조항으로서의 '설명을 요구할 권리'에 대한 실증적 연구를 시도하였다. 이들은 엡스타인(Epstein, 2015; 2018)등의 연구자에서 이용자가 편향된 검색결과를 인지하였을 때 정보의 평가에 미치는 영향을 설명방식과 연관지어 실험설계를 했다. 검색 알고리즘의 설명이 이용자에게 다르게 제공되었을 때 이용자들은 검색 알고리즘의 인지성(Awareness), 정확성(Correctness), 해석가능성(Interpretability)과 인지된 책무성(Accountability)을 평가하는 데에 차이가 있는 지를 실험을 통해 검증하였다.

연구결과, 알고리즘의 설명방식이 다르게 제공되었을 때 이용자들이 인식하는 인지성, 정확성, 해석가능성, 인지된 책무성에 통계적으로 유의미한 차이가 있음을 발견하였다. 검색 방식에 아무런 설명이 제공되지 않은 참여자들은 알고리즘의 네 가지 영역 모두 가장 높은 점수를 준 반면 알고리즘의 존재, 블랙박스의 영역(Why), 화이트박스의 영역(How) 중 하나의 정보만을 제공받은 참여자들은 보다 낮은 점수를 주었다.

2 아래 연구서에서 두 개의 실험연구와 이론적 개념을 확인할 수 있다.
황용석, 정재관, 김기태, 이현주(2018). 알고리즘 이용자 보호이슈의 방법론적 접근, 정보통신정책연구원 위탁연구보고서.

마지막으로, 알고리즘에 대해 복합적인 설명을 들은 이용자들은 검색 알고리즘에 대해 가장 비판적인 평가를 내렸다.

이 같은 연구결과는, 검색 알고리즘에 대한 설명방식이 복잡적이고 다양할수록 이용자들은 검색 결과의 정확성과 신뢰성에 대해 보다 명확히 판단할 수 있음을 의미하며, 이는 알고리즘이 이용자들의 권익을 위해 알고리즘의 논리를 일정수준에서 제공되어야 함을 시사한다. 디지털플랫폼 사업자들은 이용자들이 검색 결과를 능동적이고 비판적으로 분석하고 검색시스템의 신뢰성을 확보하기 위하여 각각의 서비스 제공자가 적용하고 있는 검색 알고리즘을 이해 가능한 형태로 포괄적으로 설명하는 노력이 필요하다. 정부 부처는 기업과의 협력을 통해 이를 지원 할 수 있는 방안을 구축할 수 있는 방안을 모색하여야 할 시점이다.

아울러 정보를 매개하는 알고리즘의 영향력이 가시화되면서, 국가단위의 정책적 대응도 나타나고 있다. 유럽연합(EU)이 2016년에 발표하고, 2018년 5월에 발효한 일반개인정보보호규정(GDPR: General Data Protection Regulation)을 통해 알고리즘 작동원리나 목적을 뚜렷이 명시할 것을 명문화하고 있다. 학계에서도 굿맨과 플렉스먼 (Goodman & Flaxman, 2016)의 논문에서 알고리즘의 설명책무성 원리를 처음 도입한 이래 이용자 보호와 알고리즘 규제에 관한 학문적인 논의를 이어오고 있다는 점을 고려해서 관련한 논의를 국내에서 활성화 시킬 필요가 있다.

참고문헌

- 김선호(2018). 필터버블, 플랫폼의 예방노력이 중요. 신문과 방송, 6-11.
- 노정규, 민영 (2012). 정치 정보에 대한 선택적 노출이 태도 극화에 미치는 효과. 한국언론학보, 56(2),226-248.
- 니콜라스 카, 최지향 옮김, 『생각하지 않는 사람들: 인터넷이 우리의 뇌 구조를 바꾸고 있다』, pp.176-177.)
- 정낙원(2014). "선택적 노출과 호모필리". 양승찬 외, 디지털 사회와 커뮤니케이션.2014 커뮤니케이션 북스.
- 황용석 (2017). "디지털과 리터러시". 디지털미디어와 사회, 나남 : 서울, 375-406.
- 황용석, 정재관, 김기태, 이현주(2018). 알고리즘 이용자 보호이슈의 방법론적 접근, 정보통신정책연구원 위탁연구보고서.
- 황용석, 김기태(2019). "개인화 서비스 진전에 따른 자동추천시스템 연구동향과 방법론적 특성연구", 사이버커뮤니케이션학보, 36(2), 221-254.
- Epstein, R., Robertson, R. E., Lazer, D., & Wilson, C. (2017). Suppressing the search engine manipulation effect (SEME). Proceedings of the ACM: Human-Computer Interaction, 1, 42.
- 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, 201320040.
- Bixby, S (2016, October 1). 'The end of Trump': how Facebook deepens millennials' confirmation bias.

Bastos, M., Mercea, D., & Baronchelli, A. (2018). The geographic embedding of online echo chambers: Evidence from the Brexit campaign. *PLoS one*, 13(11), e0206841.

Bauer, Johannes & Latzer, Michael eds (2015). *Handbook on the Economics of the Internet*. Edward Elgar.

Benkler, Y. (2006). *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University

Bright, Jonathan, Explaining the Emergence of Echo Chambers on Social Media: The Role of Ideology and Extremism (March 10, 2017). Available at SSRN: <https://ssrn.com/abstract=2839728> or <http://dx.doi.org/10.2139/ssrn.2839728>

Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of Communication*, 64(2), 317-332.

Cormen, T. H & Leiserson, C. E (2009). *Introduction to Algorithms* (3rd ed.). London, England: MIT Press.

Cota, W., Ferreira, S. C., Pastor-Satorras, R., & Starnini, M. (2019). Quantifying echo chamber effects in information spreading over political communication networks. *arXiv preprint arXiv:1901.03688*.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., ... & Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3), 554-559.

Dubois, E., & Blank, G. (2018). The echo chamber is overstated: the moderating effect of political interest and diverse media. *Information, Communication & Society*, 21(5), 729-745.

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

Feldman, M., Friedler, S. A., Moeller, J., Scheidegger, C., & Venkatasubramanian, S. (2015, August). Certifying and removing disparate impact. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 259-268). ACM.

Ferrara, Emilio (2017). "Disinformation and social bot operations in the run up to the 2017 French presidential election." *First Monday* 22(8).

Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter Bubbles, Echo Chambers, and Online News Consumption. *Public opinion*, 80(S1), 298-320.

Garrett, R. K. (2009). Echo chambers online?: Politically motivated selective exposure among internet news users. *Journal of Computer-Mediated Communication*, 14(2):265-285.

Goodman, B., & Flaxman, S. (2016). European Union regulations on algorithmic decision-making and a "right to explanation". *arXiv preprint arXiv:1606.08813*.

Gozdecka, D. A., Ercan, S. A & Kmak, M (2014). From multiculturalism to post-multiculturalism: Trends and paradoxes. *Journal of Sociology*, 50(1), 51-64.

Guan, Z., & Cutrell, E. (2007, April). An eye tracking study of the effect of target rank on web search. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 417-420). ACM.

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*, 201320040.

Latzer, M., Katharina Natascha, H., & Saurwein J. F. (2015). *The economics of algorithmic selection on the Internet*. Working Paper –Media Change & Innovation Division, University of Zurich.

Latzer, M. (2014), "Convergence, Co-evolution and Complexity in European Communications Policy". in K.Donders, C.Pauwels and J.Loisen (eds), *The Palgrave Handbook of European Media Policy*, Houndmills: Palgrave Macmillan, pp.36-53.

Pariser, E. (2011), *The Filter Bubble*, London: Penguin Books

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. London, England: Harvard University Press.

Prensky, M. (2001). Digital Natives, Digital Immigrants, Part 1. *On The Horizon*, 9, 3-6.

Resnick, M., & Albert, W. (2014). The impact of advertising location and user task on the emergence of banner ad blindness: An eye-tracking study. *International Journal of Human-Computer Interaction*, 30(3), 206-219.

Reuters Institute (2009). *Digital News Report 2019*

RHEINGOLD, HOWARD (1991). *Virtual Reality* ", by , Summit Books/Simon and Schuster.

Steel, E., & Angwin, J. (2010). On the web's cutting edge, anonymity in name only. *The Wall Street Journal*, 4.

Tapscott, D. (1998). *Growing Up Digital. The Rise of the Net Generation*. McGraw Hill.

Timmermans, Mike (2019). "The political effects of algorithms: a look at Facebook and Google". <https://www.digitmagazine.com/papers/political-effects-algorithms>.

Wojcieszak, M. (2010). 'Don't talk to me': effects of ideologically homogeneous online groups and politically dissimilar offline ties on extremism. *New Media & Society*, 12(4), 637-655.

Zafar, M. B., Valera, I., Gomez Rodriguez, M., & Gummadi, K. P. (2017, April). Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment. In *Proceedings of the 26th International Conference on World Wide Web* (pp. 1171-1180). International World Wide Web Conferences Steering Committee.

서울대학교 인공지능정책 이니셔티브 안내

서울대학교 인공지능정책 이니셔티브는 인공지능과 관련된 다양한 사회경제적, 법적, 정책적 이슈들을 연구하고 논의하기 위해 시작된 서울대학교 법과경제연구센터의 프로그램입니다. ‘소셜랩(Social Lab)’ 개념을 지향하여, 여러 배경과 관심을 가진 분들 사이의 협업과 지속적인 대화를 추구합니다. 서울대학교 법학전문대학원의 교수와 임용 교수가 함께 이끌고 있습니다.

1. 발간물 안내

서울대학교 인공지능정책 이니셔티브의 주요 발간물은 이슈페이퍼와 워킹페이퍼가 있고, 비정기적으로 발간되는 단행본 및 학술행사 자료집 등이 있습니다. 이슈페이퍼와 워킹페이퍼 등의 자료들은 홈페이지를 통해 다운로드 받으실 수 있습니다.

2. 행사 안내

서울대학교 인공지능정책 이니셔티브의 주요 행사는 이슈페이퍼를 발표하고 논의하는 행사(상반기 및 하반기 각 1회) 그리고 국내외 연구자들을 초빙하여 진행하는 대규모 국제학술대회(연 1회) 등이 있습니다. 그 이외에 비정기적으로 진행하는 행사들도 있습니다.

3. 이슈페이퍼 2019-2

이번 이슈페이퍼는 서울대학교 인공지능정책 이니셔티브의 두 번째 이슈페이퍼로, 2019.11.07. D2 Startup Factory에서 열린 행사에 맞춰 준비되었습니다.