MACHINE MANIPULATION: WHY AN AI EDITOR DOES NOT SERVE FIRST AMENDMENT VALUES

Alec Peters*

The past few years have seen increasing calls for regulation of large social media platforms, and several states have recently enacted laws regulating their content moderation, promotion, and recommendation practices. But if those platforms are exercising editorial discretion when carrying out these tasks, many of the regulations will run into constitutional concerns: the First Amendment protects the "exercise of editorial control and judgment" by publishers over their choice of content and how it is presented. However, the editorial operation of social media platforms differs significantly from traditional media, most importantly in the use of artificial intelligence (AI) for editorial decision-making. While courts have thus far not given much attention to the implications of this use, this Note argues that using AI for editorial decision-making should not be entitled to the same protection as a human decision-maker. After introducing how social media platforms employ AI in their editorial operations, it outlines the foundational values of democratic self-governance, the marketplace of ideas, and autonomy underlying the First Amendment, and assesses how the use of AI impacts those values. The Note concludes that the ability of AI to manipulate human behavior and preferences, combined with the delegation of decisional autonomy from humans to AI, harms the foundational First Amendment values. Therefore, the use of AI is not deserving of the same protection as human editorial decision-making.

---

* I would like to thank the wonderful editors of the *University of Colorado Law Review*, including Jonathan Murray, John Bellipanni, Louise Fliegel, Alexandra Nielsen, Mary Slosson, and Jenna King for their thoughtful feedback through many rounds of edits and what must have appeared to be my best efforts to make their job as difficult as possible. Many thanks also to Editor-in-Chief Caitlin Dacus and Executive Editor Casey Nelson for their care and leadership in guiding me through this process. Finally, I would like to thank Professors Helen Norton and Blake Reid for helping me organize and develop my ideas with their feedback and guidance.

INTRODUCTION

The past few years have seen increasing calls for the regulation of large social media platforms as concerns over content moderation practices, disinformation, and political polarization grow. In addition to a variety of proposed legislation at the federal level, at least four states have enacted laws regulating their content moderation, promotion, and

recommendation practices.[1] But if those platforms are exercising their editorial discretion when carrying out these tasks, many of the regulations will run into constitutional concerns: the First Amendment protects publishers' "exercise of editorial control and judgment" over their choice of content and how it is presented.[2] Indeed, the Eleventh Circuit has already struck down the majority of Florida's social media statute by relying on precedent from First Amendment cases involving traditional forms of media and expressive conduct.[3]

While social media platforms undoubtedly exercise some degree of editorial control over the content allowed on their sites and how it is presented to users, they do so in a different manner than traditional media like newspapers and television broadcasters.[4] In particular, they use artificial intelligence ("AI") to carry out a substantial portion of editorial decision-making on their platforms. This Note considers whether an editor employing AI to exercise editorial discretion should be entitled to the same protections under the First Amendment as a human editor. When applying the First Amendment to new expressive technologies, courts rely on consideration of the fundamental values underlying the freedom of speech to understand how they are impacted by the new technology.[5] This Note explains how the use of AI for editorial decision-making harms these values and concludes that it should not be entitled to the same protection as a human editor.

AI is the foundation of modern social media.[6] Yet recent legislation—inspired by partisan public discourse over

---

1. *See, e.g.*, FLA. STAT. § 501.2041(2)(b) (2021); TEX. BUS. & COM. CODE ANN. § 120.051(a) (West 2021); Algorithmic Justice and Online Platform Transparency Act, S. 1896, 117th Cong. (2021).

2. Miami Herald Pub. Co. v. Tornillo, 418 U.S. 241, 258 (1974).

3. NetChoice, LLC v. Att'y Gen., Fla., 34 F.4th 1196, 1213, 1226–27 (11th Cir. 2022), *cert. granted in part sub nom.* Moody v. Netchoice, LLC, No. 22-277, 2023 WL 6319654 (U.S. Sept. 29, 2023), and *cert. denied sub nom.* Netchoice v. Moody, No. 22-393, 2023 WL 6377782 (U.S. Oct. 2, 2023).

4. Evelyn Douek & Genevieve Laker, *Rereading "Editorial Discretion,"* KNIGHT FIRST AMEND. INST. BLOG (Oct. 24, 2022), https://knightcolumbia.org/blog/rereading-editorial-discretion [https://perma.cc/VV6A-C532] ("Recognizing that the First Amendment protects editorial discretion does not mean that figuring out when or how social media platforms actually *exercise* that editorial discretion is an easy task.").

5. *See, e.g.*, Red Lion Broad. Co. v. F.C.C., 395 U.S. 367 (1969).

6. *See, e.g.*, Rachel Metz, *Facebook's Top AI Scientist Says It's 'Dust' Without Artificial Intelligence*, CNN BUS. (Dec. 5, 2018), https://www.cnn.com/2018/12/05/tech/ai-facebook-lecun [https://perma.cc/EAL7-LK6J]; Chris Meserole, *How Do*

misinformation, hate speech, and political bias—has been primarily focused on the platforms' content moderation practices.[7] Indeed, the role of AI is generally misunderstood or ignored in the public discourse around social media.[8] Moreover, the effects of AI are underappreciated and poorly addressed by courts, which are quick to write off this aspect of social media as a "distraction" from relevant First Amendment considerations.[9]

However, a closer examination of the operation of AI within social media platforms illustrates that AI's use in this context is a significant departure from all prior editorial practices—one that requires careful thought and attention in applying existing doctrine if foundational First Amendment values are to be protected. Writing about violent video games in 2011, Justice Alito cautioned against an inflexible application of First Amendment doctrine that had been fashioned for a different time:

> In considering the application of unchanging constitutional principles to new and rapidly evolving technology, this Court should proceed with caution. We should make every effort to understand the new technology. We should take into account the possibility that developing technology may have important societal implications that will become apparent only with time. We should not jump to the conclusion that new technology is fundamentally the same as some older thing with which we are familiar.[10]

If there is any technology to which these words apply, it is AI.

This Note examines the use of AI by social media platforms and considers how it impacts the foundational values underlying

---

Recommender Systems Work On Digital Platforms?, TECHSTREAM, BROOKINGS INST. (Sept. 21, 2022), https://www.brookings.edu/techstream/how-do-recommender-systems-work-on-digital-platforms-social-media-recommendation-algorithms [https://perma.cc/XEH8-9TEQ].

7. *E.g.*, TEX. CIV. PRAC. & REM. CODE § 143A.002; *see* Rebecca Kern, *Push to Rein in Social Media Sweeps the States*, POLITICO (July 1, 2022), https://////https://www.politico.com/news/2022/07/01/social-media-sweeps-the-states-00043229 [https://perma.cc/JC2W-QU29].

8. *See, e.g.*, Meserole, *supra* note 6.

9. NetChoice, Inc. v. Paxton, 573 F. Supp. 3d 1092, 1108 (W.D. Tex. 2021), *vacated and remanded sub nom.* NetChoice, L.L.C. v. Paxton, 49 F.4th 439 (5th Cir. 2022), *cert. granted in part sub nom.* Netchoice, LLC v. Paxton, No. 22-555, 2023 WL 6319650 (U.S. Sept. 29, 2023).

10. Brown v. Ent. Merchants Ass'n, 564 U.S. 786, 806 (2011).

editorial rights protected by the First Amendment. First, AI on social media platforms is used to manipulate behavior and preferences, going beyond the capabilities of human editors in ways that are fundamentally counter to First Amendment values. Commentators are quick to imagine science fiction stories of runaway AI maliciously controlling us and taking over the world.[11] But one need not imagine future "super-intelligent" AI to see its effects today. With the amount of individual behavioral data that platforms are collecting from users, manipulation of human behavior is a relatively simple task for present-day AI, and it forms the foundation of social media platforms.[12]

Second, the platforms' delegation of their editorial discretion to AI is a release of their own autonomy over decision-making that does not warrant the same protection under the First Amendment. Employing AI for editorial decision-making, as with all uses of AI, necessarily involves a loss of control over the actions taken when compared to traditional algorithms or human decision-making. As a result, when and to the extent that a platform employs AI in its editorial operation, the platform is giving up control of the decision-making process over what content is presented to users. In sum, using AI to manipulate users through the delegation of editorial decision-making does not serve the foundational values of the First Amendment. It is not deserving of the same protections as human decision-making.

---

11. *See* Christine Moser et al., *What Humans Lose When We Let AI Decide*, MIT SLOAN MGMT. REV. (Feb. 7, 2022), https://sloanreview.mit.edu/article/what-humans-lose-when-we-let-ai-decide [https://perma.cc/M4FH-A3R6] ("But instead of worrying about futuristic sci-fi nightmares, we should instead wake up to an equally alarming scenario that is unfolding before our eyes: We are increasingly, unsuspectingly yet willingly, abdicating our power to make decisions based on our own judgment . . . ."); François Chollet, *What Worries Me About AI*, MEDIUM (Mar. 28, 2018), https://medium.com/@francois.chollet/what-worries-me-about-ai-ed9df072b704 [https://perma.cc/ML9A-LLLQ].

12. *See* Bruce Schneier, *The Coming AI Hackers*, HARVARD KENNEDY SCH., BELFER CTR. FOR SCI. & INT'L AFFAIRS (Apr. 2021), https://www.belfercenter.org/publication/coming-ai-hackers [https://perma.cc/V9L7-KQMG]; *Q&A with Michael Schrage: The Pull of Recommendation Engines*, MIT INITIATIVE ON THE DIGIT. ECON. (Aug. 31, 2020), https://ide.mit.edu/insights/qa-with-michael-schrage-the-pull-of-recommendation-engines [https://perma.cc/98GD-QD7Y]; Chollet, *supra* note 11 (explaining that "Facebook's business lies in influencing people" and that "I chose to write about mass population manipulation specifically because I see this risk as pressing and direly under-appreciated.").

This Note proceeds in four parts. Part I describes how prominent social media platforms generally use AI in their operations. Part II outlines modern editorial rights under the First Amendment. Part III explores how those rights are understood in the context of different technologies, which, in turn, require different First Amendment analyses. In considering each new technology, the Supreme Court has focused on protecting the foundational values underlying the First Amendment. Part IV examines how the editorial discretion exercised by modern social media differs from traditional media because of the use of AI and how that distinction matters for the protection of First Amendment values. Finally, this Note concludes by considering the implications for evaluating future regulations of social media platforms.

I. "ARTIFICIAL INTELLIGENCE IS THE PRODUCT": THE USE OF AI IN SOCIAL MEDIA

Social media platforms use AI for an increasingly wide variety of tasks. Before outlining the most relevant ones, it is helpful to define what is meant by the term "AI." Unless otherwise specified—such as when contrasting "AI" with "traditional algorithms"—this Note will use the terms "AI" and "algorithm" interchangeably. While not all algorithms use AI, the term "algorithm" is used here to refer to the subset of algorithms that use AI techniques.

A. What is Artificial Intelligence?

The term "AI" is used in this Note to refer to automated decision-making processes where the substance of the decision-making is not in any meaningful sense determined in advance by the programmers.[13] Specifically, this Note focuses on the current state of machine learning ("ML") techniques generally used by social media platforms. AI is not susceptible to simple definition or line drawing.[14] It is not defined by a particular

13. *See* Lawrence Lessig, *The First Amendment Does Not Protect Replicants* 4 (Harvard Pub. L. Working Paper No. 21-34, 2021).

14. *See* Darrell M. West & John R. Allen, *How Artificial Intelligence Is Transforming the World*, BROOKINGS INST. (Apr. 24, 2018), https://www.brookings.edu/research/how-artificial-intelligence-is-transforming-the-world [https://perma.cc/W6NM-YKZJ].

technology or technique, and many commentators default to "the ability of machines to perform tasks that normally require human intelligence."[15] Early attempts at getting machines to behave "intelligently" often involved programming extensive rules to tell the machine how to respond to each situation it encountered.[16] While this could provide powerful results in certain limited situations, it was brittle, cumbersome, and did not work for many tasks requiring human-level intelligence, such as image recognition or language translation.[17]

Machine learning is the technique underlying the rapid growth of modern AI.[18] ML "gives computers the ability to learn without explicitly being programmed."[19] It represents "a conceptual shift [where] we went from attempting to encode human-distilled insights into machines to delegating the learning process itself to the machines."[20] Using a wide variety of techniques—the terms "deep learning," "supervised learning," "unsupervised learning," "reinforcement learning," "neural networks," "natural language processing," and many others all fall under the umbrella of ML—a computer "trains" on large amounts of data in order to "learn" how to act in future situations.[21] This ability allows the machine to "gain insight or automate decision-making in cases where humans would not be able to."[22]

As this Note is centered around the use of AI on social media platforms, it is not based on an abstract notion of AI, but rather on the current state of the technology as demonstrated through

---

15. *Artificial Intelligence*, AIR FORCE RSCH. LAB'Y, https://afresearchlab.com/technology/artificial-intelligence [https://perma.cc/3CQT-LAT6]; *see also* Sara Brown, *Machine Learning, Explained*, MIT SLOAN SCH. OF MGMT. (Apr. 21, 2021), https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained [https://perma.cc/RT3S-WVHW].

16. *See* Brown, *supra* note 15.

17. Gideon Lewis-Kraus, *The Great A.I. Awakening*, N.Y. TIMES MAG. (Dec. 14, 2016), https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html [https://perma.cc/7GVS-4AAN]; *see also* Danny Sullivan, *FAQ: All About the Google RankBrain Algorithm*, SEARCH ENGINE LAND (June 23, 2016), https://searchengineland.com/faq-all-about-the-new-google-rankbrain-algorithm-234440 [https://perma.cc/J8BZ-B5TA] (detailing Google's shift from human-coded knowledge to AI in its search algorithm).

18. Brown, *supra* note 15.

19. *Id.* (quoting "AI pioneer" Arthur Samuel).

20. HENRY A. KISSINGER ET AL., THE AGE OF AI: AND OUR HUMAN FUTURE 61 (2021).

21. Brown, *supra* note 15.

22. *Id.*

research and commercial use. Due to the difficulty of stating with certainty which models are used by each platform, it does not focus on a single technique or model.[23] Thus, this Note uses the term "AI" to refer to the range of ML techniques in use today.[24]

B. How AI Provides the Foundation of Modern Social Media

AI is the foundation of modern social media and many of today's largest internet platforms.[25] It is important to appreciate the extent of the transformation from past iterations of social media platforms and other internet services, such as Google search, to the AI-focused platforms we see today. Facebook[26] existed as a dominant social media platform long before it began using AI.[27] Yet by 2018, its Chief AI Scientist, Yann LeCun, declared, "[i]f you take the deep learning out of Facebook today,

---

23. Even if a model is identified as being in use at a particular moment, it can always be replaced at any time. Similarly, even if a model is not currently used, there is no guarantee it was not used in the past or will not be in the future. *See, e.g.*, *Our Approach to Explaining Ranking*, META: TRANSPARENCY CTR. (June 29, 2023), https://transparency.fb.com/features/explaining-ranking [https://perma.cc/MQ3U-YL8F] ("Prediction models, the predictions they make, and their input signals are dynamic. They change frequently as the system learns and improves over time and as Meta's products are modified.").

24. This Note does not claim that every characteristic discussed will be applicable to every implementation of AI by social media platforms. Rather, it aims to illustrate the implications of using current AI technologies for the role of content curation based on the general approaches currently in use today.

25. *See, e.g.*, *First Quarter 2023 Results Conference Call*, META (Apr. 26, 2023), https://s21.q4cdn.com/399680738/files/doc_financials/2023/q1/META-Q1-2023-Earnings-Call-Transcript.pdf [https://perma.cc/QHH5-FEMG] (describing how Facebook's "massive [AI] recommendations and ranking infrastructure . . . powers all of [its] main products"); *Research Areas: Machine Learning*, SPOTIFY, https://research.atspotify.com/machine-learning [https://perma.cc/F6C6-Z4HB] ("Machine learning touches every aspect of Spotify's business.").

26. In 2021, Facebook changed its name to Meta, but both names will be used to refer to the company in this Note. *See Introducing Meta: A Social Technology Company*, META (Oct. 28, 2021), https://about.fb.com/news/2021 /10/facebook-company-is-now-meta [https://perma.cc/F9NX-UQJL].

27. Facebook began using ML in its News Feed in 2011, when it was already the second most-visited website in the world, behind Google. Victor Luckerson, *Here's How Facebook's News Feed Actually Works*, TIME (July 9, 2015), https://time.com/collection-post/3950525/facebook-news-feed-algorithm [https://perma.cc/67YE-8HED]; JVG, *Google and Facebook Reign As the Most-Visited Sites of 2011*, VENTUREBEAT (Dec. 28, 2011), https://venturebeat.com/social/google-top-web-brand [https://perma.cc/DE3M-SZ9Y].

Facebook's dust. It's entirely built around it now."[28] Twitter[29] transitioned its timelines from reverse chronological ordering to AI-determined rankings in 2016 and immediately saw increased engagement.[30] It now uses AI across its many content presentation areas.[31] Similarly, over 70 percent of the watch time on YouTube now comes from AI recommendations,[32] and company insiders have said that "the algorithm is the single most important engine of YouTube's growth."[33]

YouTube's transition to an AI foundation was part of a broader shift throughout its parent company, Google.[34] Because of its success in internet search using traditional algorithms with hard-coded human expertise, there was internal skepticism that machine learning would provide any benefits and discomfort with entrusting the company's core operations to AI.[35] But in the end, the significant advances achieved by AI won

---

28. Metz, *supra* note 6.

29. In 2023, Twitter changed its name to X, but Twitter will still be used in this Note. *Bye-Bye Birdie: Twitter Jettisons Bird Logo, Replaces It With "X"*, CBS NEWS: MONEYWATCH (July 24, 2023), https://www.cbsnews.com/news/twitter-bird-logo-replacement-x-elon-musk **[**https://perma.cc/R6N6-CCEA]**.**

30. *See Never Miss Important Tweets From People You Follow*, TWITTER (Feb. 10, 2016), https://blog.twitter.com/official/en_us/a/2016/never-miss-important-tweets-from-people-you-follow.html [https://perma.cc/9RNS-WWKV]; Nicolas Koumchatzky & Anton Andryeyev, *Using Deep Learning at Scale in Twitter's Timelines*, TWITTER ENG'G (May 9, 2017), https://blog.twitter.com/engineering/en_us/topics/insights/ 2017/using-deep-learning-at-scale-in-twitters-timelines [https://perma.cc/R7KQ-D85S] ("[O]nline experiments have also shown significant increases in metrics such as Tweet engagement, and time spent on the platform [from deep learning].").

31. *See* Motley Fool Transcribing, *Twitter (TWTR) Q4 2018 Earnings Conference Call Transcript*, MOTLEY FOOL (Apr. 16, 2019), https://www.fool.com/earnings/call-transcripts/2019/02/07/twitter-twtr-q4-2018-earnings-conference-call-tran.aspx **[**https://perma.cc/W32C-7RWU]**.**

32. Joan E. Solsman, *YouTube's AI is the Puppet Master Over Most of What You Watch*, CNET (Jan. 10, 2018), https://www.cnet.com/tech/services-and-software/youtube-ces-2018-neal-mohan [https://perma.cc/6NDY-FV3Q].

33. Paul Lewis, *'Fiction Is Outperforming Reality': How YouTube's Algorithm Distorts Truth*, THE GUARDIAN (Feb. 2, 2018), https://www.theguardian.com/technology/2018/feb/02/how-youtubes-algorithm-distorts-truth [https://perma.cc/6S84-QXCT].

34. *See* Blaise Zerega, *AI Weekly: Google Shifts From Mobile-First to AI-First World*, VENTUREBEAT (May 18, 2017), https://venturebeat.com/ai/ai-weekly-google-shifts-from-mobile-first-to-ai-first-world [https://perma.cc/UG4Y-CRAP].

35. Steven Levy, *How Google is Remaking Itself As A "Machine-Learning First" Company*, WIRED (June 22, 2016), https://www.wired.com/2016/06/how-google-is-remaking-itself-as-a-machine-learning-first-company [https://perma.cc/8F6M-CQHL]; *see* Cade Metz, *AI Is Transforming Google Search. The Rest of the Web Is Next.*, WIRED (Feb. 4, 2016), https://www.wired.com/2016/02/ai-is-changing-the-technology-behind-google-searches [https://perma.cc/U8C7-F5FV].

the day, and it has become increasingly integrated into all of Google's businesses since it was first introduced in search operations in 2015.[36]

Another example is the video-sharing app TikTok, which Connie Chan of the venture capital firm Andreessen Horowitz described in 2018 as "the first mainstream consumer app where artificial intelligence IS the product."[37] Indeed, TikTok's parent company, ByteDance, sees itself this way.[38] In 2018, the "About" page on its website described how AI was the foundation of the company: "[ByteDance founder] Yiming [Zhang] saw an opportunity to combine the power of artificial intelligence with the growth of mobile internet to revolutionize the way people consume and receive information."[39] A representative told *The Verge* that "[a]rtificial intelligence powers all of Bytedance's content platforms."[40] Thus, while the transition to AI-based platforms has not always been readily apparent to the end user, it is important to understand the extent to which the major platforms are not the same companies they were a decade ago. They now all rely on AI as the core of their business.

In particular, "[e]very major platform now relies on some version of deep learning to choose what content to display."[41] While this includes many components of a social media platform's operations, this Note will focus on the use of AI for content recommendation and presentation—such as timelines,

---

36. KISSINGER ET AL., *supra* note 20, at 101 ("In 2015, Google's search team moved from using these human-developed algorithms to implementing machine learning. This change led to a watershed moment: incorporating AI has vastly improved the quality and usability of the search engine, making it better able to anticipate questions and organize accurate results."); *see also* Jack Clark, *Google Turning Its Lucrative Web Search Over to AI Machines*, BLOOMBERG (Oct. 26, 2015), https://www.bloomberg.com/news/articles/2015-10-26/google-turning-its-lucrative-web-search-over-to-ai-machines [https://perma.cc/WX5T-YHW7]; Prabhakar Raghavan, *How AI Is Powering A More Helpful Google*, GOOGLE: THE KEYWORD (Oct. 15, 2020), https://blog.google/products/search/search-on [https://perma.cc/9KVR-U2YU] ("With recent advancements in AI, we're making bigger leaps forward in improvements to Google than we've seen over the last decade . . . .").

37. Connie Chan, *When AI is the Product*, ANDREESSEN HOROWITZ (Dec. 3, 2018), https://a16z.com/2018/12/03/ when-ai-is-the-product-the-rise-of-ai-based-consumer-apps [https://perma.cc/5TU3-49V8].

38. *See* ByteDance, *About*, INTERNET ARCHIVE WAYBACK MACHINE (Oct. 31, 2018), https://web.archive.org/web/20181031204318/http://bytedance.com#about.

39. *Id.*

40. Sam Byford, *How China's Bytedance Became the World's Most Valuable Startup*, VERGE (Nov. 30, 2018), https://www.theverge.com/2018/11/30/18107732/bytedance-valuation-tiktok-china-startup [https://perma.cc/ZFN2-WVQ8].

41. Meserole, *supra* note 6.

news feeds, or connection recommendations[42]—because these systems carry the greatest potential for manipulation and, thus, significant implications for First Amendment jurisprudence.[43]

It is difficult to know exactly how the platforms operate internally, but public statements from the companies provide a basic understanding. Content recommendation systems are generally built on a foundation of AI models that "decide how relevant a particular piece of content is to a user," with layers of manual business logic added in to achieve the platform's desired balance of content.[44] For example, Facebook explains that its content recommendation system uses thousands of signals—pieces of data about a user or a piece of content, such as the user's location and device information, the type of posts the user has liked in the past, whether the post contains a URL, etc.—and inputs these into hundreds of ML models that predict various things about the user and the post, such as how likely the user is to share it or how much time the user is going to spend engaging with it.[45] These hundreds of predictions are then fed into another ML model that calculates an overall score of how "relevant" the content is to the user and ranks posts based on this value.[46] Along the way, Facebook will insert manual business logic to filter (e.g., remove posts it predicts are likely to be offensive) and tune (e.g., to ensure there is a mix of different content types) the models to achieve what it thinks is the best balance of content on the platform.[47]

The evidence indicates that each of the major platforms uses an approach similar to the one described here, with a foundation of ML models combined with manual business logic.[48] Thus, this

---

42. *See, e.g.*, *Our Approach to Explaining Ranking*, *supra* note 23.

43. *See* Carina Prunkl, *Human Autonomy in the Age of Artificial Intelligence*, 4 NATURE MACH. INTEL. 99 (2022).

44. Kristian Lum & Tomo Lazovich, The Myth of The Algorithm: A System-Level View of Algorithmic Amplification, KNIGHT FIRST AMEND. INST. (Sept. 13, 2023), https://knightcolumbia.org/content/the-myth-of-the-algorithm-a-system-level-view-of-algorithmic-amplification [https://perma.cc/33MR-9Q4U].

45. *Our Approach to Explaining Ranking*, *supra* note 23.

46. *Id.*

47. *Id.*; *see also* Sandeep Grover & Mabel Wang, *Introducing a Way to Refresh Your For You Feed on TikTok*, TIKTOK (Mar. 16, 2023), https://newsroom.tiktok.com/en-us/introducing-a-way-to-refresh-your-for-you-feed-on-tiktok-us [https://perma.cc/6FEX-BGFH].

48. *Instagram Feed AI System*, META: TRANSPARENCY CTR. (June 29, 2023), https://transparency.fb.com/features/explaining-ranking/ig-feed [https://perma.cc/ATT6-EZF9]; *How TikTok Recommends Videos #ForYou*, TIKTOK (June 18, 2020), https://newsroom.tiktok.com/en-us/how-tiktok-recommends-videos-for-you [https://

Note will focus on the First Amendment implications of using current ML technologies for editorial decision-making.

II. EDITORIAL RIGHTS UNDER THE FIRST AMENDMENT

The scope of editorial rights protected under the First Amendment is broad, extending beyond plain speech to include editorial discretion and control. For example, a newspaper is not always publishing content written by its own employees, but its editors are still protected in their choices about what content to publish and how they choose to present that content.[49] Importantly, editorial rights are also enjoyed by publishers beyond the traditional press under the protection of expressive conduct.[50]

Nonetheless, the editorial right is best illustrated by the Court's treatment of the classic editor: a traditional print newspaper. In *Miami Herald Publishing Company v. Tornillo*, the Supreme Court considered the constitutionality of a Florida "right of reply" statute, which provided that "if a candidate for nomination or election is assailed regarding his personal character or official record by any newspaper, the candidate has the right to demand that the newspaper print, free of cost to the candidate, any reply the candidate may make to the newspaper's charges."[51] Before the case reached the Supreme Court, Florida's highest court held that the law was constitutional because it furthered the "broad societal interest in the free flow

---

perma.cc/374T-QE4N]; *Twitter's Recommendation Algorithm*, TWITTER (March 31, 2023), https://blog.twitter.com/engineering/en_us/topics/open-source/2023/twitter-recommendation-algorithm [https://perma.cc/8PR9-P2U2]; Cristos Goodrow, *On YouTube's Recommendation System*, YOUTUBE: OFFICIAL BLOG (Sept. 15, 2021), https://blog.youtube/inside-youtube/on-youtubes-recommendation-system [https://perma.cc/8P47-BKE9].

49. Miami Herald Pub. Co. v. Tornillo, 418 U.S. 241, 256, 258 (1974).

50. *See* Hurley v. Irish-Am. Gay, Lesbian & Bisexual Grp. of Bos., 515 U.S. 557, 574 (1995); NetChoice, LLC v. Att'y Gen., Fla., 34 F.4th 1196, 1210 (11th Cir. 2022) ("Laws that restrict platforms' ability to speak through content moderation therefore trigger First Amendment scrutiny. Two lines of precedent independently confirm this commonsense conclusion: first, and most obviously, decisions protecting exercises of 'editorial judgment'; and second, and separately, those protecting inherently expressive conduct."), cert. *granted in part sub nom.* Moody v. Netchoice, LLC, No. 22-277, 2023 WL 6319654 (U.S. Sept. 29, 2023), and *cert. denied sub nom.* Netchoice v. Moody, No. 22-393, 2023 WL 6377782 (U.S. Oct. 2, 2023).

51. *Tornillo*, 418 U.S. at 244.

of information to the public."[52] Similarly, proponents of the law argued that, in contrast to the "relatively easy access to the channels of communication" at the time of the country's founding, the modern press was in the hands of relatively few noncompetitive, concentrated interests with the power to shape and manipulate public opinion: the "marketplace of ideas" had become "a monopoly controlled by the owners of the market."[53]

However, the Supreme Court disagreed, concluding that concerns over the new media landscape were not sufficient to override the protections of a free press. While it was argued that monopoly control of the press did not promote our "profound national commitment to the principle that debate on public issues should be uninhibited, robust, and wide-open,"[54] the Court found that, on the contrary, the right-of-access statute would discourage editors from expressing their views in order to avoid controversy, thus inhibiting and dampening public debate.[55] It held that "any such compulsion to publish that which 'reason tells them should not be published' is unconstitutional. A responsible press is an undoubtedly desirable goal, but press responsibility is not mandated by the Constitution and like many other virtues it cannot be legislated."[56] The case resulted in the Court's clearest expression of the scope of editorial rights:

> [T]he Florida statute fails to clear the barriers of the First Amendment because of its intrusion into the function of editors. . . . The choice of material to go into a newspaper, and the decisions made as to limitations on the size and content of the paper, and treatment of public issues and public officials—whether fair or unfair—constitute the exercise of editorial control and judgment. It has yet to be demonstrated how governmental regulation of this crucial process can be

52. *Id.* at 245.
53. *Id.* at 248–51.
54. N.Y. Times Co. v. Sullivan, 376 U.S. 254, 270 (1964).
55. *Tornillo*, 418 U.S. at 257 ("[U]nder the operation of the Florida statute, political and electoral coverage would be blunted or reduced. Government-enforced right of access inescapably 'dampens the vigor and limits the variety of public debate.'") (quoting *N.Y. Times*, 376 U.S. at 279).
56. *Id.* at 256.

> exercised consistent with First Amendment guarantees of a free press as they have evolved to this time."[57]

It thus established strong protection for the variety of activities constituting the role of an editor in choosing, preparing, and presenting material.

While *Tornillo* involved a traditional newspaper, editorial rights have been applied to a broad range of entities and circumstances, bringing social media platforms well within their ambit. In *Pacific Gas & Electric Co. v. Public Utilities Commission of California*, the Court considered a challenge to a California Public Utilities Commission requirement that the public utility Pacific Gas & Electric (PG&E) include third-party content in the newsletter that it distributed with its monthly bills.[58] The Court held that the reasoning of *Tornillo* applied to the utility newsletter just like it did to "the institutional press."[59] It concluded that the Commission's requirements impermissibly interfered with PG&E's editorial discretion by forcing it to carry and associate with speech with which it disagreed.[60] Similarly, in *Hurley v. Irish-American Gay, Lesbian and Bisexual Group of Boston*, the Court applied First Amendment editorial rights to a parade organizer, writing: "Nor is the [benefit of editorial discretion] restricted to the press, being enjoyed by business corporations generally and by ordinary people engaged in unsophisticated expression as well as by professional publishers."[61] The Court described the right as the "autonomy to control one's own speech," and found that state interference with the parade organizer's editorial choices invaded that autonomy.[62]

From these cases, it is apparent that the scope of editorial rights extends well beyond traditional publishers and includes broad protection of discretion and control over one's expression. With modern social media, the platforms exercise significant discretion in choosing what content is presented to each user, as billions of pieces of content are distilled down to a curated

---

57. *Id.* at 258.
58. Pac. Gas & Elec. Co. v. Pub. Utilities Comm'n of California, 475 U.S. 1, 4–7 (1986).
59. *Id.* at 11.
60. *Id.* at 13–15.
61. Hurley v. Irish-Am. Gay, Lesbian & Bisexual Grp. of Bos., 515 U.S. 557, 574 (1995).
62. *Id.*

selection of relevant material.[63] Thus, there is a strong argument that the use of AI for editorial decision-making on social media platforms implicates First Amendment concerns.

III. HOW ARE EDITORIAL RIGHTS APPLIED TO NEW TECHNOLOGIES?

While the above cases clearly demonstrate strong protection for the editorial rights of a variety of speakers, the strength of this protection is not applied uniformly in all situations.[64] The Supreme Court has established that although "differences in the characteristics of new media justify differences in the First Amendment standards applied to them,"[65] "the basic principles of freedom of speech and the press . . . do not vary when a new and different medium for communication appears."[66] Therefore, application to a new medium of communication, type of speaker, or form of expression must first identify the relevant foundational First Amendment values and then employ a purposive approach to ensure those principles are protected in whatever manner is appropriate to the unique characteristics of the situation.[67] The following sections will outline three commonly articulated First Amendment values and explore how the use of AI in editorial decision-making impacts these values.

---

63. *See, e.g.*, *The AI Behind Unconnected Content Recommendations on Facebook and Instagram*, META AI (June 29, 2023), https://ai.facebook.com/blog/ai-unconnected-content-recommendations-facebook-instagram [https://perma.cc/2RUX-YY3U].

64. *See* Se. Promotions, Ltd. v. Conrad, 420 U.S. 546, 557 (1975) ("Each medium of expression, of course, must be assessed for First Amendment purposes by standards suited to it, for each may present its own problems."); *see also* Red Lion Broad. Co. v. F.C.C., 395 U.S. 367, 387 (1969) ("[T]he ability of new technology to produce sounds more raucous than those of the human voice justifies restrictions on the sound level, and on the hours and places of use, of sound trucks so long as the restrictions are reasonable and applied without discrimination.").

65. *Red Lion*, 395 U.S. at 386.

66. Brown v. Ent. Merchants Ass'n, 564 U.S. 786, 790 (2011) (quoting Joseph Burstyn, Inc. v. Wilson, 343 U.S. 495, 503 (1952)) (internal quotation marks omitted).

67. *See, e.g.*, David S. Han, *Constitutional Rights and Technological Change*, 54 U.C. DAVIS L. REV. 71, 100 (2020) ("Regardless of the result, the novelty presented by search engine results forces courts to undertake some sort of first-principles boundary analysis . . . .").

A. The Values Protected by First Amendment Editorial Rights

There are numerous strands of First Amendment jurisprudence, often with slightly different articulations of the underlying values being protected.[68] This Note will identify the common First Amendment values articulated in editorial rights cases and examine how those values are applied by courts when considering expression involving new technologies.

Three theories of the values underlying First Amendment jurisprudence can inform the application of editorial rights to the use of AI on social media: democracy and self-governance, the "marketplace of ideas," and autonomy.[69] In their various applications, these values serve to protect the interests of speakers, listeners, and the public at-large.[70]

First is the view that freedom of speech is a value "necessary for the effective operation of the democratic process."[71] Under this theory, free speech is protected to facilitate uninhibited debate such that the outcomes of the democratic process can be traced to the thoughts, beliefs, and reasoning of its participants.[72] In *New York Times v. Sullivan*, the Court explained that the First Amendment "was fashioned to assure unfettered interchange of ideas for the bringing about of political and social changes desired by the people."[73] It therefore considered the case "against the background of a profound national commitment to the principle that debate on public issues should be uninhibited, robust, and wide-open."[74] Based on this principle, the Court concluded that an outcome in the case that led to self-censorship would "dampen the vigor and limit the variety of public debate," which it saw as "inconsistent

---

68. *See* Toni M. Massaro & Helen Norton, *Siri-ously? Free Speech Rights and Artificial Intelligence*, 110 NW. U. L. REV. 1169, 1175 (2016); Robert C. Post, *Racist Speech, Democracy, and the First Amendment*, 32 WM. & MARY L. REV. 267, 278 (1991).

69. Massaro & Norton, *supra* note 68, at 1175.

70. The term "listener" is used in this Note—as it often is in First Amendment jurisprudence—to generically refer to listeners, readers, viewers, and any other receiver of speech or expression.

71. CBS, Inc. v. F.C.C., 453 U.S. 367, 396 (1981); *accord* Lyrissa Barnett Lidsky, *Nobody's Fools: The Rational Audience As First Amendment Ideal*, 2010 U. ILL. L. REV. 799, 839 (2010) ("It is generally agreed that a core purpose of the First Amendment is to foster the ideal of democratic self-governance.").

72. *See* Massaro & Norton, *supra* note 68, at 1176–78.

73. 376 U.S. 254, 269 (1964) (quoting Roth v. U.S., 354 U.S. 476, 484 (1957)).

74. *Id.* at 270.

with the First [Amendment]."[75] A legitimate government *of* the people must reflect the free will of those people, and decisions made by the people cannot be considered free if the ideas that can be spoken and heard are restricted.[76]

The second value is that of the "uninhibited marketplace of ideas."[77] Under this theory, the First Amendment protects the "widest possible dissemination of information" so that listeners will be exposed to the greatest variety of ideas.[78] In contrast to the value of democratic self-governance, the marketplace of ideas has as its goal the discovery of truth.[79] To that end, it values the dissemination of all information regardless of its importance to the democratic process. It relies on a conception of members of the public as rational, truth-seeking individuals who will use their reason to identify the best ideas such that "truth will ultimately prevail."[80]

Third is the value of individual autonomy.[81] "At the heart of the First Amendment lies the principle that each person should decide for himself or herself the ideas and beliefs deserving of expression, consideration, and adherence."[82] This can be understood as protecting both the speaker's "autonomy to control one's own speech"[83] and the autonomy of listeners to decide for themselves what to believe.

With regard to protecting the autonomy of the speaker, in *Hurley*, the Court described the parade organizer's editorial right as the "autonomy to control one's own speech."[84] It explained that "when dissemination of a view contrary to one's own is forced upon a speaker intimately connected with the

---

75. *Id.* at 279.

76. *See* Post, *supra* note 68, at 282–84; Julie E. Cohen, *Examined Lives: Informational Privacy and the Subject As Object*, 52 STAN. L. REV. 1373, 1426 (2000) ("The cornerstone of a democratic society is informed and deliberate self-governance.").

77. Red Lion Broad. Co. v. F.C.C., 395 U.S. 367, 390 (1969).

78. Associated Press v. U.S., 326 U.S. 1, 20 (1945).

79. *See Red Lion*, 395 U.S. at 390.

80. *Id.*; *see* Lidsky, *supra* note 71, at 816.

81. *See* Richard H. Fallon, Jr., *Two Senses of Autonomy*, 46 STAN. L. REV. 875, 875–76 (1994) ("A diverse collection of writers has identified autonomy as a central value underlying the First Amendment's commitment to free expression . . . Among other things, autonomy holds unique promise to function as the constitutional value of values.").

82. Turner Broad. Sys., Inc. v. F.C.C., 512 U.S. 622, 641 (1994).

83. Hurley v. Irish-Am. Gay, Lesbian & Bisexual Grp. of Bos., 515 U.S. 557, 574 (1995).

84. *Id.*

communication advanced, the speaker's right to autonomy over the message is compromised."[85] While an autonomy-based free speech right has intuitive plausibility to many commentators, it often eludes a precise definition, and the Court does not elaborate on what it means by the term "autonomy" in this context.[86] Borrowing from Professor Richard H. Fallon Jr.'s discussion of this dynamic, one can think of autonomy descriptively: "To be autonomous, one must be able to form a conception of the good, deliberate rationally, and act consistently with one's goals. Beyond the capacities necessary for self-government, descriptive autonomy requires freedom from coercion, manipulation, and temporary distortion of judgment."[87] Inherent in this definition is the fact that autonomy is a matter of degree: the amount of unacceptable influence, distortions in judgment, and actual rationality exercised by an individual exist on a spectrum, and there are few, if any, bright-line rules as to when a belief or act is fully autonomous.[88] A minimum definition can be thought of as the ability to express or act upon one's wishes or goals, whether or not those wishes themselves were formed without influence.[89]

Perhaps the most useful articulation of autonomy from the Court comes from *Cohen v. California*, where the Court described "[t]he constitutional right of free expression . . . [as] putting the decision as to what views shall be voiced largely into the hands of each of us . . . in the belief that no other approach would comport with the premise of individual dignity and choice upon which our political system rests."[90] In this context, autonomy over one's message is based on a conception of the speaker in control of their message, with the ability to make their own decisions and act consistent with their goals.

Turning to listener autonomy, Professor David A. Strauss argues that most First Amendment theories can be explained by what he calls the "persuasion principle," which states that "the government may not suppress speech on the ground that the speech is likely to persuade people to do something that the

85. *Id.* at 576.
86. *See* Fallon, *supra* note 81, at 875–76.
87. *Id.* at 877.
88. *See id.*; *see also* Margaret A. Somerville, *Labels Versus Contents: Variance Between Philosophy, Psychiatry and Law in Concepts Governing Decision-Making*, 39 MCGILL L.J. 179, 187 (1994).
89. *See* Somerville, *supra* note 88, at 194.
90. Cohen v. California, 403 U.S. 15, 24 (1971).

government considers harmful."[91] He argues that this principle can be justified by a theory of human autonomy, under which the First Amendment is "designed to protect the autonomy of potential listeners."[92] Through this lens, the fundamental concern of the First Amendment is preserving the "dignity and choice" of the individual.[93]

B. Government Regulation of Private Action

Two other observations regarding the application of First Amendment values are important to note. First, though the First Amendment only applies to government action, the Court's focus on the interests of listeners means that the effects of private actions on free speech values are relevant to an analysis of government regulation of those actions.[94] Second, each of the values outlined here implicitly relies on a view of the public as being made up of rational individuals using reason to freely choose which ideas to believe and accept.[95]

Editorial rights are not only, or even primarily, about an individual's autonomy to speak but rather about preserving the valuable societal function of the press for the benefit of all.[96] In the context of the media, the Court has said, "the people as a whole retain their interest in free speech by radio and their collective right to have the medium function consistently with the ends and purposes of the First Amendment. It is the right of the viewers and listeners, not the right of the broadcasters, which is paramount."[97] Thus, a First Amendment analysis of

---

91. David A. Strauss, *Persuasion, Autonomy, and Freedom of Expression*, 91 COLUM. L. REV. 334, 335 (1991).

92. *Id.* at 371.

93. *See Cohen*, 403 U.S. at 24.

94. *See, e.g.*, Associated Press v. United States, 326 U.S. 1, 20 (1945) ("Freedom of the press from governmental interference under the First Amendment does not sanction repression of that freedom by private interests.").

95. *See* Lidsky, *supra* note 71.

96. *See* Gerald G. Ashdown, *Editorial Privilege and Freedom of the Press:* Herbert v. Lando *in Perspective*, 51 U. COLO. L. REV. 303, 303 n.3 (1980). The Court has never been entirely clear on the boundaries between the right to expressive conduct as "speech" and editorial rights under freedom of the press.

97. Red Lion Broad. Co. v. F.C.C., 395 U.S. 367, 390 (1969); *see also* CBS, Inc. v. F.C.C., 453 U.S. 367, 370 (1981) (finding the challenged statute made "a significant contribution to freedom of expression by enhancing the ability of candidates to present, and the public to receive, information necessary for the effective operation of the democratic process"); Pac. Gas & Elec. Co. v. Pub. Utils. Comm'n of Cal., 475 U.S. 1, 8 (1986) (citations omitted) (expressly relying on

editorial rights must factor in the broader impacts of a given regulation on society.

Indeed, this focus on non-speaker interests means that courts will occasionally approve government restrictions on the editorial function of private entities to preserve the First Amendment values of listeners and society. In both *Associated Press* and *Red Lion*, the Court emphasized that the government did not have to sit idly by while private actors used their freedom of speech to thwart the underlying purposes of the First Amendment, writing, "[t]here is no sanctuary in the First Amendment for unlimited private censorship operating in a medium not open to all."[98] The Court observed, "[i]t would be strange indeed . . . if the grave concern for freedom of the press" meant that "the government was without power to protect that freedom."[99]

C. Rationality and Reason in First Amendment Theory

Finally, there is an underlying conception of "the power of reason" pervading each of these First Amendment values.[100] It adopts a view of humans as rational beings exercising their reason and judgment as both speakers and listeners. For speakers, the *Tornillo* court said, "any such compulsion to publish that which 'reason tells them should not be published' is unconstitutional."[101] This envisions a speaker in control of their message, exercising rational thought without interference. Indeed, the court referred to editorial discretion in terms of "editorial control and judgment."[102] As explained above, *Hurley's* vision of the First Amendment protecting the "autonomy to control one's own speech" similarly rests on a conception of the speaker in control of their expression, capable of exercising reason and judgment to "deliberate rationally, and act consistently with one's goals."[103] Thus, the First

---

"significant societal interests wholly apart from the speaker's interest in self-expression . . . . [In particular,] the public's interest in receiving information.").

98. *Red Lion*, 395 U.S. at 392.

99. Associated Press v. United States, 326 U.S. 1, 20 (1945).

100. *See generally* Lidsky, *supra* note 71.

101. Mia. Herald Pub. Co. v. Tornillo, 418 U.S. 241, 256 (1974) (quoting *Associated Press*, 326 U.S. at 20 n.18).

102. *E.g., id.* at 258; *see also* Pittsburgh Press Co. v. Pittsburgh Comm'n on Hum. Rels., 413 U.S. 376, 391 (1973).

103. Fallon, *supra* note 81, at 877.

Amendment's concern with protecting the publisher's autonomy is designed to protect the exercise of reason, judgment, and control over one's expression.

First Amendment jurisprudence also envisions a rational listener. The connection between the dissemination of information and democratic self-governance is premised on individuals exercising reason and judgment to evaluate ideas and choose those deserving acceptance.[104] As constitutional law Professor Lyrissa Lidsky explains:

> The assumption that citizens are rational is deeply embedded in democratic theory . . . . The ideal of democratic self-governance, however, makes no sense unless one assumes that citizens will generally make rational choices to govern the fate of the nation. If a majority of citizens make policy choices based on lies, half-truths, or propaganda, sovereignty lies not with the people but with the purveyors of disinformation. If this is the case, democracy is both impossible and undesirable.[105]

In *New York Times v. Sullivan*, the Court credited Justice William Brandeis with the "classic formulation" of the principle of uninhibited public debate in his concurrence from *Whitney v. California*.[106] Justice Brandeis wrote, "[t]hose who won our independence believed that the final end of the state was to make men free to develop their faculties, and that in its government the deliberative forces should prevail over the arbitrary."[107] This principle was grounded in the Founders' belief "in the power of reason as applied through public discussion . . . ."[108] Justice Brandeis saw the exercise of reason and "deliberative forces" as the foundation of democratic self-determination.[109] Similarly, the concept of "an uninhibited marketplace of ideas in which truth will ultimately prevail"[110] is premised on the capacity of individuals to exercise rational thought to evaluate the wide dissemination of ideas and discern

---

104. *See* Lidsky, *supra* note 71, at 811.
105. *Id.* at 838–39.
106. New York Times Co. v. Sullivan, 376 U.S. 254, 270 (1964).
107. Whitney v. California, 274 U.S. 357, 375 (1927) (Brandeis, J., concurring), *overruled by* Brandenburg v. Ohio, 395 U.S. 444 (1969).
108. *Id.*
109. *Id.*
110. Red Lion Broad. Co. v. F.C.C., 395 U.S. 367, 390 (1969).

the truth. The value of uninhibited public discussion in service of both self-governance and the individual search for truth lies in the ability of individuals to use their reason and judgment and to act free from coercion and manipulation.[111]

In outlining his "persuasion principle," discussed above, Professor Strauss explains the contours of the First Amendment in terms of preserving the autonomy of the individual to make rational decisions without manipulation, writing: "[T]he autonomous individual is an unmanipulated individual."[112] He explains how the Court's treatment of *false statements* and *fighting words* (defined further below) can be understood to consider such manipulation, which necessarily precludes a rational response, to be beyond the protection of the First Amendment.[113]

In *Chaplinsky v. State of New Hampshire*, the Court described fighting words as words "likely to cause an average addressee to fight" or having the "characteristic of plainly tending to excite the addressee to a breach of the peace."[114] While the Court framed the opinion in terms of preventing any breach of the peace, the holding necessarily relies on a non-autonomous theory of causation where the listener's behavior is largely dictated by the speaker.[115] The First Amendment does not protect such words because they invade the autonomy of the listener by, in the words of Professor Fallon, inducing a "temporary distortion of judgment."[116]

Similarly, in *Gertz v. Robert Welch, Inc.*, the Court said, "there is no constitutional value in false statements of fact. Neither the intentional lie nor the careless error materially advances society's interest in 'uninhibited, robust, and wide-open' debate on public issues."[117] False statements interfere with the autonomy of the listener by manipulating them to act in a desired manner that is contrary to reality, thereby precluding rational self-determination.[118] *Gertz* reveals that the

---

111. *See* Lidsky, *supra* note 71, at 815–16.
112. Strauss, *supra* note 91, at 371.
113. *See id.* at 339, 343.
114. 315 U.S. 568, 573 (1942).
115. As Professor Fallon explains, "autonomy requires freedom from coercion, manipulation, and temporary distortion of judgment." Fallon, *supra* note 81, at 877.
116. *Id.*
117. Gertz v. Robert Welch, Inc., 418 U.S. 323, 340 (1974) (quoting New York Times Co. v. Sullivan, 376 U.S. 254, 270 (1964)).
118. *See* Strauss, *supra* note 91, at 366.

value in "uninhibited" debate is not, in fact, in it being uninhibited. Rather, preserving this aspect of public debate allows individuals to participate autonomously, using their reason and judgment to advance themselves and the interests of a democratic society.

The above cases illustrate how editorial discretion is protected for the purpose of serving the underlying First Amendment values of democratic self-governance, the marketplace of ideas, and autonomy. As Professors Evelyn Douek and Genevieve Laker explain, "what the editorial discretion cases show is that the First Amendment is not concerned solely—or perhaps even primarily—with the maximization of speech per se. Instead, what it protects and facilitates is the kind of information ecosystem in which free speech values can flourish."[119] Thus, any analysis of a new editorial rights case, especially one involving new technology, requires a holistic consideration of how the various interests involved impact those values.[120]

D. Application of the First Amendment to New Technologies

The Supreme Court has historically analyzed cases involving new expressive technologies—such as social media and AI—by assessing the impact of the technology and its regulation on the above values.[121] The clearest illustration of the Court considering the unique characteristics of a new technology to determine First Amendment protections based on underlying values comes from *Red Lion Broadcasting Company v. Federal Communications Commission*. In that case, the Court considered the constitutionality of the FCC's "fairness doctrine" regulations, which required radio and television broadcasters to

---

119. Douek & Laker, *supra* note 4.

120. *See id.* ("[W]hat the editorial discretion cases implicitly illustrate is that this analysis must be purpose-driven.").

121. *See, e.g.*, FCC v. Pacifica Found., 438 U.S. 726, 748–50 (1978) (finding FCC censorship of a radio broadcast constitutional based on the special problems presented by the broadcast medium); Denver Area Educ. Telecomms. Consortium, Inc. v. F.C.C., 518 U.S. 727, 740 (1996) (rejecting a "categorial approach" to First Amendment jurisprudence that would "import law developed in very different contexts into a new and changing environment, and . . . lack the flexibility necessary to allow government to respond to very serious practical problems without sacrificing the free exchange of ideas the First Amendment is designed to protect").

give "adequate coverage to public issues," and that "each side of those issues must be given fair coverage."[122] In particular, the regulations required a broadcaster who made an attack "upon the honesty, character, integrity or like personal qualities of an identified person or group" to give them a "reasonable opportunity to respond over the licensee's facilities."[123] The regulations were promulgated under the FCC's authority to issue regulations in the public interest.[124]

While at first glance the situation appears similar to the "right-of-reply" statute applied to newspapers in *Tornillo*, the Court analyzed the cases differently.[125] In holding the regulations at issue in *Red Lion* constitutional, the Court focused on the technological differences between print and broadcast media, saying, "differences in the characteristics of new media justify differences in the First Amendment standards applied to them."[126] It reasoned that access to broadcasting must be restricted or the medium becomes unusable. Because radio frequencies are limited and anyone can broadcast over them, unregulated access causes broadcasts to overlap and conflict with each other over the airwaves.[127] Recognizing this, the Court observed that "[i]t would be strange if the First Amendment, aimed at protecting and furthering communications, prevented the Government from making radio communication possible by requiring licenses to broadcast and by limiting the number of licenses so as not to overcrowd the spectrum."[128] It emphasized that the government is not prohibited from regulating a new technology just because it happens to be used for speech.[129]

IV. HOW DOES AI EDITORIAL DECISION-MAKING IMPACT FIRST AMENDMENT VALUES?

From the above discussion, we can now consider the application of First Amendment rights to the use of AI for editorial decision-making. Two federal circuit courts recently

122. Red Lion Broad. Co. v. F.C.C., 395 U.S. 367, 369, 377 (1969).
123. *Id.* at 373–74.
124. *Id.* at 379.
125. Mia. Herald Pub. Co. v. Tornillo, 418 U.S. 241, 256 (1974).
126. *Red Lion*, 395 U.S. at 386.
127. *See id.* at 375–76.
128. *Id.* at 389.
129. *Id.* at 387.

split over how the First Amendment applies to government regulation of social media. Both cases have been appealed to the Supreme Court, and this Note does not attempt to settle the debate. Nonetheless, an examination of the use of AI by social media platforms can be helpful in understanding the divergent analyses by the two courts.

The Eleventh Circuit found the editorial decision-making by social media platforms to be "closely analogous" to that of the newspaper and cable company in *Tornillo* and *Turner I*, respectively, as each user sees a "curated and edited compilation of content."[130] The court held that Florida's social media statute restricted the platforms' exercise of editorial judgment and thus triggered First Amendment scrutiny.[131] In contrast, when considering Texas's regulation, the Fifth Circuit described the platforms as "nothing like the newspaper in *Miami Herald*," seeing them instead as conduits that "exercise virtually no editorial control or judgment."[132] Based on the findings by the United States District Court for the Northern District of Florida that "the overwhelming majority of the material never gets reviewed except by algorithms" and was therefore "invisible to the [platform],"[133] the Fifth Circuit concluded that this content was "just posted to the Platform with zero editorial control or judgment."[134] The court held that Texas's prohibition of viewpoint-based content moderation was constitutional.[135]

The reality is that content on social media is indeed invisible to the platform, but there is also substantial editorial control and judgment—it is just exercised by AI rather than humans. No human makes an editorial judgment on the vast majority of content. However, contrary to the Fifth Circuit's analysis, the platforms are still far more than conduits of speech.[136] Rather,

130. NetChoice, L.L.C. v. Att'y Gen., Fla., 34 F.4th 1196, 1213, 1204 (11th Cir. 2022), *cert. granted in part sub nom.* Moody v. Netchoice, LLC, No. 22-277, 2023 WL 6319654 (U.S. Sept. 29, 2023), and *cert. denied sub nom.* Netchoice v. Moody, No. 22-393, 2023 WL 6377782 (U.S. Oct. 2, 2023).

131. *Id.* at 1210.

132. NetChoice, L.L.C. v. Paxton, 49 F.4th 439, 459 (5th Cir. 2022), *cert. granted in part sub nom.* Netchoice, LLC v. Paxton, No. 22-555, 2023 WL 6319650 (U.S. Sept. 29, 2023).

133. NetChoice, L.L.C. v. Moody, 546 F. Supp. 3d 1082, 1091–92 (N.D. Fla. 2021), *aff'd in part, vacated in part, remanded sub nom. Att'y Gen., Fla.*, 34 F.4th 1196.

134. *Paxton*, 49 F.4th at 459.

135. *Id.* at 473.

136. *Id.* at 460.

they deliver a "curated compilation[]" of content carefully chosen for each user.[137] As the Eleventh Circuit explained, "the platforms invest significant time and resources into editing and organizing . . . users' posts into collections of content that they then disseminate to others."[138] Of course, their investment is not actually in editing or organizing; it is in building and training AI to perform the task for them.

Thus, even if social media platforms do exercise some amount of editorial discretion and are entitled to the protection of the First Amendment, the question remains whether the extent of this protection is the same as that of more traditional publishers like newspapers and cable operators.[139] In particular, is the use of AI for editorial decision-making entitled to the same First Amendment protection as human decision-making?

Editorial decision-making has historically been performed by a human. Traditional mass media exercises its editorial discretion through several layers of human vetting: First, a company hires specific journalists to create its content; second, it chooses what stories to cover and how to cover them;[140] and finally, it decides what content to actually publish or not publish as well as how and where it wants to present each piece of content.[141] Even for third-party content, such as opinion pieces in a newspaper or a cable TV channel carried by a cable operator, the media company uses human discretion to determine what it wants to publish. While this discretion could be exercised in any number of ways, it most likely entails some combination of evaluating the identity of the creator of the content (profession, relevant experience, status in the community, etc.) to understand its context, relevance, and newsworthiness, as well as evaluating the content itself—its quality, the viewpoint it

---

137. *Att'y Gen., Fla.*, 34 F.4th at 1213.

138. *Id.* at 1204–05.

139. *See* Douek & Laker, *supra* note 4 ("But when, how, and why the First Amendment protects editorial discretion is the question we should be asking, not whether it does. That latter question has been well and truly answered—in the affirmative.").

140. *See, e.g.*, *Editorial Process in Action*, KQED, https://www.kqed.org/about/editorial-process-in-action [https://perma.cc/FZH5-RCKP].

141. *See, e.g.*, David Manning White, *The "Gate Keeper": A Case Study in the Selection of News*, 27 JOURNALISM Q. 383 (1950), http://www.aejmc.org/home/wp-content/uploads/2012/09/Journalism-Quarterly-1950-White-383-90.pdf [https://perma.cc/WH6B-N5TS].

espouses, and how it fits into the broader context of the publication.[142]

Social media platforms delegate much of the decision-making in content curation to AI.[143] But, thus far, courts have had little opportunity to consider whether and how this might affect a First Amendment analysis.[144] When they do, they have been dismissive of distinctions between the use of AI and traditional human processes: the district court analyzing Texas's social media regulation found that "focusing on whether a human or AI makes those decisions is a distraction."[145]

This Note argues that the opposite is true: the use of AI for editorial decisions is a fundamental transformation from human decision-making processes that must be understood and appreciated by both courts and legislators. With the values underlying First Amendment editorial protection and past jurisprudence on new technology in mind, this Note outlines two reasons why the use of AI deserves a different analysis: manipulation and lack of control.

A. Manipulation Beyond the Ability of a Human Editor

First, AI is able to manipulate listeners more than a human editor.[146] Indeed, AI is already more effective than humans at many tasks,[147] and it has consistently demonstrated the ability

142. *See, e.g.*, Remy Tumin, *The Op-Ed Pages, Explained*, N.Y. TIMES (Dec. 3, 2017), https://www.nytimes.com/2017/12/03/insider/opinion-op-ed-explainer.html [https://perma.cc/7S6P-93SW].

143. *See* Nick Clegg, *How AI Influences What You See on Facebook and Instagram*, META (June 29, 2023), https://about.fb.com/news/2023/06/how-ai-ranks-content-on-facebook-and-instagram [https://perma.cc/LRT8-YUAM].

144. *See, e.g.*, NetChoice, LLC v. Paxton, 49 F.4th 439, 459 n.8 (5th Cir. 2022) ("The Platforms have disclosed little about their algorithms in this appeal, other than suggesting that they 'often moderate certain policy-violating content before users see it.' The Platforms never suggest their algorithms somehow exercise substantive, discretionary review akin to newspaper editors."), *cert. granted in part sub nom.* Netchoice, LLC v. Paxton, No. 22-555, 2023 WL 6319650 (U.S. Sept. 29, 2023).

145. NetChoice, L.L.C. v. Paxton, 573 F.Supp.3d 1092, 1108 (W.D. Tex. Dec. 1, 2021), *vacated and remanded sub nom. Paxton*, 49 F.4th 439.

146. *See* KISSINGER ET AL., *supra* note 20, at 193 ("AI is capable of exploiting human passions more effectively than traditional propaganda.").

147. *See, e.g.*, *id.* at 58 (describing how "AI fighter pilots have outperformed humans in simulated combat by executing maneuvers beyond the capabilities of human pilots"); Will Knight, *The Dark Secret at the Heart of AI*, MIT TECH. REV. (Apr. 11, 2017), https://www.technologyreview.com/2017/04/11/5113/the-dark-secret-at-the-heart-of-ai [https://perma.cc/4DKA-G39B]; Fergus Walsh, *AI*

to manipulate humans to achieve strategic goals.[148] Most importantly, this manipulation runs counter to the values protected by the First Amendment. First, it violates the principle of rational thought underlying the value of democratic self-governance and the marketplace of ideas.[149] Second, like false statements and fighting words, it interferes with the autonomy of the listener in order to influence beliefs and actions.[150] The impacts of editorial decision-making by AI on these First Amendment values are explained in more detail below.

1. Machine Manipulation: How an AI Editor Manipulates Users

To understand manipulation by AI, it is helpful to start with a definition. An influential characterization by Daniel Susser, Beate Roessler, and Helen Nissenbaum defines manipulation as "intentionally and covertly influencing [someone's] decision-making, by targeting and exploiting their decision-making vulnerabilities."[151] This Note considers an AI system to have intent to influence a human when it has an incentive to cause that influence and acts as if it is pursuing that incentive. In turn, an incentive to influence a human's state or behavior exists if such state or behavior increases the reward[152] the AI receives

'Outperforms' Doctors Diagnosing Breast Cancer, BBC NEWS (Jan. 2, 2020), https://www.bbc.com/news/health-50857759 [https://perma.cc/4KXF-YFBU].

148. *See infra* notes 163–166 and accompanying text.

149. *See* Helen Norton, *Manipulation and the First Amendment*, 30 WM. & MARY BILL RTS. J. 221, 238 (2021) ("[M]anipulation in public discourse additionally threatens collective harm to our democratic self-governance.").

150. *See* Council of Eur., *Declaration by the Committee of Ministers on the Manipulative Capabilities of Algorithmic Processes*, COUNCIL OF EUR. (Feb. 13, 2019), https://search.coe.int/cm/pages/result_details.aspx?objectid=090000168092dd4b [https://perma.cc/8SN2-S3NA] ("[A]lgorithmic persuasion may have significant effects on the cognitive autonomy of individuals and their right to form opinions and take independent decisions.").

151. Daniel Susser et al., *Technology, Autonomy, and Manipulation*, 8 INTERNET POL'Y REV., 1, 4 (2019).

152. The term "reward" refers specifically to reinforcement learning algorithms. In a supervised learning algorithm the equivalent concept to increasing reward would be decreasing loss. *See* Andrew G. Barto & Thomas G. Dietterich, *Reinforcement Learning and its Relationship to Supervised Learning*, in HANDBOOK OF LEARNING AND APPROXIMATE DYNAMIC PROGRAMMING 45, 51 (Si et al., 2004), https://ieeexplore.ieee.org/document/5273620 [https://perma.cc/Q5BQ-KS3J].

during training.[153] In the case of an AI recommendation system, an intent to influence behavior is consistently present: the AI gets rewarded for increased content engagement. Indeed, while a human editor chooses content for a wide variety of reasons, many of them without any intention to influence the reader, an AI system is always intending to influence because it is always seeking the reward it receives for changing human behavior.

Susser et al. describe acting covertly as attempting to influence the decision-making of a person in ways of which they are not aware and could not easily become aware.[154] AI recommendation systems act covertly for several reasons. First, the platforms do not reveal the details of their algorithms that might inform people how they are being influenced. While the platforms might release select pieces of information, they continue to obscure many of the details essential to understanding how the recommendation system operates.[155] Second is the concept of explainability. The prominent ML techniques used in recommender systems are often described as "black boxes," referring to the fact that there is generally no way for humans—even their creators—to understand how they arrive at their decisions.[156] Thus, even if the entire algorithm is known to the user, the AI still acts covertly because the specific ways in which it attempts to influence the user cannot be understood or explained.[157] To illustrate, knowing that an AI system looks at time, location, and recent activity on the platform is not the same as understanding that it is presenting content to you because it knows you're sad and tired and more vulnerable to certain viewing behavior. Third, even as explainability research progresses, there appears to be limits to its potential to eliminate manipulation. To illustrate this, consider the following situation: if Facebook's AI shows me a bicycling video, it can truthfully say that this is because I have shown an interest in bicycling in the past. But with thousands

153. Micah Carroll et al., *Characterizing Manipulation from AI Systems* (under review) (manuscript at 2), (available at https://arxiv.org/pdf/2303.09387.pdf [https://perma.cc/A9BB-MQRE]).

154. Susser et al., *supra* note 151, at 4.

155. *See, e.g.*, *Our Approach to Explaining Ranking*, *supra* note 23 (describing the recommendation system in broad terms without giving specific models or signals used).

156. *See* Chloe Xiang, *Scientists Increasingly Can't Explain How AI Works*, VICE (Nov. 1, 2022), https://www.vice.com/en/article/y3pezm/scientists-increasingly-cant-explain-how-ai-works [https://perma.cc/84G7-GVZQ].

157. *See* Carroll et al., *supra* note 153, at 3.

or millions of possible bicycle videos to choose from, it also has to explain why it chose this particular one above all others, including how each of its thousands of signals influenced its final decision. Either because there is simply too much information for humans to digest or because the AI has identified patterns and relationships in the data that have no human-comprehensible explanation, it is possible that certain AI systems will remain unexplainable.[158] For these reasons, AI recommendation systems act covertly because the user cannot know the specific ways in which the AI uses their data to influence them.

Thus, AI recommendation systems engage in manipulation because they pursue an incentive to influence human behavior by covertly exploiting an individual's traits and vulnerabilities. This is not to say that all AI recommendation systems will necessarily always manipulate, but the essential ingredients are there unless mitigated by techniques to eliminate incentives or increase transparency.

With the above theoretical framework in mind, this Note will next look at evidence of AI manipulation in practice. While platforms have made it difficult to study the effects of their content recommendation, the evidence indicates that there is significant manipulation.[159] We can see this by looking first at studies of AI manipulation in other contexts and second at the platforms' own claims about the impacts of their AI systems.

As an illustration of AI's capacity to manipulate, in 2021, researchers at Australia's national science agency created an AI that was specifically designed to exploit "vulnerabilities in the ways people make choices."[160] They found that the AI was able to direct humans to its desired choice about seventy percent of the time.[161] John Whittle, the director of the research team,

158. Bartosz Brożek et al., *The Black Box Problem Revisited. Real and Imaginary Challenges for Automated Legal Decision Making*, A.I. & L. (Apr. 4, 2023), https://link.springer.com/article/10.1007/s10506-023-09356-9 [https://perma.cc/VSU9-S4QM] ("[AI algorithms] were designed to find patterns in datasets which cannot be analyzed by humans with their limited computational capacities."); *see also* Yavar Bathaee, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, 31 HARV. J.L. & TECH. 889, 891–92 (2018).

159. *See* Carroll et al., *supra* note 153, at 1–2.

160. Jon Whittle, *AI Can Now Learn to Manipulate Human Behavior*, SCI. ALERT (Feb. 12, 2021), https://www.sciencealert.com/ai-can-now-learn-to-manipulate-human-behavior [https://perma.cc/H3WX-VLU5].

161. *Id.*

described how the research "shows machines can learn to steer human choice-making through their interactions with us."[162]

Indeed, the evidence indicates that influencing human behavior through manipulation is a common technique of today's AI, regardless of the context or the intention of its designers.[163] In 2019, researchers at Carnegie Mellon and Facebook AI Research created an AI system named Pluribus that was able to beat five of the best professional human poker players in a six-player game of Texas hold 'em, primarily through the effective use of bluffing.[164] One of its creators, Dr. Noam Brown, described how "[p]eople have this notion that [bluffing] is a very human ability – that it's about looking into the other person's eyes . . . . It's really about math, and this is what's going on here. We can create an AI algorithm that can bluff better than any human."[165] While bluffing in poker is not the same as editorial discretion, it illustrates the ability of AI to independently learn to exploit human tendencies to manipulate our behavior.[166]

---

162. *Id.*; *see also* Liesl Yearsley, *We Need to Talk About the Power of AI to Manipulate Humans*, MIT TECH. REV. (June 5, 2017), https://www.technologyreview.com/2017/06/05/105817/we-need-to-talk-about-the-power-of-ai-to-manipulate-humans [https://perma.cc/5CJ8-2TWY] ("Every behavioral change we at Cognea wanted, we got. If we wanted a user to buy more product, we could double sales. If we wanted more engagement, we got people going from a few seconds of interaction to an hour or more a day.").

163. *See* Carroll et al., *supra* note 153, at 4; Lauren E. Willis, *Deception by Design*, 34 HARV. J.L. & TECH. 115, 150 (2020) ("[M]achine learning systems can teach themselves to select and engage in deception as a strategy even when humans did not design the systems to engage in deception. When Facebook's artificial intelligence research lab used machine learning to train bots to negotiate with humans, the bots quickly adopted strategies that involved 'initially feigning interest in a valueless item, only to later "compromise" by conceding it.' The researchers did not expect this result: 'Our agents [learned] to deceive without any explicit human design, simply by trying to achieve their goals.' In 2018, an artificial intelligence system deceived its own programmers about how it was performing a task.").

164. Noam Brown & Tuomas Sandholm, *Superhuman AI for Multiplayer Poker*, 365 SCIENCE 885 (2019), https://www.science.org/doi/10.1126/science.aay2400 [https://perma.cc/N96N-7SQB]; Daniela Hernandez, *Computers Can Now Bluff Like a Poker Champ. Better, Actually.*, WALL ST. J. (July 12, 2019), https://www.cmu.edu/ambassadors/october-2019/artificial-intelligence [https://perma.cc/EXX9-VFH6].

165. Hernandez, *supra* note 164.

166. The comparison between poker bluffing and human communication is not entirely abstract: Dr. Brown and the Meta AI team have since been working on AI that can "negotiate, persuade, and work with people to achieve strategic goals similar to the way humans do" and created an AI that "achieved more than double the average score of the human players and ranked in the top 10 percent of participants" in the online game Diplomacy. *CICERO: An AI Agent that Negotiates,*

AI's manipulation of human behavior on social media is not dissimilar to its success in more constrained scenarios such as Pluribus.[167] On social media, AI is able to use the same techniques to learn how to change user behavior to increase engagement.[168] It has access to extensive behavioral data on every interaction that a user makes on the platform—including clicking, liking, commenting, amount of time watching, scrolling, pausing scrolling to view content, mouse hovering, etc.—as well as a substantial amount of data from off of the platform.[169] Indeed, "[i]t is no exaggeration to say that popular platforms . . . know [their] users better than their families and friends do."[170] With all of this data, the AI is told to maximize engagement as measured by a variety of metrics such as likes, clicks, comments, or time spent on content.[171] When it

---

Persuades, and Cooperates with People, META AI BLOG (Nov. 22, 2022), https://ai.facebook.com/blog/cicero-ai-negotiates-persuades-and-cooperates-with-people [https://perma.cc/CF5X-97G2]. The success of CICERO illustrates the ability of AI to shape human behavior with a specific goal in mind.

167. *See* Dipayan Ghosh & Ben Scott, *#DIGITALDECEIT: The Technologies Behind Precision Propaganda on the Internet*, NEW AMERICA: PUB. INT. TECH. PROGRAM (Jan. 2018), https://d1y8sb8igg2f8e.cloudfront.net/documents/digital-deceit-final-v3.pdf [https://perma.cc/P4PZ-5KW5] ("There are strong similarities between this type of advertising industry AI and the Google-developed AI that recently defeated a Go master.").

168. *See Q&A with Michael Schrage*, *supra* note 12 ("In nearly every dimension of life — relationships, health care, pop culture — machines demonstrably can offer better advice than our best friends. Something I thought would likely take a decade, happened in barely two or three years. I was surprised by how quickly recommenders got super-smart about their users.").

169. *See* Akos Lada et al., *How Does News Feed Predict What You Want to See?*, TECH AT META (Jan. 26, 2021), https://tech.facebook.com/engineering/2021/01/news-feed-ranking [https://perma.cc/5M4G-AXQU]; AMNESTY INT'L, SURVEILLANCE GIANTS: HOW THE BUSINESS MODEL OF GOOGLE AND FACEBOOK THREATENS HUMAN RIGHTS 15 (2019), https://www.amnesty.org/en/documents/pol30/1404/2019/en/ [https://perma.cc/5L9J-FQVP] ("In practice, this means people are constantly tracked when they go about their day-to-day affairs online, and increasingly in the physical world as well."); Goodrow, *supra* note 48 ("A number of signals build on each other to help inform our system about what you find satisfying: clicks, watch time, survey responses, sharing, likes, and dislikes.").

170. Georgios Petropoulos, *The Dark Side of Artificial Intelligence: Manipulation of Human Behaviour*, BRUEGEL (Feb. 2, 2022), https://www.bruegel.org/blog-post/dark-side-artificial-intelligence-manipulation-human-behaviour [https://perma.cc/4F73-Q7PU]; Ankur A. Patel, *How TikTok Uses AI to Engineer User Addiction*, ANKUR'S NEWSLETTER (May 13, 2022), https://www.ankursnewsletter.com/p/how-tiktok-uses-ai-to-engineer-user [https://perma.cc/8Z7R-UJZL].

171. *See* Ben Smith, *How TikTok Reads Your Mind*, N.Y. TIMES (Dec. 5, 2021), https://www.nytimes.com/2021/12/05/business/media/tiktok-algorithm.html [https://perma.cc/VHP7-WXBN]; Zeynep Tufekci, *It's the (Democracy-Poisoning)*

recommends a piece of content, the AI is able to observe how the user behaves in response to it and adjust its next recommendation based on how well it performed with the initial recommendation.[172] To AI, the task appears just like a game, where it must continually adjust its actions to maximize a measurable result, and our interactions on the platform become "an optimization loop for human behavior."[173] As companies gather more data on us and AI continues to improve, it will only get better at exploiting human tendencies to maximize its own goals.

AI is able to manipulate more effectively than a human editor because it is able to act on an individual level. While a human editor might select content with the goal of influencing their readers, they must do so at the population level, attempting to exploit general human tendencies without the ability to target the context or vulnerabilities of individuals. In contrast, AI selects content based on personal data gathered from its users, and it covertly uses that information to attempt to influence the user. In the words of Professor Ryan Calo, platforms "will be in a position to surface and exploit how consumers tend to deviate from rational decisionmaking on a previously unimaginable scale. Thus, [they] will increasingly be in the position to create suckers, rather than waiting for one to be born."[174]

Even compared to a hypothetical human editor who personalizes content for an individual, AI is capable of doing things well beyond the abilities of a human or any prior technology.[175] When an AI system performs a task, it operates

---

Golden Age of Free Speech, WIRED (Jan. 16, 2018), https://www.wired.com/story/free-speech-issue-tech-turmoil-new-censorship [https://perma.cc/C6VN-ZGYT].

172. *See, e.g.*, Willis, *supra* note 163, at 129.

173. Chollet, *supra* note 11.

174. Ryan Calo, *Digital Market Manipulation*, 82 GEO. WASH. L. REV. 995, 1018 (2014); *see also* Willis, *supra* note 163, at 147 ("Machine learning is ideal for [exploiting vulnerability because] it can identify relationships unexpected by humans or involving too many interacting variables for humans to assess. As one journalist concluded after interviewing marketers of scam products (e.g., sham diet pills, fake antivirus software), 'Facebook's targeting algorithm is so powerful . . . [the marketers] don't need to identify suckers themselves—Facebook does it automatically.'").

175. *See* Brown, *supra* note 15 ("In some cases, machine learning can gain insight or automate decision-making in cases where humans would not be able to, [Director of the MIT Center for Deployable Machine Learning Aleksander] Madry said. 'It may not only be more efficient and less costly to have an algorithm do this,

in a "realm just beyond human reach," choosing its own path to accomplish its goal based on the world it sees.[176] Because its perception of the world—through processing amounts of data far beyond the ability of humans to hold and process, at speeds beyond that of the human brain—is fundamentally non-human, AI is not merely the equivalent of an exceptionally skilled human.[177] Rather, it exercises its own "form of learning and logical evaluation" to identify new solutions and "capture intricate connections, including connections that can elude humans."[178] This "alien"[179] form of learning and evaluation is what allows it to discover entirely new solutions to problems that expert humans have been unable to access.[180] An example of this comes from the discovery of a new antibiotic drug called halicin:

> The AI used to identify halicin illustrates the centrality of the machine-learning process. When MIT researchers designed a machine-learning algorithm to predict the antibacterial properties of molecules, training the algorithm with a dataset of more than two thousand molecules, the result was something no conventional algorithm—and no human—could have accomplished. Not only do humans not understand the connections AI revealed between a compound's properties and its antibiotic capabilities, but even more fundamentally,

---

but sometimes humans just literally are not able to do it,' he said."); Schneier, *supra* note 12 ("It's not just a difference in degree; it's a difference in kind.").

176. KISSINGER ET AL., *supra* note 20, at 24.

177. *See* Schneier, *supra* note 12 ("AIs don't solve problems like humans do. Their limitations are different than ours . . . . [T]hey'll look at more types of solutions. They'll go down paths that we simply have not considered, paths more complex than the sorts of things we generally keep in mind.").

178. KISSINGER ET AL., *supra* note 20, at 24, 64.

179. Warr et al., *A Chat about GPT3 (and Other Forms of Alien Intelligence) with Chris Dede*, 67 TECHTRENDS 396 (2023); Andréa Morris, *The Paradox of Predicting AI*, FORBES (July 5, 2023), https://www.forbes.com/sites/andreamorris/2023/07/05/the-paradox-of-predicting-ai-unpredictability-is-a-measure-of-intelligence [https://perma.cc/8N46-UKDW].

180. *See* KISSINGER ET AL., *supra* note 20, at 8–11 ("[The AI] had a logic of its own, informed by its ability to recognize *patterns* of moves across vast sets of possibilities human minds cannot fully digest or employ . . . . The AI did not just process data more quickly than humanly possible; it also detected aspects of reality humans have not detected, or perhaps cannot detect."); Kyle Wiggers, *DeepMind Claims AI Has Aided New Discoveries and Insights in Mathematics*, VENTUREBEAT (Dec. 1, 2021), https://venturebeat.com/uncategorized/deepmind-claims-ai-has-aided-new-discoveries-and-insights-in-mathematics [https://perma.cc/8BET-QLKM].

> the properties themselves are not amenable to being expressed as rules. A machine-learning algorithm that improves a model based on underlying data, however, is able to recognize relationships that have eluded humans.[181]

Similarly, in 2015, researchers at Mount Sinai Hospital trained a machine learning model named Deep Patient to predict future disease occurrences in patients.[182] Not only was Deep Patient "just way better"[183] than existing methods for predicting disease, but it was able to predict "the onset of psychiatric disorders like schizophrenia surprisingly well. But since schizophrenia is notoriously difficult for physicians to predict, [lead researcher Joel] Dudley wondered how this was possible. He still doesn't know."[184] Deep Patient was able to identify patterns in the data that captured an aspect of reality that anticipated schizophrenia, an aspect that humans are still unable to identify.

In the context of social media, platforms have found that AI's superhuman processing of personal data makes it extremely effective at changing our behavior. Speaking at a panel in 2018, YouTube's Chief Product Officer, Neal Mohan, explained that he sees it as YouTube's job to "give the [user] a steady stream, almost a synthetic or personalized channel" using AI.[185] He said that "more than 70 percent of the time you spend watching on [YouTube], you're lured in by one of the service's AI-driven recommendations."[186] And this effectiveness is specifically attributable to AI: *Fast Company* reported in 2017 that "[s]ince YouTube turned to Google Brain (another of the company's AI-research arms) to tune its video recommendations, it has increased average watch time by 50% each of the last three years."[187] Similarly, in 2022, Meta CEO Mark Zuckerberg said that in one quarter the company "saw a more than 30% increase

---

181. *Id.* at 62.
182. Knight, *supra* note 147.
183. *Id.*
184. *Id.*
185. Joan E. Solsman, *YouTube's AI is the Puppet Master Over Most of What You Watch*, CNET (Jan. 10, 2018), https://www.cnet.com/tech/services-and-software/youtube-ces-2018-neal-mohan [https://perma.cc/C6GX-GT7L].
186. *Id.*
187. *How Apple, Facebook, Amazon, And Google Use AI to Best Each Other*, FAST COMPANY (Oct. 11, 2017), https://www.fastcompany.com/40474585/how-apple-facebook-amazon-and-google-use-ai-to-best-each-other [https://perma.cc/7QP5-3A7C].

in the time that people spent engaging with Reels across Facebook and Instagram" and "not[ed] that much of this increase was attributable to advancements in its artificial intelligence recommendations."[188]

These statistics demonstrate how AI recommendations can be effective at changing *behavior* by increasing engagement with content. However, the evidence indicates that AI recommendation systems can also learn to manipulate user *preferences*.[189] In his research, Professor Stuart Russell found that "the algorithms have learnt to manipulate people to change them so that in future they're more susceptible and they can be monetised at a higher rate."[190] While social media platforms generally respond to criticism of their recommendations by explaining that they are simply giving people the content they want and value,[191] Russell's research indicates that the AI has been able to determine that the best way to maximize engagement is by actually changing our preferences to match the content it presents, rather than changing the content to match our preferences.[192] He explains this as "a sort of a failure mode . . . of any AI system that's trying to satisfy human preferences . . . . One way to satisfy them is to change them so that they're already satisfied."[193]

This is not to say Facebook or any other platform has a malicious plan to control or manipulate us. Rather, the AI itself

---

188. Connor Perrett, *Instagram and Facebook Feeds Are About to Be Swarmed by Accounts Users Don't Follow*, BUS. INSIDER (July 28, 2022), https://www.businessinsider.com/instagram-feed-content-accounts-dont-follow-to-double-2022-7 [https://perma.cc/MES7-2Z94] (quoting Mark Zuckerberg); Q1 of 2023 saw AI recommendations drive more than a "24% increase in time spent on Instagram," *First Quarter 2023 Results Conference Call*, *supra* note 25.

189. *See* Carroll et al., *Estimating and Penalizing Induced Preference Shifts in Recommender Systems*, PROC. OF THE 39TH INT'L CONF. ON MACH. LEARNING (2022); Charles Evans & Atoosa Kasirzadeh, *User Tampering in Reinforcement Learning Recommender Systems*, Presented at the AAAI/ACM Conference on AI, Ethics, and Society (2023).

190. Robin Pomeroy, *Transcript: The Promises and Perils of AI - Stuart Russell on Radio Davos*, WORLD ECON. F. (Jan. 6, 2022), https://www.weforum.org/agenda/2022/01/artificial-intelligence-stuart-russell-radio-davos [https://perma.cc/QZ32-ZBQM].

191. *E.g.*, Goodrow, *supra* note 48.

192. *See* Stuart Russell, *Artificial Intelligence and the Problem of Control*, PERSP. ON DIGIT. HUMANISM 19, 21 (2022) ("Rather than simply adjusting their recommendations to suit human preferences, these algorithms will, in pursuit of their long-term objective, learn to manipulate humans to make them more predictable in their clicking behavior.").

193. Pomeroy, *supra* note 190.

has the power to manipulate, regardless of the platform's intentions.[194] Even when a platform only intends to influence user behavior by increasing engagement with content, the AI it creates is able to develop its own internal goals to change user preferences as a way of pursuing the platform's goals.[195]

In addition to AI developing its own internal goals, manipulation unintended by the platforms also occurs through the intentional spread of content by actors taking advantage of AI's ability to exploit us.[196] There is substantial evidence of malicious actors using social media campaigns to influence political views or recruit followers to their cause.[197] The power of AI gives people outside of the platforms an extremely effective tool for influencing public discourse in ways that would be impossible on their own.[198] In either case, the result is that we,

194. *See* Carroll et al., *supra* note 189, at 1-2.

195. *Id.*

196. AMNESTY INTERNATIONAL, *supra* note 169, at 28 ("In some cases, such impacts are directly caused by the company's technology itself; in other cases, these tools can be exploited by other actors in ways that harm rights."); *see, e.g.*, Gonzalez v. Google L.L.C., 2 F.4th 871, 937 (9th Cir. 2021) (Gould, J., concurring) ("[T]errorist organizations like ISIS have obviously played Google and YouTube like a fiddle."), *cert. granted*, 214 L. Ed. 2d 12 (Oct. 3, 2022), and *cert. granted sub nom.* Twitter, Inc. v. Taamneh, 214 L. Ed. 2d 12 (Oct. 3, 2022), and *vacated and remanded*, 143 S. Ct. 1191 (2023), and *rev'd sub nom.* Twitter, Inc. v. Taamneh, 143 S. Ct. 1206 (2023).

197. Pomeroy, *supra* note 190 ("[T]here's a massive human-driven industry that sprung up to feed this whole process . . . . So people have hijacked the ability of the algorithms to very rapidly change people because it's hundreds of interactions a day, everyone has a little nudge. But if you nudge somebody hundreds of times a day for days on end, you can move them a long way in terms of their beliefs, their preferences, their opinions. . . . [People] hijacked the process to take advantage of it and create the polarisation that suits them for their purposes."); *e.g.*, SOCIAL MEDIA MANIPULATION BY POLITICAL ACTORS AN INDUSTRIAL SCALE PROBLEM - OXFORD REPORT, UNIV. OF OXFORD (2021), https://www.ox.ac.uk/news/2021-01-13-social-media-manipulation-political-actors-industrial-scale-problem-oxford-report [https://perma.cc/EB8P-NYVG]; *see generally* SEBASTIAN BAY, ET AL., SOCIAL MEDIA MANIPULATION 2021/2022: ASSESSING THE ABILITY OF SOCIAL MEDIA COMPANIES TO COMBAT PLATFORM MANIPULATION (Monika Hanley ed., 2022), https://stratcomcoe.org/publications/download/SOCIAL-MEDIA-MANIPULATION-2021-2022-FINAL.pdf [https://perma.cc/NRA8-N56J]; *see generally* Allyson Haynes Stuart, *Social Media, Manipulation, and Violence*, 15 S.C. J. INT'L L. & BUS. 100 (2019).

198. *See* AMNESTY INTERNATIONAL, *supra* note 169, at 31 ("Google's Redirect Method . . . [is] a project that uses the company's AdWords platform (now called Google Ads) to deradicalize potential supporters of Islamic terrorism. One commentator successfully used the same tool – which is freely available online – to nudge suicidal people to call a helpline. This demonstrates that such 'social engineering' could easily be used to manipulate people's opinions and beliefs, either by the companies directly or by other actors. Although in the latter examples, such

the users of social media, are continuously manipulated and exploited to serve the ends of others through the use of AI.

2. The Impact of AI Manipulation on First Amendment Values

Manipulation of users by AI runs counter to each of the core values underlying First Amendment editorial protection. First, manipulation interferes with the foundation of rational choice underlying the value of uninhibited debate to democratic self-governance and the marketplace of ideas.[199] If debate is based not on rational thought but instead on coercive nudging and manipulation, then it provides no value to democratic self-determination. As Professor Lidsky explains, "[d]emocracy accords citizens the right to choose their collective fates: that choice is meaningless if it is coerced or manipulated. In fact, democratic decisions are simply illegitimate if they violate this fundamental principle."[200] Like the false statements of fact in *Gertz*, "there is no constitutional value" in manipulated debate because it does not "materially advance[] society's interest in 'uninhibited, robust, and wide-open' debate on public issues."[201]

Moreover, truth will not ultimately prevail in an environment curated by AI. The proliferation of fake news and falsehoods on social media renders any assertion that social media leads to truth highly doubtful.[202] More importantly,

---

influence was used for a purportedly positive objective, these tools could easily be (mis)used in ways that harm our rights, particularly if deployed at scale.").

199. *See* Schneier, *supra* note 12 ("[AI] will artificially influence what we think is normal, what we think others think. This sort of manipulation is not what we think of when we laud the marketplace of ideas, or any democratic political process.").

200. Lidsky, *supra* note 71, at 840; *see also* Daniel Susser et al., *Online Manipulation: Hidden Influences in a Digital World*, 4 GEO. L. TECH. REV. 1, 43 (2019) ("On one hand, manipulative practices undermine individual autonomy—people's capacity for self-government, their ability to pursue their own goals. . . . But perhaps more worrying are the threats to *collective* self-government. When citizens are targets of online manipulation and voter decisions rather than purchase decisions are swayed by hidden influence, democracy itself is called into question.").

201. Gertz v. Robert Welch, Inc., 418 U.S. 323, 340 (1974) (citing New York Times Co. v. Sullivan, 376 U.S. 254, 270 (1964)); *see also* Post, *supra* note 68 ("Coercion is precluded from public debate because the very purpose of that debate is the practice of self-determination.").

202. *See* Tufekci, *supra* note 171 ("John Stuart Mill's notion that a 'marketplace of ideas' will elevate the truth is flatly belied by the virality of fake news."); *see also* Peter Dizikes, *Study: On Twitter, False News Travels Faster Than True Stories*,

manipulation inherently interferes with our ability to judge and determine for ourselves the truth and value of information. Rather than coming from the rational choices of autonomous individuals, beliefs are formed by AI "covertly influenc[ing] those listeners' choices without their conscious awareness and by targeting and exploiting their vulnerabilities."[203] AI guides users down individualized information paths designed to serve its own internal goals, precluding them from freely exercising reason and judgment to arrive at their own conclusions. As Professor Julie Cohen describes, "the rational listener's presumptive autonomy increasingly is displaced by automaticity–by habitual, precognitive behaviors that require no conscious attention."[204]

This violation of autonomy is a core component of manipulation.[205] AI manipulation attempts to bypass rational decision-making and influence behavior by exploiting hidden vulnerabilities without the knowledge of the target. It is a direct assault on "the principle that each person should decide for [themselves] the ideas and beliefs deserving of expression, consideration, and adherence."[206] As Professor Fallon explains, "descriptive autonomy requires freedom from coercion, manipulation, and temporary distortion of judgment."[207] AI invades our autonomy by exploiting our personal data in an attempt to produce behavior that could not be achieved through reason or argument alone.[208] This is the opposite of appealing to "the power of reason" and persuasion through rational argument.[209] As AI "fundamentally chang[es] the nature of

---

MIT NEWS (Mar. 8, 2018), https://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308 [https://perma.cc/G35V-VC5Q].

203. Norton, *supra* note 149, at 224.

204. Julie E. Cohen, *Tailoring Election Regulation: The Platform Is the Frame*, 4 GEO. L. TECH. REV. 641, 652 (2020).

205. Carroll et al., *supra* note 153, at 4; Cass Sunstein, *Fifty Shades of Manipulation*, 1 J. MARKETING BEHAV. 213 (2016).

206. Turner Broad. Sys., Inc. v. F.C.C., 512 U.S. 622, 641 (1994).

207. Fallon, *supra* note 81, at 877.

208. *See* Shaun B. Spencer, *The Problem of Online Manipulation*, 2020 U. ILL. L. REV. 959, 963–66, 974–75 (2020); *e.g.*, Nick Statt, *Facebook Reportedly Ignored Its Own Research Showing Algorithms Divided Users*, THE VERGE (May 26, 2020), https://www.theverge.com/2020/5/26/21270659/facebook-division-news-feed-algorithms [https://perma.cc/3GSY-K5MH] (reporting on internal Facebook research that found "[o]ur algorithms exploit the human brain's attraction to divisiveness.").

209. Whitney v. California, 274 U.S. 357, 375 (1927) (Brandeis, J., concurring), *overruled by* Brandenburg v. Ohio, 395 U.S. 444 (1969). In the context of this

persuasion," we must ask, "[h]ow far does that ability need to go before it erodes the autonomy of those targeted to make decisions of their own free will?"[210]

AI recommendation systems go far beyond traditional techniques of influence; they have "radically transformed our relationship with choice."[211] While humans have always employed tactics to manipulate, AI is able to do so to a far greater degree and in ways beyond the capacity of any human.[212] "[I]t is altering the role that our minds have traditionally played in shaping, ordering, and assessing our choices and actions."[213] Our political system is premised on "individual dignity and choice,"[214] but "[r]ecommendation engines represent a global revolution in how choice can be

---

discussion, persuasion can be thought of as appealing to the rational decision-making of the listener and respecting their autonomy. *See* Susser et al., *supra* note 200, at 3.

210. Bruce Schneier, *The Peril of Persuasion in the Big Tech Age*, FOREIGN POL'Y: ARGUMENT (Dec. 11, 2020), https://foreignpolicy.com/2020/12/11/big-tech-data-personal-information-persuasion [https://perma.cc/L6QC-T2JR]; *see also* Council of Eur., *supra* note 150 ("The Committee of Ministers . . . encourages member States to . . . initiat[e] . . . open-ended, informed and inclusive public debates with a view to providing guidance on where to draw the line between forms of permissible persuasion and unacceptable manipulation. The latter may take the form of influence that is subliminal, exploits existing vulnerabilities or cognitive biases, and/or encroaches on the independence and authenticity of individual decision-making.").

211. *Q&A with Michael Schrage*, *supra* note 12; *see* Norton, *supra* note 149, at 227 ("Twenty-first-century technologies—including the use of predictive algorithms informed by the collection and analysis of huge amounts of data—thus create opportunities for manipulation different in both degree and in kind from more traditional forms of manipulation.").

212. *See* Olaf J. Groth et al., *AI Algorithms Need FDA-Style Drug Trials*, WIRED (Aug. 15, 2019, 9:00 AM), https://www.wired.com/story/ai-algorithms-need-drug-trials [https://perma.cc/67NB-8M9P]. Furthermore, despite humanity's long history of manipulation, several commentators have recently made compelling arguments for regulating the broader spectrum of digital manipulation based on the transformative capabilities of modern digital techniques. *See* Norton, *supra* note 149, at 233; Susser et al., *supra* note 200, at 43–45; Willis, *supra* note 163, at 188–90; Spencer, *supra* note 208, at 1000–01; Ryan Calo, *Digital Market Manipulation*, 82 GEO. WASH. L. REV. 995, 998–99 (2014); Ido Kilovaty, *Legally Cognizable Manipulation*, 34 BERKELEY TECH. L.J. 449, 499–501 (2019). While there may always be some difficult line-drawing problems, identifying the use of AI techniques with enough manipulative potential to warrant regulation offers a relatively clear path to addressing a wide range of problematic practices. *See* Norton, *supra* note 149, at 236 ("Indeed, algorithmic manipulation at times may be easier to identify and measure—and thus responsibly regulate—than that by manipulative humans.").

213. KISSINGER ET AL., *supra* note 20, at 53.

214. Cohen v. California, 403 U.S. 15, 24 (1971).

personalized, packaged, presented, experienced, and understood."[215] Such a dramatic transformation in our ability to understand and shape our own choices influences our autonomy in ways we have not encountered before.

It is worth noting that the "quality" of the content that AI presents to us and the particular harms it might cause—such as the spread of hate speech, fake news, and political polarization—are not the focus here because they do not impact the values protected by the First Amendment. These harms demonstrate the capabilities of AI, but it is the manipulation that does not serve the values underlying the First Amendment. Even accepting the platforms' assertions that AI content recommendation systems only present content that users value most,[216] and assuming that such content is in fact beneficial for us in the long term, AI's manipulation to achieve its own goals still runs counter to the principle of reason and rationality underlying the First Amendment.[217] Through the lens of Professor Strauss' persuasion principle, manipulation by AI denies our individual autonomy by "interfer[ing] with [our] control over [our] own reasoning processes."[218] It declares that it knows what we want and, therefore, we should accept its control over our content consumption. When a human trains a dog using food, the dog is happy to change its behavior to get the reward; food is indeed what it wants and needs most. But the fact that food is good for the dog does not change the fact that the human is in control, taking advantage of the dog's wants and needs to achieve their own ends. Similarly, AI is able to identify and exploit our individual wants and needs by analyzing vast amounts of our behavioral data.[219] Whether it is actually acting

---

215. Michael Schrage, *The Recommender Revolution*, MIT TECHN. REV. (Apr. 27, 2022), https://www.technologyreview.com/2022/04/27/1048517/the-recommender-revolution [https://perma.cc/V5SJ-SW9V].

216. *Our Approach to Facebook Feed Ranking*, META: TRANSPARENCY CENTER (June 29, 2023), https://transparency.fb.com/features/ranking-and-content [https://perma.cc/9G9Q-MPFA].

217. *See* Lidsky, *supra* note 71, at 839 ("If citizens are incapable of exercising their rational faculties to participate in public discourse, then they are equally incapable of rational self-governance. To reject the possibility of a rational citizenry, therefore, is to reject the democratic ideal.").

218. Strauss, *supra* note 91, at 354; *see also* Cohen, *supra* note 204, at 651 ("[P]latform-based, massively intermediated information environments are not designed for the rational listener. Instead, they are both systematically configured and continually reoptimized to elicit automatic, precognitive interactions with online content.").

219. *See* Norton, *supra* note 149, at 231.

in our "best interest" or not, handing control of substantial corners of public discourse over to AI pursuing its own goals does not "comport with the premise of individual dignity and choice upon which our political system rests."[220]

B. Delegation of Editorial Control

When an editor chooses to use AI to make editorial decisions, they are releasing their own editorial discretion and control over the published product. In other words, the degree of descriptive autonomy exercised over editorial decision-making is significantly decreased. "[T]he very essence of publishing is making the decision whether to print or retract a given piece of content,"[221] yet when a social media platform employs AI as an editor, no human is making publication decisions regarding the vast majority of content.[222] In considering Florida's social media regulation, the district court noted "[s]omething well north of 99% of the content that makes it onto a social media site never gets reviewed" by a person.[223] Because of this, it distinguished the situation from the more traditional publishers in *Tornillo*, *PG&E*, and *Hurley*, concluding that "it cannot be said that a social media platform, to whom most content is invisible to a substantial extent, is indistinguishable for First Amendment purposes from a newspaper or other traditional medium."[224]

---

220. Cohen v. California, 403 U.S. 15, 24 (1971); *see also* David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, U.N. Doc. A/73/348 (Aug. 29, 2018) ("In an AI-governed system, the dissemination of information and ideas is governed by opaque forces with priorities that may be at odds with an enabling environment for media diversity and independent voices.").

221. Klayman v. Zuckerberg, 753 F.3d 1354, 1359 (D.C. Cir. 2014).

222. Platforms generally do not publicly release what percentage of their content is reviewed by a human, so the evidence from trial in *NetChoice, LLC v. Moody* is a best estimate. *See, e.g.*, BRADFORD ET AL., YALE LAW SCHOOL: THE JUSTICE COLLABORATORY, REPORT OF THE FACEBOOK DATA TRANSPARENCY ADVISORY GROUP 17, 43 (2019), https://law.yale.edu/sites/default/files/area/center/justice/document/dtag_report_5.22.2019.pdf [https://perma.cc/2U87-U5TA].

223. NetChoice, LLC v. Moody, 546 F. Supp. 3d 1082, 1092 (N.D. Fla. 2021), *aff'd in part, vacated in part, remanded sub nom.* NetChoice, LLC v. Att'y Gen., Fla., 34 F.4th 1196 (11th Cir. 2022),), *cert. granted in part sub nom.* Moody v. Netchoice, LLC, No. 22-277, 2023 WL 6319654 (U.S. Sept. 29, 2023), and *cert. denied sub nom.* Netchoice v. Moody, No. 22-393, 2023 WL 6377782 (U.S. Oct. 2, 2023).

224. *Id.* at 1093. Nonetheless, the Court struck down the regulations because they were content-based and intruded on those cases "on which the platforms are most likely to exercise editorial judgment." *Id.* at 1092.

1. Delegation of Editorial Decision-Making to AI

To appreciate the extent of this delegation, we must understand the operation of current AI systems. As explained above, when a platform employs AI to decide where and when to publish content, there are no clear rules for decision-making programmed in.[225] They exercise broad, generalized control over the structure of the models, but the AI decides what content to surface in the first instance and the platform then applies its own filters and business logic on top. This is not to say that the platforms are incapable of exercising control over what they present. They are always free to change their systems, and they are constantly doing so.[226] Rather, it is to say that *when and to the extent that a platform employs AI*, the platform is giving up control of the decision-making process over what content is presented to users. The question is whether and how much someone should be able to delegate their discretion to AI without proper safeguards, even in an expressive context.

When a platform uses AI for a recommendation system, it does not know what content will be recommended to each user or how the AI chose to recommend that content.[227] A platform does not know what decisions the AI will make in specific situations.[228] Companies turn to AI because it is able to perform better than humans, but this also means that it is inherently making decisions that humans would not have made and cannot predict.[229]

---

225. *See, e.g.*, Goodrow, *supra* note 48 ("[O]ur recommendation system doesn't operate off of a 'recipe book' of what to do. It's constantly evolving, learning every day from over 80 billion pieces of information we call signals.").

226. For instance, Twitter owner Elon Musk reportedly directed his engineers to increase the promotion of his own tweets. Zoe Schiffer & Casey Newton, *Yes, Elon Musk Created a Special System for Showing You All His Tweets First*, THE VERGE (Feb. 14, 2023), https://www.theverge.com/2023/2/14/23600358/elon-musk-tweets-algorithm-changes-twitter [https://perma.cc/E9HX-8MPL].

227. *See* Knight, *supra* note 147 ("The computers . . . have programmed themselves, and they have done it in ways we cannot understand. Even the engineers who build these apps cannot fully explain their behavior.").

228. Miriam C. Buiten, *Towards Intelligent Regulation of Artificial Intelligence*, 10 EUR. J. RISK REG. 41, 50 (2019) ("In essence, in advance of their use, developers are not able to predict or explain their functioning.").

229. *See* Brożek et al., *supra* note 158, at 8 ("Given that AI algorithms – and in particular machine learning – are capable of analyzing huge datasets in ways far exceeding the abilities of the human mind, our hope is that the algorithms will produce better outcomes than humans are capable of. However, this means that these outcomes will be unexpected.").

One could argue that, while the platforms do not necessarily oversee individual content recommendations, they sufficiently define and control the AI to retain the essential editorial discretion over the content presented to users. After all, they write the algorithms, define the goals AI is incentivized to pursue, and employ substantial filtering of objectionable content.[230] This is all true, but it does not tell the whole story.

As noted above, research has shown that AI recommendation systems of the kind currently employed by social media platforms are able to learn to manipulate user preferences as a way of achieving their goals, even absent any intention to do so on the part of their creator.[231] The AI learns to pursue its human-defined goal of engagement by pursuing an internal goal it identifies independent from human intervention.[232] Indeed, in a broad range of contexts, AI systems have developed internal goals and techniques that were not intended or desired by their creators.[233] Though current AI may be far from consciousness or general human-level intelligence, this does not change the "increasingly evident fact that ML systems are not fully under human control."[234]

Loss of control over content decisions due to AI is a central issue for modern social media platforms.[235] In 2021, Twitter announced its Responsible Machine Learning Initiative to study and address harms caused by the company's use of AI, explaining that "[w]hen Twitter uses [machine learning], . . . sometimes, the way a system was designed to help could start to

230. 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 [https://perma.cc/8AY4-DBEB].

231. Evans & Kasirzadeh, *supra* note 189.

232. *Id*. at 2.

233. Willis, *supra* note 163, at 150.

234. CHAN ET AL., PROCEEDINGS OF THE 2023 ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY, HARMS FROM INCREASINGLY AGENTIC ALGORITHMIC SYSTEMS 651, 652 (2023).

235. Indeed, lack of control is a central issue in modern AI in general—more commonly referred to as the problem of "alignment." *See generally* BRIAN CHRISTIAN, THE ALIGNMENT PROBLEM: MACHINE LEARNING AND HUMAN VALUES (2020). Many of the effects of AI on social media can be thought of as alignment problems, where the AI behaves in unintended ways because of the misalignment between its training and specifications and the actual intentions of the platform. *See* Pan et al., *The Effects of Reward Misspecification: Mapping and Mitigating Misaligned Models*, at 1, arXiv:2201.03544 (Feb. 14, 2022).

behave differently than was intended."[236] Twitter recognized that it could not always control the behavior of its AI and needed to make efforts to study and monitor its behavior. Similarly, TikTok has spoken about its efforts to exercise more control over problematic content presentation decisions by its AI, explaining that "[g]etting these systems and tools right will take time and iteration."[237] While the platforms are making efforts to gain more control over their AI systems, delegation to AI necessarily entails a loss of control.[238] Furthermore, measures intended to control AI behavior can sometimes have the opposite effect. Researchers have found that some of the particular techniques platforms are using to counter problematic content—such as long-term planning[239] and human feedback[240]—can actually increase the likelihood of unintended manipulative behaviors.[241]

This lack of control over AI behavior was also a concern within Google as it contemplated transitioning its search engine to machine learning.[242] There was internal resistance to giving

236. Jutta Williams & Rumman Chowdhury, *Introducing Our Responsible Machine Learning Initiative*, TWITTER: BLOG (Apr. 14, 2021), https://blog.twitter.com/en_us/topics/company/2021/introducing-responsible-machine-learning-initiative [https://perma.cc/6DKL-3PMT].

237. *An Update on Our Work to Safeguard and Diversify Recommendations*, TIKTOK: NEWSROOM (Dec. 16, 2021), https://newsroom.tiktok.com/en-us/an-update-on-our-work-to-safeguard-and-diversify-recommendations [https://perma.cc/7GZF-598Q].

238. The importance of such efforts also underscores the need for regulation. *See* Renée DiResta et al., *It's Time to Open the Black Box of Social Media*, SCI. AM. (Apr. 28, 2022), https://www.scientificamerican.com/article/its-time-to-open-the-black-box-of-social-media [https://perma.cc/R4WM-GK7W] ("[P]latforms have assured legislators that they are taking steps to counter misinformation and disinformation by flagging content and inserting fact-checks. Are these efforts effective? Again, we would need access to data to know. Without better data, we can't have a substantive discussion about which interventions are most effective and consistent with our values. We also run the risk of creating new laws and regulations that do not adequately address harms or of inadvertently making problems worse.").

239. *The AI Behind Unconnected Content Recommendations on Facebook and Instagram*, *supra* note 63.

240. *The New AI-powered Feature Designed to Improve Feed for Everyone*, META AI (Oct. 5, 2022), https://ai.meta.com/blog/facebook-feed-improvements-ai-show-more-less [https://perma.cc/GE2S-5MTG] (describing how it uses human feedback to generalize about what people want to see).

241. *See* Carroll et al., *supra* note 153, at 8 (discussing how optimizing for long-term engagement can, ironically, increase incentives to manipulate); Perez et al., *Discovering Language Model Behaviors with Model-Written Evaluations*, ARXIV at 7–8 (2022) (discussing how reinforcement learning with human feedback ("RLHF") can increase polarization and pursuit of subgoals).

242. Metz, *supra* note 35.

up the precision and control over search results, as well as the ability to fix problems that came up.[243] With traditional non-AI algorithms, when a platform wanted to change behavior regarding specific types of content, engineers could add new rules to the code to tell the algorithm what to do in each situation.[244] With AI, however, it is much more difficult to change behavior in precise ways.[245] Researchers refer to this as the "CACE principle: Changing Anything Changes Everything," because "machine learning models . . . mak[e] the isolation of improvements effectively impossible" and "[n]o inputs are ever really independent."[246] Faced with this problem, Google decided that, though AI "technologies sacrifice some control, . . . the benefits outweigh that sacrifice."[247] In the end, each of the platforms has, like Google, chosen to shift the foundation of their content presentation over to AI because it is good for their business.[248]

While social media platforms have devoted significant resources to training their AI to achieve desired results, it nonetheless functions more like an independent partner than a tool.[249] Instead of telling the machine what actions to take, with AI, one simply gives it an end goal and it figures out for itself the best way to get there.[250] But AI does not take the same path a

---

243. *See id.*

244. *See, e.g.*, *id.* ("Google's search engine was always driven by algorithms that automatically generate a response to each query. But these algorithms amounted to a set of definite rules. Google engineers could readily change and refine these rules. And unlike neural nets, these algorithms didn't learn on their own.").

245. *Id.* ("The concern—-as described by some former Google employees—-was that it was more difficult to understand why neural nets behaved the way it did, and more difficult to tweak their behavior.").

246. D. Sculley et al., *Machine Learning: The High-Interest Credit Card of Technical Debt*, SOFTWARE ENG'G FOR MACH. LEARNING, at 2 (2014), https://storage.googleapis.com/pub-tools-public-publication-data/pdf/43146.pdf [https://perma.cc/7VZF-KY7Z].

247. Metz, *supra* note 35.

248. *See* Meserole, *supra* note 8 ("For large social networks and digital platforms, the performance gains of deep learning recommendation algorithms far outweigh both the cost of developing them and the corresponding decline in interpretability.").

249. *See* KISSINGER ET AL., *supra* note 20, at 20; CHAN ET AL., *supra* note 234 (highlighting "the increasingly evident fact that ML systems are not fully under human control").

250. *See* Knight, *supra* note 147 ("Instead of a programmer writing the commands to solve a problem, the program generates its own algorithm based on example data and a desired output. . . . [T]he machine essentially programs itself."); *e.g.*, Willis, *supra* note 163, at 128–29 (describing how "algorithmic marketing"

human would.[251] This allows it to achieve super-human performance but also produces unpredictable results.[252] The numerous stories of harmful content being recommended by AI systems[253]—against the policies and despite the intentions of the platforms—demonstrate how the platforms are, to a large extent, simply unable to effectively control the behavior of their systems.[254]

2. The Impact of Delegation on First Amendment Values

The delegation of editorial decision-making to AI both distorts public debate and decreases a platform's autonomy over its message, and thus is not entitled to the same First Amendment protection as human decision-making. First, the delegation of editorial decision-making to AI interferes with the uninhibited public debate necessary for both the operation of the democratic process and individual self-determination. Far from being uninhibited, public discourse on social media is largely curated by AI as it selects each piece of content a user sees on the platform.[255] Dissemination of information is heavily

---

enables marketers to simply specify a business goal and let the AI determine how best to interact with customers).

251. *See* KISSINGER ET AL., *supra* note 20, at 11.

252. *See* Kaye, *supra* note 220 ("This lack of predictability holds the true promise of AI as a transformational technology, but it also illuminates its risks: as humans are progressively excluded from defining the objectives and outputs of an AI system, ensuring transparency, accountability and access to effective remedy becomes more challenging, as does foreseeing and mitigating adverse human rights impacts."); Schneier, *supra* note 12 (observing that because "AIs don't solve problems in the same way people do, they will invariably stumble on solutions we humans might never have anticipated—and some will subvert the intent of the system.").

253. *See, e.g.*, Guillaume Chaslot, *The Toxic Potential of YouTube's Feedback Loop*, WIRED (July 13, 2019), https://www.wired.com/story/the-toxic-potential-of-youtubes-feedback-loop [https://perma.cc/RZ6U-AQKQ].

254. *See* Williams & Chowdhury, *supra* note 236; Llansó et al., Artificial Intelligence, Content Moderation, and Freedom of Expression 16 (Feb. 26, 2020) (unnumbered working paper) (on file with the University of Amsterdam Institute for Information Law), https://www.ivir.nl/publicaties/download/AI-Llanso-Van-Hoboken-Feb-2020.pdf [https://perma.cc/8NXN-NU9U] ("[A]mplification of harmful content through recommendation systems is not necessarily fully intentional or expected on the part of the platform; it may occur without the platform's full knowledge, intent and control, since these systems operate in complex and dynamic networks of multiple invisible actors and incentives.").

255. *See* Kaye, *supra* note 220; Schneier, *supra* note 12 ("[AI will] artificially influence what we think is normal, what we think others think. This sort of manipulation is not what we think of when we laud the marketplace of ideas, or

structured by the opaque decisions of an AI trying to maximize its own, internal goals. The platforms have delegated the very structure of the marketplace of ideas to AI.[256] As mentioned above, this AI decision-making has extensive corollary effects on public discourse well beyond the intentions of the platforms.[257] If the value of uninhibited public debate is to be preserved, maintaining effective structures of communication is essential.[258] While AI could certainly play an important role in achieving such structures, it is entirely dependent on how that AI is designed and regulated.[259]

Of course, the information space has always been curated and filtered in various ways by those in control of key media sources. One could argue that curation by AI is no different than that exercised by the editors running Fox News or the *New York Times*. But, as explained in section IV.A above, AI recommendation systems perform tasks in a manner fundamentally different from humans, and they do so by exploiting individual vulnerabilities in ways that no traditional editor is capable of. As social media platforms have already demonstrated, the effects are both far-reaching and difficult to predict.[260] Moreover, for editorial decisions made by humans, the autonomy of the speaker to exercise reason and judgment to contribute to public discourse—"the principle that each person should decide for [themselves] the ideas and beliefs deserving of expression"[261]—is a value the First Amendment seeks to

---

any democratic political process."). Of course, as noted earlier, the platforms are free to hand-pick content if they want and might do so in certain cases. This argument is based on the general principle, as presented by the platforms themselves, that content presentation is generally done with AI models.

256. *See* Meserole, *supra* note 8 ("If accurate and reliable information is the lifeblood of democracy, recommender systems increasingly serve as its heart.").

257. *See supra* notes 197–198 and accompanying text.

258. *See* Post, *supra* note 68, at 282 ("[T]he notion that [democratic] self-determination requires the maintenance of a structure of communication open to all commands a wide consensus.").

259. Indeed, some scholars have suggested that AI speech could well serve the interests of listeners under First Amendment theory. *See, e.g.*, Massaro & Norton, *supra* note 68; Toni M. Massaro et. al., *Siri-ously 2.0: What Artificial Intelligence Reveals About the First Amendment*, 101 MINN. L. REV. 2481 (2017).

260. *See* KISSINGER ET AL., *supra* note 20, at 21 ("In [the information] space . . . AI sometimes operates in ways even its designers can only elaborate in general terms. As a result, the prospects for free society, even free will, may be altered. Even if these evolutions prove to be benign or reversible, it is incumbent on societies across the globe to understand these changes so they can reconcile them with their values, structures, and social contracts.").

261. Turner Broad. Sys., Inc. v. F.C.C., 512 U.S. 622, 641–42 (1994).

protect. In contrast, the First Amendment value in a speaker's interest in decisions made by AI is significantly diminished.

Compared to a human decision-making process, delegation to AI decreases the autonomy a platform exercises over its message and, consequently, the value of First Amendment protection. When it employs AI to perform editorial decision-making, the platform is no longer using its own reason or judgment to make editorial decisions—many of the decisions are invisible to it, instead made by AI.[262] The protection of that which "reason tells the publisher to do"—given to the newspapers in *Tornillo* and *Associated Press*—does not apply.

Similarly, the foundational principle of speaker autonomy, that "each person should decide for himself or herself the ideas and beliefs deserving of expression,"[263] is not implicated because the platform is not actually making the decisions as to what deserves expression. Indeed, AI continually makes decisions that the platforms expressly do not think are deserving of expression, thus forcing them to fix the problems that arise.[264]

Professor Fallon's definition of autonomy requires the ability to "deliberate rationally, and act consistently with one's goals."[265] In this case, no human is actually deliberating about how to handle a given piece of content; that task falls to AI. In the words of the Court in *Cohen v. California*, "the decision as to what views shall be voiced"[266] is given to AI instead of "each of us." Indeed, the delegation of decision-making to AI renders the platforms unable to carry out their own editorial goals because they do not have sufficient control over its decision-making process.[267] The platforms have given up a significant portion of

---

262. *See* Kaye, *supra* note 220 ("AI shapes the world of information in a way that is opaque to the user and often even to the platform doing the curation.").

263. *Turner Broad. Sys.*, 512 U.S. at 641–42.

264. *See* Nathalie Maréchal & Ellery Roberts Biddle, *Algorithmic Transparency: Peeking into the Black Box, Ranking Digital Rights*, NEW AM. (Mar. 17, 2020), https://www.newamerica.org/oti/reports/its-not-just-content-its-business-model/algorithmic-transparency-peeking-into-the-black-box [https://perma.cc/25VD-UT22].

265. Fallon, *supra* note 81, at 877.

266. Cohen v. California, 403 U.S. 15, 24 (1971).

267. *See* Shannon Bond, *Facebook, YouTube Warn of More Mistakes as Machines Replace Moderators*, NPR (Mar. 31, 2020), https://www.npr.org/2020/03/31/820174744/facebook-youtube-warn-of-more-mistakes-as-machines-replace-moderators [https://perma.cc/W5RL-RAQ7].

their "autonomy to control one's own speech"[268] by delegating that task to AI.

It is true that delegation of editorial control to other humans does not decrease First Amendment protection.[269] However, delegation to a human is fundamentally different from delegation to AI. Delegation to a human entails a whole host of underlying assumptions and guarantees, whether implicit or explicit (e.g., contractual), that cannot be relied upon for AI.[270] AI does not operate like a human, it does not have the same constraints inherent in being human, and delegation of decision-making to it without human oversight raises significant concerns.[271] Thus, while social media platforms can still assert an ascriptive right to autonomy free from government intervention, the intentional relinquishment of a substantial amount of descriptive autonomy renders this interest significantly less compelling.

In sum, the use of AI for editorial decision-making is not entitled to the same protection as a human editor because manipulating public discourse and delegating our foundational communication structures to AI does not serve the values underlying the First Amendment.

CONCLUSION

The above discussion illustrates how the use of AI for editorial decision-making has significant potential to harm our foundational First Amendment values. Of course, consideration of a technology's unique characteristics does not mean abdicating the constitutional duty to uphold the freedom of speech. It means thoughtfully and carefully protecting those

268. Hurley v. Irish-Am. Gay, Lesbian & Bisexual Grp. of Bos., 515 U.S. 557, 574 (1995).

269. *See* Turner Broad. Sys., Inc. v. F.C.C., 512 U.S. 622, 636 (1994) (finding cable operators are entitled to First Amendment protection when exercising discretion over which stations to carry, even though the stations then make the actual programming decisions).

270. *See* Schneier, *supra* note 12 ("[N]o humans maximize their own interests without constraint. Even sociopaths are constrained by the complexities of society and their own contradictory impulses. They're concerned about their reputation, or punishment."); KISSINGER ET AL., *supra* note 20, at 79 ("AI cannot reflect; the significance of its actions is up to humans to decide. Humans, therefore, must regulate and monitor the technology.").

271. *See* Schneier, *supra* note 12 ("AIs don't solve problems like humans do. Their limitations are different than ours.").

values. Indeed, the capabilities and impacts of AI outlined in this Note highlight the potential for government and private misuse alike. For instance, the recent state laws that single out political speech[272] or hate speech[273] for regulation raise serious constitutional questions due to their focus on restricting certain types of content. But regulations aimed at addressing the serious present and future harms from manipulation by AI deserve full consideration in light of their impact on the values underlying the First Amendment.

Imposing transparency requirements around the operation of AI on social media is "perhaps the most widely supported policy priority"—such requirements are generally viewed as necessary for any effective regulation to occur.[274] Many recommendations emphasize the importance of allowing independent researchers access to data to help us understand how AI actually operates.[275] Indeed, a fundamental problem in the effective regulation of AI is that we do not understand how AI works due to the lack of transparency and companies' refusal to allow access for independent research.[276] Effective governance of a complex technology requires a deep understanding of the issue, and tailoring regulations to address concrete problems while preserving our fundamental freedoms

---

272. FLA. STAT. § 501.2041(2)(h) (2021).

273. CAL. AB 587 (2022).

274. Llansó et al., *supra* note 254, at 21; *see also* Daphne Keller & Max Levy, *Getting Transparency Right*, LAWFARE (July 11, 2022), https://www.lawfareblog.com/getting-transparency-right [https://perma.cc/4TDD-U43Q]; John Breeden II & Navrina Singh, *Expert Analysis of Dangerous Artificial Intelligences in Government*, NEXTGOV (Nov. 14, 2022), https://www.nextgov.com/emerging-tech/2022/11/expert-analysis-dangerous-artificial-intelligences-government/379690 [https://perma.cc/XDJ7-BFTT] ("The importance of transparency reporting and system assessments cannot be overstated as a critical foundation for AI governance for all organizations. . . . Reporting allows policymakers to start to evaluate different approaches, and potentially opens the door for benchmarking—reporting is the step that gets us to standards that can be enforced.").

275. *E.g.*, DiResta, *supra* note 238 ("[W]e need access to data on the structures of social media, such as platform features and algorithms, so we can better analyze how they shape the spread of information and affect user behavior.").

276. *See* Gideon Lewis-Kraus, *How Harmful Is Social Media?*, NEW YORKER (June 3, 2022), https://www.newyorker.com/culture/annals-of-inquiry/we-know-less-about-social-media-than-we-think [https://perma.cc/6B2K-2DUS] (quoting Dartmouth political scientist Brendan Nyhan saying "[w]e're years into this, and we're still having an uninformed conversation about social media. It's totally wild."); Maréchal & Biddle, *supra* note 264.

will only be possible if policymakers can see how AI actually operates.

Addressing potential transparency mandates imposed on social media, Evelyn Douek and Genevieve Laker explain that:

> [W]e must look to the purpose of the protection of editorial discretion . . . . Such a purposive approach means that to figure out whether transparency mandates are constitutional, we need to first understand when these mandates—and any possible chilling effect they may have on speech—threaten First Amendment values. In other words, do First Amendment values require social media platforms' power over speech to be completely opaque and unbounded? Or are there, perhaps, First Amendment values protected by understanding how platforms curb people's speech that also need to be weighed in the balance?[277]

The discussion in this Note offers a closer examination of how the use of AI in editorial decision-making impacts First Amendment values to better inform an analysis of transparency requirements—or, indeed, any regulations—in the context of social media.

Nothing in this discussion should be read to conclude that AI is inherently harmful. Rather, it is what we make it. It can be extremely valuable in bearing the burden of difficult human tasks[278] and realizing substantial improvements in a wide variety of human endeavors. There is enormous potential for benefits to individuals and society and much to be gained by encouraging responsible development of AI.[279] Nonetheless, examination of foundational First Amendment values shows that the use of AI for editorial decision-making implicates significantly different concerns from human editing and is not deserving of the same protection.

---

277. *See* Evelyn Douek & Genevieve Laker, *Rereading Herbert v. Lando*, KNIGHT FIRST AMEND. INST. BLOG (May 26, 2022), https://knightcolumbia.org/blog/rereading-herbert-v-lando [https://perma.cc/8URL-SSQ2].

278. *See, e.g.*, Casey Newton, *The Trauma Floor: The Secret Lives of Facebook Moderators in America*, THE VERGE (Feb. 25, 2019), https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona [https://perma.cc/PPJ2-8HJG].

279. *See* Kaye, *supra* note 220 ("(AI) technologies may enable broader and quicker sharing of information and ideas globally, a tremendous opportunity for freedom of expression and access to information.").