The Case for Establishing a Collective Perspective to Address the Harms of Platform Personalization

Ayelet Gordon-Tapiero,[*] Alexandra Wood,[**] & Katrina Ligett[***]

ABSTRACT

Personalization on digital platforms drives a broad range of harms, including misinformation, manipulation, social polarization, subversion of autonomy, and discrimination. In recent years, policy

[*] Research Fellow, the Federmann Cyber Security Center, Postdoctoral Research Fellow, the Rachel and Selim Benin School of Computer Science and Engineering, the Hebrew University of Jerusalem.

[**] Fellow, Berkman Klein Center for Internet & Society at Harvard University.

[***] Associate Professor, the Rachel and Selim Benin School of Computer Science and Engineering, the Hebrew University of Jerusalem.

The Authors would like to thank Gordon Berlin, Michael Birnhack, Ella Corren, Dafna Dror, Elizabeth Edenberg, Alon Harel, Daniel Ho, Neta Livneh, Kobbi Nissim, Gideon Parchomovsky, Matt Prewitt, Amnon Reichman, Gal Shahaf, Thomas Streinz, Bo Waggoner, and Eyal Zamir for their invaluable insights. In addition, the Authors are grateful for useful comments provided by the participants in a working group on data co-ops hosted by the Simons Institute for the Theory of Computing in Spring 2019; the Weizmann University program on Societal Concerns in Algorithms and Data Analysis; the Tel Aviv University Law Workshop in Information Technology and Law; a seminar held at the Berkman Klein Center for Internet & Society at Harvard University; the Harvard Privacy Tools Project; the Boston University Cyber Security, Law, and Society Alliance; the Colloquium of Microsoft Research New England; the JerusML Show; a series of workshops on data co-ops organized at Georgetown University and the Hebrew University of Jerusalem; Intel Jerusalem Techweek; the Tel Aviv University Workshop on AI, Law and Agency in the Age of Machine Learning; the RadicalXChange Annual Conference; the colloquium of the Max Planck Institute for Software Systems; the DIMACS Workshop on the Co-Development of Computer Science and Law; the Privacy in Machine Learning (priML)/Privacy-Preserving Machine Learning (PPML) Workshop; the 8th Privacy, Cyber and Technology Workshop at Tel Aviv University; and the Cyber Forum at the Haifa Center for Cyber, Law and Policy. This work has been supported by a workshop grant from the Simons Foundation, a gift to the McCourt School of Public Policy and Georgetown University, Simons Foundation Collaboration 733792, and Israel Science Foundation (ISF) grants 1044/16 and 2861/20. Part of Ligett's work was done while the author was visiting the Simons Institute for the Theory of Computing. An abridged version of this Article appeared as a proceedings paper for the 2022 ACM Symposium on Computer Science and Law (CSLAW '22). Ayelet Gordon-Tapiero, Alexandra Wood & Katrina Ligett, *The Case for Establishing a Collective Perspective to Address the Harms of Platform Personalization*, PROC. 2022 ACM SYMP. COMPUT. SCI. & L. (2022).

makers, civil society advocates, and researchers have proposed a wide range of interventions to address these challenges. This Article argues that the emerging toolkit reflects an individualistic view of both personal data and data-driven harms that will likely be inadequate to address growing harms in the global data ecosystem. It maintains that interventions must be grounded in an understanding of the fundamentally collective nature of data, wherein platforms leverage complex patterns of behaviors and characteristics observed across a large population to draw inferences and make predictions about individuals.

Using the lens of the collective nature of data, this Article evaluates various approaches to addressing personalization-driven harms under current consideration. It also frames concrete guidance for future legislation in this space and for meaningful transparency that goes far beyond current transparency proposals. It offers a roadmap for what meaningful transparency must constitute: a collective perspective providing a third party with ongoing insight into the information gathered and observed about individuals and how it correlates with any personalized content they receive across a large, representative population. These insights would enable the third party to understand, identify, quantify, and address cases of personalization-driven harms. This Article discusses how such transparency can be achieved without sacrificing privacy and provides guidelines for legislation to support the development of such transparency.

TABLE OF CONTENTS

I. INTRODUCTION

Platforms' ability to personalize content for each of their users has recently given rise to several controversies, including the Facebook-Cambridge Analytica data scandal;[1] the "emotional contagion" experiment to study the influence of Facebook posts on users' moods;[2] research uncovering leading platforms' discriminatory presentation of job and housing ads on the basis of race, gender, and age;[3] and, most recently, the Wall Street Journal's investigative

1. *See* Nicholas Confessore, *Cambridge Analytica and Facebook: The Scandal and the Fallout So Far*, N.Y. TIMES (Apr. 4, 2018), https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html [https://perma.cc/6A2E-8NAL].

2. *See* Adam D.I. Kramer, Jamie E. Guillory & Jeffrey T. Hancock, *Experimental Evidence of Massive-Scale Emotional Contagion Through Social Networks*, 111 PROC. NAT'L ACAD. SCIS. 8788, 8788 (2014).

3. *See, e.g.*, Basileal Imana, Aleksandra Korolova & John Heidemann, *Auditing for Discrimination in Algorithms Delivering Job Ads*, 2021 PROC. WEB CONF. 3767, 3767–70 (2021) (demonstrating that presentation of ads on Facebook and Google can be skewed by gender); Alexia Fernández Campbell, *Job Ads on Facebook Discriminated Against Women and Older Workers, EEOC Says*, VOX (Sept. 25, 2019, 2:20 PM) https://www.vox.com/identities/2019/9/25/20883446/facebook-job-ads-discrimination [https://perma.cc/S46D-55CA] (finding that Facebook presented ads in a way that discriminated against women and older users). *See generally* Anja Lambrecht & Catherine Tucker, *Apparent Algorithmic Discrimination and Real-Time Algorithmic Learning in Digital Search Advertising* (Apr. 12, 2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3570076 [https://perma.cc/M45L-DEXS] (finding that GoogleAds presented users who

reporting on The Facebook Files.[4] In response to the testimony of Facebook whistleblower Frances Haugen before Congress in October 2021, a bipartisan group of US lawmakers signified that "Facebook and Big Tech are facing a Big Tobacco moment,"[5] joining a chorus of voices in the United States and around the world calling for stronger regulation of platforms.[6]

Economic, social, political, and cultural activities are increasingly mediated by platforms,[7] representing a shift "from industrial to information capitalism."[8] As the process of digitization has

had previously searched for Black names with ads for disadvantageous jobs compared to users who had previously searched for White names).

For an introduction to platforms' approaches to personalization, see Kimberly Rhum, *Information Fiduciaries and Political Microtargeting: A Legal Framework for Regulating Political Advertising on Digital Platforms*, 115 NW. U. L. REV. 1829, 1831 (2021) (detailing how a variety of platforms offer their users personalized experiences).

4. *See The Facebook Files*, WALL ST. J., https://www.wsj.com/articles/the-facebook-files-11631713039 [https://perma.cc/C4SE-9QNR] (last visited Apr. 1, 2023).

5. Cecilia Kang, *Lawmakers See Path to Rein in Tech, but It Isn't Smooth*, N.Y. TIMES, https://www.nytimes.com/2021/10/09/technology/facebook-big-tobacco-regulation.html [https://perma.cc/8QNM-3XYV] (Oct. 12, 2021).

6. *See* Adam Satariano, *Facebook Hearing Strengthens Calls for Regulation in Europe*, N.Y. TIMES (Oct. 6, 2021), https://www.nytimes.com/2021/10/06/technology/facebook-european-union-regulation.html [https://perma.cc/GFE4-BY3C].

7. Several definitions of the term "platform" have been offered in the literature. For example, Lina Kahn emphasizes platforms' role as intermediaries of economic activities, likening them to bank holding companies. Lina M. Kahn, *Amazon's Antitrust Paradox*, 126 YALE L.J. 710, 795 (2017). Other definitions focus on the fact that platforms do not only mediate economic transactions, but "in a broader social sense of comprising the basic infrastructure of modern society." K. Sabeel Rahman, *The New Utilities: Private Power, Social Infrastructure, and the Revival of the Public Utility Concept*, 39 CARDOZO L. REV. 1621, 1641 (2018). Perhaps one of the most important areas in which platforms have had a transformative role is that of data production and collection. Indeed, Cohen recognizes that platforms' greatest interest lies in "data extracted from people as they invest, work, operate businesses, socialize, and engage in innumerable other activities." JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM 38 (2019); *see also* Priscilla M. Regan, *A Design for Public Trustee and Privacy Protection Regulation*, 44 SETON HALL LEGIS. J. 487, 496 (2020) ("It is widely recognized that the business models of large internet companies rely upon the collection, use, and analysis of personal information.").

In this Article, the Authors build on Cohen's recognition of the central role of data in the business models and activities of platforms, using the term to refer to entities that collect, store, process, analyze, or act upon data pertaining to individuals (for example, in the provision of content, services, recommendations, or ads), and whose presence is primarily in the digital realm. The Authors use the term *users* to denote individuals who use the services of the platforms. The term *individuals* describes people (who have not necessarily signed up to use a certain platform or agreed to its terms of service). Finally, the term *data ecosystem* refers to platforms, individuals, and any other entities participating in exchanging, transacting, and acting on data pertaining to individuals. *See* Salomé Viljoen, *A Relational Theory of Data Governance*, 131 YALE L.J. 573, 612 (2021) (noting "[t]he combination of relational and aggregate effects from data production drives companies to collect as much data as possible from data subjects").

8. COHEN, *supra* note 7, at 5.

enabled increased datafication—the ability to render into data many aspects of the world that have never been quantified before[9]—platforms' power and control over modern marketplaces for social interactions have grown.[10] To manage and leverage the growing amount of electronic data they possess, platforms have developed and implemented artificial intelligence and machine learning algorithms, which, in turn, demand large volumes of data as inputs.[11] Common across platforms' various business models is a strong incentive to collect and analyze massive quantities of data about individuals—and to use this information to present individuals with personalized content.[12]

To achieve these ends, platforms harness their ability to capture, analyze, and act upon data measuring the behavior of large groups; detect patterns of behavior and previously unanticipated clusters of users; make predictions about how individuals and groups of individuals will respond to personalized content; infer deeply personal attributes that an individual has not expressly disclosed; and act upon these predictions and inferences.[13] Such personalization—i.e., platforms' ability to show each user content specifically chosen for them—can benefit users, but it also contributes to a broad range of

9. Kenneth Cukier & Viktor Mayer-Schoenberger, *The Rise of Big Data: How It's Changing the Way We Think About the World*, 92 FOREIGN AFFS. 28, 29 (2013).

10. *See* SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER 184–85 (2019); COHEN, *supra* note 7, at 28 (detailing an example of how platforms became involved in the field of consumer finance).

11. *See* Josep Lluís Berral-Garcia, *A Quick View on Current Techniques and Machine Learning Algorithms for Big Data Analytics*, INT'L CONF. TRANSPARENT OPTICAL NETWORKS, 2016 (explaining that in order to manage big data the development of machine learning algorithms is necessary); Jafar Alzubi, Anand Nayyar & Akshi Kumar, *Machine Learning from Theory to Algorithms: An Overview*, 1142 J. PHYSICS: CONF. SERIES 1, 13 (2018) (observing that "machine learning algorithms require large volumes of data to be accurate and efficient").

12. *See* Julie E. Cohen, *Law for the Platform Economy*, 51 U.C. DAVIS L. REV. 133, 140, 160 (2017) (discussing how ongoing collection of large amounts of data is an important part of platforms' market power); Shoshana Zuboff, *Big Other: Surveillance Capitalism and the Prospects of an Information Civilization*, 30 J. INFO. TECH. 75, 75–77 (2015) (explaining that the ability to collect large amounts of data is a significant part of *surveillance capitalism*); Hal R. Varian, *Computer Mediated Transactions*, 100 AM. ECON. REV. 1, 2 (2010) (identifying relatively early on in the development of the internet that facilitating personalization was one of the substantial impacts of computer mediated transactions); Brent Mittelstadt, *Auditing for Transparency in Content Personalization Systems*, 10 INT'L J. COMM. 4991, 4991 (2016) ("Content personalization systems display information tailored to individual users, often based on perceived preferences or past behaviors.").

13. *See* Eduard Fosch-Villaronga, Adam Poulsen, Roger Andre Søraa & Bart Custers, *A Little Bird Told Me Your Gender: Gender Inferences in Social Media*, 58 INFO. PROCESSING & MGMT. 1, 1 (2021) (demonstrating that platforms can infer an individual's gender even when the individual has not provided it).

data-driven harms, including misinformation,[14] manipulation,[15] social polarization,[16] subversion of autonomy,[17] and discrimination.[18]

As a consequence, much of the early optimism that the internet would evolve to be a "liberating and democratic social force"[19] has faded, and in recent years policy makers, civil society advocates, and researchers around the world have increasingly turned their attention to addressing data-driven harms in the modern data ecosystem.[20]

---

14. *See* Ashley Smith-Roberts, *Facebook, Fake News, and the First Amendment*, 95 DENV. L. REV. F. 118, 125 (2018).

15. *See* Zeynep Tufekci, *Algorithmic Harms Beyond Facebook and Google: Emergent Challenges of Computational Agency*, 13 COLO. TECH. L.J. 203, 204 (2015).

16. *See* Christopher A. Bail, Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M.B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout & Alexander Volfovsky, *Exposure to Opposing Views on Social Media Can Increase Political Polarization*, 115 PROC. NAT'L ACAD. SCIS. 9216, 9216 (2018).

17. *See* Daniel Susser, Beate Roessler & Helen Nissenbaum, *Technology, Autonomy, and Manipulation*, 8 INTERNET POL'Y REV. 1, 3 (2019).

18. For example, US antidiscrimination laws prohibit discrimination in housing and employment advertising. *See* 42 U.S.C. §§ 804, 2000e-3(b). Section 804 of the Fair Housing Act served as the basis for the US Department of Housing and Urban Development's charge of discrimination against Facebook in 2019, alleging discrimination in the presentation of ads for housing on the platform. Charge of Discrimination, U.S. Dep't Hous. & Urb. Dev. v. Facebook, Inc., FHEO No. 01-18-0323-8 (2019). Section 2000e of Title VII of the Civil Rights Act of 1964 served as the basis for a decision by the US Equal Employment Opportunity Commission, finding that seven employers had violated federal law when advertising jobs on Facebook in a way that excluded women and older workers from seeing the ads. *In Historic Decision on Digital Bias, EEOC Finds Employers Violated Federal Law When They Excluded Women and Older Workers from Facebook Job Ads*, ACLU (Sept. 25, 2019, 11:00 AM), https://www.aclu.org/press-releases/historic-decision-digital-bias-eeoc-finds-employers-violated-federal-law-when-they [https://perma.cc/YL3G-M38K] (reporting on the decision); *Compiled U.S. Equal Emp. Opportunity Comm'n Letters of Determination*, OUTTEN & GOLDEN, LLP (July 5, 2019), https://www.onlineagediscrimination.com/sites/default/files/documents/eeoc-determinations.pdf [https://perma.cc/4LNE-F3N5]. Researchers have also demonstrated that numerous platforms present housing and employment ads in a discriminatory manner. *See, e.g.*, Muhammad Ali, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove & Aaron Rieke, *Discrimination Through Optimization: How Facebook's Ad Delivery Can Lead to Skewed Outcomes*, 3 PROC. ASS'N COMPUTING MACH. ON HUM.-COMPUT. INTERACTION, art. 199, at 1, 1–2 (2019) (observing significant skews in the presentation of ads for housing and employment along gender and racial lines); Imana et al., *supra* note 3 (demonstrating that presentation of ads on Facebook and Google can be skewed by gender).

19. ZUBOFF, *supra* note 10, at 74.

20. Explosive growth in the global data ecosystem has led to the recent adoption of a number of data protection and consumer privacy laws. *See, e.g.*, Regulation 2016/679, of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC, 2016 O.J. (L 119) [hereinafter GDPR]; California Consumer Privacy Act, CAL. CIV. CODE §§ 1798.100 –1798.199.100; Virginia Consumer Data Protection Act, VA. CODE ANN. §§ 59.1-575 to 59.1-585; Colorado Privacy Act, COLO. REV. STAT. §§ 6-1-1301 to 6-1-1313.

In the United States, several legislative proposals have targeted the harms stemming from platform personalization. *See, e.g.*, Honest Ads Act, H.R. 4077, 115th Cong. (2017); Deceptive

This Article argues that current regulatory and technological proposals reflect an individualistic view of personal data and data-driven harms, and that such a framing will likely fail to adequately address the harms stemming from platform personalization. Instead, interventions must be grounded in an understanding of the fundamentally collective nature of data[21]—that is, an understanding that recognizes that the personalized content a user receives is strongly driven by rich data gathered about other users around the globe.[22] Many platform-driven challenges such as social polarization and discrimination do not arise with respect to one isolated individual; such harms, as well as the ability to define and detect them, inherently and inextricably exist within a broader social context.[23] Furthermore, the only parties that may currently possess a full picture of this personalization landscape are the platforms themselves.[24] Proposals that seek to enhance individual control are ineffective

---

Experiences to Online Users Reduction (DETOUR) Act, S. 1084, 116th Cong. (2019) ("To prohibit the usage of exploitative and deceptive practices by large online operators and to promote consumer welfare in the use of behavioral research by such providers"); Social Media Addiction Reduction Technology (SMART) Act, S. 2314, 116th Cong. (2019); Filter Bubble Transparency Act, S. 2763, 116th Cong. (2019); Children and Media Research Advancement (CAMRA) Act, S. 971, 117th Cong. (2021); Protecting Americans from Dangerous Algorithms Act, H.R. 2154, 117th Cong. (2021); Justice Against Malicious Algorithms Act of 2021, H.R. 5596, 117th Cong. (2021); Health Misinformation Act of 2021, S. 2448, 117th Cong. (2021); Social Media Disclosure and Transparency (DATA) Act, H.R. 3451, 117th Cong. (2021); Platform Accountability and Transparency Act (PATA), S. 5339, 117th Cong. (2021); Algorithmic Justice and Online Platform Transparency Act, S. 1896, 117th Cong. (2021).

In Europe, several initiatives to address the challenges of personalization have been introduced. *See, e.g.*, *Proposal for a Regulation of the Parliament and of the Council on a Single Market for Digital Services and amending Directive 2000/31/EC*, COM (2020) 825 final (Dec. 15, 2020) [hereinafter *Proposal for DSA*] (aiming to "establish a powerful transparency and a clear accountability framework for online platforms"); EU, CODE OF PRACTICE ON DISINFORMATION (2018), https://digital-strategy.ec.europa.eu/en/library/2018-code-practice-disinformation [https://perma.cc/6S8D-CJCK] (adopting self-regulatory standards to combat disinformation).

21. *See* Martin Tisne, *The Data Delusion: Protecting Individual Data Isn't Enough When the Harm Is Collective*, LUMINATE, July 2020, at 1, 1, 2 ("The collective nature of big data means people are more impacted by other people's data than by data about them. Like climate change, the threat is societal and personal."); Regan, *supra* note 7, at 501 ("There is no question that regulators are struggling and not doing very well in this struggle.").

22. *See* Sandra Wachter & Brent Mittelstadt, *A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI*, 2019 COLUM. BUS. L. REV. 494, 613–14 (2019).

23. *See* Simon A. Levin, Helen V. Milner & Charles Perrings, *The Dynamics of Political Polarization*, 118 PROC. NAT'L ACAD. SCIS. 1, 1 (2021) (acknowledging that phenomena such as polarization "are inherently systems-level phenomena, involving interactions among multiple component parts and the emergence of broader scale features").

24. *See* Yochai Benkler, *Degrees of Freedom, Dimensions of Power*, 145 DAEDALUS 18, 23 (2016) ("Big data collection and processing, combined with ubiquitous sensing and connectivity, create extremely powerful insights on mass populations available to relatively few entities.").

because platforms can infer personal attributes an individual has not expressly disclosed based on the data provided by other users, thereby nullifying an individual's decision to withhold their data from platforms.[25]

Additionally, many approaches to transparency fail to provide sufficient visibility into personalization-based harms. Carefully constructed experiments have demonstrated that platforms induce discriminatory personalization of certain content, such as ads for employment.[26] Such experiments, however, are inherently limited in scope and can identify only instances of the particular harm researchers sought to measure.[27] Adequate transparency, furthermore, requires far more than disclosing ad-targeting criteria or ad-funding details, as in the Honest Ads Act;[28] creating databases of ads divorced from the personal information of those who received them, as in the Digital Services Act (DSA);[29] or focusing primarily on ads, as in the Social Media DATA Act.[30] These approaches do not provide the kind of transparency that third parties must have in order to investigate collective-level data-driven harms.

Without meaningful, effective transparency, society lacks the essential tools to properly understand the role that personalization plays in generating and amplifying various harms. At present, there is uncertainty regarding even the most basic questions, such as whether

---

25. *See* Viljoen, *supra* note 7, at 578, 607.

26. *See, e.g.*, Muhammad Ali, Piotr Sapiezynski, Aleksandra Korolova, Alan Mislove & Aaron Rieke, *Ad Delivery Algorithms: The Hidden Arbiters of Political Messaging*, 14 PROC. ASS'N COMPUTING MACH. INT'L CONF. ON WEB SEARCH & DATA MINING, Mar. 8–12, 2019, at 13, 14 (finding that "Facebook preferentially shows users political ads whose contents Facebook predicts are aligned with their political views."); Imana et al., *supra* note 3 (demonstrating that presentation of ads on Facebook and Google can be skewed by gender); Amit Datta, Michael Carl Tschantz & Anupam Datta, *Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination*, PROCS. ON PRIV. ENHANCING TECHS., 2015, at 92, 93 (demonstrating that changing one's self-reported gender influences the job ads one sees).

27. Basileal Imana, Aleksandra Korolova & John Heidemann, *Having Your Privacy Cake and Eating it Too: Platform-Supported Auditing of Social Media Algorithms for Public Interest*, 7 PROC. ASS'N COMPUTING MACH. HUM.-COMPUT. INTERACT., art. 134, at 5 (2023) (arguing that existing research methods "are difficult to generalize," "have high cost," and "are reaching hard limits in terms of what they can reliably and provably learn about the role of platforms' algorithms" without better access to platform data).

28. Honest Ads Act, H.R. 4077, 115th Cong. (2017).

29. *Proposal for a Regulation of the Parliament and of the Council on a Single Market for Digital Services and amending Directive 2000/31/EC*, COM (2020) 825 final (Dec. 15, 2020).

30. Social Media Disclosure and Transparency (DATA) Act, H.R. 3451, 117th Cong. (2021).

personalization is contributing to polarization or defusing it.[31] This Article offers a roadmap for what meaningful transparency must constitute: ongoing visibility into the information that platforms gather and observe about individuals and how that information correlates with the personalized content these users receive—across a large, representative population. Then, it discusses how a third party can achieve meaningful transparency without sacrificing privacy, and, finally, it provides guidelines for future legislation to support the development of such transparency.

The Article proceeds as follows: Part II describes the structure of the data ecosystem, explains the financial incentives driving platforms' extensive data collection, and introduces novel terminology that captures the different flows of content between users and platforms. It also highlights the various ways in which data is collective and demonstrates how information about one person can allow a platform to learn about another. Part III uses this lens of the collective nature of data to help analyze various regulatory and technical approaches that have been designed to address personalization-driven harms. Part IV presents design principles that can facilitate effective intervention. It advocates for meaningful transparency—namely, by generating a collective perspective that would allow a third party to view the data of large groups of users—and offers ways regulation could facilitate the creation of such a perspective.

II. THE STRUCTURE OF THE DATA ECOSYSTEM

This Part provides an overview of the structure of the data ecosystem and the incentives that drive platforms' activities therein.[32] In particular, platforms' business models have created powerful incentives—and capabilities—for them to design their services, content, and interfaces to increase opportunities for impactful personalized

31. *Compare* Levi Boxell, Matthew Gentzkow & Jesse M. Shapiro, *Is the Internet Causing Political Polarization? Evidence from Demographics* 3 (Nat'l Bureau of Econ. Rsch., Working Paper No. 23258, 2017) (demonstrating that the age group exhibiting the highest level of polarization was the group aged seventy-five and older, i.e., the age bracket with the least exposure to the internet and social media), *with* Bail et al., *supra* note 16 (describing concerns that social media exacerbates polarization).

32. Platforms' business models vary based on numerous criteria, such as whether individuals pay to access the service, to what extent advertising is a significant part of the platform's revenue, what type of data the platform gathers, which parties it shares data with, what information services the platform provides, and how personalized the offered services are. In this Article, the Authors refer to all platforms as defined in *supra* note 7, regardless of their business model.

advertising, thereby boosting profitable revenue streams.[33] This Part introduces terminology to describe the flows of content between users and platforms and explains how these flows create a feedback loop: data collected by platforms serves as a basis for personalizing content for users, whose activity then generates more data for platforms to collect. In addition, this Part demonstrates why it is critical to recognize the collective nature of data when considering the suitability of interventions to address personalization-driven harms.

A. The Data Ecosystem's Outgoing and Incoming Vectors

In the data ecosystem, information flows between users and platforms in both directions. In one direction, data flows from users to platforms along what this Article terms the *outgoing vector*. Along the outgoing vector, platforms collect vast quantities of data about users and their activities,[34] including interactions each user has directly with the platform (e.g., noting groups users belong to and pages and other content they "like"), interactions among users (e.g., commenting on a friend's post, retweeting, and sharing media), and users' online activity outside the platform (e.g., identifying other web sites they have visited).[35] In some cases, platforms also collect information about users'

33. COHEN, *supra* note 7, at 41. One of the byproducts of platforms' ability to personalize ads and other content—indeed, of informational capitalism as a broad phenomenon—is a deepening of social inequality. Platforms have amassed power while society has seen the emergence of a "seemingly permanent economic underclass." COHEN, *supra* note 7, at 180; *see also* Tim Berners-Lee, *One Small Step for the Web. . .*, MEDIUM (Sept. 29, 2018), https://medium.com/@timberners_lee/one-small-step-for-the-web-87f92217d085 [https://perma.cc/XKH9-7YTC] (observing that "for all the good we've achieved, the web has evolved into an engine of inequity and division; swayed by powerful forces who use it for their own agendas").

34. *See* Datta et al., *supra* note 26, at 92 ("Colossal amounts of collected data are used, sold, and resold for serving targeted content, notably advertisements, on websites."); Jack M. Balkin, *Information Fiduciaries and the First Amendment*, 49 U.C. DAVIS L. REV. 1183, 1185 (2016) (acknowledging the widespread collection of personal data); Shira Ovide, *What's Behind the Apple-Facebook Feud?*, N.Y. TIMES, https://www.nytimes.com/2021/04/26/technology/apple-facebook-feud.html [https://perma.cc/CDC7-USS2] (June 11, 2021) ("Currently, Facebook and companies like it track the ways people use their phones, picking up bits of information such as how often they open their yoga app and what they buy at Target. Facebook then uses that information to help companies target their ads."); Till Speicher, Muhammad Ali, Giridhari Venkatadri, Filipe Nunes Ribeiro, George Arvanitakis, Fabrício Benevenuto, Krishna P. Gummadi, Patrick Loiseau & Alan Mislove, *Potential for Discrimination in Online Targeted Advertising*, 81 PROC. MACH. LEARNING RSCH. 1, 3 (2018) ("Facebook gathers and infers several hundreds of attributes for all of its users.").

35. When a user signs into a third-party service with their Facebook account, Facebook is made aware of their activity, even though it takes place outside the Facebook platform. Additionally, when a Facebook user visits a site with the 'like' button embedded in it, Facebook collects information about that visit regardless of whether the user clicked the 'like' button. *See*

offline activity that is provided by their devices, such as their location data,[36] or by third parties, such as information about users' shopping habits, credit scores, public records, voter registration data, and more.[37] Platforms collect and analyze this data in order to draw a detailed profile about each user and, at times, to make the data available to third parties.[38] This Article views privacy as predominantly an

---

Jonathan R. Mayer & John C. Mitchell, *Third-Party Web Tracking: Policy and Technology*, INST. ELEC. & ELEC. ENG'RS SYMP. ON SEC. & PRIV., 2012, at 413, 419 (2012); Dina Srinivasan, *The Antitrust Case Against Facebook: A Monopolist's Journey Towards Pervasive Surveillance in Spite of Consumers' Preference for Privacy*, 16 BERKELEY BUS. L.J. 39, 41, 62 (2019). Additionally, Google keeps track of news articles that its users read, even if they are not accessed via a Google search. *See* Brian X. Chen, *I Downloaded the Information That Facebook Has on Me. Yikes*, N.Y. TIMES (Apr. 11, 2018), https://www.nytimes.com/2018/04/11/technology/personaltech/i-downloaded-the-information-that-facebook-has-on-me-yikes.html [https://perma.cc/S8SR-Y7Y5] ("Google kept a history of many news articles I had read. . . . I didn't click on ads for either of these stories, but the search giant logged them because the sites had loaded ads served by Google.").

36. *See* Chen, *supra* note 35 ("On some days, [Facebook] even logged my locations, like when I was at a hospital two years ago or when I visited Tokyo last year."); Irfan Faizullabhoy & Aleksandra Korolova, Facebook's Advertising Platform: New Attack Vectors and the Need for Interventions 1 (2018) (unpublished manuscript) (on file with Cornell University) ("Social media websites such as Facebook, Google, and Pinterest record and learn from user behavior . . . such as location."); John Herrman, *Google Knows Where You've Been, But Does It Know Who You Are?*, N.Y. TIMES (Sept. 12, 2018), https://www.nytimes.com/2018/09/12/magazine/google-maps-location-data-privacy.html [https://perma.cc/89XZ-QW7N] ("Some Google apps automatically store time-stamped location data without asking.") (internal quotation marks omitted).

37. *See* Giridhari Venkatadri, Piotr Sapiezynski, Elissa M. Redmiles, Alan Mislove, Oana Goga, Michelle L. Mazurek & Krishna P. Gummadi, *Auditing Offline Data Brokers via Facebook's Advertising Platform*, WORLD WIDE WEB CONF., May 2019, at 1920, 1920 ("Recently, data brokers and online services have begun partnering together, allowing for the data collected about users online to be linked against data collected offline. This enables online services to provide advertisers with targeting features that concern users' offline information."); Pauline T. Kim & Sharion Scott, *Discrimination in Online Employment Recruiting*, 63 ST. LOUIS U. L.J. 93, 97 (2018) ("Facebook also purchases information from data brokers to learn about users' offline behavior, including income and spending habits."); Kalev Leetaru, *The Data Brokers So Powerful Even Facebook Bought Their Data - But They Got Me Wildly Wrong*, FORBES (Apr. 5, 2018, 4:08 PM), https://www.forbes.com/sites/kalevleetaru/2018/04/05/the-data-brokers-so-powerful-even-facebook-bought-their-data-but-they-got-me-wildly-wrong [https://perma.cc/WU68-3574] ("In essence, Facebook recognized that many of the most useful data points on our daily lives come not from the utopian image of perfection we project on Facebook, but from the actual mundane reality of our daily lives, from what we purchase at the grocery store to where we live to our financial status."); Kashmir Hill, *Facebook Is Tracking What Users Buy In Stores To See Whether Its Ads Work*, FORBES (Sept. 26, 2012, 5:47 PM), https://www.forbes.com/sites/kashmirhill/2012/09/26/facebook-is-tracking-what-users-buy-in-stores-to-see-whether-its-ads-work [https://perma.cc/53QB-8F6T].

38. *See* Gabriel J.X. Dance, Nicholas Confessore & Michael LaForgia, *Facebook Gave Device Makers Deep Access to Data on Users and Friends,* N.Y. TIMES (June 3, 2018), https://www.nytimes.com/interactive/2018/06/03/technology/facebook-device-partners-users-friends-data.html [https://perma.cc/5HVX-9W8Y] ("Facebook has reached data-sharing partnerships with at least 60 device makers—including Apple, Amazon, BlackBerry, Microsoft and Samsung—over the last decade, starting before Facebook apps were widely available on smartphones, company officials said."); *see also* Venkatadri et al., *supra* note 37; Nizan Geslevich

outgoing-vector concern, related to mitigating the platform-mediated flow of data pertaining to individuals.[39]

In the other direction, information flows from platforms to users along what this Article calls the *incoming vector*. This encompasses all content that platforms present or suggest to users based on the detailed profile that the platform has created about them.[40] For example, platforms send notifications, resurface old posts as memories, compile photos and other user-generated content into custom videos, and present posts or videos to users in their feed (and decide on the order in which they are presented). They also suggest groups to join as well as other content users may be interested in (e.g., news articles, physical gatherings, and other users to connect with). Platforms employ personalization along the incoming vector with the goal of increasing user engagement and time spent on the platform. The more time a user spends interacting with the platform, the more data the platform collects, allowing it to present personalized content that is increasingly tailored to the user's inferred interests.

Packin, *Show Me the (Data About the) Money!*, 2020 UTAH L. REV. 1277, 1310 (2020) ("FinTech apps collect more data than needed, save it in an unsafe way, and sell it to third-parties."). Whereas platforms derive enormous profits from users' data, users do not enjoy a portion of these financial benefits. Scholars, activists, and technologists have proposed changes in data governance to overcome this imbalance of power online; two central suggestions include treating data as property and providing "fundamental-rights protections to data as an extension of personal selfhood." *See* Viljoen*, supra* note 7, at 617.

39. While the Authors observe that approaches to privacy in practice tend to focus primarily on addressing outgoing-vector concerns, the Authors recognize that some dimensions of privacy and data protection, such as the principles of purpose limitation and data minimization, among others, are also relevant to incoming-vector concerns. *See* discussion *infra* Section III.A.

40. This information flow also includes the order in which the newsfeed or timeline is presented, and content such as compiling photos and other content into a friendship anniversary movie, suggestions to join groups, and more. *See* Sandra Wachter, *Affinity Profiling and Discrimination by Association in Online Behavioral Advertising*, 35 BERKELEY TECH. L.J. 367, 369 (2020) (describing platforms' ability to personalize content for users based on platforms' knowledge of users' personal attributes).

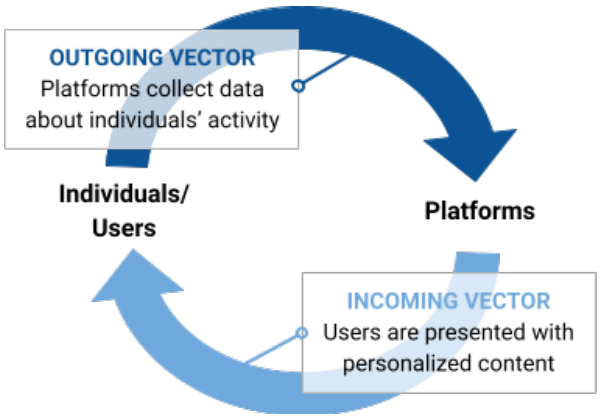

Figure 1. Schematic of the flows of information via the outgoing and incoming vectors.[41]

As illustrated in Figure 1, the outgoing and incoming vectors create a feedback loop: data collected along the outgoing vector is analyzed as a basis for content personalization along the incoming vector, and individuals' interactions with personalized content presented to them along the incoming vector then generate more information along the outgoing vector for the platforms to collect and analyze.

B. The Collective Nature of Data

This Section describes the fundamentally collective nature of data within the data ecosystem, whereby data about one individual can enable platforms to learn about another individual, and patterns of data detected across groups of users also provide insight into the behavior or characteristics of others. This Section argues that recognition of the collective nature of data should inform any intervention to address the harms stemming from incoming-vector personalization.[42]

Each platform user is associated with an extensive record of behavior, such as searches conducted, links clicked, posts liked,

41. Ayelet Gordon-Tapiero, Alexandra Wood & Katrina Ligett, *The Case for Establishing a Collective Perspective to Address the Harms of Platform Personalization*, 1 PROC. ASS'N COMPUTING MACH. SYMP. ON COMP. SCI. & L. 119 (2022).

42. See discussion *infra* in this Section on the different ways in which the information about one user can teach a platform about another user. *See also* Lars Backstrom, Cynthia Dwork & Jon Kleinberg, *Wherefore Art Thou R3579X?: Anonymized Social Networks, Hidden Patterns, and Structural Steganography*, 16 PROC. ASS'N COMPUTING MACH. INT'L CONF. ON WORLD WIDE WEB, May 8–12, 2007, at 181, 181–82 (describing a family of attacks that can enable an adversary to learn of connections between specific users in a network).

messages sent, photos posted, social connections formed, and more.[43] Although there is a tendency to think of such data as belonging to a single user,[44] the reality is much blurrier,[45] as a message sent pertains to both the sender and to the recipient,[46] and a photo in which a friend is tagged pertains both to the poster and to the subject.[47] Such associations can also reveal information that carries consequences extending beyond the context of social media and digital platforms. For example, a Google search for information about a rare genetic disease may have implications not only for the searcher, but also for the searcher's genetic relatives.[48]

In other words, platforms analyze user data not to recognize each individual's uniqueness but to examine how individuals fit into patterns, clusters, and trends.[49] Solon Barocas and Karen Levy call these relationships *privacy dependencies* and present three categories that describe how one can infer the personal attributes of a user based on their social, physical, or electronic ties with others[50]: tie-based dependencies, similarity-based dependencies, and difference-based dependencies.[51]

---

43. *See* Kim & Scott, *supra* note 37 (observing that "Facebook systematically collects large amounts of data about users' activities on the site, such as who their friends are, when they 'like' something, and what links they click").

44. *See* RadicalXChange Foundation, The Data Freedom Act 1 (2020) (draft proposal), https://www.radicalxchange.org/media/papers/data-freedom-act.pdf [https://perma.cc/2AA8-VU57] [hereinafter The Data Freedom Act].

45. *See* Solon Barocas & Karen Levy, *Privacy Dependencies*, 95 WASH. L. REV. 555, 556 (2020) (arguing that "it can be practically difficult to disentangle whether the information 'belongs' to Alice or to Bob and which of them ought to have control over disclosure decisions"); Viljoen, *supra* note 7, at 580 ("[D]ata production in the digital economy is fundamentally relational.").

46. *See* RadicalXChange Foundation, *supra* note 44 ("Data about people is always the output of a network of social activity. Even apparently 'individual' data, such as a particular consumer's shopping habits or travel itinerary, is a product of the social world in which that person lives. . . . Text or email conversations, group photos, calendar entries for meetings, and many other records of social life, record many peoples' activities—not only those of the person who chooses to divulge or exploit the records.").

47. *See* Gergely Biczok & Pern Hui Chia, *Interdependent Privacy: Let Me Share Your Data*, INT'L CONF. FIN. CRYPTOGRAPHY & DATA SEC., 2013, at 335, 338 (describing one user tagging another in a photo as an example of the interdependent nature of data online).

48. *See* Sylvie Delacroix & Neil D. Lawrence, *Bottom-Up Data Trusts: Disturbing the 'One Size Fits All' Approach to Data Governance*, 9 INT'L DATA PRIV. L. 236, 249 (2019) ("Genetic data presents particular challenges because our genome encodes not only information about ourselves but our relatives too: sensitive information can leak through other individuals sharing their genomic data.").

49. *See* RadicalXChange Foundation, *supra* note 44, at 2 (describing the intertwined nature of seemingly personal data); Viljoen, *supra* note 7, at 578, 607.

50. *See* Barocas & Levy, *supra* note 45, at 559.

51. *See id.*

The most intuitive way that one person's information can provide details about another is if the second user is captured in the first user's data incidentally based on their social, physical, or electronic ties, in what Barocas and Levy term *tie-based dependencies*.[52] When Alice uploads a photo from a party she attended, the platform learns about her friend Bob who appears in the photo. Similarly, if Alice uses a virtual assistant or a video-integrated doorbell, the platform may capture information about Bob through his physical interactions with Alice without his knowledge. The platform may also directly prompt Alice to provide information about Bob; for example, when downloading Facebook's Messenger app, users (sometimes unwittingly) give Facebook permission to collect their entire contact list.[53] If enough of Bob's friends join the service, Facebook will be able to construct a web of Bob's social ties even though he himself has provided no information to the platform and may even prefer to avoid the platform altogether.

In some cases, the disclosure of information by one individual enables a platform to indirectly learn something about another, whether because the disclosure highlights a similarity between the two users (in a *similarity-based dependency*) or because it shines a light on the way that the one user differs from the other (in a *difference-based dependency*).[54] By analyzing the behavior of an individual and comparing it to patterns of behavior common to many users, platforms are able to make predictions about individual users and infer a broad range of personal attributes that users have not expressly provided.[55] For example, when Alice conducts her shopping on a platform, the platform gains knowledge of her personal attributes as well as her shopping habits. If another user with attributes similar to Alice's were to start shopping on the platform, the platform may offer her some of the same products that Alice purchased. Similarly, if a new user

52. *See id*.

53. *See* Chen, *supra* note 35.

54. *See* Barocas & Levy, *supra* note 45, at 559; Alessandro Mantelero, *Personal Data for Decisional Purposes in the Age of Analytics: From an Individual to a Collective Dimension of Data Protection*, 32 COMPUT. L. & SEC. REV. 238, 239 (2016) (acknowledging the collective dimension of data, in particular in the context of privacy and data protection); Julie E. Cohen, *The Biopolitical Public Domain: The Legal Construction of the Surveillance Economy*, 31 PHIL. TECH. 213, 225 (2018) (acknowledging the importance of pattern detection in platforms' ability to make predictions about their users).

55. *See* Wachter & Mittelstadt, *supra* note 22, at 506–07 (describing how platforms can infer data about individuals even if they did not provide it); Solon Barocas & Helen Nissenbaum, *Big Data's End Run around Anonymity and Consent*, *in* PRIVACY, BIG DATA, AND THE PUBLIC GOOD: FRAMEWORKS FOR ENGAGEMENT 44, 55 (Julia Lane, Victoria Stodden, Stefan Bender, & Helen Nissenbaum eds., 2014) (observing that "insights drawn from big data can furnish additional facts about an individual (in excess of those that reside in the database) without any knowledge of their specific identity or any identifying information").

demonstrates similar shopping patterns to Alice, the platform may infer that she shares an economically relevant personal attribute with Alice.

Platforms are often able to infer a broad range of potentially sensitive personal attributes such as race, sexual orientation, income, political affiliation, and opinions from users' behavior, interests, and social connections.[56] For example, the detection of patterns across groups serves as the basis for gender classification systems that platforms employ.[57] These systems analyze user data such as pictures, videos, likes, and language patterns, drawing insights from patterns among users who provided their gender in order to infer the gender of users who did not. Users who did not disclose their gender to the platform but whom the platform classified as belonging to a certain gender may feel that their privacy, dignity, and autonomy have been violated, and, in some communities, such inferences may even put individuals at risk of harm, including discrimination and oppression.[58] Further exacerbating autonomy concerns, due to platforms' ability to infer user attributes that the user has not expressly disclosed, an individual cannot prevent a platform from learning about her by refusing to disclose her data. Effectively, this means that no single individual can decide to withhold her data from platforms.

As Salomé Viljoen argues, relationships among users who belong to a group enable platforms to use data about one user to infer characteristics of another member of the same group.[59] Furthermore, in order to learn something about a group of people it is enough that a small minority has provided their data. In fact, this is precisely the mechanism that allows researchers to generalize the results of a study involving a small number of participants in order to draw conclusions about a larger population of similar individuals. For example, if a study finds that people who rank low on agreeableness are more likely to

---

56. *See* Kristen M. Altenburger & Johan Ugander, *Monophily in Social Networks Introduces Similarity Among Friends-of-Friends*, 2 NATURE HUM. BEHAV. 284, 284 (2018) (finding that "even if an individual does not disclose private attribute information about themselves (such as their gender, age, race or political affiliation), methods for relational learning can leverage attributes disclosed by that individual's similar friends to possibly predict their private attributes").

57. *See* Yingxiao Wu, Yan Zhuang, Xi Long, Feng Lin & Wenyao Xu, *Human Gender Classification: A Review*, INT'L J. BIOMETRICS 1, 6 (2016) (describing gender classification systems and how they operate); Michal Kosinski, David Stillwell & Thore Graepel, *Private Traits and Attributes are Predictable from Digital Records of Human Behavior*, 110 PROC. NAT'L ACAD. OF SCIS. 5802, 5802 (2013) (demonstrating "that easily accessible digital records of behavior, Facebook Likes, can be used to automatically and accurately predict a range of highly sensitive personal attributes including: sexual orientation, ethnicity, religious and political views, personality traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender").

58. *See* Viljoen, *supra* note 7, at 581.

59. *See id.* at 578.

exhibit compulsive buying behavior,[60] this finding could enable researchers to infer the buying behavior of individuals who did not participate in the study but whose ranking on the agreeableness scale is known. This Article argues that users' interests as a collective are currently severely underrepresented in regulatory discourse, despite the significant role that relationships between users—and the consequential and invasive inferences they enable—have played in the development of the data ecosystem.[61]

In summary, the collective, interdependent nature of personal data means that no single individual can decide on their own how much data they want to disclose to platforms, or what data they want to keep private. Therefore, any intervention in the data ecosystem must be characterized by a deep understanding of the strong collective nature of data and the various dependencies that characterize data. As the following Parts will discuss in detail, one substantial policy implication of this finding is the need to generate a collective perspective within the data ecosystem.

III. APPROACHES TO OVERCOMING HARMS FROM INCOMING-VECTOR PERSONALIZATION

Recent sessions of Congress have produced an abundance of bills that aim to address the harms associated with incoming-vector personalization,[62] alongside a vast array of other regulatory and technical proposals being introduced around the world.[63] This Part analyzes the tools presented in a selection of recent proposals as a reflection more broadly of emerging regulatory approaches to overcoming challenges created by incoming-vector personalization.[64] It

60. *See* Kiran Shehzadi, Muhammad Ahmad-ur-Rehman, Anam Mehmood Cheema & Alishba Ahkam, *Impact of Personality Traits on Compulsive Buying Behavior: Mediating Role of Impulsive Buying*, 9 J. SERV. SCI. & MGMT. 416 (2016).

61. *See* Viljoen*, supra* note 7, at 613.

62. *See, e.g.*, Justice Against Malicious Algorithms Act of 2021, H.R. 5596, 117th Cong. (2021). For a detailed review of a wide range of initiatives adopting a proprietarian rationale for regulating data collection and use proposed by technologists, economists, legal scholars, politicians and even a presidential candidate, see Viljoen, *supra* note 7, at 617.

63. *See, e.g.*, GDPR, *supra* note 20.

64. While many of the harms these interventions seek to address are personalization-driven, others stem from non-personalization-related design choices made by platforms. Two examples of the latter category are addictive features and the use of so-called dark patterns to manipulate user behavior, which are the focus, for example, of the Deceptive Experiences to Online Users Reduction (DETOUR) Act, S. 1084, 116th Cong. (2019), as well as of the Social Media Addiction Reduction Technology (SMART) Act, S. 2314, 116th Cong. (2019). In this Article, the Authors recognize that personalization of addictive design features or dark

begins by discussing challenges of liability and enforcement, including the enforcement of antidiscrimination laws against harmful advertising, the enforcement of data protection laws against harmful platform personalization, and platforms' liability vis-à-vis Section 230 of the Communications Decency Act. It then discusses recent regulatory proposals and analyzes the extent to which they incorporate a collective perspective to adequately combat the harms they are intended to address. Finally, this Part presents a selection of technological solutions that policy makers have proposed to address the challenges created by incoming-vector personalization.

A. Liability and Enforcement Mechanisms

Enforcing existing laws—such as antidiscrimination, consumer protection, and privacy and data protection laws—and reforming existing liability protections for platforms may address certain aspects of harmful platform personalization.

1. Antidiscrimination Law

US antidiscrimination laws, for instance, prohibit discrimination in ads for housing and job opportunities based on protected attributes such as race, sex, age, and religion.[65] In some cases, the content of the ads may not be inherently discriminatory, but the targeting criteria produce discriminatory effects by excluding certain groups on the basis of protected characteristics. For example, Pauline T. Kim and Sharion Scott have identified at least three potential ways in which employment recruiting via targeted advertising can produce discriminatory effects.[66] The first occurs when advertisers use protected attributes as their targeting criteria—for example, by selecting an audience of only men aged eighteen to forty or by excluding people belonging to an ethnic minority.[67] The second occurs when an advertiser

patterns can substantially amplify the harms they create. However, the non-personalization-driven aspects of these features are not the main focus of this Article; rather, the Authors limit their focus to the harms arising from incoming-vector content that is personalized for different users based on data collected along the outgoing vector.

65. *See, e.g.*, 42 U.S.C. § 3604(c) (prohibiting discrimination in advertising for housing opportunities); 42 U.S.C. § 2000e-3(b) (prohibiting discrimination in job advertisements based on protected characteristics); 29 U.S.C. § 623(e) (prohibiting discrimination in advertising of job opportunities on the basis of age).

66. *See* Kim & Scott, *supra* note 37, at 98.

67. *See id.* In 2016, ProPublica reported on how the Facebook ad targeting platform allows advertisers to place housing ads that explicitly exclude from their targeting criteria users with African American, Asian American, or Hispanic affinity. *See* Julia Angwin & Terry Parris, Jr.,

selects targeting criteria based on seemingly mundane attributes, such as ZIP code or expressed interests, which are strongly correlated with, and in effect serve as a proxy for, a protected attribute.[68] While such a method of targeting may result in discriminatory effects, it may be difficult to anticipate *ex ante*.[69] The third occurs when the job's advertiser uses a tool like Facebook's "lookalike audience" feature to identify and screen for a relevant audience based on a sample group, such as the employer's current workforce.[70] If the sample group is biased, this tool will produce an audience that reflects the same bias.[71]

A strong argument exists that the first source of discriminatory impact is prohibited by laws such as Title VII of the Civil Rights Act of 1964 and the Age Discrimination in Employment Act of 1967.[72] However, because platforms rarely disclose their incoming and outgoing vector flows of information, regulators and watchdog groups lack the evidence they would need to launch meaningful investigations into ad

---

Facebook Lets Advertisers Exclude Users by Race, PROPUBLICA (Oct. 28, 2016, 1:00 PM), https://www.propublica.org/article/facebook-letsadvertisers-exclude-users-by-race [https://perma.cc/F4JV-RUPA]. Lawsuits have also alleged that the Facebook ad platform enables the placement of discriminatory advertising. *See, e.g.*, Bradley v. T-Mobile US, Inc., No. 17-cv-07232-BLF, 2020 U.S. Dist. LEXIS 44102 (N.D. Cal. Mar. 13, 2020) (dismissing a class action lawsuit against T-Mobile and Amazon for allegedly routinely using ad-targeting criteria that exclude users over the age of forty from being presented with job ads they placed on Facebook, but outlining criteria for the plaintiffs to file a new complaint and allowing for additional discovery). In response to the reports of discrimination, Facebook announced changes to its targeting mechanism in order to comply with existing antidiscrimination laws. *See* Julia Angwin, *Facebook Says it Will Stop Allowing Some Advertisers to Exclude Users by Race*, PROPUBLICA (Nov. 11, 2016, 10:00 AM), https://www.propublica.org/article/facebook-to-stop-allowing-some-advertisers-to-exclude-users-by-race [https://perma.cc/MX6P-22DH]. In 2017, ProPublica found that Facebook still enabled discriminatory targeting of housing ads. *See* Julia Angwin, Ariana Tobin & Madeleine Varner, *Facebook (Still) Letting Housing Advertisers Exclude Users by Race*, PROPUBLICA (Nov. 21, 2017, 1:23 PM), https://www.propublica.org/article/facebook-advertising-discrimination-housing-race-sex-national-origin [https://perma.cc/2P42-X8WW].

68. *See* Kim & Scott, *supra* note 37, at 98.

69. For example, in areas with a high degree of residential segregation, location, particularly ZIP code, may serve as a proxy for race. *See id.* In other cases, the demographic characteristics of the audience created by the selected combination of targeting criteria may be more difficult to predict. *See id.* at 99.

70. *See id.* at 98; *About Lookalike Audiences*, META, https://www.facebook.com/business/help/164749007013531 [https://perma.cc/ATV5-GK4Q] (last visited Apr. 1, 2023).

71. Targeting potential employees based on a "lookalike" audience criterion could also be seen as similar to recruiting via word of mouth. *See* Kim & Scott, *supra* note 37, at 116; Speicher et al., *supra* note 34, at 7, 11. In *Thomas v. Washington County School Board*, the court found that advertising for job applicants using existing employees' word of mouth had a discriminatory effect and "serve[s] to freeze the effects of past discrimination," whether the employer had discriminatory intent or not. 915 F.2d 922, 925 (4th Cir. 1990).

72. *See* Kim & Scott, *supra* note 37, at 113.

targeting, especially concerning discrimination of the second or third type.[73]

2. Consumer Protection Law

The US Federal Trade Commission (FTC), exercising its investigatory power and its authority to bring enforcement actions against companies that engage in unfair and deceptive trade practices,[74] has also been active in addressing incoming-vector harms.[75] For example, in 2019, the FTC brought an enforcement action against Devumi, a business that sold fake social media followers, views, and likes to buyers seeking to inflate their influence metrics on platforms—a practice that can facilitate the spread of fake product reviews, spam, manipulation, and disinformation.[76] In its complaint, the FTC alleged that Devumi violated the FTC Act by enabling its customers to mislead the public, thereby providing them with the "means and instrumentalities" to commit deceptive acts or practices.[77] Additionally, in December 2020, the FTC launched an investigation into the practices of nine social media companies, requiring them to disclose information about content moderation and the effects of their practices on children and teenagers, among other things.[78] Facebook whistleblower Frances Haugen claimed that the company's internal research demonstrated knowledge of personalization-driven harms.[79]

---

73. *See id*. at 116.

74. *See* 15 U.S.C. § 45(a)(1) (providing that "unfair or deceptive acts or practices in or affecting commerce . . . are . . . declared unlawful"); 15 U.S.C. § 46(b) (providing the Commission with the authority to require certain entities engaged in commerce to file "annual or special . . . reports or answers in writing to specific questions").

75. *See* FED. TRADE COMM'N, SOCIAL MEDIA BOTS AND DECEPTIVE ADVERTISING (2020).

76. *See* Complaint at 1, Fed. Trade Comm'n v. Devumi, LLC, No. 9:19cv81419 (S.D. Fla. Oct. 18, 2019); Stipulated Order for Permanent Injunction and Monetary Judgment at 1, Fed. Trade Comm'n v. Devumi, LLC, No. 19-81419-CIV-ALTMAN/Brannon (S.D. Fla. Oct. 22, 2019).

77. Complaint at 5, *Devumi, LLC*, No. 9:19cv81419. The court order settling this complaint imposed a $2.5 million judgment against Devumi's owner. *See* Stipulated Order at 3, *Devumi*, No. 19-81419-CIV-ALTMAN/Brannon.

78. *See* Press Release, *FTC Issues Orders to Nine Social Media and Video Streaming Services Seeking Data About How They Collect, Use, and Present Information*, FED. TRADE COMM'N (Dec. 14, 2020) https://www.ftc.gov/news-events/press-releases/2020/12/ftc-issues-orders-nine-social -media-video-streaming-services [https://perma.cc/Z74U-9LEZ].

79. *See* John D. McKinnon & Brent Kendall, *Federal Trade Commission Scrutinizing Facebook Disclosures*, WALL ST. J., https://www.wsj.com/articles/facebook-ftc-privacy-kids-11635289993 [https://perma.cc/QX8T-ZUQR] (Oct. 27, 2021, 12:38 PM).

In light of this, FTC staffers are reportedly exploring whether Facebook engaged in deceptive or unfair trade practices.[80]

3. Privacy and Data Protection Law

Enforcing existing data protection regulations, such as the European Union's GDPR, offers an additional mechanism for addressing certain incoming-vector harms. Personalization ostensibly encroaches on privacy rights,[81] for example, by undermining individuals' right to be left alone,[82] by curtailing their right to play a meaningful part in their self-determination, and by negatively affecting their ability to "maintain relational ties and to develop critical perspectives on the world around them."[83]

Data protection principles, such as data minimization and purpose limitation,[84] likely serve to curb platforms' collection, use, and

---

80. *See id.*; EUROPEAN DATA PROTECTION BOARD, GUIDELINES 8/2020 ON THE TARGETING OF SOCIAL MEDIA USERS 5 (2020) ("Targeting of social media users may involve uses of personal data that go against or beyond individuals' reasonable expectations and thereby infringes applicable data protection principles and rules."); EUROPEAN DATA PROTECTION SUPERVISOR, EDPS OPINION 3/2018 ON ONLINE MANIPULATION AND PERSONAL DATA 15 (2018) ("The concern of using data from profiles for different purposes through algorithms is that the data loses its original context. Repurposing of data is likely to affect a person's informational self-determination, further reduce the control of data subjects over their data, thus affecting the trust in digital environments and services."). Other jurisdictions have also recently enacted data protection regulations influenced by the GDPR, such as the Lei Geral de Proteção de Dados (LGPD) in Brazil, the proposed Digital Charter Implementation Act in Canada, and the California Consumer Privacy Act (CCPA) as amended by the California Privacy Rights Act. The Authors discuss the rights provided by the GDPR as a reflection of general regulatory trends with respect to data protection. *See* The General Law for the Protection of Personal Data (LGPD), Law no. 13,709/2018 (Brazil); Digital Charter Implementation Act, 2022, 44th Parliament, 1st Session (Canada); California Consumer Privacy Act, Cal. Civ. Code § 1798.100 et seq.

81. *See* Daniel J. Solove, *"I've Got Nothing to Hide" and Other Misunderstandings of Privacy*, 44 SAN DIEGO L. REV. 745, 756–57 (2007) (including data-driven harms under the umbrella of "privacy" such as "problems of information processing . . . [that] frustrate the individual by creating a sense of helplessness and powerlessness, but they also affect social structure by altering the kind of relationships people have with the institutions that make important decisions about their lives").

82. *See* Samuel D. Warren & Louis D. Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193, 195 (1890); Daniel J. Solove, *Conceptualizing Privacy*, 90 CAL. L. REV. 1087, 1105 (2002) (characterizing the right to be left alone as capturing a common understanding of privacy); Bart van der Sloot, *The Right to be Let Alone by Oneself: Narrative and Identity in a Data-Driven Environment*, 13 LAW, INNOVATION & TECH. 223, 226 (2021) (proposing a reformulation of "the right to privacy that also includes a right to be protected from information-communication to oneself—a right to be let alone by oneself").

83. *See* Julie E. Cohen, *What Privacy is For*, 126 HARV. L. REV. 1904, 1906 (2013); *see also infra* Section IV.A.

84. *See, e.g.*, GDPR, *supra* note 20, at art. 5(1)(b)–(c) (providing that the collection of personal data must be limited to "specified, explicit and legitimate purposes and not further

retention of large quantities of fine-grained user data to target individuals with highly personalized content.[85] However, numerous scholars have observed that these principles seem incompatible with the analytics at the heart of platform personalization, which require platforms to disclose to users how they intend to use their data in the future, which may be intrinsically unforeseeable.[86] In addition, data controllers and processors must demonstrate an applicable legal basis, such as consent or legitimate interests,[87] to justify processing users' personal data for platform targeting, which may pose challenges, particularly in contexts in which profiling and tracking persist across multiple platforms.[88]

Further, the processing of special categories of personal data, namely "personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation" is prohibited,[89] unless an enumerated exception, such as "explicit consent . . . for one or more specified purposes," applies.[90] This creates challenges, since a seemingly innocuous data point, like a user's

processed in a manner that is incompatible with those purposes" and providing that personal data must be "adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed").

85. *See, e.g.*, Tal Zarsky, *Incompatible: The GDPR in the Age of Big Data*, 47 SETON HALL L. REV. 995, 1005 (2017).

86. *See id.* at 1006; Chris Jay Hoofnagle, Bart van der Sloot & Frederik Zuiderveen Borgesius, *The European Union General Data Protection Regulation: What It Is and What It Means*, 28 INFO. & COMMC'NS TECH. L. 65, 77–78 (2019) (observing that "[p]urpose limitation strikes at the heart of information-intensive industries, because companies so frequently find utility for data by using and repurposing the data in unforeseeable ways" and that, "[i]ndeed, the very purpose of machine learning is to discover patterns not anticipated or even perceivable to people").

87. GDPR, *supra* note 20, at art. 6(1)(a), (f).

88. *See* EUROPEAN DATA PROTECTION BOARD, *supra* note 80, at 16 (noting that the Article 29 Working Party "has previously considered that it would be difficult for controllers to justify using legitimate interests as a legal basis for intrusive profiling and tracking practices for marketing or advertising purposes, for example those that involve tracking individuals across multiple websites, locations, devices, services or data-brokering.") (citing Commission Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679 (2018)).

89. GDPR, *supra* note 20, at art. 9(1).

90. *Id.* at art. 9(2)(a).

geolocation information, may serve as a proxy to uncover protected personal characteristics like race and ethnicity.[91]

Recent European legislation posits that users have specific rights in the context of automated decision-making, which could potentially help restrict platform personalization.[92] With respect to automated decision-making, an individual has the rights to "obtain human intervention," "express his or her point of view," "contest the decision," to know of "the existence of automated decision-making, including profiling," and to receive "meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject."[93] Scholars have expressed doubt that these requirements will have a "significant practical impact on automated profiling," but they could conceivably apply where "advertising involves blatantly unfair discrimination in the form of web-lining and the discrimination has non-trivial economic consequences," particularly in situations where this conduct occurs on a repeated basis.[94]

4. Reforms to Platform Liability Protection

For decades, platforms have enjoyed legal protection from liability for harmful content posted by their users under laws such as Section 230 of the Communications Decency Act.[95] Such protection has been both heralded as integral to online free speech and criticized as "an ill-conceived shield for scoundrels."[96] There are growing calls to amend Section 230 to remove platforms' protection from liability in certain circumstances, seeking to hold platforms responsible for the

---

91. *See* Zarsky, *supra* note 85, at 1013; EUROPEAN DATA PROTECTION BOARD, *supra* note 80, at 5 ("Recent research suggests that the potential for discriminatory effects exists also without using criteria that are directly linked to special categories of personal data in the sense of Article 9 of the GDPR.") (citing Speicher et al., *supra* note 34).

92. *See* GDPR, *supra* note 20, at art. 22(1)–(2). Also note that the DMA prohibits sharing data between jointly owned platforms, which enables "deep consumer profiling." *See* Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act)

[hereinafter DMA].

93. *See* GDPR, *supra* note 20, at art. 9.

94. *See* Isak Mendoza & Lee A. Bygrave, *The Right Not to be Subject to Automated Decisions Based on Profiling*, *in* EU INTERNET LAW: REGULATION AND ENFORCEMENT 77, 89 (Tatiana-Eleni Synodinou, Philippe Jougleux, Christiana Markou & Thalia Prastitou eds., 2017).

95. 47 U.S.C. § 230.

96. David S. Ardia, *Free Speech Savior or Shield for Scoundrels: An Empirical Study of Intermediary Immunity Under Section 230 of the Communications Decency Act*, 43 LOY. L.A. L. REV. 373, 379–80 (2010).

active role they take in promoting harmful content to users who are likely to respond to it, which, in turn, amplifies its presence and impact.[97] Examples include the proposed Justice Against Malicious Algorithms Act,[98] the proposed Health Misinformation Act,[99] and the proposed Protecting Americans from Dangerous Algorithms Act,[100] each of which would remove immunity from liability for large platforms whose algorithms amplify particular forms of severely problematic content.[101]

B. Individual Control Via Notice and Consent

In addressing the harms stemming from platforms' ability to manipulate users and undermine their autonomy,[102] many proposals focus on individual control-based approaches,[103] such as notice-and-consent mechanisms. Such approaches often seek to

97. Note that the approaches adopted by the collection of legislative proposals discussed in this Section differ significantly from the proposal in the DSA, which places responsibility for content moderation on the platform, by requiring the largest online platforms to set up a notice-and-action mechanism allowing users to report content they believe is illegal. *See* Regulation 2022/2065 of the European Parliament and of the Council on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act), 2022 O.J. (L 277) [hereinafter DSA]. While platforms would not be liable for the content, upon receiving notice of the presence of allegedly unlawful content, platforms would be obligated to remove it and notify the poster that it had been removed. *See id.* at art. 5(1)(b).

98. Justice Against Malicious Algorithms Act of 2021, H.R. 5596, 117th Cong. (2021).

99. Health Misinformation Act of 2021, S. 2448, 117th Cong. (2021).

100. Protecting Americans from Dangerous Algorithms Act, H.R. 2154, 117th Cong. (2021).

101. Press Release, Anna G. Eshoo, Congresswoman, California's 16th Congressional District, *Reps. Eshoo and Malinowski Reintroduce Bill to Hold Tech Platforms Accountable for Algorithmic Promotion of Extremism* (Mar. 24, 2021), https://eshoo.house.gov/media/press-releases/reps-eshoo-and-malinowski-reintroduce-bill-hold-tech-platforms-accountable [https://perma.cc/EKX4-NP3X]; *see* H.R. 2154 § 2; S. 2448 § 3(a)(1)(B). The Secretary of Health and Human Services would be responsible for determining what content should be considered health misinformation. *See* S. 2448 § 3(b).

102. Such questions were raised in Europe in the context of the DSA's disclosure requirement in EUROPEAN DATA PROTECTION SUPERVISOR, EDPS OPINION 1/2021 ON THE PROPOSAL FOR A DIGITAL SERVICES ACT 17 (2021) (recognizing that "including information about the recommender system parameters and options in the terms and conditions would only make them difficult to find and understand for data subjects.").

103. Enhancement of individual control is one of the rationales underlying the fair information practice principles that have inspired many privacy and data protection regulations, such as the GDPR. *See* Julie E. Cohen, *Turning Privacy Inside Out*, 20 THEORETICAL INQUIRIES L. 1, 10 (2019) (discussing the challenges raised by privacy as control). Note, however, that some privacy scholars disagree with this framing of the GDPR. *See, e.g.*, Meg Leta Jones & Margot E. Kaminski, *An American's Guide to the GDPR*, 98 DENVER L. REV. 93, 93 (2021) ("We endeavor to correct common misconceptions about the GDPR: that it is primarily founded on individual consent (it is not); . . . and that it is primarily about individual rights and control (it is equally about risk management and corporate compliance).").

enhance individual control without recognizing data's collective nature nor providing meaningful insight into the role that personalization plays.[104] Instead, they often burden individuals with uninterpretable, empty choices, rendering the sense of control they convey a mirage.[105]

Individual autonomy is foundational to modern liberal societies and is a prerequisite for the realization of basic human rights such as freedom of expression, the capacity to shape opinions and values, and the choice between right and wrong.[106] The question of whether an action subverts individual autonomy does not always have a clear-cut answer; in fact, manipulative behavior extends across a spectrum.[107] At one end of the spectrum is mildly manipulative behavior, which platforms illustrate through personalized suggestions to post, for example, a "happy birthday" message to a friend's feed or to add another user to one's list of friends. While users may not understand exactly the information on which such recommendations are based, they likely recognize this as content that the platform creates, and the final decision whether to act on these recommendations remains within the user's discretion. At the other end of the spectrum are actions platforms take that users can neither discern nor avoid, such as conducting an experiment to manipulate users' moods without their informed consent.[108]

---

104. *See* Viljoen, *supra* note 7, at 582, 617 ("Individualist theories of informational interests result in legal proposals that . . . practically fall back on individuals to adjudicate between legitimate and illegitimate information production. This not only leaves certain social information harms unrepresented . . . [individualist theories] reduce legal interests in information to individualist claims subject to individualist remedies.").

105. *See* Woodrow Hartzog, *The Case Against Idealising Control*, 4 EUR. DATA PROT. L. REV. 423, 425 (2018) (detailing the limitations of the privacy as control paradigm).

106. *See* GERALD DWORKIN, THE THEORY AND PRACTICE OF AUTONOMY 10 (1988) ("As a political ideal, autonomy is used as a basis to argue against the design and functioning of political institutions that attempt to impose a set of ends, values, and attitudes upon the citizens of a society."); Susser et al., *supra* note 17, at 4–6 (defining manipulation as hidden interference that deprives us of authorship over our own choices); Julie E. Cohen, *Examined Lives; Informational Privacy and the Subject as an Object*, 52 STAN. L. REV. 1373, 1426 (2000) (suggesting that autonomy is a prerequisite for participation in the governance of a community); Antoinette Rouvroy & Yves Poullet, *The Right to Informational Self-Determination and the Value of Self-Development: Reassessing the Importance of Privacy for Democracy*, *in* REINVENTING DATA PROTECTION? 45, 47 (2009) ("Self-determination is an elementary functional condition of a free democratic community based on its citizens' capacity to act and to cooperate.").

107. *See* Tess M. Wilkinson, *Nudging and Manipulation*, 61 POL. STUD. 341, 342 (2013) (recognizing that there are different levels of manipulation); YOCHAI BENKLER, THE WEALTH OF NETWORKS: HOW SOCIAL PRODUCTION TRANSFORMS MARKETS AND FREEDOM 141 (2006) ("We experience some decisions as being more free than others.").

108. *See* Kramer et al., *supra* note 2 (reporting the Facebook emotional contagion experiment and its outcomes).

Notice-and-consent mechanisms might, at face value, seem to support individual autonomy; however, an extensive body of behavioral research calls their effectiveness into question.[109] This research demonstrates that individuals often fail to read or understand the implications of platforms' terms of service. Moreover, such agreements are contracts of adhesion, offered on a take-it-or-leave-it basis, precluding the ability of individual users to negotiate changes to their terms;[110] individuals, as participants in a knowledge-based economy, lack a meaningful choice to opt out of the use of digital platforms altogether. Therefore, investing the time and effort to read and understand these documents would be inefficient.[111] Consequently, proposals relying on control-based mechanisms such as notice and consent burden individuals with a pseudo-choice that they are not equipped to make and, in the process, absolve platforms of responsibility for the harms created by platform personalization.[112]

The European Commission has recently passed two legislative initiatives reflecting control-based approaches: the DSA and the Digital

---

109. *See, e.g.*, Joel R. Reidenberg, Travis Breaux, Lorrie Faith Cranor, Brian French, Amanda Grannis, James T. Graves, Fei Liu, Aleecia McDonald, Thomas B. Norton, Rohan Ramanath, N. Cameron Russell, Norman Sadeh & Florian Schaub, *Disagreeable Privacy Policies: Mismatches between Meaning and Users' Understanding*, 30 BERKELEY TECH. L. J. 39, 41 (2015); Ian Ayres & Alan Schwartz, *The No-Reading Problem in Consumer Contract Law*, 66 STAN. L. REV. 545, 606 (2014); Samuel I. Becher & Tal Z. Zarsky, *Minding the Gap*, 51 CONN. L. REV., 69, 73 (2019); David A. Hoffman, *Relational Contracts of Adhesion*, 85 U. CHI. L. REV. 1395, 1396–98 (2018); Kevin Litman-Navarro, Opinion, *We Read 150 Privacy Policies. They Were an Incomprehensible Disaster.,* N.Y. TIMES, June 12, 2019, https://www.nytimes.com/interactive/2019/06/12/opinion/facebook-google-privacy-policies.html [https://perma.cc/JS9L-UJY9] (last visited Apr. 2, 2023); Uri Benoliel & Samuel I. Becher, *The Duty to Read the Unreadable*, 60 B.C. L. REV. 2255, 2257–58 (2019); Yannis Bakos, Florencia Marotta-Wurgler & David R. Trossen, *Does Anyone Read the Fine Print? Consumer Attention to Standard-Form Contracts*, 43 J. LEG. STUD. 1, 6 (2014).

110. *See* Clayton P. Gillette, *Rolling Contracts as an Agency Problem*, 2004 WIS. L. REV. 679, 680 (2004) (arguing that "failure to read may be perfectly rational, especially given the inability to negotiate around terms").

111. Research has suggested that if every user read every privacy policy they agreed to in a year, it would result in $781 billion in lost productivity. *See* Aleecia M. McDonald & Lorrie Faith Cranor, *The Cost of Reading Privacy Policies*, 4 I/S: J. L. & POL'Y FOR INFO. SOC'Y 543, 564 (2008); Melvin A. Eisenberg, *Behavioral Economics and Contract Law*, *in* THE OXFORD HANDBOOK OF BEHAVIORAL ECONOMICS AND THE LAW 438 (Eyal Zamir & Doron Teichman eds., 2014) (noting that "analyzing [the terms of standard form contracts] would often be unduly costly"); Lorrie Faith Cranor, Candice Hoke, Pedro Giovanni Leon & Alyssa Au, *Are They Worth Reading? An In-Depth Analysis of Online Advertising Companies' Privacy Policies*, CONF. ON COMM., INFO. & INTERNET POL'Y (TPRC 2014) (finding a lack of transparency in the privacy policies of seventy-five online tracking companies and a confusing lack of consistent terminology).

112. *See* Julie E. Cohen, *How (Not) to Write a Privacy Law*, KNIGHT FIRST AMENDMENT INST. (Mar. 23, 2021), https://knightcolumbia.org/content/how-not-to-write-a-privacy-law [https://perma.cc/Q67X-MA8V].

Markets Act (DMA).[113] The DSA recognizes the tremendous power that platforms wield from their ability to control the content they present to users and demands that platforms exercise this power responsibly.[114] In line with this approach, the DSA requires very large platforms to provide notice in their terms of service that content has been algorithmically generated and to detail the main parameters used by recommender systems.[115] These platforms must also allow their users the ability to modify the parameters used by recommender systems—for instance, by providing at least one option to opt out of recommendations based on profiling.[116] In contrast, the focus of the DMA is the functioning and competitiveness of the market, not the rights of a particular user. Its disclosure mandates aim to increase platform transparency vis-à-vis advertisers for the purpose of promoting competitive markets rather than for understanding platform-related harms.[117]

In contrast to an omnibus legislative proposal like the DSA, which seeks to address a wide range of incoming-vector harms, regulatory proposals in the United States tend to focus on combating specific categories of harms, such as those stemming from (1) platform experimentation and (2) filter bubbles, which platforms create by manipulating their presentation of personalized content.

1. Experimentation

The unique position of platforms within the data ecosystem enables them to experiment with the presentation of different types of

113. *See* DSA, *supra* note 97, at art. 29, rec. 62 (providing that "very large online platforms should ensure that recipients are appropriately informed, and can influence the information presented to them"). This approach is consistent with recent trends in EU data protection law as reflected in the GDPR. The DSA aims to bring EU regulation of the data ecosystem up to date and in particular will modernize Directive 2000/31/EC of the European Parliament and Council of 8 June 2000 on Certain Legal Aspects of Information Society Services, in Particular Electronic Commerce, in the Internal Market ('Directive on electronic commerce') O.J. (L 178); DMA, *supra* note 92; Caroline Cauffman & Catalina Goanta, *A New Order: The Digital Services Act and Consumer Protection*, 12 EUR. J. OF RISK REG. 758, 760 (2021).

114. Eline Chivot, *The New EU Rulebook for Online Platforms: How to Get it Right, Who Will it Impact and What Else is Needed?* 20 EUR. VIEW 121, 124 (2021); *see* DSA, *supra* note 97, at art. 29(1), rec. 62.

115. *See* DSA, *supra* note 97, at art. 29; EUROPEAN DATA PROTECTION SUPERVISOR, *supra* note 102, at 17 (suggesting that including information in platforms' terms and conditions is unlikely to enable users to become exposed to them or understand them better, and, instead, "[t]he EDPS strongly recommends to require that such information concerning the role and functioning of recommender systems to be presented separately, in a manner that should be easily accessible, clear for average users and concise").

116. *See* DSA, *supra* note 97, at art. 29(1), rec. 62.

117. *See* DMA, *supra* note 92, at art. 5(9).

content and observe how various categories of users respond.[118] Platforms continuously run such experiments, aiming to refine their personalization algorithms, boost the impact of content presented to users, and make ongoing changes to their interfaces in order to generate increased engagement. However, individuals may not be aware that platforms are experimenting on them, nor that the content they view is based on past experimentation that leveraged platforms' unique perspective within the data ecosystem. For one example, Facebook's mood manipulation experiment, which studied whether users' emotional states could be influenced by the content they were shown on the platform, sparked widespread criticism from civil society, academics, and regulators alike,[119] prompting Facebook COO Sheryl Sandberg to apologize for how the company had communicated the experiment to the public.[120] Critics claimed that Facebook's experiments effectively subverted its users' deliberative capacities,[121] treating them as "tools and fools" and insulting their dignity.[122]

---

118. *See* Zeynep Tufekci, *Engineering the Public: Big Data Surveillance and Computational Politics*, 19 FIRST MONDAY, no. 7 (2014) (arguing that platforms use computational politics to advance their own interests); Kramer et al., *supra* note 2 (reporting the Facebook emotional contagion experiment and its outcomes); Evan Selinger & Woodrow Hartzog, *Facebook's Emotional Contagion Study and the Ethical Problem of Co-opted Identity in Mediated Environments Where Users Lack Control*, 12 RSCH. ETHICS 35, 35 (2016) (describing the problematic aspects of the Facebook experiment). Before introducing its new "care" button during the COVID-19 pandemic, Facebook experimented with its use on a subgroup of its users. Several years earlier it conducted a similar experiment for adding a flower reaction before Mother's Day in several markets. While the care button was a success, the flower button was not. Andrew Hutchinson, *Facebook's Testing a New COVID-19-Themed Reaction Emoji*, SOCIALMEDIATODAY (Mar. 31, 2020), https://www.socialmediatoday.com/news/facebooks-testing-a-new-covid-19-themed-reaction-emoji/575152 [https://perma.cc/2LHK-XRR9].

119. *See* Kashmir Hill, *Facebook Manipulated 689,003 Users' Emotions for Science,* FORBES (June 28, 2014, 2:00 PM), https://www.forbes.com/sites/kashmirhill/2014/06/28/facebook-manipulated-689003-users-emotions-for-science [https://perma.cc/8NPR-WKEQ] (reporting that Facebook acknowledged the nature of the experiment).

120. Facebook COO Sheryl Sandberg clarified that the company was not apologizing for the experiment itself but rather for the way it was communicated. *See* Gail Sullivan, *Sheryl Sandberg Not Sorry for Facebook Mood Manipulation Study,* WASH. POST (July 3, 2014, 6:21 AM), https://www.washingtonpost.com/news/morning-mix/wp/2014/07/03/sheryl-sandberg-not-sorry-for-facebook-mood-manipulation-study/ [https://perma.cc/XPQ8-KATJ] ("[Sandberg] expressed regret over how the company communicated its 2012 mood manipulation study of 700,000 unwitting users, but she did not apologize for conducting the controversial experiment. It's just what companies do, she said.").

121. *See* CASS R. SUNSTEIN, THE ETHICS OF INFLUENCE: GOVERNMENT IN THE AGE OF BEHAVIORAL SCIENCE 86 (2016) (explaining that behavior that "subverts the target's rational capacities" can be manipulative).

122. *See* Wilkinson, *supra* note 107, at 345 ("To manipulate people is to treat them as both tools and fools.").

An example of a recent legislative proposal to address the harms of platform experimentation is the Deceptive Experiences To Online Users Reduction (DETOUR) Act, introduced in 2019 and 2021 by Sens. Mark Warner (D-Va.), Deb Fischer (R-Nev.), and their colleagues. In 2022 the bill's sponsors announced new endorsements from scientific and nonprofit organizations and from academics.[123] The bill seeks "[t]o prohibit the usage of exploitative and deceptive practices by large online operators."[124] In particular, it obligates platforms that conduct psychological or behavioral experiments on their users to receive users' informed consent and to periodically disclose to users and the general public any experiments being conducted by the platform.[125]

However, because the scope of the DETOUR Act is limited to a narrow subset of personalization constructed from psychological or behavioral experiments, this proposal likely fails to address harms from other, similar types of testing. Further, because the bill relies solely on tools that enable the individual user to exercise control over the content she sees, it fails to recognize the collective nature of data. For example, if Alice has opted out of platform experimentation, but Bob, a friend of Alice (or someone judged by the platform to be in some way similar to Alice), has not, Alice might still see content Bob has interacted with as part of the experiment because platforms base their recommendations to an individual on the content viewed by that person's social connections on the platform.

2. Filter Bubbles

Scholars, politicians, and the media have expressed concern regarding platforms' role in amplifying extremism and polarization by channeling progressively more extreme content to users based on their interests and opinions.[126] Because users are shown content that

123. Press Release, Mark R. Warner, U.S. Senator, Virginia, *Lawmakers Announce Additional Support for Bipartisan, Bicameral Legislation to Ban Manipulative 'Dark Patterns'* (June 15, 2022), https://www.warner.senate.gov/public/index.cfm/2022/6/lawmakers-announce-additional-support-for-bipartisan-bicameral-legislation-to-ban-manipulative-dark-patterns [https://perma.cc/N5U2-Q75W].

124. Deceptive Experiences to Online Users Reduction (DETOUR) Act, S. 1084, 116th Cong. (2019).

125. *See id*. The bill also addresses other non-personalization-driven harms, in particular certain aspects of addiction, by prohibiting design features aimed at cultivating compulsive usage of the platform in children under the age of thirteen. *See id.* § 3(a)(1)(C).

126. *See* Julie E. Cohen, *Tailoring Election Regulation: The Platform Is the Frame*, 4 GEO. L. & TECH. REV. 641, 647 (2020); Luke Munn, *Angry by Design: Toxic Communication and Technical Architectures*, 7 HUM. & SOC. SCI. COMMC'N., no. 53, at 6 (2020) ("Recommending content based on engagement, then, often means promoting incendiary, controversial, or polarizing

increasingly reaffirms their existing beliefs and reflects the opinions of users similar to them, each newsfeed can quickly turn into an echo chamber[127] or filter bubble, in which users face little or no exposure to opinions or news reports that contradict their beliefs.[128] "Interactional polarization" and social fragmentation are vital concerns,[129] as

content."); Joseph B. Bak-Coleman, Mark Alfano, Wolfram Barfuss, Carl T. Bergstrom, Miguel A. Centeno, Iain D. Couzin, Jonathan F. Donges, Mirta Galesic, Andrew S. Gersick, Jennifer Jacquet, Albert B. Kao, Rachel E. Moran, Pawel Romanczuk, Daniel I. Rubenstein, Kaia J. Tombak, Jay, J. Van Bavel & Elke U. Weber, *Stewardship of Global Collective Behavior*, 118 PROC. NAT'L ACAD. SCI. 1, 5 (2021) (describing how algorithmic decision-making can facilitate and increase polarization, extremism, and inequality); Center for Humane Technology, *A New Agenda for Tech,* VIMEO (Apr. 25, 2019), https://vimeo.com/332532972 (describing the ways in which platforms encourage extremism); Manuel Ricardo Torres-Soriano, *The Dynamics of the Creation, Evolution, and Disappearance of Terrorist Internet Forums*, 7 INT'L J. CONFLICT & VIOLENCE 164, 167–68 (2013) (explaining how online forums help promote radical jihadist positions); Jeff Horowitz & Deepa Seetharaman, *Facebook Executives Shut Down Efforts to Make the Site Less Divisive,* WALL ST. J. (May 26, 2020, 11:38 AM), https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499 [https://perma.cc/UB9B-GTJK] (reporting that Facebook acknowledges that its algorithms "exploit the human brains' attraction to divisiveness"); *Protecting Kids Online: Testimony from a Facebook Whistleblower: Hearing Before the Subcomm. on Consumer Prot., Prod. Safety, and Data Sec. of the S. Comm. on Com., Sci. & Transp.*, 117th Cong. (2021) (written statement of Frances Haugen) [hereinafter *Frances Haugen, Written Testimony*] ("The result has been a system that amplifies division, extremism, and polarization—and undermining societies around the world.").

127. *See* Dominic Spohr, *Fake News and Ideological Polarization: Filter Bubbles and Selective Exposure on Social Media*, 34 BUS. INFO. REV. 150, 151 (2017) ("The key issue here is that these groups, convinced of the echo that surrounds them with their own views and preconceptions, in a sense lose the inclination to proactively discuss ideas with people or groups of a different opinion.").

128. *See* Bail et al., *supra* note 16 at 9216 ("Social media sites are often blamed for exacerbating political polarization by creating "echo chambers" that prevent people from being exposed to information that contradicts their preexisting beliefs."); Guy Aridor, Duarte Gonçalves & Shan Sikdar, *Deconstructing the Filter Bubble: User Decision-Making and Recommender Systems*, 14 ASS'N COMPUTING MACH. CONF. ON RECOMMENDER SYS. 82, 82 (2020) (describing that platforms that offer personalized suggestions can lead users "into filter bubbles where they effectively get isolated from a diversity of viewpoints or content"). Exposure to others teaches individuals about themselves and to shape their opinions. HANNAH ARENDT, THE HUMAN CONDITION 50 (1998) ("The presence of others who see what we see and hear what we hear assures us of the reality of the world and ourselves.").

129. *See* Moran Yarchi, Christian Baden & Neta Kligler-Vilenchik, *Political Polarization on the Digital Sphere: A Cross-Platform, Over-Time Analysis of Interactional, Positional, and Affective Polarization on Social Media*, 38 POL. COMMC'N. 98 (2021) (explaining that interactional polarization "focuses on a process whereby participants in a debate increasingly interact with like-minded individuals, while disengaging from interactions with others who hold opposing viewpoints"); Robert M. Bond, Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle & James H. Fowler, *A 61-Million-Person Experiment in Social Influence and Political Mobilization*, 489 NATURE 295 (2012) (reporting the results of an experiment showing that Facebook users who were presented with a message encouraging them to vote and information about Facebook friends of theirs who had voted, participated in the election at higher rates than people who were only presented with a message encouraging them to vote, without the social context).

deliberation, persuasion, and compromise with opposing views—central to democratic functions[130]—are precluded by the very nature of the personalized experience each user encounters on online platforms.[131]

Against this backdrop, the Filter Bubble Transparency Act (FBTA) was introduced in 2019 and 2021 by Senator John Thune (R-S.D.) and colleagues.[132] The bill seeks to implement disclosure and consent requirements to address the rise of filter bubbles on large platforms.[133] In particular, it would require large platforms to disclose to users that they use algorithms based on users' data (collected and inferred) to select the content presented to them and the order in which it is presented.[134] Additionally, the FBTA would require platforms to enable users to opt out of the filter bubble and instead view an input-transparent version of the platform—for example, a newsfeed that was not algorithmically personalized based on user-provided content by presenting content in reverse chronological order.[135] This approach is similar to the DSA's requirement for platforms using recommender systems to notify their users and enable them to opt out of seeing content based on profiling.[136]

---

130. *See* ROBERT HUCKFELDT, PAUL E. JOHNSON & JOHN SPRAGUE, POLITICAL DISAGREEMENT: THE SURVIVAL OF DIVERSE OPINIONS WITHIN COMMUNICATION NETWORKS 1–24 (2004) (explaining that political deliberation between people has the potential to enhance democratic aspects); Diana C. Mutz, *Cross-Cutting Social Networks: Testing Democratic Theory in Practice*, 96 AM. POL. SCI. REV. 111, 111 (2002) ("Political talk is central to most current conceptions of how democracy functions.").

131. *See* Spohr, *supra* note 127 at 151; Cohen, *supra* note 83, at 1907 ("In its ideal form, the liberal self-possesses both abstract liberty rights and the capacity for rational deliberation and choice and is capable of exercising its capacities in ways uninfluenced by cultural context."); Cohen, *supra* note 126, at 659 (discussing the presumption that more information will lead people to in depth discourse which is in and of itself a noble goal, but noting that the reality is far from this ideal); Robert Post, *The Constitutional Status of Commercial Speech*, 48 UCLA L. REV. 1, 7 (2000) ("Public discourse is comprised of those processes of communication that must remain open to the participation of citizens if democratic legitimacy is to be maintained."); FRANK PASQUALE, THE BLACK BOX SOCIETY 61 (2015) ("The power to include, exclude, and rank is the power to ensure which public impressions become permanent and which remain fleeting.").

132. *See* Filter Bubble Transparency Act, S. 2763, 116th Cong. (2019).

133. *See* Adi Robertson, *The Senate's Secret Algorithms Bill Doesn't Actually Fight Secret Algorithms*, VERGE (Nov. 5, 2019, 8:01 AM), https://www.theverge.com/2019/11/5/20943634/senate-filter-bubble-transparency-act-algorithm-personalization-targeting-bill [https://perma.cc/4XC3-7XM9].

134. *See* S. 2763 § 3(b)(1)(A) ("The person provides notice to users of the platform that the platform uses an opaque algorithm that makes inferences based on user specific data to select the content the user sees.").

135. *See id.* § 3(b)(1)(B) ("The person makes available a version of the platform that uses an input-transparent algorithm and enables users to easily switch between [the two versions].").

136. *See* DSA, *supra* note 97, at art. 29; *id.* rec. 62 (requiring that "very large online platforms . . . ensure that recipients are appropriately informed, and can influence the information presented to them").

One question this bill raises is what criteria platforms would use to generate an alternative newsfeed not based on user-provided content.[137] For instance, would users of social networks still see content posted, liked, or shared by their social contacts, content from groups they belong to, or pages they have liked? If so, this outcome would challenge the assumption that social media users could remove themselves from filter bubbles while both remaining active and connected with other users that continue to utilize personalization algorithms. Further, this assumption reflects a lack of understanding of the collective nature of data—an individual who "opts out" would still see a newsfeed laced with content that personalization algorithms have promoted to her social media contacts.[138]

Without an overhaul of the current approach to control-based mechanisms, such mechanisms are unlikely to provide greater protection of individual autonomy.[139] In particular, an effective consent-based mechanism must ensure individuals are able to make meaningful and consequential choices regarding authorized uses of their data, including permissible types of personalization.[140] Additionally, individuals must be presented with more than one viable option to choose from, the consent process must not be overly burdensome, and individuals must be meaningfully informed about the

137. One option discussed in this context is that the default feed would be similar to the sparkle icon option on Twitter. Since 2018, Twitter has provided users with two options to view their newsfeed: either Twitter's choice of top Tweets, or, for those users who opt out of this view by selecting the sparkle icon, tweets from accounts they follow in reverse chronological order. @TwitterSupport, TWITTER (Dec. 19, 2018, 4:39 PM), https://twitter.com/twittersupport/status/1075506037820579841 [https://perma.cc/XS5J-TFAR]; Will Oremus, *Twitter Has Finally Made It Easy to Set Your Timeline to Reverse-Chronological,* SLATE (Dec. 18, 2018, 12:15 PM), https://slate.com/technology/2018/12/twitter-reverse-chronological-timeline-setting.html [https://perma.cc/4P97-HXUY].

138. *See* Natali Helberger, Max van Drunen, Sanne Vrijenhoek & Judith Möller, *Regulation of News Recommenders in the Digital Services Act: Empowering David Against the Very Large Online Goliath*, INTERNET POL'Y REV. (Feb. 26, 2021), https://policyreview.info/articles/news/regulation-news-recommenders-digital-services-act-empowering-david-against-very-large [https://perma.cc/NM3J-UNB6].

139. *See* SUSAN BENESCH, DANGEROUS SPEECH PROJECT, PROPOSALS FOR IMPROVED REGULATION OF HARMFUL ONLINE CONTENT 22 (2020).

140. *See* Daniel Solove, *Introduction: Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1880, 1894 (2013) (recognizing the challenges of a notice and consent regime while expressing concern that regulation compelling certain privacy choices may be too paternalistic); Viljoen, *supra* note 7, at 594 ("Notice and consent structures the basic legal relationship between the individual consumer (the data subject) and the digital service provider (the data processor).").

ramifications of each choice.[141] Inasmuch as platforms' incentives remain fixed, however, countering the harmful effects of platform personalization will require entrusting a third-party body with a collective perspective, as Part IV outlines below.

C. Transparency Mandates

A third category of interventions includes mandates for platforms to disclose certain information regarding personalization to third parties for transparency and accountability purposes. Disclosure requirements that recognize the collective aspect of data are critical to constructing policies that address platform harms such as disinformation and discrimination. The need to mandate transparency by statute has been underscored by platforms' recent attempts to block third parties from collecting information about outgoing- and incoming-vector content. In August 2021, Facebook shut down the accounts of three New York University researchers who had initially been granted access to conduct a study regarding political ads on the platform[142] on the grounds that they had violated the platform's terms of service prohibiting the use of automated scraping tools. Such scraping, Facebook alleged, posed risks to individual privacy.[143] Facebook has responded to criticism about its lack of transparency by making certain data available to researchers. However, researchers have noted that their access has been too limited to enable them to effectively study harms such as disinformation and manipulation on the platform;

---

141. *See* DSA, *supra* note 97, at art. 12(1). The DSA seeks to establish a standard for increased clarity for users with regards to the terms and services provided by platforms. The DSA requires platforms to include certain information in "clear and unambiguous language" and "in an accessible format" in policies regarding content moderation as well as information about platforms' use of recommender systems. *See id.* arts. 12, 29.

142. *See* Laura Edelson & Damon McCoy, Opinion, *We Research Misinformation on Facebook. It Just Disabled Our Accounts*, N.Y. TIMES (Aug. 10, 2021), https://www.nytimes.com/2021/08/10/opinion/facebook-misinformation.html [https://perma.cc/S2D2-YXQJ]. This action followed previous efforts by Facebook to thwart third-party transparency tools, including those from ProPublica, Mozilla, and AlgorithmWatch. *See* Jeremy B. Merrill & Ariana Tobin, *Facebook Moves to Block Ad Transparency Tools — Including Ours*, PROPUBLICA (Jan. 28, 2019, 4:29 PM), https://www.propublica.org/article/facebook-blocks-ad-transparency-tools [https://perma.cc/Z723-S49H]; Nicolas Kayser-Bril, *AlgorithmWatch Forced to Shut Down Instagram Monitoring Project After Threats from Facebook*, ALGORITHMWATCH (Aug. 13, 2021), https://algorithmwatch.org/en/instagram-research-shut-down-by-facebook [https://perma.cc/Z7XW-9WBL].

143. *See* Mike Clark, *Research Cannot Be the Justification for Compromising People's Privacy*, META (Aug. 3, 2021), https://about.fb.com/news/2021/08/research-cannot-be-the-justification-for-compromising-peoples-privacy/ [https://perma.cc/KS29-T5ND].

consequently, they contend that federal legislation mandating platform data sharing is urgently needed.[144]

Many proposals promoting transparency require platforms and advertisers to disclose advertisers' targeting metrics as well as other considerations that impact the presentation of ads to users.[145] Understanding these criteria may support efforts to address harms of discrimination and disinformation on social media platforms.

1. Discrimination

The Algorithmic Justice and Online Platform Transparency Act,[146] introduced by Senator Markey (D-Mass.) and Representative Matsui (D-Cal.-06), seeks to mandate transparency as a way to combat platforms' ability to use their algorithms in order to promote content in a discriminatory fashion.[147] Platforms would need to retain a record containing data about their algorithmic processes and, upon request, provide the FTC with access to the record.[148] These databases would store information about the personal data that platforms collect and how they use it, as well as information about what data was used for training platforms' algorithms and how platforms audit their algorithms to prevent discrimination.[149] If the algorithm promotes ads for services such as housing, education, employment, insurance, or credit, the platform would also need to assess whether the algorithm creates a disparate outcome based on a protected attribute.[150] The bill would also require platforms to publish a publicly available annual report of their content moderation practices.[151]

The annual report would include details about the number of content moderation decisions that platforms have made pursuant to the

144. *See* Simon Hegelich, *World View: Facebook Needs to Share More with Researchers*, 579 NATURE 473 (2020); Nathaniel Persily & Joshua A. Tucker, *Report: How to Fix Social Media? Start with Independent Research*, BROOKINGS INSTITUTION (Dec. 1, 2021), https://www.brookings.edu/research/how-to-fix-social-media-start-with-independent-research [https://perma.cc/A4XH-M5QL].

145. *See, e.g.*, Social Media Disclosure and Transparency (DATA) Act, H.R. 3451, 117th Cong. (2021); Algorithmic Justice and Online Platform Transparency Act, S. 1896, 117th Cong. (2021); Honest Ads Act, H.R. 4077, 115th Cong. (2017).

146. S. 1896.

147. The bill also promotes tools of disclosure to users; for example, it requires platforms to clearly disclose to users the categories of personal information collected, how it is collected, and what method the platform's algorithms use to promote or withhold content from users. *See* S. 1896 § 4(a)(1)(A).

148. *See id*. § 4(a)(2)(C).

149. *See id*. § 4(a)(2)(A).

150. *See id*.

151. *See id*.

Act.[152] This data would be sorted to convey information about the types of content moderation decisions the platform made, whether these decisions were made by human labor or by algorithms, and the aggregate demographic information of the users who created the content that was subject to content moderation.[153]

2. Disinformation

Disinformation campaigns have interfered in democratic elections and engendered mistrust in democratic institutions and in democracy itself.[154] Such content can incite individuals to harm democratic symbols,[155] commit violent acts, or even participate in genocide.[156] Although disinformation is not an exclusively personalization-driven harm, the harmful effects of disinformation are substantially amplified by platforms' ability to present such content to users who are more susceptible to believing and acting upon it.[157] The spread of disinformation online can also indirectly impact individuals who do not actively participate on digital platforms.[158]

While existing US laws seek to increase transparency by requiring disclosure of the sponsors of political ads on TV, radio, and

---

152. *See id.* § 4(a)(2)(b).

153. *See id.*

154. *See* ROBERT S. MUELLER, III, U.S. DEP'T OF JUST., REPORT ON THE INVESTIGATION INTO RUSSIAN INTERFERENCE IN THE 2016 PRESIDENTIAL ELECTION 4 (2019) [hereinafter MUELLER REPORT].

155. *See* Sheera Frenkel, *The Storming of Capitol Hill Was Organized on Social Media*, N.Y. TIMES (Jan. 6, 2021), https://www.nytimes.com/2021/01/06/us/politics/protesters-storm-capitol-hill-building.html [https://perma.cc/34KX-PDCP].

156. *See* The World Staff, *In Myanmar, Fake News Spread on Facebook Stokes Ethnic Violence*, WORLD (Nov. 1, 2017, 3:15 PM), https://www.pri.org/stories/2017-11-01/myanmar-fake-news-spread-facebook-stokes-ethnic-violence [https://perma.cc/CF9P-63ZS] (describing how fake news posted on Facebook allegedly had a role in facilitating the genocide of Rohingya Muslims in Myanmar); Alexandra Stevenson, *Facebook Admits It Was Used to Incite Violence in Myanmar*, N.Y. TIMES (Nov. 6, 2018), https://www.nytimes.com/2018/11/06/technology/myanmar-facebook.html [https://perma.cc/HM63-CFWA] (reporting that Facebook acknowledged it had a certain role in the events). Fake news was also alleged to have incited violent attacks in Sri Lanka in 2018. *See* Amanda Taub & Max Fisher, *Where Countries Are Tinderboxes and Facebook Is a Match*, N.Y. TIMES (Apr. 21, 2018), https://www.nytimes.com/2018/04/21/world/asia/facebook-sri-lanka-riots.html [https://perma.cc/L9SS-3R3V].

157. *See* Tomer Shadmy, *Content Traffic Regulation: A Democratic Framework for Addressing Misinformation*, 63 JURIMETRICS J. 1, 10–11 (2022).

158. *See Frances Haugen, Written Testimony*, *supra* note 126, at 3 ("Right now, Facebook chooses what information billions of people see, shaping their perception of reality. Even those who don't use Facebook are impacted by the radicalization of people who do. A company with control over our deepest thoughts, feelings and behaviors needs real oversight.").

satellite, such requirements do not apply to ads placed online.[159] Following findings of Russian involvement in the 2016 US presidential election, Senators Warner (D-Va.), Klobuchar (D-Minn.), and Graham (R-S.C.) introduced the Honest Ads Act in order to uphold the Supreme Court's directive in *Buckley v. Valeo* that disclosure should "provide . . . the electorate with information" and "insure [sic] that voters are fully informed" of the identity of who they are listening to.[160] The bill seeks to expand the applicability of the existing disclosure requirements for political ads under the Federal Election Campaign Act of 1971[161] to online media, thereby requiring platforms to accompany political advertisements with a clear statement disclosing who is financing them.[162] In addition, the bill would require platforms to maintain a publicly accessible database disclosing different details about the political ads they host, including, *inter alia*, "a description of the audience targeted by the advertisement."[163] This requirement seeks to establish a collective point of view regarding the ability to detect personalization; however, because it would require disclosure of only the targeting criteria (as collected along the outgoing vector) and not data about the actual presentation of the content (as presented along the incoming vector), it would not enable a third party to detect correlations between outgoing- and incoming-vector content.

The tools employed by the Honest Ads Act and their focus on transparency about political *ads* but not other types of content render it unlikely that the Act will achieve its goal of preventing manipulation of political processes due to the influence other types of content have on elections. For example, the Mueller report found that much of the disinformation spread online in the period leading up to the 2016 US presidential campaign did not appear in the form of ads.[164] Twitter acknowledged, for instance, that approximately 1.4 million Twitter users had been exposed to content generated by almost four thousand Twitter accounts controlled by the Russian Internet Research Agency (IRA)[165] and consequently spread by unsuspecting Twitter users.[166]

---

159. *See The Honest Ads Act*, OFFICE OF U.S. SENATOR MARK. R. WARNER, https://www.warner.senate.gov/public/index.cfm/the-honest-ads-act [https://perma.cc/X3S5-AAYZ] (May 2019).

160. Buckley v. Valeo, 424 U.S. 1, 66–67 (1976); Honest Ads Act, H.R. 4077, 115th Cong. (2017).

161. 52 U.S.C. § 301.

162. H.R. 4077, § 8(a).

163. *Id.*

164. *See* MUELLER REPORT, *supra* note 154, at 14.

165. *See id.* at 15.

166. *See id.* at 27–28.

Similarly, the IRA used fake profiles on Facebook to promote political rallies and to invite reporters to attend these events.[167]

3. Researcher Access

To further achieve meaningful transparency, policy makers have put forward proposals requiring platforms to disclose their data in repositories that researchers could access and analyze. For example, in the European Union, the DSA would require platforms to create a repository of ads presented on platforms' interface, including a copy of the ad itself, information about the targeting criteria used,[168] and aggregate information about the number of users actually presented with the ad (but not information about their personal attributes).[169] The DSA would further require platforms to provide vetted researchers with information that would enable researchers to identify systemic risks, such as discrimination and polarization, created by platform activity.[170] Because the DSA would not require disclosure of key outgoing-vector data about users who saw the ads, however, it cannot properly identify correlations within the data that would enable researchers to detect patterns of personalization—the same flaw that compromised the Honest Ads Act.

Similarly, in the United States, the Social Media DATA Act,[171] sponsored by Representative Trahan (D-Mass.-3), would mandate that platforms provide academic researchers and the FTC with access to all ads placed by advertisers,[172] together with details about their targeting

---

167. *See id.* at 29. Furthermore, public figures and social media influencers may also be involved in spreading political messaging other than political ads, and other types of content may be posted initially for free and then promoted in order to increase the audience size. *See* Anna Reepschlager & Elizabeth Dubois, *New Election Laws Are No Match for the Internet*, POL'Y OPTIONS (Jan. 2, 2019), https://policyoptions.irpp.org/fr/magazines/january-2019/new-election-laws-no-match-internet [https://perma.cc/Y2HA-CX8R].

168. *See Proposal for DSA*, *supra* note 20, at art. 30(1)–(2); *Assessment of the Code of Practice on Disinformation—Achievements and Areas For Further Improvement*, at 5, SWD (2020) 180 final (Sept. 10, 2020) [hereinafter *Disinformation Assessment*] (stating that the European Democracy Action Plan will also regulate the presentation and transparency requirements of political advertising).

169. *See Proposal for DSA*, *supra* note 20, at art. 30(2)(e).

170. *See id.* at art. 26(1); *Disinformation Assessment*, *supra* note 168 (stating that the European Democracy Action Plan will also regulate the presentation and transparency requirements of political advertising).

171. Social Media Disclosure and Transparency (DATA) Act, H.R. 3451, 117th Cong. (2021).

172. *See id.* § 2(a)(1)(B). A similar requirement appears in the DSA. *See Proposal for DSA*, *supra* note 20, at art. 24.

and presentation criteria and about the demographics of their ultimate audience.[173] Another proposal, the Platform Accountability and Transparency Act (PATA), introduced in December 2021 by US Senators Chris Coons (D-Del.), Rob Portman (R-Ohio), and Amy Klobuchar (D-Minn.), takes a somewhat different approach.[174] PATA would enable researchers to submit research proposals to the National Science Foundation, and, upon approval, the relevant platforms would be required to provide the data requested.[175] Additionally, the PATA would enable the FTC to require ongoing transparency about certain data, even if no particular request has been made by researchers.[176]

4. Privacy Protection

In many cases, regulatory proposals mandating transparency are paired with safeguards to protect individual privacy. When platforms are required to disclose data, they often cite privacy concerns as a rationale for denying data requests from third parties.[177] PATA includes provisions requiring researchers to submit their research results to the FTC prior to publication in order to ensure that final research products do not compromise privacy or other confidential business information.[178]

The Social Media DATA Act envisions that the FTC would establish a working group "tasked with providing guidance on how independent research using social media data can be done in a way that protects academic researcher independence and consumer's [sic] rights to privacy"[179]—guidance that would consider "[u]nder what circumstances privacy preserving techniques such as differential privacy and statistical noise could be used."[180] Differential privacy is a mathematical technique that intentionally perturbs

---

173. H.R. 3451, § 2(a)(1)(F). The Algorithmic Justice and Online Platforms Transparency Act, S. 1896, 117th Cong. § 4(c) (2021), also requires that platforms create a library of advertisements including, *inter alia*, the content of the advertisement, the targeting criteria used and information about the identity of the advertiser and the cost of the advertisement.

174. Platform Accountability and Transparency Act, S. 5339, 117th Cong. (2021).

175. *See id.* § 4.

176. *See id.* § 10(e)–(f).

177. *See, e.g.*, Amanda Holpuch, *Airbnb Refuses to Comply with State Order to Hand Over Users' Data*, GUARDIAN (Oct. 8, 2013, 3:20 PM), https://www.theguardian.com/world/2013/oct/08/airbnb-new-york-users-data [https://perma.cc/884V-N4KM].

178. S. 5339 § 5.

179. OFFICE OF CONGRESSWOMAN LORI TRAHAN, FACT SHEET: THE SOCIAL MEDIA DATA ACT OF 2021 2 (2021), https://trahan.house.gov/uploadedfiles/social_media_data_act_two-pager.pdf [https://perma.cc/HRR6-ZAKR].

180. Social Media Disclosure and Transparency (DATA) Act, H.R. 3451, 117th Cong. § 2(c)(4)(C)(ii)(II) (2021).

computations—e.g., by adding a small, random amount of noise to their results—to mask the influence of any single individual's data on the outcome.[181] This Article agrees that this technique is well suited to compile the aggregate statistics necessary to audit for problematic personalization, as Part IV discusses in more detail below.

The proposals described in this Section offer various transparency-increasing mechanisms. The Authors argue that creating a database that includes demographic characteristics collected along the outgoing vector and ad-targeting criteria and information about ad presentation along the incoming vector could enable researchers to generate a meaningful collective perspective. This perspective will allow researchers to more effectively detect cases of unfair treatment or illegal discrimination. In addition, including information about ad sponsorship in such a database would play a critical role in limiting the ability of malicious parties to spread disinformation.

D. Self-Regulation

Many platforms have developed and adopted internal policies of self-regulation to remove, block, or restrict content the platforms deem problematic.[182] At times, they have received criticism for their removal

---

181. *See, e.g.*, Cynthia Dwork, Frank McSherry, Kobbi Nissim & Adam Smith, *Calibrating Noise to Sensitivity in Private Data Analysis*, 3 THEORY CRYPTOGRAPHY CONF. 265, 266 (2006) (introducing the notion of differential privacy); Alexandra Wood, Micah Altman, Aaron Bembenek, Mark Bun, Marco Gaboardi, James Honaker, Kobbi Nissim, David R. O'Brien, Thomas Steinke & Salil Vadhan, *Differential Privacy: A Primer for a Non-Technical Audience*, 21 VAND. J. ENT. & TECH. L. 209, 212 (2018) (introducing the notion of differential privacy to a law audience); Ori Heffetz & Katrina Ligett, *Privacy and Data-Based Research*, 28 J. ECON. PERSP. 75, 82 (2014) (explaining the theory and application of differential privacy to a non-technical audience). As another example, the Data Governance Act lists a few privacy preserving techniques that could be used in data sharing: such as anonymization, pseudonymization, differential privacy, generalization, or suppression and randomization. *See Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act)*, at rec. 6, COM (2020) 767 final (Nov. 25, 2020) [hereinafter *Proposal for DGA*]; *see also* Imana et al., *supra* note 3, at 3771 (proposing a framework for auditing platform algorithms while protecting user information using differential privacy, with results demonstrating that it is "feasible to both audit for fairness and protect user privacy and platforms' business interests").

182. The Authors use the term self-regulation to denote restrictions put in place by platforms themselves, rather than by an external regulator. *See* Molly Cohen & Arun Sundararajan, *Self-Regulation and Innovation in the Peer-to-Peer Sharing Economy*, 82 U. CHI. L. REV. 116, 123–24 (2017). While self-regulation could occur at the exclusive initiative of the self-regulating body, it could also be developed in the shadow of the possibility of external regulation. For example, the DSA encourages the European Commission and the European Board for Digital Services (established under Article 47 of the DSA) to develop voluntary industry standards, codes of conduct, and crisis protocols to be adopted by platforms as part of their self-regulation. *See Proposal for DSA*, *supra* note 20, at arts. 34–37. Various civil society

of content in certain contentious cases[183]—most notably, when Twitter and Facebook decided to block US President Donald Trump from their platforms.[184] Facebook has implemented a third-party fact-checking program[185] that aims to limit the spread of disinformation by reviewing suspicious posts and identifying the source of the disinformation where possible. In addition, Facebook established an oversight board and entrusted it with the authority to make binding decisions about what content Facebook should remove from its platform.[186] There are also examples where platforms have aimed to address incoming-vector harms by introducing new user-facing design features; for example, Instagram recently announced a new tool to encourage its users to "Take a Break" in an effort to address criticisms that the platform is intentionally designed to be addictive.[187]

Other self-regulatory initiatives include voluntary codes of conduct. For example, in 2018, the European Commission signed a Code of Practice on Disinformation together with such leading platforms as Facebook, Google, Twitter, and Mozilla. Microsoft, TikTok, and related advertisers joined soon thereafter.[188] Representing the first time that platforms and advertisers agreed to adhere to self-regulatory standards to fight disinformation online,[189] the Code recognizes the harms caused by the amplification of disinformation and commits to mitigate these

---

organizations have also formulated voluntary codes of conduct that platforms and their workers are encouraged to adopt. For example, the Integrity Institute has developed a Code of Conduct and Integrity Institute Oath for platform workers who are part of the Institute's goal to create "an internet that helps individuals, societies and democracies thrive." *See About Us*, INTEGRITY INST., https://integrityinstitute.org/about-us [https://perma.cc/GZ8B-4FYM] (last visited Apr. 1, 2023). The Oath includes a commitment to put the public first and an acknowledgement that protecting the public is their first job. *See id.*

183. *See, e.g.*, KALINA BONTCHEVA, JULIA POSETTI, DENIS TEYSSOU, TRISHA MEYER, SAM GREGORY, CLARA HANOT & DIANA MAYNARD, BALANCING ACT: COUNTERING DIGITAL DISINFORMATION WHILE RESPECTING FREEDOM OF EXPRESSION (Kalina Bontcheva & Julie Posetti eds., 2020).

184. *See* Mike Isaac & Sheera Frenkel, *Facebook Says Trump's Ban Will Last at Least 2 Years*, N.Y. TIMES, https://www.nytimes.com/2021/06/04/technology/facebook-trump-ban.html [https://perma.cc/GJ7K-SHGY] (June 7, 2021).

185. *See Meta's Third-Party Fact-Checking Program*, META https://www.facebook.com/journalismproject/programs/third-party-fact-checking [https://perma.cc/YMU3-UESP] (last visited Apr. 1, 2023).

186. *See* META OVERSIGHT BD., https://www.oversightboard.com [https://perma.cc/H849-HJA2] (last visited Apr. 1, 2023).

187. *See* Andrew Hutchinson, *Instagram Tests New 'Take A Break' Feature to Encourage Users to Limit Time in the App,* SOCIALMEDIATODAY (Nov. 10, 2021), https://www.socialmediatoday.com/news/instagram-tests-new-take-a-break-feature-to-encourage-users-to-limit-time/609854/ [https://perma.cc/ZW99-3NN2].

188. *See* CODE OF PRACTICE ON DISINFORMATION, *supra* note 20.

189. *See id*.

harms while retaining individuals' freedom of expression.[190] The signatories commit to dilute "the visibility of disinformation"[191] by providing users with tools to customize their own content, discover content, and "find diverse perspectives about topics of public interest."[192] In line with the mechanism proposed in the DSA,[193] the Code requires that platforms provide their users with tools to report content they believe to be disinformation,[194] as well as an explanation as to why users have been presented with particular content.[195] It also recognizes that technology will be an integral part of overcoming disinformation and requires parties to invest in technological solutions that will prioritize "relevant, authentic[,] and authoritative information."[196]

Following in the footsteps of the Commission, in July 2019, the Australian government published a report offering twenty-three recommendations "to promote competition, enhance consumer protection[,] and support a sustainable Australian media landscape in the digital age,"[197] some of which encouraged platforms to develop a voluntary code of conduct on disinformation.[198] This resulted in the Australian Code of Practice on Disinformation and Misinformation, which launched in February 2021 and was adopted by leading platforms such as Apple, Facebook, Google, Microsoft, TikTok, and Twitter.[199] In an attempt to encourage self-regulation, the EU Artificial Intelligence Act has petitioned EU member states to develop voluntary codes of conduct that hold platforms accountable to content policies broader than those the Act strictly requires.[200]

The first-year assessment of the European Code of Practice on Disinformation found that it served as an important basis for dialogue among stakeholders and provided transparency into platforms' policies

---

190. *Id.* at art. I.

191. *See id.* at art. I, sec. ix.

192. *See id.* at art. II.D.

193. *See Proposal for DSA*, *supra* note 20, at art. 14.

194. *See* CODE OF PRACTICE ON DISINFORMATION, *supra* note 20, at art. I, sec. x.

195. *See id.* at art. II.D.

196. *See id.*

197. AUSTL. GOV'T, REGULATING IN THE DIGITAL AGE: GOVERNMENT RESPONSE AND IMPLEMENTATION ROADMAP FOR THE DIGITAL PLATFORMS INQUIRY 3 (2019).

198. *See id.* at 12.

199. *See About the Code*, DIGIT. INDUS. GRP. INC., https://digi.org.au/disinformation-code/ [https://perma.cc/4U8W-UJGG] (last visited Apr. 3, 2023).

200. *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*, at art. 69, COM (2021) 206 final (Apr. 21, 2021).

on disinformation.[201] However, the assessment recognized that a substantial shortcoming of the Code involves a lack of access to platform data, preventing third parties from conducting their own independent evaluation of "emerging trends and threats posed by online disinformation."[202] In May 2021, the signatories to the Australian Code submitted their first transparency reports. These reports largely detailed that the platforms' policy framework conformed with the Code's requirements.[203]

While platforms' self-regulatory efforts may be a complementary step in the right direction, as detailed in the next Section, this type of effort involved very little external oversight and has been criticized as "little more than a symbolic activity."[204] Creating a transparency-increasing mechanism, which would enable third parties to observe platforms' behavior and track their adherence to the standards created by self-regulation, can help alleviate these concerns, as Part IV describes in more detail below.

E. Technical Approaches

Several recent projects looking to address the harms of the data ecosystem take a strongly control-driven perspective, seeking to keep each individual's data in a location controlled by that person and allowing software under their personal control to dictate whether outside platforms and apps would gain access to their data.[205] While there is both a role and a need for better control of data, such an individualistic perspective misses the nuances of the fundamentally collective nature of data and thus cannot meaningfully intervene to prevent incoming-vector harms.

Francis Fukuyama and other scholars have recently begun to explore a structural intervention they call "middleware": software that would enable users to choose the type of content they want to see, what order they would want to see it in, and the sources they trust to present

---

201. *See Disinformation Assessment*, *supra* note 168.

202. *See id.* at 19.

203. *See Transparency Reports*, DIGIT. INDUS. GRP. INC., https://digi.org.au/disinformation-code/transparency/ [https://perma.cc/BWM2-AJM9] (last visited Apr. 3, 2023).

204. John Braithwaite & Brent Fisse, *Self-Regulation and the Control of Corporate Crime*, 23 PRIV. POLICING 221, 224 (1987).

205. *See, e.g.*, SOLID, https://solidproject.org [https://perma.cc/RXE4-JZU2] (last visited Apr. 1, 2023) (explaining that the project enables individuals to "store their data securely in decentralized data stores called Pods . . . [enabling the individual to] control which people and applications can access it").

them with such content.[206] The Middleware proposal would dilute the power that platforms currently have over public and political discourse. The proposal is technologically situated to minimize friction with the existing ecosystem.[207] However, it is not clear how individual preferences would interact with platform-driven content promotion or personalization in the Middleware model. Furthermore, as framed, the Middleware proposal does not seek to provide insight into patterns of personalization or their impacts.

A handful of recent technical projects have explicitly aimed to make personalization along the incoming vector more transparent. Several carefully constructed studies have analyzed incoming- and outgoing-vector data at a fixed point in time to reveal instances of problematic, discriminatory presentation of advertising content by platforms.[208] Because no ongoing infrastructure exists for collecting such data, however, these studies are limited by the time and effort it takes to conduct them, and they can only provide insight into one isolated issue at a particular point in time.

Another approach, the Mozilla Rally project,[209] allows individual users of the Mozilla Firefox web browser to sign up to volunteer information about themselves (i.e., outgoing-vector content such as demographic characteristics or answers to surveys), allow Mozilla to gather content related to their browsing (such as the URLs of the pages they browse, page content, and how much time they spend on each page), and opt in to allow preapproved research projects access to their

206. *See* FRANCIS FUKUYAMA, BARAK RICHMAN, ASHISH GOEL, ROBERTA R. KATZ, A. DOUGLAS MELAMED & MARIETJE SCHAAKE, MIDDLEWARE FOR DOMINANT DIGITAL PLATFORMS: A TECHNOLOGICAL SOLUTION TO A THREAT TO DEMOCRACY, 2, 6 (2020) ("Middleware's primary benefit is that it dilutes the enormous control that dominant platforms have."); Francis Fukuyama, *Making the Internet Safe for Democracy*, 32 J. DEMOCRACY 37, 43 (2021) ("[Large platforms] possess not only enormous wealth . . . but also something of a chokehold over the communications channels that facilitate democratic politics.").

207. *See* Fukuyama, *supra* note 206, at 43.

208. *See* sources cited *supra* note 26; Ali et al., *supra* note 18; *see also* Joshua Asplund, Motahhare Eslami, Hari Sundaram, Christian Sandvig & Karrie Karahalios, *Auditing Race and Gender Discrimination in Online Housing Markets*, 14 PROC. INT'L ASS'N ADVANCEMENT A.I. CONF. ON WEB & SOC. MEDIA 24, 25 (2020) (demonstrating differential treatment in the presentation of housing ads and property recommendations based on users' race and gender); Anja Lambrecht & Catherine Tucker, *Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads*, 65 MGMT. SCI. 2966, 2966 (2018) (finding that ads promoting job opportunities in the science, technology, engineering, and math fields were presented less often to women, who constitute a prized demographic, and thus a more expensive target-audience for ads. An algorithm that simply optimizes cost effectiveness in ad delivery may deliver ads in an apparently discriminatory way, even if the ads were intended to be gender neutral).

209. *See* MOZILLA RALLY, https://rally.mozilla.org [https://perma.cc/WDQ7-E5KF] (last visited Apr. 1, 2023).

relevant data.[210] This effort, if widely adopted, could potentially provide broad, meaningful transparency into platform personalization from a collective perspective, due to its access to both incoming- and outgoing-vector data. One downside is the project's lack of formal privacy guarantees for the sensitive data that it gathers. However, the high-level idea offers an incredibly promising model for future development.

In summary, this Article finds that many current approaches seeking to overcome incoming-vector harms adopt an individualistic approach. This finding is consistent with many scholars' observations that privacy and data protection have traditionally been conceptualized as individual rights,[211] largely focused on individuals' ability to control the flow of their data through the data ecosystem.[212]

This framing of data, however, ignores the current reality, in which the process of datafication creates unjust results on a social level.[213] Platforms collect and analyze massive amounts of data from millions of individuals to personalize content effectively, a process that subordinates and manipulates the individual and generates collective harm.[214] An individual acting on her own cannot counteract either end of this problem: she alone cannot effectively withhold her data along the outgoing vector, and she cannot effectively protect herself from the harms of incoming-vector personalization.

IV. RECOMMENDED DESIGN PRINCIPLES FOR EFFECTIVE INCOMING-VECTOR INTERVENTIONS

The previous Parts describe the collective nature of data[215] and discuss how outgoing-vector content provided by one individual can serve (along with the data of many others) to personalize

210. *Take Control Over Your Data with Rally, A Novel Privacy-First Data Sharing Platform*, MOZILLA RALLY: DISTILLED (June 25, 2021), https://blog.mozilla.org/en/mozilla/take-control-over-your-data-with-rally-a-novel-privacy-first-data-sharing-platform [https://perma.cc/53AN-9XWM].

211. *See* Alessandro Mantelero, *From Group Privacy to Collective Privacy: Towards a New Dimension of Privacy and Data Protection in the Big Data Era*, in GROUP PRIVACY 139 (Linnet Taylor, Luciano Floridi & Bart van der Sloot eds., 2017) (noting that "informational privacy and data protection have been protected as individual rights").

212. *See, e.g.*, Viljoen, *supra* note 7, at 593. This approach is exemplified by the fair information privacy principles, which have strongly influenced the development of privacy and data protection frameworks in the United States and European Union and around the world. *Id.*

213. *See id.* at 617.

214. *See id.* at 631.

215. *See supra* Section II.B.

incoming-vector content for other users.[216] Additionally, the previous Parts of this Article survey various approaches to counter the many harms of incoming-vector personalization, evaluating their strengths and weaknesses using the lens of the collective nature of data, and finding that, despite the strengths of certain proposals, the general principles driving many approaches are highly individual-centric.[217]

This Part proposes an alternative path forward for addressing the harms of personalization. In particular, it argues the need for a form of transparency that the Authors refer to as a *collective perspective*: transparency that allows visibility into correlations between the incoming and outgoing vectors with respect to a large number of people.

Non-platform actors cannot effectively overcome the harms of platform personalization without meaningful transparency; society cannot properly understand the role that personalization plays in generating or amplifying various harms without it. At present, there is a lack of clarity regarding even the most basic of questions, such as whether platform personalization contributes to polarization or defuses it.[218] Furthermore, at present, it is nearly impossible to detect or measure patterns of personalization.

A. What Information Is Needed to Achieve Meaningful, Effective Transparency?

First, due to the collective nature of data, meaningful transparency must provide visibility into the personalized content presented to a *large number* of individuals, not just one or a handful. Indeed, it is only possible to define some of the harms that may be induced by incoming-vector personalization within a broader social context.

For example, if Jane were the only person using a service, it might be categorically impossible for the service to provide Jane with polarizing or discriminatory content because there would be no other users with whom Jane could be contrasted or compared. More crucially, though, given any definition of what constitutes problematic personalization (such as illegal discrimination), the data of only a single person or a small number of people cannot generally be used to

216. *See supra* Section II.A.

217. *See supra* Part III.

218. *See* Bail et al., *supra* note 16, at 2916.

determine the presence or extent of a problem.[219] For example, if one wished to show that a platform had displayed a particular ad for housing in a manner that disproportionately excluded Black individuals, she would need to do more than simply observe that the platforms had displayed the ad to a particular White person or not shown it to a particular Black person.

Instead, one would need to know the rate of display on a representative sample of the relevant White and Black populations, and one would need enough observations such that measured differences in the rate of display would be statistically significant. Similarly, one may want to detect the presentation of misleading, polarizing, or incendiary content.[220] To detect such personalization, one would need to analyze a broad sample of individuals' incoming-vector content.

The precise number of people's perspectives needed in order to detect patterns of problematic personalization depends on several parameters—for example, the number of types of problematic personalization one wishes to audit for, the size of the population one wishes to study, the prevalence of the problematic phenomenon, and the severity of the phenomenon one wishes to detect.

Thus, when there are more questions to be studied, one must increase the number of observations in order to maintain the statistical validity of the conclusions. If one wishes to detect discrimination against a tiny group, it may be difficult to get enough observations of that group. Finally, fewer observations are necessary to detect cases of extreme discrimination compared to the large number of observations that would be necessary to detect subtle discrimination. In practice, the

---

219. CHRISTIAN SANDVIG, KEVIN HAMILTON, KARRIE KARAHALIOS & CEDRIC LANGBORT, AUDITING ALGORITHMS: RESEARCH METHODS FOR DETECTING DISCRIMINATION ON INTERNET PLATFORMS, 1, 6 (2014), http://social.cs.uiuc.edu/papers/pdfs/ICA2014-Sandvig.pdf [https://perma.cc/G6W5-GAC8] (proposing that "normative concerns that have been raised involving algorithmic discrimination . . . demand an audit of online platforms," meaning "a program of research should be undertaken to audit important Internet-based intermediaries with large data repositories (e.g., YouTube, Google, Facebook, Netflix, and so on) to ascertain whether they are conducting harmful discrimination by class, race, gender, and to investigate the operation of their algorithms consequences on other normative concerns"); Mathias Lecuver, Riley Spahn, Yannis Spiliopolous, Augustin Chaintreau, Roxana Geambasu & Daniel Hsu, *Sunlight: Fine-Grained Targeting Detection as Scale with Statistical Confidence*, 22 PROC. ASS'N COMPUTING MACH. SPECIAL INT. GRP. SEC. AUDIT & CONTROL CONF. COMP. & COMM. SEC. 554, 556 (2014) (detailing the shortcomings of past experiments checking, for example, discriminatory pricing and advertising, detailing that in order to generate meaningful input, "*experiments must be run at large scale*" and presenting a system satisfying this requirement); Imana et. al, *supra* note 27, at 2–3 (observing that many "types of harm can be invisible to end-users and require systematic study by experts to detect" and proposing a new framework for platform-supported auditing that provides researchers with access to platform data at scale while protecting privacy).

220. *See* Soroush Vosoughi, Deb Roy & Sinan Aral, *The Spread of True and False News Online*, 359 SCI. 1146 (2018) (finding that false news stories spread faster than true ones).

actual number of individuals needed to form a useful collective perspective could range from the dozens to the tens or hundreds of thousands.[221]

Second, meaningful transparency must expose patterns and correlations that *relate outgoing-vector content* (such as individual characteristics and actions taken) *to incoming-vector content* at an aggregate level. Visibility into only incoming-vector content could reveal that a certain piece of content was or was not displayed and how many times, but it would be blind to how the decision to present content was *personalized*. The individual characteristics and behaviors revealed along the outgoing vector—potentially indicating each individual's age, gender, location, race, religion, political affiliation, income, occupation, medical history, and more—form the basis of such personalization.[222] Hence, the ability to relate the outgoing vector to the incoming vector is a crucial component of meaningful transparency.[223]

It is also important to be able to determine the source used for personalizing incoming-vector content. For example, a rule could restrict the data that platforms use to personalize content along the incoming vector. This could be done, for example, by permitting personalization based on data explicitly provided by the user but prohibiting it based on inferred characteristics.[224] If one wished to detect a violation of such a rule, one would need the ability to determine the *source of incoming-vector information*.

Insights into incoming-vector personalization must also clearly be *ongoing*—that is, they cannot form meaningful conclusions from one-off measures from any one point in time, as personalization algorithms and their content (and hence their harms) are constantly changing and evolving. Furthermore, some concerns, such as platforms' promoting *increasingly* polarized content, have an inherent longitudinal aspect. Detection and analysis of such trends require a collective perspective.

---

221. Existing experimental studies, such as those mentioned *supra* note 26 provide some insight into the size of cohorts that have been required to detect specific instances of problematic personalization. See, for an example, Ali et al., *supra* note 26, which used a cohort of tens of thousands to hundreds of thousands of participants.

222. For the definition of *outgoing vector*, see discussion *supra* Section II.A.

223. JOSHUA A. TUCKER, ANDREW GUESS, PABLO BARBERÁ, CRISTIAN VACCARI, ALEXANDRA SIEGEL, SERGEY SANOVICH, DENIS STUKAL & BRENDAN NYHAN, SOCIAL MEDIA, POLITICAL POLARIZATION, AND POLITICAL DISINFORMATION: A REVIEW OF THE SCIENTIFIC LITERATURE 64 (2018) (reviewing current literature that analyzes the relationship between social media, political polarization, and disinformation).

224. *See* Wachter & Mittelstadt, *supra* note 22, at 610.

In sum, meaningful transparency requires far more than disclosing ad-targeting criteria or ad-funding details,[225] creating databases of ads divorced from the actual outgoing-vector data of those who received them,[226] or focusing primarily on ads.[227] To be effective, transparency with respect to algorithmic personalization must constitute a genuine collective perspective. Such a perspective must be based on ongoing insights into the information users provide and platforms observe along the outgoing vector and on how that information correlates with personalized content that a large, representative population receives along the incoming vector.

B. What Body Could Be Tasked with Establishing a Collective Perspective?

Currently, platforms are the only actors in the data ecosystem privy to the full range of incoming- and outgoing-vector content, enabling them to hold a collective perspective. However, past analyses, such as that of Lina Khan and David Pozen, suggest that the incentives of platforms are so misaligned with those of individual users and the public at large that platforms should not and cannot be assigned sole responsibility for detecting, measuring, and mitigating the harms inflicted by the personalized content they purvey.[228] It is therefore worth exploring alternative bodies that could be entrusted with the collective perspective.[229]

There are at least two senses in which policy makers would need to trust a third party with a collective perspective to detect and measure personalization-driven harms. First, policy makers would need to trust the body to carry out its duties of observation in the best interest of platform users in particular and of society more generally. Second, since its analyses could pertain to quite sensitive information about some

225. Honest Ads Act, H.R. 4077, 115th Cong. (2017).

226. DSA, *supra* note 97.

227. Social Media Disclosure and Transparency (DATA) Act, H.R. 3451, 117th Cong. (2021).

228. Lina M. Kahn & David E. Pozen, *A Skeptical View of Information Fiduciaries*, 133 HARV. L. REV. 497, 498 (2019); *see also Frances Haugen, Written Testimony*, *supra* note 126 ("I saw Facebook repeatedly encounter conflicts between its own profits and our safety. Facebook consistently resolves these conflicts in favor of its own profits."); Nathaniel Persily, *Facebook Hides Data Showing It Harms Users. Outside Scholars Need Access*, WASH. POST (Oct. 5, 2021, 7:20 AM), https://www.washingtonpost.com/outlook/2021/10/05/facebook-research-data-haugen-congress-regulation [https://perma.cc/L3EN-47X7].

229. *See* Margot Kaminski, *Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability*, 92 S. CAL. L. REV. 1529, 1533 (2019) ("Collaborative governance is described, in brief, as a better way to govern fast-changing, risky systems with a high degree of technological complexity.").

individuals, the third party would need to maintain the privacy of the users involved. Fortunately, meaningful transparency often does not require direct access to individuals' sensitive information, but just to statistical aggregates that can be computed with small, intentional perturbations in order to provide formal privacy guarantees. To address privacy concerns, the body could employ technological tools for collecting aggregate, privacy-preserving measurements (such as metrics quantifying gender disparity in the delivery of a certain type of ad), thereby avoiding the need to access raw data containing individuals' demographic information or information about the personalized content they are shown.

The Social Media DATA Act offers a useful framework for future development in this area, as it recognizes the potential of modern technology to resolve apparent conflicts between transparency and privacy.[230] A body holding a collective perspective can use local differential privacy to add noise to personal data before it is collected so that the individual-level data is close to meaningless, but the aggregate-level data (such as the level of correlation between an ad being shown and the race of the viewer) can still serve as a basis for generating meaningful insight.[231] Secure multiparty computation additionally provides a modern cryptographic toolkit that can remove the need to entrust one monolithic body with correctly and safely monitoring platforms for harmful patterns of personalization.[232] In the secure multiparty computation model, a few trusted parties share responsibility for carrying out the necessary computations, and cryptographic guarantees ensure that no small coalition of these parties would be able to corrupt the computation or gain inappropriate access to personal information. Of course, regulatory and contractual safeguards could provide an additional layer of protection.

A third party could source the information needed to establish the collective perspective through several possible models, including, potentially, by directly intermediating between individuals and platforms, receiving information primarily from individuals, or receiving information primarily from platforms. As one example, when the FTC asks a platform to disclose all relevant data as part of an investigation of possible unfair or deceptive trade practices, it could require that such disclosure be sufficient to generate a collective

230. H.R. 3451 § 2.

231. *See* sources cited *supra* note 181.

232. *See generally* Yehuda Lindell, *Secure Multiparty Computation*, INT'L ASS'N CRYPTOLOGIC RSCH., no. 300, 2020 (providing an accessible but detailed introduction to the guarantees that secure multiparty computation provides).

perspective. Additionally, in interpreting the DMA's requirement that platforms inform the European Commission of their profiling techniques, the European Commission could require a disclosure of data that provides a collective perspective. Similarly, the DSA's mandate that platforms make available public repositories of the online ads they display could also require the publication of additional information necessary for a collective perspective. This approach could also enhance various legislative proposals like the DETOUR Act, which seeks to establish an Independent Review Board (IRB) that would be responsible for approving platform-run experiments. With a view of the collective perspective, the IRB could incorporate a review of the consequences for groups and society, not just for the individuals directly affected.

The governance, funding, and structure of the entity could also take a range of forms, from a government body to a private for-profit or nonprofit service heavily regulated by law. In addition, as Section IV.D details below, there are several possible entities that might receive access to the insights afforded by the collective perspective for enforcement purposes.

The proposed EU Data Governance Act (DGA) provides one useful model for establishing trustworthy intermediating bodies.[233] The proposal would require "data intermediaries" to maintain neutrality and use individuals' data solely for the purpose of promoting its lawful exchange.[234] Intermediaries' business models must "assure that there are no misaligned incentives that encourage individuals to make more data available for processing than what is in the individuals' own interest."[235] Furthermore, intermediaries would owe a fiduciary duty to those data holders whose data-sharing they facilitate.[236] The DGA would also recognize data cooperatives—entities that would assist users in making informed and meaningful choices over their data and its sharing, *inter alia,* by enabling "mechanisms to exchange views on data processing" that would best represent members' interests.[237] Such a body may potentially be positioned to establish the needed collective perspective.

233. *See Proposal for DGA, supra* note 181; Thomas Streinz, *The Evolution of European Data Law, in* THE EVOLUTION OF EU LAW 902, 935 (Paul Craig & Gráinne de Búrca eds., 2021).

234. *Proposal for DGA*, *supra* note 181, at art. 11(1).

235. *Id.* at rec. 23.

236. *See id.* at rec. 26.

237. *Id*. at art. 9 (1)(c).

C. How Can Regulation Support the Establishment of the Necessary Collective Perspective?

Legislation must take an active role in establishing or identifying an intermediating body that will establish the collective perspective. In order to ensure that such a body enjoys the trust of the public, legislation must also ensure that it will have unencumbered access to the information that it needs, that it establishes mechanisms for the harms that come to light, and that it provides enforcement mechanisms against those harms.

Regulation should helpfully tie the hands of the intermediating body. It should restrict the body's ability to share any data and any derivatives of that data to which it receives access (whether for profit or not), and it should mandate the use of modern cryptographic and statistical techniques (as discussed above in Section IV.B) to minimize the exposure and gathering of sensitive data.

Regulatory intervention will also likely be necessary in order to oblige platforms to cooperate with the monitoring and data collection required in order to establish the collective perspective. This is in line with—although more demanding than—the various transparency mandates currently under discussion, as mentioned in Section III.C.

Legislation must also support the intermediating body in gaining access to the information it needs. For example, one might consider laws allowing users to install software that enables a third party to collect information about users' interaction with a platform.[238] Such regulation would facilitate direct, non-intermediated access to user data. Currently, platforms restrict users' ability to share content outside the platform in their terms of service and do not allow third parties to scrape content from the platform.[239] Indeed, Facebook has filed lawsuits against individuals and organizations that scraped content from its platforms in violation of its terms of service.[240] Care must be taken to ensure that the intermediating body does not use privacy and security concerns (whether real or fictional) and corresponding legislation, including the Computer Fraud and Abuse Act,[241] as an excuse to sabotage its own effectiveness.

---

238. A similar proposal appears in the Platform Accountability and Transparency Act. S. 5339, 117th Cong. (2021)

239. *See* Persily, *supra* note 228.

240. *See* Jessica Romero, *Combating Scraping by Malicious Browser Extensions***,** META (Jan. 14, 2021), https://about.fb.com/news/2021/01/combating-scraping-by-malicious-browser-extensions [https://perma.cc/5V4T-NGY6].

241. 18 U.S.C. § 1030.

Legislation should also determine who would have the right to query or access the collective perspective. Analogous to the approach taken by PATA,[242] one possible model would provide academic researchers—who are subject to oversight by an institutional board and have applied for and received approval to carry out studies on the data—with the right to interrogate whatever body holds the collective perspective. Academic researchers who discover cases of harmful personalization could share their research findings with the appropriate oversight body to initiate potential investigatory and enforcement actions. Alternatively, or additionally, access to the collective perspective could be made available to journalists for investigative reporting purposes. An advantage of either of these first two models is that granting academic researchers and journalists access to the collective perspective opens up the possibility of identifying instances of newly emerging informational harms that are problematic but permissible under existing law. A third model would involve making the collective perspective directly available to a government agency with investigation and enforcement authority, such as the US FTC, the US Department of Housing and Urban Development, or the US Equal Employment Opportunity Commission (EEOC). In cases where such a body identified instances of illegal personalization, it could file a complaint, as the EEOC did when it alleged that Facebook facilitated the discriminatory presentation of job ads.[243] Finally, an independent, cooperative entity, such as a data cooperative or data trust, could be established with the explicit purpose of monitoring platforms for unacceptable personalization.

D. What Is the Expected Impact of the Collective Perspective?

The collective perspective, once established, would shed light on the mechanisms by which personalization is contributing to known harms, enable quantification of the severity of harms, and potentially also draw attention to previously unrecognized personalization-driven harms. This would ultimately provide a basis for informed discourse

242. S. 5339 § 5.

243. In 2019, the EEOC found that seven employers had violated federal law when advertising jobs on Facebook in a way that excluded women and older workers from getting the ads. *In Historic Decision on Digital Bias, EEOC Finds Employers Violated Federal Law when they Excluded Women and Older Workers from Facebook Ads*, *supra* note 18 (reporting on the decision). Additionally, the Fair Housing Act, 42 U.S.C. § 804, prohibits discrimination in advertising for housing opportunities. This section served as the basis for the US Department of Housing and Urban Development's charge of discrimination against Facebook in 2019, alleging discrimination in the presentation of ads for housing on the platform. *See* Charge of Discrimination, U.S. Dep't Hous. & Urb. Dev. v. Facebook, Inc., FHEO No. 01-18-0323-8 (2019).

among academics, policy makers, and society at large, enabling them to grapple with myriad questions such as: How severe is the discrimination in digital advertising of housing opportunities, and what role does platform personalization play? Does personalization on the basis of inferred characteristics contribute more to the amplification of misinformation than personalization on the basis of characteristics a user has explicitly provided for the purpose of content-tailoring? How significant is the contribution of algorithmic personalization to the rapid spread of incendiary content?

Establishing a collective perspective would foster public debates over the character of appropriate interventions to address personalization-driven harms. In some cases, legislation could establish enforcement mechanisms against problematic personalization, such as requirements for flagging, deprioritizing, or blocking content that reflects such personalization, backed by civil or even criminal penalties for platforms. In other cases, the collective perspective could provide evidence that federal agencies can use to enforce existing laws. The third party could potentially wield existing enforcement authority—or share insights with an enforcement body with such authority—in order to respond to harms detected using the collective perspective. For example, the FTC could initiate an enforcement action against a company engaging in unfair or deceptive trade practices based on evidence from the collective perspective.[244] The collective perspective could also provide evidence for enforcement under various regulations seeking to moderate certain potentially harmful aspects of platform activity.[245] Beyond regulatory enforcement, the collective perspective could inform the work of bodies tasked with developing norms of content moderation, such as Susan Benesch's proposal for creating local independent councils that would set "ethical standards specific to the online distribution of content and cover topics such as terms and conditions, community guidelines, and the content regulation practices of social media companies."[246]

One could also promote adherence to norms regarding personalization by providing measurements derived from the collective perspective to individual users, regulators, or the public. These norms

244. For a discussion outlining examples of FTC investigations and enforcement actions with respect to incoming-vector harms, see *supra* Section III.A.

245. For example, the DSA requires member states to establish national Digital Service Coordinators to be in charge of "application and enforcement" of the DSA. *See* DSA, *supra* note 97, art. 38. As part of ensuring these national bodies are in a position to effectively carry out their supervisory role, they are granted broad authority to request access to necessary data from platforms. *See id.* art. 41.

246. Benesch, *supra* note 139, at 19.

could take the form of regulatory standards, but they could also constitute community norms adopted by individuals who wish to adhere to certain standards—even those that go beyond the legal requirements. For example, a group of users may not want to see content that has been personalized based on their political position, or they may not want to be gender stereotyped in the personalized content presented to them. Meaningful transparency into algorithmic personalization could give people the power to pressure platforms to live up to the desired standards of their users.

In summary, legislators must intervene to establish a collective perspective to enable society to collectively understand, detect, study, quantify, and respond to problematic personalization. Without such intervention, harmful personalization will continue to harm individuals and society, unchecked and largely unobserved.

V. CONCLUSION

This Article offers an analysis of the structure of the data ecosystem and the incentives that shape it. It identifies the importance of and the relationship between the outgoing vector and the incoming vector and offers terminology that enables researchers to discuss flows of information between platforms and individuals.[247] This terminology provides not only a framework for describing these two data flows, but also the ability to analytically evaluate the various challenges and opportunities presented by each.[248] In surveying existing and proposed regulatory and technological approaches designed to address the harms stemming from incoming-vector personalization, this Article finds that nearly all of these approaches will likely be ineffective in their ability to combat incoming-vector harms, and it demonstrates that the lack of sufficient recognition of the collective nature of data explains this failure.[249]

Finally, this Article offers a path forward that involves a radical new level of transparency around platform personalization.[250] In particular, it calls for a specific form of transparency—a *collective perspective*—that affords continuous visibility into correlations between the incoming and outgoing vectors across a large, representative population.[251] The establishment of this collective perspective would provide a basis for society to better understand the harms of the

247. *See supra* Part II.
248. *See supra* Part II.
249. *See supra* Part III.
250. *See supra* Part IV.
251. *See supra* Part IV.

incoming vector and the impact of these harms on society.[252] Such visibility would transform society's ability to develop regulations, enforcement mechanisms, and other interventions to address platform-driven harms and promote a more just data ecosystem.[253]

252. *See supra* Part IV.

253. *See supra* Part IV.