Algorithmic Personalization Features and Democratic Values: What Regulation Initiatives Are Missing

By Sharon Bassan*

2024 was poised to be the largest election year in history, with pivotal elections in Asia, Europe, and the Americas encompassing regional, legislative, and presidential contests, capturing the attention of half the globe.[1] *In an era dominated by social media, these elections were influenced by information dissemination through digital platforms.*[2]

Over the last two decades, the landscape of public discourse in matters of civic concern has undergone a transformative shift, moving from traditional media to personalized digital social media outlets. Algorithmic features now selectively match content to users, fostering engagement but also giving rise to issues such as echo chambers, filter bubbles, and sensationalized content. While much attention has been devoted to the challenges posed by personalized discourse, this paper sheds light on a critical aspect that has been overlooked in regulatory paradigms: the erosion of an open, public sphere for discourse due to individualized manipulated content matching.

In a well-functioning democratic system, an informed citizenry is paramount. However, amplification algorithms and recommendation systems—referred to as "personalization features" here—employ manipulated partial and even contradicting messages designed to targeted audience, while excluding users who may object, correct or protest against them. This paper aims to address the question of whether democratic public discourse can be preserved amidst the individualized flow of manipulated content on social media.

Existing regulatory approaches—including content regulation, algorithm regulation, and privacy regulation—primarily focus on data profiling, algorithmic content matching, and the nature of discourse within personalized spheres. They prove inadequate in mitigating harm to the public sphere resulting from the selective distribution of individualized manipulated content, hindering open public discussion. Even regulatory initiatives specifically targeting accessible public discourse fall short in addressing both the individualized and manipulated aspects of personalized speech.

---

* In the memory of Alex Geisinger, a mentor, a friend, an inspiration.

Dr. Bassan serves as the Head of Innovation Law, Policy and Ethical Governance at The Louis Brandeis Institute for Society, Economy, and Democracy.

1. Koh Ewe, *The Ultimate Election Year: All the Elections Around the World in 2024*, TIME (Dec. 28, 2023, 12:00 AM), https://time.com/6550920/world-elections-2024/ [perma.cc/4MDT-3NDC].

2. Robert Hart, *Here's How AI and Tech Giants Say They'll Moderate Content in 2024—The World's Biggest Election Year*, FORBES (Jan. 17, 2024, 09:06 AM), https://www.forbes.com/sites/roberthart/2024/01/17/heres-how-ai-and-tech-giants-say-theyll-moderate-content-in-2024-the-worlds-biggest-election-year/?sh=76b942b54280 [perma.cc/6EQ3-ZQD9].

The research findings highlight a pressing concern—the inherent contradiction between the individualized manipulated flow of content and the principles of democratic deliberation. Urgent regulatory frameworks are needed to safeguard a public space for deliberation that is accessible to all, facilitating joint decisions and the construction of pluralistic democratic societies. Potential solutions may involve establishing alternative spaces for public discourse and recognizing distinct considerations for issues in the public sphere. Additionally, a multifaceted strategy beyond online discourse or algorithm regulation is proposed, aiming to foster constructive dialogue, respect, and tolerance in the digital ecosystem, complemented by public education.

I. SOCIAL MEDIA AND POWER OVER PUBLIC DISCOURSE

According to most democratic theories, deliberation stands as the bedrock of democratic values.[3] Elections gain democratic legitimacy based on the assumption that voting stems from a robust deliberative process.[4] Political and other public affairs must be informed by an ongoing background of public discourse, serving as the pathway for individuals to uncover truth and to make educated choices about representatives and their policies through exchange and consideration of diverse viewpoints.[5] This deliberative process is seen as crucial for fostering a marketplace

3. Giles Howdle, *Microtargeting, Dogwhistles, and Deliberative Democracy*, 42 TOPOI 445, 449 (2023) (distinguishing between "'vote-centric' or 'aggregative' and . . . 'talk-centric' or 'deliberative' theories of democracy. According to the former . . . democracy is to be understood almost entirely in terms of voting. This conception of democracy has its roots in Rousseau's idea that the results of democratic processes reflects [sic] a 'general will' . . . . [A]ccording to this conception, [voting] is a necessary and sufficient condition for its result to possess democratic legitimacy—a democratic mandate . . . Deliberative Democracy. A voting process confers democratic legitimacy on its results only if the voting body is provided sufficient opportunity to engage in public deliberation of an adequate standard.").

4. *Id.*

5. FORUM ON INFORMATION & DEMOCRACY, PLURALISM OF NEWS AND INFORMATION IN CURATION AND INDEXING ALGORITHMS 49 (2023) [hereinafter PLURALISM OF NEWS], https://informationdemocracy.org/wp-content/uploads/2023/02/Report_Pluralism-in-algorithms.pdf [perma.cc/7WUM-XD7U].

of ideas where the best arguments can prevail.[6] The public sphere has the potential to provide a forum for citizens to engage in discussions about political and social issues, exchange diverse perspectives, test and refine ideas, and ultimately contribute to a more robust democracy. It is within this public sphere that citizens can come together to deliberate, where narratives could compete against one another for acceptance, ultimately leading to the "truth."[7] Through this open exchange, societies can advance knowledge, uncover new insights and effective solutions, expose falsehoods, hold their governments accountable, make fair and just decisions, and work towards more inclusive and equitable frameworks.[8]

In order for diverse citizens to coexist, civil society needs to commit to a pluralistic, open-to-all public discourse. The exchange of ideas and opinions with others in the community who may not share one's beliefs is an important mode of active participation in public and political life. A well-functioning public sphere fosters the development of common values, promotes social cohesion by fostering a sense of belonging, and accommodates the diversity inherent in democratic life.[9]

Open dialogue and debate ensure that a well-informed citizenry can actively participate in shaping their collective future, guided by a commitment to upholding common values.[10] This exposure equips individuals with the tools to subject their own traditions to rigorous and rational examination, fostering a climate of mutual respect for alternative ways of life and a receptivity to different perspectives of truth.[11] When individuals from diverse backgrounds actively participate in open and respectful dialogues, they play a pivotal role in bridging societal divides and nurturing a sense of community and tolerance. This participation facilitates the resolution of political conflicts and promotes the stability of institutional frameworks.[12] Moreover, in diverse societies, pluralistic discourse helps protect the rights of minority groups by (1) ensuring that minority voices are heard and

---

6. McCullen v. Coakley, 573 U.S. 464, 476, 503 (2014) (describing public forums as places to engage in persuasion and influence within the broader community—i.e., to take at least some small part in the formation of public opinion. For that, one needs to speak to those with whom one disagrees. Observing that public streets and sidewalks "remain one of the few places where a speaker can be confident that he is not simply preaching to the choir"); *see also* Nick Clegg, *You and the Algorithm: It Takes Two to Tango*, MEDIUM (Mar. 31, 2021), https://nickclegg.medium.com/you-and-the-algorithm-it-takes-two-to-tango-7722b19aa1c2 [perma.cc/W9LH-2LVZ]; *cf.* Cornelius v. NAACP Legal Def. & Educ. Fund, Inc., 473 U.S. 788, 815 (1985) (Blackmun, J., dissenting) ("Government property often provides the only space suitable for large gatherings, and it often attracts audiences that are otherwise difficult to reach.").

7. Erin L. Miller, *Amplified Speech*, 43 CARDOZO L. REV. 1, 30–31 (2021).

8. *See id.* at 32 n.131.

9. *See generally* Todd Fraley, Mediated Communication, Democracy, and The Public Sphere: Critical Media Consciousness Within Progressive Social Movements (2004) (Ph.D. dissertation, University of Georgia) (on file with University) (suggesting an elaboration of ideas concerning the fundamental components for a critical media consciousness, and providing support for the contention that to expand the democratic potential of communication technologies for the common good, the democratic process necessitates democratic communications structures and practices that increase citizen participation through the building of strong, equitable, and sustainable social relations among diverse peoples).

10. Miller, *supra* note 7, at 29.

11. *Id.* at 27.

12. Gary J. Simson & Rosalind S. Simson, *Rescuing* Roe, 24 N.Y.U. J. OF LEGIS. & PUB. POL'Y 313, 350 (2022).

considered, and (2) challenging prevailing norms and power structures.[13] Dialogue across cultural boundaries and multiple perspectives decreases indifference or hostility to others because it addresses differences, instils an ethic of mutual respect and recognition, and enhances a critical attitude toward one's own culture and a tolerant attitude toward others.[14] In the public sphere differences are acknowledged and addressed, and society can adapt to changing circumstances and challenges by drawing on the creativity and expertise of its members. Conversely, in the absence of respect for a plurality of conceptions of the good, intolerance can take root, accompanied by a steadfast refusal to acknowledge the inherently partial nature of all truths.[15] This is true for any public matters, but particularly for pre-election discourse.

Social media and user-generated content create a marketplace of ideas and have the potential to enable meaningful exchange, where public figures and commentators can engage, express themselves, and interact with other users in new and enhanced ways.[16] As speakers, social media enables users to participate in shaping published content by actively reporting news at scale, expressing their opinions, discussing, sharing, criticizing freely, and engaging in the public sphere without barriers to entry (apart from an internet connection) or the mediation previously imposed by the gatekeepers of the traditional media industry.[17] For

---

13. *See generally* Janna Anderson & Lee Rainie, *Many Tech Experts Say Digital Disruption Will Hurt Democracy*, PEW RSCH. CTR. (Feb. 21, 2020), https://www.pewresearch.org/internet/2020/02/21/concerns-about-democracy-in-the-digital-age/ [perma.cc/RU8C-9NUJ].

14. Jaeho Cho et al., *Do Search Algorithms Endanger Democracy? An Experimental Investigation of Algorithm Effects on Political Polarization*, 64 J. BROAD. & ELEC. MEDIA 150, 165-67 (2020) (demonstrating that when individuals encounter information that contradicts their existing beliefs, known as counter-attitudinal information, they may experience heightened ambivalence. This increased ambivalence has the potential to mitigate affective polarization, thus fostering a more conducive environment for productive discourse).

15. *Id.* at 168 (demonstrating that expanding the range of search terms employed, beyond those closely aligned with an individual's pre-existing beliefs and identity, holds the potential to diminish political selectivity and promote a more inclusive and balanced information landscape).

16. For the definition of "platform," see Ayelet Gordon-Tapiero, Alexandra Wood & Katrina Ligett, *The Case for Establishing a Collective Perspective to Address the Harms of Platform Personalization*, 25 VAND. J. ENT. & TECH. L. 635, 638 (2023). For platforms as intermediaries of economic activities, see Lina M. Khan, *Amazon's Antitrust Paradox*, 126 YALE L.J. 710, 795–96 (2017). For platforms "in a broader social sense of comprising the basic infrastructure of modern society," see 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). For platforms as 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, see 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, 495 (2020) ("It is widely recognized that the business models of large internet companies rely upon the collection, use, and analysis of personal information.").

17. David Ardia, Evan Ringel, Victoria Smith Ekstrand & Ashley Fox, *Addressing the Decline of Local News, Rise of Platforms, and Spread of Mis- and Disinformation Online*, CTR. INFO. TECH. PUBLIC LIFE, https://citap.unc.edu/news/local-news-platforms-mis-disinformation/[perma.cc/6TED-EKBP]; Leslie Gielow Jacobs, *Is There an Obligation to Listen?*, 32 U. MICH. J. L. REFORM 489, 493 (1998). *But see* Efrat Nechushtai & Seth C. Lewis, *What Kind of News Gatekeepers Do We Want Machines to Be? Filter Bubbles, Fragmentation, and the Normative Dimensions of Algorithmic Recommendations*, 90 COMPUTS. IN HUM. BEHAV. 298, 298 (2018) ("[T]he most recommended five news organizations comprised 69% of all recommendations. Five news organizations alone accounted for 49% of the total number of recommendations collected."); Miller, *supra* note 7, at 66 (2021) ("For many citizens lacking wealth and resources, speaking at certain public places or times, or in certain manners, are roughly their

hundreds of millions of people, social media creates an environment where as listeners, they have an opportunity to be exposed to a wider range of viewpoints and experiences from diverse backgrounds. In recent years, audiences have come to consume news on social media platforms and less directly from broadcast, cable, and print media.[18] "Two-thirds of Americans (67 percent) get at least some news [from] social media,"[19] and 25 percent of adults in the United States regularly consume political content via YouTube.[20] In this sense the internet, and social media in particular, has the potential to be the new town square,[21] the marketplace of ideas,[22] where freedom of speech can flourish, facilitate the exchange of information, and manifest the values of democracy and pluralism.[23]

Social media also has the potential to foster engagement in online conversation between the public and political actors. For example, government representatives' engagement with citizens through the public sphere has intrinsic benefits.[24] It can lead to better and more democratic, transparent, inclusive, legitimate, and accountable policy making, and enhances public trust in government and democratic institutions.[25] By giving citizens a role in public decision making, open discourse on social media platforms promotes individuals' autonomy and self-governance, demonstrating that decisions are not imposed from above but are the result of a participatory process.[26]

---

only gateways into broader public discourse and the formation of public opinion. Letters to the editor of newspapers are often not accepted; internet blogs and webpages will not be visited unless they are picked up by amplifying algorithms.").

18. PLURALISM OF NEWS, *supra* note 5, at 34 ("Habits of direct news consumption (such as flipping through a newspaper or watching a continuous public broadcast) have been overtaken by aggregated news consumption (such as scrolling through a feed of articles from different publishers, or watching a playlist of recommended videos from various sources)."); *see also* Jeffrey Gottfried & Elisa Shearer, *News Use Across Social Media Platforms 2016*, PEW RSCH. CTR. (May 26, 2016), https://www.pewresearch.org/journalism/2016/05/26/news-use-across-social-media-platforms-2016/ [perma.cc/49RL-PVH7].

19. Kristen Bialik & Katerina Eva Matsa, *Key Trends in Social and Digital News Media*, PEW RSCH. CTR. (Oct. 4, 2017), https://www.pewresearch.org/short-reads/2017/10/04/key-trends-in-social-and-digital-news-media/ [perma.cc/3JJ4-GK7E].

20. Hazem Ibrahim, Nouar AlDahoul, Sangjin Lee, Talal Rahwan & Yasir Zaki, *YouTube's Recommendation Algorithm Is Left-Leaning in the United States*, 2 PNAS NEXUS 1, 1 (2023).

21. As suggested by Bill Gates in 1995. Tom Huddleston Jr., *Here's What Bill Gates Said About the Internet in a Microsoft Internal Memo 25 Years Ago Today: It's a 'Tidal Wave*,*'* CNBC (May 26, 2020, 5:02 PM), https://www.cnbc.com/2020/05/26/how-bill-gates-described-the-internet-tidal-wave-in-1995.html [perma.cc/BSG6-ZK6S].

22. As suggested by Mark Zuckerberg in 2019 in the name of Facebook and other websites (defending Facebook's refusal to censor false political speech). Tony Romm, *Zuckerberg: Standing for Voice and Free Expression*, WASH. POST (Oct. 17, 2019, 4:22 PM), https://www.washingtonpost.com/technology/2019/10/17/zuckerberg-standing-voice-free-expression/ [perma.cc/4LQ9-JJDE] (text of Georgetown University speech); *see also* Packingham v. North Carolina, 137 S. Ct. 1730, 1737 (2017) (adopting the "modern public square" metaphor); Stanley Ingber, *The Marketplace of Ideas: A Legitimizing Myth*, 1984 DUKE L.J. 1, 4 (1984) ("[C]ourts that invoke the marketplace model of the [F]irst [A]mendment justify free expression because of the aggregate benefits to society, and not because an individual speaker receives a particular benefit.").

23. Gielow Jacobs, *supra* note 17, at 493.

24. *See, e.g.*, Anna P. Kambhampaty, *Securing the TikTok Vote*, N.Y. TIMES (June 22, 2023), https://www.nytimes.com/2022/03/19/style/tiktok-political-campaigns-midterm-elections.html [perma.cc/7PJT-C4DM] (discussing politicians' use of social media to develop a public image).

25. OECD, OECD GUIDELINES FOR CITIZEN PARTICIPATION PROCESSES 3 (2022), https://www.oecd.org/en/publications/oecd-guidelines-for-citizen-participation-processes_f765caf6-en/full-report.html [perma.cc/GJ23-DXR7].

26. *Id.*

But social media platforms also affect and engineer the public discourse through their algorithms.[27] Platforms employ algorithms that curate, recommend, and distribute content to users, prioritizing certain types of content higher than others.[28] The criteria algorithms employ must prioritize certain content, and by default, suppress or filter out other content, unless the user deliberately searches for it.[29] This is inherent, therefore unavoidable. Any decision, or lack thereof, regarding the management of content found on the platforms will inevitably create some prioritization that affects content flow.[30] The design of content distribution on social media platforms exerts a profound influence on the values central to pluralistic-democratic societies.[31]

Considering the significance of public discourse for democracy, this paper aims to pinpoint a specific aspect that regulations overlook: the degradation of discourse in a universally accessible public sphere caused by individually tailored algorithmic content matching. It questions whether maintaining democratic discourse is attainable when each content is selectively matched to each user.

Part I explains how personalizing digital platforms stifles meaningful public discourse, which in turn harms democratic values. The content feed on each user's social media is the result of two related content matching features, amplifying algorithms and recommendation systems, which I will name "personalization features." It reviews the way personalization features affect information flow, its social consequences, and the power dynamic they reflect. Part II highlights differences between traditional forms of media (e.g., printed and broadcast media) and digital social media in methods, volumes, legal and ethical responsibilities, and the individualized selective exposure to content, some of which are under-addressed by existing regulation. Part III discusses existing legal mechanisms that address aspects of public discourse online. Part IV evaluates whether and how legal mechanisms could address the unique problems created by selective exposure to content. Part V suggests that potential solutions could take two directions. First, creating alternative platforms for public discussion, acknowledging that addressing issues in the public domain might necessitate distinct considerations. Second, developing a comprehensive approach that goes beyond regulating online discussions or algorithms, exploring ways to promote positive dialogue, respect, and tolerance in the digital environment, along with a focus on public education.

---

27. Sofia Grafanaki, *Platforms, the First Amendment and Online Speech: Regulating the Filters*, 39 PACE L. REV. 111, 115 (2018); *see also* John Villasenor, *Why Creating an Internet "Fairness Doctrine" Would Backfire*, BROOKINGS (June 24, 2020), https://www.brookings.edu/blog/techtank/2020/06/24/why-creating-an-internet-fairness-doctrine-would-backfire/ [perma.cc/R4GG-65WA] (quoting Sen. Josh Hawley (R-Mo.): "Big tech companies would have to prove to the FTC by clear and convincing evidence that their algorithms and content-removal practices are politically neutral.").

28. Grafanaki, *supra* note 27, at 125.

29. *Id.* at 125–26.

30. *Id.*; *see also* TARLETON GILLESPIE, CUSTODIANS OF THE INTERNET: PLATFORMS, CONTENT MODERATION, AND THE HIDDEN DECISIONS THAT SHAPE SOCIAL MEDIA 5 (2018) ("The fantasy of a truly 'open' platform is powerful, resonating with deep, utopian notions of community and democracy—but it is just that, a fantasy. There is no platform that does not impose rules, to some degree. Not to do so would simply be untenable.").

31. Grafanaki, *supra* note 27, at 111.

A. The Role of Personalizing Features in Exposure to Content

The platforms' business models are based on the attention economy, which allows the platforms to quantify users' attention to specific content and monetize it through exposure to advertisements.[32] Amplifying algorithms, also known as content curation algorithms, determine which content to prioritize by controlling its order and prominence in users' feeds or search results.[33] To support their business model, algorithms are designed to choose the content that provokes users to engage and react.[34] Amplified content that is being displayed more prominently in users' feeds, receives more engagement and tends to keep users on the platforms for longer, which results in higher advertising revenues.[35] Amplification algorithms are extremely beneficial to commercial, political, and other entities who hope to amplify specific content to specific audiences.[36] The speakers, data brokers, and advertisers included are the clients— not the users.

Recommendation systems are designed to provide personalized suggestions or recommendations to users.[37] Recommendation systems tailor the content that users see in their news feeds based on their individual preferences, behaviors, and interests, typically based on user data such as browsing history, purchase history, ratings, likes, and demographic information.[38] More satisfied users will have better

---

32. CHANTAL LINE CARPENTIER, UNITED NATIONS, NEW ECONOMICS FOR SUSTAINABLE DEVELOPMENT: ATTENTION ECONOMY 1, https://www.un.org/sites/un2.un.org/files/attention_economy_feb.pdf [perma.cc/RGD8-4MPY].

33. For the definition of "amplifying algorithms," see Daphne Keller, *Amplification and Its Discontents: Why Regulating the Reach of Online Content is Hard*, 1 J. FREE SPEECH L. 227, 231–32 (2021). Within the definition of "amplifying algorithms" Keller includes various platform features, like recommended videos on YouTube or the ranked newsfeed on Facebook, that increase people's exposure to content organically created by other users (not the content of advertisements) in consumer-facing platforms like Facebook or YouTube (not infrastructure providers like CloudFlare or Amazon Web Services). This definition includes both "pull" models like the search results a user requests from Google and "push" models like YouTube video recommendations. It includes both actions platforms take in response to specific content (like demoting news items identified as false by fact checkers) and global algorithmic changes (like Google's 2017 shift to reduce ranking of content including "hoaxes and unsupported conspiracy theories"). Excluded is user-initiated viral content shared on platforms like WhatsApp, as well as the additional algorithmic boost platforms might provide.

34. Motahhare Eslami, Karrie Karahalios, Christian Sandvig, Kristen Vaccaro, Aimee Rickman, Kevin Hamilton & Alex Kirlik, *First I" Like" It, Then I Hide It: Folk Theories of Social Feeds*, *in* PROCS. CHI CONF. ON HUM. FACTORS IN COMPUTING SYS. 2373 (2016); PIER LUIGI PARCU ET AL., EUROPEAN COMMISSION, DIRECTORATE-GENERAL FOR COMMUNICATIONS NETWORKS, STUDY ON MEDIA PLURALITY AND DIVERSITY ONLINE: FINAL REPORT 76–77 (2022), https://data.europa.eu/doi/10.2759/529019 [perma.cc/LUD2-82YV] (describing how Google search works). *See, e.g.*, Mark Scott, *YouTube Algorithms Promote Climate Change-Denying Videos: Report*, POLITICO (Jan. 16, 2020, 10:05 AM), https://www.politico.eu/article/youtube-avaaz-climate-change-manipulation-global-warming-algorithm/ [perma.cc/A8QC-N4CL] .

35. Gordon-Tapiero et al., *supra* note 16, at 643, 646; *see* Stuart Minor Benjamin, *The First Amendment and Algorithms*, *in* THE CAMBRIDGE HANDBOOK OF THE LAW OF ALGORITHMS 624–25 (Woodrow Barfield ed., 2021).

36. CHANTAL LINE CARPENTIER, *supra* note 32, at 2; *see, e.g.*, *About Lookalike Audiences*, FACEBOOK, https://www.facebook.com/business/help/164749007013531 [perma.cc/AQR4-4MDZ] (last visited Dec. 27, 2023) (explaining that advertisers can identify a relevant audience and choose targeting criteria based on attributes, such as ZIP code or expressed interests).

37. Zeshan Fayyaz, Mahsa Ebrahimian, Dina Nawara, Ahmed Ibrahim & Rasha Kashef, *Recommendation Systems: Algorithms, Challenges, Metrics, and Business Opportunities*, APPLIED SCI., Nov. 2020, at 1.

38. *Id.*

experience, and engage more with the platform. Therefore, if a user frequently interacts with certain types of content, recommendation systems will then be more likely to prioritize similar content in the future and suggest new content, goods, and activities.[39] Some algorithms use the behavior and preferences of other users who have a similar profile to suggest relevant content. For instance, if a user shows interest in specific types of music, the algorithm may suggest similar music that other users with similar interests have listened to. If less meaningful content would lead to less engagement—the goal is to make sure you see what you find most meaningful.[40]

The functions of amplifying algorithms and recommendation systems are different. One focuses on pushing specific content above others, and the other focuses on matching to users' specific content out of the existing general pool.[41] They prioritize content differently. Amplifying algorithms determine what content matches users according to the content's engagement metrics, popularity, and the potential to attract traffic in order to increase user engagement, such as likes, comments, shares, and click-through, rather than enhancing users' experience.[42] Thus, the size of the audience reached depends on how much attention the speaker can draw to her speech, and for how long.[43] Recommendation systems prioritize based on users' search terms, personal network, and consumption of information to best reflect users' interests and preferences.[44] But both recommendation systems and amplifying algorithms are feeding each other and contributing to individualized users' exposure to certain content—ideas, trends, or opinions. Amplified content is likely to increase user engagement, and the ways that users engage with social media will affect the kind of content available to users through recommendation systems.[45] This project exposes the practice of personalizing features and their impact on the public discourse, which is required for maintaining democratic values.

Both amplifying algorithms and recommendation systems use the same practices to function and match specific type of content to users. First, to understand their interests, algorithms gather data regarding users' online activity and preferences, browsing history, search history, personal attributes, and social, physical, or electronic ties and activities.[46] The information points for profiling may

---

39. Keller, *supra* note 33, at 230; *see, e.g.*, GILLESPIE, *supra* note 31, at 21–23 (describing in detail how Facebook and other social media platforms comprehensively filter, sort, and structure the content that flows between users).

40. Clegg, *supra* note 6.

41. Keller, *supra* note 33, at 232.

42. Clegg, *supra* note 6.

43. Miller, *supra* note 7, at 17 (arguing that the speaker's audience "depends on her use of additional amplifiers").

44. Cho et al., *supra* note 14, at 152 ("Since search terms reveal users' underlying interests and preferences, algorithms can make informed guesses about each user and personalize searches and recommendations."). Broadening search terms could be possible when individuals are motivated to explore ideas encountered in their social networks, leading to more diverse algorithmic recommendations, and potentially reducing self-reinforcement and opinion polarization. *Id.* at 167. *See also* Judith Möller, Damian Trilling, Natali Helberger & Bram van Es, *Do Not Blame It on the Algorithm: An Empirical Assessment of Multiple Recommender Systems and Their Impact on Content Diversity*, 21 INFO., COMM. & SOC'Y 959, 961–62 (2018) (generally describing the basics of algorithmic design).

45. *See* Matthew S. Levendusky, *Why Do Partisan Media Polarize Viewers?*, 57 AM. J. POL. SCI. 611, 611–12 (2013).

46. *See* Gordon-Tapiero et al., *supra* note 16, at 637–43; *see also* Kai Shu et al., *The Role of User Profiles for Fake News Detection*, *in* PROCS. 2019 IEEE/ACM INT'L CONF. ON ADVANCES SOC.

also be gathered through less obvious tactics, such as a quiz to determine which "Friends" character the user is.[47] The analysis of users' information helps platforms examine not only each individual's uniqueness but also how users fit into profiles, patterns, clusters, and trends.[48] Second, based on the gathered data, the algorithms create detailed profiles of users.[49] Thousands of reference points forming these profiles are the basis for personalizing features to match between speakers' content and the individual listener's features. Algorithms analyze metadata of web pages, search engines, social media posts, advertising networks, marketing analytics providers, and videos, including other user-generated content; suggest groups to join; and present posts or videos (e.g., news articles, physical gatherings, and other users to connect with) to boost for maximum engagement.[50] In this paper I will use the term *personalization features* to describe these practices.

B. Personalization Features and the Nature of Public Discourse in Social Media

Building on previous scholarship that focused on how the shift of public discourse to social media has sensationalized the nature of discourse, this section will focus on how personalization features have contributed to its increasing sensationalism. Social media platforms offer a multitude of specialized communities which often facilitate the organization of groups or individuals, each catering to specific interests, ideologies, and affiliations; foster a sense of belonging; and provide spaces for like-minded individuals to connect.[51] Social media can raise awareness and provide emotional support, which may enhance social connectedness and reduce loneliness for certain individuals.[52] At the same time, the way in which algorithms are encoded leads to the formation of feedback loops, often called "echo chambers" or "bubble filters," which have significant impact on information flow and polarization.[53] For example, frequent engagement with posts from conservative

---

NETWORKS ANALYSIS & MINING 436, 437 (2019) (arguing for a relationship between user profiles on social media and the proliferation of fake news).

47. Shu et al., *supra* note 46, at 438.

48. Gordon-Tapiero et al., *supra* note 16, at 639.

49. *Id.* at 645.

50. *Id.* at 646; *see also* Emma Llansó, Joris van Hoboken, Paddy Leerssen & Jaron Harambam, Artificial Intelligence, Content Moderation, and Freedom of Expression 14 (Transatlantic Working Group, Working Paper, 2020).

51. Tessie Waithira, *Belonging Online*, MEDIUM (Oct. 5, 2023), https://medium.com/spur-collective/belonging-online-cd7960f9cd53 [perma.cc/EC8T-VSL5] ("Whether it's a niche subreddit, a Facebook group for cat lovers, a hashtag to follow on Instagram, or a fandom dedicated to a celebrity, communities offer a sense of connection and shared interests.").

52. *See generally* Hunt Allcott, Luca Braghieri, Sarah Eichmeyer & Matthew Gentzkow, *The Welfare Effects of Social Media*, 110 AM. ECON. REV. 629, 630 (2020).

53. Frederik J. Zuiderveen Borgesius, Damien Trilling, Judith Möller, Balázs Bodó, Claes H. de Vreese & Natali Helberger, *Should We Worry About Filter Bubbles?*, 5 INTERNET POL'Y REV. 1, 2–3 (2016) (distinguishing self-selected, in which the user chooses to encounter like-minded views and opinions—i.e., an "echo chamber"—and pre-selected, which is driven by platforms without the user's deliberate choice, input, knowledge or consent—i.e., a "filter bubble"); Anna Gausen, Wayne Luk & Ce Guo, *Using Agent-Based Modelling to Evaluate the Impact of Algorithmic Curation on Social Media*, 15 ACM J. DATA & INFO. QUALITY 1, 18 (2022).

sources[54] will lead to users receiving more content from conservative perspectives.[55] Likewise, if certain information does not match users' preferences, it is hidden from sight and other information is pushed, further intensifying the establishment of filter bubbles.[56] Thus, personalization features may narrow the range of information individuals are exposed to, rather than broaden it.[57]

The content chosen by the algorithm will most likely be the content that can generate higher levels of engagement and maximize advertisement revenue.[58] Provocative sensational content, while harmful, is attention-grabbing and elicits strong emotional responses, such as outrage or controversy, making it more profitable and monetizable for the platform.[59] Since algorithms prioritize this type of content, speakers who seek an audience will adjust to the form of discourse encouraged by the algorithms. Therefore, speakers may fuel their content with severe forms of expression, inappropriate comments and curses, hate speech, violence and pornography, misinformation, and false news in an attempt to gain

54. *See generally* Cass R. SUNSTEIN, #REPUBLIC: DIVIDED DEMOCRACY IN THE AGE OF SOCIAL MEDIA (2017); TAINA BUCHER, IF . . . THEN: ALGORITHMIC POWER AND POLITICS 78–79 (2018) (noting that Facebook, Twitter, Instagram, and YouTube all use algorithms that promote content to users at least in part based on what they have previously liked); *see, e.g.*, Derek O'Callaghan, Derek Greene, Maura Conway, Joe Carthy & Pádraig Cunningham, *Down the (White) Rabbit Hole: The Extreme Right and Online Recommender Systems*, 33 SOC. SCI. COMPUT. REV. 459, 460 (2015) (users accessing an extreme right YouTube video are likely to be recommended further extreme right content, leading to immersion in an ideological bubble).

55. *See, e.g.*, Eytan Bakshy, Solomon Messing & Lada A. Adamic, *Exposure to Ideologically Diverse News and Opinion on Facebook*, 348 SCI. 1130, 1130 (2015) (noting that almost 71% of new information presented to a user by Facebook in the newsfeed shows opinions that are aligned with the ideology of the user in question).

56. Researchers at MIT have found that more moderate or balanced perspectives are often marginalized. However, "empirical research tends to focus on a single platform—YouTube—or is observational or anecdotal in nature." Joe Whittaker, Seán Looney, Alastair Reed & Fabio Votta, *Recommender Systems and The Amplification of Extremist Content*, 10 INTERNET POL'Y REV. 1, 2 (2021); Itai Himelboim, Stephen McCreery & Marc Smith, *Birds of a Feather Tweet Together: Integrating Network and Content Analyses to Examine Cross-Ideology Exposure on Twitter*, 18 J. COMPUTER-MEDIATED COMM. 154, 156 (2013) (describing how on Twitter, citizens tend to interact with other users who are similar to themselves within homogeneous groups, where discussions take place between people with similar positions, and where different opinions are generally marginalized or ignored).

57. *See* Brent Kitchens, Steven L. Johnson & Peter Gray, *Understanding Echo Chambers and Filter Bubbles: The Impact of Social Media on Diversification and Partisan Shifts in News Consumption*, 44 M.I.S. QUARTERLY 1619, 1620 (2020) ("Personalization technology is sensitive to personal preferences; once a user engages with opinion-reinforcing content, algorithmic filtering may constrain further exposure to a narrower, more closely aligned range of content."); *see also* Gielow Jacobs, *supra* note 17, at 493 (discussing the purposes of free speech).

58. Peter Dizikes, *Study: On Twitter, False News Travels Faster Than True Stories*, MIT NEWS (Mar. 8, 2018), https://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308 [perma.cc/4B9W-WQ94] (stating that since provocative content travels six times faster than truthful information, its dissemination is significantly more extensive).

59. *See generally* Steve Rathje, Jay J. Van Bavel & Sander van der Linden, *Out-Group Animosity Drives Engagement on Social Media*, 118 PROCS. NAT'L ACAD. SCI. 1 (2021) (analyzing nearly three million social media posts finding that posts about the political outgroup were more likely to be shared than those about the political ingroup. Each additional outgroup word (e.g., "liberal" or "Democrat" for a Republican post) increased the odds of that post being shared by 67% and increased the volume of "angry" reactions on Facebook); *see also* Caroline Mala Corbin, *The First Amendment Right Against Compelled Listening*, 89 B.U. L. REV. 939, 962 (2009) (noting that studies show that certain forms of speech promoted online produce a range of negative emotional, psychological, and physical effects, "including feelings of guilt, shame, anxiety, fear, vulnerability, inferiority, personal inadequacy, and degradation," which affect mostly children and students).

influence and reach a larger audience.[60] Indirectly, algorithms contribute to the fact that many groups use hateful language, exacerbating extreme forms of expression, sensationalism, and emotionally provocative content, eventually affecting the nature of discourse as a whole.[61]

Alongside the ability to express oneself and connect, the echo chamber creates an illusion of widespread agreement among its members, not necessarily reflective of the sphere beyond the echo chamber.[62] This distortion, through practices that are skewed towards particular perspectives, is poisonous for coexistence in any society.[63] People exercising their own rights to freedom of expression—through, for example, unrestricted hate speech —threaten the free speech of others.[64] In such an environment, on the one hand, users feel more comfortable to express less

60. Grafanaki, *supra* note 27, at 119; Mark S. Kende, *Social Media, the First Amendment, and Democratic Dysfunction in the Trump Era*, 68 DRAKE L. REV. 273, 286–87 (2020); *see, e.g.*, Sheera Frenkel & Kate Conger, *Hate Speech's Rise on Twitter Is Unprecedented, Researchers Find*, N.Y. TIMES (Dec. 2, 2022), https://www.nytimes.com/2022/12/02/technology/twitter-hate-speech.html [perma.cc./9M2L-6BLL]; Strategic Communications, *Hate Speech Poisons Societies and Fuels Conflicts*, EUROPEAN UNION EXTERNAL ACTION (June 18, 2022), https://www.eeas.europa.eu/eeas/hate-speech-poisons-societies-and-fuels-conflicts_en [perma.cc./FS6D-NX9W] (according to the European Commission Statement, during the coronavirus pandemic, "hate speech and extremist ideologies have increasingly been spreading on the Internet").

61. For example, "Facebook's friend- and group-recommendation algorithms are said to have brought together violent right-wing extremists, one of whom ultimately shot and killed two people in Kenosha, Wisconsin." Keller, *supra* note 33, at 230; *see also* Megan A. Brown, Jonathan Nagler, James Bisbee, Angela Lai & Joshua A. Tucker, *Echo Chambers, Rabbit Holes, and Ideological Bias: How YouTube Recommends Content to Real Users*, BROOKINGS (Oct. 13, 2022), https://www.brookings.edu/research/echo-chambers-rabbit-holes-and-ideological-bias-how-youtube-recommends-content-to-real-users/ [perma.cc./NHK8-NBFS]; Tanya Basu, *YouTube's Algorithm Seems To Be Funneling People to Alt-Right Videos*, MIT TECH. REV. (Jan. 29, 2020), https://www.technologyreview.com/2020/01/29/276000/a-study-of-youtube-comments-shows-how-its-turning-people-onto-the-alt-right/ [perma.cc./L425-678F].

62. Whittaker et al., *supra* note 56, at 3.

63. *See* SUNSTEIN, *supra* note 54, at 271 n.2 (stating that personalized content, driven by algorithms and filter bubbles, contributes to the fragmentation and polarization of public discourse. Sunstein argues that social media platforms and their algorithms often show users content that aligns with their preexisting beliefs and preferences, creating echo chambers where individuals are exposed primarily to information and opinions that reinforce their existing views, which can further deepen political and ideological divisions).

64. Steven J. Heyman, *The Conservative-Libertarian Turn in First Amendment Jurisprudence*, 117 W. VA. L. REV. 231, 334 (2014). Identifying, quantifying, and understanding hate incidents can be challenging. First, there may be divergent views on which hate incidents should be counted and why they matter. Defining hate and establishing clear criteria for what constitutes a hate incident can be subjective and vary across contexts and jurisdictions. Identifying and quantifying hate incidents may involve subjective judgments. Different organizations, advocacy groups, or individuals may have varying interpretations of what qualifies as hate based on their objectives, societal impact, or perception of urgency, leading to discrepancies in reporting and measurement. For example, hate can manifest in various forms, including hate speech, hate crimes, online harassment, discrimination, and systemic biases. Which content should be removed? Second, hate incidents are often underreported due to various reasons, including fear of retaliation, lack of trust in authorities, or individuals not recognizing their experiences as hate related. As a result, reporting systems reflect inconsistent information to obtain a comprehensive view of the prevalence and nature of hate. Limited data availability can hinder accurate tracking and understanding of hate levels in society. Finally, the impact and significance of hate incidents can vary depending on the social, cultural, and historical context. *See* Matteo Vergani, Barbara Perry, Joshua Freilich, Steven Chermak, Ryan Scrivens & Rouven Link, *PROTOCOL: Mapping the Scientific Knowledge and Approaches to Defining and Measuring Hate Crime, Hate Speech, and Hate Incidents*, 18 CAMPBELL SYST. REVS. 1, 1–3 (2022).

tolerant views.[65] On the other hand, peer pressure, or the desire to conform to the dominant group, can discourage individuals from engaging in open-minded dialogue and expressing what they really think about controversial issues for fear of harsh criticism, retaliation, or "being cancelled" as cultural responses suggest.[66] Effectively, algorithms promote small numbers of aggressive people, established and influential accounts or entities, "trolls" and provocateurs, making their content more visible and far-reaching while silencing moderate citizens.[67]

When individuals are directed to information solely from the sources they find most engaging,[68] their passive exposure to a narrow spectrum of opinions and ideas can hinder critical thinking and the development of a well-rounded comprehension of intricate societal issues.[69] Members may avoid exploring alternative perspectives, further isolating dissenting voices within the community.[70] Placing users into information silos based on their personal characteristics serves as a barrier to

65. For particular regulatory dilemmas and challenges posed by online as compared to offline hate speech, see DANIELLE KEATS CITRON, HATE CRIMES IN CYBERSPACE (2014); Alexander Tsesis, *Hate in Cyberspace: Regulating Hate Speech on the Internet*, 38 SAN DIEGO L. REV. 817 (2001); Barbara Perry & Patrik Olsson, *Cyberhate: The Globalization of Hate*, 18 INFO. & COMM. TECH. L. 185 (2009); Julian Baumrin, *Internet Hate Speech and the First Amendment. Revisited*, 37 RUTGERS COMPUT. & TECH. L.J. 223 (2011); Richard Delgado & Jean Stefancic, *Four Observations About Hate Speech*, 44 WAKE FOREST L. REV. 353 (2009); RAPHAEL COHEN-ALMAGOR, CONFRONTING THE INTERNET'S DARK SIDE: MORAL AND SOCIAL RESPONSIBILITY ON THE FREE HIGHWAY (2015); Alexander Brown, *What is so Special About Online (as Compared to Offline) Hate Speech?*, 18 ETHNICITIES 297 (2017), https://journals.sagepub.com/doi/10.1177/1468796817709846 [perma.cc/F9EK-K2DJ].

66. The Editorial Board, *America Has a Free Speech Problem*, N.Y. TIMES (Mar. 18, 2022), https://www.nytimes.com/2022/03/18/opinion/cancel-culture-free-speech-poll.html [perma.cc/NC7T-9M42].

67. Alexander Bor & Michael Bang Petersen, *The Psychology of Online Political Hostility: A Comprehensive, Cross-National Test of the Mismatch Hypothesis*, 116 AM. POL. SCI. R. 1, 2 (2022); Jonathan Haidt, *Why the Past 10 Years of American Life Have Been Uniquely Stupid*, ATLANTIC (Apr. 11, 2022) https://www.theatlantic.com/magazine/archive/2022/05/social-media-democracy-trust-babel/629369/[perma.cc/3AKN-G897] ("Additional research finds that women and Black people are harassed disproportionately, so the digital public square is less welcoming to their voices."); *see also* Villasensor, *supra* note 27 (stating that many critics of the current form of Section 230 believe that tech leadership has been partial, favoring certain political stances and expressions over others).

68. Walé Azeez, *YouTube: We've Learnt Lessons From Christchurch Massacre Video*, YAHOO FINANCE U.K. (May 15, 2019), https://uk.finance.yahoo.com/news/you-tube-weve-learnt-lessons-from-christchurch-massacre-video-163653027.html [perma.cc/9WPK-UFZQ] (showing that YouTube's public policy director told a U.K. House of Commons select committee that around 70% of content watched on the platform was derived from recommendations rather than users' organic searches.).

69. Mihaela Popescu & Lemi Baruh, *Captive but Mobile: Privacy Concerns and Remedies for the Mobile Environment*, 29 THE INFO. SOC'Y 272, 278 (2013). *See, e.g.*, Samuel C. Rhodes, *Filter Bubbles, Echo Chambers, and Fake News: How Social Media Conditions Individuals to be Less Critical of Political Misinformation*, 39 POL. COMMC'N 1, 2–3 (2022) ("[H]omogeneous information streams increase beliefs in fake news and that there is evidence of heuristic information processing after exposure to filter bubbles."); William Hart, Dolores Albarracín, Alice H. Eagly, Inge Brechan, Matthew J Lindberg & Lisa Merrill, *Feeling Validated Versus Being Correct: A Meta-Analysis of Selective Exposure to Information*, 135 PSYCH. BULL. 555, 555 (2009) (exploring the distinction between two motivations behind selective exposure: the desire to feel validated in one's existing beliefs and the desire to be factually correct. They find that people are more likely to selectively expose themselves to information that aligns with their pre-existing attitudes and opinions, seeking validation for their beliefs, rather than prioritizing factual accuracy).

70. Virginia Morini, Laura Pollacci & Giulio Rossetti, *Toward a Standard Approach for Echo Chamber Detection: Reddit Case Study*, 11 APPLIED SCIS. 1, 2 (2021).

constructive dialogue and may contribute to a propensity for conflict and violence.[71] The absence of face-to-face interaction can make it more challenging to foster mutual understanding, empathy, and respectful dialogue.[72]

Moreover, the decentralized and anonymous nature of some online communities make these spaces susceptible to the spread of misinformation (the spread of inaccurate or misleading information) and disinformation (the strategic creation and dissemination of false information with the intent to deceive or manipulate).[73] Free speech on social media often promotes false information that should not be trusted, but when the full picture is concealed, users may trust such information nevertheless. Lacking editorial oversight, fact-checking, or verification mechanisms, false or misleading information can easily circulate within online communities.

Misinformation,[74] manipulation,[75] subversion of autonomy,[76] and discrimination[77] erode values such as inclusion and diversity and make it difficult to establish a shared understanding of facts.[78] Such discourse contributes to polarization and intolerance rather than connectedness and hinders the formation of a common factual basis. Despite that, repetitive exposure to content insulated from rebuttal is often viewed as more believable, digestible, and of higher quality,

---

71. Nerra Chandhoke, *The Advantages of Plural Societies*, *in* CONTESTED SECESSIONS: RIGHTS, SELF-DETERMINATION, DEMOCRACY, AND KASHMIR 126, 153–55 (2011). For the way narrow discourse will erode respect for democratic decisions, see Robert Post, *Religion and Freedom of Speech: Portraits of Muhammad*, 14 CONSTELLATIONS 72, 73–77 (2007) [hereinafter *Portraits of Muhammad*]; *see also* Ronald Dworkin, *The Right to Ridicule*, N.Y. REV. BOOKS, Mar. 23, 2006, at 44.

72. Martha Newson, Yi Zhao, Marwa El Zein, Justin Sulik, Guillaume Dezecache, Ophelia Deroy & Bahar Tunçgenç, *Digital Contact Does Not Promote Wellbeing, but Face-To-Face Contact Does: A Cross-National Survey During the COVID-19 Pandemic*, 26 NEW MEDIA & SOC'Y 426, 428 (2024); *see also* Brown et al., *supra* note 61; *see, e.g.*, Basu, *supra* note 61.

73. Rhodes, *supra* note 69; *see, e.g.*, Eric Fehrnstrom, *A Punch-Drunk Jeb Bush Carries On*, BOSTON GLOBE (Feb. 10, 2016, 3:15 PM), https:/www.bostonglobe.com/opinion/2016/02/10/punch-drunk-jeb-bush-carries/AFJZk3IF7UwmOsd4jqMfIP/story.html [perma.cc/EB2F-FP6Z]; Bradford Richardson, *Trump Rips Bush Over 'Act Of Love' Remarks on Illegal Immigration*, THE HILL (Aug. 31, 2015, 1:31 PM), https://thehill.com/blogs/ballot-box/presidential-races/252325-trump-rips-bush-over-act-of-love-remarks-on-illegal/ [web.archive.org/web/20160220100558/http:/thehill.com/blogs/ballot-box/presidential-races/252325-trump-rips-bush-over-act-of-love-remarks-on-illegal].

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

75. Zeynep Tufekci, *Algorithmic Harms Beyond Facebook and Google: Emergent Challenges of Computational Agency*, 13 COLO. TECH. L.J. 203, 216 (2015).

76. For more on the effects of online manipulation on autonomy, see Daniel Susser, Beate Roessler & Helen Nissenbaum, *Technology, Autonomy and Manipulation*, 8 INTERNET POL'Y REV. 1, 5 (2019) ("Manipulators can frame information in a way that disposes us to a certain interpretation of the facts.").

77. Press Release, American Civil Liberties Union, In Historic Decision on Digital Bias, EEOC Finds Employers Violated Federal Law When They Excluded Women and Older Workers From Facebook Ads (Sept. 25, 2019, 11:00 AM), https://www.aclu.org/press-releases/historic-decision-digital-bias-eeoc-finds-employers-violated-federal-law-when-they [perma.cc/8FFQ-ZMKM]; Fair Housing Act of 1968, 42 U.S.C. §§ 3601-3619, 3604(c); Charge of Discrimination, U.S. Dep't of Hous. & Urb. Dev., FHEO No. 01-18-0323-8 (2019); *see, e.g.*, Kunal Relia, Zhengyi Li, Stephanie H. Cook & Rumi Chunara, *Race, Ethnicity and National Origin-Based Discrimination in Social Media and Hate Crimes Across 100 U.S. Cities*, 13 PROC. OF THE INT'L AAAI CONF. ON WEB AND SOC. MEDIA 417 (2019).

78. European Commission Against Racism and Intolerance (ECRI), *Hate Speech and Violence*, COUNCIL OF EUROPE, https://www.coe.int/en/web/european-commission-against-racism-and-intolerance/hate-speech-and-violence# [perma.cc/JC4Z-DPKM].

which facilitates the processing of it.[79] Users perceive these popular messages as more credible or valid and may act passionately according to them.[80] In the past it has incited acts of violence and hate crimes which have been negatively impacting individuals and communities, undermining social cohesion, as well as democracy itself and the rule of law. For example, the Capitol insurrection on January 6, 2021, when thousands of Donald Trump supporters thronged the U.S. Capitol to halt the constitutionally mandated certification of President Biden's election victory, was managed and spread through Facebook. Facebook's internal report showed 40,000 user reports of "false news" per hour prior to the event, including from the President's official Instagram account.[81]

In summary, algorithmic personalization hampers the free flow of information and ideas based on reasoned and informed debate; narrows the exposure to alternative, diverse, and dissenting opinions, rather than broadening it and encouraging critical thinking.[82] Interactions between different filter bubbles are not systematically designed into the algorithms; therefore, the development of a broader, inclusive democratic discourse is limited.[83] The spread of harmful, misleading, or skewed content—which is also fragmented, decentralized, and scattered—undermines broader societal interests in constructive public dialogue across and within various communities necessary for effective democratic decision-making.[84]

C. The Ramifications of Information Overload on the Public Discourse

Another implication of personalization features on the public discourse is the amount of content on social media, and the contribution of personalization features to divisiveness.[85] More information is created every two years in the modern world

79. Arthur G. Miller, John W. McHoskey, Cynthia M. Bane & Timothy G. Dowd, *Attitude Polarization Phenomenon: Role of Response Measure, Attitude Extremity, and Behavioral Consequences of Reported Attitude Change*, 64 J. PERSONALITY & SOC. PSYCH. 561 (1993); COUNCIL OF THE EUROPEAN UNION, THE ROLE OF ALGORITHMIC AMPLIFICATION IN PROMOTING VIOLENT AND EXTREMIST CONTENT AND ITS DISSEMINATION ON PLATFORMS AND SOCIAL MEDIA (2020), https://data.consilium.europa.eu/doc/document/ST-12735-2020-INIT/en/pdf [perma.cc/5UE5-A5M9] (assigning polarization to recommendation systems).

80. 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 (n.d.), https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/cpc-middleware_ff_v2.pdf [perma.cc/V8GS-C68B].

81. Craig Timberg, Elizabeth Dwoskin & Reed Albergotti, *Inside Facebook, Jan. 6 Violence Fueled Anger, Regret Over Missed Warning Signs*, WASH. POST (Oct. 22, 2021), https://www.washingtonpost.com/technology/2021/10/22/jan-6-capitol-riot-facebook/ [perma.cc/985C-MCYZ].

82. 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. SCI. 9216 (2018); Bakshy et al., *supra* note 55, at 1130–32.

83. SUNSTEIN, *supra* note 54, at 155–56 ("To the extent that the communications market becomes more personalized, it reduces the range of widely shared experiences and at the same time fails to confer some of the benefits that come when individuals receive information, often more helpful to others than to themselves, that they would not have chosen in advance. If the role of public forums and general-interest intermediaries is diminished, and if good substitutes do not develop, those benefits will be diminished as well, with harmful results for democratic ideals.").

84. *Id.*; *see* Brown et al., *supra* note 61; Basu, *supra* note 61.

85. For studies that have analyzed the effects of algorithmically recommended content on political opinion polarization and their conflicting results, see Eliza Mitova, Sina Blassnig, Edina

than had previously been created in all of human history,[86] or, as Eric Schmidt, then CEO of Google, put it in 2010, we create as much information in two days as in all of history up until 2003.[87] 995 photos are uploaded to Instagram, 6,000 new tweets are published, and 41,000 posts are made on Facebook every second. Susan Gunelius found that every minute Facebook users shared nearly 2.5 million pieces of content, Twitter users tweeted nearly 300,000 times, Instagram users posted nearly 220,000 new photos, YouTube users uploaded 72 hours of new video content, Apple users downloaded nearly 50,000 apps, and email users sent over 200 million messages.[88] Proliferation of poisonous narratives is part of this information overload.[89]

Many narratives, in different levels of accuracy and depth, target internet users. Users are unable to take in all the information that is out there. When overwhelmed by the complexity of content overload, most people cannot judge what constitutes misinformation, skewed or manipulated information beyond their immediate environment.[90] To confront this, users often rely on mechanisms that screen for what is "real" and "true," that are simpler than applied rationality.[91] These mechanisms embody tribal narratives,[92] because they are both simple and emotionally satisfying.[93] They undermine rationality as basis of legitimacy, and base individuals' perspectives on immediacy and emotion. While this strategy reduces the need to process incoming information, the reliance on over-simplified tribal

---

Strikovic, Aleksandra Urman, Aniko Hannak, Claes H. de Vreese & Frank Esser, *News Recommender Systems: A Programmatic Research Review*, 47 ANNALS INT'L COMMUC'N ASS'N 84, 95 (2023); Cho et al., *supra* note 14, at 165–67.

86. Bernard Marr, *Big Data: 20 Mind-Boggling Facts Everyone Must Read*, FORBES (Sept. 30, 2015, 2:19 AM), https://www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/#2162010c17b1 [perma.cc/PZ7E-UFFQ]. Naturally, there is a huge argument about what constitutes "information," but the fundamental point is that virtually all information processing mechanisms, including human cognitive systems and institutions, are operating in an information environment for which they were never intended, and which is overwhelmingly complex, rapid, and unintelligible to existing entities.

87. M.G. Siegler, *Eric Schmidt: Every Two Days We Create as Much Information as We Did Up to 2003*, TECHCRUNCH (Aug. 4, 2010, 4:58 PM), https://techcrunch.com/2010/08/04/schmidt-data [perma.cc/5VSJ-9FQJ].

88. Susan Gunelius, *The Data Explosion in 2014 Minute by Minute—Infographic*, ACI (July 12, 2014), https://www.newstex.com/blog/the-data-explosion-in-2014-minute-by-minute-infographic.

89. *See* Braden R. Allenby, *Information Technology and the Fall of the American Republic*, 59 JURIMETRICS 409, 415–16 (2019) ("'[W]eaponized narrative,' which is defined as 'the use of information and communication technologies, services, and tools to create and spread stories intended to subvert and undermine an adversary's institutions, identity, [] civilization,' . . . 'exacerbating complexity, confusion, and political and social schisms.'").

90. *Id.* at 435 n.155.

91. *Id.* at 419.

92. *Id.* at 419, 435. For the understanding of "tribal narratives," see Amy Chua, *Tribal World: Group Identity Is All*, FOREIGN AFFS. (July 2018), https://www.foreignaffairs.com/articles/world/2018-06-14/tribal-world [perma.cc/6NS3-29P6] ("In recent years, the United States has begun to display destructive political dynamics much more typical of developing and non-Western countries: the rise of ethno-nationalist movements, eroding trust in institutions and electoral outcomes, hate-mongering demagoguery, a popular backlash against both 'the establishment' and outsider minorities, and, above all, the transformation of democracy into an engine of zero-sum political tribalism.").

93. Allenby, *supra* note 89, at 419; ROBERT KAPLAN, THE RETURN OF MARCO POLO'S WORLD: WAR, STRATEGY, AND AMERICAN INTERESTS IN THE TWENTY-FIRST CENTURY 242–49 (2018); FRANCIS FUKUYAMA, IDENTITY: THE DEMAND FOR DIGNITY AND THE POLITICS OF RESENTMENT 69–70 (2018).

narratives pushes the discourse towards more simplistic identity political discourse that provides pre-made meanings .[94] Partial and arbitrary tribal narratives reinforce pre-existing beliefs and polarization, and create division where individuals are more likely to view those from opposing groups as enemies rather than fellow citizens.[95] When political discourse revolves around tribal identity demands, the focus shifts away from universal values and a nuanced discourse that serves social and collective interests towards identity lines and group-specific concerns that contribute to the fragmentation of society. Identities can be mapped and individually attacked, becoming a geopolitical battlespace to be designed.[96] Moreover, identity politics can lead to a zero-sum mentality where one group's gains are seen as another's losses.[97] When a political discourse ends in "winning" of one narrative, it leads to exclusion and contributes to political polarization, where certain groups assert their interests and rights at the expense of others.[98]

Such limited narratives cannot support or represent socially shared values. They hinder unity and replace the sense of a shared national or societal values to the point that it can be difficult to conduct productive political discourse or accommodate diverse interests and perspectives, which are crucial for democratic decision-making.[99] Instead of reducing anxiety, divisiveness increases social tension and alarm by generating fear about social change and sharpening of social divides between different groups in society. Emphasizing social and collective interests over narrow tribal affiliations is essential for fostering a more cohesive and pluralistic

---

94. DANIEL KAHNEMAN, THINKING, FAST AND SLOW 324 (2011); Allenby, *supra* note 89, at 423, 435 ("Does it agree with what I and those like me believe?" Rather than "it is true"); Amy Chua, *How America's Identity Politics Went From Inclusion to Division*, THE GUARDIAN (Mar. 1, 2018, 6:00 AM), https://www.theguardian.com/society/2018/mar/01/how-americas-identity-politics-went-from-inclusion-to-division [perma.cc/VQ25-2E6P].

95. SUNSTEIN, *supra* note 54, at 14; *see also Post-Truth Politics: Art of the Lie*, THE ECONOMIST (Sept. 10, 2016), https://www.economist.com/leaders/2016/09/10/art-of-the-lie [perma.cc/7LW8-Z94A]; Allenby, *supra* note 89, at 435; Francis Fukuyama, *Against Identity Politics: The New Tribalism and the Crisis of Democracy*, 97 FOREIGN AFF. 90 (2018).

96. Braden R. Allenby, *World Wide Weird: Rise of the Cognitive Ecosystem*, 37 ISSUES SCI. & TECH. 34, 36 (2021).

97. Walter Benn Michaels, *Identity Politics: A Zero-Sum Game*, 19 NEW LAB. F. 8, 11 (2010); *see* Aníbal Rosario Lebrón, *The Ouroboros of Identity Politics*, 30 TUL. J.L. & SEXUALITY 131, 132 (2021).

98. PAUL M. BARRETT, JUSTIN HENDRIX & J. GRANT SIMS, FUELING THE FIRE: HOW SOCIAL MEDIA INTENSIFIES U.S. POLITICAL POLARIZATION — AND WHAT CAN BE DONE ABOUT IT (2021), https://static1.squarespace.com/static/5b6df958f8370af3217d4178/t/613a4d4cc86b9d3810eb35aa/1631210832122/NYU+CBHR+Fueling+The+Fire_FINAL+ONLINE+REVISED+Sep7.pdf [perma.cc/GU7H-T65Q]; Kende, *supra* note 60, at 281; *see id.* at 284 (mentioning that at a symposium, "Jack Balkin described the nation as the most politically polarized since the Civil War"). For studies analyzing the effects of algorithmically recommended content on political opinion polarization and their conflicting results, see Mitova et al., *supra* note 85, at 95; Cho et al., *supra* note 14, at 165–67 (suggesting that political self-reinforcement, as indicated by the political emotion-ideology alignment, and affective polarization are heightened by political videos—selected by the YouTube recommender algorithm—based on participants' own search preferences).

99. Chua, *supra* note 92 (noting that the United States, with its perception of individuals as holders of natural rights on the one hand and as citizens who are the source of legitimacy for the modern state on the other, is particularly blind to the rise of tribalism. "In recent years, the United States has begun to display destructive political dynamics much more typical of developing and non-Western countries: the rise of ethnonationalist movements, eroding trust in institutions and electoral outcomes, hate-mongering demagoguery, a popular backlash against both 'the establishment' and outsider minorities, and, above all, the transformation of democracy into an engine of zero-sum political tribalism."); Chua, *supra* note 94.

society. Therefore, normative criticism should not focus solely on problematic content but also consider, irrespective of content, how personalization features threaten to impede adequate democratic deliberation and may be inherently contradictory to it.[100]

II. WHAT IS THE DIFFERENCE? CONTROLLED DISTRIBUTION VS. CONTROLLED CONTENT GENERATION

Real public discourse has tiers defined by levels of amplification—for example, organic level, paid level, and viral level. Additionally, space and time constraints inevitably subject all public discourse to tiers of amplification. After all, only a select few speakers can be heard by a large number of people. For broader distribution of information and ideas, speakers need the right platform, which can reach a wide audience and have the potential to change the beliefs and conduct of others. Amplifying algorithms play a crucial role in facilitating the flow of information to an audience.[101] Without mass amplification, individuals would probably remain isolated in their own informational bubbles, lacking access to diverse perspectives and challenges to their views.[102] While both traditional media and digital social media are platforms used for mass amplification and distribution, there is a difference between what each can offer.[103]

In broad terms, it can be asserted that traditional media platforms typically exert control over content because they employ journalists and editors that create content, but have less control over distribution of their prints or broadcast. Once their content is published, anyone can see it and pass it on. In contrast, digital social media platforms tend to have less control over content, which can be generated practically by any user. However, personalization features play a pivotal role in the selective distribution of content providing the capability and distribution methods to control exposure to particular content more precisely according to users' profiles, and maintaining the discourse in an individualized manipulated sphere rather than the public sphere. Traditional news broadcast sources—which concentrate ownership of amplification decisions in the hands of a select few gatekeepers—can be accused of inflicting many of the same harms as algorithmic personalizing features, such as catering to specific target audiences, facilitating the formation of communities based on specific interests and ideologies (creating echo chambers), and manipulating audiences' exposure to content.[104] While these concerns are not unique to the digital sphere, their manifestation online differs significantly in both scale and effect. Traditional broadcast and print media operate within more visible and regulated frameworks. In contrast, social media platforms deploy opaque

100. Howdle, *supra* note 3, at 445.

101. Grafanaki, *supra* note 27, at 115.

102. Miller, *supra* note 7, at 25–26. *Id.* at 12 (defining mass amplification as at least a few million people.).

103. Quantifying the exact impact of newspapers versus social media is complex, as it can vary depending on the specific context, geographical location, target audience, and research methodology employed.

104. C. Edwin Baker, *Scope of the First Amendment Freedom of Speech*, 25 UCLA L. REV. 964, 965–66 (1978) ("Because of monopoly control of the media, lack of access of disfavored or impoverished groups, techniques of behavior manipulation, irrational response to propaganda, and the nonexistence of value-free, objective truth, the marketplace of ideas fails to achieve the desired results.").

technical infrastructure that imposes structural barriers to participation in the "marketplace of ideas,"[105] foster a false appearance of neutral or rational information exposure,[106] and selectively distribute individualized manipulated content without meaningful legal or professional accountability.[107] This Section examines how these systemic differences reshape the informational landscape.

A. Manipulated, Individualized Content Distribution

Advancements across a wide range of disciplines, including psychology, cultural intelligence studies, behavioral economics, marketing, and neuroscience, are accelerating capabilities to distribute individualized content in a way that has the power to manipulate users, communities, institutions, and states.[108] The next sections explain the difference between these capabilities and those of traditional print and broadcast media.

1. Autonomy Traps and Non-Transparent Content Manipulation Capabilities

Traditional forms of media have a limited reach both in scope (more localized readership and limited to subscribers or specific regions) and in access compared to global social media platforms, presenting a singular message to their entire viewing audience.[109] In comparison, the widespread sharing of content on social media platforms facilitates broader and more rapid spread of information to a wider audience than it might find in print media, often leading to the viral dissemination of content.[110] Therefore, online speech is likely to lead to much more severe and

105. *See id.* Not everyone may have equal access to the required devices, stable internet connections, or the literacy to navigate virtual reality environments. This can result in exclusion and a potential digital divide, where certain groups or individuals are disproportionally excluded from democratic deliberation and decision-making processes, exacerbating existing inequalities in democratic participation.

106. *See* Stanley Ingber, *The Marketplace of Ideas: A Legitimizing Myth*, 1984 DUKE L.J. 1, 7 (1984) ("On the whole, current and historical trends have not vindicated the market model's faith in the rationality of the human mind[.]"); Harry H. Wellington, *On Freedom of Expression*, 88 YALE L.J. 1105, 1130 (1979) ("It is naive to think that truth will *always* prevail over falsehood in a free and open encounter, for too many false ideas have captured the imagination of man."); *see also* Baker, *supra* note 104, at 976 ("Emotional or 'irrational' appeals have great impact."); Ingber, *supra*, at 35–36 (explaining that social sciences have established the irrational elements of persuasiveness, so that the "market model's reliance on public rationality is, at best, misplaced").

107. Cent. Hudson Gas & Elec. Corp. v. Pub. Serv. Comm'n, 447 U.S. 557, 592 (1980) (Rehnquist, J., dissenting) ("There is no reason for believing that the marketplace of ideas is free from market imperfections any more than there is to believe that the invisible hand will always lead to optimum economic decisions in the commercial market."); Ingber, *supra* note 106, at 15–16 (suggesting that the marketplace of ideas is as flawed as the economic market).

108. Rachel Weintraub-Reiter, *Hate Speech over the Internet: A Traditional Constitutional Analysis or a New Cyber Constitution?*, 8 BOS. PUB. INT'L L.J. 145, 165 (1998) ("[T]he Internet is not as intrusive as television or radio."); *see also* Allenby, *supra* note 89, at 415–16.

109. KAREN ROSS & VIRGINIA NIGHTINGALE, MEDIA AND AUDIENCES: NEW PERSPECTIVES 16 (2003). Newspapers have expanded their online presence and adapted to the digital era by developing their own websites, mobile apps, and social media accounts. This blurs the line between traditional newspapers and online news consumption. Operating in environments that prioritize engagement and sensationalism over impartiality and fact-checking affects the nature and quality of content generated by news media organizations.

110. Keller, *supra* note 33, at 232; *see also* PLURALISM OF NEWS, *supra* note 5, at 34.

wide-spread consequences than offline speech.[111] If I say something controversial to my neighbor, only my neighbor will be able to respond. Maybe word will get around the neighborhood, but the overall reach of the statement (and any consequences/retaliation) will probably be relatively limited. If I say something controversial online, millions of people with no relation to me can potentially find it and try to retaliate against me.

Platforms use systemic and behavioral capabilities, linking together ever more powerful networks to exploit cognitive biases and encourage the wide spread of certain content.[112] Cognitive bias steer users towards preferred choices by using dark patterns, such as visual cues, color schemes, manipulative language, or creating a false sense of urgency by displaying countdown timers.[113] For example, most platforms primarily seek to raise revenue or will favor paid speech.[114] Through the availability bias (the tendency to rely on readily available information when making judgments), platforms can position content prominently in paid high-visibility placements, like margins or banners, elevating the "rank" of content to ensure it appears at the forefront or top of a user's primary information feed (i.e., Facebook news feed, Google search results, or recommended videos on YouTube).[115] Recency bias, giving more weight to recent information, creates a perception of reality that is based on popular or recent trends rather than a comprehensive understanding of the broader context. Both biases hide undesirable or unpaid content, making it less practically available to users.

Powerful competitive forces drive forward the widespread use of commercially motivated design, dark patterns and manipulative tactics. Difficult-to-detect dark patterns and non-transparent practices create an autonomy trap, different from the practice used in most traditional media, because they can manipulate consumers without their knowledge or specific consent so that people are unable to make informed choices about what they consume or how they behave.[116] Users may believe they are in control of their content flow while not fully

---

111. Zachary Laub, *Hate Speech on Social Media: Global Comparisons*, COUNCIL ON FOREIGN RELS. (June 7, 2019, 3:51 PM), https://www.cfr.org/backgrounder/hate-speech-social-media-global-comparisons [perma.cc/5P2N-MRXH].

112. Allenby, *supra* note 96.

113. Keller, *supra* note 33, at 231. *See* Miller, *supra* note 7, at 14. For example, content that is prominently positioned in Facebook's user interface (like the middle of the screen) is amplified compared to other parts of the page (like the bottom edge). *See also* James Grimmelmann, *Listeners' Choices*, U. COLO. L. REV. 365, 379 (2019) (describing how Facebook's algorithmic News Feed, for example, "nudges users into reading friends' posts as well as Pages and Groups you have chosen to follow or join in a default, Facebook-selected order").

114. Miller, *supra* note 7, at 13; Kende, *supra* note 60, at 282 ("[S]ocial media sites have financial incentives to use algorithms that channel consumers to the most inflammatory sites possible to solidify the user's interest."); COMM. OF LEGAL AFFS., DRAFT REPORT WITH RECOMMENDATIONS TO THE COMMISSION ON A DIGITAL SERVICES ACT 5, 23 (2020), https://www.europarl.europa.eu/doceo/document/JURI-PR-650529_EN.pdf [perma.cc/4Y56-2HRQ]; *see also* Ben Popken, *As Algorithms Take Over, YouTube's Recommendations Highlight a Human Problem*, NBC NEWS (Apr. 19, 2018), https://www.nbcnews.com/tech/social-media/algorithms-take-over-youtube-s-recommendations-highlight-human-problem-n867596 [perma.cc/SS28-NLZW] (according to a EU Parliament report, a platform that optimizes for ad revenue has reason to prioritize sensational, emotional-provoking content); Louis Michael Seidman, *Can Free Speech Be Progressive?*, 118 COLUM. L. REV. 2219, 2235–36 (2018).

115. Miller, *supra* note 7, at 14.

116. Press Release, Fed. Trade Comm'n, FTC Report Shows Rise in Sophisticated Dark Patterns Designed to Trick and Trap Consumers (Sept. 15, 2022), https://www.ftc.gov/news-events

appreciating the effect of content compelled upon them by algorithms, or know how design techniques are employed to steer their behavior towards actions they may not have intended or desired, that are not in their best interests, or that go against their values. It is one thing to be persuaded to change one's belief; it is quite another to have it changed by evading the ordinary means of rational thought that ideally shape one's beliefs.[117] Recognizing the ability of content that is made by bots, not even necessarily generated or moderated by humans to steer users' behavior is crucial for understanding the potential dangers associated with the impact of multiple metaverses and personalized online experiences on democratic deliberation.

Similar concerns about ethics, transparency, and effectiveness have been raised in the past regarding subliminal messaging or advertising.[118] Subliminal messaging involves the presentation of information below the threshold of conscious perception, intended to bypass deliberate awareness and embed itself in the subconscious mind in a manner that is virtually undetectable.[119] The deliberate goal of this technique is to influence attitudes or behavior—particularly consumer behavior—without the recipient being aware that they are receiving any message at all. Common methods include flashing words or images on a screen for extremely brief durations (e.g., as short as 1/2000 of a second),[120] embedding messages in audio tracks at frequencies difficult to detect or through reversed lyrics,[121] and incorporating subtle or hidden visuals, such as words formed by clouds in the background of an image.[122] When discovered, the use of subliminal messaging has been found to damage corporate reputation and has been linked to the spread of harmful stereotypes, perpetuation of inequalities, and erosion of trust in advertising.[123]

Many countries have specific laws to ban subliminal marketing focused on the deceptive result.[124] In the United States, their use falls under federal law

/news/press-releases/2022/09/ftc-report-shows-rise-sophisticated-dark-patterns-designed-trick-trap-consumers [perma.cc/Y5UT-SV28].

117. Larry Alexander, *Compelled Speech*, 23 CONST. COMMENT. 147, 154 (2006).

118. Derived from the Latin, "sub" meaning below and "limen" meaning threshold, "subliminal exists just below the threshold of conscious awareness." *Subliminal*, MERRIAM-WEBSTER DICT., https://www.merriam-webster.com/dictionary/subliminal [perma.cc/U3R7-RE9X] (last visited Dec. 27, 2023).

119. *Id.*

120. *The History of Subliminal Messages*, UNIV. MICH., https://websites.umich.edu/~onebook/pages/frames/historySet.html [perma.cc/M6XS-YHCN] (last visited Dec. 27, 2023); *see, e.g.*, *Subliminal McDonald's Ad on Food Network*, YOUTUBE (Jan. 22, 2007), https://youtu.be/amnZX-jjBD8 [perma.cc/GWG3-DQUK].

121. In 1990, a 1978 cover version by heavy metal band Judas Priest was the subject of a much-publicized "subliminal message trial." The lawsuit alleged that the band's recording contained hidden messages that were responsible for influencing a pair of young men in Sparks, Nevada, to make a suicide pact in 1985. The case was eventually dismissed. *Vance v. Judas Priest*, No. 86-3939, 1990 WL 130920, at *1 (Nev. Dist. Ct. Aug. 24, 1990).

122. *See, e.g.*, *Subliminal Messaging Examples*, UNIV. MICH., http://websites.umich.edu/~onebook/pages/frames/usesSet.html [perma.cc/7QBG-HT52] (last visited Dec. 27, 2023).

123. *See Ethical Considerations in Advertising and Price Manipulation*, FASTERCAPITAL (June 2, 2024), https://fastercapital.com/topics/ethical-considerations-in-advertising.html/1 (providing advice for businesses on how to advertise).

124. *Are Subliminal Messages Legal?*, UNIV. MICH., https://websites.umich.edu/~onebook/pages/frames/legalF.html [perma.cc/9APS-U2B4] (last visited Aug. 27, 2024). Britain and Australia, for example, ban subliminal advertising for any reason. The United States does not expressly forbid the

enforcement jurisdiction.[125] The Federal Communications Commission (FCC) issued Public Notice FCC 74–78 (1974), and an Information Bulletin titled "Subliminal Projection,"[126] stating that the broadcast license of any company that uses subliminal advertising techniques in its broadcast will be revoked because the techniques are deceptive by nature, and run counter to the purpose of the FCC—"to promote connectivity and ensure a robust and competitive market."[127] At the same time, algorithms that have been proven to be able to lead to the same results are protected under free speech.[128] Sections 5 and 12 of the Federal Trade Commission Act, which forbid any advertising that is deceptive or unfair, could also be unsuitable for regulating dark patterns.[129] Determining whether dark patterns are "fair" or "non-deceptive" is hard, because personalization features include many required, inevitable functions, such as design or editing choices. It is therefore harder to prove deceptive or unfair use to qualify with Sections 5 and 12 of the Federal Trade Commission Act.[130]

2. Individualized Outlet

Any media has power over democracy as controller of the architecture of public discourse. This is not unique to social media platforms. The ownership and control of traditional media outlets has a significant impact on editorial decisions and the narratives they promote.[131] They have the power to frame news stories to influence how the public perceives political issues and events and to impact the outcome of elections.[132] Critics consistently raise concerns about media ownership,

---

use of subliminal messages in advertisements, though subliminal messaging was acknowledged in 1955 to be a cause of public concern. The Federal Communications Commission statement also states that broadcasters should approach the technique cautiously.

125. For laws on subliminal marketing, *see* Stephanie Dube Dwilson, *Laws on Subliminal Marketing*, CHRON, https://smallbusiness.chron.com/laws-effect-against-spam-64389.html [perma.cc/8KLT-JTCQ] (last visited Dec. 27, 2023).

126. Subliminal Projection, 44 Fed. Reg. 36389 (June 22, 1979).

127. *Federal Communications Commission*, U.S. GOV, https://www.usa.gov/federal-agencies/federal-communications-commission [perma.cc/9MHZ-HZK3] (last visited Dec. 27, 2023).

128. *See* Stuart Minor Benjamin, *Algorithms and Speech*, 161 U. PA. L. REV. 1445, 1457–58 (2013) (discussing *Brown v. Ent. Merchs. Assn.*, 564 U.S. 786 (2011), among others).

129. Section 5 of the Federal Trade Commission Act (FTC Act) prohibits "unfair or deceptive acts or practices in or affecting commerce." 15 U.S.C. § 45. Section 12 prohibits the dissemination of any false advertisement that is likely to induce the purchase of food, drugs, devices, services, or cosmetics. 15 U.S.C. § 52.

130. Dwilson, *supra* note 125. However, in a 1989 case, a Nevada judge ruled that subliminal messages aren't protected by the speakers' First Amendment rights and constitute an invasion of privacy. In that particular case, no one had proven that subliminal messages could actually move someone to act against their will.

131. Miller, *supra* note 7, at 61–62. *See e.g.*, Daniel Muise, Homa Hosseinmardi, Baird Howland, Markus Mobius, David Rothschild, & Duncan J. Watts, *Quantifying Partisan News Diets in Web and TV Audiences*, 8 SCI. ADVANCES 1, 1 (2022) (analyzing billions of browsing and viewing events between 2016 and 2019, the study estimates that 17% of Americans are partisan-segregated through television versus roughly 4% online. Also, television news consumers are several times more likely to maintain their partisan news diets month-over-month. TV viewers' news diets are far more concentrated on preferred sources. Therefore, the study suggests that television is the top driver of partisan audience segregation among Americans).

132. Cho et al., *supra* note 14, at 154 ("The heightened selectivity as a result of the user-algorithm interaction induces a personally tailored information environment for each user which, in turn, shapes the user's opinion.").

both from the perspective of speakers and news subjects who want to convey a message, and from the perspective of audiences seeking unbiased news coverage that suits their interests.[133]

It could be argued that as intermediaries between political actors and the citizens, traditional media has caused the same harms as algorithmic personalizing features. Historically, newspapers, television, and radio have played a crucial role in informing the public about political issues, candidates, and government policies.[134] Traditional media outlets as gatekeepers set the agenda of the public discourse by determining which issues to cover and how they are framed, which voices and perspectives are heard, and which are marginalized. However, due to the differences between traditional and online personalized news delivery methods, personalization features pose a novel threat, as they enable the delivery of individualized content streams to each member of the audience, carrying distinct strengths and limitations in fostering public discourse.

Unlike the traditional news framework, where one headline serves all readers (listeners in traditional First Amendment terminology), on digital social media each member gets an individualized output based on targeting points such as race, ethnicity, gender, religion, sexual orientation, as well as individual social interactions and behavior online.[135] Intersectionality adds complexity to the specific compilation of content curated for each individual. An individualized content stream allows different and hyper-specific messages to be amplified, contributing to a lack of shared reality and eroding notions of common facts.[136] Effectively, there are billions of front page headlines, each personalized at the individual level, and no uniform content that all users see and can relate to.[137] No two individuals receive the same output of content.[138]

---

133. In the United States, Americans increasingly distrust traditional news mediums over the last several decades; a 2016 Gallup poll found that a majority of Americans hold very little or no confidence whatsoever in newspapers, television news, or online news. *Confidence in Institutions*, GALLUP, http://www.gallup.com/poll/1597/confidenceinstitutions.aspx [perma.cc/W5HU-L6TS] (last visited Dec. 27, 2023).

134. *See The Presidency in the Television Era*, UVA MILLER CTR., https://millercenter.org/the-presidency/teacher-resources/recasting-presidential-history/presidency-television-era [perma.cc/CJ3P-BMYR] (last visited Dec. 27, 2023).

135. SUNSTEIN, *supra* note 54, at 19.

136. *See* MERCY CORP., THE WEAPONIZATION OF SOCIAL MEDIA (2019), https://www.mercycorps.org/sites/default/files/2020-01/Weaponization_Social_Media_FINAL_Nov2019.pdf [perma.cc/5BR5-LS5N].

137. Allenby, *supra* note 89, at 432.

138. Will Oremus, Chris Alcantara, Jeremy B. Merrill, & Artur Galocha, *How Facebook Shapes Your Feed*, WASH. POST (Oct. 26, 2021, 7:00 AM), https://www.washingtonpost.com/technology/interactive/2021/how-facebook-algorithm-works/ [perma.cc/NGT4-BULW] (reporting a huge variety of factors the algorithm uses to determine what each person sees. Recommender systems consider ten thousand different factors, which suggests that it would be extremely unlikely for any two people to see the same picture); *see generally* ELI PARISER, THE FILTER BUBBLE: WHAT THE INTERNET IS HIDING FROM YOU (2011) (reporting a personal experience wherein two people ended up with different outcomes from an online search for "BP," with one getting investment news about British Petroleum and the other getting information about the Deepwater Horizon oil spill); Aniko Hannak, Balachander Krishnamurthy, Piotr Sapieżyński, David Lazer, Christo Wilson, Arash Molavi Kakhki, & Alan Mislove, *Measuring Personalization of Web Search*, 22 INT'L WORLD WIDE WEB CONF. 527, 536 (2013) (measuring the effect of search personalization on Google, concluding that 11.7% of Google search results differed between users due to personalization—a finding that the authors describe as "significant personalization").

Individualized selective exposure has important implications for how people consume and process information. The individualized compilation of content chosen by the algorithm manipulates and constructs, on an individual level, the perception of reality and as a result, how each user interprets information in political affairs.[139] For example, it can strengthen ideological associations, reinforcing existing beliefs and emotions (ideological reinforcement), or amplify the polarized emotions toward in-group and out-group political actors (affective polarization).[140] Individuals tend to have a structured set of attitudes, with specific opinions rooted in broader political beliefs such as ideology.[141] Confirmatory information aligns with individual existing beliefs and attitudes, reinforcing a psychological phenomenon known as motivated reasoning, which serves various psychological functions.[142] Attitude-consistent information is easier to process because it reaffirms and strengthens opinions, bolsters self-esteem, maintains a coherent and positive sense of self, and reduces cognitive dissonance that require additional cognitive effort (the discomfort that arises from holding conflicting beliefs or attitudes).[143] People are more likely to accept information that flatters their self-image and reject information that threatens it. Thus, attitude-consistent information is prioritized over factual accuracy and tends to strengthen intolerance towards opposing views.[144]

Through personalization features, platforms can leverage these patterns to strategically push and amplify specific content through their algorithms, thereby influencing public sentiment and shape outcomes, whether in service of advertisers'

139. Jaeho Cho, Saifuddin Ahmed, Martin Hilbert, Billy Liu, & Jonathan Luu, *Do Search Algorithms Endanger Democracy? An Experimental Investigation of Algorithm Effects on Political Polarization*, 64 J. BROAD. & ELEC. MEDIA 150, 153–54, 165–67 (2020); BARRETT ET AL., *supra* note 98; Jay J. Van Bavel, Steve Rathje, Elizabeth Harris, Claire Robertson, & Anni Sternisko, *How Social Media Shapes Polarization*, 25 TRENDS COGNITIVE SCIS. 913, 913 (2021) (showing methodology and deduction process flaws in research that aims to show that greater internet use is not associated with faster growth in political polarization).

140. Cho et al., *supra* note 14, at 153–54, 165–67 (2020); *see also* Van Bavel et al., *supra* note 139, at 913 (showing methodologies and deduction process flaws in research that aims to show that greater internet use is not associated with faster growth in political polarization).

141. Ziva Kunda, *The Case for Motivated Reasoning*, 108 PSYCH. BULL. 480 (1990).

142. *See id.*

143. *See id.* at 481 (1990); *see generally* ANDREW GUESS, BRENDAN NYHAN, BENJAMIN LYONS & JASON REIFLER, AVOIDING THE ECHO CHAMBER ABOUT ECHO CHAMBERS 1–25 (2018); NATALIE JOMINI STROUD, NICHE NEWS: THE POLITICS OF NEWS CHOICE (2011); CASS SUNSTEIN, ECHO CHAMBERS: *BUSH V. GORE*, IMPEACHMENT, AND BEYOND (2001) (arguing information that contradicts existing beliefs can be challenging to process and takes more time and effort).

144. Miller, *supra* note 7, at 38; *see, e.g.*, Cho et al., *supra* note 14, at 165–67 (noting that people tend to defend and favor their in-group while disliking the out-group, especially in a partisan context).

interests or political agendas.[145] Facebook,[146] for example, has the power both to influence users' moods by the content they are shown on the platform,[147] and when they are in a certain mood, tailor the right content at the right location and the right hour in the day.[148] In the same way, new technological capabilities provide social media platforms powers to control political discourse and with it impose new, worrisome risks to pluralistic democratic values.

Political psychology research shows that social media platforms and the incorporation of personalized algorithmic recommendations have a significant impact on users' political information reception and political opinions.[149] It can give rise to political microtargeting, making certain speakers and speeches appear more popular than they actually are, and silence others to promote platforms' business model.[150] By choosing amplified speakers, platforms can crowd out non-chosen speakers, so the chosen ones have little need to respond to lesser amplified opponents.[151] By segmenting audiences with precision, microtargeting strategies

145. Eli J. Finkel, Christopher A. Bali, Mina Cikara, Peter H. Ditto, Shanto Iyengar, Samara Klar, Lilliana Mason, Mary C. McGrath, Brendan Nyhan, David G. Rand, Linda J. Skitka, Joshua A. Tucker, Jay J. Van Bavel, Cynthia S. Wang, & James N. Druckman, *Political Sectarianism in America*, 370 SCI. MAG. 533, 534 (2021) ("In recent years, social media companies like Facebook and Twitter have played an influential role in political discourse, intensifying political sectarianism."); *see generally* Judith Moeller, Damian Trilling, Natali Helberger, Kristina Irion, & Claes De Vreese, *Shrinking Core? Exploring the Differential Agenda Setting Power of Traditional and Personalized News Media*, 18 EMERALD GRP. PUBL'G LTD., 26–41 (2016); R. Lance Holbert, Brian E. Weeks, & Sarah Esralew, *Approaching the 2012 U.S. Presidential Election from a Diversity of Explanatory Principles: Understanding, Consistency, and Hedonism*, 57 AM. BEHAV. SCIENTIST 1663, 1663–87 (2013) (providing empirical evidence and insights into how the principles of understanding (the desire to make informed and rational choices), consistency (the tendency to align choices with previously held beliefs and party affiliations), and hedonism (the role of emotional factors in voting decisions based on personal feelings, or perceptions of the candidates' likability) influenced voters' choices and decision-making processes).

146. Although the Facebook company is now "Meta," I use Facebook and Meta interchangeably for consistency with previous scholarship. *What Are the Meta Products?*, FACEBOOK, https://www.facebook.com/help/1561485474074139 [perma.cc/2WVU-YKJD] (last visited Dec. 27, 2023).

147. 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 (2014). For criticism from civil society, academics, and regulators alike, 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 [perma.cc/K74T-FHAM] (reporting that Facebook acknowledged the nature of the experiment).

148. Popescu & Baruh, *supra* note 69, at 278; *see, e.g.*, Rebecca Rosen, *Is This the Grossest Advertising Strategy of All Time?*, ATLANTIC (Oct. 3, 2013), https://www.theatlantic.com/technology/archive/2013/10/is-this-the-grossest-advertising-strategy-of-all-time/280242/ [perma.cc/FBX4-77L6]; Paul Armstrong, *Facebook is Helping Brands Target Teens Who Feel "Worthless,"* FORBES (May 1, 2017, 11:53 AM), https://www.forbes.com/sites/paularmstrongtech/2017/05/01/facebook-is-helping-brands-target-teens-who-feel-worthless/#b810c8a344ed [perma.cc/ZFH6-NSGC].

149. *Compare* Michael A. Beam, *Automating the News*, 41 COMMC'N RSCH. 1019, 1019–41 (2014) (finding that user-driven news customization leads to a higher elaboration, i.e., more thoughtful processing of the news message and higher attention to its content, which positively affects political knowledge), *with* Lucien Heitz, Juliane A. Lischka, Alena Birrerd, Bibek Paudel, Suzanne Tolmeijer, Laura Laugwitzc & Abraham Bernstein, *Benefits of Diverse News Recommendations for Democracy: A User Study*, 10 DIGITAL JOURNALISM 1, 1–21 (2022) (suggesting that the diversity-optimizing news recommenders that foster the exposure to opposing political viewpoints do not enhance political participation).

150. *See generally* Eslami et al., *supra* note 34, 2371–82. YouTube algorithms following searches for the moon landing or global warming quickly lead to flagrantly false information. *See* Scott, *supra* note 34.

151. Miller, *supra* note 7, at 6.

enable political parties to emphasize different issues for different voters. [152] This means that even within the same political campaign, individuals may be exposed to entirely distinct narratives or priorities, depending on their data profiles. A voter may receive an overwhelming amount of content about a single issue, leading to the mistaken belief that it is a central concern of the campaign.[153] At the same time, microtargeting allows parties to exclude certain voter groups from specific messages or suppress the visibility of particular content, deepening divisions, even within the same demographic group.[154] Thus, variation in messaging occurs not only between voter groups but also among individuals within a single group.[155]

The Cambridge Analytica scandal sheds light on the manipulation of personalized and targeted advertising to influence political campaigns.[156] Cambridge Analytica, a political consulting firm, sought to influence voter behavior and shape political opinions. During the months leading up to the election, Cambridge Analytica orchestrated a campaign based on a combination of intensive survey research, data modeling, and performance-optimizing algorithms. The company improperly harvested data from millions of Facebook users without their consent.[157] This data was then used to create detailed psychological profiles.[158] According to ProPublica, Facebook has at least 52,235 categories to sort users into demographic slices.[159] The campaign involved the creation and dissemination of a staggering ten thousand distinct customized messages, each meticulously tailored to resonate with specific demographic groups, delivered to individual voters based on

152. Frederik J. Zuiderveen Borgesius, Judith Möller, Sanne Kruikemeier, Ronan Ó Fathaigh, Kristina Irion, Tom Dobber, Balazs Bodo, & Claes de Vreese, *Online Political Microtargeting: Promises and Threats for Democracy*, 14 UTRECHT L. REV. 82, 88–89 (2018).

153. *Id.* at 86.

154. *Id.* at 88; Cho et al., *supra* note 14, at 166; 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.").

155. Sandra González-Bailón, David Lazer, Pablo Barberá, Meiqing Zhang, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Deen Freelon, Matthew Mentzkow, Andrew M. Guess, Shanto Iyengar, Young Mie Kim, Neil Malhortra, Devra Moehler, Brendan Nyhan, Jennifer Pan, Carlos Velasco Rivera, Jaime Settle, Emily Thorson, Rebekah Tromble, Arjun Wilkins, Magdalena Wojcieszak, Chad Kiewiet de Jonge, Annie Franco, Winter Mason, Natalie Jomini Stroud, & Joshua A. Tucker, *Asymmetric Ideological Segregation in Exposure to Political News on Facebook*, 381 SCI. MAG. 392, 392–98 (2023) (arguing that Facebook does enable ideological segregation in political news consumption, with conservatives having a more distinct and isolated space within the platform's news ecosystem, and a higher prevalence of misinformation within that segment).

156. 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 [perma.cc/XT5Z-NGA4].

157. Paolo Zialcita, *Facebook Pays $643,000 Fine For Role In Cambridge Analytica Scandal*, NPR (Oct. 30, 2019), https://www.npr.org/2019/10/30/774749376/facebook-pays-643-000-fine-for-role-in-cambridge-analytica-scandal [perma.cc/9RXX-DV2Y] ("Facebook breached data protection laws by failing to keep users' personal information secure, allowing Cambridge Analytica to harvest the data of up to 87 million people without their consent worldwide.").

158. Philip Bump, *All The Ways Trump's Campaign Was Aided By Facebook, Ranked By Importance*, WASH. POST (Mar. 22, 2018), https://www.washingtonpost.com/news/politics/wp/2018/03/22/all-the-ways-trumps-campaign-was-aided-by-facebook-ranked-by-importance/ [perma.cc/ARL2-WTXB].

159. Julia Angwin, Surya Mattu & Terry Parris Jr., *Facebook Doesn't Tell Users Everything It Really Knows About Them*, PROPUBLICA (Dec. 27, 2016, 9:00 AM), https://www.propublica.org/article/facebook-doesnt-tell-users-everything-it-really-knows-about-them [perma.cc/9QF8-X6BN].

their unique profiles.[160] Tailoring messages based on individuals' personality traits, fears, and preferences, Cambridge Analytica made "personal identity and political behavior a design space to be manipulated, often without awareness on the part of the target."[161] For those residing in areas with a strong likelihood of supporting Donald Trump, the advertisements featured a triumphant image of the nominee, accompanied by practical information on locating their nearest polling station.[162] In contrast, individuals whose geographic data suggested a less fervent inclination towards Trump, such as swing voters, were exposed to images of high-profile Trump supporters.[163] Cambridge Analytica employed paid Google ads for "persuasion search advertising" on their main search platform.[164] This approach entailed boosting search results favoring Donald Trump while discrediting Hillary Clinton.[165] For example, users seeking information on topics like "Trump Iraq War" encountered sponsored results highlighting Trump's opposition to the Iraq War in contrast to Hillary's support, with an arrow pointing to a result that prominently stated: "Hillary Voted for the Iraq War—Donald Trump opposed it." This comprehensive strategy exemplified Cambridge Analytica's ability to tune their messaging for maximum impact across various voter segments.[166] Such targeting is more likely to reach the relevant audience.

As the process of digitization and datafication becomes more prevalent, social media platforms gain increasing power to shape speaker-listener matching by making different suggestions on an individual basis and thereby mediating economic, social, and cultural interactions.[167] Access to these advanced

160. Paul Lewis & Paul Hilder, *Leaked: Cambridge Analytica's Blueprint for Trump Victory*, GUARDIAN (Mar. 23, 2018, 8:53 AM), https://www.theguardian.com/uk-news/2018/mar/23/leaked-cambridge-analyticas-blueprint-for-trump-victory [perma.cc/W6YB-ADQV].

161. Allenby, *supra* note 89, at 431.

162. Paolo Zialcita, *Facebook Pays $643,000 Fine for Role in Cambridge Analytica Scandal*, NPR (Oct. 30, 2019, 1:16 PM), https://www.npr.org/2019/10/30/774749376/facebook-pays-643-000-fine-for-role-in-cambridge-analytica-scandal [perma.cc/NJM7-PSNT].

163. Lewis & Hilder, *supra* note 160. One of the most impactful ads took the form of native advertising on the widely read political news website Politico. This interactive graphic, cleverly designed to resemble a legitimate piece of journalism, developed in-house by Politico's team responsible for sponsored content, enumerated "10 inconvenient truths about the Clinton Foundation." It was strategically displayed over several weeks to individuals from a carefully curated list of pivotal swing states when they visited the Politico site and achieved "average engagement time of four minutes," making it a standout achievement within their campaign.

164. Zialcita, *supra* note 162.

165. *See id.*

166. *Id.*

167. JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM 38 (2019); FRANCIS FUKUYAMA, POLITICAL ORDER AND POLITICAL DECAY 7 (2014). *See* SEAN MCFATE, THE MODERN MERCENARY: PRIVATE ARMIES AND WHAT THEY MEAN FOR WORLD ORDER 74–75 (2014); *See generally* Guy Schleffer & Benjamin Miller, *The Political Effects of Social Media Platforms on Different Regime Types*, 4 TX NAT'L SEC. REV. 77, 78 (2021) The effect of social media depends on how major actors use social media as well as on a state capacity and political regime type, and will vary accordingly. The relevant major actors were found to be: domestic opposition, external forces, and the governing regime. Depending on how these three actors use social media, there are four different effects that social media can have: it can have a weakening effect on strong democratic regimes, an intensifying effect on strong authoritarian regimes, a radicalizing effect on weak democratic regimes, and a destabilizing effect on weak authoritarian regimes. *See e.g.*, Grimmelmann, *supra* note 113, at 378; Michael Warren, *Trump Hits Jeb on "Act of Love"*, WEEKLY STANDARD (Aug. 31, 2015, 12:32 PM), http://www.weeklystandard.com/trump-hits-jeb-on-act-of-love/article/1023012 [web.archive.org/web/20160225064955/http:/www.weeklystandard.com/trum

technological capabilities is far from equitable. Speakers cannot gain mass attention simply by posting content on social media. Truly mass reaching media sources are rare, and access to them is characterized by significant and non-random disparities.[168] The 2022 Global Internet Phenomena Report highlights that six major companies—Google, Facebook (which have further consolidated their influence by acquiring rivals like Instagram and YouTube), Netflix, Amazon, Apple, and Microsoft—dominate the majority (56.96%) of all internet traffic, significantly shaping content consumption.[169] The ability to amplify and distribute content places great power in the hands of those platforms that are in a near-monopoly position in developing the distribution mechanisms that steer content consumption, and can selectively amplify specific voices.[170] Platforms provide services to political parties at their own rate and discretion, in an unequal way, and lack stable incentives to prioritize diversity and balanced views when deciding who gets the most amplification.[171]

Furthermore, unequal access to amplification that privileges powerful speakers over others is a concerning issue for citizens.[172] For the non-famous speaker, payment becomes the most reliable avenue to access these platforms. However, advertising prices on mass platforms are often beyond the means of the average citizen. Certain forms of microtargeting are so cost-prohibitive that larger parties with greater resources and expertise gain an unfair advantage, influencing people's choices and actions to their benefit.[173] For example, it costs approximately a dollar per click to advertise on Facebook, "over $400,000 for a thirty-second advertisement during a popular television show, and about $1,000 per column inch

---

p-hits-jeb-on-act-of-love/article/1023012]. In the first Republican Presidential debate, held on August 6, 2015, the moderator asked candidate Jeb Bush if he stood by a statement made the previous April that illegal entry into the U.S. by undocumented migrants is "an act of love," to which Bush replied that he did. Almost immediately thereafter, the Trump campaign posted his comment as part of a video showing mugshots of illegal immigrants who committed violent crimes in the United States, intercut with footage of Bush using the phrase. This tactic marked a turning point in Trump's campaign. The Clinton campaign used social media to advertise Trump's use of fake news and potential Russian intervention.

168. Data available in the United States suggests that American media attention is highly concentrated, flowing largely to a few media companies. *See* Patrick J. Kennedy & Andrea Prat, *Where Do People Get Their News?*, 34 ECON. POL'Y 5, 29–30 (2019) ("On average, the top five media organizations in a country control about a third of the total attention share.").

169. SANDVINE, PHENOMENA: GLOBAL INTERNET PHENOMENA REPORT 14 (2022), https://www.sandvine.com/hubfs/Sandvine_Redesign_2019/Downloads/2022/Phenomena%20Reports/GIPR%202022/Sandvine%20GIPR%20January%202022.pdf [perma.cc/5AU6-J7HP] (stating that six companies generate the majority (56.96%) of all web traffic: Google, Facebook, Netflix, Amazon, Apple, and Microsoft). *See generally* ELI M. NOAM & THE INT'L MEDIA CONCENTRATION COLLABORATION, WHO OWNS THE WORLD'S MEDIA?: MEDIA CONCENTRATION AND OWNERSHIP AROUND THE WORLD 3 (2016).

170. Miller, *supra* note 7, at 25.

171. *See id.* at 60–61.

172. *See* LAURENCE H. TRIBE, AMERICAN CONSTITUTIONAL LAW 786 (2d ed. 1988) ("Especially when the wealthy have more access to the most potent media of communication than the poor, how can we be sure that 'free trade in ideas' is likely to generate truth?"); Baker, *supra* note 104, at 978 (explaining the marketplace of ideas is improperly biased in favor of dominant groups); Ingber, *supra* note 106, at 38 ("Restriction of entry to the economically advantaged quells voices today that might have been heard in the time of the town meeting and the pamphleteer."); Grimmelmann, *supra* note 114, at 374 ("There are disparities among speakers, to be sure, of wealth, class, power, nationality, gender, language, education, and many others.").

173. Borgesius et al., *supra* note 152, at 88.

to advertise in *The New York Times*."[174] The larger targeted relevant audience reached, the higher the cost, and for multiple spots on multiple platforms or repetition, a speaker must be prepared to pay more.[175] Savvy entities like political campaigns are investing significant resources in such strategies.[176] As platforms increase engagement and drive traffic to specific sites or pages, their profits soar.[177] When only a selectively relevant audience is reached, and access is contingent to a costly mechanism, democracy faces significant threats.[178] For example platforms may cater to specific ideological preferences and decline to collaborate with political parties that lack substantial resources, thereby limiting the visibility of marginalized communities or lesser-known perspectives and constraining their ability to participate meaningfully in democratic processes.

The thought of social media as bypassing access barriers and creating a more open discourse than traditional media is at least inaccurate. Effectively, not all speakers and ideas enjoy an equal opportunity to be heard, and the power to decide who the amplified speakers are is in the hands of quasi monopolistic platforms which are profit-driven. This interplay is integral to comprehending the multifaceted challenges of the current media environment, as it has the potential to skew the public discourse, particularly in relation to the financial resources available to political parties. Erin Miller argues that the marketplace of ideas serves two crucial functions in the democratic process: (1) informing voters, and (2) legitimizing political decisions.[179] Meaningful information is crucial for democracy. The ultimate source of political power rests with citizens' unrestricted access to clear and accurate information, which enables individuals in democratic societies to make political choices.[180] Individual autonomy is paramount to this framework, supported by the cognitive skills and critical capacities necessary to resist groupthink, indoctrination, and manipulation.[181] An informed citizenry is a foundational pillar of a well-functioning democratic system[182] Both mass amplification and microtargeting have

174. Miller, *supra* note 7, at 13. *See also* Akvile DeFazio, *How Much Do Facebook Ads Cost in 2021? (+ 3 Ways to Save)*, WORDSTREAM (July 20, 2021), https://www.wordstream.com/blog/ws/2021/07/12/facebook-ads-cost [perma.cc/G7FV-U7RQ]; Julia Stoll, *Cost of a 30-Second TV Spot During This Is Us in the United States from 2016/17 to 2020/21 TV Season (in U.S. Dollars)*, STATISTA (Jan. 13, 2021), https://www.statista.com/statistics/756867/this-is-us-ad-price-usa [perma.cc/2UCS-BWGA]; *2021 Newspaper Rates*, N.Y. TIMES, https://assets.ctfassets.net/jxri9wzjewim/6NJQdCKs26V2pbwIdNRKeV/399525471157f9f4649a377fe5b72b87/901453_Newspaper_Rate_Card_2021_AW11.pdf [perma.cc/A5P3-6KKD] (click "Category Rate Cards" and choose "Newspaper Rates" from dropdown menu).

175. Miller, *supra* note 7, at 13–14.

176. Miller, *supra* note 7, at 14.

177. JEFF KOSSEFF, THE TWENTY-SIX WORDS THAT CREATED THE INTERNET 212 (2019).

178. Miller, *supra* note 7, at 6.

179. Miller, *supra* note 7, at 6.

180. Howdle, *supra* note 3, at 445; *see also* Borgesius et al., *supra* note 152.

181. Tetyana Hoggan-Kloubert & Chad Hoggan, *Post-Truth as an Epistemic Crisis: The Need for Rationality, Autonomy, and Pluralism*, 73 ADULT EDUC. Q. 3, 12 (2023).

182. In alignment with the International Declaration on Information and Democracy, the "global communication and information space should serve the exercise of freedom of expression and opinion and respect the principles of pluralism, freedom, dignity, tolerance, and the ideal of reason and understanding." PLURALISM OF NEWS, *supra* note 5, at 27. This assertion underscores the intrinsic importance of knowledge, which is indispensable for human beings to develop their biological, psychological, social, political, and economic capacities. *See* Sarah Eskens, Natali Helberger & Judith Moeller, *Challenged by News Personalisation: Five Perspectives on the Right to Receive Information*, 9 J. OF MEDIA L. 259 (2017) (considering the right to receive information).

distinct and detrimental effects on democracy by disrupting citizens' access to information. However, sespite its potential distortions of democratic values and its impact on equal participation in the democratic process, the European Court of Human Rights (ECHR), has affirmed that "publishing information with the [intent] to influence" voters is an exercise of freedom of expression.[183]

Effective participation in civic matters can only occur when individuals engage in meaningful dialogue with others and make autonomous political choices. According to Miller, for the system of freedom of expression to operate effectively, it must incorporate three key features: diversity, to ensure a broad range of perspectives reach the public; mobility, to allow ideas a fair opportunity and a realistic chance to gain traction; and at least occasional antagonism, to facilitate vibrant, responsive, and sometimes confrontational exchanges among differing viewpoints in close proximity.[184] Mass-amplified speech deliberately disrupts all these preconditions. Individualized flow of content denies the public access to relevant information, diminishing shared reality and public accountability, which are vital elements for democratic deliberation.[185] The delivery of individualized skewed content enables the dissemination of a diverse array of highly precise and conflicting messages, tailored to specific individuals, without any one individual encountering messages that are not tailored to them.[186] According to Giles Howdle, accessible publicity, or the extent to which salient political information is accessible to the public as a whole, is an ethical essence of fostering deliberative democracy, enabling public scrutiny, and providing a platform for opposing views to be "heard out."[187] Howdle points out that a traditionally targeted political advertisement in a left-wing newspaper, while primarily seen by left-wing voters, remains accessible to right-wing voters and media. This accessibility ensures that the placement of such an advertisement contributes to increasing the publicity of the information conveyed, even for those with differing viewpoints.[188] In comparison, individualized flow of content undermines the public discourse because it disrupts publicity of salient information. Echo chambers and filter bubbles remain only between their members. Microtargeted content may rarely reach individuals outside the intended audience. As a result, not only are uninformed audiences excluded, but the content also receives less scrutiny from those who hold opposite views.

When the same outlet is open to all listeners, the overlap is large enough to make salient information accessible to the public as a whole. Individualized manipulated outlets are not posted anywhere for the general public to see, invisible to the public as a whole. Microtargeted messaging selectively exposes individuals to persuasive content tailored to influence their choices and decisions at the individual level, thus restricting users' access to critical political information, such as campaign promises and claims made by politicians, and effectively, erode the sphere for public discourse. Such methods not only undermine democratic values such as autonomy and the ability to make decisions based on personal preferences and values, but also

183. Bowman v. UK, App. No. 24839/94, ¶ 47, (19 February 1998), https://hudoc.echr.coe.int/fre?i=001-58134 [perma.cc/E9C5-FSP7].

184. Miller, *supra* note 7, at 6.

185. Miller, *supra* note 7, at 6; Howdle, *supra* note 3, at 451–52.

186. Howdle, *supra* note 3, at 447.

187. *Id.* at 450.

188. *Id.* at 454.

inhibit the possibility of genuine public discourse, because some users are never granted access to the same scope of information as their fellow members. When users discuss reality, they address different factual bases, where the mere existence of global, political, or scientific phenomena are questionable, unaware that the distribution of the information they rely on is being manipulated, not necessarily by humans. This environment of "the tower of Babel outlets," leaves less and less room for public sphere with a clear public discourse. In such a scenario, voters may legitimately complain that their understandings of the issues were partial and lacked deliberation, compromising the democratic process and its legitimacy.[189] Hence, both mass amplification and micro-targeting, by employing manipulated partial exposure designed to prevent the public from accessing information, work in anti-democratic ways.

B. Content Generation and Corresponding Legal and Ethical Responsibilities, Obligations, and Accountabilities

Traditional newspaper and news broadcasting typically involve controlled content generation by professional reporters and editors, albeit with less control over distribution of prints or broadcasts. Once the content is published, it is open to all. Anyone can see it and pass it on. The information flow in traditional media is mostly one-way, from professional journalists and news companies to a listening audience. Individuals receive content controlled by editors, without active participation or the ability to directly contribute to the content. The freedom of press granted by the First Amendment protects editorial decisions of newspapers on what to print and what not to print.[190] But in traditional media, the industry (both publishers and journalists) undertake legal duties and responsibilities for professional reporting.

Providing accurate and relevant news coverage and facilitating a well-informed citizenry is a fundamental duty of newspaper publishing companies, aimed to serve the public interest by informing the public in a meaningful way. A newspaper publishing company must adhere to defamation laws, privacy laws, copyright laws, and intellectual property laws. For example, journalists must avoid reporting information they know to be false that harms the reputation of an individual or organization, and publishers have a duty to avoid publishing such statements.[191]

---

189. *Id.* at 452.

190. *See* Miami Herald Pub. Co. v. Tornillo, 418 U.S. 241 (1974). In *Miami Herald*, the Miami Herald challenged a Florida statute that required newspapers to allow political candidates who had been criticized by the press the right to have their responses to the criticisms published by the outlet that criticized them. *Id.* The newspaper claimed that such a requirement is a violation of the free press clause of the First Amendment. *Id.* The circuit court held the statute unconstitutional as infringing on the freedom of the press and dismissed the action. *Id.* The Florida Supreme Court reversed, holding that the statute did not violate constitutional guarantees, and that civil remedies, including damages, were available. *Id.* The Supreme Court stated that the First Amendment erects an insurmountable barrier between government and print media, insofar as government tampering, in advance of publication, where news and editorial content is concerned. *Id.* at 259 (White, J., concurring). Many tech companies use the editorial decision defense in *Miami Herald* to validate their First Amendment protection when it comes to defending their decision to choose what would be published on their platform. This principle extends even to cases that lead to harmful consequences. *See* Daphne Keller, *Platform Transparency and he First Amendment*, 4 J. FREE SPEECH L. 1, 41 (2023).

191. In the past, TV and radio broadcasters had to carry politically balanced coverage and opinion under the FCC's fairness doctrine, which was abolished. *See* Francis Fukuyama, *Making the*

Threatening, harassing, or defamatory speech would usually be subject to civil or, under certain circumstances, criminal liability.[192] Content created by a third party lays secondary liability on the publisher.[193] Individuals or groups affected by the content may sue the print industries for damages through tort law if the published statements are proven to be false and damaging, subject to First Amendment limitations on libel suits brought by public figures.[194] Additionally, states' privacy laws generally require journalists and publishers to respect an individual's right to privacy, including avoiding the unauthorized publication of private information, intrusion into someone's personal life, and the publication of embarrassing or private facts without legitimate public interest.[195] Obviously, publishers must avoid publishing classified or sensitive national security information that could potentially harm the country.[196] They must refrain from publishing copyrighted materials without permission and obtain appropriate licenses or permissions when using copyrighted works, including text, images, music, videos, and the like.[197] Publishers that include advertisements in their publications must adhere to advertising laws

---

Internet Safe for Democracy, 32 J. DEMOCRACY 37, 40-41 (2021). The constitutionality of this intrusion into private speech was challenged in the 1969 case *Red Lion Broadcasting Co. v. FCC*, 395 U.S. 367 (1969). *Id.* The Supreme Court upheld the commission's authority to compel a radio station to carry replies to a conservative commentator. *Id.* Republican presidents repeatedly vetoed Democratic attempts to turn it into a statute, and the FCC itself rescinded the doctrine in 1987. *Id.*

192. *See, e.g.*, Kristi Nickodem, *Dealing with Harassment and Threats Towards Local Government Officials and Employees*, COATES' CANONS, NC LOCAL GOVERNMENT LAW, UNC (July 28, 2022), https://canons.sog.unc.edu/2022/07/dealing-with-harassment-and-threats-towards-local-government-officials-and-employees/ [perma.cc/F9NK-39JL] (discussing possible courses of action for public servants receiving threats or harassing messages).

193. *See* Stratton Oakmont, Inc. v. Prodigy Servs. Co., No. 31063/94, 1995 WL 323710 (N.Y. Sup. Ct. May 24, 1995).

194. *See, e.g.*, N.Y. Times Co. v. Sullivan, 376 U.S. 254 (1964) (demonstrating an analysis of the "actual malice" standard under an Alabama libel law); Gertz v. Welch, Inc., 418 U.S. 323, 351 (1974) (finding that a newspaper who publishes falsehoods about an individual who is neither a public figure nor a public official may not claim a constitutional privilege against liability for injury inflicted by those statements). *See also Balancing Act: Free Speech & Misinformation*, CIVIC GENIUS (Mar 2, 2022), https://www.ourcivicgenius.org/learn/balancing-act-free-speech-misinformation/ [perma.cc/ZM8Y-J3L5]. Defamation refers to any statement that harms the reputation of an individual, group, or organization. *Id.* Internet defamation occurs when a defamatory statement is published on the internet, including social media platforms, websites, and blogs. *Id.* Individuals who believe they have been the victim of online defamation can file a lawsuit under the laws of their state. *Id.* To prove defamation, the plaintiff must show that the statement was false, that it was communicated to others, and that it caused harm to their reputation. *Id.*

195. *See Publishing Personal and Private Information*, Digital Media Law Project (Sept. 10, 2023), https://www.dmlp.org/legal-guide/publishing-personal-and-private-information [perma.cc/U9NU-W3DW].

196. Ojan Aryanfard, *National Security*, THE FIRST AMENDMENT ENCYCLOPEDIA, FREE SPEECH CTR. M. TN. STATE UNIV. (August 7, 2023), https://firstamendment.mtsu.edu/article/national-security/ [perma.cc/24T2-FFR9] ("Despite the absolute language of the First Amendment, wars, threats of wars, and perceived risks to national security have prompted the government to, at times, restrict freedom of speech and other First Amendment freedoms throughout U.S. history.").

197. *See* Geoffrey P. Hull, *Copyright*, THE FIRST AMENDMENT ENCYCLOPEDIA, FREE SPEECH CTR. M. TN. STATE UNIV. (August 7, 2023), https://firstamendment.mtsu.edu/article/copyright/ [perma.cc/L5S3-EDH5]. Copyright, by its nature, restricts speech to some extent. *Id.* Copyright owners possess a bundle of rights to their literary and artistic work. *Id.* 17 U.S.C. covers copyright law. *Id.*

and regulations.[198] They should avoid deceptive or misleading advertisements and comply with guidelines from the Federal Trade Commission (FTC) regarding disclosure requirements.[199] Finally, publishers may be subject to various regulations depending on the nature of their publications, such as those related to specific industries (e.g., finance, health, or pharmaceuticals) or targeted audiences (e.g., children).[200]

In addition to legal duties, traditional news providers also face professional duties. In liberal democracies, news providers and journalists typically adhere to specific professional content production standards, which set them apart from other content creators.[201] News outlets have developed voluntary codes of ethics and guidelines to promote journalistic integrity and professional standards to aim for objectivity.[202] Other news organizations and media outlets have their own sets of guidelines and ethical policies that their journalists are expected to follow. Editorial policies and guidelines often encompass issues such as standards of reporting, conflicts of interest, correction procedures, and the separation of news and opinion sections. They provide a professional norm that news organizations should aim to present a comprehensive, balanced and unbiased view of events; seek diverse perspectives voices and communities; and provide fair coverage of different viewpoints.[203] Publishing companies and journalists must report accurate and credible information, and should have processes to ensure that the news articles are based on reliable sources and accurate information before publication.[204] If errors or inaccuracies are found in reporting, journalists should correct them promptly and

198. *Truth in Advertising*, FED. TRADE COMM'N, https://www.ftc.gov/news-events/topics/truth-advertising [perma.cc/CK2Q-7JAB] (last visited Dec. 27, 2023). Truth-in-advertising laws are enforced by the FTC. *Id.*

199. *Id.*

200. *See, e.g.*, Brown v. Ent. Merchs. Ass'n, 564 U.S. 786 (2011) (assessing California's interest in protecting children from violent video games).

201. *See generally What the Codes Say: Code Provisions by Subject*, SOC'Y PRO. JOURNALISTS, https://www.spj.org/ethicscode-provisions.asp [perma.cc/WJ8Z-CWDG] (last visited Dec 27, 2023).

202. *See, e.g.*, *SPJ Code of Ethics*, SOC'Y PRO. JOURNALISTS, https://www.spj.org/ethicscode.asp [perma.cc/2EMS-9HDV] (last visited Dec. 27, 2023) (describing a large organization that represents journalists in the United States, whose Code of Ethics outlines principles such as seeking truth and accuracy, minimizing harm, acting independently, and being accountable and transparent); *News Values and Principles*, ASSOC. PRESS, https://www.ap.org/about/news-values-and-principles/ [perma.cc/Z6A5-TEHE] (last visited Dec. 27, 2023) ("Associated Press (AP) agency has developed a comprehensive set of standards and guidelines that cover various aspects of journalism, including accuracy, fairness, objectivity, and responsible sourcing."); *About Us*, NEWS LEADERS ASS'N, https://members.newsleaders.org/about-us [perma.cc/KCG5-7C8C] ("Dedicated to promoting journalism excellence, American Society of News Editors (ASNE) has published principles of ethical conduct for journalists, which emphasize accuracy, fairness, accountability, and transparency."); *Code of Ethics*, RADIO TELEVISION DIGIT. NEWS ASS'N, https://www.rtdna.org/ethics [perma.cc/A3EN-Z7R7] (last visited Dec. 27, 2023) (The Radio Television Digital News Association (RTDNA), has established codes of ethics that guide journalists in their work.); *NPR Ethics Handbook*, NPR, https://www.npr.org/series/688409791/npr-ethics-handbook [perma.cc/FT8U-G8TE] (last visited Dec. 27, 2023) (providing guidance to NPR journalists and staff, covering topics such as accuracy, fairness, transparency, conflicts of interest, and maintaining the trust of the audience).

203. *SPJ Code of Ethics*, *supra* note 202.

204. *NPR Ethics Handbook: Accuracy*, NPR https://www.npr.org/about-npr/688139552/accuracy [perma.cc/FL74-4QN6] (last visited Dec. 27, 2023) ("Our purpose is to pursue the truth. Diligent verification is critical.").

transparently.[205] In order to remain impartial, the journalist should avoid undue influence or censorship from outside sources such as governments, owners, advertisers, sponsors, or personal biases that would result in conflicts of interest (which should be disclosed when they exist).[206] Moreover, journalists must be sensitive to the impact of their reporting on individuals, particularly in cases of tragedy or trauma, and avoid unnecessary harm.[207] While not legally enforceable, these professional and ethical responsibilities are crucial for maintaining public trust and credibility in journalism.

In contrast, digital media platforms tend to have less control over the generation of content, which can be generated practically by any user. However, platforms have the capability and distribution methods to control the exposure to particular content more precisely, on an individual level according to users' profiles. Social media platforms offer non-professional users the ability to actively engage with content, share opinions, and participate in discussions. Anyone and everyone can serve as an information source through a blog, X account, or Facebook group without having taken a single course on basic journalistic standards or the fundamentals of American government.[208] Establishing shared facts and fostering diverse perspectives is not required for discussion. "Popularity, advertising interests, and commercial agreements have become the primary drivers for prominence online, rather than integrity and diversity of news and information."[209] Reporting is often anecdotal and may limit the depth of coverage on certain topics. It often leads to compromising high levels of accuracy, which make it difficult to discern the reliability and credibility of information or the identity of content creators.[210]

Relatedly, the commercial entities responsible for personalization features in social media are bound by fewer, if any, legal or professional obligations to the information distributed than traditional media. Since they are not involved in investigative journalism or reliant on confidential sources in the same way as newspapers, the duty to protect sources may not directly apply to them. Individuals can sue the person who posted defamatory, violent, or harassing speech, but not the platform where such speech is published.[211] Moreover, unlike traditional journalists who follow a professional ethics code, neither users nor platforms are committed by a meta-framework that all communities necessarily believe in, governing ethical or legal principles such as human rights, equality, or democratic principles. On the contrary, competition with an increasing number of voices, creators, and producers who do not consistently uphold these standards makes it hard to maintain credibility.[212]

---

205. *Code of Ethics*, *supra* note 202 ("Ethical journalism requires owning errors, correcting them promptly and giving corrections as much prominence as the error itself had.").

206. *Id.* ("[I]ndependence from influences that conflict with public interest remains an essential ideal of journalism.").

207. *Id.* ("Minimizing harm, particularly to vulnerable individuals, should be a consideration in every editorial and ethical decision.").

208. Rhodes, *supra* note 69, at 4.

209. PLURALISM OF NEWS, *supra* note 5, at 31.

210. *Id.* at 28 (showing, for example, how media organizations engaged in partisan conflicts attract significant audiences and can be highly profitable).

211. *See supra* text accompanying note 191.

212. PLURALISM OF NEWS, *supra* note 5, at 28.

In recent years, regulators have tried to level the playing field—not always successfully—between media organizations and platforms.[213] Illustrative examples of relevant interventions can be found around the world in (1) media plurality and diversity frameworks,[214] (2) the commercial relationships between news providers and platforms,[215] and (3) attempts at addressing information disorder.[216] Furthermore, the Guiding Principles on Business and Human Rights have set forth global standards of expected conduct for companies, across geographical boundaries.[217] But creation of legislation is challenging, because it requires establishing whether social media platforms should be treated as common carriers, editors, or publishers.[218] Big tech companies analogize their practice to bulletin boards, unable to control users' content, rather than traditional media, which reviews or edits prior to publishing and can ensure the accuracy and quality of content that is posted on their sites.[219]

The Communications Decency Act of 1996 classifies platforms as "interactive computer services" rather than "publishers."[220] Under this classification, Section 230 allows users to freely post user generated content on internet platforms, while delegating platforms (interactive computer services)[221] the power of judgment

213. *See* Villasenor, *supra* note 27 ("In 1987, the year the fairness doctrine was abolished, the U.S. broadcast market had grown to include 'more than 1,300 television stations and more than 10,000 radio stations' . . ."). *See also* Robert D. Hershey Jr., *F.C.C. Votes Down Fairness Doctrine in a 4-0 Decision*, N.Y. TIMES, Aug. 5, 1987, at A1, https://www.nytimes.com/1987/08/05/arts/fcc-votes-down-fairness-doctrine-in-a-4-0-decision.html [perma.cc/GSE5-BKNB].

214. *Guiding Principles on Diversity of Content Online*, GOV'T OF CAN. (June 2021), https://www.canada.ca/en/canadian-heritage/services/diversity-content-digital-age/guiding-principles.html [perma.cc/8ZKT-3WW4].

215. *The Online News Act*, GOV'T OF CAN., https://www.canada.ca/en/canadian-heritage/services/online-news.html [perma.cc/JZ66-YZTT] (last visited Oct. 9, 2022); *News Media Bargaining Code*, AUSTRALIAN COMPETITION & CONSUMER COMM'N, https://www.accc.gov.au/focus-areas/digital-platforms/news-media-bargaining-code [perma.cc/5UZ6-72TP] (last visited Oct 9, 2022).

216. *See e.g.*, *The Journalism Trust Initiative*, JOURNALISM TR. INITIATIVE, https://www.journalismtrustinitiative.org [perma.cc/3A2R-UG73] (last visited Dec. 27, 2023). The Journalism Trust Initiative has developed a standardization instrument for evaluating news media organizations. This instrument is in line with ISO protocols and is published by the European Committee of Standardization. DIGITAL INDUSTRIAL GROUP INC., AUSTRALIAN CODE OF PRACTICE ON DISINFORMATION AND MISINFORMATION (2021), https://digi.org.au/wp-content/uploads/2021/10/Australian-Code-of-Practice-on-Disinformation-and-Misinformation-FINAL-WORD-UPDATED-OCTOBER-11-2021.pdf [perma.cc/8NGG-RX9T].

217. United Nations Human Rights Office of the High Commissioner, Guiding Principles on Business and Human Rights (June 16, 2011), https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf [perma.cc/29EW-JJDR] (on file with the U.C. Irvine Law Review).

218. GILLESPIE, *supra* note 30.

219. Christopher Kullenberg, Frauke Rohden, Anders Björkvall, Fredrik Brounéus, Anders Avellan-Hultman, Johan Järlehed, Sara Van Meerbergen, Andreas Nord, Helle Lykke Nielsen, Tove Rosendal, Lotta Tomasson & Gustav Westberg, *What Are Analog Bulletin Boards Used for Today? Analysing Media Uses, Intermediality and Technology Affordances in Swedish Bulletin Board Messages Using a Citizen Science Approach*, 13.8 PLOS ONE 1 (2018).

220. 47 U.S.C. § 230(f) ("[A]ny information service, system, or access software provider that provides or enables computer access by multiple users to a computer server . . . ."). *See* 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, 407–11 (2010) (chronicling the history of the lead-up to Section 230, including the Stratton Oakmont, Inc. v. Prodigy Servs. Co. decision).

221. 47 U.S.C. § 230(f) ("[A]ny information service, system, or access software provider that provides or enables computer access by multiple users to a computer server, including specifically a

regarding what they allow published on their sites.[222] Free expression is often considered a democratic value as it promotes the principles of individual liberty and the free flow of information and ideas. In the United States, and in any democratic society, free speech and expression without fear of censorship or punishment is a primary right, a founding principle that weaves the political fabric of the nation.[223] Under the First Amendment, governmental interference restricting certain users' content, albeit privately administered, would be considered censorship infringing free speech, and thus unconstitutional.[224] Section 230 therefore does not limit the freedom of speech, and at the same time, enables platforms to establish their own voluntary content-related self-regulation, compete over rules for their communities, and enforce them (even imperfectly), therefore avoiding practical and constitutional problems of government supervision of content. Section 230 exempts platforms from liability for content moderation decisions done in "good faith" and provides them immunity,[225] whether they over- or under-remove user generated content.[226]Most online platforms, through terms of service/terms of use, clarify community standards regarding the types of content that are allowed or prohibited on their platforms, what forms of expression are acceptable, and which type of expression will be removed.[227] Platforms' terms of use do not reflect a specific legal system but often overlap with local law regarding illegal content online to prevent harm and liability. Social media platforms are expected to have mechanisms to remove or moderate content that violates community standards or terms of service—especially if it is reported by users or deemed to be in violation of applicable laws and regulations related to defamation, copyright, privacy, and

---

service or system that provides access to the Internet and such systems operated or services offered by libraries or educational institutions.").

222. Adam Holland, Chris Bavitz, Jeff Hermes, Andy Sellars, Ryan Budish, Michael Lambert & Nick Decoster, *Intermediary Liability in the United States*, GLOB. NETWORK INTERNET & SOC'Y RSCH. CTRS., https://publixphere.net/i/noc/page/OI_Case_Study_Intermediary_Liability_in_the_United_States [perma.cc/8672-ZTXM]. ("[T]o apply Section 230's protection [and avoid legal repercussions for removing or limiting content], a defendant must show (1) that it is a provider or user of an interactive computer service; (2) that it is being treated as the publisher of content (though not with respect to a federal crimes, intellectual property, or communications privacy law); and (3) that the content is provided by another information content provider.").

223. Kyler Baier, Note, *Replacing What Works With What Sounds Good: The Elusive Search For Workable Section 230 Reform*, 26 ILL. BUS. L. J. 40, 52 ("The first and last consideration must be free speech.").

224. Keller, *supra* note 33, at 238.

225. While the CDA does not explicitly define "good faith," courts have interpreted it as reasonable and sincere efforts to address objectionable or unlawful content without engaging in intentional misconduct or acting with malicious intent. Courts consider factors such as the platform's moderation efforts, consistency in policy enforcement, and the reasonableness of their actions when determining if a platform has acted in good faith.

226. James Grimmelmann, *The Virtues of Moderation*, 17 YALE J.L. & TECH. 42, 103 (2015) (immunizing moderators "both for the content they moderate and the content they miss").

227. Jon Bateman, Natalie Thompson & Victoria Smith, *How Social Media Platforms' Community Standards Address Influence Operations*, CARNEGIE ENDOWMENT FOR INT'L PEACE (Apr. 1, 2021), https://carnegieendowment.org/2021/04/01/how-social-media-platforms-community-standards-address-influence-operations-pub-84201 [perma.cc/B8KC-6238] (explaining that problems caused in recent years by the Cambridge Analytica targeting, the COVID-19 pandemic, the protests of George Floyd's murder, and the contestation and insurrection related to the 2020 U.S. election lead internet platforms to moderate what content can be posted on their sites); Evelyn Douek, *Governing Online Speech: From "Posts-As-Trumps" to Proportionality and Probability*, 121 COLUM. L. REV. 759, 800–04 (2021).

intellectual property rights. They may employ fact-checking measures, promote reliable sources to combat misinformation, identify false or misleading information, and prohibit content promoting hate speech, violence, or discrimination. In this spirit, platforms like Facebook,[228] YouTube,[229] and X[230] have taken voluntary self-regulated mechanisms to promote reliability in news stories, demote conspiracy-related content, and flag factually false tweets based on user feedback.

This does not end the dilemma. Platforms enjoy immunity as long as the service provider does not act as the publisher or author of the statement,[231] responsible for the "development" of the content at issue.[232] Courts consider the publisher liable only if they "know[] or ha[ve] reason to know" of the statement's defamatory or tortious character, or if online platforms present information or transactions in a way that would lead a consumer to believe that the online platform itself were providing the information or service.[233] A similar regulation exists in the European Union (EU) as well.[234] Platforms are confronting the moderator dilemma.

---

228. *See* Emily Dreyfuss & Issie Lapowsky, *Facebook Is Changing News Feed (Again) to Stop Fake News*, WIRED (Apr. 10, 2019, 1:00 PM), https://www.wired.com/story/facebook-click-gap-news-feed-changes/ [perma.cc/8T8K-4ZUH].

229. *See, e.g.*, MARC FADDOUL, GUILLAUME CHASLOT & HANY FARID, A LONGITUDINAL ANALYSIS OF YOUTUBE'S PROMOTION OF CONSPIRACY VIDEOS (2020), https://arxiv.org/pdf/2003.03318.pdf [perma.cc/5JWE-T4Y5].

230. Gilad Edelman, *Twitter Finally Fact-Checked Trump. It's a Bit of a Mess*, WIRED (May 27, 2020, 12:21 PM), https://www.wired.com/story/twitter-fact-checked-trump-tweets-mail-in-ballots/ [perma.cc/Q5JG-M8YE].

231. Berin Szóka & Ari Cohn, *The Wall Street Journal Misreads Section 230 and the First Amendment*, LAWFARE (Feb. 3, 2021, 3:43 PM) https://www.lawfaremedia.org/article/wall-street-journal-misreads-section-230-and-first-amendment [perma.cc/N9YB-8TQK]. Actions that would hold platforms accountable include cases when the intermediary creates content itself, editorial functions of a third-party content and materially altering its meaning to make it actionable (remove, or edit content); paying a third party to create or submit content; allowing users to respond to forms or drop-downs to submit content; promises to remove material and keeping content online even after being notified the material is unlawful. *See* Adam Holland, Chris Bavitz, Jeff Hermes, Andy Sellars, Ryan Budish, Michael Lambert & Nick Decoster, *Intermediary Liability in the United States*, GLOB. NETWORK INTERNET & SOC'Y RSCH. CTRS., nn.42–51 and accompanying text, https://publixphere.net/i/noc/page/OI_Case_Study_Intermediary_Liability_in_the_United_States [perma.cc/C6RU-LVHP].

232. 47 U.S.C. § 230(f)(3) ("[A]ny person or entity that is responsible, in whole or in part, for the creation or development of information provided through the Internet or any other interactive computer service.").

233. Alan Z. Rozenshtein, *Interpreting the Ambiguities of Section 230*, BROOKINGS (Oct. 26, 2023), https://www.brookings.edu/articles/interpreting-the-ambiguities-of-section-230/ [perma.cc/36C5-25TU]. The court in *Stratton Oakmont* liable because it had voluntarily assumed an editorial role by moderating and screening messages on its bulletin board. The court further stated that Prodigy could have avoided liability if it had a policy of removing defamatory messages upon receiving actual knowledge of their defamatory content. However, because Prodigy did not have such a policy in place, the court held that Prodigy was liable for the defamatory message. This case set an important precedent for online service providers and clarified the standard for liability in cases of online defamation. It established that online service providers can be held liable for defamatory content if they have actual knowledge of its defamatory nature and fail to remove it. RESTATEMENT (SECOND) OF TORTS § 581. *See also* Steinbuch v. Hachette Book Grp., 2009 WL 963588, at *3 (E.D. Ark. April 8, 2009); Lee v. Penthouse Int'l Ltd., 1997 WL 33384309, at *8 (C.D. Cal. March 19, 1997).

234. In the European Union, in order to establish that a referencing service provider is neutral and its liability may be limited under Article 14 of the E-Commerce Directive. Céline Castets-Renard, *Algorithmic Content Moderation on Social Media in EU Law: Illusion of Perfect Enforcement*, U. ILL. J.L. TECH. & POL'Y 283, 296 (2020). *See* Cases C-236/08 to C-238/08, Google France v. Louis Vuitton, ECLI:EU:C:2010:159, ¶ 114 (Mar. 23, 2010) ("[I]t is necessary to examine whether the role played by that service provider is neutral, in the sense that its conduct is merely technical, automatic and passive,

By editing, changing or interfering with users generated content on a voluntary basis, the online platform may be deemed to be playing an active role and therefore lose the benefit of the liability exemption.[235] Therefore, without the protections afforded by Section 230, platforms would have little incentive to engage in content moderation, as any such action could render them "aware or having a reason to be aware" of the user-generated content.

The Section 230 exemption for internet platforms has historic justifications, normative and practical. The law has never been about platforms' lack of interference with speech, but rather about promoting innovation. Historically, when it was not logistically feasible to support a large number of small communications firms, the United States government entered into "regulatory deals" with select providers. In exchange for intermediary immunity, these firms agreed to terms intended to serve the public interest, including adopting non-discriminatory policies, servicing unprofitable markets, and assuming additional liability.[236] It is unclear whether these commitments still stand and whether platforms should be considered neutral carriers or active actors who should be liable for personalizing algorithms, but this question remains beyond the scope of this Article.[237]

The above-mentioned differences between traditional media and digital social media (the contribution of personalization features to the public discourse, prioritizing and matching content, undertaking editorial functions, and the corresponding legal liability) reflect tensions between certain business models and

---

pointing to a lack of knowledge or control of the data which it stores."); Case C-324/09, L'Oreal et al. v. eBay, ECLI:EU:C:2011:474, ¶ 116 (July 12, 2011) ("Where, by contrast, the operator has provided assistance which entails, in particular, optimising the presentation of the offers for sale in question or promoting those offers, it must be considered not to have taken a neutral position between the customer-seller concerned and potential buyers but to have played an active role of such a kind as to give it knowledge of, or control over, the data relating to those offers for sale. It cannot then rely, in the case of those data, on the exemption from liability. . . ."); Case C-484/14, Mc Fadden v. Sony Music Ent., ECLI:EU:C:2016:689, ¶ 62 (Sept. 15, 2016). Joined Cases C-682/18 & C-683/18, Peterson v. Google, ECLI:EU:C:2021:503, sec. 10(42) (June 22, 2021) ("The exemptions from liability established in this [d]irective cover only cases in which the activity of the information society service provider is . . . of a mere technical, automatic and passive nature, which implies that that service provider has neither knowledge of nor control over the information which is transmitted or stored.").

235. ALEXANDRE DE STREEL, ELISE DEFREYNE, HERVÉ JACQUEMIN, MICHÈLE LEDGER & ALEJANDRA MICHEL, ONLINE PLATFORMS' MODERATION OF ILLEGAL CONTENT ONLINE: LAW, PRACTICES AND OPTIONS FOR REFORM 20 (2020); *see also* Celine Castets-Renard, *supra* note 234, at 308 (arguing that "voluntary monitoring could generate awareness and knowledge of facts or circumstances from which the illegal activity or information is apparent and known by the platform. In this case, the platform must act promptly and remove the illegal content. Consequently, platforms could lose the benefit of the liability exemption regime. Therefore, concerns related to losing the benefit of the liability exemption should neither deter nor preclude the application of the effective proactive voluntary measures that this Communication seeks to encourage." (footnotes omitted)).

236. Adam Candeub, *Bargaining for Free Speech: Common Carriage, Network Neutrality, and Section 230*, 22 YALE J.L. & TECH. 391, 396, 407–08, 412–13 (2020).

237. FEDERAL TRADE COMMISSION, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE: RECOMMENDATIONS FOR BUSINESSES AND POLICYMAKERS 56 (2012) http://www.ftc.gov/os/2012/03/120326privacyreport.pdf [perma.cc/Z34H-GWFM] (expressing particular concern about large platform providers such as Internet Service Providers (ISPs), operating systems, or browsers, the FTC determined that "while companies such as Google and Facebook are expanding their reach rapidly, they currently are not so widespread that they could track a consumer's every movement across the Internet."). For The Dual Identity Problem: Publisher Speech vs. Platform Immunity, see Sharon Bassan, *Transparency ≠ Accountability? Rethinking Voluntary Vs. Mandatory Content Moderation Reports*, SSRN (Feb. 18, 2025), https://ssrn.com/abstract=5143075.

the public interest in public discourse. Social media platforms are classified as interactive service providers, rather than publishers or editors. Unlike traditional media, they do not exert editorial control over users' content and,[238] paradoxically, despite the contribution of personalization features to the nature of public discourse, social media platforms are not held legally responsible for the consequences of that lack of control.[239] This happens in a landscape, where power dynamics between the public and private sectors have shifted. The growing influence of private, public, non- and quasi-governmental has positioned this sector as a near-monopoly in the provision of certain services, blurring the boundaries between them to the extent that the government is no longer the dominant center of power.[240] Platforms at times exercising authority over individuals comparable to that of the state. This shift in power and messaging raises concerns similar to those associated with government control of free speech. While liberal, democratic governments and traditional media often have limits over their powers to control speech, corporations do not.[241] Platforms are private entities in the commercial sphere who are free to decide how content is matched to users as they see fit, with all the implications such freedom entails.[242]

---

238. Subcommittee on Antitrust, Commercial and Administrative Law, *Heads of Facebook, Amazon, Apple & Google Testify on Antitrust Law*, C-SPAN (July 29, 2020), https://www.c-span.org/video/?474236-1 [perma.cc/S5QM-7NSQ] (Representative David Cicilline (D-RI): "When a television station runs a false political advertisement, they're held liable for that. Why should Facebook or any other platform be different? While you may not be a publisher, you're responsible maybe not for the first posting, but you then take that posting and you apply a set of algorithms that decide how you will disseminate that, which is a business decision, not a first amendment decision. And it's hard to understand why Facebook shouldn't be responsible for those business decisions.").

239. Force v. Facebook, Inc., 934 F.3d 53 (2nd Cir. 2019) (holding that Facebook was not liable even though its algorithms helped terrorists collaborate to attack U.S. citizens in Israel. Chief Judge Robert Katzmann concurred in part but criticized the algorithms as not being neutral, given the hate-based linkage, and cited data showing social media algorithms have contributed to political polarization.). *But see* Kende, *supra* note 60, at 283 (arguing that some studies have taken the opposite view, though these studies are problematic).

240. Allenby, *supra* note 89, at 422–23.

241. Scholars and individuals cognizant of historical power structures, encompassing government authority in regulating free expression, either through media oversight or through the governance of media under authoritarian regimes, should harbor analogous apprehensions concerning the manipulation exercised by platforms that derive financial gain from catering to clients' requests to distort information, irrespective of the democratic or non-democratic nature of said clients. The degree of disquiet is a matter of uncertainty, as both forms of manipulation warrant significant concern. However, it is evident that both the presence and absence of legislative interventions may empower one of the stakeholders to exert greater influence over the shaping of our perception of political events.

242. Christopher E. Peterson, User-Generated Censorship: Manipulating the Maps of Social Media (June 2013) (Ph.D. dissertation, Massachusetts Institute of Technology), https://dspace.mit.edu/bitstream/handle/1721.1/81132/858280891-MIT.pdf?sequence=2&isAllowed=y ("Case studies reveal that these platforms, far from being neutral pipes through which information merely travels, are in fact contingent sociotechnical systems upon and through which users effect their politics" by strategically pulling the levers which make links to sites more or less visible. The tools designed to help make information more available have been repurposed and reversed to make it less available.); *see also* GILLESPIE, *supra* note 30, at 7 (describing in detail how Facebook and other social media platforms comprehensively filter, sort, and structure the content that flows between users); Oremus et al., *supra* note 138 ("One takeaway is that Facebook's algorithm isn't a runaway train. The company may not directly control what any given user posts, but by choosing which types of posts will be seen, it sculpts the information landscape according to its business priorities.").

III. CURRENT REGULATORY INITIATIVES ADDRESSING PERSONALIZATION FEATURES

The United States does not have comprehensive federal legislation specifically targeting personalization features. However, various discussions, proposals, and initiatives address some of the above-mentioned harms. Current regulatory approaches suggest that societal concerns centered on filter bubbles, echo chambers, extreme content, and hate crimes, with a primary focus on the discourse enabled by personalization features. However, this Article suggests that these phenomena may be symptoms of a deeper structural problem: the erosion and fragmentation of public sphere discourse, and the unprecedented capacity to manipulate individuals within society on a personalized basis. When reviewing current regulatory initiatives, several relevant questions arise: At what stage of the process should regulation interfere—data collection, algorithm processing (content-neutral), or content output? Should moderation be governmentally mandated or part of platforms' self-regulation? This Part examines three regulatory models: content regulation, content-neutral regulation, and data-collection focused regulation. It reviews the rationales and limitations of each model, as well as their potential to fix the previously discussed problems.

A. Content Regulation

Content regulation focuses on the nature of the material that constitutes online discourse, including content matched to users through personalization features, as well as amplified narratives.[243] This Part therefore explores the kind of content to be personalized and targeted at users.

Not everything is allowed to be published. Some categories of speech may be restricted or prohibited.[244] Courts have shaped the scope of the right to publish certain speech, primarily by emphasizing what publishers may lawfully decline to publish. For example, the publisher's refusal to disseminate content that incites imminent lawless action,[245] print articles in a school newspaper over administrative objections,[246] or promote illegal drug use at a school-sponsored event[247] does not constitute a First Amendment violation. Additionally, while free speech protections

243. Jacobs, *supra* note 17, at 536.

244. At one end of the scope, regulation could limit particular types of communications that threaten democratic-pluralistic values. This could be done without the listeners ever being exposed to these expressions, which may decrease exposure to pluralistic ideas, or after first having heard them, by blocking or removing them. Alternatively, regulation supporting autonomous decision-making could accommodate speech according to users' preferences or strengthen users' right to make decisions regarding sharing their information in order to decrease the possibility to profile and personalize. Finally, promoting the opportunity to accommodate any view, the government could refrain from regulating individuals' decision not to listen. *See, e.g.*, European Comm'n, *Extending EU crimes to Hate Speech and Hate Crime*, https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/extending-eu-crimes-hate-speech-and-hate-crime_en [perma.cc/2FS7-TC6P]; *see also* Ellen P. Goodman & Ryan Whittington, *Section 230 of the Communications Decency Act and the Future of Online Speech* 7-8 (Rutgers L. Sch. Research Paper, 2019), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3458442 [perma.cc/VDE7-UWLL] (suggesting eliminating protection for deep fakes and creating a tort for failure to remove deep fakes (Senator Mark Warner), and for drug trafficking (Senator Joe Machin)).

245. Brandenburg v. Ohio, 395 U.S. 444 (1969).

246. Hazelwood Sch. Dist. v. Kuhlmeier, 484 U.S. 260 (1988).

247. Morse v. Frederick, 551 U.S. 393 (2007).

remain robust, the broad immunity granted by Section 230 is subject to statutory exceptions. Platforms are required to monitor specific categories of illegal speech.[248] Narrow and well-defined exceptions to First Amendment protection include speech covered by federal criminal law, intellectual property law, electronic communications privacy statutes, and sex trafficking legislation (notably expanded by the 2018 FOSTA-SESTA Act)).[249] These categories fall outside the constitutional shield traditionally afforded to publishers.

The delivery of certain types of content might be restricted, based on factors such as the reach or invasiveness of the content, as in traditional broadcasting (i.e., the legal doctrine of obscenity).[250] The doctrine of obscenity suggests that preventing the spread of certain types of content may constitute a valid justification for restricting content, even under the First Amendment. The Communications Decency Act of 1996 regulates obscene or offensive online content and criminalizes certain activities, such as broadcasting obscene language,[251] prohibiting the production or transfer of obscene matter with intent to distribute or sell, including through interactive computer services,[252] or distribution of obscene content via radio communications, cable, or subscription television.[253] Similarly, the Communications Decency Act, as amended by the PROTECT Act of 2003,[254] determines that it is a criminal offense to use a telecommunications device to send or display obscene or indecent content to a recipient who is under 18 years old.[255] Commercial website operators are legally obligated to take reasonable measures to restrict access by minors to content that is harmful for them.[256] While courts have recognized obscenity as a category of speech unprotected by the Constitution ,[257] the Supreme Court has held that the government may not invoke a broad definition of obscenity or rely solely on the potential harm to minors to restrict speech that is not obscene for adults.[258] Although the government retains authority to penalize the online distribution of obscene material, the emergence of technologies enabling precise audience targeting, which personalization features theoretically allow (e.g., by restricting access to such content only to people above 18 years old), may render

248. For categories of "illegal speech" that Section 230 does not cover, *see, e.g.*, Steven H. Shiffrin, *Racist Speech, Outsider Jurisprudence, and the Meaning of America*, 80 CORNELL L. REV. 43, 50–64 (1994); Steven Heyman, *Spheres of Autonomy: Reforming the Content Neutrality Doctrine in First Amendment Jurisprudence*, 10 WM. & MARY BILL RTS. J. 647, 651 (2002) ("[T]he Court has carved out two exceptions to the neutrality doctrine.")

249. 47 U.S.C. § 230(e)(5).

250. FCC v. Pacifica Found., 438 U.S. 726 (1978).

251. 18 U.S.C. § 1464 (broadcasting obscene language).

252. 18 U.S.C. § 1465 (transportation of obscene matters for sale or distribution); 18 U.S.C. § 1466 (engaging in the business of selling or transferring obscene matter).

253. 118 U.S.C. § 1468 (distributing obscene material by cable or subscription television).

254. Prosecutorial Remedies and Other Tools to End the Exploitation of Children Today Act of 2003, Pub. L. 108-21, 117 Stat. 650 (2003).

255. Communications Decency Act of 1996, 47 U.S.C. § 223(d), *amended by*, the PROTECT Act of 2003.

256. Child Online Protection Act of 1998, 47 U.S.C. § 231.

257. Roth v. United States, 354 U.S. 476, 492 (1957) (holding that "obscenity is not expression protected by the First amendment"); Bethel Sch. Dist. No. 403 v. Fraser, 478 U.S. 675, 685–86 (1986) (holding that the First Amendment did not prohibit schools from prohibiting vulgar and lewd speech since such discourse was inconsistent with the "fundamental values of public-school education").

258. Reno v. Am. C.L. Union, 521 U.S. 844 (1997).

some forms of speech that were previously unlawful more legally defensible.[259]The government could reconsider the scope of legal and illegal speech and mandate platforms to remove certain categories of content. Expanding the categories of illegal content risks undermining open discourse and eroding the foundations of democratic-pluralistic values. It will have several implications: First, courts have struck down prior attempts to outlaw hate speech,[260] "communist political propaganda,"[261] misogynic posts, and false representations,[262] as violations of free speech.[263] Legally, even demonstrating harm is not sufficient to overcome First Amendment protection (except narrowly defined harmful categories like child pornography).[264] The very premise of free speech is that people are not limited to expressing only nice and correct things. Offensive and dangerous expression such

259. United States v. Playboy Ent. Grp., Inc., 529 U.S. 803, 815 (2000) (rejecting the cable pornography-blocking law in *Playboy*, for example, because Congress could have chosen alternate approaches that would have avoided making cable companies block lawful pornography transmissions. The Supreme Court emphasized that if new technologies permit more accurate targeting of unlawful speech, Congress should use them. Noting the "key difference between cable television and the broadcasting media, which is the point on which this case turns: Cable systems have the capacity to block unwanted channels on a household-by-household basis"); *see also* Ashcroft v. Am. C.L. Union, 542 U.S. 656, 667–70 (2004). If other laws establish the possibility of less restrictive responses to similar problems—as the DMCA arguably does for any changes to CDA 230—then that, too, is relevant to courts' scrutiny. Denver Area Educ. Telecomms. Consortium, Inc. v. F.C.C., 518 U.S. 727, 756 (1996) (holding that the law requiring cable companies to segregate offensive content on leased access channels violates the First Amendment).

260. Adam Liptak, *Unlike Others, U.S. Defends Freedom to Offend in Speech*, N.Y. TIMES (June 12, 2008), https://www.nytimes.com/2008/06/12/us/12hate.html [perma.cc/382N-WF5G].

261. Lamont v. Postmaster Gen., 381 U.S. 301, 307 (1965). *Lamont* dealt with Section 305(a) of the Postal Service and Federal Employees Salary Act of 1962, which required the Postal Service to detain "communist political propaganda" mailed into the United States from certain countries and deliver it "only on the addressee's request." *Id.* at 302. The Postal Service implemented the statute by screening all mail from those countries and sending a reply card to the addressee of any piece of mail determined to be statutory agitprop. If the recipient filled out the card and returned it to the Postal Service, it would then deliver the mail; but if the card was not returned within twenty days, the mail would be destroyed. In both cases, the American recipients filed suit challenging the constitutionality of Section 305 rather than return the reply card. The Court held that the statute "amounts in our judgment to an unconstitutional abridgment of the *addressee's* First Amendment rights" (emphasis added). *Id.* at 307.

262. United States v. Alvarez, 567 U.S. 709, 715 (2012).

263. *See, e.g.*, R.A.V. v. City of St. Paul, 505 U.S. 377, 414 (1992) (White, J., concurring) ("The mere fact that expressive activity causes hurt feelings, offense, or resentment does not render the expression unprotected."); Pro-Choice Network v. Schenck, 67 F.3d 377, 395 (2d Cir. 1995), *aff'd in part*, *rev'd in part*, 519 U.S. 357 (1997) (Winter, J., concurring) ("[S]hame is a form of persuasion . . . [that] must be tolerated."); N.A.A.C.P. v. Clairborne Hardware Co., 458 U.S. 886, 910 (1982) ("Speech does not lose its protected *character* . . . simply because it may embarrass others or coerce them into action.") (emphasis added); Int'l Soc'y for Krishna Consciousness, Inc. v. Lee, 505 U.S. 672, 712 (1992) (Souter, J., dissenting) ("The First Amendment inevitably requires people to put up with annoyance and uninvited persuasion."); Eisenstadt v. Baird, 405 U.S. 438, 459 (1972) (Douglas, J., concurring) ("The First Amendment protects the opportunity to persuade to action whether that action be unwise or immoral, or whether the speech incites to action.").

264. Boos v. Barry, 485 U.S. 312, 322 (1988) ("[C]itizens must tolerate insulting, and even outrageous, speech in order to provide 'adequate "breathing space" to the freedoms protected by the First Amendment.'" (quoting Hustler Mag., Inc. v. Falwell, 485 U.S. 46, 56 (1988)).

as racist hate speech,[265] racist fighting words,[266] white supremacist posts,[267] lies,[268] vile and traumatizing remarks,[269] and under certain categories even threats against specific individuals[270] may be still fall under protected speech, unless restrictions pass strict scrutiny.[271] Regardless of its consequences, the Supreme Court has consistently affirmed that such speech is not only constitutionally protected but also that restricting a speaker's access to mass media may itself constitute a violation of their First Amendment rights.[272] Platforms rely on this jurisprudence to justify their continued refusal to censor controversial content.

Second, if platforms are required to moderate content, they are will likely encounter several practical challenges. The vast volume of online content on platforms makes it difficult to assess every item due to the scale of publication.[273] Platforms may delegate content moderation functions to algorithmic or automated detection technologies that filter out speech to deal with this large scale.[274] A growing body of empirical literature demonstrates that automated decision-making systems have limitations in accuracy, and may misidentify legal contexts. In many cases, it is impossible to determine the legality of the content without considering

265. Matal v. Tam, 582 U.S. 218, 246 (2017).

266. *City of St. Paul*, 505 U.S. at 390.

267. *See* The Editorial Board, *We Have a White Nationalist Terrorist Problem*, N.Y. TIMES (Aug. 4, 2019), https://www.nytimes.com/2019/08/04/opinion/mass-shootings-domestic-terrorism.html [perma.cc/PX5K-C9X4]; Daisuke Wakabayashi, *Legal Shield for Websites Rattles Under Onslaught of Hate Speech*, N.Y. TIMES (Aug. 6, 2019), https://www.nytimes.com/2019/08/06/technology/section-230-hate-speech.html?action=click&module=Well&pgtype=Homepage&section=Technology [perma.cc/BZ9U-SLAC].

268. United States v. Alvarez, 567 U.S. 709 (2012).

269. Snyder v. Phelps, 562 U.S. 443 (2011).

270. Elonis v. United States, 575 U.S. 723, 741 (2015); *see also* Counterman v. Colorado, 600 U.S. 66 (2023) (describing what falls into the categories).

271. *See* Barr v. Am. Ass'n of Pol. Consultants, 591 U.S. 610 (2020) (describing strict scrutiny of content-based regulations); Reed v. Town of Gilbert, 576 U.S. 155 (2015) (Justice Thomas clarified that strict scrutiny should always be applied when a law is content-based on its face).

272. The ability to use mass media to disseminate one's message, viewed as a fundamental aspect of the right to free speech, is being protected in a manner similar to the content of the speech itself. Miller, *supra* note 7, at 19–21 (discussing whether one speech amplified to a thousand has the same constitutional weight as a thousand separate speeches or as just one speech act). Speakers have a "freedom of reach" under the First Amendment, and government regulations are scrutinized vis-à-vis what alternative audience they allow speakers to access. Courts often show more concern for individuals or entities with significant resources when it comes to their ability to amplify their speech without government interference. *See* Weinberg v. City of Chicago, 310 F.3d 1029, 1041 (7th Cir. 2002) ("[A]n alternative [audience] is not adequate if it 'foreclose[s] a speaker's ability to reach one audience even if it allows the speaker to reach other groups.'") (footnote omitted). Moreover, government may have even less power to regulate virtual speech. Garnier v. O'Connor-Ratcliff, 41 F.4th 1158, 1181–82 (9th Cir. 2022), *vacated*, 601 U.S. 205 (2024), *abrogated by*, Linde v. Freed, 601 U.S. 187 (2024) (noting that government's complaint that too much social media speech would result in "clutter" was judicially noncognizable as a government interest).

273. Richard A. Wilson & Molly K. Land, *Hate Speech on Social Media: Content Moderation in Context*, 52 CONN. L. REV. 1029, 1064–65 (2021) (exploring the difficulties faced by social media platforms in effectively moderating hate speech due to factors such as the vast scale of user-generated content, the subjective nature of determining what constitutes hate speech, and the balance between freedom of expression and the need to protect users from harm).

274. Alexander Brown, *Averting Your Eyes in the Information Age: Online Hate Speech and the Captive Audience Doctrine*, 12 CHARLESTON L. REV. 1, 34–35 (2017); *see also* Hannah Bloch-Wehba, *Global Platform Governance: Private Power in the Shadow of the State*, 72 SMU L. REV. 27, 29 (2018); Castets-Renard, *supra* note 234, at 294.

its broader context.[275] The line defining prohibited speech is not easy to recognize.[276] The legality of almost any content will depends on context.[277] For example, violent words may require different treatment depending on whether they are intended to incite imminent violence, are being reported on by a news website, or are serving as part of a political protest.[278] First Amendment protections are intended to avoid the need for detailed decision-making about what can and cannot be published (at least as a matter of Constitutional law).[279] To identify undesirable content, all the major speech platforms use humans to assess context,[280] ensure accuracy,[281] and deliberate the enforcement and improvement of their content rules,[282] but human moderators may face similar challenges. To determine which content may be lawfully removed, a clear distinction must be made between protected speech, classified as legal under the First Amendment, and unprotected

275. For doubts regarding submitting content moderation function entirely to algorithmic systems, see Castets-Renard, *supra* note 234, at 292, 294, 315–16; Daphne Keller, *Internet Platforms: Observations on Speech, Danger, and Money* 4 (Hoover Institution's Aegis Paper Series No. 1807, 2018). (arguing that automated decision-making systems merely create an illusion of monitoring while being inefficient). Facebook employees say that Facebook's AI cannot consistently identify all harmful content, such as first-person shooting videos, racist rants, and even the difference between cockfighting and car crashes, according to the documents. Jeff Horowitz, *The Facebook Files*, WALL ST. J., Oct. 1, 2021, at 6 https://www.wsj.com/articles/the-facebook-files-11631713039 [perma.cc/8VLL-XUH3].

276. *See* Daphne Keller, *One Law, Six Hurdles: Congress's First Attempt to Regulate Speech Amplification in PADAA*, CIS BLOG (Feb. 1, 2021, 7:00 AM), https://cyberlaw.stanford.edu/blog/2021/02/one-law-six-hurdles-congresss-first-attempt-regulate-speech-amplification-padaa [perma.cc/NMR4-ANHK]. 18 U.S. Code § 2339B makes it unlawful, within the United States, or for any person who is subject to the jurisdiction of the United States anywhere, to knowingly provide material support to a foreign terrorist organization. 18 U.S.C. § 2339B. For platforms, the official list of designated organizations makes the task of identifying prohibited content comparatively clear, but not all content is as clear. Keller, *supra.*

277. Sharon Bar-Ziv & Niva Elkin-Koren, *Behind the Scenes of Online Copyright Enforcement: Empirical Evidence on Notice & Takedown*, 50 CONN. L. REV. 339, 343 (2017); Daphne Keller, *Inception Impact Assessment: Measures to Further Improve the Effectiveness of the Fight Against Illegal Content Online* (Mar. 29, 2018), https://papers.ssrn.com/abstract=3262950 [perma.cc/YZR7-U2PU] (arguing that in many cases, it is impossible to appreciate the legality of the content without consideration of the context).

278. *See* Brandenburg v. Ohio, 395 US 444 (1969) (applying a two-pronged test to evaluate speech acts: (1) speech can be prohibited if it is "directed to inciting or producing imminent lawless action," and (2) it is "likely to incite or produce such action." *Id.* at 447. The test concerns the regulation of speech that incites illegal content. But it does not cover all forms of information and makes it difficult to convict individuals for inciting harmful speech on social media.); Keller, *supra* note 33, at 252 (explaining that under the seminal Brandenburg "incitement to violence" standard, for example, words that are permitted in one's home may become illegal when spoken to an angry mob as they are unlikely to entice an action. Such a claim will be made ex post, against the speakers themselves, rather than against the platforms for their "speech," i.e., their amplification features. For example, many of the communications used to organize the violent January 6, 2021 attack on the Capitol, which left five people dead, are protected under the Brandenburg standard); For "fake news," see GILLESPIE, *supra* note 30, at 202 ("The clamor about 'fake news' may tend to erase important distinctions between propaganda, overstatement, partisan punditry, conspiracy theories, sensationalism, clickbait, and downright lies.").

279. CIVIC GENIUS, *supra* note 194.

280. Castets-Renard, *supra* note 234, at 313.

281. *Id.* at 294.

282. *Id.* at 291, 313 (arguing that platform monitors recognize when human verification is "appropriate" and when it is relevant to provide a detailed assessment of the context, as technical filters cannot assess context); *see also* Tim Wu, *Will Artificial Intelligence Eat the Law? The Rise of Hybrid Social-Ordering Systems*, 119 COLUM. L. REV. 2001, 2013 (2019).

or illegal speech.[283] Many of those judgments are open for interpretation and could easily be appealed. Therefore, such regulation would require a mechanism to review disputes raised by affected speakers.[284]

Third, a narrower scope of free speech will not necessarily prevent targeting harmful ideas at users. In practice restricted speech can often be pushed beyond the margins of suspicion by using coded language or symbols to evade prohibitions.[285] Alternatively, other speakers or posts conveying the same message are likely be re-targeted to users' profiles.[286] Additionally, while certain forms of speech may be subject to legal restrictions,[287] such laws raise equal constitutional concerns like equal protection issues, (where restrictions disproportionately affect specific groups based on race, religion, gender, or political affiliation) and Fourth Amendment issues (where law enforcement agencies conduct mass surveillance of individuals' online communications without adequate justification).[288]

Finally, content-based laws will create pervasive legal risk for platforms with potential repercussions for the entire landscape of public discourse. Platforms could face a continuous stream of lawsuits from users claiming that the platform violated their free speech rights. More influential speakers, whose content tends to spread more rapidly, are likely to be disproportionately affected.[289] Even in cases where platforms are likely to prevail, the financial and operational costs of defending against First Amendment lawsuits can be substantial.[290] In order to avoid liability or

283. Keller, *supra* note 33, at 245; *see also* Keller, *supra* note 276; Dan L. Burk, *Algorithmic Fair Use*, 86 U. CHI. L. REV. 283 (2019); *See generally* KATHLEEN HALL JAMIESON, CYBERWAR: HOW RUSSIAN HACKERS AND TROLLS HELPED ELECT A PRESIDENT—WHAT WE DON'T, CAN'T, AND DO KNOW (2018) (noting that there is an ongoing academic debate about exactly what fake news is, and whether to even call it as such).

284. Keller, *supra* note 33, at 242. As no regulator has the capacity to provide nuanced analysis of restriction and removal questions, down-stream disputes requiring adequately resolving interpretive questions would require administrative capacity well beyond that of any current U.S. regulator. Courts or administrative agencies will have to decide which speech is illegal, providing due process to the speaker, and serve platforms with court orders.

285. Keller, *supra* note 33, at 256 ("According to Facebook, no matter what line the company draws in prohibiting content like misinformation or racist language, users are incentivized to create and post a high volume of material that comes close to crossing the line without quite doing so."). *See* Whittaker et al., *supra* note 56, at 19 (noting a case of borderline content).

286. Keller, *supra* note 33, at 235.

287. U.N. Special Rapporteurs on the Promotion and Protection of the Right to Freedom of Opinion and Expression et al., *Joint Communication to Pakistan and Lao People's Democratic Republic*, U.N. Doc. OL PAK 8/2016 & OL LAO 1/2014; ASSOCIATION FOR PROGRESSIVE COMMUNICATIONS, UNSHACKLING EXPRESSION: A STUDY ON LAWS CRIMINALISING EXPRESSION ONLINE IN ASIA (2017).

288. Mike Masnick, *NY Times Joins Lots of Other Media Sites in Totally and Completely Misrepresenting Section 230*, TECHDIRT (Aug. 7, 2019, 9:34 AM), https://www.techdirt.com/2019/08/07/ny-times-joins-lots-other-media-sites-totally-completely-misrepresenting-section-230/ [perma.cc/CA2D-CJMD].

289. Keller, *supra* note 33, at 262.

290. Keller, *supra* note 33, at 271 ("Such errors in demoting content, like errors in deleting it, can be checked by better processes within the platform or before courts or administrative agencies. But the better those processes are, the greater their cost, for the platform and for any agencies or courts tasked with resolving disputes."); *see, e.g.*, Green v. Am. Online, 318 F.3d 465 (3d Cir. 2003); Prager Univ. v. Google LLC, 951 F.3d 991 (9th Cir. 2020) (noting Prager University, a nonprofit that creates short videos advocating conservative viewpoints, filed suit against Google and YouTube alleging violation of the First Amendment. The court held that the AOL is immunized from suit and could not be held liable for an alleged negligent failure to police its network for content provided by its users.

even just litigation expenses, platforms may be disincentivized from offering open forums for user engagement altogether,[291] or may instead choose to broadly suppress categories of content—not only clearly unlawful material but also speech that falls into legally ambiguous or subjective grey areas.[292] Over-enforcement of potentially lawful speech risks further eroding legitimate and public discourse.[293]

Ultimately, these proposed solutions are not only difficult to implement, they also miss the core issue. Attempts to regulate content are often constitutionally untenable under the First Amendment. More importantly, content-based regulation typically seeks to suppress harmful or undesirable speech, often targeting specific ideas, potentially those that distort reality or promote misinformation and hate speech. Such regulation fails to address the deeper structural problem: the individualized manipulated distribution of content enabled by personalization features. While restricting certain types of content may contribute to a safer online environment, it does not confront the mechanisms by which platforms deliver personalized content to individual users, mechanisms that can distort perception, reinforce bias, and shield users from competing viewpoints even with lawful

---

Section 230(c) does not require AOL to restrict speech; rather it allows AOL to establish standards of decency without risking liability for doing so). Freedom Watch, Inc. v. Google, Inc. 816 F. App'x 497 (D.C. Cir. 2020) (discussing the Sherman Antitrust Act and the District of Columbia Human Rights Act. Freedom Watch had brought a complaint against major technology firms arguing that Google, Facebook, Twitter, and Apple intentionally and willfully conspired to suppress politically conservative content which resulted in a dramatic loss of its viewership and user engagement. The Appellate Court held that Freedom Watch's claim concerning the defendants' anti-competitive conduct was invalid as no agreement occurred between the platforms, the defendants were not "public accommodations" within the ambit of DC's anti-discrimination laws, and that the defendants were not quasi-state actors capable of being sued for First Amendment violations for suppression of speech. The Court noted that general allegations of conspiracy to suppress speech without concrete facts was insufficient to warrant a judgment in favor of the appellants).

291. Daphne Keller, *Facebook Filters, Fundamental Rights, and the CJEU's* Glawischnig-Piesczek *Ruling*, 69 GRUR INT'L, 616, 623 (2020). *See, e.g.*, Midwest Video Corp. v. F.C.C., 571 F.2d 1025, 1056–57 (8th Cir. 1979), *aff'd on other grounds*, 440 U.S. 689 (1979) (rejecting an FCC regulation requiring cable operators to restrict unlawful content from programmers. "The Commission made the cable operator both judge and jury, and subjected the cable user's First Amendment rights to decision by an unqualified private citizen, whose personal interest in satisfying the Commission enlists him on the "safe" side—the side of suppression." The Court rejected both common carriage obligations for cable companies and imposition of liability for obscene and other unlawful content on public access channels). The Eighth Circuit ruled only on statutory grounds, despite asserting that "[w]ere it necessary to decide the issue, the present record would render the intrusion represented by the present rules constitutionally impermissible." *Id.* at 1056.

292. Keller, *supra* note 33, at 246. *See* Smith v. People of the State of Cal., 361 U.S. 147 (1959) (holding unanimously that the First Amendment rights of a Los Angeles bookstore owner had been violated when he was held criminally liable and sentenced to 30 days in jail in 1956 for selling the pulp novel Sweeter Than Life, by Mark Tryon. The Court rejected strict obscenity liability, noting that the bookseller's resulting "timidity" from such unbounded liability can lead it to "restrict the public's access to forms of the printed word which the State could not constitutionally suppress directly"); *See* Keller, *supra* note 33, at 248 (explaining a similar regulation that only incentivizes, rather than mandates excessive caution, will result in a similar way. It will cause platforms to remove or demote, self-censure and hold back on recommendation of lawful, albeit harmful content, which may be considered unjustified interference with users' fundamental rights).

293. *See* U.N. Special Rapporteurs on the Promotion and Protection of the Right to Freedom of Opinion and Expression et al., *Joint Communications to Malaysia, the Russian Federation, the United Arab Emirates, Bahrain, Singapore, and the Russian Federation*, U.N. Docs. OL MYS 1/2018; UA RUS 7/2017; UA ARE 7/2017; AL BHR 8/2016; AL SGP 5/2016; & OL RUS 7/2016. For example, Azerbaijan prohibits propaganda of terrorism, religious extremism, and suicide: Azerbaijan submission.

content. Crucially, the legality of the content is beside the point of selective distribution due to personalization features. The core concerns lie not only in the nature of content itself, but in its selective, opaque distribution and the fragmented effects it produces across audiences. Personalization enables the delivery of divergent, partial and even contradictory messages to targeted users of both lawful and unlawful speech, intensifying their manipulative potential regardless of the content's legal status. Each user receives only a curated slice of reality. This grants platforms a powerful role in shaping public perception and controlling the flow of information—particularly in areas most vital to democratic governance. As long as distribution remains individualized, excluding others who might challenge, contextualize, or correct them, it will keep eroding the public discourse.

B. Content-Neutral Regulation

Since most content regulation would constitute a violation of the First Amendment, regulators may, instead, address the technology used by online platforms to distribute speech and personalize content. Algorithm regulation addresses a message's *volume* rather than its *content*, meaning it is content-neutral.[294] Such regulation may include factors such as, dismissing any form of amplification, showing user posts in reverse-chronological order, or "circuit breakers" laws that slow the spread of highly viral content, minimizing the flow of potentially harmful information.[295] If highly popular content can do more damage than less distributed content, "circuit breakers" could reduce the weight given to engagement metrics such as likes and shares, restrict the number of times an item is displayed to users, cap the hourly rate of increase in viewership, and eventually reduce its visibility.[296]

Content-neutral models have several advantages. Distribution rules avoid the need to legally define restricted speech and allow choosing more nuanced responses than the binary take-down/leave-up choices recognized under most laws today. If the court accepts that a ban on platform amplification is content-neutral, perhaps it can stand the constitutional scrutiny of the First Amendment and mediate the power of gatekeeping platforms as distributers and their capacity to affect the discourse.[297] But platforms object, arguing that amplifying algorithms are in

294. Keller, *supra* note 33, at 261.

295. *Id.* at 254; Keller, *supra* note 276 (explaining that at least one proposed U.S. law has taken this approach, eliminating immunity under CDA 230 for terrorism or civil rights lawsuits about amplified content); Petition for Rulemaking of the National Telecommunications and Information Administration (Fed. Commc'ns Comm'n July 27, 2020) https://www.ntia.gov/files/ntia/publications/ntia_petition_for_rulemaking_7.27.20.pdf [perma.cc/W8KA-BB3B] (describing a proposal backed by the Trump administration that would have stripped platforms of immunity for user content that they algorithmically "promoted"); *Draft Report of the Committee of Legal Affairs with Recommendations to the Commission on a Digital Services Act*, at 23 (Apr. 22, 2020), https://www.europarl.europa.eu/doceo/document/JURI-PR-650529_EN.pdf [perma.cc/M9GU-CZM9] (describing an EU Parliament draft report that discussed restricting "the amplification of content that is attention-seeking or sensationalist in nature.") *see, e.g.*, Justice Against Malicious Algorithms Act of 2021, H.R. 5596, 117th Cong. (2021); Health Misinformation Act of 2021, S. 2448, 117th Cong. (2021) (suggesting limiting a social media company's immunity from liability if it promotes certain content on its platform).

296. Keller, *supra* note 33, at 254.

297. *Id.* at 260.

themselves protected speech.[298] Constitutionally, platforms' right to decide what kind of content to entertain, recommend, personalize, amplify, or offer users is part of their own commercial speech, protected by the First Amendment.[299] When amplifying certain content, platforms convey the following information and ideas: "I predict that you'll like this" or "I think this is what you're looking for."[300] Restricting amplification will interfere with platforms' speech—as speakers making such recommendations, or as editors sharing, omitting, or prioritizing content for users through their algorithms.[301] Moreover, protecting algorithm-based decisions aligns with the broader protection of human expression, because depending on the algorithm, algorithm-based decisions (algorithms output) constitute an extension of the developer's self-expression, autonomy, and meaningful thought, which is a human message protected in itself.[302]

Challenging platforms' constitutional defenses is difficult, but some legal approaches have sought to weaken platforms' claims that algorithmic content delivery is constitutionally protected expression, either by limiting the scope of protection afforded to commercial speech,[303] or by questioning whether

298. Gilad Edelman, *How Facebook Gets the First Amendment Backward*, WIRED (Nov. 7, 2019, 5:29 PM), https://www.wired.com/story/facebook-first-amendment-backwards [perma.cc/9J9S-YWZV].

299. U.S. COPYRIGHT OFF., CIRCULAR 14, COPYRIGHT IN DERIVATIVE WORKS AND COMPILATIONS (2020), https://www.copyright.gov/circs/circ14.pdf [perma.cc/NHA8-8Y2K]; *see also* Keller, *supra* note 33, at 247–48 ("[T]he difference between user speech and platform speech is analogous to the difference between an essay and the anthology that contains it—each of which is deemed a distinct creative work under U.S. copyright law, with the latter receiving its own protection based on the anthologist's selection and arrangement of third-party speech."). For the case establishes the right of a company to choose what expressions it carries, see Pac. Gas & Elec. Co. v. Pub. Utilities Comm'n, 475 U.S. 1 (1986) (describing a regulatory authority that wanted Pacific Gas to add information to the envelope sent to consumers in order to maximize efficiency. The Court found the order to be unconstitutional because the company's freedom of speech includes the choice what not to say).

300. Keller, *supra* note 33, at 247.

301. *See id.* at 229 (arguing that arriving at well-crafted regulation focusing on amplification may cause more problem, because they have much in common with human or constitutional rights hurdles that models which define platforms' responsibility for content posted by users often face); *see also* Berin Szóka & Ari Cohn, *The Wall Street Journal Misreads Section 230 and the First Amendment*, LAWFARE (Feb. 3, 2021, 3:43 PM), https://www.lawfaremedia.org/article/wall-street-journal-misreads-section-230-and-first-amendment [perma.cc/8DPG-DU3N] (noting that websites have the same constitutional right as newspapers to choose whether or not to carry, publish or withdraw the expression of others. The value of Section 230 is instrumental, it "ensures that courts will quickly dismiss lawsuits that would have been dismissed anyway on First Amendment grounds").

302. Benjamin, *supra* note 35, at 624 ("The problem is that many algorithm-based decisions similarly involve the creator's self-expression and autonomy.").

303. For suggestions to exclude free speech protection from speech made by corporations solely for their own benefit, or to give lower standard of constitutional protection to commercial speech, see, for example, C. Edwin Baker, *Commercial Speech: A Problem in the Theory of Freedom* 62 IOWA. LAW REV. 1, 3 (1976) ("[G]iven the existing form of social and economic relationships in the United States, a complete denial of first amendment protection for commercial speech is not only consistent with, but is required by, first amendment theory."); C. R. SUNSTEIN, DEMOCRACY AND THE PROBLEM OF FREE SPEECH 123, 127 (1993) (arguing for lesser protection of advertising because it does not contribute to democratic deliberation). Some have supported lesser protection for commercial speech even previously to algorithmic advertising. *See* Va. State Bd. of Pharmacy v. Va. Citizens Consumer Council, Inc., 425 U.S. 748 (1976) (acknowledging that commercial speech is not immune from government regulation, but entitled to less protection than political speech.); In re R.M.J., 455 U.S. 191, 203 (1982) (noting that restrictions can be imposed if the content or method of advertising is inherently misleading or subject to abuse, or if there is a substantial government interest at stake).

algorithmically generated outputs qualify as protected speech.[304] If the value behind free speech is the speakers' participation in the public political and social life, it makes sense to protect speech made by individuals differently than by non-human speakers, or decisions made by algorithms. But the success of such an argument depends on the rationale behind free speech: Is it the speaker's right to speak, or the audience's interest to be informed? Diminished protection for non-human speech may undermine consumers' interest in being informed, which is not contingent on whether the speech originates from an algorithm or an individual.[305] For the time being, the Supreme Court has ruled that both human and non-human speech should be afforded the same constitutional protections, emphasizing the value of such speech to listeners rather than the humanity of the speaker.[306]

Beyond speakers and listeners, Erin L. Miller argues that the rationale behind First Amendment protections is not solely grounded in individual rights but also in the structural role that freedom of expression plays in supporting democratic values and safeguarding the public interest.[307] While individual expression is essential, Miller emphasizes that the First Amendment also serves broader systemic goals—namely, maintaining the integrity of democratic values and public interests.[308] These structural interests pertain not just to individual acts of speech, but to the overall functioning of the expressive ecosystem, including its value to third parties and society at large.[309] Building on this logic, some regulation proposals have sought to limit social media platforms' immunity when they algorithmically amplify certain types of content.[310] However, Miller's framework presents a practical and

304. *See* Benjamin, *supra* note 35, at 624 (arguing that focusing on an individual's expression is a way to limit the scope of the First Amendment suggesting that communications by corporations via their algorithms should not constitute speech).

305. First Nat'l Bank of Boston v. Bellotti, 435 U.S. 765 (1978) (declaring unconstitutional a state law restricting banks and corporations from spending money to influence referendum votes on unrelated issues. The Court recognized the public's right to hear all views on matters of public importance, irrespective of whether the speech was from a corporation or an individual).

306. Benjamin, *supra* note 35, at 606; Sorrell v. IMS Health Inc., 564 U.S. 552, 557 (2011). In *Sorrell v. IMS Health Inc.*, the Supreme Court discussed Vermont's right to regulate the selling of users' economic activity and the flow of information about such activity. Vermont legislated Act 80 to restrict pharmacies from selling prescribers' identifying information without prescribers' consent because of privacy considerations. Medical data miners argued that the law violated their free speech. The Court of appeal reversed the district Court's decision and determined that the Act violated the First Amendment by unduly burdening the commercial speech of pharmaceutical marketers and data miners. The decision in *Sorrell* establishes that algorithms are a form of expression that falls within the First Amendment's free speech rights despite the underlying regulation having an economic motive. Big tech companies use the free speech defense from *Sorrell* when dealing with decisions based on data mining made by their algorithms and other systems.

307. Miller, *supra* note 7, at 25.

308. Miller, *supra* note 7, at 69 ("[C]ertain speech-restrictive measures can be taken to preserve the integrity of democratic discourse without infringing on individuals' free speech rights. These measures, from campaign finance laws to media regulations, target primarily speech amplified over mass-media platforms. These platforms are scarce, and access to them is granted to no small degree due to socioeconomic advantages. Thus, laws designed to ensure, in service of democratic discourse, that a diversity of voices are heard over those platforms, should, if narrowly tailored, survive constitutional scrutiny.").

309. *Id.* at 28.

310. Protecting Americans from Dangerous Algorithms Act, H.R. 2154, 117th Cong. (2021) (sponsored by Reps. Malinowski (D-N.J.-7) and Eshoo (D-Cal.-18) with the aim to limit social media companies' immunity for interactive computer services for certain claims if it amplified user content that violates the specified federal statutes on its platform. Specifically, the bill removes this immunity

constitutional challenge for application, similar to content-based regulation poses: unless regulation can clearly differentiate between speech that upholds democratic integrity and that which undermines it, efforts to impose such distinctions risk arbitrariness or constitutional infirmity.

While still protected, there is a strong argument that algorithms should be subjected to a lower standard of constitutional protection, as the value of their speech is not the same as political speech.[311] The extent of a lower standard of protected amplification is unclear and difficult to implement.[312] Assessment of a lower standard of First Amendment protections will be hard to determine in the context of a huge for-profit corporation, which often owns the products of employees' work. Compelling disclosure of proprietary algorithms involves a complex intersection of intellectual property rights, including trade secrets and copyrights.[313] Copyright protection extends to expressions of ideas, methods, or systems.[314] Algorithms, being creative expressions of ideas, may be eligible for copyright protection, which helps prevent unauthorized use and ensures that companies can maintain a competitive advantage. Courts will need to balance the interests of advancing innovation with consumer protection, cybersecurity, and the potential societal benefits of algorithmic transparency. Additionally, this assessment should include an evaluation of the potential harm to the developer's autonomy and meaningful thought, as recognized in the protection of self-expression. Since algorithm-based speech forms the foundation of many platforms' business models,

---

from a social media company with more than 10 million monthly users if it utilizes an algorithm, model, or other computational process to amplify or recommend content to a user that is directly relevant to a claim involving (1) interference with civil rights, (2) neglect to prevent interference with civil rights, or (3) acts of international terrorism.); *see also* Protecting Americans from Dangerous Algorithms Act, H.R. 8636, 116th Cong. (2020), https://www.congress.gov/bill/116th-congress/house-bill/8636/text [perma.cc/7JCD-U6E3]. For criticism of the bill, see Keller, *supra* note 276 (arguing that the bill ensures over-enforcement against lawful expression by assigning risk-averse platforms to interpret the law, with no procedural protections for user rights and no effort to avoid driving platforms to use flawed automation).

311. Benjamin, *supra* note 35, at 623. *See generally* Grafanaki, *supra* note 27 (noting that algorithms that propagate false information, extreme hate speech, and polarization may not deserve the same level of protection. If algorithms have any Constitutional status, it is derived from the value they provide to their audience as a delivery mechanism of information. If constitutional protection is based on delivery of information, content navigation algorithms may not qualify to be protected under the core First Amendment doctrine).

312. Cent. Hudson Gas and Elec. Corp. v. Pub. Serv. Comm'n of N.Y., 447 U.S. 557 (1980) (holding that a state's regulation of commercial speech should be "no more extensive than necessary" to achieve its intended purpose). But this formulation was considered too burdensome on the states. *See* Bd. of Trs. of State Univ. of N.Y. v. FOX, 492 U.S. 469 (1989). A state-university regulation of on-campus business activity effectively prevented a seller of household goods from holding "Tupperware parties" in the dormitories. *Id.* at 472. Although the company's representatives not only sold goods but also made presentations on home economics, the Court concluded that the speech was commercial. Keller, *supra* note 33, at 255 ("[D]raw[ing] conclusions about that trade-off, we need to be clearer about what the resulting platform services would actually do. What counts as 'amplification,' and what do we believe an authentic, un-amplified service looks like? Depending on our answers to those questions and our own policy priorities, we may have different ideas about which platform design choices are legitimate and authentic, when or why algorithmic intervention is warranted, and whether we would be better off without those algorithms.").

313. Bowman v. UK, App. No. 24839/94, ¶ 47, (19 February 1998), https://hudoc.echr.coe.int/fre?i=001-58134 [perma.cc/5D99-8ALN].

314. CHRISTOPHER T. ZIRPOLI, CONG. RSCH. SERV. LEGAL SIDEBAR, LSB10922, GENERATIVE ARTIFICIAL INTELLIGENCE AND COPYRIGHT LAW (2023).

restricting commercial activities may suppress the industry or prevent healthy competition.[315]

Moreover, in a reality of unprotected algorithms, an un-personalized feed may miss the core issue, as it is lacking the capacity to demote or filter out potentially harmful, extreme, or fake material.[316] In the absence of algorithmic intervention, there would be no mechanism to address or curtail inauthentic user behavior or the spread of harmful content,[317] which may subject users to numerous repetitions and potential spam, resulting in a poor user experience and reduced attention to the social discourse.[318] In such circumstances, it is hard to justify such a regulatory model, even if one accepts the need to restrict amplification features.[319]

C. Privacy Rights Approaches

Because approaches based on content and algorithmic regulation appear to run into constitutional dead ends or solutions incompatible with existing business models, it is also worthwhile to explore options grounded in privacy or consumer rights.[320] Privacy-based regulation concentrates on safeguarding user data and privacy in the digital sphere. It aims to protect individual privacy by restricting collection, use, and sharing of personal information by online platforms, which can

315. *See* Keller, *supra* note 33, at 236; Kende, *supra* note 60, at 295.

316. Keller, *supra* note 33, at 256. Two studies have tried to audit the algorithms using "puppet accounts" to provide improved understanding of the impact they have on the information users see. *See* Jack Bandy & Nicholas Diakopoulos. *More Accounts, Fewer Links: How Algorithmic Curation Impacts Media Exposure in Twitter Timelines*, 5 PROCS. ACM ON HUM.-COMPUT. INTERACTION 1, 1–28 (2021) (finding that the curation algorithm, when compared to a chronological newsfeed, resulted in an increase in unique accounts and partisan echo chambers and a decrease in dominance of accounts with high number of followers and exposure to links); Nathan Bartley, Andres Abeliuk, Emilio Ferrara & Kristina Lerman, *Auditing Algorithmic Bias on Twitter*, *in* 13 PROCS. ACM WEB SCI. CONF. 65, 65–73 (2021) (finding that the curation algorithm showed more popular and less recent tweets compared to a chronological timeline).

317. Keller, *supra* note 33, at 258; Andrew M. Guess et al., *How Do Social Media Feed Algorithms Affect Attitudes and Behavior in an Election Campaign?*, 381 SCI. MAG. 398 (2023). In a study with chronologically ordered feed, the amount of political and untrustworthy content they saw increased on both platforms, the amount of content classified as uncivil or containing slur words they saw decreased on Facebook, and the amount of content from moderate friends and sources with ideologically mixed audiences they saw increased on Facebook. Despite these substantial changes in users' on-platform experience, the chronological feed did not significantly alter levels of issue polarization, affective polarization, political knowledge, or other key attitudes during the three-month study period.

318. Keller, *supra* note 33, at 256; *see also* Michael Beam, *Automating the News: How Personalized News Recommender System Design Choices Impact News Reception*, 41 COMMC'N RSCH. 1019 (2014) (examining the impact of design choices in personalized news recommender systems on users' news consumption behavior and attitudes. Key Findings suggest that users tended to engage more with personalized news recommendations compared to non-personalized ones. Therefore, they were more likely to encounter content that aligned with their preexisting beliefs and preferences. Users appeared to accept these trade-offs, as they valued the convenience and personalization provided by such systems. Personalization enhanced the overall user experience of the news platform. Users were aware of the potential for algorithmic bias but did not seem to be overly concerned about it).

319. Keller, *supra* note 33, at 251.

320. *Id.* at 271. *See* THE WHITE HOUSE, BLUEPRINT FOR AN AI BILL OF RIGHTS 3 (2022), https://bidenwhitehouse.archives.gov/ostp/ai-bill-of-rights/ [perma.cc/AV76-MG36] (focusing on privacy violation and surveillance when conceptualizing the problem: "Too often, these tools are used to limit our opportunities and prevent our access to critical resources or services. . . . The President has spoken forcefully about the urgent challenges posed to democracy today and has regularly called on people of conscience to act to preserve civil rights—including the right to privacy").

be exploited for personalization purposes. Privacy-based approaches often adopt an individualistic framework, emphasizing a person's right to control how their personal data is collected, used, and shared.[321] These approaches may overlap with consumer protection rights and attempt to subject companies' business conduct to oversight.[322]

Controlled data collection mechanisms aim to give consumers upstream control of data that is generated by or about them, before that data is used for profiling and personalization.[323] Opt-out approaches allow individuals to have control over certain tracking activities, such as which personal information is collected, how it is used, with whom it is shared, and the conditions of exposure to it.[324] For example, the California Consumer Privacy Act (CCPA) includes provisions that allow consumers to opt out of the sale of their personal information, including for the purposes of personalized advertising.[325] Data minimization mechanisms focus on collecting the least amount of personal data.[326] "Do Not Track" rejects cookies in the browser, signaling to websites that the user does not want to be tracked.[327] Additionally, many social media platforms and online services offer users technological features, tools and privacy settings that allow users to filter out unwanted types of content, such as ad blockers, or filters that block certain keywords or topics, or limit who can see your posts.[328] Users can choose whom to follow, what advertisements to view, and which posts or users to hide.

These methods could assist with the problem of individualized content matching. Preventing the flow of users' information that platforms could rely on interferes with the role information plays in profiling and therefore limits the option to personalize content efficiently. Privacy-based approaches often frame privacy as a market product to be competed over.[329] However, the market for privacy is not very competitive. Given the profitability attached to targeting, it is unclear what kind

---

321. For a privacy-based approach see, for example, Council 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 Council Directive 95/46/EC, 1995 OJ (L 281) 31, 31–50 (EU) [hereinafter GDPR].

322. Gordon-Tapiero et al., *supra* note 16, at 654–55; *see, e.g.*, Honest Ads Act, H.R. 4077, 115th Cong. (2017); GDPR art. 5.

323. Grimmelmann, *supra* note 113, at 349.

324. *See, e.g.*, Popescu & Baruh, *supra* note 69, at 283 (noting that consent-based models may require platforms to obtain user consent before collecting and processing their personal data. Privacy policies do not allow the users to renegotiate the privacy contract "on the go" without costly expenditure of attention and time).

325. *California Consumer Privacy Act (CCPA)*, ATT'Y GEN. CA. (May 10, 2023), https://oag.ca.gov/privacy/ccpa [perma.cc/65WY-UYA8].

326. *What is Data Minimization and Why Is It Important?*, KITEWORKS, https://www.kiteworks.com/risk-compliance-glossary/data-minimization/ [perma.cc/PNH7-XAC2] (last visited Dec. 24, 2023); ANN CAVOUKIAN & JEFF JONAS, PRIVACY BY DESIGN IN THE AGE OF BIG DATA (2012), https://jeffjonas.typepad.com/Privacy-by-Design-in-the-Era-of-Big-Data.pdf; *see also* American Data Privacy and Protection Bill, H.R. 8152, 117th Cong. (2022) (requiring most companies to limit the collection, processing, and transfer of personal data to that which is reasonably necessary to provide a requested product or service and to other specified circumstances).

327. *What Is Do Not Track?*, ONETRUST (Sept. 10, 2024), https://my.onetrust.com/s/article/UUID-ddce2f5c-d01c-add4-26eb-c105b086217d?language=en_US [perma.cc/48CS-USLV].

328. *See How to Filter, Block, and Report Harmful Content on Social Media*, RAINN, https://rainn.org/articles/how-filter-block-and-report-harmful-content-social-media [perma.cc/6MY9-B346].

329. Popescu & Baruh, *supra* note 69, at 274.

of incentives would drive companies to outperform their competitors with respect to their privacy practices.[330]

Moreover, opting out of data collection is insufficient to prevent content matching. Withholding information from the platforms does not meaningfully hinder algorithms from inferring user characteristics, nor does it prevent the platform from delivering individually tailored content.[331] In the complex landscape of data collection and profiling, there may be indirect ways for platforms to target users even after they have opted out and rejected data collection. Platforms may still have access to identifiers associated with a user, such as device IDs, geo-tagging, IP addresses, or cookies.[332] Additionally, statistical inference techniques allow platforms to make educated guesses or predictions about users, based on aggregated data and patterns through dependencies based on online connections, [333] activities of other users with similar characteristics or behaviors, and data synchronization between devices.[334] When users opt out, it would typically mean that their data has been deleted or anonymized and personal identifiers were removed from data to eliminate any profiling or targeting before similar information is gathered again.[335] Yet, a large body of scholarship has shown that anonymization cannot prevent re-identification, especially when data is combined or linked with other available datasets.[336] Finally, while a platform may have erased a user's data within its own systems, other platforms or third-party data brokers also collect and store users' data. With data-sharing agreements in place, third parties may continue using previously obtained data for profiling and targeting purposes. Individualized distribution will persist despite restrictions on data collection, whether based on user profiles or other mechanisms. As long as platforms tailor content on an individual basis, public discourse will be undermined unless platforms are deliberately designed to present balanced or representative discourse.

Privacy regulation alone cannot fully address the negative consequences of personalization, since the impact of personalization extends beyond individual privacy and affects the integrity of the shared public sphere. Gordon-Tapiero et al.

---

330. *Id.* at 279.

331. Gordon-Tapiero et al., *supra* note 16, at 641–42.

332. Popescu & Baruh, *supra* note 69, at 282.

333. *See* Solon Barocas & Karen Levy, *Privacy Dependencies*, 95 WASH. L. REV. 555 (2020) (explaining tie-based dependencies).

334. *See* Gordon-Tapiero et al., *supra* note 16, at 648; Kristen M. Altenburger & Johan Ugander, *Monophily in Social Networks Introduces Similarity Among Friends-of-Friends*, 2 NATURE HUM. BEHAV. 284, 287 (2018).

335. *See* Gordon-Tapiero et al., *supra* note 16, at 661; *see also* Gaetano DiNardi, *How To Opt Out of Data Broker Sites (With Examples)*, IDENTITY GUARD (Aug. 23, 2023), https://www.identityguard.com/news/how-to-opt-out-of-data-broker-sites [perma.cc/8NPE-K376] ("Depending on the language of the opt-out, you may block your information from being sold, or delete[d].").

336. Sarah Zhang, *Scientists Are Just as Confused about the Ethics of Big-Data Research as You*, WIRED (May 20, 2016, 5:07 PM), https://www.wired.com/2016/05/scientists-just-confused-ethics-big-data-research/ [perma.cc/7TFK-QXUD]. To address this scholarship, differential privacy models involve adding noise to data to protect user privacy while still allowing content moderators to analyze the data for harmful content. It makes it harder for individuals to be re-identified from the data. *See* 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 (2018).

point out a need to adopt a collective perspective.[337] They argue that the harms caused by personalization are not confined to individual experiences but contribute to shaping a broader social context in which personalization operates.[338] I would argue that this argument requires two additional steps. First, recognizing that individuals are being profiled and targeted as part of a collective, as Gordon-Tapiero et al. suggest, and second, addressing the possibility that collective data can be used to manipulate each member of the collective individually. Individual user's privacy settings typically govern data sharing between platforms and third parties, but they do not grant users control over content-matching criteria or communicative interaction between the user and the platform itself, the type of interaction that platforms profit from.[339]

To summarize Part III, while content regulation, algorithm regulation, and privacy-based regulation each serve important purposes, the erosion of the public sphere for discourse caused by selective exposure to content remains largely overlooked.[340] Current regulatory initiatives are also ill-equipped to effectively address this issue. They primarily address partial aspects of personalization features and may fall short in comprehensively addressing the broader impact personalization features have on public discourse (see Table 1 below). These regulatory approaches face legal and practical difficulties. Approaches that aim to expand the scope of illegal content face free speech hurdles, as they qualify as censorship of users' content and are not nuanced enough to account for context. Approaches that aim to restrict amplifying algorithms arguably violate the platforms' right to free speech as well as meddle with commercial secrets and business models. Approaches focused on individual rights, user privacy, and consumer protection, particularly those aimed at giving individuals control over their data, are insufficient to prevent profiling or regulate the individualized flow of content.

All these approaches also face common enforcement challenges. Regardless of how a law is framed (content, amplification, or privacy regulation), or who defines the scope of new speech restrictions (the government or the platforms), it is up to the major platforms to actively decide which content should be entirely removed from the platform or demoted through the algorithm, or whether or not to respect privacy settings. Any legal framework effectively delegates enforcement steering wheel to the platforms themselves, thereby subjecting its implementation to their commercial interests and internal company values systems.[341] Platforms are

337. *See* Gordon-Tapiero et al., *supra* note 16, at 687 (suggesting the adoptions of a collective perspective to address the negative effects of personalization, either by regulatory standards or community norms beyond legal requirements. Legislation could establish enforcement mechanisms such as flagging, deprioritizing, or blocking problematic personalization, or assigning criminal or civil penalties for platforms that engage in harmful personalization).

338. *Id.* at 641.

339. Popescu & Baruh, *supra* note 69, at 273 (arguing the FTC "context of interaction" standard focuses exclusively on user's informational choices, privacy as data appropriation and private data disclosure).

340. Fukuyama, *supra* note 95.

341. Benjamin, *supra* note 35, at 606—31; *see also* PLURALISM OF NEWS, *supra* note 5, at 28. ("[I]t is hard to convince corporate executives to see the upside of adopting technologies that serve the sector or society as a whole, with limited or no impact on their bottom line, because they already see themselves as leaders in their field."); Oremus et al., *supra* note 138 ("Yet the news feed ranking system is not a total mystery. Two crucial elements are entirely within the control of Facebook's human

well positioned to moderate harmful content voluntarily. First, as private entities, platforms are not government actors bound by the First Amendment, and are therefore free to decide what content may be published on their services without any obligation to safeguard users' freedom of speech.[342] A constitutional right protects citizens from government interference, but it does not protect them from private entities, who are free to take as much leverage as they would like when deciding what they publish on their own websites. Second, platforms have the technical capability to design interfaces that amplify and recommend content to users, they control the information about the impact of their algorithms, and have administrative control over their business model.[343] Third, they have a business interest in monitoring content. For example, Facebook argues that pushing users towards extreme content is against their business interest—financially or reputationally—because the vast majority of its revenue comes from advertising, and advertisers do not want their brands and products displayed next to extreme or hateful content because that would alienate potential customers.[344] According to Facebook, advertisers and members routinely hold platforms accountable for the content that they allow on their services.[345] Meaning, it is in part because of—not despite—their ad-driven revenue model that platforms are incentivized to remove harmful content.

However, many criticize self-enforcement as being unenforceable,[346] depends on users possessing digital privacy literacy,[347] and has limited impact on algorithmic profiling practices without offering, perhaps deliberately, to provide users with meaningful control over personalization and targeting criteria.[348] Self-regulated

---

employees, and depend on their ingenuity, their intuition and ultimately their value judgments. Facebook employees decide what data sources the software can draw on in making its predictions. And they decide what its goals should be—that is, what measurable outcomes to maximize for, and the relative importance of each.")

342. Prager Univ. v. Google LLC, 951 F.3d 991, 998 (9th Cir. 2020) (arguing that YouTube had violated his First Amendment rights by demonetizing and otherwise disfavoring his videos. The Ninth Circuit firmly rejected that claim); CIVIC GENIUS, *supra* note 194.

343. Gordon-Tapiero et al., *supra* note 16, at 643–44.

344. Clegg, *supra* note 6.

345. *Our Advertising Principles*, FACEBOOK, https://www.facebook.com/business/help/2001034850142726.

346. Tal Z. Zarsky, *Incompatible: The GDPR in the Age of Big Data*, 47 SETON HALL L. REV. 995 (2016); *see* Castets-Renard, *supra* note 234, at 321 (noting that users rights should be ensured by governmental laws and not only soft laws or internal rules from the platforms).

347. Popescu & Baruh, *supra* note 69, at 273. For literacy regarding consent forms online; *see, e.g.*, Joel R. Reidenberg et al., *Disagreeable Privacy Policies: Mismatches Between Meaning and Users' Understanding*, 30 BERKELEY TECH. L. J. 39 (2015); Ian Ayres & Alan Schwartz, *The No-Reading Problem in Consumer Contract Law*, 66 STAN. L. REV. 545, 600 (2014); Shmuel 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 (2018); Kevin Litman-Navarro, *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 [perma.cc/N4TC-CSKQ]; Uri Benoliel & Shmuel I. Becher, *The Duty to Read the Unreadable*, 60 B.C. L. REV. 2255, 2257 (2019); Yannis Bakos, Florencia Marotta-Wurgler & David R. Trossen, *Does Anyone Read the Fine Print? Consumer Attention to Standard-Form Contracts*, 43 J. LEGAL STUD. 1, 6 (2014).

348. Popescu & Baruh, *supra* note 69, at 279. DE STREEL ET AL., *supra* note 235, at 40–41 (arguing that measures used by online platforms are insufficiently effective in moderating illegal content online and removing provocative or inflammatory content. In the context of safeguarding fundamental rights, most NGOs noted that online platforms' moderating practices should increase moderation transparency, access to data, and information regarding platforms' decision-making processes).

restrictions are market-oriented and have not been effectively self-enforced in the past when they conflicted contradicted with corporate business models, diminishing their credibility.[349] Most anti-social behaviors remain unmoderated.[350] A Wall Street Journal series exposed that Facebook removes only a sliver of the posts that violate its own community standards.[351] Facebook was accused of actively contributing to harm, often in ways only the company fully understands.[352] Despite significant resistance from a large group of employees, Facebook has failed at removing some horrific child pornography.[353] Facebook's policy is not to remove false political ads, which played an active role in political polarizations such as the Facebook-Cambridge Analytica data scandal,[354] and had contributed to divisive, inter-religious conflict in India.[355] That Facebook has been the target of this investigation does not imply that they are the only platform to conduct itself in this way.[356] Instagram is harmful for, most notably, teenage girls, yet the company does not take a strong response despite employees flagging negative practices.[357] Platforms frequently, and conspicuously, fail to live up to our expectations.[358] Platforms are willing to risk the "threat" that advertisers would not want their products associated with extreme content, because the profits from distributing such harmful content outweigh potential advertiser losses.[359]

There are ways to do better, but platforms may not be well-motivated to try harder.[360] For example, internal Facebook documents show that people inside the company have long discussed a systematic approach to restrict features that disproportionately amplify incendiary and divisive posts. In 2018, Mark Zuckerberg announced that his product teams would focus not only on serving people the most *relevant* content for engaging users, but also on helping them have more *meaningful* social interactions—primarily by promoting content from friends, family, and groups they are part of rather than content from pages they follow.[361] Zuckerberg recognized explicitly that this shift would lead to people spending less time on

---

349. *See* Gilles Hilary & Clive Lennox, *The Credibility of Self-Regulation: Evidence from the Accounting Profession's Peer Review Program*, 40 J. ACCT. & ECON. 211, 212 (2005).

350. Joon Sung Park, Joseph Seering, & Michael S. Bernstein, *Measuring the Prevalence of Anti-Social Behavior in Online Communities*, 6 PROCS. ACM ON HUM.-COMPUT. INTERACTION 1, 1 (2022) ("[M]oderators only removed one in twenty violating comments in 2016, and one in ten violating comments in 2020. Personal attacks were the most prevalent category of norm violation; pornography and bigotry were the most likely to be moderated, while politically inflammatory comments and misogyny/vulgarity were the least likely to be moderated.").

351. Horowitz, *supra* note 275 (reviewing internal Facebook documents, research reports, online employee discussions and drafts of presentations to senior management).

352. *Id.*

353. Kende, *supra* note 60, at 289–90.

354. Confessore, *supra* note 156.

355. Horowitz, *supra* note 275.

356. *See*, *e.g.*, Kende, *supra* note 60, at 291–93 (illustrating that Twitter, despite announcing it will prohibit false political advertising globally, still allows problematic statements).

357. *Id.*

358. *Id.*

359. Halisia Hubbard, *Twitter Has Lost 50 of Its Top 100 Advertisers Since Elon Musk Took Over, Report Says*, NPR (Nov. 25, 2022, 7:37 AM), https://www.npr.org/2022/11/25/1139180002/twitter-loses-50-top-advertisers-elon-musk [perma.cc/73RB-56JH].

360. Keller, *supra* note 33, at 239.

361. Clegg, *supra* note 6.

Facebook.[362] The prediction proved correct, as the change led to a decrease of 50 million hours' worth of time spent on Facebook per day, and prompted a loss of billions of dollars in the company's market cap.[363] Facebook rejected those efforts because they would impede the platform's usage and growth; it uses a case-by- case approach instead (i.e., may use different standards based on specific profitability potential).[364] As long as platforms' incentives remain unchanged, effectively addressing the harmful effects of personalization will require, at least, entrusting oversight responsibility to an independent external body.[365]

IV. EXAMPLES OF APPROACHES THAT ADDRESS PERSONALIZATION OF CONTENT

This Article identifies a gap in current regulatory initiatives: it fails to address the impact of individualized, manipulated content matching on the public discourse. The next Part focuses on this unique concern and examines whether regulation could mitigate the harms to public discourse caused by individualized manipulated content flow. It explores two contrasting regulatory approaches: On the one hand, a less individualized content flow that enforces a unified narrative, illustrated by the example of The Great Firewall of China. On the other hand, choice-based models, based on transparency and middleware market, which aim to address the manipulation of content delivery and the current lack of user control over the criteria used to match content to them.

A. The Great Firewall of China

The Great Firewall of China, officially known as the Golden Shield Project, is a comprehensive system of internet censorship and control, implemented by the Chinese government to regulate online content and ensure that public discourse aligns with national narratives.[366] It accomplishes this purpose through a combination of technological, legal, and administrative measures. According to the government statement, China's Cyberspace Administration provisions aim to protect the rights of the public, by guiding and governing the ways in which technology companies operate.[367] China prevents Chinese internet users from freely accessing popular international social media platforms such as Facebook, X, Instagram, and YouTube, which are blocked within the country.[368] To replace

362. *Id.*

363. *Id.*

364. Horowitz, *supra* note 275.

365. Gordon-Tapiero et al., *supra* note 16, at 667.

366. RONALD DEIBERT, JOHN PALFREY, RAFAL ROHOZINSKI & JONATHAN ZITTRAIN, ACCESS DENIED: THE PRACTICE AND POLICY OF GLOBAL INTERNET FILTERING 263–68 (2008).

367. Zhaohui Su, Barry L. Bentley, Dean McDonnel, Ali Cheshmehzangi, Junaid Ahmad, Sabina Segalo, Claudimar Pereira da Veiga & Yu-Tau Xiang, *China's Algorithmic Regulations: Public-Facing Communication Is Needed*, 12 HEALTH POL'Y & TECH. 1, 3 (2023).

368. Matt Sheehan, *The Chinese Way of Innovation: What Washington Can Learn from Beijing About Investing in Tech*, FOREIGN AFFS. (Apr. 21, 2022), https://www.foreignaffairs.com/articles/china/2022-04-21/chinese-way-innovation; FRIEDRICH-EBERT-STIFTUNG, CHINA'S REGULATIONS ON ALGORITHMS: CONTEXT, IMPACT, AND COMPARISONS WITH THE EU 1, 2–5 (2023), https://library.fes.de/pdf-files/bueros/bruessel/19904.pdf [perma.cc/E9TB-YYW5]; Lionel Lavenue, Joseph Myles, Andrew Schneider, Finnegan Henderson, Garrett Farabow & Dunner, *Evaluating China's New 'Internet Information Service Algorithmic Recommendation Management' Regulations*, LAW.COM (April 21, 2022, 7:10 AM), https://www.law.com/legaltechnews/2022/04/21/evaluating-chinas-new-internet-i

international social media platforms, China has developed its own domestic social media platforms, such as Weibo, WeChat, and Douyin (TikTok), which are subject to strict government regulations and surveillance, allowing authorities to exert control over the information and discussions taking place on these platforms.[369] According to the Chinese government, its control over the internet has allowed domestic technology companies to thrive within a regulated environment, fostering economic growth and innovation within the country. But Chinese platforms are not free from legal responsibility.[370] Chinese technology companies are required to cooperate with government authorities in enforcing censorship mandates. The regulatory framework imposes "primary responsibilities" on platforms for the governance of online content, thereby promoting platform accountability.[371]

The Chinese government has enacted a series of laws and regulations that grant state authorities broad powers to control online content and to take legal action against individuals or entities disseminating information that contradicts official national narratives.[372] In March 2022, China released the Internet Information Service Algorithmic Management Regulations to regulate "algorithmic recommendation services" across the internet, and to facilitate a balance between Internet openness and censorship.[373] The Chinese government employs a sophisticated content control infrastructure, which involves extensive surveillance, Domain Name System (DNS) filtering, and deep packet inspection, all designed to align online information with the government's political and social objectives.[374] As part of the Great Firewall, keyword filtering is employed to detect and obstruct specific words or phrases associated with prohibited topics.[375] These measures extend to websites addressing themes such as human rights, democracy, criticisms of the government, and other subjects that diverge from the government's official ideology. Chinese platforms are required to ensure that algorithms "present information conforming to mainstream values," "prevent or reduce controversies or disputes," and avoid publishing "fake news."[376] The system is routinely assessed for efficacy, fairness, and security. Content deemed politically sensitive, whether

---

nformation-service-algorithmic-recommendation-management-regulations/ [perma.cc/FC6Y-GP48] (explaining that regulations apply to "personalized recommendations in mobile applications" and require that "algorithm recommendation services" providers uphold certain "user rights").

369. Christina Lu, *China's Social Media Explosion*, FOREIGN POL'Y (Nov. 11, 2021, 10:15 AM), https://foreignpolicy.com/2021/11/11/china-social-media-tech-linkedin-wechat-censorship-privacy-regulation/#cookie_message_anchor/ [perma.cc/Q6RH-DL36].

370. FRIEDRICH-EBERT-STIFTUNG, *supra* note 368, at 4 ("[T]he party-state has recently come to view the consumer internet as a largely unproductive and overcapitalized economic sector, which diverts talent and investment from core technologies critical for economic development—all while threatening social stability.").

371. JUFANG WANG, REGULATION OF DIGITAL MEDIA PLATFORMS: THE CASE OF CHINA 1 (2020).

372. HARRIET MOYNIHAN & CHAMPA PATEL, RESTRICTIONS ON ONLINE FREEDOM OF EXPRESSION IN CHINA: THE DOMESTIC, REGIONAL AND INTERNATIONAL IMPLICATIONS OF CHINA'S POLICIES AND PRACTICES 1 (2021), https://www.chathamhouse.org/sites/default/files/2021-03/2021-03-17-restrictions-online-freedom-expression-china-moynihan-patel.pdf [perma.cc/BAK8-CXQV].

373. FRIEDRICH-EBERT-STIFTUNG, *supra* note 368, at 1.

374. ALINA POLYAKOVA & CHRIS MESEROLE, EXPORTING DIGITAL AUTHORITARIANISM: THE RUSSIAN AND CHINESE MODELS 1, 10 (2019).

375. DEIBERT ET AL., *supra* note 366.

376. FRIEDRICH-EBERT-STIFTUNG, *supra* note 368, at 3.

because it challenges the government's authority, ideological control, or contradicts national interests, is subject to removal or blocking.[377] As the Chinese Communist Party (CCP) relies on morality-based discourses to legitimize internet regulation, platforms frequently suppress content categorized as vulgar, harmful to public morals, or spreading rumors or disinformation.[378]

The China Firewall presents a single narrative that adheres to official governmental values and non-individualized output, censoring diverse outputs disregarding concerns about privacy, or rights that are considered fundamental in the western world.[379] Internet censorship can help maintain political stability without triggering conventional military responses, promote social harmony by preventing the spread of information that may challenge the government's authority, and protect national security by preventing the dissemination of sensitive information that could incite social unrest, violence, or hate speech.[380] This can be seen as an advantage from the perspective of the ruling regime. Yet, preventing individualized content comes at a cost of having only the government-approved information seen. From the perspective of citizens, not only does China's internet policy restrict freedom of expression, limit access to information, and inhibit open dialogue and exchange of ideas, the vague language of the regulations gives the government broad discretion in enforcement and the ability to target differently users who are on a political blacklist.[381] To deter the spread of misinformation or online harassment, the regulation includes a real-name registration and verification policy, so users can be easily identified.[382] But the regulation cannot necessarily combat "fake news" when promoted by the government and may still shape public opinion in ways liberal society would consider unreliable.[383] Moreover, China's internet censorship isolates its population from global discourse and information, which can limit its citizens' understanding of the world and lead to a sense of cultural

---

377. *See generally* Gary King, Jennifer Pan & Margaret E. Roberts, *How Censorship in China Allows Government Criticism but Silences Collective Expression*, AM. POL. SCI. REV., May 2013, at 1 (having "devised a system to locate, download, and analyze the content of millions of social media posts originating from nearly 1,400 different social media services all over China before the Chinese government is able to find, evaluate, and censor").

378. WANG, *supra* note 371.

379. Jack Linchuan Qiu, *Chapter 4: The Internet in China: Technologies of Freedom in a Statist Society*, *in* NETWORK SOC'Y: CROSS-CULTURAL PERSP. 99, 114 (Manuel Castells ed., 2004); FRIEDRICH-EBERT-STIFTUNG, *supra* note 368, at 5.

380. *See* Allenby, *supra* note 89, at 413–26; *see, e.g.*, Strategic Communications, *Hate Speech Poisons Societies and Fuels Conflicts*, EUR. UNION EXTERNAL ACTION (June, 18, 2022), https://www.eeas.europa.eu/eeas/hate-speech-poisons-societies-and-fuels-conflicts_en [perma.cc/R5AX-9QCX] ("[I]n Belarus, every Sunday, state-controlled media features a section called the 'Order of Judas' . . . [which] targets the so-called Belarusians 'regime traitors'—the list is long and includes Belarusian opposition leaders, activists as well as social media influencers, singers and artists, journalists and media personalities, former regime officials, and diplomats.").

381. FRIEDRICH-EBERT-STIFTUNG, *supra* note 368, at 3.

382. WANG, *supra* note 371.

383. Su et al., *supra* note 367 (stating that the primary criticism of the regulations is their apparent lack of enforcement protocol and a need in a public-facing portal that would allow users to seek the government's help regarding violations of the regulations); *see, e.g.*, Daisuke Wakabayashi, Tiffany May & Claire Fu, *As China Looks to Broker Gaza Peace, Antisemitism Surges Online*, N.Y. TIMES (Oct. 28, 2023), https://www.nytimes.com/2023/10/28/world/asia/china-israel-hamas-antisemitism.html (discussing China's heavily censored internet, inflammatory speech critical of Israel is rampant, with emboldened commenters censored).

and intellectual insularity.[384] It is neither a pluralistic nor a democratic solution. According to the International Declaration on Information and Democracy, access to a multitude of information sources and viewpoints is an unequivocal fundamental right.[385] Many Chinese citizens use Virtual Private Networks (VPNs) to bypass the Great Firewall. This leads to a cat-and-mouse game where the government attempts to block VPNs, and citizens find new ways to access restricted content. In a democratic society, efforts to create a safer and more inclusive online environment, especially when it comes at the expense of the above-mentioned values such as diversity of voices, should remain subject to ongoing public deliberation.

B. A Competitive Market

Today, governments in both the United States and Europe are increasingly taking actions against Big Tech platforms under existing antitrust law, aiming to curb the dominance of digital platforms and open the market to greater consumer choices.[386] Francis Fukuyama observes that the reason we do not object when the New York Times declines to publish work by a particular author is because the newspaper market is decentralized and competitive, offering that author many alternative outlets.[387] In contrast, decisions by platforms like Facebook or YouTube to exclude a specific speaker are far more consequential due to their near-monopolistic control over online discourse.[388] Such concentrated power cannot be exercised responsibly unless dispersed within a competitive marketplace. A fair competition framework is essential to ensure that platforms compete fairly with each other, foster pluralistic discourse, and refrain from using internal policies to secure unfair advantage over competitors and users.[389]

Within this framework, two critical requirements frequently emerge: transparency and explainability.[390] Transparency involves making the algorithms, their inputs, and their decision-making processes accessible and understandable to relevant stakeholders in order to ensure accountability, fairness, and ethical considerations. Explainability goes further, demanding clear and comprehensible reasons for algorithmic decisions, enabling users, regulators, and other stakeholders to assess their legitimacy and fairness. Yet, these safeguards only have practical meaning in an environment where users and speakers have meaningful alternatives. The next Part discusses transparency and availability of multiple choices.

---

384. *See supra* notes 353–359 and accompanying text.

385. PLURALISM NEWS, *supra* note 5, at 27.

386. Keller, *supra* note 33, at 263. *But see* FUKUYAMA ET AL., *supra* note 80, at 4 (explaining that antitrust law is designed to redress economic harms where competition is currently lacking, it is not likely to fundamentally reduce the size of the major platforms or to require material changes in their business models. Antitrust enforcement is therefore unlikely to provide an effective remedy for unique political threats to democracy created by platform scale).

387. Fukuyama, *supra* note 191.

388. *Id.*

389. Rebecca Kern, *Push to Rein in Social Media Sweeps the States*, POLITICO (July 1, 2022, 4:30 AM), https://www.politico.com/news/2022/07/01/social-media-sweeps-the-states-00043229. By July 2022, 34 states, both red and blue, pushed bills about online companies handling of User Generated Content (UGC). *Id.*

390. Algorithmic "transparency and explainability" refers to the extent to which the inner workings of an algorithm can be understood and interpreted by individuals, particularly those affected by the algorithm's decisions or those responsible for overseeing its use.

1. Transparency

Users typically lack access to accurate information about the input data used to generate a particular algorithmic output. This opacity hinders the ability to understand, evaluate, or influence the criteria that determine algorithmic decisions and limits the detection of potential dark patterns, manipulations, and embedded biases within these algorithms.[391] The underlying "educate the users" rationale behind transparency suggests that, when better informed, users can make better decisions. The European Union's 2018 General Data Protection Regulation (GDPR) provides an example for an existing regulation of internet algorithms that supports transparency.[392] The GDPR entrenches new user rights, including the right to be informed about the existence of automated decision-making (including profiling),[393] the right to object to the processing of personal data,[394]and the right to obtain human intervention in decisions made solely through automated processing, including the ability to express their point of view and challenge the decision. The call for transparency about data that is fed into algorithms includes how, why, and when algorithms recommend material to various profiles across representative populations.[395]

Within the European Union, the Digital Markets Act (DMA) imposes explicit demands on gatekeepers to publish transparency reports, providing insights into their operations and practices. Transparency obligations include how they treat business users, the data they use in the process of ranking products and services, and the methodologies governing their ranking systems in platform search results. The DMA applies to companies designated as "gatekeepers" by the European Commission, based on three cumulative criteria: (1) a significant impact on the internal market, (2) control of an important gateway for business users to reach end users; and (3) an entrenched and durable market position. Social media platforms are likely to qualify as "gatekeepers" due to their substantial market influence.[396] A

391. GILLESPIE, *supra* note 30. Tarleton Gillespie sheds light on the often obscured and hidden decision-making processes that platforms employ for content moderation. He argues that the lack of transparency in these processes can lead to concerns about accountability and the potential for biases to affect content moderation outcomes.

392. Regulation 2016/679, art. 22(3), General Data Protection Regulation, 2016 O.J. (L 119) 1 (EU) (stating that individuals have the right to obtain human intervention—express their point of view and challenge the decision—in decisions made about individuals based solely on automated processing (e.g., algorithms) that could have a significant impact on them); *id.* at art. 13(2)(f) (stating that individuals have the right to know why their data is being processed and what benefits or reasons the data controller or third party has for processing their data); *id.* at art. 14(2)(g) (stating that individuals have a right to be informed about the recipients or categories of recipients of their personal data, especially when it's collected from a source other than an individual); *id.* at art. 15(1)(h) (stating that individuals have the right to be informed about the existence of automated decision-making—including profiling—and meaningful information about the logic involved, as well as the significance and envisaged consequences of such processing).

393. GDPR art. 22(1) ("The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.").

394. *Id.* at art. 21.

395. Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343, 1348 (2018); PLURALISM NEWS, *supra* note 5, at 77.

396. *Questions and Answers: Digital Markets Act: Ensuring Fair and Open Digital Markets*, EURP. COMM'N (Sept. 5, 2023), https://ec.europa.eu/commission/presscorner/detail/en/qanda_20_2349.

single company may be designated as a gatekeeper for multiple core platform services.

The European Digital Services Act (DSA), passed in July 2022, has introduced few types of transparency measures for online platforms.[397] Article 24(2) of the DSA requires that very large online platforms (VLOPs) and very large online search engines (VLOSEs) inform users in their terms of service if content users are interacting with has been algorithmically generated or manipulated. Under Article 27 of the DSA, platforms that use recommender systems must provide meaningful information about the main parameters used to determine which content is shown to users.[398] According to Article 26(3) of the DSA, platforms are required to provide users with clear information about why they are being targeted with specific advertisements, as well as offer guidance on how to change the parameters used advertisement targeting. Online platforms must also allow their users to influence the parameters used by recommender systems, including offering at least one option to opt out of recommendations based on profiling.[399] The DSA further mandates routine and comprehensive transparency reporting to be shared with independent auditors, supervisory authorities, and researchers from academia and civil society, to support the identification and mitigation of systemic risks.[400] Finally, the DSA requires transparent disclosures to users about online advertising practices. These obligations are not uniformly imposed all platforms but are tailored proportionally to the platform's size, function, and societal impacts that different types of platforms have. Alongside these European initiatives, national legislators are introducing measures to enhance the transparency of these digital services.[401]

In the United States, several similar federal legislative proposals aimed at enhancing transparency are currently pending.[402] For example, the Filter Bubble

---

397. *Commission Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services* (Digital Services Act) and amending Directive 2000/31/EC, COM (2020) 825 final (Dec. 15, 2020), https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM%3A2020%3A825%3AFIN (noting that this applies to hosting services, online marketplaces, and social media networks, platforms must report their efforts to combat illegal content, including the number of removal requests received and the actions taken).

398. EUROPEAN DATA PROTECTION SUPERVISOR, OPINION 1/2021 ON THE PROPOSAL FOR A DIGITAL SERVICES ACT 16–17 (2021), https://edps.europa.eu/system/files/2021-02/21-02-10-opinion_on_digital_services_act_en.pdf ("[M]odify or influence th[e] main parameters" of the recommender system including "at least one option [for engaging with the service] which is not based on profiling.").

399. Gordon-Tapiero et al., *supra* note 16, at 661.

400. *See also* EU COMM'N, *Shaping Europe's Digital Future: AI Act* (Nov. 15, 2023), https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai [perma.cc/J4BE-L6HC] (describing that the Commission is proposing a risk-based legal framework on AI, using clear rules for online platforms to address illegal content, including hate speech, terrorist content, and more. It can also ban decision-making algorithms in cases where they pose a threat to "safety, livelihoods and rights of people"). *But see* Inioluwa Deborah Raji, Timnit Gebru, Margaret Mitchell, Joy Buolamwini, Joonseok Lee & Emily Denton, *Saving Face: Investigating the Ethical Concerns of Facial Recognition Auditing*, *in* PROCS. AAAI/ACM CONF. ON A.I., ETHICS, & SOC'Y 145, 146, 147 (2020) (noting that algorithmic auditing presents ethical concerns).

401. *See, e.g.*, PLURALISM NEWS, *supra* note 5, at 40 (reporting that in the UK, the Online Safety Bill is expected to introduce new transparency requirements for platforms and their recommender systems, particularly as it relates to combating illegal content and legal, but harmful content).

402. For Federal bills, see, for example, Algorithmic Justice and Online Platform Transparency Act, S. 2325, 118th Cong. (2023) (proposing to increase transparency of algorithms used to amplify and moderate content); Digital Services Oversight and Safety Act of 2022, H.R. 6796, 117th Cong. (2022)

Transparency Act requires social media platforms to disclose that the content users see and the order in which it is appears is determined by an algorithm and based on user-specific data or inferences drawn from such data. The Act seeks to provide users with the option to access content in a format not filtered or ranked by opaque algorithmic processes, or allow users to opt out of the so-called "filter bubble" and instead offer an input-transparent version of the platform's services (i.e., a newsfeed not algorithmically personalized based on user-provided information).[403]

On the users' end, transparency alone is unlikely to meaningfully enhance public understanding of how current industry practice affect their daily lives unless the information disclosed is clear and explainable. However, even where transparency requirements are met, they fail to address the significant switching costs that users face when trying to move between platforms, in response to disclosed practices they may find objectionable.[404] By the time users become aware of problematic uses of their data, they are already entrenched within a particular platform's digital ecosystem through interconnected applications and devices.[405] The EU's DMA directly addresses these barriers by focusing on large online platforms that function as gatekeepers. Its primary objective is to ensure fair competition in the digital market and prevent anti-competitive behaviors. The DMA prohibits gatekeepers from exploiting their dominant positions by, for example, affording preferential treatment to their own proprietary products, content, or information through various means including their ranking and recommender systems, at the expense of third-party offerings on their platforms.[406] It further mandates interoperability obligations, requiring platforms to allow third-party providers to integrate and interact with the platform's own services, thereby reducing user lock-in and enhancing market openness.

Moreover, to meaningfully address the harms of personalization, transparency must be accompanied by baseline agreements on acceptable standards for personalization. Absent such standards, platforms could follow unethical standards as long as they disclose them. Moreover, transparency alone is insufficient without robust enforcement mechanisms and institutional oversight.[407] Legal frameworks can play a critical role in promoting ethical guidelines for algorithmic design that go beyond disclosure.[408] Regulation may even require a mandate to limit algorithmic

---

(covers recommender systems, including requirements for transparency and opt-in consent.); Filter Bubble Transparency Act, S. 2024, 117th Cong. (2021) [hereinafter Filter Bubble Transparency Act]; Platform Accountability and Transparency Act, S. 1876, 118th Cong. (2023) (requiring increased transparency of algorithms used to recommend content.); Kids Online Safety Act, S. 1409, 118th Cong. (2023) (requiring platforms to disclose information about recommendation systems and give minors a way to opt out of recommended content).

403. Filter Bubble Transparency Act § 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."); *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]."); *see also* Gordon-Tapiero et al., *supra* note 16, at 665.

404. Popescu & Baruh, *supra* note 69, at 282.

405. *Id.*

406. PLURALISM NEWS, *supra* note 5, at 41.

407. Gordon-Tapiero et al., *supra* note 16, at 667; PLURALISM NEWS, *supra* note 5, at 40.

408. *See, e.g.*, *Diversity of Content Online*, GOV'T OF CAN. (2021), https://www.canada.ca/en/canadian-heritage/services/diversitycontent-digital-age.html. The Canadian government's recent development of non-binding guiding principles to promote diverse and pluralistic sources of news and

reliance on popular engagement metrics and reduce discriminatory impacts.[409] For example, in cooperation with The International Conference of Data Protection and Privacy Commissioners (ICDPPC), the Organization for Economic Co-operation and Development (OECD), United Nations Educational, Scientific and Cultural Organization (UNESCO), and the Public Voice project, along with other international organizations, civil society leaders, and government officials, published ethical recommendations focused on design of systems, titled Universal Guidelines for Artificial Intelligence.[410] The Guidelines adopt users' right to transparency; to know the basis of an AI decision that concerns them, including access to the factors, logic, and techniques that produced the outcome; a right to a final determination made by a person to ensure a more nuanced understanding of individual circumstances; fair and non-discriminatory practices based on accurate data; and a prohibition on secret profiling, opaque scoring systems, and autonomous systems that preclude human oversight. Such transparency is crucial to hold institutions accountable for the consequences of their automated decisions making.

Striking an effective balance between personalization features, exposure to diverse viewpoints, respectful dialogue, and democratic deliberation can incentivize user engagement and foster a healthier social media environment.[411] However, the pursuit of fairness in algorithmic systems face significant challenges rooted in both biases embedded in training data and broader societal inequalities.[412] Modifying algorithms to align with fairness objectives may produce unforeseen consequences, including reinforcing existing biases or introducing new ones. The complexity of algorithmic systems often obscures their inner workings, making oversight and correction of bias more difficult and raising questions of legal responsibility when algorithmic decisions yield harmful or discriminatory outcomes.[413] Tensions

information through recommender systems is a noteworthy step in fostering a healthier public discourse. It underscores the importance of diverse cultural content and information in a democratic society.

409. *See generally* Daniel E. Ho & Alice Xiang, *Affirmative Algorithms: The Legal Grounds for Fairness as Awareness*, 2020 U. CHI. L. REV. ONLINE 134 (2020). Ho and Xiang propose a new framework for fairness, which they term "fairness as awareness," in the context of algorithmic decision-making, to mitigate potential discriminatory impacts, beyond traditional nondiscrimination approaches such as anti-discrimination laws and privacy regulations. They argue that providing transparency and understanding of algorithmic processes can help identify and rectify algorithmic biases that may lead to disproportionate harm to certain groups and empower individuals to contest unfair decisions and promotes accountability.

410. Public Voice, *Universal Guidelines for Artificial Intelligence*, BOSTON GLOB. F. (Oct. 23, 2018), https://bostonglobalforum.org/initiative/aiws-and-the-age-of-global-enlightenment/7-layer-model-of-ai-world-society/universal-guidelines-for-artificial-intelligence-of-epic/ (explaining that the Public Voice coalition was established in 1996 by the Electronic Privacy Information Center (EPIC) to promote public participation in decisions concerning the future of the internet).

411. Gausen et al., *supra* note 53, at 18.

412. SASHA CONSTANZA-CHOCK, DESIGN JUSTICE: COMMUNITY-LED PRACTICES TO BUILD THE WORLDS WE NEED 44 (2020). *See, e.g.*, Sara Chodosh, *Courts Use Algorithms to Help Determine Sentencing, but Random People Get the Same Results*, POPULAR SCI. (Jan. 18, 2018, 9:00 PM), https://www.popsci.com/recidivism-algorithm-random-bias/ [perma.cc/TM82-2E9F]; Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 PROC. MACH. LEARNING RSCH. 1 (2018) (discussing studies that algorithms discriminate based on race and gender).

413. Nicol Turner Lee, Paul Resnick & Genie Barton, *Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms*, BROOKINGS (May 22, 2019), https://ww

between fairness and other crucial design goals, such as accuracy, efficiency, and profitability, add additional layers of complexity, raising questions about accountability in instances where algorithmic outputs produce adverse consequences.[414] Moreover, the contextual nature of fairness further complicates matters. Critics contend that algorithmic intervention might inadvertently perpetuate systemic discrimination without effectively addressing its underlying causes.[415] While some scholars advocate for fairness frameworks that address the broader societal context, ethical design principles may inadvertently reflect dominant perspectives, potentially sidelining alternative viewpoints.[416]

The feasibility of aligning algorithmic design with principles of transparency and explainability remains uncertain, and depends on their voluntary nature, competing business interests and limited institutional accountability. When guidelines are adopted on a voluntary basis by entities that fund, develop, and deploy algorithmic systems, a complex challenge arises: the lack of binding obligations raises questions about the incentives for embracing transparency and ethical standards. In the absence of enforceable mandates, commercial priorities may undermine efforts to implement meaningful oversight and safeguard public interest.

2. Middleware Market

To address asymmetric power relations between companies and consumers in a non-competitive environment, we should facilitate competition in alignment with market-based dynamic. Fukuyama et al. suggest creating a new market for recommender systems unbundled from the hosting platforms, thereby enabling users to choose among competing content-moderation and curation services.[417] Middleware, a software operating between users and major platforms provided by a third party, could offer services that supplement those currently provided by the major platforms.[418] Services such as fact-checking, customized news rankings (how users would like it to be ranked), relevance prioritization, identification of trustworthy sources, privacy controls, information filtering, and features that promote security, safety, and pluralism could be integrated into any platform, allowing users to tailor editorial judgments, thereby decentralizing content curation. By shifting control over personalization and filtering from platforms to an open, competitive market of middleware providers, this model could reduce monopolistic control, incentivize transparency, flexibility and user autonomy over personalization

---

w.brookings.edu/articles/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/ [perma.cc/9CMC-JVV7]. *See also* Chodosh, *supra* note 412; Villasenor, *supra* note 27.

414. Melissa D. McCradden, Shalmalia Joshi, Mjaye Mazwai & James A. Anderson, *Ethical Limitations of Algorithmic Fairness Solutions in Health Care Machine Learning*, 2 LANCET 221, 221 (2020).

415. *See, e.g.*, Pauline T. Kim, *Auditing Algorithms for Discrimination*, 166 U. PA. L. REV. ONLINE 189, 191 (2017) (arguing that algorithms should be audited for bias "because the causes of bias often lie not in the code, but in broader social processes"); McCradden et al., *supra* note 414, at 221.

416. McCradden et al., *supra* note 414, at 222; Deborah Hellman, *Measuring Algorithmic Fairness*, 106 VA. L. REV. 811, 814 (2020). *See also* Sina Fazelpour, Zachary C. Lipton & David Danks, *Algorithmic Fairness and the Situated Dynamics of Justice*, 52 CAN. J. PHIL. 44, 44 (2020).

417. FUKUYAMA ET AL., *supra* note 80.

418. *Id.* at 3.

settings, terms of service, privacy and other services that users care about.[419] Practically, Fukuyama's proposal is a technologically compatible with the current digital ecosystem, thereby enhancing its practical feasibility while minimizing friction with the existing digital ecosystem.[420]

The rationale behind the middleware market stems from the asymmetry inherent in algorithmic personalization. Platforms design and control the algorithm that target users and prioritize what they see, leaving users subject to decisions outside their own control.[421] At the individual level, users have limited control over what content the algorithm prioritizes for them and cannot influence the variables that shape algorithmic outputs.[422] With few options to choose from, users' choices are neither meaningful, nor reflect an informed or tailored user experience. Moreover, assumptions about users' identities will likely play a major role in determining what kind of information each user will be matched with and exposed to.[423] These assumptions are based on limited interaction and behavior of people online may not necessarily represent their full personhood and interests. For instance, a single online search may be interpreted as an enduring preference, even though it does not necessarily indicate that in real life the user lacks interest in diverse points of view. People may value less skewed information and pursue accuracy, even when their immediate online behavior suggests otherwise.[424] Regulations should protect users from the consequences of wrong assumptions algorithms make about them, or at least affirm their right not to be defined or constrained by such wrong assumptions.

A fully developed market of middleware systems would dilute platforms' centralized control over the organization of online information and decentralize decision-making regarding what users see. The market would allow each user to choose customized content-matching values and impact content-matching filters upstream. In this model, users autonomy plays a role because users themselves would be handed the power to choose their preferred recommender system, albeit from among the options offered by competing market actor,[425] akin to selecting mobile applications via app stores on operating systems such as Android or iOS.[426] Each user could select providers that reflect their informational priorities, normative values, and have earned their trust, thus fostering a more pluralistic and responsive

419. *Id.* at 6, 8. A business model for middleware providers would have to be sufficiently attractive to induce an adequate supply. Fukuyama et al. propose that the most logical approach would be to establish revenue sharing arrangements between the dominant platforms and the third-party providers of middleware. If a middleware product enhances the value of the platforms to users, the platforms might be able to generate increased advertising (or maybe, in the future, user fee) revenues that could be shared with the middleware provider. Alternatively, the middleware provider might be able to charge user fees or sell advertising directly.

420. Fukuyama, *supra* note 191, at 37–44.

421. Keller, *supra* note 33, at 248.

422. Batya Friedman & Helen Nissenbaum, *Bias in Computer Systems*, 14 ACM TRANSACTIONS ON INFO. SYS. 330, 339 (1996).

423. SUNSTEIN, *supra* note 54, at 116.

424. Jaeho Cho, et al., *supra* note 14, at 166. *See also* William Hart et al., *supra* note 69, at 555–88.

425. Fukuyama, *supra* note 191, at 42; FUKUYAMA ET AL., *supra* note 80, at 11–12.

426. FUKUYAMA ET AL., *supra* note 80, at 7. *See also* Keller, *supra* note 33, at 263.

information environment.[427] Such a market would encourage critical engagement, prompting users to evaluate which recommender system(s) best serve their needs and to make deliberate choices about how content is curated and presented to them.[428]

However, the establishment of a middleware market faces significant structural and normative challenges, ranging from regulatory complexities to resistance from dominant platforms. First, the introduction of middleware does not inherently guarantee the creation of more diverse and nuanced profiles than those produced by existing platform-driven personalization systems. Within the problem of "individualized manipulated content-matching," the middleware market addresses the "manipulated" dimension. Since the objective behind the middleware concept is not to suppress content-matching but rather to give users a choice about how it is done, it will not necessarily prevent the clustering of like-minded consumers.[429] The act of selecting among middleware providers, each with distinct targeting logic, requires some sort of "self-profiling," rather than one that is done by platforms. Self-profiling might inadvertently reproduce the same dynamics of echo chambers, identity-based segmentation, and selective exposure, the same problems that personalization by platforms currently facilitates. The model raises further questions: would users be required to select a profile or recommender system, or could they simply opt into the default system provided by the platform? Since users seem to be comfortable with the current system, they may lack the incentive to switch, even when alternatives are available. While the middleware model introduces choice and decentralization, in the absence of regulatory safeguards against cognitive and ideological fragmentation to counteract the reinforcement of existing biases and the creation of identity-based divisions among users, the middleware market is also unlikely to remedy the full scope of the problem.[430]

Second, establishing a functioning middleware market is a complex task that requires specialized technical expertise and a regulatory framework.[431] Enabling such a market would require congressional action to either empower an existing agency or create a new regulatory body overseeing integration protocols, revenue-sharing arrangements, and compliance standards for middleware providers.[432] Absent legislative authorization and regulatory oversight, the dominance of existing platform giants could suppress competition, thereby preempting the development of a viable middleware ecosystem.[433] Safeguards must be established to ensure a

427. Lauren Jackson & Desiree Ibekwe, *Jack Dorsey on Twitter's Mistakes*, N.Y. TIMES (Aug. 19, 2020), https://www.nytimes.com/2020/08/07/podcasts/the-daily/Jack-dorsey-twittertrump.html [perma.cc/3H7X-LLJN] ("We need to open up and be transparent around how our algorithms work and how they're used, and maybe even enable people to choose their own algorithms to rank the content or to create their own algorithms, to rank it.").

428. PLURALISM NEWS, *supra* note 5, at 41.

429. Fukuyama, *supra* note 191, at 42.

430. CONSTANZA-CHOCK, *supra* note 412 (suggesting that promoting inclusive design principles, transparency in how algorithms make decisions, preventing discriminatory practices, accessibility standards, and measures that ensure the representation of diverse voices and perspectives, and foster accountability for the impact of AI on democratic processes).

431. *Id.* at 10.

432. *Id.* at 11.

433. Fukuyama, *supra* note 191, at 43.

healthy and competitive environment, interoperability, and non-discriminatory treatment of competing middleware providers. Without such protections, middleware firms risk becoming extensions of the very dynamics they were designed to counteract, incentivized to maximize engagement rather than uphold user agency or content diversity.

Finally, for a middleware software market to thrive and fulfill its intended function to decentralize control and enhance user autonomy, it must be characterized by genuine plurality and should not fall under the dominance of a single entity. The emergence of a single dominant middleware provider would defeat the purpose of introducing competition and choice, and may replicate the power dynamic users currently have with platforms, undermining the rationale for middleware altogether.[434] Given the potential disruption to the status quo and a challenge to the current business model of platforms, entrenched platforms may engage in significant lobbying and political friction to resist the regulatory and structural shifts as well as the loss of control over content required to accommodate a middleware market. Addressing these concerns is essential to realizing middleware's potential as a meaningful corrective to platform power.

V. CAN WE MAINTAIN BOTH PERSONALIZATION FEATURES AND DEMOCRATIC PUBLIC DISCOURSE?

The purpose of this Article is to highlight a regulatory blind spot: the impact of personalization features on public discourse, particularly as shaped by social media algorithms. It examines how two specific personalizing mechanisms, amplification algorithms and recommendation systems, interact to produce a largely overlooked and under-theorized harm: the delivery of individualized manipulated content to each member of the audience. This form of algorithmic tailoring operates at the level of the individual, shaping perceptions and discourse in ways that evade traditional regulatory frameworks and challenge existing conceptions of democratic discourse.

Problems from individualized manipulated content matching may have existed in traditional media and past public discourse; however, this Article underscores a key distinction between those earlier methods and individualized targeting capabilities in the digital age. Today's personalization relies on non-transparent technical practices that create capabilities to selectively distribute individualized manipulated content, different in scale, precision, and societal impact, with minimal to no legal, ethical, or professional accountability. Does that mean that the very concept of personalization is inherently incompatible with democratic values? Between the technical imperative to tailor content to each individual, and the regulatory challenge posed to the public sphere by both individualizing features and manipulation options—perhaps it is. Perhaps certain forms of personalization undermine the foundations of democratic discourse.

This Article builds on the growing field of algorithm regulation, by examining the feasibility of preserving democratic public discourse in the face of algorithms' capacity to selectively distribute individualized manipulated content to each member

434. FUKUYAMA ET AL., *supra* note 80, at 7 ("[R]eforming antitrust's [sic] consumer welfare standard would cause damaging policy incoherence, invite politically motivated and unprincipled actions, and erode support for procompetitive policies.").

of the audience. Current regulatory initiatives, which often address specific aspects of personalization and challenges in the personalized sphere, fall short of effectively confronting the full scope of the harm to public discourse: these initiatives address the nature of personalized content, but not the mechanism of its selective distribution; the scale of amplification, but not the logic behind content ranking for individualizing matching; and user control over personal data collected to profile them, but not the systemic manipulation of sub-populations within the collective database. Each approach addresses a component of the problem, but fails to engage with its structural totality.

Additionally, attempts to regulate either online discourse or algorithmic personalization features often raise significant First Amendment concerns. Even targeted interventions that address specifically individualized manipulated content-matching by algorithms—such as transparency mandates, ethical algorithmic design, and the creation of a middleware market—are limited in their capacity to address the broader erosion of public sphere and the risk of social indoctrination through individualized manipulated content-matching and microtargeting. These approaches frame the issue as lack of awareness or autonomy trapping, implying that greater awareness or choice can mitigate the risks. Yet, while autonomy and informed citizenry are essential to democratic functioning, such solutions overlook the deeper structural harm: the fragmentation of shared discourse through the systematic "divide and target" strategies embedded in platform architectures. These approaches also disregard the entrenched economic incentives of platform operators whose business models depend on maximizing engagement through manipulation and segmentation. As long as algorithmic infrastructures remain optimized for individual-level manipulation and narrative-based communities, the ethical and deliberative foundations of democratic society remain vulnerable. Addressing this threat requires confronting not only the technological mechanics but also the political economy that sustains them.

	What elements the regulation addresses	**Potential and Limitations**
Content regulation	Addresses the content of expressions matched to users: extreme content and illegal expression.	Can affect the content and nature of the discourse, but does not address targeting practices, and risks First Amendment violations.
Content-neutral regulation	Addresses the algorithm that matches content.	Can potentially affect criteria for content-matching or content removal, but cannot prevent manipulated individualized content flow and risks First Amendment violations.
Privacy Rights approach	Addresses the data used for profiling: Data collection and sharing.	Can minimize data for profiling but doesn't change neither the criteria for individual content-matching nor manipulative targeting.
China's Firewall – Cohesive Narrative	Addresses the public discourse by preventing multiple narratives.	Allows for cohesive narratives, but non-democratic and violates privacy. The narrative is not individualized, but it is manipulated.

Middleware market – Choice between personalization criteria	Addresses users' lack of control when third parties personalize content to them, in particular content-matching criteria.	Cannot prevent manipulated individualized content flow, but has the potential to refine the criteria for matching.

Table 1

It is, perhaps, imprecise to say that regulation fails to address the impact of personalization features on public discourse but is more accurate to say it is ill-equipped to manage it. This Article does not purport to offer a definitive regulatory blueprint or singular solution, which is well beyond the scope of this project. However, it aims to highlight critical considerations for any path forward. First, to address the erosion of public discourse and pluralistic-democratic values, it is imperative to nurture a public sphere for public discourse. Solutions grounded solely in the private or personalized user sphere are unlikely to remedy problems that originate in the structural degradation of public deliberation. The public sphere, as the heart of democratic societies, plays a crucial role in facilitating collective understanding, accommodating pluralism, and fostering social cohesion. Therefore, legal scholarship should shift its focus from merely regulating the features that harm individual users' sphere to imagining and supporting alternatives capable of sustaining democratic discourse. In forging such alternatives, two key challenges must be acknowledged. First, a relevant alternative public sphere might be susceptible to the same distortions that plague the current personalized ecosystem. Therefore, efforts to construct or support such a space should be guided by understanding of the potential power dynamics that shape and potentially compromise impartiality and openness. Second, this approach may prove difficult to reconcile with prevailing interpretations of the First Amendment. Indeed, classical First Amendment doctrine—developed in an era of slow communication, information scarcity, and state-centered power—is increasingly out of step with the realities of a digital information environment dominated by private platforms and individualized content feeds. Many of the rationales underlying the protection of free speech when applied to new technologies cannot ensure that people are exposed to new ideas, get informed, or are able to express themselves freely.[435] Contemporary information governance blurs the boundaries between public regulation and private ordering in ways that the traditional constitutional framework is ill-suited to address. In today's environment, when as a practical matter free speech is effectively mediated through platform terms of service and algorithmic systems engineered to exploit cognitive biases, the traditional constitutional framework offers limited protection for meaningful expression or exposure to diverse viewpoints. This doctrinal obsolescence suggests that prevailing interpretations of the First Amendment may need to be revisited. There are calls to steer away from the individualized sphere. Balkin, for example, supports the view that content moderation is less a problem of constitutional law [. . .] and more a problem of technology and administrative regulation, acknowledging that the administrative agency in this case is a private company.[436] Douek argues that the

435. *See supra* Section III.A: Content Regulation.

436. Jack M. Balkin, *The Future of Free Expression in a Digital Age,* 36 PEPP. L. REV. 427, 441

narrow approach to content moderation that focuses on a singular post and represented as an aggregation of many individual adjudications, is misguided. Instead, she suggests we should talk about content moderation as a project of mass speech administration through a systems thinking approach, not a free speech problem.[437]

The second thing to keep in mind is that any viable solution may be outside the scope of law alone and require a multi-layered interdisciplinary approach. Addressing the fragmentation of public discourse and the erosion of pluralistic-democratic values requires a broader societal response that encompasses ethical design, technological reform, public education, and cultural engagement. Goals that regulation, in isolation, is ill-equipped to achieve. To preserve the integrity of public discourse and pluralistic-democratic values, a profound examination of how technology shapes collective emotions, perceptions, and civic engagement, how to foster constructive dialogue, respect, and tolerance within the digital ecosystem, is required. This examination must be coupled with public education initiatives that illuminate the social and political implications of algorithmic personalization. In this sense, the path forward involves not only legal innovation but also a fundamental reassessment of the role of technology in shaping the public sphere. Efforts in this direction will inform policy development and public understanding of the trade-offs between personalization features and collective democratic values. Ultimately, this may foster governance models better aligned with the ethical imperatives and societal needs of the digital age.

(2009). *See also* Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598, 1621 (2018); Marvin Ammori, *The "New" New York Times: Free Speech Lawyering in the Age of Google and Twitter*, 127 HARV. L. REV 2259, 2262 (2014).

437. Evelyn Douek, *Content Moderation as Systems Thinking*, 136 HARV. L. REV. 526, 531 (2022).