No. 21-1333

IN THE

SUPREME COURT OF THE UNITED STATES

REYNALDO GONZALEZ, *et al.*,

Petitioners,

v.

GOOGLE LLC,

Respondent.

On Writ of Certiorari
to the United States Court of Appeals
for the Ninth Circuit

**BRIEF OF THE INTEGRITY INSTITUTE AND
ALGOTRANSPARENCY AS *AMICI CURIAE*
IN SUPPORT OF NEITHER PARTY**

Jim Davy
Counsel of Record
Hannah Mullen
ALL RISE TRIAL & APPELLATE
P.O. Box 15216
Philadelphia, PA 19125
(215) 792-3579
jimdavy@allriselaw.org

Counsel for Amici Curiae

TABLE OF CONTENTS

ii

TABLE OF AUTHORITIES

Cases

Other Authorities

v

vi

vii

INTEREST OF *AMICI CURIAE*[1]

Integrity Institute is a nonprofit membership organization comprised of engineers, product managers, researchers, data scientists, operations specialists, policy experts, and others with decades of experience working at technology companies. Integrity Institute members have worked at Facebook (and Instagram and WhatsApp), Twitter, PayPal, YouTube, and almost every other prominent technology company, including Google itself. Members have observed, and often helped build, the architecture of the social internet. Integrity Institute rests on the belief that the internet can help people and society thrive, but that platforms can cause serious harm to individuals and society. Some of those harms give rise to suits against technology companies—indeed, Integrity Institute members have worked for the defendants in many prominent Section 230 cases.

AlgoTransparency is a nonprofit that educates people about algorithms that shape our access to information. Its team has values and a research vision that address the intersection of artificial intelligence, human rights, misinformation, and policy. Guillaume Chaslot, AlgoTransparency's founder, previously worked at Microsoft and Google, and contributed to YouTube's recommendation algorithm. AlgoTransparency formed with the mission to educate the public about how recommendation algorithms show people content across platforms.

Amici and their members have unique expertise in design practices, responsible platform construction, and the effects that internal platform decisions have on individuals, institutions, and society. *Amici* have an interest in ensuring that the Court has a comprehensive and accurate understanding of how algorithms operate. They believe that independent, honest explanations of technology are valuable to non-experts making law. *Amici* offer their expertise and take no position on the outcome of this case.

SUMMARY OF ARGUMENT

In the certiorari-stage briefing, the Parties repeatedly discuss algorithms, or "algorithm-based recommendations," without offering any meaningful explanation of how they work. *See, e.g.*, Cert. Pet. at 6, 9, 10, 15, 18, 19; Br. in Opp. at 4, 5, 8, 15, 17, 19, 20, 22. Petitioner asks whether Section 230 immunity applies to "recommendations … implemented through automated algorithms," Cert. Pet. at 3; Respondent characterizes the issue as whether that immunity "should apply to algorithms that display content writ large." Br. in Opp. at 2.

But the Parties' imprecise discussions of the relationship between Section 230 immunity and platforms' use of algorithms risk leading this Court astray. Technology companies and platforms use algorithms in a variety of ways across different contexts, and different uses of recommender algorithms cause different harms. Some recommender algorithms help platforms break laws, often in service of increasing profit. Other recommender algorithms allow platforms simply to function. As decisions of the Courts of Appeals demonstrate, the key inquiry when assessing Section 230 immunity is not whether an

algorithm is involved. Rather, it is whether alleged illegality comes from the platform's own conduct or from underlying third-party content. That fact-intensive inquiry should be conducted on a case-by-case basis, in reference to specific allegations about a platform's conduct and the role of a particular algorithm. In deciding this case, *Amici* urge this Court to carefully distinguish algorithms based upon the nature of the recommendations they make, and to ground its opinion in specific factual allegations. While *Amici* take no position on the outcome here, they offer clarity and nuance about the underlying technology.

ARGUMENT

I. Large technology companies use recommendation algorithms differently depending on context.

The mere presence of an "algorithm" in Google's alleged conduct does not resolve the legal issue here. Algorithms are used in all computer software. Recommender systems, colloquially called "algorithms," undergird all online experiences. There is no single "algorithm" for a platform: Instead, multiple pieces of software (called machine learning models) work together to determine how any user sees different content. They serve different functions and, depending on their use and implementation, give rise to different benefits and harms. Algorithms allow platforms to offer a functional user experience and moderate some undesirable content; they also allow platforms to make decisions that may violate statutes and ordinances across fields of law. To avoid a sweeping result that inadvertently reshapes how people experience the internet, this Court should

distinguish the content recommendation algorithms at issue in this case from other types of algorithms that are not before it.

A. Recommendation algorithms are a core component of online platforms.

This case is about recommendation algorithms, as distinct from other sorts of algorithms used by platforms and technology companies. These algorithms play a significant role in users' experiences on these platforms.[2] On one hand, recommendation algorithms make the user experience functional and enjoyable. On the other hand, they can facilitate serious harms to users and societies, including through the promotion, demotion, or removal of content. Recommendation algorithms should be treated carefully, on their own terms.

1. Billions of people use the most popular online platforms.

Online platforms are a prominent part of everyday life for most Americans.[3] Defined at a high level, platforms are internet enterprises that facilitate interactions between individuals. Platforms generate

---

[2] YouTube, for example, reports that more than 70% of views of videos on their platform are generated by recommendations. *See* Ashley Rodriguez, *YouTube's recommendation algorithms drive 70% of what we watch*, Quartz (January 13, 2018), https://qz.com/1178125/youtubes-recommendations-drive-70-of-what-we-watch.

[3] Andrew Perrin and Sara Atske, *About three-in-ten U.S. adults say they are 'almost constantly' online*, Pew Research Center (March 21, 2021), https://www.pewresearch.org/fact-tank/2021/03/26/about-three-in-ten-u-s-adults-say-they-are-almost-constantly-online/.

profits for companies that operate them by capturing some of the value generated by those interactions. *See generally* Paul Gowder, The Networked Leviathan (Cambridge University Press, forthcoming). Some platforms, like eBay or Amazon Marketplace, are two-sided marketplaces that take commissions on economic transactions between users. Others, like Facebook, Instagram, YouTube, and TikTok, are social media platforms that profit from advertising revenue and the use of data generated by users posting and interacting with content.

Billions of people use the most popular online platforms. Facebook recently reported 2.93 billion monthly active users, with 1.97 billion average active users every day.[4] YouTube[5] and Instagram[6] both boast over 2 billion monthly active users, and TikTok has around 1 billion monthly active users.[7]

---

[4] *Meta Reports Second Quarter 2022 Results*, Facebook (July 27, 2022), https://s21.q4cdn.com/399680738/files/doc_news/Meta-Reports-Second-Quarter-2022-Results-2022.pdf.

[5] Jacob Carpenter, *YouTube has always been an afterthought in the streaming wars. Here's how it's finally changing tactics*, Fortune (Nov. 1, 2022), https://fortune.com/2022/11/01/youtube-paramount-starz-primetime-channels-streaming-television-movies/.

[6] Alex Barinka, *Meta's Instagram Users Reach 2 Billion, Closing in on Facebook*, Bloomberg (Oct. 26, 2022), https://www.bloomberg.com/news/articles/2022-10-26/meta-s-instagram-users-reach-2-billion-closing-in-on-facebook?leadSource=uverify%20wall.

[7] Jessica Bursztynsky, *Tiktok says 1 billion people use the app each month*, CNBC (Sept. 27, 2021), https://www.cnbc.com/2021/09/27/tiktok-reaches-1-billion-monthly-users.html.

Unsurprisingly, billions of users generate immense amounts of content—sharing photos, uploading videos, offering goods for sale, and otherwise engaging with platforms. The information flood poses a challenge for platforms: How can platforms create a functional experience for users who prefer to see content that interests them and avoid deluging those users with irrelevant, inappropriate, or otherwise undesirable content? Without a mechanism for ordering content, online platforms would be unusable.

Imagine, for example, a college student who plans to visit her grandfather over winter break. When she types a search query into Google for "weather in Fort Lauderdale, Florida in December," she hopes to see reputable sources—like local weather forecasts—describing the climate. But without a scheme for ordering those results, she would receive an overwhelming mass of content, most irrelevant. She might shut her laptop in frustration, unable to find useful answers.

Similarly, most social media users follow too many accounts to view all those accounts' content. Unless online platforms rank content in order of the users' anticipated interest level, users will miss out on relevant content they would enjoy.

> **2. Online platforms rely on recommendation algorithms to build a functional and enjoyable user experience.**

Platforms address the volume problem by relying on recommendation systems—algorithms that sort and rank content by anticipated user interest. These systems facilitate the healthy and sustainable

functioning of online platforms. Without them, there is simply too much content for users to have an enjoyable experience.

Though details differ by platform, recommendation systems share a common purpose: to gather content, score that content, and produce a final ranked list of content to display to users. Platforms choose which objectives to prioritize in their algorithms, and the recommendation system serves those objectives. Most companies want to achieve maximum user engagement with content, which can include users clicking, sharing, commenting, or otherwise reacting to it. All those actions correlate with users spending more time on the platform and, accordingly, users spending more time viewing advertising.[8]

Recommendation systems also operate similarly across platforms.[9] For each user, the following process

---

[8] YouTube's algorithm, for example, was long optimized to maximize watch time, as the company explained publicly in 2012. *See* Eric Meyerson, *YouTube Now: Why We Focus on Watch Time*, YouTube Official Blog (Aug. 10, 2012), https://blog.youtube/news-and-events/youtube-now-why-we-focus-on-watch-time/.
Similarly, TikTok's algorithm "learns a viewer's tastes with every second they watch, pause or scroll," which informs its functioning. Drew Harwell, *How TikTok ate the internet*, Wash. Post (Oct. 14, 2022), https://www.washingtonpost.com/technology/interactive/2022/tiktok-popularity/.

[9] *See Ranking and Design Transparency: Data, Datasets, and Reports to Track Responsible Algorithmic and Platform Design*, Integrity Institute (Sept. 28, 2021), https://static1.squarespace.com/static/614cbb3258c5c870264975 77/t/617834ea6ee73c074427e415/1635267819444/Ranking+and+

unfolds. First, the system gathers all relevant content on the platform that the user could see. This is called the "inventory." That content could be posts, videos, ads, or even the entire internet, depending on the platform. Each piece of content is a "candidate" for placement before that user. Then, for each candidate piece of content, the system analyzes data about the candidate, the user, and interactions between the user and other relevant actors (like the candidate content's producer). Take, for example, a user we'll call John, who retweets a post by the Chicago Cubs.[10] The system must decide at which position (if any) his friend Steve would see the post: at the top of his page— or buried 200 posts down? The recommendation system considers whether Steve has liked, retweeted, or viewed posts by the Cubs before; if Steve has interacted with John's posts; whether users similar to Steve have liked, retweeted, or viewed the Cubs' content; whether Steve has liked, retweeted, or viewed content similar to the Cubs' post (perhaps posts about baseball or by other teams); and whether the content includes external validation from other sources.

With that information, a large group of algorithms will predict two types of things: various outcomes and various qualities about the content. A model of the

_____

Design+Transparency+%28EXTERNAL%29.pdf; *see also* Naomi Shiffman, *Feed Ranking and Social Harms: A Trustworthy AI Problem*, George Wash. Univ. (Oct. 31, 2022), https://static1.squarespace.com/static/614cbb3258c5c870264975 77/t/638911f4f3cd084e67a5a4ff/1669927418263/Trustworthy+AI _+Ranked+Feeds.pdf.

[10] *See, e.g.*, Chicago Cubs (@cubs), Twitter (Oct. 5, 2022, 8:00 PM), https://twitter.com/Cubs/status/1577810968595533824.

first type might forecast the likelihood of John liking certain images, resharing other posts, and clicking on ads for various products. A model of the second type might analyze whether the content involved is harmful, as defined by law or platform policy, or high quality, according to platform standards.

Lastly, the system aggregates these probabilities and places weights on how important each probability might be. (For instance, if Steve also retweets John's post, the system might value that more than if he merely lingers on it.) By combining these weights and probabilities, a platform's recommendation system generates a final ranking of content to show Steve, optimized towards the highest expected value to the company as quantified by business logic. Each user receives this type of individualized experience on the platform, but the specific predictions (and content inventory) vary.

In sum: Every online platform—Twitter, Instagram, TikTok, Amazon Marketplace, countless more—uses recommendation systems to sort mountains of content, setting aside posts that an individual user would find uninteresting or undesirable and presenting the user with content tailored to them. That sorting is necessary to platforms' survival and functionality. Without a recommendation system, users—unwilling to sift through millions of irrelevant posts to find ones they enjoy—would spend increasingly less time on a platform, eventually abandoning it.

B. The different ways that online platforms use recommendation algorithms cause a range of harms and benefits.

Platforms' different uses of recommendation algorithms give rise to different harms and benefits for users. While platforms' purposes sometimes overlap, this Court should distinguish between the different uses of recommendation algorithms—and the different attendant harms that private lawsuits or public enforcement seek to address. Among different types of platform uses, this case addresses only algorithms that platforms employ to recommend content to users, which are discussed first below.

1. Content recommendations

Platforms use recommendation systems to direct users to content—posts, tweets, videos, product listings, among others—that has the highest expected value to the company. In addition to determining the content users see and the order in which they see it, platforms use a version of the information-gathering and ranking processes described above to suggest new-to-the-user accounts or channels to follow, groups to join, or other sources of content. John the Cubs fan, for example, might find more Chicago-related or baseball-related content in his Twitter feed after he retweets the Cubs' post.

Most social media platforms optimize for engagement, meaning that the platform suggests content because the platform predicts that the user

will interact with it at a relatively high rate.[11] So, Twitter could be more likely to recommend that John follow a Cubs fan account with 100 followers than a Cubs fan account with 100,000 followers, if it expects John to engage more with the former.

The same methods can be used to optimize for other characteristics. A recommendation system might optimize for content "quality," for example, by looking to the structural features of the content, like whether it includes a domain name of a "reputable" website, how many links around the internet point to that content, and quality assessments made by a panel of raters using standardized definitions of quality.

The most common recommendation systems currently used by online platforms are extremely personalized, showing users content based upon the users' precise past behavior, plus their expressed and inferred preferences. That means the algorithms don't merely infer broad categories about a user's interests (e.g. "Cubs Fan"), but suggest content that the recommendation algorithms think the user will want to see, tailored to their past behavior on the platform.

Harms caused by content recommendation: Content recommendation algorithms that optimize for engagement cannot effectively prevent a user from consuming violent, graphic, misleading, illegal, or otherwise harmful content if the user actively

---

[11] *See, e.g.*, Gilad Edelman, *How Facebook Could Break Free From the Engagement Trap*, Wired (Nov. 19, 2021), https://www.wired.com/story/jeff-allen-interview-facebook-engagement-trap/.

interacts with that content.[12] Once the user indicates interest in a type of harmful content by viewing it or even minimally engaging with it, recommendation algorithms optimized for engagement will continue to push similar content to them. A negative feedback loop emerges: The more someone engages with harmful content, the more they'll be shown harmful content, with which they will then engage, ad infinitum.[13]

The business models for Facebook, Instagram, YouTube, Twitter, TikTok, and other social platforms depend upon advertising to a large number of active users. Users who spend more time on a platform see more ads, which advertisers pay to place. Users who post more content are more likely to engage their friends and followers, who keeps them on the platform to see ads. As a result, those platforms design their recommendation systems to prioritize content that generates user engagement and ad revenue.[14]

---

[12] *See generally* Jeff Allen, *Sculpting the Future of Social Media through Incentives and Regulation*, Keynote Address, EU Radicalisation Awareness Network (RAN) Policy Support meeting (June 23, 2022), https://static1.squarespace.com/static/614cbb3258c5c870264975 77/t/6389121c9e67f72f4ad34179/1669927456179/Sculpting+the+ Future+of+Social+Media+through+Incentives+and+Regulation. pdf.

[13] Ranking and Design Transparency, Integrity Institute, *supra* note 9.

[14] *See generally* Jeff Allen, *Social Media and the Spread of Harmful Content,* Grand Rounds in the Oakland Department of Psychiatry (June 22, 2022), https://static1.squarespace.com/static/614cbb3258c5c870264975 77/t/6389125816e7a1745103a09c/1669927516632/Social+Media +and+the+Spread+of+Harmful+Content.pdf.

13

According to Facebook's own internal research, as content grows closer to violating an internal policy—such as policies against incitement to violence, self-harm, or misinformation—engagement grows. This is a familiar dynamic; cable news and tabloids also leverage sensational content and headlines to increase viewership.[15] As Mark Zuckerberg has explained, content that risks violating a Facebook policy is "more sensationalist or provocative," and therefore leads more users to view, click, like, and share it—and possibly see additional ads as a result.[16]
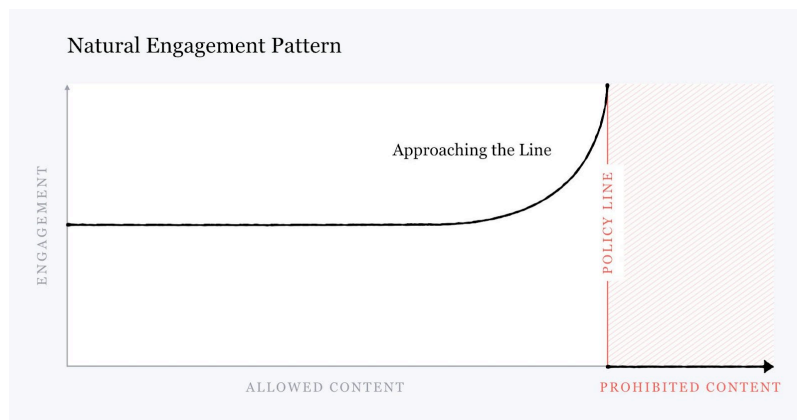
Image 1: Graph showing the relationship between engagement and content that comes close to violating platform policies.[17]

---

[15] *See* Mark Zuckerberg, *A Blueprint for Content Governance and Enforcement*, Facebook (last edited May 5, 2021), facebook.com/notes/751449002072082/.

[16] *See id.*

[17] Zuckerberg, *supra* note 15.

This means that, when a social media platform makes the reasonable business decision to show users content predicted to increase engagement, that decision likely pushes more problematic content atop users' feeds. This dynamic exists wherever a platform draws its line on harmful content, regardless of how people self-identify their content interests. As Mark Zuckerberg said, "Our research suggests that no matter where we draw the lines for what is allowed, as a piece of content gets close to that line, people will engage with it more on average—even when they tell us afterwards they don't like the content."[18]

Put another way, optimizing for engagement means that harmful content will rise to the top of recommendation feeds.[19] This happens in part because in their capacity as creators—rather than consumers—of content, users have strong incentives to post content that garners more engagement from other users.[20] This is true for users whose job is

---

[18] Zuckerberg, *supra* note 15.

[19] Sahar Massachi, *Ranked Feeds and the Harms They Pose*, Institute for Strategic Dialogue's Digital Policy Lab Working Group on "Algorithmic recommendation systems," Keynote Address (July 6, 2022), https://static1.squarespace.com/static/614cbb3258c5c870264975 77/t/638912c1bf77317a3162865c/1669927632600/Ranked+Feeds +and+the+Harms+They+Pose.pdf.

[20] *Cf.* Steven Lee Myers, *How Social Media Amplifies Misinformation More Than Information*, N.Y. Times (Oct. 13, 2022), https://www.nytimes.com/2022/10/13/technology/misinformation -integrity-institute-report.html; Cristiano Lima, *Facebook's most popular posts show it's vulnerable to exploitation, report finds*, Wash. Post (Mar. 30, 2022),

content creation, and for users who simply feel happier when more people like their posts. But given the incentives, no matter how a platform responds, users will work to evade any specific restriction. In fact, the more a platform prevents harmful content from spreading, the greater the engagement-related rewards for content-posters who can bypass the barricade. A platform's choice to optimize for engagement inevitably rewards bad behavior.

Benefits caused by content recommendation: Despite incentivizing harmful content, content recommendation isn't all bad. In fact, recommendation algorithms are ubiquitous across platforms because they are essential to a positive user experience. Recall the college student seeking information about the weather in Fort Lauderdale. Recommenders ensure that she receives worthwhile search results by pushing useful content—like well-regarded weather forecasts—to the top of her search results. Critically, the recommender pushing search results to her is not optimized for engagement, but for quality and relevance.

Recommendation algorithms create social benefits, too. A dating app like Tinder might decide that, rather than suggesting romantic matches based upon location, it will suggest matches based upon expressed interests and dating preferences. Such a choice would make users more likely to encounter people with whom they are compatible and less likely

---

https://www.washingtonpost.com/politics/2022/03/30/facebooks-most-popular-posts-show-how-its-vulnerable-exploitation-report-finds/.

to experience harassment.[21] Again, in this hypothetical, the recommender has an objective—lack of harassment—that is *not* engagement.

Recommenders also improve the user experience in more subtle ways unrelated to engagement optimization. Web browsers often auto-fill URLs for users based on their previous browsing history, so that the user doesn't have to manually type the URL every time they visit the same website.[22] That is a "content recommendation," albeit in a completely different context than social media, and straightforwardly benign.

2. Content moderation and safety

Platforms also use the underlying technology of recommendation algorithms to flag, remove, and re-rank content likely to violate platform policies or laws. Common types of undesirable content include spam, pornography, misinformation, and content that incites violence or promotes self-harm. Content moderation ensures that users have an enjoyable experience on a platform. After all, if John the Cubs fan was barraged with gory videos when he scrolled his Twitter feed hoping to see the score of the latest baseball game, he would probably spend much less time on Twitter.

To perform content moderation at scale, platforms use the same machine learning processes that they use

---

[21] Steve Liu, *Personalized User Recommendations at Tinder*, MLconf (Nov. 10, 2017), https://mlconf.com/sessions/personalized-user-recommendations-at-tinder-the-t/.

[22] *See, e.g.*, *Address bar autocomplete suggestions in Firefox*, Support: Mozilla, https://support.mozilla.org/en-US/kb/address-bar-autocomplete-firefox.

to optimize for other metrics. First, they collect all posts across the platform and analyze their features. The analysis focuses on information like whether a post was created by a user who has previously violated a platform's content policies, whether its poster is behaving in likely harmful ways (e.g. spamming), and whether the post has similarities to previously-identified harmful content.[23] Then, machine learning models use that information to estimate the likelihood that the content violates a platform policy. Posts that are deemed likely to violate policies are subject to platform intervention: The posts may be pushed down in all ranked feeds so fewer people see them, they may be removed, the platform may add a warning screen or a disclaimer to the posts, or the platform may impose another interstitial measure before users can view them.

This process is probabilistic and therefore imperfect; algorithms inevitably generate false negatives (harmful posts misclassified as benign) and false positives (benign posts misclassified as harmful). As a result, content moderation is a tradeoff between precision and accuracy.[24] Precision refers to the

---

[23] This is the Actor Behavior Content framework most associated with Camille Francois. *See* Camille Francois, *Actors, Behaviors, and Content: A Disinformation ABC*, U.S. House of Representatives Committee on Science, Space and Technology (Sept. 20, 2019), https://science.house.gov/imo/media/doc/Francois%20Addendum%20to%20Testimony%20-%20ABC_Framework_2019_Sept_2019.pdf.

[24] *See* Nafia Chowdhury, *Automated Content Moderation: A Primer*, The Program on Platform Regulation, https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/automated_content_moderation_a_primer.pdf.

likelihood that a machine learning model will correctly identify a single piece of content as harmful. Accuracy refers to the likelihood that a machine learning model will catch all pieces of harmful content in a pool. The higher the precision, the lower the accuracy, and vice versa.

Harms and benefits of content moderation: A platform prioritizing accuracy over precision—attempting to remove all harmful content on the platform—would have a high false positive rate. Under that scheme, a platform would remove lots of benign content, resulting in a frustrating user experience. A user's own content might be wrongly removed and they would miss out on some wrongly removed desirable content. That frustration would drive users off the platform. Prioritizing accuracy may also disproportionately remove benign content from marginalized users.[25] Besides driving users off the platform, that might run afoul of non-discrimination laws. *See* Section II, *infra*. And even if not illegal, that disparity might still bother platforms and users.

Platforms typically strike the opposite balance, prioritizing precision over accuracy. Prioritizing precision benefits users because it reduces the likelihood of false positives, allowing most users to post content without it being wrongfully flagged as violating a platform's policies. But insufficient accuracy causes extensive and well-documented harms. In one gruesome example, a man in New

---

[25] *Double Standards in Social Media Content Moderation*, Brennan Center for Justice (Aug. 4, 2021), https://www.brennancenter.org/our-work/research-reports/double-standards-social-media-content-moderation.

Zealand live-streamed himself on Facebook for at least 17 minutes as he violently attacked a mosque.[26] Foreign actors have exploited platforms' precision prioritization to sow misinformation, discord, and distrust in American democracy.[27] And, as in this case, extremist groups exploit under-moderation to post recruitment materials online.[28] In short, harmful content regularly slips into user's feeds.

3. Advertising and commerce

Platforms also rely on recommendation systems to serve targeted ads to users through a technique called "retargeting," which takes advantage of the platform's compilation of information about users' expressed or inferred interests.[29]

---

[26] Donie O'Sullivan, *Facebook says it's policing its platform, but it didn't catch a livestream of a massacre. Why?*, CNN (Mar. 15, 2019), https://www.cnn.com/2019/03/15/tech/facebook-new-zealand-content-moderation.

[27] Sheera Frenkel and Katie Benner, *To Stir Discord in 2016, Russians Turned Most Often to Facebook*, N.Y. Times (Feb. 17, 2018), https://www.nytimes.com/2018/02/17/technology/indictment-russian-tech-facebook.html.

[28] *See* Bennett Clifford, *Moderating Extremism: The State of Online Terrorist Content Removal Policy in the United States*, George Wash. Univ. Program on Extremism (Dec. 2021), https://extremism.gwu.edu/sites/g/files/zaxdzs2191/f/Moderating%20Extremism%20The%20State%20of%20Online%20Terrorist%20Content%20Removal%20Policy%20in%20the%20United%20States.pdf.

[29] *What are retargeting ads? A guide to customer retargeting*, Google Ads, https://ads.google.com/intl/en_uk/home/resources/retargeting-ads/.

Retargeting is why users often see ads for a website they recently visited when they're browsing a social media platform. Put simply, a social media platform—say, Facebook—creates a digital passport for the user, and every time the user visits a website that also buys advertising on Facebook, the user receives a stamp in their digital passport.[30] Facebook inventories a user's passport stamps when deciding which ads to show the user when the user is on the platform. So, if a user starts the day by browsing Target's website for new appliances, they are likely to see Target ads when next on Facebook, assuming Target purchases advertising on Facebook.

In addition to ads informed by retargeting, platforms serve ads based on accounts and content with which users engage, queries users make, and other information. When advertisers submit ads to a platform, they include criteria about the types of users they want to see the ads, like demographics, geography, and predicted interests. In using machine learning processes to determine which ads to show to a user, a platform first collects all possible ads for which the user meets those advertiser criteria. It then narrows that pool by using algorithms that focus on user activity on and off the platform, including clicking on similar ads, installing an app, or product purchasing. When the platform anticipates that users will likely click on a particular ad, it will put the ad in their feed.

Harms and benefits of targeted advertising: Targeted advertising primarily benefits platforms

---

[30] The stamp in the passport is called a "cookie."

because users' increased engagement with advertising directly results in increased revenue. That potential for increased revenue creates incentives for platforms to target ads to users increasingly narrowly and directly. Such targeting requires collecting more and more data about users—some of which are the sort that this Court has recognized can "convey far more than previously possible" about people. *Riley v. California*, 573 U.S. 373, 393 (2014). And targeted advertising that improperly relies on information about users' races, nationalities, religions, sexes, or other information can violate civil rights, consumer protection, and other law.

* * *

Above all, companies have a choice about the objectives for which they optimize. Those choices have consequences for the kinds of outcomes—positive or harmful—they generate for users. And this is all fact-intensive, which drives the analysis as to applying Section 230 immunity in a particular case. Immunity turns on a platform's conduct and how the platform used its specific recommender algorithm.

II. Differences in recommender algorithms affect application of Section 230 immunity—and the Court should decide only this case.

The uses, benefits, and harms of recommender algorithms vary widely. The Courts of Appeals have recognized this nuance and, in applying Section 230, generally distinguish between suits that make allegations about the platform's conduct and suits that ultimately take issue with third-party content. Suits alleging illegality of third-party content attack a platform's role as a publisher and fall within the

heartland of Section 230's statutory immunity. Suits alleging illegality caused by a recommender algorithm through the platform's own conduct, separate from the underlying third-party content, fall outside of Section 230 immunity. Courts of Appeals strike this balance based on the text and purpose of Section 230 and because platforms may be held liable for their own conduct and choices. Extending Section 230's immunity further risks seriously undermining generally applicable laws; withdrawing it too far risks the functioning of platforms and the internet.

Amici take no position in this case for two reasons. First, *Amici*'s expertise best helps the Court understand the nature of the recommender algorithms at play. Second, and more substantively, *Amici* understand the application of recommendation algorithms to content ranking, as here, to present a complicated case for Section 230 application. On one hand, platforms fall well within Section 230 immunity when the alleged illegality comes from the third-party content, regardless of the nature of that content. On the other, they fall well outside Section 230 immunity when the alleged illegality comes from the algorithm itself, regardless of the nature of that content. Allegations about a platform's own choices may fairly give rise to liability. Regardless, *Amici*'s main concern is that a sweeping decision about recommender algorithms generally would affect numerous important uses of those algorithms, including content moderation. Ultimately, *Amici* urge the Court to recognize the potentially far-reaching effects of a decision here across algorithmic uses, and to carefully decide only this case.

A. Section 230 does not immunize platforms' use of algorithms when the harm comes from the algorithm itself, rather than from third-party content.

Platforms' use of recommender algorithms can cause harms regardless of the substance of third-party content posted there, and harms allegedly caused by platforms' own conduct fall outside of Section 230 immunity. Platforms' use of recommender algorithms gives effect to platforms' own choices—business, moral, aesthetic, and others. When those choices cause the alleged harm, platforms may fairly face liability. Illegality from the platforms' uses of recommender algorithms includes where platforms' use of recommender algorithms causes or materially contributes to violations of federal, state, or local laws, including antidiscrimination laws, consumer protection laws, or other civil or criminal statutes. Courts of Appeals, federal agencies, and platforms themselves have recognized that platforms cannot evade liability for their own intentional conduct, and the presence of a recommender algorithm creates no exception to that common-sense rule. This Court should not disturb that key analytical distinction between illegality caused by a platform's own conduct and illegality ultimately based on third-party content.

To provide an example of illegality coming directly from a platform's own use of a recommender algorithm: Some platforms have used algorithms to engage in business practices that violate state or local laws. Courts of Appeals and District Courts consistently refuse to extend Section 230 immunity to platforms that use recommender algorithms for those purposes. In *Airbnb, Inc. v. City & County of San*

Francisco, 217 F. Supp. 3d 1066, 1073 (N.D. Cal. 2016) and *HomeAway.com, Inc. v. City of Santa Monica*, 918 F.3d 676, 683 (9th Cir. 2019), courts allowed online vacation rental platforms to face potential liability for violating local ordinances against unlicensed rentals. In those cases, the platforms' alleged conduct violating the ordinances had nothing to do with the third-party content—the actual rental listings—at all. The platforms made decisions about where, how, and to whom to offer listings, in ways that violated local law. "[P]latforms did not face liability for the content of their listings; rather liability arose from facilitating unlicensed booking transactions." *Dyroff v. Ultimate Software Group, Inc.*, 934 F.3d 1093, 1098 (9th Cir. 2019) (discussing *HomeAway.com*). Put another way, if a platform could modify its own use of recommender algorithms—its own conduct—to comply with applicable law without reference to users' content, the illegality comes from the platform's choices, and Section 230 immunity does not apply.

Platforms' choices in implementing recommender algorithms can violate generally applicable anti-discrimination or consumer protection laws. The Ninth Circuit held that a website that specifically solicited racial preferences from users seeking roommates could not rely on Section 230 immunity in civil-rights lawsuits arising out of that conduct. Running a business that excludes protected groups from a platform's accommodations, facilities, advantages, and services is "something the law prohibits" in its own right. *Fair Housing Council v. Roommates.com*, 521 F.3d 1157, 1167 (9th Cir. 2008) (en banc). The Fourth Circuit recently held Section 230 immunity did not shield platforms from suits alleging violations of the Fair Credit Reporting Act

based upon aggregating and selling user data. *Henderson v. The Source for Public Data, L.P.*, 53 F.4th 110 (4th Cir. 2022). Such claims do "not seek to impose liability on Public Data as a speaker or publisher of any information." *Id.* at 130. The alleged illegality came not from third-party data, but from the platform's choice to sell the data. Relatedly, platforms' intentional use of recommendation algorithms may violate statutes like FCRA or the Equal Credit Opportunity Act and may constitute unfair or deceptive trade practices under the FTC Act.[31]

Other federal agencies have similarly distinguished between harms caused by platform conduct and harms caused by third-party content. The Department of Housing and Urban Development sued Facebook for "provid[ing] a toggle button that enable[d] advertisers to exclude men or women from seeing an ad, a search-box to exclude people who do not speak a specific language from seeing an ad, and a map tool to exclude people who live in a specified area from seeing an ad by drawing a red line around that area." *HUD. v. Facebook*, Charge of Discrimination, FHEO No. 01-18-0323-8 at 4. The Federal Reserve observed that new techniques for ad targeting and steering raise "a range of consumer protection and

---

[31] Elisa Jillson, *Aiming for truth, fairness, and equity in your company's use of AI*, FTC (April 19, 2021), https://www.ftc.gov/news-events/blogs/business-blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai (discussing "targe[ting] consumers most interested in buying their product … by considering race, color, religion, and sex").

financial concerns."[32] By using recommender algorithms to serve ads to users in discriminatory ways, platforms violated the Fair Housing Act and other non-discrimination laws without reference to the content of the advertisements. Indeed, the problem was that women, or people who spoke languages besides English, *wanted* to see the third-party advertisements, and the platform chose not to let them.

Because Section 230 does not immunize a platform when its own actions cause harm, a platform's use of recommender algorithms to incentivize dangerous conduct may lead to liability. For example, platforms may face product liability claims for the design of their recommender algorithms based upon those algorithms' encouragement of particular user conduct and the algorithms' own content creation. *See, e.g.*, *Lemmon v. Snap*, 995 F.3d 1085 (9th Cir. 2021) (allowing potential liability for a platform allegedly inducing teens to drive unsafely). In *Lemmon*, the plaintiffs alleged "an incentive system within Snapchat that encouraged its users to . . . drive at speeds exceeding 100 MPH." *Id.* at 1091-92. The Ninth Circuit held that the platform could fairly face liability for its own design choices. *Id.* at 1094. That case illustrates the key distinction: If the basis of a plaintiff's claim is that a platform should have made a different decision on whether to publish or not publish

---

[32] Carol Evans and Westra Miller, *From Catalogs to Clicks: The Fair Lending Implications of Targeted, Internet Marketing*, Consumer Compliance Outlook: Third Issue, Federal Reserve Board (2019), https://consumercomplianceoutlook.org/2019/third-issue/from-catalogs-to-clicks-the-fair-lending-implications-of-targeted-internet-marketing/.

certain third-party content, or that it should have ranked third-party content lower because that content is illegal, Section 230 immunity attaches, notwithstanding anything about the underlying third-party content. By contrast, if platforms use recommender algorithms to generate independent legally cognizable injuries, Section 230 does not insulate the platform from liability.

Those independently cognizable injuries can take different forms. For example, a platform is not entitled to Section 230 immunity when it uses recommender algorithms to connect minor users with adult users in ways that seriously harm the minors. *A.M. v. Omegle.com, LLC*, No. 3:21-cv-01674, 2022 WL 2713721(D. Or. July 13, 2022). In *A.M.*, the plaintiff alleged that a recommender algorithm connected a child with a sexual predator, causing injuries separate from hosting or publishing the predator's content (unlawful or otherwise). The harm came from "designing a product" that "match[es] minors and adults," and the platform could have changed its own conduct without affecting third-party content at all. *Id.* at *3. Nor does Section 230 immunize a platform from liability when it implements a recommender algorithm to sell defective goods, even if some of the marketing content about those products derived from third parties. *Erie Ins. Co. v. Amazon.com, Inc.*, 925 F.3d 135 (4th Cir. 2019). In those kinds of cases, the illegality and resulting harm comes from a platform's own choices and conduct.

Platforms have occasionally argued that using a recommender algorithm to analyze or process third-party content necessarily falls within Section 230 immunity. But Section 230 "does not provide a general

immunity against all claims derived from third party content." *Doe v. Internet Brands, Inc.*, 824 F.3d 846, 853 (9th Cir. 2016). Such far-reaching immunity would "exceed the scope of the immunity provided by Congress.*" Id.* at 853. Platforms sometimes engage in conduct involving third-party content, but that nonetheless causes harms separate from the third-party content—those, too, fall outside of Section 230. Imagine a platform whose recommender algorithm predicts, based upon the location of rental housing in an ad, that primarily white people have interest in the rental, and so displays the advertisement disproportionately to white users. Or a platform whose algorithm delivers a classified ad for construction jobs only to men, because of predictions based on internal data. Both situations involve ads whose content comes from third parties and the platform's use of a recommender algorithm. But extending Section 230 to the platform simply because its illegal conduct involved using a recommender algorithm threatens the ability of civil rights laws to address discriminatory conduct online. "Providing immunity every time a website uses data initially obtained from third parties would eviscerate [the statute]." *Roommates.com*, 521 F.3d at 1171). Platforms themselves recognize this, too—Facebook, for example, changed some of its algorithmic targeting tools for housing, employment, and credit ads as part of a settlement of civil rights litigation. ACLU, *Summary of Settlements Between Civil Rights Advocates and Facebook* (Mar. 19, 2019), https://www.aclu.org/other/summary-settlements-between-civil-rights-advocates-and-facebook.

**B. Section 230 does immunize platforms'
use of recommender algorithms to
moderate content and protect users.**

By contrast, Section 230 immunity protects
platforms' use of algorithms from suits alleging
illegality rooted in underlying third-party content.
"The prototypical service qualifying for [Section 230]
immunity is an online messaging board (or bulletin
board) on which Internet subscribers post comments
and respond to comments posted by others." *Kimzey v.
Yelp! Inc.*, 836 F.3d 1263, 1266 (9th Cir. 2016) (quoting
FTC v. Accusearch Inc., 570 F.3d 1187, 1195 (10th Cir.
2009)). The internet has evolved far beyond message
boards, but the key consideration remains whether the
alleged liability from which a platform seeks
immunity arises from content created by a third party.
See Dyroff, 934 F.3d at 1097 (9th Cir. 2019) (rejecting
plaintiffs' challenge to underlying dangerous content
that a forum had not actively censored); *see also
Barnes v. Yahoo!, Inc.*, 570 F.3d 1096, 1102 (9th Cir.
2009). If it does, then the platform has immunity. For
example, although platforms are not immune if their
recommender algorithms deliver ads only to male
users, they are immune if the ad itself—third-party
content—offers a job only to men.

Claims against platforms that ultimately depend
upon third-party content take different forms. In one
version, someone aggrieved by a post on a message
board maintained by a platform sues the platform for
the harm caused by the content. That platform has
clear immunity under Section 230. *See, e.g., Kimzey*,
836 F.3d at 1266. Another version inverts the claim:
Someone aggrieved by a post sues the platform for not
taking an offending post down. That platform has

Section 230 immunity, too. *Zeran v. America Online*, 129 F.3d 327, 330 (4th Cir. 1997). Both claims depend not on the recommender algorithm that helped display that content to platform users, but rather, on the underlying content of the posts.

Those outcomes reflect Section 230's clear command that platforms should not face liability for trying to make their spaces hospitable or functional for users. Section 230 immunizes platforms that "voluntarily make good faith efforts to restrict access to material they consider to be objectionable, for example, obscene, excessively violent, or harassing." *Green v. America Online*, 318 F.3d 465, 472 (3d Cir. 2003). Platforms need not engage in content moderation, but if they do moderate content for the benefit of their users, they do not "risk[] liability for doing so." *Id.* Immunity for claims related to content moderation includes immunity for imperfect content moderation. *See Carafano v. Metrosplash.com, Inc.*, 339 F.3d 1119, 1124 (9th Cir. 2003) (applying immunity for platform failure to address identity theft and impersonation via third-party content). Not holding platforms liable when a user publishes defamatory content, and the platform fails to catch it quickly and fails to avert any harm that results, serves an "important purpose … to encourage service providers to self-regulate the dissemination of offensive material over their services." *Zeran*, 129 F.3d at 331. If platforms faced liability for failing to act quickly enough, or comprehensively enough, or as to specific content, platforms would likely decline to engage in moderation at all.

In sum, Section 230 immunity protects platforms that engage in content moderation, imperfect or

otherwise, because lawsuits attacking platforms for engaging in moderation turn on the third-party content itself. Platforms facing suits alleging liability for "deciding whether to publish, withdraw, postpone or alter content" get immunity. *Id.* at 330; *see also Green*, 318 F.3d at 471 (3d Cir. 2003) (immunizing "the monitoring, screening, and deletion of content from its network" as "quintessentially related to a publisher's role"). What *Zeran* and *Green* describe—taking down content that runs afoul of user guidelines, putting up warnings, restricting visibility for younger users or in particular contexts—forms the basis of most content moderation strategies. *See Green*, 318 F.3d at 469 (discussing requirement of users adhering to platform guidelines). Those choices and conduct by platforms, including when implemented by recommender algorithms, get immunity.

C. Suits about content recommendation are intensely fact-bound and the result in this case should be limited accordingly.

Judges on the Courts of Appeals considering whether platforms that use algorithms to make content recommendations—as distinct from algorithms used for other purposes—have reached divergent conclusions about applying Section 230 immunity. Petitioner discussed some of those outcomes in the certiorari petition in this case. *See* Cert. Pet. at 3 (discussing various opinions in *Dyroff*, 934 F.3d at 1093, and *Force v. Facebook, Inc.*, 934 F.3d 53 (2d Cir. 2019)). Judges have taken different views in part because these lawsuits turn on specific, distinct facts and legal claims. Indeed, the outcome of any individual case depends on the details of a particular

complaint, including but not limited to: the choice of statute or ordinance under which to sue; the specific allegations about the content recommendation algorithm, such as the purpose it serves and how it was designed; the alleged role that the platform played in causing the underlying harm, potentially including empirical evidence about the functioning of the platform; and the platform's alleged ability to have prevented that harm through different actions separate from third-party content. Specific allegations may include details from government investigations or external expert research, *see* Section I, *supra*, or could come from increased transparency on the part of platforms themselves,[33] because facts about how these platforms work are not presently well understood by the public. More transparency would make resolution of cases like this one easier for parties and courts alike, and maybe help forestall harms giving rise to lawsuits in the first place. But the point is that these suits vary substantially based on the specifics of each case.

So too here. The correct outcome in this case substantially depends on the allegations in the complaint and facts that might come out in discovery.

---

[33] *See, e.g.*, *Blueprint for an AI Bill of Rights*, The White House Office of Science and Technology Policy (Oct. 2022), https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf; *see also Metrics & Transparency: Data and Datasets to Track Harms, Design, and Process on Social Media Platforms*, Integrity Institute (Sept. 22, 2021), https://static1.squarespace.com/static/614cbb3258c5c870264975 77/t/617834d31bcf2c5ac4c07494/1635267795944/Metrics+and+T ransparency+-+Summary+%28EXTERNAL%29.pdf (describing straightforward ways that platforms could increase transparency, including to consumers and the legal system).

Content recommendation, as one particular use of recommender algorithms, reflects a choice (or a set of choices) by a platform. Many aspects of algorithmic design fall within a platform's control and a legal claim about it might not turn on the third-party content at all. A platform might use internally collected data to determine that users consume more of a particular kind of content, and thus stay on the platform longer and see more advertisements, and so design its algorithm to incentivize such content even if the platform knows that it causes foreseeable harm. This is like a platform making more money by showing advertisements to (for example) only white users, or by asking roommate seekers to express prohibited preferences, even though both violate anti-discrimination law. Alternatively, a platform's choice merely to increase or decrease the ranking or visibility of a specific piece of content—in effect, publishing that content to more or fewer people, or publishing that content higher or lower in users' feeds—would be protected by Section 230 immunity. Moderation depends on recommendation algorithms pushing down potentially dangerous or harmful content, while also highlighting content that may be higher quality in part because it poses less danger to users. Content recommendation algorithms may implement platform choices that violate some law and do not get Section 230 immunity, may implement choices more akin to content moderation that do get immunity, or may do a bit of both. It depends very much on the facts.

Amici ultimately take no position on the facts of this case. But regardless of the instant result, *Amici* urge this Court to ensure that its opinion is limited. The opinion should not hold or suggest that the presence of a recommender algorithm affords blanket

immunity to a platform under Section 230; nor should the opinion hold or suggest that a platform making any decisions at all about content recommendation, including especially content moderation, excludes the platform from the ambit of Section 230 immunity. This Court should let Courts of Appeals continue to answer novel questions about recommendation algorithms as they arise, *see, e.g.*, *Opiotennione v. Bozzuto*, No. 21-1919 (4th Cir.), *pending*; *Vargas v. Facebook*, No. 21-16499 (9th Cir.), *pending*; *In re: Apple Inc. App Store Casino-Style Games Litigation*, No. 5:21-md-2985 (N.D. Cal.), *interlocutory certification pending*, and should continue to allow platforms to moderate content to benefit users without worrying that using an algorithm to do so risks their immunity under Section 230. For over two decades, the internet has grown to play a role in more areas of our lives, and much of it depends on algorithms. *Amici* respectfully request that the Court treads carefully when writing about them.

CONCLUSION

Amici take no position on the ultimate outcome of this case. They offer the Court their expertise on the algorithms used by large platforms, in the hopes of nuanced treatment of those algorithms. *Amici* urge the Court to reinforce the distinction between claims about platforms' own conduct and claims about third-party content, and to underscore both the continued statutory immunity for a platform's moderation of third-party content to benefit users, and the possible liability that a platform can face for violating generally applicable laws by virtue of its own conduct and choices. *Amici* hope their expertise proves helpful.

Respectfully submitted,

Jim Davy
Counsel of Record
Hannah Mullen
ALL RISE TRIAL & APPELLATE
P.O. Box 15216
Philadelphia, PA 19125
(215) 792-3579
jimdavy@allriselaw.org

December 6, 2022   *Counsel for Amici Curiae*